Density modification for macromolecular phase improvement

Kevin D. Cowtan [a,*], Kam Y.J. Zhang [b]

[a]*Department of Chemistry, University of York, York YO10 5DD, UK*
[b]*Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, Seattle, WA 90109, USA*

Abstract

Density modification provides a simple and largely automatic tool for improving phase estimates for observed structure factors. The phase information arises from a combination of the known structure factor magnitudes, the current phase estimates, and stereochemical information. The magnitudes, the current phase estimates, and stereochemical information. The addition of these phase information derived from theoretical sources renders new structures amenable to solution, and reduces the effort required to solve other structures. A diverse array of techniques which have been applied to the phase improvement problem are reviewed. © 1999 Published by Elsevier Science Ltd. All rights reserved.

1. Introduction

The rapid increase of the number of known three-dimensional (3D) macromolecular structures is largely due to the success of X-ray crystallography. X-ray methods reveal the 3D structure of macromolecules using the diffraction phenomenon caused by the interaction of X-rays with electrons in the macromolecules that are arranged in a 3D crystal lattice. To obtain the image of the macromolecules, requires the knowledge of both the amplitudes and phases of the diffracted X-rays. However, it is only the amplitudes of the diffracted X-rays that are observed, the phase information is missing. The solution to the phase problem is one of the major challenges in the determination of 3D structures of macromolecules by X-ray crystallography.

* Corresponding author.

The phase information for macromolecules are generally derived experimentally using either the multiple isomorphous replacement (MIR) method or the multiple anomalous dispersion (MAD) method. These methods require additional measurements of diffraction data and are sometimes very time consuming. In the MIR method phase information is deduced by the introduction of heavy atoms to the crystal and comparison of the resulting diffraction pattern. However, the introduction of new atoms alters the crystal packing to various degrees, consequently the derived phases may suffer from inaccuracies due to lack of isomorphism in the derivative crystal and rarely reaches the full resolution of the native data. The electron density maps derived from the phases estimated from these experiments may not be of sufficient quality into which an atomic structure can be built. These drawbacks have stimulated efforts to improve the accuracy of the experimentally derived phases and to extend them to the full resolution of the native data. As a result, various techniques that seek to improve the quality of the electron density map and thereby the phases by imposing some known physical constraints on the electron density were proposed, the earliest being the 'phase correction' method of Hoppe and Gassmann (1968).

Density modification is a tool for generating improved phase estimates, and therefore improved electron density maps, when a set of experimental structure factor magnitudes and some initial phase estimates are available. Often this includes calculation of phases for previously unphased reflections. The calculation of weights, which indicate a degree of confidence in the new phase estimates, is an important part of the calculation. Additional information arising from chemical knowledge about the structure is combined with the information from the initial phase estimates in order to obtain the improved phases.

Sources of information about a structure include the following.

1. The observed native magnitudes.
2. Experimental phase estimates based on isomorphous/anomalous differences.
3. Solvent content in the crystal. The solvent can usually be located from very weak phase information, and thus, this part of the crystal structure can be solved.
4. Density histogram: the ideal electron density distribution in the protein region of the map can serve as a constraint for phase improvement.
5. Atomicity: this can be employed strongly as Sayre's equation at high resolution and as a known density histogram or through atomization at lower resolutions.
6. Molecular packing: the association of molecules into an oligomer either in solution or on crystallization provides redundancy in the map, which is informative in phasing.
7. Chain connectivity can be exploited through skeletonization, which allows enhancement of connectivity and protein-like features in the density.

Since some of this information is expressed in real space and some in reciprocal space, the phase improvement calculation must also span both spaces. In order for phase improvement to be integrated with statistical phasing and refinement procedures, the improved phases must also be expressed in terms of phase probability distributions for each reflection. However, the real-space constraints are difficult to express in statistical terms, therefore there must also be a transformation into and out of the statistical phase representation in each cycle of the calculation. The resulting approach is as follows.

The centroid of the initial phase probability distribution is calculated, giving rise to a 'best'

Reciprocal space   **Real space**

$|F_{obs}|, P(\phi)$ ρ_{init}

Weighted map calculation (FFT)

Phase combination
$P_{comb}(\phi) = P_{obs}(\phi)P_{mod}(\phi)$

$P_{mod}(\phi)$

Map modification

Phase weighting
$P_{mod}(\phi) = f(F_{obs}, F_{mod})$

$|F_{mod}|, \phi_{mod}$ Structure factor calculation (FFT) ρ_{mod}
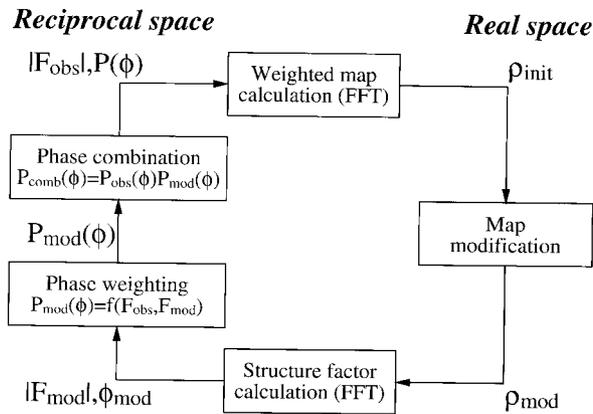
Fig. 1. Density modification calculation, showing iterative application of real-space and reciprocal space constraints.

phase and weight that can be used to calculate a minimally noisy map. All the real-space constraints are then applied as modifications to this map to bring it into agreement with those constraints. The modified map is then back-transformed to produce a set of magnitudes and phases. The agreement between the observed magnitudes and the values from the map is then used to estimate the level of phase error in the modified phases. This error estimate is transformed back into a phase probability distribution, which is then combined with the original experimental phase probability distribution. This process is shown diagrammatically in Fig. 1.

2. Density modification methods

2.1. Solvent flattening

Biological molecules are typically irregular in shape, often taking rough globular forms. Therefore when they are packed regularly to form a crystal lattice there are gaps between them, with the spaces filled with the solvent in which the crystallization was performed. This solvent is a disordered liquid, and thus apart from small regions near the surface of the protein the arrangement of atoms in the solvent regions varies between unit cells.

The X-ray image forms an average of electron density over many cells, and so to a good approximation the electron density over much of the solvent region appears to be constant.

