

## 15.1. PHASE IMPROVEMENT BY ITERATIVE DENSITY MODIFICATION

## 15.1.3. Reciprocal-space interpretation of density modification

Density modification, although mostly performed in real space for ease of application, can be understood in terms of reciprocal-space constraints on structure-factor amplitudes and phases.

Main & Rossmann (1966) showed that the NCS-averaging operation in real space can be expressed in reciprocal space as the convolution of the structure factors and the Fourier transform of the molecular envelope and the NCS matrices. Similarly, the solvent-flattening operation can be considered a multiplication of the map by some mask,  $g_{sf}(\mathbf{x})$ , where  $g_{sf}(\mathbf{x}) = 1$  in the protein region and  $g_{sf}(\mathbf{x}) = 0$  in the solvent region. Thus

$$\rho_{\text{mod}}(\mathbf{x}) = g_{sf}(\mathbf{x}) \times \rho(\mathbf{x}). \quad (15.1.3.1)$$

This assumes that the solvent level is zero, which can be achieved by suitable adjustment of the  $F(000)$  term.

If we transform this equation to reciprocal space, then the product becomes a convolution; thus

$$F_{\text{mod}}(\mathbf{h}) = (1/V) \sum_{\mathbf{k}} G_{sf}(\mathbf{k}) F(\mathbf{h} - \mathbf{k}), \quad (15.1.3.2)$$

where  $G_{sf}(\mathbf{k})$  is the Fourier transform of the mask  $g_{sf}(\mathbf{x})$ . The solvent mask  $g_{sf}(\mathbf{x})$  shows the outline of the molecule with no internal detail, so must be a low-resolution image. Therefore, all but the lowest-resolution terms of  $G_{sf}$  will be negligible.

The convolution expresses the relationship between phases in reciprocal space from the constraint of solvent flatness in real space.

Since only the terms near the origin of  $G_{sf}$  are nonzero, the convolution can only relate phases that are local to each other in reciprocal space. Thus, it can only provide phase information for structure factors near the current phasing resolution limit.

This reasoning may also be applied to other density modifications. Histogram matching applies a nonlinear rescaling to the current density in the protein region. The equivalent multiplier,  $g_{hm}(\mathbf{x})$ , shows variations of about 1.0 that are related to the features in the initial map. The function  $G_{hm}(\mathbf{h})$  for histogram matching is, therefore, dominated by its origin term, but shows significant features to the same resolution as the current map or further, as the density rescaling becomes more nonlinear. Histogram matching can therefore give phase indications to twice the resolution of the initial map or beyond, although phase indications will be weak and contain errors related to the level of error in the initial map.

$$\rho_{\text{mod}}(\mathbf{x}) = g_{ncs}(\mathbf{x}) (1/N_{ncs}) \sum_i \rho_i(\mathbf{x}). \quad (15.1.3.3)$$

Averaging may be described as the summation of a number of reoriented copies of the electron density within the region of the averaging mask (Main & Rossmann, 1966), *i.e.* where  $\rho_i(\mathbf{x})$  is the initial density,  $\rho(\mathbf{x})$ , transformed by the  $i$ th NCS operator and  $g_{ncs}(\mathbf{x})$  is the mask of the molecule to be averaged. This summation is repeated for each copy of the molecule in the whole unit cell. The reciprocal-space averaging function,  $G_{ncs}(\mathbf{h})$ , is the Fourier transform of a mask, as for solvent flattening, but since the mask covers only a single molecule, rather than the molecular density in the whole unit cell, the extent of  $G_{ncs}(\mathbf{h})$  in reciprocal space is greater.

Sayre's equation is already expressed as a convolution, although in this case the function  $G(\mathbf{h})$  is given by the structure factors  $F(\mathbf{h})$  themselves. It is, therefore, the most powerful method for phase extension. However, as resolution decreases, more of the reflections required to form the convolution are missing, and the error increases.

The functions  $g(\mathbf{x})$  and  $G(\mathbf{h})$  for these density modifications are illustrated in Fig. 15.1.3.1 for a simple one-dimensional structure.

