shown in equation (15.1.3.2). If the origin term of G is set to zero, then the modified structure factor, $F_{\rm mod}(\mathbf{h})$, 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 $g_\gamma(\mathbf{x})$:

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

then

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

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 $g_\gamma(\mathbf{x})$:

$$\rho_{\rm mod}(\mathbf{x}) = \mathrm{g}_\gamma(\mathbf{x}) \times \rho(\mathbf{x}). \tag{15.1.4.9}$$

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, γ, can be determined, then the flipping calculation may alternatively be performed by subtracting a copy of the initial map scaled by γ from the modified map. This is the γ correction of Abrahams (1997). This approach may be generalized to arbitrary density-modification methods by use of the perturbation γ (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 γ. The perturbation γ provides effective bias correction for any combination of solvent flattening, histogram matching and averaging. γ 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, 1990*b*; Zhang & Main, 1990*b*).

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

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

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, 1990*b*)

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

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, 1990*b*).

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

For a system of nonlinear equations of electron density,

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

where

$$\mathbf{F}(\rho(\mathbf{x})) = [F_1(\rho(\mathbf{x}))\ F_2(\rho(\mathbf{x}))\ \ldots\ F_m(\rho(\mathbf{x}))]^T,$$
$$\rho(\mathbf{x}) = [\rho_1\ \rho_2\ \ldots\ \rho_n]^T,$$

$\mathbf{0}$ 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,

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

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

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial F_1}{\partial \rho_1} & \dfrac{\partial F_1}{\partial \rho_2} & \cdots & \dfrac{\partial F_1}{\partial \rho_n} \\ \dfrac{\partial F_2}{\partial \rho_1} & \dfrac{\partial F_2}{\partial \rho_2} & \cdots & \dfrac{\partial F_2}{\partial \rho_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial \rho_1} & \dfrac{\partial F_m}{\partial \rho_2} & \cdots & \dfrac{\partial F_m}{\partial \rho_n} \end{bmatrix}, \tag{15.1.5.5}$$

ε is a vector of residuals to equation (15.1.5.3) for a trial solution, $\rho(\mathbf{x})$, and $\delta\rho(\mathbf{x})$ is a vector of shifts to the density. Hence, the solution for $\rho(\mathbf{x})$ is achieved in an iterative manner,

$$\rho^{i+1}(\mathbf{x}) = \rho^{i}(\mathbf{x}) + \delta\rho(\mathbf{x}). \tag{15.1.5.6}$$

Therefore, the problem of solving a system of nonlinear equations (15.1.5.3) is transformed into solving a system of linear equations (15.1.5.4), which forms one cycle of Newton–Raphson iteration.

If there are more equations than unknowns ($m > n$), the unknowns are obtained through a least-squares solution to equations (15.1.5.4),

$$\mathbf{J}^T\mathbf{J}\delta\rho(\mathbf{x}) = -\mathbf{J}^T\varepsilon. \tag{15.1.5.7}$$

Theoretically, the above system of equations could be solved by matrix multiplication and inversion, *i.e.*

$$\delta\rho(\mathbf{x}) = -\left(\mathbf{J}^T\mathbf{J}\right)^{-1}\mathbf{J}^T\varepsilon. \tag{15.1.5.8}$$

However, the amount of calculation involved in setting up the normal matrix of least squares is huge for the problem presented by protein structures. This can be completely avoided by using the conjugate-gradient technique for solving the system of linear equations.

15.1.5.2.1. *The conjugate-gradient method*

The conjugate-gradient method does not require the inversion of the normal matrix, and therefore the solution to a large system of linear equations can be achieved very quickly.