When solving a structure the contents of the unit cell are usually known, and so an estimate can usually be formed of how much of the cell volume is taken up by solvent. If the solvent region can be located in the cell, then we can improve an electron density map by setting all the electron density in this region to the expected mean solvent density. This in itself is not very useful, however, 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 by the following features:

- the mean electron density in the solvent region should be lower than in the protein region; note that this information will come from the low-resolution data, which dictates long range density variations over the unit cell;
- the variation in density in the flat solvent region should be much smaller than in the ordered protein where the map should show sharp atomic features, the sharp features of the protein region will come from the high-resolution data.

A good method for locating the solvent region will therefore take into account information from both low and higher resolution structure factors.

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

$$\rho_{\text{trunc}}(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}), & \rho(\mathbf{x}) > \rho_{\text{solv}} \\ 0, & \rho(\mathbf{x}) < \rho_{\text{solv}} \end{cases}. \tag{1}$$

The electron density is simply truncated at the expected solvent value ρ_{solv}, however, as 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 over a surrounding sphere of the truncated density. This has the effect of smoothing the map. We can write this operation as a convolution with a spherical top hat function,

$$\rho_{\text{av}}(\mathbf{x}) = \sum_{\mathbf{r}} g(\mathbf{y})\rho_{\text{trunc}}(\mathbf{x} - \mathbf{y}), \tag{2}$$

where

$$g(\mathbf{y}) = \begin{cases} 1 & |\mathbf{y}| < r \\ 0 & |\mathbf{y}| > r \end{cases}. \tag{3}$$

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 region. All points in the map where $\rho_{\text{av}}(\mathbf{x}) < \rho_{\text{cut}}$ can then be assumed to be in the solvent region.

Leslie (1987) noted that the convolution operation required in Eq. (2) can be more efficiently performed in reciprocal space, thus making the calculation very quick to perform on modern computers. A section from a typical mask obtained from an MIR map by this means, and the final model, is shown in Fig. 2.

Once the envelope is available, solvent flattening is performed by simply changing the density in the solvent region as follows;

$$\rho_{\text{mod}}(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}) & \rho_{\text{av}}(\mathbf{x}) > \rho_{\text{cut}} \\ \rho_{\text{solv}} & \rho_{\text{av}}(\mathbf{x}) < \rho_{\text{cut}} \end{cases}, \tag{4}$$

where ρ_{solv} is the expected electron density in the solvent region.

A related method is solvent flipping, developed by Abrahams and Leslie (1996). In their
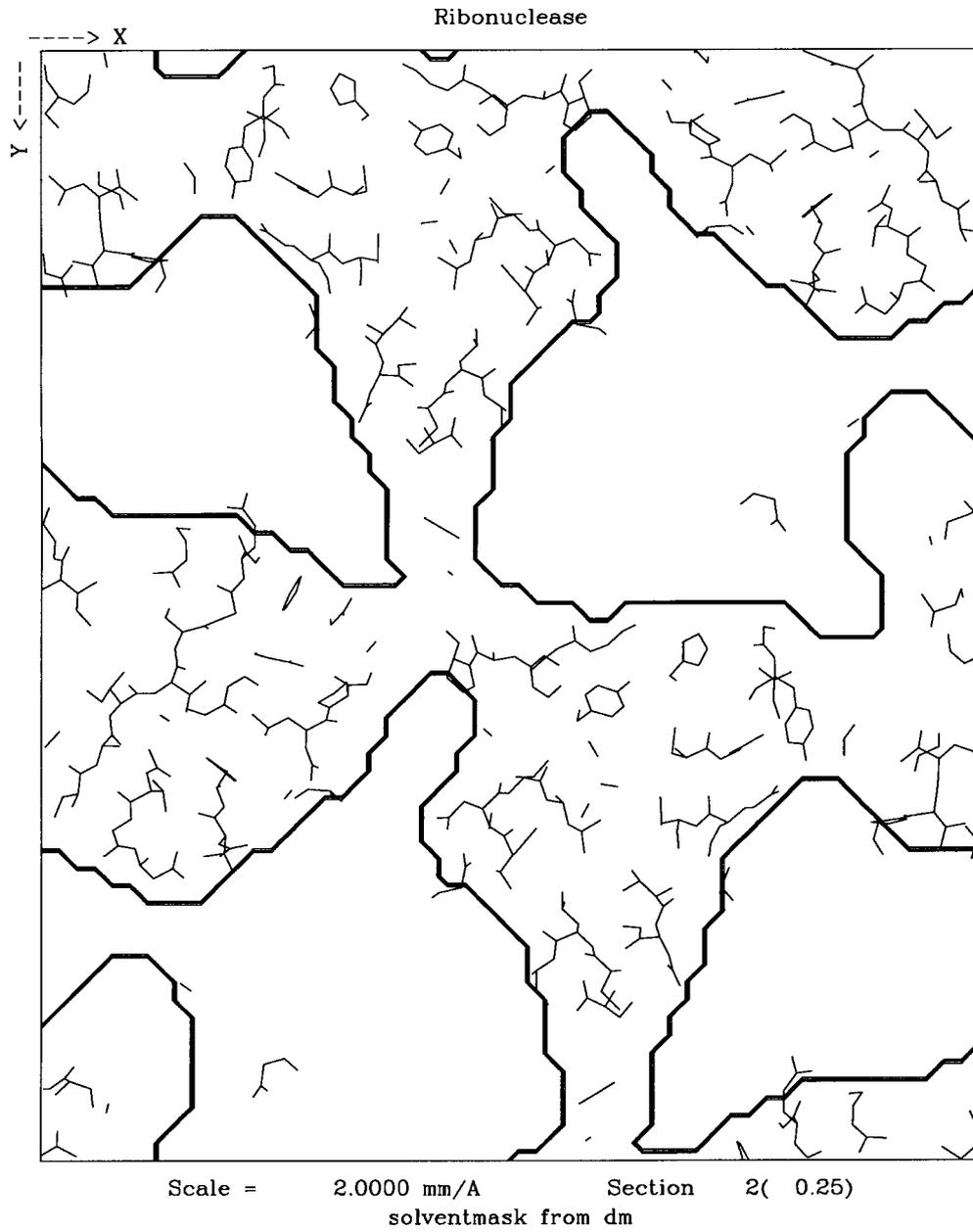
Fig. 2. Solvent mask determined from a map by Wang's method.

approach the flattening operation is modified by the introduction of a relaxation factor γ,

$$\rho_{\mathrm{mod}}(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}), & \rho_{\mathrm{av}}(\mathbf{x}) > \rho_{\mathrm{cut}} \\ \rho(\mathbf{x}) + \gamma[\rho_{\mathrm{solv}} - \rho(\mathbf{x})], & \rho_{\mathrm{av}}(\mathbf{x}) < \rho_{\mathrm{cut}} \end{cases}, \tag{5}$$

where the optimum value of γ is greater than 1, effectively 'flipping' the density in the solvent region. The effect of this modification is to correct the problem of independence in phase combination, and is discussed in section 3.3.