## 15.1.4. Phase combination

Phase combination is used to filter the noise in the modified phases and eliminate the incorrect component of the modified phases through a statistical process. The observed structure-factor amplitudes are used to estimate the reliability of the phases after density modification. The estimated probability of the modified phases is combined with the probability of observed phases to produce a more reliable phase estimate,

$$P_{\text{new}}[\varphi(\mathbf{h})] = P_{\text{obs}}[\varphi(\mathbf{h})] P_{\text{mod}}[\varphi(\mathbf{h})]. \quad (15.1.4.1)$$

Once a modified map has been obtained, modified phases and amplitudes may be derived from an inverse Fourier transform. The modified phases are normally combined with the initial phases by multiplication of their probability distributions. The probability distribution for the experimentally observed phases is usually de-

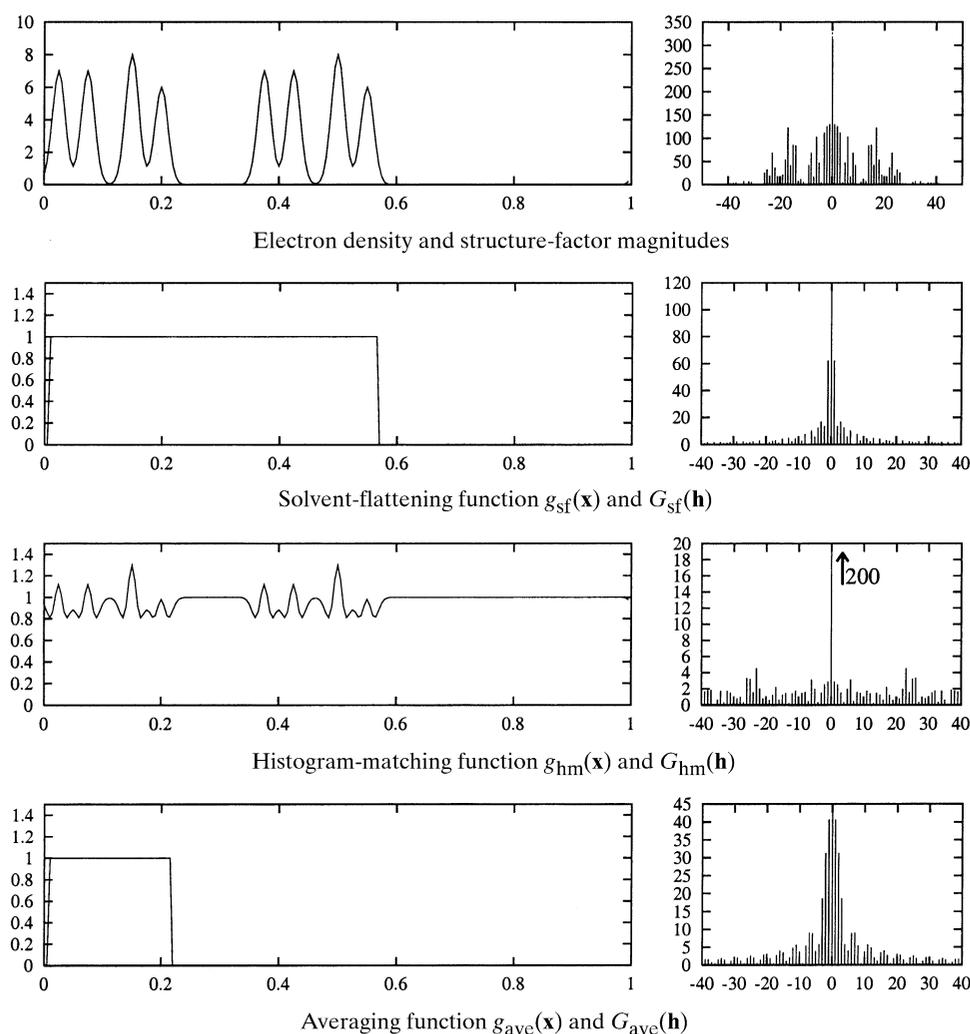


Fig. 15.1.3.1. The functions  $g(\mathbf{x})$  and  $G(\mathbf{h})$  for solvent flattening, histogram matching and averaging.

## 15. DENSITY MODIFICATION AND PHASE COMBINATION

scribed in terms of a best phase and figure of merit (Blow & Rossmann, 1961) or by Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970). In order to estimate a unimodal probability distribution for the modified phase, some estimate of the associated error must be made; this is usually achieved using the Sim weighting scheme (Sim, 1959).

