15. DENSITY MODIFICATION AND PHASE COMBINATION

15.1. Phase improvement by iterative density modification

BY K. Y. J. ZHANG, K. D. COWTAN AND P. MAIN

15.1.1. Introduction

Density modification is a technique for improving the quality of an approximate electron-density map based on some conserved features of the correct electron-density map. These conserved features are independent of the unknown fine detail of the structural conformation. They are often expressed as constraints on the electron density in various forms, either in real or reciprocal space. Since the structure-factor amplitudes are known, these constraints restrict the values of phases and can therefore be used for phase improvement.

The structure-factor amplitudes and phases are independent of each other if we know nothing about the electron density. Therefore, the phases are indeterminable given only the amplitudes (Baker, Krukowski & Agard, 1993). The information about the electron density provides the missing link between structure-factor amplitudes and phases. It is only through the knowledge of the chemical or physical properties of the electron density that the phases can be retrieved. Density modification is usually the most straightforward application of the constraints on electron density. However, this is only a matter of convenience in implementation. Sometimes the constraints can be more readily implemented in reciprocal space on structure factors.

Density-modification methods are usually implemented as an iterative procedure that alternates between density modification in real space and phase combination in reciprocal space. This paradigm was first proposed by Hoppe & Gassmann (1968) in their 'phase correction' method. This approach takes advantage of the particular properties of the constraints and uses them in a way that is most convenient to implement.

Density-modification methods usually require an initial map with substantial phase information. In most cases, these phases are obtained from multiple isomorphous replacement (MIR) or multi-wavelength anomalous dispersion (MAD), but it is also possible to improve maps from other sources, such as molecular replacement. The amount of information in the initial map is dependent on phase accuracy, data resolution and completeness. As more powerful constraints are incorporated, the density modification can be initiated from lower-resolution maps with less accurate phases. *Ab initio* phasing would be achieved if a density-modification method could start from a map generated from random phases. Therefore, density modification can potentially lead to *ab initio* phasing methods, although it does not seek direct solution to the phase problem as its immediate goal.

There are two major components in a density-modification procedure. One is the type of electron-density constraints. The other is the way the constraints are exploited. These two components combined determine the phasing power of the procedure. In this chapter, we will review various electron-density constraints and the way they are exploited for phase improvement.

15.1.2. Density-modification methods

The aim of density-modification calculations is to obtain new or improved phase estimates for observed structure-factor amplitudes. Often, this includes calculation of phases for previously unphased reflections, for example, in the case of phase extension. The calculation of weights, which indicate the degree of confidence in the new phase estimates, is also an important part of the calculation. Improved phase estimates are obtained by bringing the initial phase estimates into consistency with additional sources of structural information.

One difficulty in combining information from various sources is that the amplitudes and phases are represented in reciprocal space and include good estimates of error, whereas the other constraints are in real space and in general, represent expectations about the structure which may be hard to quantify. As a result, the method that has been adopted is iterative and divided into real- and reciprocal-space steps. A weighted map is calculated and used as a basis for applying all the real-space modifications. The modified map is then back-transformed to produce a set of amplitudes and phases. The agreement between the observed amplitudes and the amplitudes calculated from the modified map is then used to estimate weights for the modified phases, which are used to combine the modified phases with experimental phases to produce new phases. This process is shown diagrammatically in Fig. 15.1.2.1.

A broad range of techniques have been applied to electron-density maps to impose chemical or physical information. Some sources of information used in density modification are summarized in Table 15.1.2.1. The list included here is not exhaustive, but covers the most widely used methods. Here, we describe some of the constraints and the techniques through which these constraints are implemented for phase improvement.

15.1.2.1. *Solvent flattening*

Solvent flattening exploits the fact that the electron density in the solvent region is flat at medium resolution, owing to the high thermal motion and disorder of solvent molecules. The flattening of the solvent region suppresses noise in the map and therefore improves phases.

15.1.2.1.1. *Introduction*

Biological molecules are typically irregular in shape, often taking roughly globular forms. When they are packed regularly to form a crystal lattice, there are gaps left between them, and these
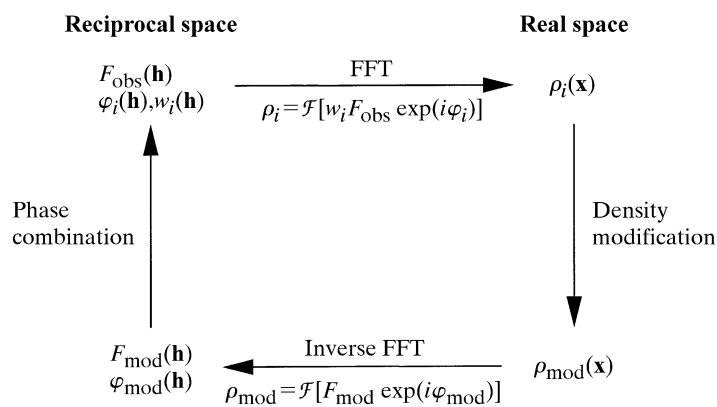

Fig. 15.1.2.1. Density-modification calculation showing iterative application of real-space and reciprocal-space constraints.

Table 15.1.2.1. *Constraints used in density modification*

Constraints	Use	Effectiveness and limitation
(1) Solvent flatness	Solvent flattening	Works best at medium resolution. Relatively resolution insensitive. Good for phase refinement. Weak on phase extension.
(2) Ideal electron-density distribution	Histogram matching	Works at a wide range of resolutions. More effective at higher resolution. Very effective for phase extension.
(3) Equal molecules	Molecular averaging	Works better at low to medium resolution. Its phasing power increases with the number of molecules in the asymmetric unit.
(4) Protein backbone connectivity	Skeletonization	Requires near atomic resolution to work.
(5) Local shape of electron density	Sayre's equation	The equation is exact at atomic resolution. It can be used at non-atomic resolution by choosing an appropriate shape function. Its phasing power increases quickly with resolution. Very powerful for phase extension.
(6) Atomicity	Atomization	If the initial map is good enough, iteration could lead to a final model.
(7) Structure-factor amplitudes	Sim weighting	Can be used to estimate the reliability of the calculated phases after density modification. It assumes the random distribution of errors that caused the discrepancy between the calculated and observed structure-factor amplitudes.
(8) Experimental phases	Phase combination	This can be used to filter out the incorrect component of the estimated phases. Most phase-combination procedures assume independence between the calculated and observed phases.

spaces are filled with the solvent in which the crystallization was performed. This solvent is a disordered liquid, and thus the arrangement of atoms in the solvent regions varies between unit cells, except in those small regions near the surface of the protein. The X-ray image forms an average of electron density over many cells, so the electron density over much of the solvent region appears to be constant to a good approximation.

The existence of a flat solvent region in a crystal places strong constraints on the structure-factor phases. The constraint of solvent flatness is implemented by identifying the molecular boundaries and replacing the densities in the solvent region by their mean density value.

When solving a structure, the contents of the unit cell are usually known, and so an estimate can be formed of how much of the cell volume is taken up by solvent (Matthews, 1968). If the solvent region can be located in the cell, then we can improve an electron-density map by setting the electron density in this region to the expected constant solvent density. Once the resulting modified phases are combined with the experimental data, an improvement can often be seen in the protein regions of the map (Bricogne, 1974).

The solvent region of a unit cell may usually be determined even from a poor MIR map using the following features:

(1) The mean electron density in the solvent region should be lower than that in the protein region. Note that this information will come from the low-resolution data, which dictate long-range density variations over the unit cell.

(2) The variation in density in the flat solvent region should be much smaller than that in the ordered protein region containing isolated clumps of density. The 'peakiness' of the protein region comes from the high-resolution data.

A good method for locating the solvent region therefore takes into account information from both low- and high-resolution structure factors. Many methods have been proposed to locate the protein–solvent boundary. The first of these were the visual identification methods. The boundary was identified by digitizing a mini-map with the aid of a graphic tablet (Hendrickson *et al.*, 1975; Schevitz *et al.*, 1981). The hand-digitizing procedure was very time-consuming and prone to subjective judgmental errors. Nevertheless, these methods demonstrated the potential of solvent flattening and stimulated further improvement on boundary-identification methods. An automated method using a linked, high-density approach was first proposed by Bhat & Blow (1982). Based on the fact that the densities are generally higher in the protein region than in the solvent region, they defined the molecular boundary by locating the protein as a region of linked, high-density points.