2.2. Histogram matching

A complementary method to solvent flattening, which may be applied to the protein region of a map, was proposed by Zhang and Main (1990a). This technique was histogram matching, a method widely used in image processing. Biological macromolecules in general and polypeptide structures in particular display a broadly similar atomic composition, and the way in which these atoms bond is also conserved across a wide range of structures. These similarities in the construction of different protein structures can be used, even when positional information for individual atoms is not available.

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, and contains some useful information. The electron density histogram has been used in phase refinement and extension (Harrison, 1988; Zhang and Main, 1990a), retrieval of the values of low-angle structure factors whose amplitudes have not been measured during X-ray experiment (Lunin, 1988) and even ab initio phase determination at low resolution (Lunin et al., 1998, 1995, 1990). 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.

Zhang and Main (1990a) demonstrated that at better than 4 Å resolution the histogram for an MIR map is generally significantly different from a theoretical distribution calculated from atomic coordinates. The obvious course is therefore to alter the map in order to make its density histogram equal to the theoretical distribution. Unfortunately, there are an infinite number of maps corresponding to any chosen density distribution, so it is necessary to 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 source 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. Let these frequency distributions be normalized, such that:

$$\int_0^\infty P(\rho) \, \mathrm{d}\rho = 1, \tag{6}$$

$$\int_0^\infty P'(\rho) \, \mathrm{d}\rho = 1. \tag{7}$$

Next the cumulative distribution functions $N(\rho)$ and $N'(\rho)$ are calculated:

$$N(\rho) = \int_0^\rho P(\rho_1) \, \mathrm{d}\rho_1, \tag{8}$$

$$N'(\rho) = \int_0^\rho P'(\rho_1) \, \mathrm{d}\rho_1. \tag{9}$$

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 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. This is because the cumulative distribution of the density value in the initial density histogram should equal the cumulative distribution of the modified density value in the ideal density histogram, i.e.,

$$N(\rho) = N'(\rho_{\mathrm{mod}}). \tag{10}$$

Thus, given a density value ρ from the initial distribution, the modified value ρ' is obtained by:

$$\rho_{\mathrm{mod}} = N'^{-1}[N(\rho)]. \tag{11}$$

The distribution of ρ_{mod} will then match the desired distribution. The transformation of an electron density value by this method is illustrated in Fig. 3.

The same envelope, which was used for isolating the solvent region, can be used to determine the protein region of the cell. The combination of solvent flattening and histogram matching may therefore be applied to update the density over the whole cell.

An alternate approach is to define separate solvent and protein masks, with uncertain regions excluded from either mask and allowed to keep their unmodified values.

2.3. Multiresolution modification

Multiresolution modification is an extension of the solvent flattening and histogram matching techniques which used the knowledge of electron density histograms over a wide range of resolutions to control the detail in the map at these resolutions. The method is applied as follows: a low-resolution map is calculated from a truncated set of reflections. This map is modified by solvent flattening and histogram matching using the electron density
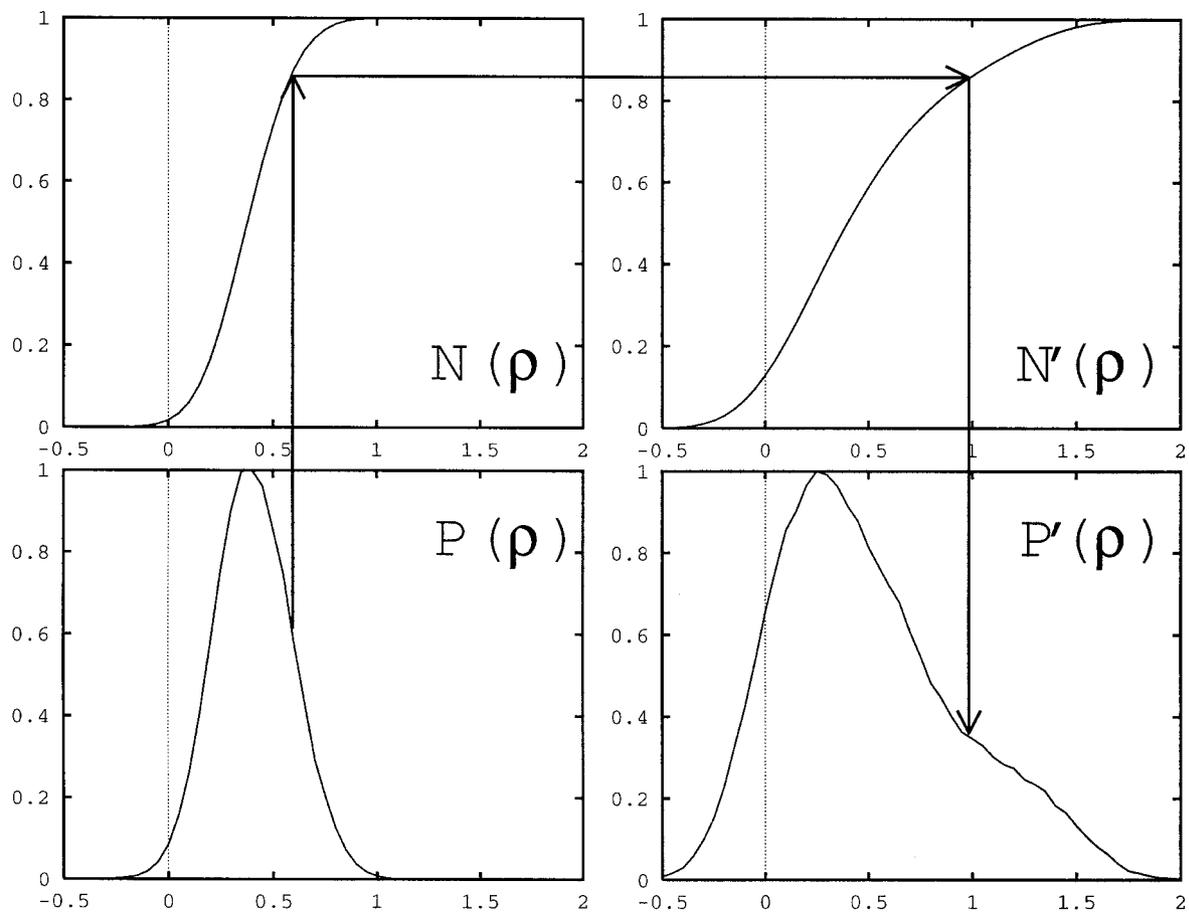
Fig. 3. Histogram matching a density distribution.