Recombination with the initial phases assumes independence between the initial and modified phases and is a source of difficulties. However, in the absence of some form of phase constraint, most density-modification constraints are too weak to guarantee convergence to a reasonable solution. The exception is when high-order NCS is present; in this case, the combination of NCS and observed amplitudes is sufficient to determine the phases (Chapman *et al.*, 1992; Tsao *et al.*, 1992), and phase combination may be omitted; however, weighting of the phases is still necessary. In this case, it is also possible to restore missing reflections in both amplitude and phase.

### 15.1.4.1. *Sim and $\sigma_a$ weighting*

The phase probability distribution for the density-modified phase is conventionally generated under assumptions that were made for the combination of a partial atomic model with experimental data. It assumes that the calculated amplitudes and phases arise from a density map in which some atoms are present and correctly positioned, and the remainder are completely absent (Sim, 1959). Thus, the difference between the true structure factor and the calculated value must be the effective structure factor due to the missing density alone. If the phase of this quantity is random and the amplitude is drawn from a Wilson distribution (Wilson, 1949), the following expression is obtained:

$$P_{\text{mod}}(\varphi) = \exp[A \cos \varphi + B \sin \varphi], \quad (15.1.4.2)$$

where

$$\begin{aligned} A &= X \cos \varphi_{\text{exp}} \\ B &= X \sin \varphi_{\text{exp}} \end{aligned} \quad (15.1.4.3)$$

and

$$X = 2|F_{\text{exp}}||F_{\text{mod}}|/\Sigma_Q, \quad (15.1.4.4)$$

where  $\Sigma_Q$  is the variance parameter in the Wilson distribution for the missing part of the structure. The figure of merit,  $w$ , can be derived from

$$w = I_1(X)/I_0(X), \quad (15.1.4.5)$$

where  $I_0$  and  $I_1$  are zero- and first-order modified Bessel functions. A similar argument follows for centric reflections.

The error estimate for the phase depends on the effective amount of missing structure that is estimated on the basis of the agreement of the modified amplitudes with their measured values, where  $\Sigma_Q$  may be estimated by a number of means, for example (Bricogne, 1976),

$$\Sigma_Q = \langle |F_{\text{obs}}|^2 - |F_{\text{mod}}|^2 \rangle, \quad (15.1.4.6)$$

where the average is normally taken over all reflections at a particular resolution. A more sophisticated approach is the  $\sigma_a$  method of Read (1986), which allows for errors in the atomic model and has also been used in density modification (Chapter 15.2).

Although these approaches have been applied with some success, the assumption in equation (15.1.4.1) that the density-modified amplitudes and phases are independent of the initial values is invalid. Since the density constraints are typically under-determined, it is possible to achieve an arbitrarily good agreement between the model amplitudes and their observed values without

improving the phases. As a result, phase weights from density modification are typically overestimated.

This problem has traditionally been addressed by limiting the number of cycles of density modification in which weakly phased reflections are included. Typically, density modification is started with only some subset of the data, such as those reflections well phased from MIR data. Only these reflections are included in the phase recombination, with other reflections set to zero. As the calculation progresses, more reflections are introduced until all the data are included. The figures of merit of reflections that undergo fewer cycles of phase recombination will be correspondingly smaller (*e.g.* Leslie, 1987; Zhang & Main, 1990a). In averaging calculations where considerable phase information is available from high-order NCS, it is still typically necessary to perform phase extension over hundreds of cycles and to add a very thin resolution shell of new reflections at each cycle.

The phases and figure of merit generated from density modification are more suited to the calculation of weighted  $F_o$  maps than  $2mF_o - F_c$  maps. The  $2mF_o - F_c$  map is designed to aid the structure completion from a partial model (Main, 1979). The  $2mF_o - F_c$  map will restore features missing from the current model at full weight if the following conditions are fulfilled. First, the model phases must be close to their true values. Secondly, the difference between the model and observed amplitudes is a good indicator of the phase error and the difference between the calculated and observed amplitudes decreases as the phases approach their true values. Neither of these assumptions are necessarily true for density modification, since it may be applied to very poor maps with almost random phases, and under most density-modification schemes the structure-factor amplitudes may be over-fitted to the observed values.