Convolution techniques were subsequently adopted as an efficient method of molecular-boundary identification. Reynolds *et al.* (1985) proposed a high mean absolute density value approach. The electron density within the protein region was expected to have greater excursions from the mean density value than the solvent region, which is relatively featureless. The molecular boundary was located based on the value of a smoothed 'modulus' electron

Scale =   1.0000 mm/A   Section   0
electron density

Scale =   1.0000 mm/A   Section   0
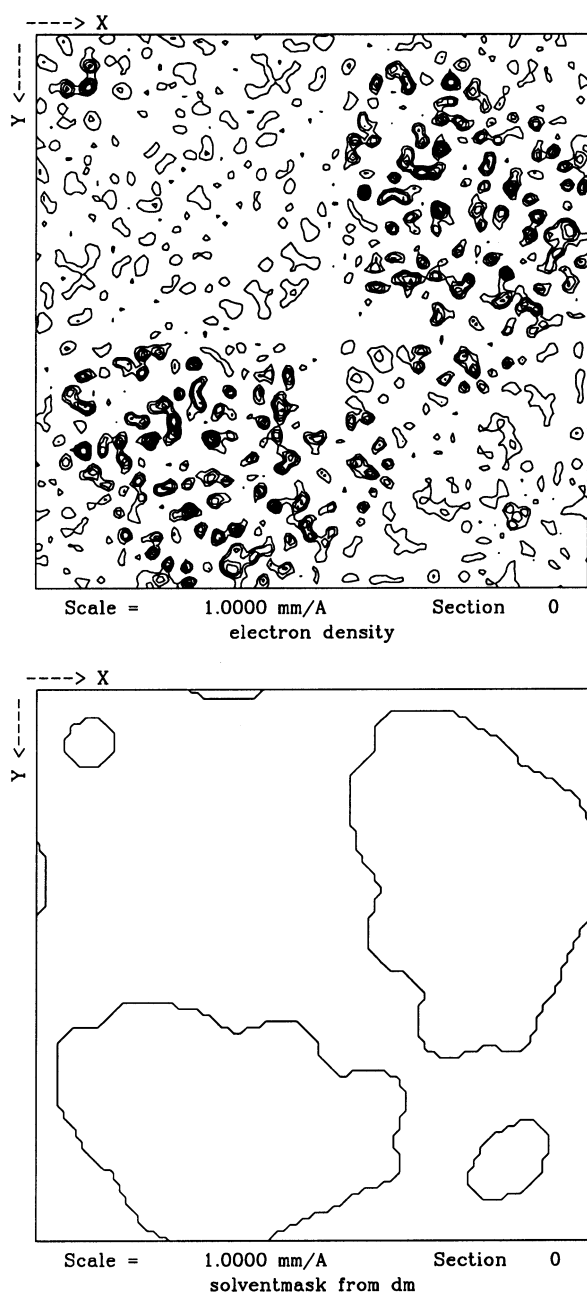solventmask from dm

Fig. 15.1.2.2. Solvent mask determined from a map by Wang's method.

density, which is the sum of the absolute values of all density points within a small box.

15.1.2.1.2. *The automated convolution method for molecular-boundary identification*

Wang (1985) suggested an automated convolution method for identifying the solvent region which has achieved widespread use. His method involved first calculating a truncated map:

$$\rho_{\text{trunc}}(\mathbf{x}) = \begin{cases} \rho\mathbf{x}, & \rho(\mathbf{x}) > \rho_{\text{solv}} \\ 0, & \rho(\mathbf{x}) < \rho_{\text{solv}} \end{cases}. \qquad (15.1.2.1)$$

The electron density is simply truncated at the expected solvent value, ρ_{solv}; however, since the variations in density in the protein region are much larger than the variations in the solvent region, it is generally only the protein region which will be affected. Thus, the mean density over the protein region is increased. Similar results may be obtained using the mean-squared difference of the density from the expected solvent value.

A smoothed map is then formed by calculating at each point in the map the mean density over a surrounding sphere of radius R. This operation can be written as a convolution of the truncated map, ρ_{trunc}, with a spherical weighting function, $w(\mathbf{r})$,

$$\rho_{\text{ave}}(\mathbf{x}) = \sum_{\mathbf{r}} \mathrm{w}(\mathbf{r})\rho_{\text{trunc}}(\mathbf{x} - \mathbf{r}), \qquad (15.1.2.2)$$

where

$$w(\mathbf{r}) = \begin{cases} 1 - |\mathbf{r}|/R, & |\mathbf{r}| < R \\ 0, & |\mathbf{r}| > R \end{cases}. \qquad (15.1.2.3)$$

Leslie (1987) noted that the convolution operation required in equation (15.1.2.2) can be very efficiently performed in reciprocal space using fast Fourier transforms (FFTs),

$$\rho_{\text{ave}}(\mathbf{x}) = \mathscr{F}^{-1}\{\mathscr{F}[\rho_{\text{trunc}}(\mathbf{x})]\mathscr{F}[w(\mathbf{r})]\}, \qquad (15.1.2.4)$$

where $\mathscr{F}$ denotes a Fourier transform, and $\mathscr{F}^{-1}$ represents an inverse Fourier transform.

The Fourier transform of the truncated density can be readily calculated using FFTs. The Fourier transform of the weighting function can be calculated analytically by

$$g(s) = \mathscr{F}[w(\mathbf{r})] = \frac{3[\sin(2\pi Rs) - 2\pi Rs\cos(2\pi Rs)]}{(2\pi Rs)^3}$$
$$- \frac{3\{4\pi Rs\sin(2\pi Rs) - [(2\pi Rs)^2 - 2]\cos(2\pi Rs) - 2\}}{(2\pi Rs)^4},$$

$$(15.1.2.5)$$

where

$$s = 2\sin\theta/\lambda.$$

Therefore, the averaging of the truncated electron density by a spherical weighting function can be achieved by two FFTs. This greatly reduced the time required for calculating the averaged density. Other weighting functions may be implemented by the same approach.

A cutoff value, ρ_{cut}, is then calculated, which divides the unit cell into two portions occupying the correct volumes for the protein and solvent regions. All points in the map where $\rho_{\text{ave}}(\mathbf{x}) < \rho_{\text{cut}}$ can then be assumed to be in the solvent region. A typical mask obtained from an MIR map by this means, and the modified map, are shown in Fig. 15.1.2.2.

The radius of the sphere, R, used in equation (15.1.2.3) for the averaging of electron densities is generally around 8 Å. The molecular envelope derived from such an averaged map tends to lose details of the protein molecular surface. Paradoxically, a large averaging sphere is required for the identification of the protein–solvent boundary based on the difference between the mean density of the protein and solvent, which is very small and can only be distinguished when a sufficiently large area of the map is averaged. Abrahams & Leslie (1996) proposed an alternative method of molecular-boundary identification that uses the standard deviation of the electron density within a given radius relative to the overall mean at every grid point of a map. The local-standard-deviation map is the square root of a convolution of a sphere and the squared map, which can be calculated in reciprocal space in a similar way to the procedure described in equations (15.1.2.4) and (15.1.2.5) as proposed by Leslie (1987). By integrating the histogram of the local-standard-deviation map, the cutoff value of the local standard deviation corresponding to the solvent fraction can be calculated. Using this procedure, a molecular envelope that contains more details of the protein molecular surface can be obtained, since the radius of the averaging sphere can be as low as 4 Å (Abrahams & Leslie, 1996).

313

15.1.2.1.3. *The solvent-flattening procedure*

Once the envelope has been determined, solvent flattening is performed by simply setting the density in the solvent region to the expected value, ρ_{solv}:

$$\rho_{mod}(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}), & \rho_{ave}(\mathbf{x}) > \rho_{cut} \\ \rho_{solv}, & \rho_{ave}(\mathbf{x}) < \rho_{cut} \end{cases}. \qquad (15.1.2.6)$$

If the electron density has not been calculated on an absolute scale, the solvent density may be set to its mean value.

A related method is solvent flipping, developed by Abrahams & Leslie (1996). In this approach, the flattening operation is modified by the introduction of a relaxation factor, γ, where γ is positive, effectively 'flipping' the density in the solvent region.

$$\rho_{mod}(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}), & \rho_{ave}(\mathbf{x}) > \rho_{cut} \\ \rho_{solv} - [\gamma/(1-\gamma)][\rho(\mathbf{x}) - \rho_{solv}], & \rho_{ave}(\mathbf{x}) < \rho_{cut} \end{cases}.$$
$$(15.1.2.7)$$

The effect of this modification is to correct for the problem of independence in phase combination and is discussed in Section 15.1.4.3.

15.1.2.2. *Histogram matching*

Histogram matching seeks to bring the distribution of electron-density values of a map to that of an ideal map. The density histogram of a map is the probability distribution of electron-density values. It provides a global description of the appearance of the map, and all spatial information is discarded. The comparison of the histogram for a given map with that expected for an ideal map can serve as a measure of quality. Furthermore, the initial map can be improved by adjusting density values in a systematic way to make its histogram match the ideal histogram.