histogram at that resolution. The resulting map coefficients (which extend to higher resolution) are averaged with the initial map coefficients. These new map coefficients are then used to calculate a higher resolution map. The process is then repeated at higher resolutions until all the data has been included. Normally it is sufficient to use just two resolutions.

2.4. Averaging

Non-Crystallographic Symmetry (NCS) arises in crystals when two or more molecules are related to each other by a symmetry operation that does not relate the whole crystal lattice onto itself. Such symmetries are therefore local, as they only apply within a region of a single unit cell. A five-fold axis, for example, must be non-crystallographic, as it is not possible to tessellate objects with five-fold symmetry. Since the symmetry does not map the crystal lattice back onto itself, the individual molecules that are related by the non-crystallographic symmetry will be in different environments, therefore, the symmetry relationships are only approximate.

Non-crystallographic 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 an asymmetric unit holding multiple copies of the molecule must be proportionally larger than one holding a single copy, the number of independent diffraction magnitudes available to any given resolution is also proportionally larger. This redundancy in sampling the molecular transform leads to additional phase information.

The self-rotation and translation function are now routinely solved by use of the Patterson rotation function and translation function, typically using software such as 'AMORE' (Navaza, 1994). In cases where the Patterson translation function is unsolvable, either the whole electron density map, or a region which may be expected to contain a molecule, may be rotated using the rotation solution and used as a search model in a phased translation function (Read and 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 that 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, since every operator will map one copy of the molecule onto another copy.

Special cases arise when there are combinations of crystallographic and non-crystallographic symmetries, of proper and improper symmetries, or when a non-crystallographic 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.

Once the mask and matrices are determined, then the electron density map may be modified by averaging. This may be achieved in one or two steps. In single step averaging the density for each copy of the molecule in the asymmetric unit is replaced by the averaged density from every copy, however this becomes slow for high-order NCS (Fig. 4(a)). In two step averaging a single averaged copy of the molecule is created in an artificial cell (referred to by Rossmann as an H-cell (Rossmann et al., 1992)), and then the unit cell is reconstructed by replacing each copy of the molecule with the averaged density from the H-cell (Fig. 4(b)). This is more

(a) 1–step averaging   (b) 2–step averaging
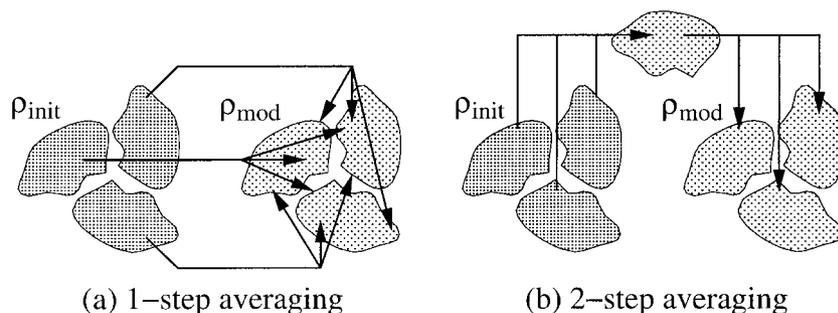
Fig. 4. Types of non-crystallographic symmetry averaging calculation.

efficient for high-order NCS, but additional errors are introduced in the use of a second interpolation step.

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 and 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 in the protein region of the solvent mask. In other cases it may be necessary to apply multi-domain averaging (Schuller, 1996), in this case the protein is divided up 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, in this case density modification is performed in each crystal form simultaneously, with averaging of the molecular density across all copies of the molecule in all the crystal forms. This is a powerful technique for phase improvement even when no phasing is available in some crystal forms.

2.5. Skeletonization

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 is 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 et al., 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.

Baker et al. (1993) have used this method to add new features to a partial model of a molecule. An efficient alternative algorithm for tracing density ridges is given by Swanson (1994).

2.6. Sayre's equation

Sayre's equation (Sayre, 1974) expresses the constraint that the atoms of a structure are equal and resolved, and has formed the foundation of direct methods for many years. In protein calculations the resolution is generally too poor for atoms to be resolved, this is apparent in Sayre's equation since the bulk of the terms required to calculate the equation are beyond the resolution limit.

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

$$\rho(\mathbf{x}) = \frac{V}{N} \sum_{\mathbf{y}} \rho^2(\mathbf{y})\psi(\mathbf{x} - \mathbf{y}), \tag{12}$$

where

$$\psi(\mathbf{x} - \mathbf{y}) = \frac{1}{V} \sum_{\mathbf{h}} \theta(\mathbf{h}) \ \exp[2\pi i\mathbf{h}(\mathbf{x} - \mathbf{y})]. \tag{13}$$

The equations are more frequently expressed in reciprocal space as a system of equations relating structure factors in magnitude and phase

$$F(\mathbf{h}) = \frac{\theta(\mathbf{h})}{V} \sum_{\mathbf{k}} F(\mathbf{k})F(\mathbf{h} - \mathbf{k}), \tag{14}$$

where

$$\theta(\mathbf{h}) = \frac{f(\mathbf{h})}{g(\mathbf{h})}$$

is the ratio of scattering factors of real and 'squared' atoms and V is the unit cell volume, and is the Fourier transform of the shape function, $\psi(\mathbf{x})$.