### 15.1.4.2. *Reflection omit*

The modified map may be made more independent of the original map, as was assumed when multiplying the phase probability distributions in equation (15.1.4.1), through a reciprocal-space analogue of the omit map, the reflection-omit method.

The reflections are divided into (typically 10 or 20) sets and density-modification calculations are performed, excluding each set in turn from the calculation of the starting map, in a manner similar to a free- $R$ -value calculation (Brünger, 1992). Density modification is applied to each map in turn, and the modified reflections from each of the free sets are combined to give a new, complete data set. This data set should be less dependent on the original amplitudes; therefore, the amplitudes may be expected to give a better indication of the quality of the modified phases.

The resulting maps obtained using solvent flattening and/or histogram matching are dramatically improved using the reflection-omit method (Cowtan & Main, 1996). In the case of averaging calculations, however, the reflection-omit approach makes little difference, since omitted reflections tend to be restored through noncrystallographic symmetry relationships to other regions of reciprocal space. It is possible that further improvements may be achieved by selecting reflection sets that approximately obey the NCS relationships.

### 15.1.4.3. *The $\gamma$ correction and solvent flipping*

Abrahams & Leslie (1996) have shown that solvent flipping is dramatically more effective as a density modification than solvent flattening. This may be shown to be theoretically equivalent to performing a reflection-omit calculation for each reflection individually (Abrahams, 1997).

Solvent flattening is represented in reciprocal space by convolution of the structure factors with a function,  $G(\mathbf{h})$ , as

## 15.1. PHASE IMPROVEMENT BY ITERATIVE DENSITY MODIFICATION

shown in equation (15.1.3.2). If the origin term of  $G$  is set to zero, then the modified structure factor,  $F_{\text{mod}}(\mathbf{h})$ , will depend on the values of all the structure factors except itself; this is equivalent to performing a reflection-omit calculation with that reflection alone omitted.

Let the origin-removed  $G$  be called  $G_\gamma(\mathbf{h})$  and its Fourier transform  $g_\gamma(\mathbf{x})$ :

$$G_\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h} = 0 \\ G(\mathbf{h}), & \mathbf{h} \neq 0 \end{cases}, \quad (15.1.4.7)$$

then

$$g_\gamma(\mathbf{x}) = g(\mathbf{x}) - \overline{g(\mathbf{x})}. \quad (15.1.4.8)$$

The convolution of the reflection data with  $G_\gamma(\mathbf{h})$  is equivalent to performing a reflection-omit calculation, omitting every reflection in turn. However, the convolution may still be performed in real space; thus, the full omit calculation becomes a simple multiplication of the map by  $g_\gamma(\mathbf{x})$ :

$$\rho_{\text{mod}}(\mathbf{x}) = g_\gamma(\mathbf{x}) \times \rho(\mathbf{x}). \quad (15.1.4.9)$$

In a solvent-flattening calculation,  $g_\gamma(\mathbf{x})$  will be equal to  $g(\mathbf{x})$  minus the fraction of the cell that is protein. In the case of a cell with 50% solvent,  $g_\gamma(\mathbf{x})$  has a value of 0.5 in the protein and  $-0.5$  in the solvent. Multiplication of the map by this function results in flipping of the solvent.

If the origin term of the  $G$  function,  $\gamma$ , can be determined, then the flipping calculation may alternatively be performed by subtracting a copy of the initial map scaled by  $\gamma$  from the modified map. This is the  $\gamma$  correction of Abrahams (1997). This approach may be generalized to arbitrary density-modification methods by use of the perturbation  $\gamma$  (Cowtan, 1999). In this approach, a random perturbation is applied to the starting data. Density modification is applied to both the perturbed and unperturbed maps. The relative size of the perturbation signal in the modified map gives an estimate for  $\gamma$ . The perturbation  $\gamma$  provides effective bias correction for any combination of solvent flattening, histogram matching and averaging.  $\gamma$  may also be estimated as a function of resolution, allowing successful application to multi-resolution modification and possibly atomization as well.