15.1.2.2.1. *Introduction*

Histogram matching is a standard technique in image processing. It is aimed at bringing the density distribution of an image to an ideal distribution, thereby improving the image quality. The first attempt at modifying the electron-density distribution was that by Hoppe & Gassman (1968), who proposed the '3-2' rule. The electron density was first normalized to a maximum of 1 and modified by imposing positivity. Subsequently, the electron density was modified by $\rho_{mod} = 3\rho^2 - 2\rho^3$. Podjarny & Yonath (1977) used the skewness of the density histogram as a measure of quality of the modified map. Harrison (1988) used a Gaussian function as the ideal histogram in his histogram-specification method for protein phase refinement and extension. The choice of the Gaussian function as the ideal electron-density distribution was based on theoretical arguments instead of experimental evaluation. The Gaussian function was also made independent of resolution. Lunin (1988) used the electron-density distribution to retrieve the values of low-angle structure factors whose amplitudes had not been measured during an X-ray experiment. The electron-density distribution was thought to be structure specific and was derived from a homologous structure. Moreover, the histogram was derived from the entire unit cell, including both the protein and the solvent. Zhang & Main (1988) systematically examined the electron-density histogram of several proteins and found that the ideal density histogram is dependent on resolution, the overall temperature factor and the phase error. It is, however, independent of structural conformation. The sensitivity to phase error suggests that the density histogram could be used for phase improvement. The structural conformation independence made it possible to predict the ideal histogram for unknown structures.

15.1.2.2.2. *The prediction of the ideal histogram*

Polypeptide structures in particular, and biological macromolecules in general, display a broadly similar atomic composition, and the way in which these atoms bond together is also conserved across a wide range of structures. These similarities between different protein structures can be used to predict the ideal histogram even when positional information for individual atoms is not available in a map. If the positional information is removed from an electron-density map, then what remains is an unlabelled list of density values. This list is the histogram of the electron-density distribution, which is independent of the relative disposition of these densities. The shape of the histogram is primarily based on the presence of atoms and their characteristic distances from each other. This is true for all polypeptide structures.

The frequency distribution, $P(\rho)$, of electron-density values in a map can be constructed by sampling the map and counting the density values in different ranges. In practice, once the electron-density map has been sampled on a discrete grid, this frequency distribution becomes a histogram, but for convenience, it is treated here as a continuous distribution.

At resolutions of better than 6.0 Å and after exclusion of the solvent region, the frequency distribution of electron-density values for protein density over a wide range of proteins varies only with resolution and overall temperature factor to a good approximation. If the overall temperature factor is artificially adjusted, for example, by sharpening to $B_{overall} = 0$, then the frequency distributions may be treated as a function of resolution only. Therefore, once a good approximation to the molecular envelope is known, the frequency distribution of electron densities in the protein region as a function of resolution may be assumed to be known. Therefore, the ideal density histogram for an unknown map at a given resolution can be taken from any known structure at the same resolution (Zhang & Main, 1988, 1990*a*).

The ideal electron-density histogram can also be predicted by an analytical formula (Lunin & Skovoroda, 1991; Main, 1990*a*). The method adopted by Main (1990*a*) represents the density histogram by components that correspond to three types of electron density in the map. The first component is the region of overlapping densities, which can be represented by a randomly distributed background noise. The second component is the region of partially overlapping densities. The third component is the region of non-overlapping atomic peaks, which can be represented by a Gaussian.

The histogram for the overlapping part of the density can be represented by a Gaussian distribution,

$$P_o(\rho) = N \exp\left[-(\rho - \overline{\rho})^2/2\sigma^2\right], \qquad (15.1.2.8)$$

where $\overline{\rho}$ is the mean density and σ is the standard deviation. The region of partially overlapping densities can be modelled by a cubic polynomial function,

$$P_{po}(\rho) = N\left(a\rho^3 + b\rho^2 + c\rho + d\right). \qquad (15.1.2.9)$$

The histogram for the non-overlapping part of the density can be derived analytically from a Gaussian atom,

$$P_{no}(\rho) = N(A/\rho)[\ln(\rho_0/\rho)]^{1/2}, \qquad (15.1.2.10)$$

where ρ_0 is the maximum density, N is a normalizing factor and A is the relative weight of the terms between equation (15.1.2.8) and equation (15.1.2.10).

314

If we use two threshold values, ρ_1 and ρ_2, to divide the three density regions, the complete formula can be expressed as

$$P(\rho) = \begin{cases} N\exp\left[-(\rho-\overline{\rho})^2/2\sigma^2\right] & \text{for} \quad 2\rho \leq \rho_2 \\ N(a\rho^3+b\rho^2+c\rho+d) & \text{for} \quad 2\rho_2 < \rho \leq \rho_1. \\ N(A/\rho)[\ln(\rho_0/\rho)]^{1/2} & \text{for} \quad 2\rho_1 < \rho \leq \rho_0 \end{cases}$$
$$(15.1.2.11)$$

The parameters a, b, c, d in the cubic polynomial are calculated by matching function values and gradients at ρ_1 and ρ_2. The parameters in the histogram formula, $\overline{\rho}$, σ, A, ρ_0, ρ_1, ρ_2, can be obtained from histograms of known structures.

15.1.2.2.3. *The process of histogram matching*

Zhang & Main (1990*a*) demonstrated that, at better than 4 Å resolution, the histogram for an MIR map is generally significantly different from the ideal distribution calculated from atomic coordinates. The obvious course is therefore to alter the map in such a way as to make its density histogram equal to the ideal distribution. Unfortunately, there are an infinite number of maps corresponding to any chosen density distribution, so we must choose a systematic method of altering the map.

The conventional method of performing such a modification is to retain the ordering of the density values in the map. The highest point in the original map will be the highest point in the modified map, the second highest points will correspond in the same way, and so on.

Mathematically, this transformation is represented as follows. Let $P(\rho)$ be the current density histogram and $P'(\rho)$ be the desired distribution, normalized such that their sums are equal to 1. The cumulative distribution functions, $N(\rho)$ and $N'(\rho)$, may then be calculated:

$$N(\rho) = \int_{\rho_{\min}}^{\rho} P(\rho)\,\mathrm{d}\rho,$$
$$N'(\rho') = \int_{\rho_{\min}}^{\rho'} P'(\rho)\,\mathrm{d}\rho.$$
$$(15.1.2.12)$$

The cumulative distribution function of a variable transforms a value chosen from the distribution into a number between 0 and 1, representing the position of that value in an ordered list of values chosen from the distribution.

The transformation may, therefore, be performed in two stages. A density value is taken from the initial distribution and the cumulative distribution function of the initial distribution is applied to obtain the position of that value in the distribution. The inverse of the cumulative distribution function for the desired distribution is applied to this value to obtain the density value for the corresponding point in the desired distribution. Thus, given a density value, ρ, from the initial distribution, the modified value, ρ', is obtained by

$$\rho' = N'^{-1}[N(\rho)]. \qquad (15.1.2.13)$$

The distribution of ρ' will then match the desired distribution after the above transformation. The transformation of an electron-density value by this method is illustrated in Fig. 15.1.2.3. The transformation in equation (15.1.2.13) can be achieved through a linear transform represented by

$$\rho'_i = a_i\rho_i + b_i, \qquad (15.1.2.14)$$

where $i = \{1, \ldots, n\}$ and n is the number of density bins. The above linear transform is sufficient if the number of density bins is large enough. An n value of about 200 is usually quite satisfactory.

Various properties of the electron density are specified in the density histogram, such as the minimum, maximum and mean density, the density variance, and the entropy of the map. The mean density of the ideal map can be obtained by

$$\overline{\rho} = \int_{\rho_{\min}}^{\rho_{\max}} \rho P(\rho)\,\mathrm{d}\rho. \qquad (15.1.2.15)$$

The variance of the density in the ideal map can be obtained by

$$\sigma(\rho) = \left(\overline{\rho^2} - \overline{\rho}^2\right)^{1/2}, \qquad (15.1.2.16)$$

where

$$\overline{\rho^2} = \int_{\rho_{\min}}^{\rho_{\max}} \rho^2 P(\rho)\,\mathrm{d}\rho. \qquad (15.1.2.17)$$

The entropy of the ideal map can be calculated by

$$S = -\int_{\rho_{\min}}^{\rho_{\max}} P(\rho)\rho\ln(\rho)\,\mathrm{d}\rho. \qquad (15.1.2.18)$$

Therefore, the process of histogram matching applies a minimum and a maximum value to the electron density, imposes the correct mean and variance, and defines the entropy of the new map. The order of electron-density values remains unchanged after histogram matching.