The residual of Sayre's equation in real space can be expressed as:

$$r_1(\mathbf{x}) = \frac{V}{N} \sum_{\mathbf{y}} \rho^2(\mathbf{y})\psi(\mathbf{x} - \mathbf{y}) - \rho(\mathbf{x}). \tag{15}$$

The gradients of this residual may be used in a Newton–Raphson calculation for the solution of the equation. To apply Sayre's equation to a particular structure, it is necessary to add additional constraints particular to that structure. Constraining the structure factor magnitudes leads to a large number of local minima, thus the approach adopted is to add a second set of linear equations, constraining the density values to equal the values obtained by density modification,

$$r_2(\mathbf{x}) = w_{\mathrm{mod}}(\mathbf{x})[\rho_{\mathrm{mod}}(\mathbf{x}) - \rho(\mathbf{x})], \tag{16}$$

where $\rho_{\mathrm{mod}}(\mathbf{x})$ is the density at a grid point obtained after averaging, solvent flattening, etc. and $w_{\mathrm{mod}}(\mathbf{x})$ is a weight indicating the strength of constraint for that grid point.

Least squares solution of the two sets of equations together may be performed rapidly using a conjugate gradient approach (Zhang and Main, 1990b) and gives rise to a map which is substantially in agreement with both Sayre's equation and the other density modification constraints.

2.7. Atomization

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

This approach has been extended in ARP/wARP program (Lamzin and 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 from poor molecular replacement models, or even directly from MIR/MAD maps (Weeks et al., 1995).

2.8. Low-density elimination

The electron density values in the protein region of the map should be positive at atomic resolution. However, the density values may be negative due to phase error or the missing of high resolution structure factors in the Fourier synthesis for non-atomic resolution data. Shiono and Woolfson (1992) proposed a low-density elimination method for phase refinement and extension. It modifies the density by,

$$
\begin{cases}
\rho'(\mathbf{r}) = \rho(\mathbf{r}) & \text{for} \quad \rho(\mathbf{r}) > 0.2\rho_c \\
\rho'(\mathbf{r}) = 0 & \text{for} \quad \rho(\mathbf{r} \leq 0.2\rho_c
\end{cases}, \tag{17}
$$

where ρ_c is the expected height of a light-atom peak in the protein region.

To eliminate the discontinuity in the density due to the above procedure, Refaat and Woolfson (1993) introduced another formula that produced a smoother modification function on the electron density,

$$
\begin{cases}
\rho'(\mathbf{r}) = \dfrac{\rho(\mathbf{r})^{n+1}}{\rho_c(\mathbf{r})^n + \rho(\mathbf{r})^n} & \text{for} \quad \rho(\mathbf{r}) > 0 \\
\rho'(\mathbf{r}) = 0 & \text{for} \quad \rho(\mathbf{r}) \leq 0
\end{cases}, \tag{18}
$$

where n is some integer greater than unity. This modification makes $\rho'(\mathbf{r})$ both smooth and continuous. The effectiveness of this low-density elimination process has been illustrated by phase refinement for the structure of avian pancreatic peptide at 1.17 Å resolution. The mean phase error was reduced from $74°$ to $39°$.

2.9. Double histogram method

In the conventional (1D) electron density histogram matching method (Zhang and Main, 1990a), a one-to-one mapping is made on the original electron density to the new electron density so that the density histogram of the modified map matches that of the ideal histogram

(see Fig. 3). The order of the electron density values is retained after histogram matching. Two grid points with the same electron density value will have the same density value after histogram matching. Therefore, the pattern of peaks and troughs in the modified map is similar to that in the original map. This is necessary in the histogram matching process, since there are many alternative ways of adjusting the electron density values for matching the electron density distribution to that of an ideal one. However, this feature is undesired, especially when spurious electron peaks need to be removed and new electron densities corresponding to missing atoms need to be generated. This coupling of original and modified maps is broken during phase combination or when other constraints are introduced. Alternative ways of decoupling the electron density order between the modified and original maps are the incorporation of other features of electron density in addition to the electron density distribution.

Refaat and Woolfson proposed a double-histogram matching method where the density modification takes into account not only the current density values at a grid point but also some characteristic of the environment of that grid point within some distance (Refaat et al., 1996). They investigated several local density environments, such as local maximum density, local minimum density and local variance. In their double-histogram matching procedure, the grid points are divided into ten groups containing the same number of grid points in each group over ten different value ranges of the local characteristic. The ten groups are modified to give different histograms, each corresponding to that obtained under the same circumstances from a structure similar to the one under investigation. Comparison of the double-histogram matching method with the normal histogram matching method with weighting of structure factors and damping, the improvement is usually an order of $4°$ in mean phase error and an increase of 0.06–0.08 in the map correlation coefficient.

2.10. Histogram moment method

The electron density distribution can be defined by the moments of electron density. In general, it requires infinite order of moments to completely define a distribution. However, the distribution can be adequately defined with only a few lower order moments. Gu et al. proposed a histogram moment method for phase improvement, which implements the density histogram constraint in reciprocal space (Gu et al., 1996). The phase improvement process was carried out by minimizing the difference between the electron density moment and its target value using Fourier transforms. The target values of moments were derived from known proteins. The protein region and the solvent region were treated separately. Tests have been carried out on two moderate sized proteins. They found that the refinement using the third moment is most effective and that refinement with higher order moments, or in solvent region, added noting useful. They suggested that the histogram moment method, which changes phases in reciprocal space, is independent of the normal histogram matching method, which changes density in real space. In general, the efficiency of the moment method for phase refinement is similar to, but slightly worse than, the normal histogram matching method. The normal histogram matching method, which matches all moments simultaneously, brings more information to bear and therefore should be more powerful. However, since the histogram

moment method modifies phase in reciprocal space rather than density in real space, it may have something extra to offer when combined with other histogram matching methods.

2.11. Ab initio phasing and density modification

Density modification is an iterative procedure that modifies the initial electron density so that it conforms to a set of given constraints corresponding to the characteristics of an ideal electron density map. The requirement of an initial electron density, thereby an initial set of phase estimates, is not due to the limitation of the density modification procedure but rather due to the phasing power of the constraints it used. Density modification methods can be used in ab initio phasing, in which case the starting map is generated from random phases. The phasing power of the electron density constraints currently available is not powerful enough to enable the phase refinement and extension at medium to high resolution starting from randomly generated phases by density modification methods. However, the constraints used in density modification have been shown to be useful in obtaining phases ab initio at very low resolution (Lunin et al., 1990, 1995, 1998). Moreover, density modification has been successfully used for ab initio phasing of small proteins at atomic resolution as part of the Shake and Bake method (Miller et al., 1993).