Starting from a trial solution to equations (15.1.5.4), such as a null vector,

$$\delta\rho_0(\mathbf{x}) = \mathbf{0}, \tag{15.1.5.9}$$

the initial residual is

$$\mathbf{r}_0 = -\mathbf{J}^T(\varepsilon - \mathbf{J}\delta\rho_0(\mathbf{x})) \tag{15.1.5.10}$$

and the initial search step is

$$\mathbf{p}_0 = \mathbf{r}_0. \tag{15.1.5.11}$$

The iterative process is as follows. The new shift to the density is

$$\delta\rho_{k+1}(\mathbf{x}) = \delta\rho_k(\mathbf{x}) + \alpha_k\mathbf{p}_k, \tag{15.1.5.12}$$

where

$$\alpha_k = \mathbf{r}_k^T\mathbf{p}_k/\mathbf{q}_k^T\mathbf{q}_k \tag{15.1.5.13}$$

and

$$\mathbf{q}_k = \mathbf{J}\mathbf{p}_k. \tag{15.1.5.14}$$

The new residual is

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\mathbf{s}_k, \tag{15.1.5.15}$$

where

$$\mathbf{s}_k = \mathbf{J}^T\mathbf{q}_k. \tag{15.1.5.16}$$

The next search step which conjugates with the residual is

$$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_k\mathbf{p}_k, \tag{15.1.5.17}$$

where

$$\beta_k = -\mathbf{r}_{k+1}^T\mathbf{s}_k/\mathbf{q}_k^T\mathbf{q}_k. \tag{15.1.5.18}$$

The process is iterated by increasing k until convergence is reached, when

$$|\mathbf{r}_{k+1} - \mathbf{r}_k| \Rightarrow 0.$$

The number of iterations required for an exact solution is equal to the number of unknowns, because the search vector at each step is orthogonal with all the previous steps. However, a very satisfactory solution can normally be reached after very few iterations. This makes the conjugate-gradient method a very efficient and fast procedure for solving a system of equations. Note that the normal matrix never appears explicitly, although it is implicit in (15.1.5.10) and (15.1.5.16). The inversion of the normal matrix and matrix multiplication is completely avoided. Most of the calculation comes from the formation of the matrix-vector products in (15.1.5.10), (15.1.5.14), and (15.1.5.16). These can be expressed as convolutions and can be performed using FFTs, thus saving considerably more time.

The solution to $\delta\rho(\mathbf{x})$ at the end of conjugate-gradient iteration is substituted into equation (15.1.5.6) to get a new solution for $\rho(\mathbf{x})$. The solution to the system of nonlinear equations (15.1.5.3) is obtained when the Newton–Raphson iteration has reached convergence.

15.1.5.2.2. *The full-matrix solution*

The equations to be solved for the electron-density shifts, $\delta\rho(\mathbf{x})$, are from the Jacobian of equation (15.1.5.2),

$$\begin{cases} (2V/N)\sum\limits_{\mathbf{y}}\rho(\mathbf{y})\psi(\mathbf{x}-\mathbf{y}) - \delta\rho(\mathbf{x}) = \Delta\rho(\mathbf{x}) \\ \delta\rho(\mathbf{x}) = \Delta H(\mathbf{x}) \end{cases}, \tag{15.1.5.19}$$

where $\Delta\rho(\mathbf{x})$ is the residual to Sayre's equation,

$$\Delta\rho(\mathbf{x}) = \rho(\mathbf{x}) - (V/N)\sum_{\mathbf{y}}\rho^2(\mathbf{y})\psi(\mathbf{x}-\mathbf{y}), \tag{15.1.5.20}$$

and $\Delta H(\mathbf{x})$ is the residual to the linear density-modification equations,

$$\Delta H(\mathbf{x}) = H(\mathbf{x}) - \rho(\mathbf{x}). \tag{15.1.5.21}$$

Starting from a trial solution of $\delta\rho_0(\mathbf{x}) = \mathbf{0}$, the initial residual vector is

$$\mathbf{r}_0(\mathbf{x}) = (2/V)\rho(\mathbf{x})\sum_{\mathbf{h}}\theta\left(\overline{\mathbf{h}}\right)\Delta F(\mathbf{h})\exp(-2\pi i\mathbf{h}\mathbf{x}) - \Delta\rho(\mathbf{x}) + \Delta H(\mathbf{x}), \tag{15.1.5.22}$$

where

$$\Delta F(\mathbf{h}) = F(\mathbf{h}) - \theta(\mathbf{h})G(\mathbf{h}), \tag{15.1.5.23}$$

$$G(\mathbf{h}) = (V/N)\sum_{\mathbf{y}}\rho^2(\mathbf{y})\exp(2\pi i\mathbf{h}\mathbf{y}) \tag{15.1.5.24}$$

and

$$\Delta\rho(\mathbf{x}) = (1/V)\sum_{\mathbf{h}}\Delta F(\mathbf{h})\exp(-2\pi i\mathbf{h}\mathbf{x}). \tag{15.1.5.25}$$

Thus, only three FFTs are required to calculate the initial residual. The residual of Sayre's equation is given in equation (15.1.5.23).