Histogram matching is complementary to solvent flattening since it is applied to the protein region of a map, whereas
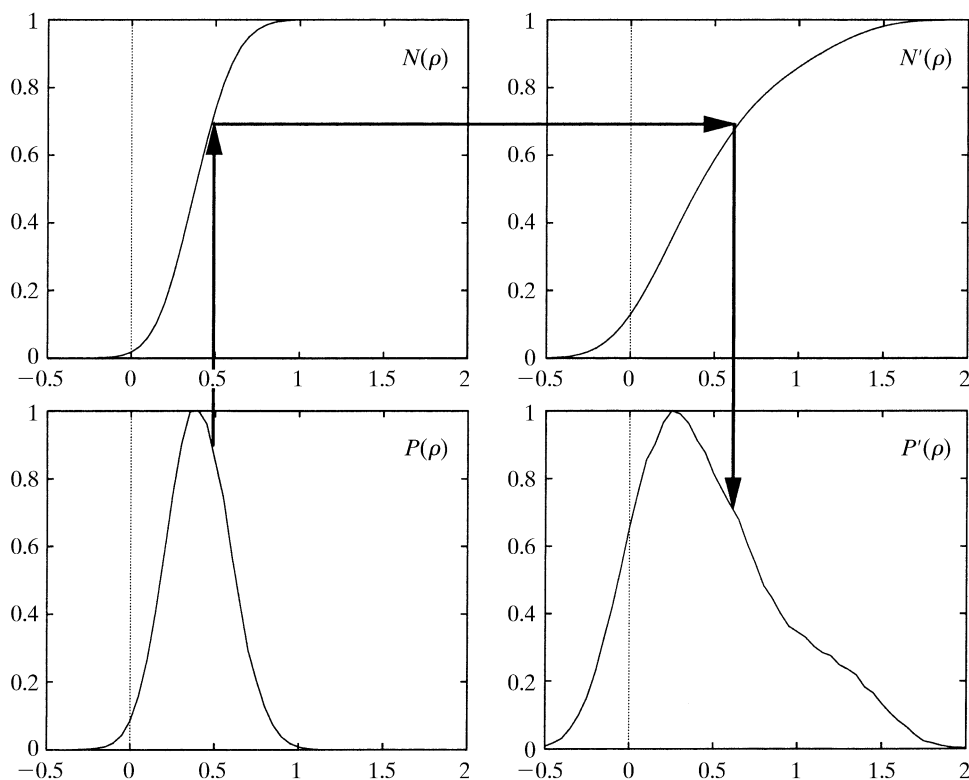

Fig. 15.1.2.3. Transformation of density ρ to ρ'_{mod} by histogram matching.

315

solvent flattening only operates on the solvent region of the map. The same envelope that was used for isolating the solvent region can be used to determine the protein region of the cell. An alternative approach is to define separate solvent and protein masks, with uncertain regions excluded from either mask and allowed to keep their unmodified values.

15.1.2.2.4. *Scaling the observed structure-factor amplitudes according to the ideal density histogram*

In the process of density modification, electron density or structure factors from different sources are compared and combined. It is, therefore, crucial to ensure that all the structure factors and maps are on the same scale. The observed structure factors can be put on the absolute scale by Wilson statistics (Wilson, 1949) using a scale and an overall temperature factor. This is accurate when atomic or near atomic resolution data are available. The scale and overall temperature factor obtained from Wilson statistics are less accurate when only medium- to low-resolution data are available. A more robust method of scaling non-atomic resolution data is through the density histogram (Cowtan & Main, 1993; Zhang, 1993).

The ideal density histogram defines the mean and variance of an electron density, as shown in equations (15.1.2.15) and (15.1.2.16). We can scale the observed structure-factor amplitudes to be consistent with the target histogram using the following formula, obtained from the structure-factor equation and Parseval's theorem. The mean density and the density variance of the observed map can be calculated as

$$\overline{\rho}' = (1/V)F(000), \tag{15.1.2.19}$$

$$\sigma'(\rho) = (1/V)\left[\sum_{\mathbf{h}}|F(\mathbf{h})|^2\right]^{1/2}. \tag{15.1.2.20}$$

The mean and variance of the electron-density map at the desired resolution are calculated using the target histogram, the mean value of the solvent density, $\overline{\rho}_{\text{solv}}$, and the solvent volume of the cell, V_{solv}. The $F(000)$ term can then be evaluated from equations (15.1.2.15) and (15.1.2.19):

$$F(000) = (V - V_{\text{solv}})\overline{\rho} + V_{\text{solv}}\overline{\rho}_{\text{solv}}. \tag{15.1.2.21}$$

The scale of the observed amplitudes can be obtained from equations (15.1.2.16) and (15.1.2.20),

$$F'(\mathbf{h}) = KF(\mathbf{h}), \tag{15.1.2.22}$$

where

$$K = \left[\left(\overline{\rho^2} - \overline{\rho}^2\right)\right]^{1/2} \Big/ \left\{(1/V)\left[\sum_{\mathbf{h}}|F(\mathbf{h})|^2\right]^{1/2}\right\}. \tag{15.1.2.23}$$

This method is adequate for scaling observed structure factors at any resolution.

15.1.2.3. *Averaging*

The averaging method enforces the equivalence of electron-density values between grid points in the map related by noncrystallographic symmetry. The averaging procedure can filter noise, correct systematic error and even determine the phases *ab initio* in favourable cases (Chapman *et al.*, 1992; Tsao *et al.*, 1992).

15.1.2.3.1. *Introduction*

Noncrystallographic symmetry (NCS) arises in crystals when there are two or more of the same molecules in one asymmetric unit. Such symmetries are local, since they only apply within a sub-region of a single unit cell. A fivefold axis, for example, must be noncrystallographic, since it is not possible to tessellate objects with fivefold symmetry. Since the symmetry does not map the crystal lattice back onto itself, the individual molecules that are related by the noncrystallographic symmetry will be in different environments; therefore, the symmetry relationships are only approximate.

Noncrystallographic symmetries provide phase information by the following means. Firstly, the related regions of the map may be averaged together, increasing the ratio of signal to noise in the map. Secondly, since the asymmetric unit must be proportionally larger to hold multiple copies of the molecule, the number of independent diffraction amplitudes available at any resolution is also proportionally larger. This redundancy in sampling the molecular transform leads to additional phase information which can be used for phase improvement.

15.1.2.3.2. *The determination of noncrystallographic symmetry*

The self-rotation symmetry is now routinely solved by the use of a Patterson rotation function (Rossmann & Blow, 1962). The translation symmetry can be determined by a translation function (Crowther & Blow, 1967) when a search model, either an approximate structure of the protein to be determined or the structure of a homologous protein, is available. The searches of the Patterson rotation and translation functions are achieved typically using fast automatic methods, such as *X-PLOR* (Brünger *et al.*, 1987) or *AMoRe* (Navaza, 1994). In cases where no search model is available or the Patterson translation function is unsolvable, either the whole electron-density map, or a region which is expected to contain a molecule, may be rotated using the rotation solution and used as a search model in a phased translation function (Read & Schierbeek, 1988).

Once the averaging operators are determined, the mask can be determined using the local density correlation function as developed by Vellieux *et al.* (1995). This is achieved by a systematic search for extended peaks in the local density correlation, which must be carried out over a volume of several unit cells in order to guarantee finding the whole molecule. The local correlation function distinguishes those volumes of crystal space which map onto similar density under transformation by the averaging operator. Thus, in the case of improper NCS, a local correlation mask will cover only one monomer. In the case of a proper symmetry, a local correlation mask will cover the whole complex (Fig. 15.1.2.4*a*,*b*).
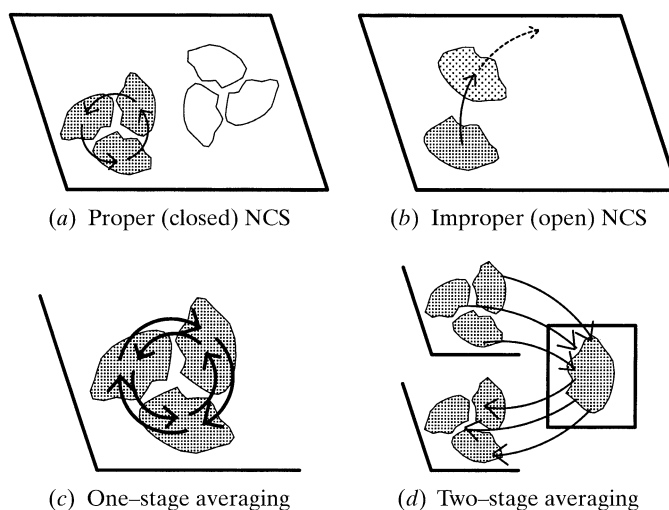

(*a*) Proper (closed) NCS   (*b*) Improper (open) NCS

(*c*) One–stage averaging   (*d*) Two–stage averaging

Fig. 15.1.2.4. Types of noncrystallographic symmetry and averaging calculation.