2.11.1. Density histogram as a constraint in reciprocal space phasing

The direct low resolution phasing method proposed by Lunin et al. (1990) used the ideal electron density histogram at very low resolution to select the best phase set among a large number of randomly generated phase sets. Due to the degeneracy of the density histogram, many phase sets may have density histograms close to the ideal histogram. Cluster analysis was used to group the viable solutions to different subsets. It was possible to identify the cluster that corresponds to the correct solution according to its compactness. Test phasing of 29 low-resolution reflections for a model structure of two molecules of carboxypetidase A resulted in a correlation coefficient of 0.94 and a mean phase error of $40°$ compared with the correct phases.

2.11.2. Structure factor amplitudes as a constraint for ab initio phasing in the Few-Atom-Model method

An alternative method to the direct low-resolution phasing method that uses the phases as variables is the Few-Atom-Model (FAM) method, which uses the positions of a very small number of large Gaussian scatters or 'pseudo-atoms' as variables (Lunin et al., 1995, 1998). The best models are selected by choosing those with the highest correlation of structure factor amplitudes corresponding to the model with the observed amplitudes. The phases calculated from these best models are analyzed by a clustering procedure leading to a few possible solutions, from which the correct solution can be recognized by simple additional criteria. The FAM method has been successfully applied to data calculated from a model ribosome crystal at 60 Å resolution and to the neutron diffraction data of the AspRS-tRNA$^{\text{Asp}}$ complex at 50 Å resolution.

2.11.3. Shake and Bake method

The phase problem can be formulated as a constrained global minimization to determine the phases of the structure factors given only their amplitudes.

For small molecules at far atomic resolution, the system is overdetermined and the constraints of positivity and atomicity are sufficient to solve the phase problem. This is embodied in the tangent formula,

$$\tan(\phi_{\mathbf{H}}) = \frac{-\sum_{\mathbf{K}} |E_{\mathbf{K}} E_{\mathbf{H+K}}|\ \sin(\phi_{\mathbf{K}} + \phi_{-\mathbf{H-K}})}{\sum_{\mathbf{K}} |E_{\mathbf{K}} E_{\mathbf{H+K}}|\ \cos(\phi_{\mathbf{K}} + \phi_{-\mathbf{H-K}})}, \tag{19}$$

which formed the basis for almost all of the Direct Method programs for small molecular structure determination. Here $K_{\mathbf{K}}$ and $E_{\mathbf{H+K}}$ are normalized structure factors, $\phi_{\mathbf{K}}$ and $\phi_{-\mathbf{H-K}}$ are phases and the index $\mathbf{K}$ represents those reciprocal space vectors with known phases. The unknown phase $\phi_{\mathbf{H}}$ is estimated individually from a set of known phases through the tangent formula.

Protein crystals rarely diffract to atomic resolution and they contain more atoms in one asymmetric unit than small molecules. The non-atomic resolution has disqualified the positivity and atomicity constraints. The large number of atoms in the asymmetric unit weakened the power of tangent formula that estimates an unknown phase using only a small set of known phases. Protein crystals also contain a large volume of contiguous solvent molecules. This also undermines the underlying assumption of random distribution of atoms in the statistically based direct methods.

In order to extend the power of direct methods to solve the phase problem of macromolecules, the minimum principle has been proposed (Hauptman, 1991). A global minimization protocol, Shake and Bake (Miller et al., 1993), has been developed to determine the phase ab initio by solving the minimum function (Debaerdemaeker and Woolfson, 1983; Hauptman, 1991),

$$R(\phi) = \frac{\sum_{\mathbf{H,\ K}} \kappa_{\mathbf{HK}} \left[\cos\ \phi_{\mathbf{HK}} - \frac{I_1(\kappa_{\mathbf{HK}})}{I_0(\kappa_{\mathbf{HK}})} \right]}{\sum_{\mathbf{H,\ K}} \kappa_{\mathbf{HK}}}, \tag{20}$$

where $\kappa_{\mathbf{HK}}$ is a triple product of normalized structure factors, $\phi_{\mathbf{HK}}$ is the phase associated with the triplet of reflections, $\mathbf{H}$, $\mathbf{K}$ and $\mathbf{H+K}$. The quotient of the two Bessel functions,

$$\frac{I_1(\kappa_{\mathbf{HK}})}{I_0(\kappa_{\mathbf{HK}})},$$

represents the expected value of $\cos \phi_{\mathbf{HK}}$.

The Shake and Bake method alternates between reciprocal space phase refinement, via the reduction of the minimum function, and the real space density modification by means of a peak-picking protocol that serves to impose electron density constraints. It starts from the set

of structure factors with phases generated from randomly positioned atoms. The starting phases are refined by minimizing the minimum function using a global binary search routine. Density filtering has been accomplished by a discrete electron-density modification protocol consisting of selecting a specified number of the highest peaks on the electron density map calculated from the refined phases. The picked atoms are used to calculate new structure factors, the phases of which are used as the starting point for a new round of global minimization. The Shake and Bake method, and its variants, such as the Half Baked method (Sheldrick and Gould, 1995), have been shown to be effective for solving the phases of small proteins at near atomic resolution (Hauptman, 1997; Weeks et al., 1994, 1995). Its potential and limitations remain to be fully explored.

2.12. Reciprocal space interpretation of density modification

Hendrickson and Lattman (1970) showed that solvent flattening can be described in terms of relationships amongst neighboring reflections in reciprocal space.

Consider the solvent flattening operation as 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:

$$\rho_{mod}(\mathbf{x}) = g_{st}(\mathbf{x}) \times \rho(\mathbf{x}) \tag{21}$$

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

If we take the Fourier transform of this equation then the product becomes a convolution, thus:

$$F_{mod}(\mathbf{h}) = \frac{1}{V} \sum_{\mathbf{k}} G_{sf}(\mathbf{k}) F(\mathbf{h} - \mathbf{k}), \tag{22}$$

where $G_{sf}(\mathbf{k})$ is the Fourier transform of the mask g_{sf}. The solvent mask g_{sf} shows the outline of the molecule with no internal detail, and 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 arising from the constraint of solvent flatness in real space. As only the terms near the origin of G_{sf} are non-zero, 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 non-linear rescaling to the current density in the protein region. Therefore the equivalent multiplier $g_{hm}(\mathbf{x})$ shows variations about 1.0, which 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. However it shows significant features to the same resolution as the current map, or beyond as the density rescaling becomes more complex. Histogram matching can therefore give phase indications to twice the resolution of the initial map, although phase indications will be weak and contain errors depending on the level of error in the initial map.