The calculation of $\mathbf{q}_k$ in equation (15.1.5.14) is achieved in a similar manner using FFTs,

$$\mathbf{q}_k = \mathbf{J}\mathbf{p}_k = \left\{\frac{(1/V)\sum_{\mathbf{h}}[2a(\mathbf{h})\theta(\mathbf{h}) - b(\mathbf{h})]\exp(-2\pi i\mathbf{hx})}{p_k(\mathbf{x})}\right\} = \left[\frac{Q_k(\mathbf{x})}{p_k(\mathbf{x})}\right], \quad (15.1.5.26)$$

where the vector is partitioned as shown above, and

$$a(\mathbf{h}) = (V/N)\sum_{\mathbf{y}}\rho(\mathbf{y})p_k(\mathbf{y})\exp(2\pi i\mathbf{hy}), \quad (15.1.5.27)$$

$$b(\mathbf{h}) = (V/N)\sum_{\mathbf{y}} p_k(\mathbf{y})\exp(2\pi i\mathbf{hy}). \quad (15.1.5.28)$$

Similarly, vector $\mathbf{s}_k$ in equation (15.1.5.16) is obtained from

$$\mathbf{s}_k = \mathbf{J}^T\mathbf{q}_k = (2/V)\rho(\mathbf{x})\sum_{\mathbf{h}}\theta(\overline{\mathbf{h}})[2a(\mathbf{h})\theta(\mathbf{h}) - b(\mathbf{h})]\exp(-2\pi i\mathbf{hx}) - Q_k(\mathbf{x}) + p_k(\mathbf{x}), \quad (15.1.5.29)$$

where $Q_k(\mathbf{x})$ is defined in equation (15.1.5.26).

The remaining calculations in equations (15.1.5.12), (15.1.5.13), (15.1.5.15), (15.1.5.17) and (15.1.5.18) require either the inner product of a pair of vectors or a linear combination of vectors, both of which are very quick to calculate. Each iteration of the conjugate gradient requires four FFTs, as described in equations (15.1.5.26–15.1.5.29).

15.1.5.2.3. *The diagonal approximation*

The full-matrix solution to equation (15.1.5.4) requires a significant amount of computing, although it can be achieved using FFTs. The diagonal approximation to the normal matrix has been used as an alternative method of solution to the electron-density shift in equation (15.1.5.4) (Main, 1990*b*). As with the full-matrix calculation, it can be done entirely by FFTs and a linear combination of vectors.

The diagonal element of the normal matrix, $\mathbf{J}^T\mathbf{J}$, in equation (15.1.5.7) is

$$d_0(\mathbf{x}) = (4/N)\rho(\mathbf{x})\left[\rho(\mathbf{x})\sum_{\mathbf{h}}|\theta(\mathbf{h})|^2 - \sum_{\mathbf{h}}\theta(\mathbf{h})\right] + 2. \quad (15.1.5.30)$$

The right-hand side of equation (15.1.5.7), $-\mathbf{J}^T\varepsilon(\mathbf{x})$, is identical to the residual vector, $r_0(\mathbf{x})$, which can be calculated from equation (15.1.5.22). Therefore, the solution to the electron-density shift, $\delta\rho(\mathbf{x})$, can be calculated from

$$\delta\rho(\mathbf{x}) = r_0(\mathbf{x})/d_0(\mathbf{x}). \quad (15.1.5.31)$$

Compared with the full-matrix solution, all the calculations involved in between equations (15.1.5.12) and (15.1.5.18) and the subsequent iterations are spared in the diagonal approximation. This makes calculation by the diagonal approximation much faster than by the full-matrix method.