Special cases arise when there are combinations of crystallographic and noncrystallographic symmetries, of proper and improper symmetries, or when a noncrystallographic symmetry element maps a cell edge onto itself. In the latter case, the volume of matching density is infinite, and arbitrary limits must be placed upon the mask along one crystal axis.

15.1.2.3.3. *The refinement of noncrystallographic symmetry*

The initial NCS operation obtained from rotation and translation functions or heavy-atom positions can be fine-tuned by a density-space R-factor search in the six-dimensional rotation and translation space. The density-space R factor is defined as

$$R = \sum_{\mathbf{r}} |\rho(\mathbf{r}) - \rho(\mathbf{r}')| / \sum_{\mathbf{r}} |\rho(\mathbf{r}) + \rho(\mathbf{r}')|, \qquad (15.1.2.24)$$

where $\mathbf{r} = \{xyz\}$ is the set of Cartesian coordinates, $\mathbf{r}' = \Omega\mathbf{r}$ is the NCS-related set of coordinates of $\mathbf{r}$ and Ω represents the NCS operator.

The six-dimensional search is very time-consuming. The search rate can be increased by using only a representative subset of grid points. The NCS operation is systematically altered to find the lowest density-space R factor for the selected subset of grid points.

The solution of the NCS operation from the six-dimensional search can be further refined by the following least-squares procedure. If $\rho(\mathbf{r})$ is related to $\rho(\mathbf{r}')$ by the NCS operation, Ω,

$$\rho(\mathbf{r}') = \rho(\Omega\mathbf{r}). \qquad (15.1.2.25)$$

Here, Ω is a function of ω, $\Omega = f(\omega)$, where $\omega = \{\alpha, \beta, \gamma, t_x, t_y, t_z\}$ represents the rotation and translation components of the NCS operation. The solution to the NCS parameters, ω, can be obtained by minimizing the density residual between the NCS-related molecules,

$$\varepsilon(\mathbf{r}) = \rho(\mathbf{r}) - \rho(\Omega\mathbf{r}), \qquad (15.1.2.26)$$

using a least-squares formula of the form

$$\left(\frac{\partial\rho}{\partial\omega}\right)^T \left(\frac{\partial\rho}{\partial\omega}\right)\Delta\omega = \left(\frac{\partial\rho}{\partial\omega}\right)^T \varepsilon(\mathbf{r}), \qquad (15.1.2.27)$$

where $\Delta\omega$ is the shift to the NCS parameters. Here,

$$\frac{\partial\rho}{\partial\omega} = \frac{\partial\rho}{\partial\mathbf{r}}\frac{\partial\mathbf{r}}{\partial\omega}. \qquad (15.1.2.28)$$

The partial derivatives, $\partial\rho/\partial\mathbf{r} = \{\partial\rho/\partial x,\ \partial\rho/\partial y,\ \partial\rho/\partial z\}$, can be calculated by Fourier transforms,

$$\frac{\partial\rho}{\partial x} = -\frac{2\pi i}{V}\sum_{hkl} hF_{hkl}\exp[-2\pi i(hx + ky + lz)]$$

$$\frac{\partial\rho}{\partial y} = -\frac{2\pi i}{V}\sum_{hkl} kF_{hkl}\exp[-2\pi i(hx + ky + lz)] \qquad (5.1.2.29)$$

$$\frac{\partial\rho}{\partial z} = -\frac{2\pi i}{V}\sum_{hkl} lF_{hkl}\exp[-2\pi i(hx + ky + lz)],$$

or more efficiently with a single Fourier transform by the use of spectral B-splines (Cowtan & Main, 1998). $\partial\mathbf{r}/\partial\omega$ is derived analytically based on the relationship between the Cartesian coordinates, r, and the rotational and translational coordinates of the NCS operation, ω,

15.1.2.3.4. *The averaging of NCS-related molecules*

Once the mask and matrices are determined, the electron-density map may be modified by averaging. This can be achieved in one or two stages: The density for each copy of the molecule in the asymmetric unit may be replaced by the averaged density from every copy; however, this becomes slow for high-order NCS (Fig. 15.1.2.4c). Alternatively, a single averaged copy of the molecule may be created in an artificial cell [referred to by Rossmann *et al.* (1992) as an *H*-cell], and then each copy of the molecule may be reconstructed in the asymmetric unit from this copy (Fig. 15.1.2.4d). This is more efficient for high-order NCS, but additional errors are introduced in the second interpolation.

Interpolation of electron-density values at non-map grid sites is usually required, since the NCS operators will not normally map grid points onto each other. To obtain accurate interpolated values, either a fine grid or a complex interpolation function are required; suitable functions are described in Bricogne (1974) and Cowtan & Main (1998). Solvent flattening and histogram matching are frequently applied after averaging, since histogram matching tends to correct for any smoothing introduced by density interpolation.

In the case of flexible proteins, it may be necessary to average only part of the molecule, in which case the averaging mask will exclude some parts of the unit cell which are indicated as protein by the solvent mask. In other cases, it may be necessary to apply multi-domain averaging; in this case, the protein is divided into rigid domains which can appear in differing orientations. Each domain must then have a separate mask and set of averaging matrices.

Averaging may also be performed across similar molecules in multiple crystal forms (Schuller, 1996); in this case, density modification is performed on each crystal form simultaneously, with averaging of the molecular density across all copies of the molecule in all crystal forms. This is a powerful technique for phase improvement, even when no phasing is available in some crystal forms.

15.1.2.4. *Skeletonization*

The skeletonization method enhances connectivity in the map. This is achieved by locating ridges of density, constructing a graph of linked peaks, and then building a new map using cylinders of density around the graph peaks.

At worse than atomic resolution, the density peaks for bonded atoms are no longer resolved, and so interpretation of the density in terms of atomic positions involves recognition of common motifs in the pattern of ridges in the density. Skeletonization was a tool developed by Greer (1985) to assist model building by tracing high ridges in the electron density to describe the connectivity in the map.

Skeletonization has more recently been adapted to the problem of density modification (Baker, Bystroff *et al.*, 1993; Bystroff *et al.*, 1993; Wilson & Agard, 1993). A skeleton is constructed by tracing the ridges in the map. The resulting ridges form connected 'trees'. These trees may be pruned to remove small unconnected fragments and break circuits to select for protein-like features. A new map may then be built by building density around the links of the skeleton using the profile of a cylindrically averaged atom at the appropriate resolution.

The skeletonization method has been used to add new features to a partial model of a molecule (Baker, Bystroff *et al.*, 1993). An

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\sin\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}. \qquad (15.1.2.30)$$

317

efficient alternative algorithm for tracing density ridges is given by Swanson (1994).

15.1.2.5. *Sayre's equation*

Sayre's equation constrains the local shape of electron density. It provides a link between all structure-factor amplitudes and phases. It is an exact equation at atomic resolution in an equal-atom system. It is, therefore, very powerful for phase refinement and extension for small molecules at atomic resolution (Sayre, 1952, 1972, 1974). However, its power diminishes as resolution decreases. It can still be an effective tool for macromolecular phase refinement and extension if the shape function can be modified to accommodate the overlap of atoms at non-atomic resolution (Zhang & Main, 1990*b*).

15.1.2.5.1. *Sayre's equation in real and reciprocal space*

Sayre's equation (Sayre, 1952, 1972, 1974) expresses the constraint on structure factors when the atoms in a structure are equal and resolved, and the equation has formed the foundation of direct methods. In protein calculations, the resolution is generally too poor for atoms to be resolved, and this is reflected in the bulk of the terms required to calculate the equation for any particular missing structure factor.

For equal and resolved atoms, squaring the electron density changes only the shape of the atomic peaks and not their positions. The original density may therefore be restored by convoluting with some smoothing function, $\psi(\mathbf{x})$, which is a function of atomic shape,

$$\rho(\mathbf{x}) = (V/N)\sum_{\mathbf{y}}\rho^2(\mathbf{y})\psi(\mathbf{x} - \mathbf{y}), \qquad (15.1.2.31)$$

where

$$\psi(\mathbf{x} - \mathbf{y}) = (1/V)\sum_{\mathbf{h}}\theta(\mathbf{h})\exp[2\pi i\mathbf{h}\cdot(\mathbf{x} - \mathbf{y})]. \qquad (15.1.2.32)$$

Here, $\theta(\mathbf{h})$ is the ratio of scattering factors of real, $f(\mathbf{h})$, and 'squared', $g(\mathbf{h})$, atoms, and V is the unit-cell volume, *i.e.*,

$$\theta(\mathbf{h}) = f(\mathbf{h})/g(\mathbf{h}). \qquad (15.1.2.33)$$

Sayre's equation states that the convolution of the squared electron density with a shape function restores the original electron density. It can be seen from equation (15.1.2.31) that Sayre's equation puts constraints on the local shape of electron density. The local shape function is the Fourier transform of the ratio of scattering factors of the real and 'squared' atoms.

