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

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

for isomorphous-replacement data and

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

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

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

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*

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

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

parameters contributing to F_{PH}, including the derivative-to-native scaling parameters. In all cases, the calculated derivative amplitudes $F_{PH_{(\mathrm{calc})}}$ are obtained from equation (25.2.1.9). To use this procedure, one must have an estimate of the protein phase φ. Several variations of this method, all available in *PHASIT*, are described below and are generally referred to as 'phase refinement'.

25.2.1.6.2.1. *'Classical' phase refinement*

With this option, one minimizes

$$\sum_j W_j[F_{PH_{(\mathrm{obs})}} - F_{PH_{(\mathrm{calc})}}(\varphi)]^2 \qquad (25.2.1.13)$$

for isomorphous-replacement data or

$$\sum_j W_j\left\{(F^+_{PH} - F^-_{PH})_{\mathrm{obs}} - [F^+_{PH}(\varphi) - F^-_{PH}(\varphi)]_{\mathrm{calc}}\right\}^2 \qquad (25.2.1.14)$$

for anomalous-scattering data with respect to the desired parameters. Typically, the weights are taken as the reciprocal of the 'standard error' (expected lack of closure) or its square. The summations are taken over all reflections for which the protein phase is thought to be reasonably valid, usually implied by a figure of merit of 0.4 or higher. The protein phase estimate usually comes from the centroid of the appropriate combined phase probability distribution given by equation (25.2.1.10); however, one has the option of including all data sets when combining the distributions, or including all *except* that for the derivative being refined. Once new heavy-atom and scaling parameters are obtained, new individual SIR or SAS phase probability distributions are computed and combined to provide new protein phases, and these phases are used to update the standard error estimates as described earlier. Then the individual distributions are recomputed once more using the new standard error estimates, and these distributions are combined again to give new protein phase estimates. The process is then iterated using the new phases and new heavy-atom parameters to start another round of refinement. After several iterations, the heavy-atom parameters, standard error estimates and protein phase estimates converge to their final values.

25.2.1.6.2.2. *Approximate-likelihood method*

This variation, also available in *PHASIT*, is similar to the classical phase refinement described above, except that instead of using only a single value for the protein phase φ during the calculation of F_{PH}, all possible values are considered, with each contribution weighted by the corresponding protein phase probability (Otwinowski, 1991). One minimizes

$$\sum_j W_j \sum_i P_i[F_{PH_{(\mathrm{obs})}} - F_{PH_{(\mathrm{calc})}}(\varphi_i)]^2 \qquad (25.2.1.15)$$

with respect to the desired parameters for isomorphous-replacement data, where P_i is the protein phase probability and the inner summation is over all allowed protein phase values, stepped in intervals of 5° (or 180° for centric reflections). For anomalous-scattering data, a similar modification is made to equation (25.2.1.14). The weights may be as in the classical phase refinement case or unity. Since each contribution is weighted by its phase probability regardless, there is no need to use a high figure-of-merit cutoff, as was done earlier. In fact, very good results are usually obtained using unit weights for W_j (that is, only the probability weighting) and a figure-of-merit cutoff of around 0.2 for inclusion of reflections in the summations. This variation has been found to increase stability in the refinement and works considerably better than conventional phase refinement when the phase probability distributions are strongly multimodal. Parameter refinement and phasing iterations proceed as described earlier. The combination of probability weighting during refinement with probability weighting during standard error estimation enables the key features of maximum-likelihood refinement to be carried out, although only approximately.

25.2.1.6.2.3. *Using external phase information*

When using either the conventional phase refinement or approximate-likelihood methods, protein phase estimates are required. In the former case, only a single value is used, whereas in the latter, information about all possibilities is provided by way of the phase probability distribution. Normally, this information comes from a prior phasing calculation; thus, the estimates are typically SIR, SAS, MIR *etc.* phases. However, in *PHASIT*, an option allows one to read in the protein phase information from an external source. This enables parameter refinement (by either conventional or approximate-likelihood methods) using protein phase estimates that are improvements over the initial ones. For example, one could get the best phases by one of the previously described methods, but then improve them by density-modification procedures, such as solvent flattening or negative-density truncation and/or NC symmetry averaging. Using these improved phases in the calculation of F_{PH} when refining should then lead to more accurate heavy-atom and scaling parameters, which in turn will produce still better protein phases. These new protein phases can either be treated as final and used to produce an electron-density map for interpretation, or be used to initiate another round of phase improvement by density modification. There are several cases where this type of refinement has been beneficial, and it is particularly useful for the refinement of derivative-to-native scaling parameters.

25.2.1.6.3. *Rigid-group refinement*

Although *GREF* can be used to refine individual heavy-atom or anomalous-scatterer parameters against isomorphous or anomalous structure-factor difference magnitudes, it is actually a group refinement program. Thus, all entities to be refined are treated as rigid bodies such that only group orientations, positions, scaling and temperature parameters can be refined. The groups, however, can be defined arbitrarily. For individual heavy-atom sites, they are simply defined as single atom 'groups', and no orientation parameters are selected for refinement. This enables the program to serve two additional roles. In the case where the heavy-atom reagent is known to contain a rigid group, it can be properly treated. Also, if one chooses the target values to be native structure-factor amplitudes instead of difference magnitudes and inputs an entire protein molecule or domain, then conventional rigid-body or segmented rigid-body refinement can be carried out. The output consists of the refined parameters and a Fourier-coefficient file suitable for map or phase combination calculations.

25.2.1.7. *Origin and hand correlation, and completing the heavy-atom substructure*

Several programs are provided to enable the computation and analysis of various types of difference-Fourier maps as an aid to completing the heavy-atom structure by picking up additional sites. They are also used to correlate the origin and hand between derivatives and to determine the absolute configuration. During phasing calculations in *PHASIT*, files suitable for isomorphous or Bijvoet difference-Fourier calculations are automatically produced for each derivative or data set and can be used directly in program *FSFOUR*. The procedures used are described below.

25.2.1.7.1. *Difference and cross-difference Fourier syntheses*

The files produced by *PHASIT* for isomorphous data sets contain the information needed to produce the Fourier coefficients

references