### 15.1.5. Combining constraints for phase improvement

The chemical and physical information of the underlying structure that the electron density represents serves as constraints on the phases. For small molecules, the constraints of positivity and atomicity are sufficient to solve the phase problem *ab initio* (Hauptman, 1986; Karle, 1986; Woolfson, 1987), because crystals of small molecules generally diffract to atomic resolution. However, no single constraint at our disposal is powerful enough to render the macromolecular phase problem determinable, because macromolecule crystals rarely diffract to atomic resolution. Therefore, individual constraints are combined to produce a more powerful density-modification protocol. This is because these constraints represent different characteristic features of the electron density and they contain independent phasing information.

The phasing power of a method increases with the number of independent constraints employed, the number of density points affected and the amplitude of changes imposed on the electron density. It also depends on the physical nature and accuracy of the constraints and how the constraints are applied. One obvious way of implementing several constraints is to apply them one after the other to the electron density. This sequential application, although easy to implement, suffers some drawbacks. The cyclic application of all constraints may not converge easily, since some constraints

may contain contradicting information as to how the density should be modified. An alternative way of implementing various constraints is simultaneous application. The density solution that satisfies all the constraints is obtained by a global minimization procedure (Main, 1990b; Zhang & Main, 1990b).

#### 15.1.5.1. The system of nonlinear constraint equations

The constraints used in *SQUASH/DM* can be divided into three categories. The first category comprises the linear constraints, such as solvent flatness, density histogram and equal molecules. The second category comprises the nonlinear constraints, such as the local shape of electron density as expressed in Sayre's equation. The third category comprises the available structural data, such as the observed structure-factor amplitudes and the experimental phases. The first and second categories of constraints are used to solve new electron-density values. The third category of constraints is used as a means to filter the modified phases.

The modification to the density value at a grid point by a linear constraint is independent of the values at other grid points. These constraints include solvent flattening, histogram matching and molecular averaging. These density-modification methods construct an improved map directly from an initial density map as expressed by

$$\rho(\mathbf{x}) = H(\mathbf{x}), \quad (15.1.5.1)$$

where  $H(\mathbf{x})$  is the target electron density produced by these linear constraints.

The new electron density that satisfies both the linear constraints represented by equation (15.1.5.1) and the nonlinear constraints expressed by Sayre's equation (15.1.2.31) can be obtained by solving the systems of simultaneous equations (Zhang & Main, 1990b)

$$\begin{cases} (V/N) \sum_{\mathbf{y}} \rho^2(\mathbf{y}) \psi(\mathbf{x} - \mathbf{y}) - \rho(\mathbf{x}) = 0 \\ H(\mathbf{x}) - \rho(\mathbf{x}) = 0 \end{cases}. \quad (15.1.5.2)$$

Equation (15.1.5.2) represents a system of nonlinear simultaneous equations with as many unknowns as the number of grid points in the asymmetric unit of the map and with twice as many equations as unknowns. The functions  $H(\mathbf{x})$  and  $\psi(\mathbf{x} - \mathbf{y})$  are both known. The least-squares solution, using either the full matrix or the diagonal approximation, is obtained using the Newton–Raphson technique with fast Fourier transforms, as described in the next section (Main, 1990b).

#### 15.1.5.2. Least-squares solution to the system of nonlinear constraint equations

For a system of nonlinear equations of electron density,

$$\mathbf{F}(\rho(\mathbf{x})) = \mathbf{0}, \quad (15.1.5.3)$$

where

$$\begin{aligned} \mathbf{F}(\rho(\mathbf{x})) &= [F_1(\rho(\mathbf{x})) \ F_2(\rho(\mathbf{x})) \ \dots \ F_m(\rho(\mathbf{x}))]^T, \\ \rho(\mathbf{x}) &= [\rho_1 \ \rho_2 \ \dots \ \rho_n]^T, \end{aligned}$$

$\mathbf{0}$  is a null vector,  $n$  is the number of grid points and  $m$  is the number of equations, the Newton–Raphson method of solution is to find a set of shifts,  $\delta\rho(\mathbf{x})$  to  $\rho(\mathbf{x})$ , through a system of linear equations,

$$\mathbf{J}\delta\rho(\mathbf{x}) = -\varepsilon, \quad (15.1.5.4)$$

where  $\mathbf{J}$  is a matrix of partial derivatives of  $\mathbf{F}$  with respect to  $\rho(\mathbf{x})$  and is called the Jacobian matrix,