Sayre's equation is more frequently expressed in reciprocal space as a system of equations relating structure factors in amplitude and phase:

$$F(\mathbf{h}) = [\theta(\mathbf{h})/V]\sum_{\mathbf{k}}F(\mathbf{k})F(\mathbf{h} - \mathbf{k}). \qquad (15.1.2.34)$$

The reciprocal-space expression of Sayre's equation can be obtained directly from a Fourier transformation of both sides of equation (15.1.2.31) and the application of the convolution theorem.

15.1.2.5.2. *The application of Sayre's equation to macromolecules at non-atomic resolution – the $\theta(\mathbf{h})$ curve*

Sayre's equation is exact for an equal-atom structure at atomic resolution. The reciprocal-space shape function, $\theta(\mathbf{h})$, can be calculated analytically from the ratio of the scattering factors of real and 'squared' atoms, which can both be represented by a Gaussian function. At infinite resolution, we expect $\theta(\mathbf{h})$ to be a spherically symmetric function that decreases smoothly with increased $\mathbf{h}$. However, for data at non-atomic resolution, the $\theta(\mathbf{h})$

curve will behave differently because atomic overlap changes the peak shapes. Therefore, a spherical-averaging method is adopted to obtain an estimate of the shape function empirically from the ratio of the observed structure factors and the structure factors from the squared electron density using the formula

$$\theta(s) = V\left\langle F(\mathbf{h})\Big/\sum_{\mathbf{k}}F(\mathbf{k})F(\mathbf{h} - \mathbf{k})\right\rangle_{|\mathbf{h}|}, \qquad (15.1.2.35)$$

where the averaging is carried out over ranges of $|\mathbf{h}|$, *i.e.*, over spherical shells, each covering a narrow resolution range. Here, s represents the modulus of $\mathbf{h}$.

The empirically derived shape function only extends to the resolution of the experimentally observed phases. This is sufficient for phase refinement. However, there are no experimentally observed phases to give the empirical $\theta(s)$ for phase extension. Therefore, a Gaussian function of the form

$$\theta(s) = K\exp(-Bs^2) \qquad (15.1.2.36)$$

is fitted to the available values of $\theta(s)$, and the parameters K and B are obtained using a least-squares method. The shape function $\theta(s)$ for the resolution beyond that of the observed phases is extrapolated using the fitted Gaussian function. The derivation of the shape function $\theta(s)$ from a combination of spherical averaging and Gaussian extrapolation is the key to the successful application of Sayre's equation for phase improvement at non-atomic resolution (Zhang & Main, 1990*b*).

15.1.2.6. *Atomization*

The atomization method uses the fact that the structure underlying the map consists of discrete atoms. It attempts to interpret the map by automatically placing atoms and refining their positions.

Agarwal & Isaacs (1977) proposed a method for the extension of phases to higher resolutions by interpreting an electron-density map in terms of 'dummy' atoms. These are so called because at the initial resolution of 3.0 Å, true atom peaks could not be resolved. The placement of 'dummy atoms' is subject to constraints of bonding distance and the number of neighbours. The coordinates and temperature factors of these dummy atoms may then be refined against all the available diffraction amplitudes. Structure factors may then be calculated from the refined coordinates to provide phases for the high-resolution reflections and to improve the phases of the starting set.

The atomization approach has been extended in the *ARP* program (Lamzin & Wilson, 1997) by the use of difference-map criteria to test dummy-atom assignments, with the aim of removing wrong atoms and introducing missing atoms. With modern refinement algorithms, this technique has become very effective for the solution of structures at high resolution from a poor molecular-replacement model, or even directly from an MIR/MAD map.

Map improvement has also been demonstrated at intermediate resolutions by Perrakis *et al.* (1997) using a multi-solution variant of the *ARP* method, and by Vellieux (1998).

The interpretation of an approximately phased map has also been applied very successfully as part of the 'Shake n' Bake' direct-methods procedure (Miller *et al.*, 1993; Weeks *et al.*, 1993). The alternating application of phase refinement by the minimum principle in reciprocal space ('Shake') and atomization in real space ('Bake') has proved to be a very powerful method for solving small protein structures at atomic resolution using only structure-factor amplitudes.

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

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

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

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

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.

$$\rho_{mod}(\mathbf{x}) = g_{ncs}(\mathbf{x})(1/N_{ncs})\sum_{i}\rho_i(\mathbf{x}). \qquad (15.1.3.3)$$

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 ith 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,

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

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-
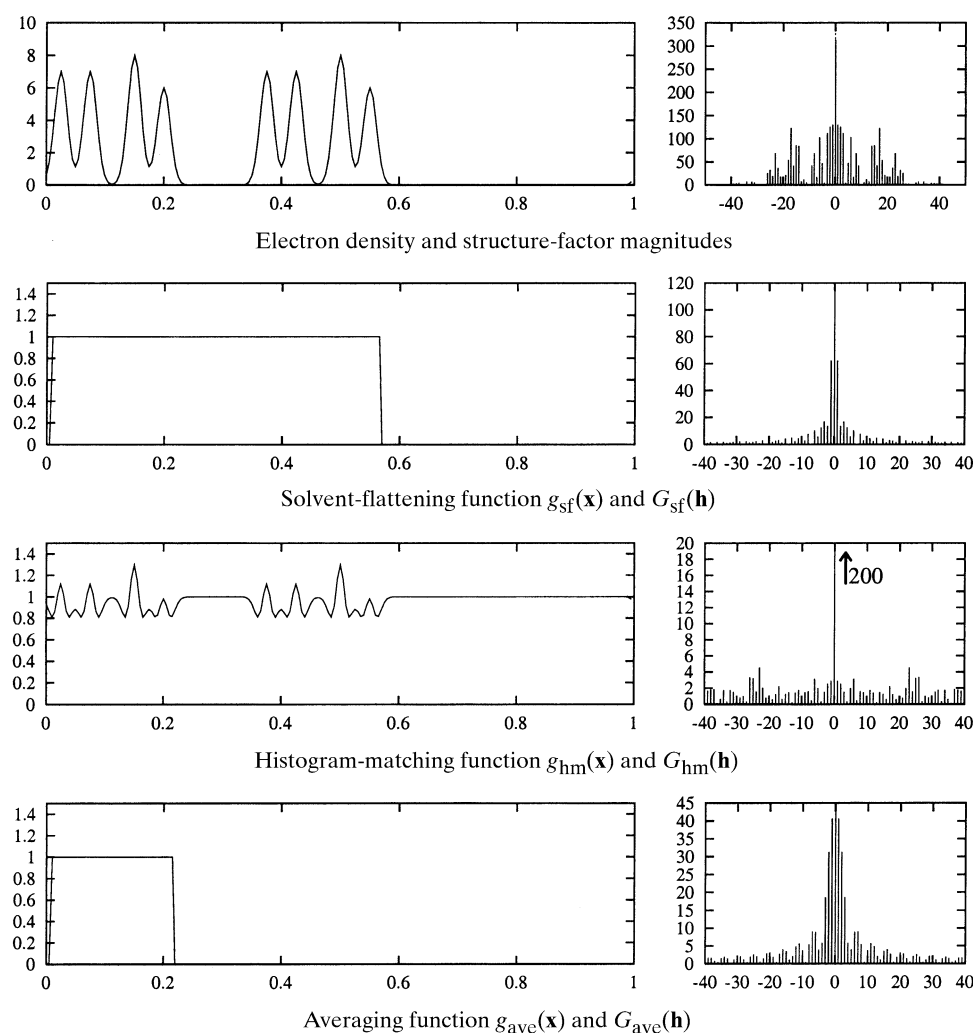

Fig. 15.1.3.1. The functions $g(\mathbf{x})$ and $G(\mathbf{h})$ for solvent flattening, histogram matching and averaging.