Averaging may be described as the summation of a number of reoriented copies of the

electron density within the region of the averaging mask (Rossmann et al., 1992), i.e.

$$\rho_{\mathrm{mod}}(\mathbf{x}) = g_{\mathrm{ncs}}(\mathbf{x}) \sum_i \rho_i(\mathbf{x}),\tag{23}$$

where $\rho_i(\mathbf{x})$ is the initial density $\rho(\mathbf{x})$ after transformation by the i'th NCS operator. This summation is repeated for each copy of the molecule in the whole unit cell. The function $G_{\mathrm{ncs}}(\mathbf{h})$ 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 $G_{\mathrm{ncs}}(\mathbf{h})$ in reciprocal space is greater.

Electron density and structure factor magnitudes

Solvent flattening function $g_{sf}(\mathbf{x})$ and $G_{sf}(\mathbf{h})$

Histogram matching function $g_{hm}(\mathbf{x})$ and $G_{hm}(\mathbf{h})$

Averaging function $g_{av}(\mathbf{x})$ and $G_{av}(\mathbf{h})$
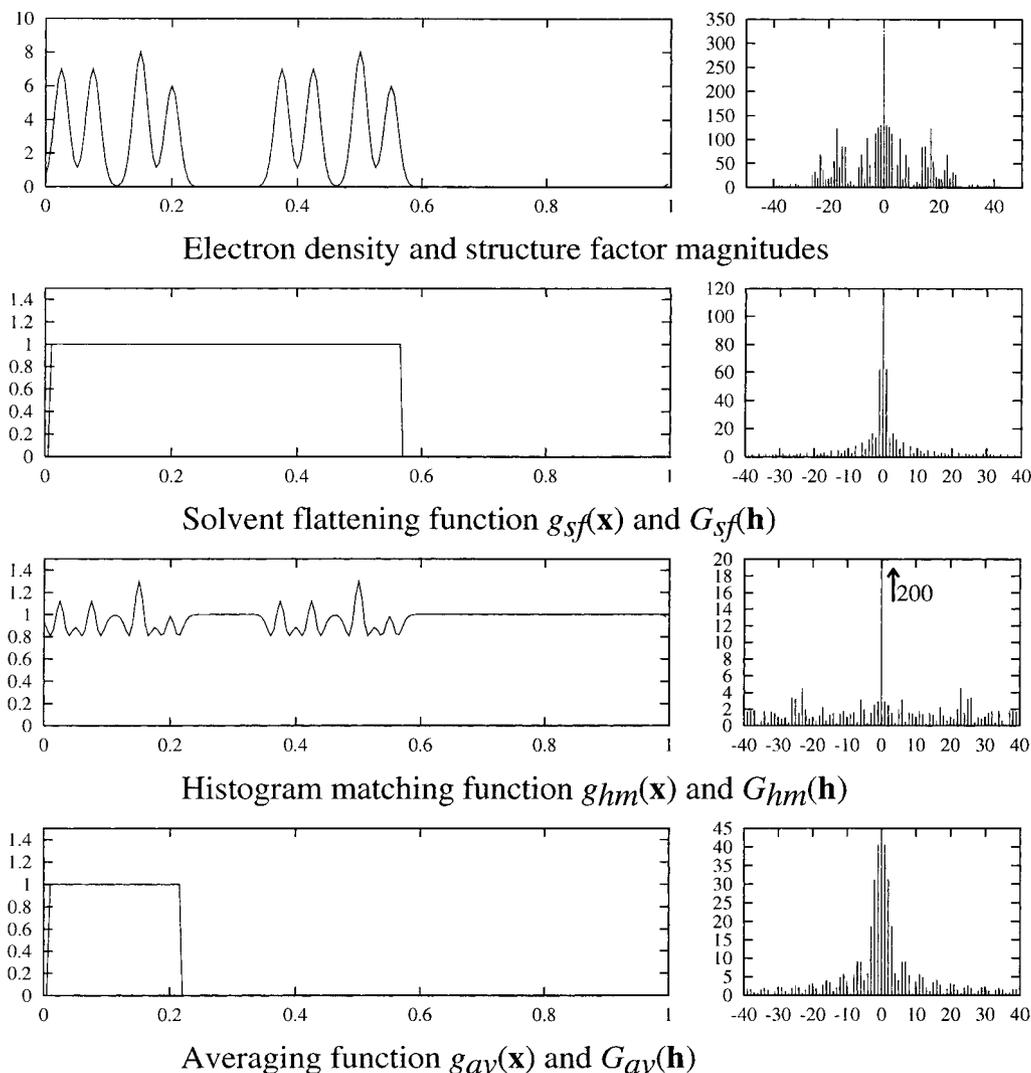
Fig. 5. The functions of $g(\mathbf{x})$ and $G(\mathbf{h})$ for solvent flattening, histogram matching, and averaging.

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. However as higher resolution terms are considered, 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. 5 for a simple one-dimensional structure.

3. Phase combination

Once a modified map has been obtained, modified phases and magnitudes may be obtained by inverse Fourier transforms. The modified phases are usually combined with the initial phases by multiplication of their probability distributions.

$$P_{\text{new}}(\phi(\mathbf{h})) = P_{\text{init}}(\phi(\mathbf{h}))P_{\text{mod}}(\phi(\mathbf{h})). \tag{24}$$

The probability distribution for the initial phases is usually described in terms of Hendrickson–Lattman coefficients (Hendrickson and Lattman, 1970) or by a centroid phase and figure-of-merit. 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 some form of Sim weighting scheme.

Recombination with the initial phases assumes independence between the initial and modified phases, and is a source of difficulties. Most density modification constraints are underdetermined and thus phase recombination with the initial data is vital. However this leads to a strong bias with respect to the initial phases. The exception is when high-order NCS is present, in this case the combination of NCS and observed magnitudes is sufficient to determine the phases, 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 magnitude and phase.