15.1.6. Example

To demonstrate the effect of different constraints on phase improvement, various density-modification techniques were applied to an MIR data set for which the refined structure coordinates are available. The test structure is 5-carboxymethyl-2-hydroxymuconate isomerase, solved by Wigley *et al.* (1989). MIR phases were available to 3.7 Å, with SIR information to 2.6 Å. Density modification was used to improve and extend phases to the limit of the data at 2.1 Å. The structure includes threefold noncrystallographic symmetry.

The MIR and density-modified phases are compared by plotting the mean of the cosine of the phase error, weighted by the figure of merit and structure-factor amplitude, as a function of resolution (Zhang *et al.*, 1997),

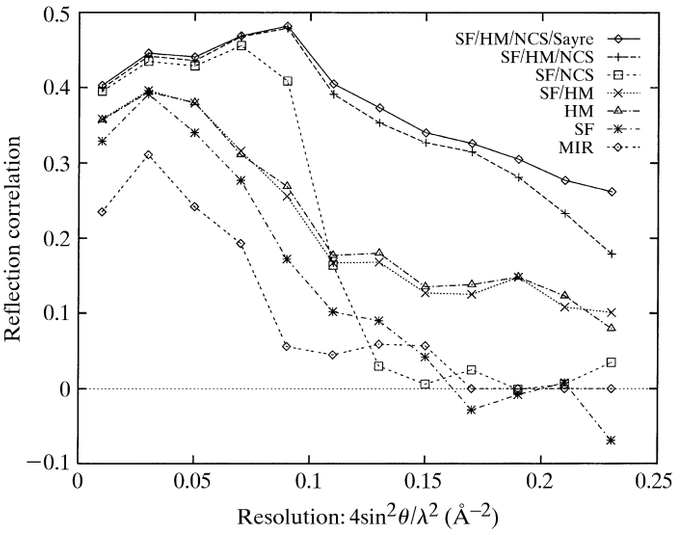

Fig. 15.1.6.1. Phase correlations after different combinations of density modifications.

$$C_f = \left\langle w|F|^2\cos(\varphi - \varphi_0)\right\rangle \Big/ \left(\left\langle w^2|F|^2\right\rangle\left\langle|F|^2\right\rangle\right)^{1/2}. \quad (15.1.6.1)$$

This phase correlation over all reflections is equivalent to map correlation. The results of density modification by various techniques, using the reflection-omit method for phase combination, are shown in Fig. 15.1.6.1.

Solvent flattening alone has slightly improved the phases at low resolution but has not lead to significant phase extension. The solvent-flattening function in Fig. 15.1.3.1 only has nonzero amplitudes close to the origin. It relates structure factors only in a very thin resolution shell. Therefore, solvent flattening is weak on phase extension.

Histogram matching alone improves the low-resolution phases and gives significant phase extension to higher resolutions. The histogram-matching function in Fig. 15.1.3.1 showed much stronger high-resolution amplitudes. Therefore, it could relate structure factors in a larger resolution shell. Moreover, there is always an ideal histogram specified at a given target resolution for phase extension. These two reasons combined make histogram matching a more powerful technique in phase extension than solvent flattening.

The combination of histogram matching and solvent flattening is slightly more powerful than histogram matching alone; since histogram matching sharpens the protein density, it implies an element of solvent flattening. Solvent flattening and averaging give a significant improvement at low resolution, but little phase extension. Averaging is powerful for phase refinement, but is weak for phase extension if no special precautions are taken. If there are flexible loop regions on the protein surface, these regions should be excluded from the molecular mask for averaging. The phasing power of averaging weakens at high resolution when the differences between NCS-related molecules become significant. Solvent flattening, histogram matching and averaging combined give a dramatic improvement at all resolutions. The addition of Sayre's equation gives a slight further improvement at high resolution.

Sayre's equation is very effective for phase refinement and extension at atomic or near atomic resolution. It becomes ineffective at low resolution or when the initial map is poor. Under these circumstances, it is better to apply other density-modification methods first to refine the phases and extend them to a higher resolution before Sayre's equation is applied. Sayre's

references