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 σ_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:

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

where

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

and

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

where Σ_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

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

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 Σ_Q may be estimated by a number of means, for example (Bricogne, 1976),

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

where the average is normally taken over all reflections at a particular resolution. A more sophisticated approach is the σ_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 underdetermined, 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, 1990*a*). 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 γ 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

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

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

then

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

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

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

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

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

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*)

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

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,

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

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,

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

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

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial F_1}{\partial \rho_1} & \dfrac{\partial F_1}{\partial \rho_2} & \cdots & \dfrac{\partial F_1}{\partial \rho_n} \\[2mm] \dfrac{\partial F_2}{\partial \rho_1} & \dfrac{\partial F_2}{\partial \rho_2} & \cdots & \dfrac{\partial F_2}{\partial \rho_n} \\[1mm] \vdots & \vdots & \ddots & \vdots \\[1mm] \dfrac{\partial F_m}{\partial \rho_1} & \dfrac{\partial F_m}{\partial \rho_2} & \cdots & \dfrac{\partial F_m}{\partial \rho_n} \end{bmatrix}, \qquad (15.1.5.5)$$

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,

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

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),

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

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

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

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,

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

the initial residual is

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

and the initial search step is

$$\mathbf{p}_0 = \mathbf{r}_0. \qquad (15.1.5.11)$$

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

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

where

$$\alpha_k = \mathbf{r}_k^T\mathbf{p}_k/\mathbf{q}_k^T\mathbf{q}_k \qquad (15.1.5.13)$$

and

$$\mathbf{q}_k = \mathbf{J}\mathbf{p}_k. \qquad (15.1.5.14)$$

The new residual is

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

where

$$\mathbf{s}_k = \mathbf{J}^T\mathbf{q}_k. \qquad (15.1.5.16)$$

The next search step which conjugates with the residual is

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

where

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

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),

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

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

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

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

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

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

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

where

$$\Delta F(\mathbf{h}) = F(\mathbf{h}) - \theta(\mathbf{h})G(\mathbf{h}), \qquad (15.1.5.23)$$
$$G(\mathbf{h}) = (V/N)\sum\limits_{\mathbf{y}}\rho^2(\mathbf{y})\exp(2\pi i\mathbf{h}\mathbf{y}) \qquad (15.1.5.24)$$

and

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

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,

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

$$= \left[\frac{Q_k(\mathbf{x})}{p_k(\mathbf{x})} \right], \qquad (15.1.5.26)$$

where the vector is partitioned as shown above, and

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

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

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

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

$$-Q_k(\mathbf{x}) + p_k(\mathbf{x}), \qquad (15.1.5.29)$$

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

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

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

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

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
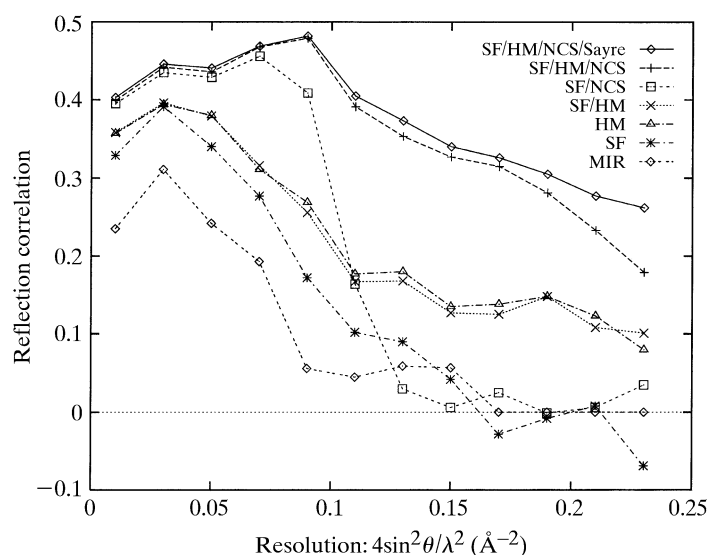

Fig. 15.1.6.1. Phase correlations after different combinations of density modifications.

merit and structure-factor amplitude, as a function of resolution (Zhang *et al.*, 1997),

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

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

equation also decreases in power as the solvent content increases, since it is only applicable to the protein regions of the map.

The fact that the best results were obtained when all the constraints were combined indicates that each constraint contains some degree of independent phasing information. Moreover, it also suggests that the strengths of these constraints are complementary. Each constraint, when applied in isolation, may introduce systematic errors that are difficult to overcome when a different constraint is subsequently applied. This problem is greatly reduced when the constraints are applied simultaneously and the combined process iterates much further towards the desired density map.

Density-modification methods have become sufficiently powerful that it is possible to solve structures from comparatively poor initial maps. This has reduced the amount of effort required to find more heavy-atom derivatives and to collect additional diffraction data sets. Density modification may simplify the process of map interpretation, even when good phase information is available. Density modification can also be used to obtain phases *ab initio* when high-order noncrystallographic symmetry is present.

Acknowledgements

Parseval's theorem) to minimizing the difference between the model electron density and the density corresponding to the map coefficients $|\mathbf{F}_O|\exp(i\alpha_C)$; a lower residual is obtained either by making the model look more like the true structure, or by making the model-phased map look more like the model through the introduction of systematic phase errors.

A number of strategies are available to reduce the degree or impact of refinement bias. The overestimation of phase accuracy has been overcome in a new version of *SIGMAA* that is under development (Read, unpublished). Cross-validation data, which are normally used to compute R_{free} as an unbiased indicator of refinement progress (Brünger, 1992), are used to obtain unbiased σ_A estimates. Because of the high statistical error of σ_A estimates computed from small numbers of reflections, reliable values can only be obtained by exploiting the smoothness of the σ_A curve as a function of resolution. This can be achieved either by fitting a functional form or by adding a penalty to points that deviate from the line connecting their neighbours. Lunin & Skovoroda (1995) have independently proposed the use of cross-validation data for this purpose, but as their algorithm is equivalent to the conventional *SIGMAA* algorithm, it will suffer severely from statistical error.

The degree of refinement bias can be reduced by placing less weight on the agreement of structure-factor amplitudes. Anecdotal evidence suggests that the problem is less serious, in structures refined using *X-PLOR* (Brünger *et al.*, 1987), when the Engh & Huber (1991) parameter set is used for the energy terms. In this new parameter set, the deviations from standard geometry are much more strictly restrained, so in effect the pressure on the agreement of structure-factor amplitudes is reduced. The use of maximum-likelihood targets for refinement (discussed below) also helps to reduce overfitting.

If errors are suspected in certain parts of the structure, 'omit refinement' (in which the questionable parts are omitted from the model) can be a very effective way to eliminate refinement bias in those regions (James *et al.*, 1980; Hodel *et al.*, 1992).

If MIR or MAD (multiwavelength anomalous dispersion) phases are available, combined phase maps tend to suffer less from refinement bias, depending on the extent to which the experimental phases influence the combined phases. Finally, it is always a good idea to refer occasionally to the original MIR or MAD map, which cannot suffer at all from model bias or refinement bias.

15.2.9. Maximum-likelihood structure refinement

In the past, conventional structure refinement was based on a least-squares target, which would be justified if the observed and calculated structure-factor amplitudes were related by a Gaussian probability distribution. Unfortunately, the relationship between $|\mathbf{F}_O|$ and $|\mathbf{F}_C|$ is not Gaussian, and the distribution for $|\mathbf{F}_O|$ is not even centred on $|\mathbf{F}_C|$. Because of this, it was suggested (Read, 1990; Bricogne, 1991) that a maximum-likelihood target should be used instead, and that it should be based on probability distributions such as those described above.

Three implementations of maximum-likelihood structure refinement have now been reported (Pannu & Read, 1996; Murshudov *et al.*, 1997; Bricogne & Irwin, 1996). As expected, there is a decrease in refinement bias, as the calculated structure-factor amplitudes will not be forced to be equal to the observed amplitudes. Maximum-likelihood targets have been shown to work much better than least-squares targets, particularly when the starting models are poor.

Prior phase information can also be incorporated into a maximum-likelihood target (Pannu *et al.*, 1998). Tests show that even weak phase information can have a dramatic effect on the success of refinement, and that the amount of overfitting is even further reduced (Pannu *et al.*, 1998).

Acknowledgements

References

15.1

Abrahams, J. P. (1997). *Bias reduction in phase refinement by modified interference functions: introducing the γ correction.* Acta Cryst. D**53**, 371–376.

Abrahams, J. P. & Leslie, A. G. W. (1996). *Methods used in the structure determination of bovine mitochondrial F_1 ATPase.* Acta Cryst. D**52**, 30–42.

Agarwal, R. C. & Isaacs, N. W. (1977). *Method for obtaining a high resolution protein map starting from a low resolution map.* Proc. Natl Acad. Sci. USA, **74**(7), 2835–2839.

Baker, D., Bystroff, C., Fletterick, R. J. & Agard, D. A. (1993). *PRISM: topologically constrained phase refinement for macromolecular crystallography.* Acta Cryst. D**49**, 429–439.

Baker, D., Krukowski, A. E. & Agard, D. A. (1993). *Uniqueness and the ab initio phase problem in macromolecular crystallography.* Acta Cryst. D**49**, 186–192.

Bhat, T. N. & Blow, D. M. (1982). *A density-modification method for the improvement of poorly resolved protein electron-density maps.* Acta Cryst. A**38**, 21–29.

Blow, D. M. & Rossmann, M. G. (1961). *The single isomorphous replacement method.* Acta Cryst. **14**, 1195–1202.

Bricogne, G. (1974). *Geometric sources of redundancy in intensity data and their use for phase determination.* Acta Cryst. A**30**, 395–405.

Bricogne, G. (1976). *Methods and programs for direct-space exploitation of geometric redundancies.* Acta Cryst. A**32**, 832–847.

Brünger, A. T. (1992). *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature (London)*, **355**, 472–475.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Crystallographic R factor refinement by molecular dynamics. Science*, **235**, 458–460.

Bystroff, C., Baker, D., Fletterick, R. J. & Agard, D. A. (1993). *PRISM: application to the solution of two protein structures.* Acta Cryst. D**49**, 440–448.

Chapman, M. S., Tsao, J. & Rossmann, M. G. (1992). *Ab initio phase determination for spherical viruses: parameter determination for spherical-shell models.* Acta Cryst. A**48**, 301–312.

Cowtan, K. D. (1999). *Error estimation and bias correction in phase-improvement calculations.* Acta Cryst. D**55**, 1555–1567.

Cowtan, K. D. & Main, P. (1993). *Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints.* Acta Cryst. D**49**, 148–157.

Cowtan, K. D. & Main, P. (1996). *Phase combination and cross validation in iterated density-modification calculations.* Acta Cryst. D**52**, 43–48.

Cowtan, K. D. & Main, P. (1998). *Miscellaneous algorithms for density modification.* Acta Cryst. D**54**, 487–493.

Cowtan, K. D. & Zhang, K. Y. J. (1999). *Density modification for macromolecular phase improvement. Prog. Biophys. Mol. Biol.* **72**, 245–270.

Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure.* Acta Cryst. **23**, 544–548.

15.1 (*cont.*)

Greer, J. (1985). *Computer skeletonization and automatic electron-density map analysis.* In *Diffraction methods for biological macromolecules*, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, Vol. 115, pp. 206–224. Orlando: Academic Press.

Harrison, R. W. (1988). *Histogram specification as a method of density modification. J. Appl. Cryst.* **21**, 949–952.

Hauptman, H. (1986). *The direct methods of X-ray crystallography. Science*, **233**, 178–183.

Hendrickson, W. A., Klippenstein, G. L. & Ward, K. B. (1975). *Tertiary structure of myohemerythrin at low resolution. Proc. Natl Acad. Sci. USA*, **72**(6), 2160–2164.

Hendrickson, W. A. & Lattman, E. E. (1970). *Representation of phase probability distributions for simplified combination of independent phase information. Acta Cryst.* B**26**, 136–143.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Karle, J. (1986). *Recovering phase information from intensity data. Science*, **232**, 837–843.

Lamzin, V. S. & Wilson, K. S. (1997). *Automated refinement for protein crystallography. Methods Enzymol.* **277**, 269–305.

Leslie, A. G. W. (1987). *A reciprocal-space method for calculating a molecular envelope using the algorithm of B. C. Wang. Acta Cryst.* A**43**, 134–136.

Lunin, V. Yu. (1988). *Use of the information on electron density distribution in macromolecules. Acta Cryst.* A**44**, 144–150.

Lunin, V. Yu. & Skovoroda, T. P. (1991). *Frequency-restrained structure-factor refinement. I. Histogram simulation. Acta Cryst.* A**47**, 45–52.

Main, P. (1979). *A theoretical comparison of the β, γ' and $2F_o - F_c$ syntheses. Acta Cryst.* A**35**, 779–785.

Main, P. (1990a). *A formula for electron density histograms for equal-atom structures. Acta Cryst.* A**46**, 507–509.

Main, P. (1990b). *The use of Sayre's equation with constraints for the direct determination of phases. Acta Cryst.* A**46**, 372–377.

Main, P. & Rossmann, M. G. (1966). *Relationships among structure factors due to identical molecules in different crystallographic environments. Acta Cryst.* **21**, 67–72.

Matthews, B. W. (1968). *Solvent content of protein crystals. J. Mol. Biol.* **33**, 491–497.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *On the application of the minimal principle to solve unknown structures. Science*, **259**, 1430–1433.

Navaza, J. (1994). *AMoRe: an automated package for molecular replacement. Acta Cryst.* A**50**, 157–163.

Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Cryst.* D**53**, 448–455.

Podjarny, A. D. & Yonath, A. (1977). *Use of matrix direct methods for low-resolution phase extension for tRNA. Acta Cryst.* A**33**, 655–661.

Read, R. J. (1986). *Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Cryst.* A**42**, 140–149.

Read, R. J. & Schierbeek, A. J. (1988). *A phased translation function. J. Appl. Cryst.* **21**, 490–495.

Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L. & Matthews, B. W. (1985). *Structure of a serine protease from rat mast cells determined from twinned crystals by isomorphous and molecular replacement. Acta Cryst.* B**41**, 139–147.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst.* **15**, 24–31.

Rossmann, M. G., McKenna, R., Tong, L., Xia, D., Dai, J.-B., Wu, H., Choi, H.-K. & Lynch, R. E. (1992). *Molecular replacement real-space averaging. J. Appl. Cryst.* **25**, 166–180.

Sayre, D. (1952). *The squaring method: a new method for phase determination. Acta Cryst.* **5**, 60–65.

Sayre, D. (1972). *On least-squares refinement of the phases of crystallographic structure factors. Acta Cryst.* A**28**, 210–212.

Sayre, D. (1974). *Least-squares phase refinement. II. High-resolution phasing of a small protein. Acta Cryst.* A**30**, 180–184.

Schevitz, R. W., Podjarny, A. D., Zwick, M., Hughes, J. J. & Sigler, P. B. (1981). *Improving and extending the phases of medium- and low-resolution macromolecular structure factors by density modification. Acta Cryst.* A**37**, 669–677.

Schuller, D. J. (1996). *MAGICSQUASH: more versatile non-crystallographic averaging with multiple constraints. Acta Cryst.* D**52**, 425–434.

Sim, G. A. (1959). *The distribution of phase angles for structures containing heavy atoms. II. A modification of the normal heavy-atom method for non-centrosymmetrical structures. Acta Cryst.* **12**, 813–815.

Swanson, S. M. (1994). *Core tracing: depicting connections between features in electron density. Acta Cryst.* D**50**, 695–708.

Tsao, J., Chapman, M. S. & Rossmann, M. G. (1992). *Ab initio phase determination for viruses with high symmetry: a feasibility study. Acta Cryst.* A**48**, 293–301.

Vellieux, F. M. D. (1998). *A comparison of two algorithms for electron-density map improvement by introduction of atomicity: skeletonization, and map sorting followed by refinement. Acta Cryst.* D**54**, 81–85.

Vellieux, F. M. D. A. P., Hunt, J. F., Roy, S. & Read, R. J. (1995). *DEMON/ANGEL: a suite of programs to carry out density modification. J. Appl. Cryst.* **28**, 347–351.

Wang, B. C. (1985). *Resolution of phase ambiguity in macromolecular crystallography.* In *Diffraction methods for biological macromolecules*, edited by H. W. Wyckoff, C. H. W. Hirs & S. N. Timasheff, Vol. 115, pp. 90–113. Orlando: Academic Press.

Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Applications of the minimal principle to peptide structures. Acta Cryst.* D**49**, 179–181.

Wigley, D. B., Roper, D. I. & Cooper, R. A. (1989). *Preliminary crystallographic analysis of 5-carboxymethyl-2-hydroxymuconate isomerase from Escherichia coli. J. Mol. Biol.* **210**, 881–882.

Wilson, A. J. C. (1949). *The probability distribution of X-ray intensities. Acta Cryst.* **2**, 318–321.

Wilson, C. & Agard, D. A. (1993). *PRISM: automated crystallographic phase refinement by iterative skeletonization. Acta Cryst.* A**49**, 97–104.

Woolfson, M. M. (1987). *Direct methods – from birth to maturity. Acta Cryst.* A**43**, 593–612.

Zhang, K. Y. J. (1993). *SQUASH – combining constraints for macromolecular phase refinement and extension. Acta Cryst.* D**49**, 213–222.

Zhang, K. Y. J., Cowtan, K. D. & Main, P. (1997). *Combining constraints for electron density modification.* In *Macromolecular crystallography*, edited by C. W. Carter & R. M. Sweet, Vol. 277, pp. 53–64. New York: Academic Press.

Zhang, K. Y. J. & Main, P. (1988). *Histogram matching as a density modification technique for phase refinement and extension of protein molecules.* In *Improving protein phases*, edited by S. Bailey, E. Dodson & S. Phillips. Report DL/SCI/R26, pp. 57–64. Warrington: Daresbury Laboratory.

Zhang, K. Y. J. & Main, P. (1990a). *Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. Acta Cryst.* A**46**, 41–46.

Zhang, K. Y. J. & Main, P. (1990b). *The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Cryst.* A**46**, 377–381.

15.2

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Bricogne, G. (1991). *A multisolution method of phase determination by combined maximization of entropy and likelihood. III. Extension to powder diffraction data. Acta Cryst.* A**47**, 803–829.

references