3.1. Sim/Sigma-a weighting

The phase probability distributions for density modified phases have generally been estimated under assumptions that were made for the combination of a partial atomic model with experimental data. The assumption is that the calculated magnitudes and phases arise from a density map in which some atoms are present and correctly positioned, and the remaining atoms are completely absent (Sim, 1959). Thus the difference between the true structure factor and the current value must be the effective structure factor due to the missing density alone. If the phase of this quantity is random and the magnitude is drawn from a Wilson distribution (Wilson, 1949) the following expression is obtained:

$$P_{\text{model}}(\phi) = \exp(X \, \cos(\phi - \phi_{\text{model}})). \tag{25}$$

The error estimate for the phase depends on the effective amount of missing structure is estimated on the basis of the agreement of the modified magnitudes with their measured values

$$X = \frac{2 \mid F_{\text{obs}} \mid\mid F_{\text{model}} \mid}{\sum_Q} \qquad (26)$$

where $\sum_Q$ may be estimated by a number of means, for example, (Bricogne, 1976):

$$\sum_Q = \langle \mid F_{\text{obs}} \mid^2 - \mid F_{\text{model}} \mid^2 \rangle, \qquad (27)$$

where the average in Eq. (27) is normally taken over all reflections at a particular resolution. A more sophisticated approach is the σ_{a} method (Read, 1986) which allows for errors in the atomic model and has also been used in density modification.

Although these approaches have been applied with some success, the assumption in Eq. (24) that the density modified magnitudes and phases are independent of the initial values is invalid. Since the density constraints are typically under-determined it is possible to achieve an arbitrarily good agreement between the model magnitudes 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 (for example, 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 is included. The figures of merit of reflections that undergo fewer cycles of phase recombination will be correspondingly smaller (Leslie, 1987; Zhang and Main, 1990a). In averaging calculations where considerable phase information is available from high order NCS, it is still typically necessary to perform phase extension over hundreds of cycles and adding a very thin resolution shell of new reflections at each cycle (Chapman, 1998).

3.2. Reflection-omit

In order for the modified phases to be independent of the initial phases, the new information must be determined from a different source from the initial phases. Since phasing calculations are performed on a reflection-by-reflection basis, the initial phase estimates for different reflections from a MIR or MAD experiment are independent. Thus the modified phase for a reflection can depend on the initial phases of every other reflection in the diffraction pattern. This is achieved through the reflection-omit scheme.

The reflections are divided into (typically 10 or 20) sets, and density modification calculations are performed excluding all reflections from each set in turn. The reflections from each of the omitted sets are combined to give a complete new data set. This data should be less dependent on the original magnitudes, therefore the modified magnitudes 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 (Cowtan and Main, 1996). In the case of averaging calculations however, the reflection-omit approach makes little difference, since the averaging constraints

usually contain sufficient phase information to make the modified map largely independent of the initial map. The resulting phase estimates and their probability distributions are unbiased after a single cycle of density modification, but on subsequent cycles the phases of individual reflections are no longer independent, and so bias remains a problem if the calculation is repeated over many cycles.

The free R-value (Brünger, 1992) has been used in a cross-validated density modification protocol to prevent over-fitting of data and to optimize the parameters in density modification, such as solvent flattening (Roberts and Brünger, 1995). This approach is subject to the same problems of bias as the estimation of phase errors, and so is of limited use.

3.3. The gamma-correction and solvent flipping

Abrahams and 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 in the case where the dependence between the initial and modified maps is linear (Abrahams, 1997).

In Eq. (22) solvent flattening is represented in reciprocal space by convolution of the structure factors with a function $G(\mathbf{h})$. If the origin term of G is set to zero, then the modified structure factor $F_{\text{mod}}(\mathbf{h})$ will depend upon 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 $g_\gamma(\mathbf{x})$:

$$G_\gamma(\mathbf{h}) = \begin{cases} 0, & h = 0 \\ G(\mathbf{h}), & h \neq 0 \end{cases} \tag{28}$$

then

$$g_\gamma(\mathbf{x}) = g(\mathbf{x}) - \overline{g(\mathbf{x})}. \tag{29}$$

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

$$\rho_{\text{mod}}(\mathbf{x}) = g_\gamma(\mathbf{x}) \times \rho(\mathbf{x}). \tag{30}$$

In the case of a solvent flattening calculation, $g_\gamma(\mathbf{x})$ will be equal to $g(\mathbf{x})$ minus the fraction of the cell which is protein, thus in the case of a protein 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 can be determined, then the flipping calculation may be alternately performed by subtracting this value, multiplied by the initial map, from the final map. The origin term of G is the γ-correction of Abrahams (1997).

Extension of the γ-correction to averaging calculations is trivial, however other density modifications present some problems. The theoretical formula for γ of Abrahams fails for histogram matching. Cowtan (1999) describes a method for estimating the γ-correction for an

arbitrary density modification by applying a perturbation to the initial map and examining the scale of the resulting perturbation in the modified map. This perturbation γ method works well for histogram matching and multi-resolution modification, and any combination of these methods with solvent flattening or averaging.

3.4. The simultaneous application of 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. Therefor, Zhang and Main suggested a protocol that combines individual constraints for phase improvement (Zhang, 1993; Zhang et al., 1997; Zhang and Main, 1990b).

One obvious way of implementing several constraints is to apply them one after the other on 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 how the density should be modified. An alternative way of implementing various constraints is through simultaneous application. The density solution that satisfies all the constraints is obtained by a global minimization procedure (Main, 1990; Zhang and Main, 1990b).

In order for a modified electron density map to satisfy all constraints at the same time, we have to solve a system of simultaneous equations which includes all the residual Eqs. (15) and (16),

$$r_1(\mathbf{x}) = \frac{V}{N} \sum_{\mathbf{y}} \rho^2(\mathbf{y}) \psi(\mathbf{x} - \mathbf{y}) - \rho(\mathbf{x}),$$

$$r_2(\mathbf{x}) = w_{\mathrm{mod}}(\mathbf{x})[\rho_{\mathrm{mod}}(\mathbf{x}) - \rho(\mathbf{x})] \tag{31}$$

where w_{mod} is the relative weight between Sayre's Eq. (14) and the constraints from density modification (16).

Eqs. (31) represent a system of non-linear simultaneous equations with as many unknowns as the number of grid points in the asymmetric unit of the map and twice as many equations as unknowns. The functions $\rho_{\mathrm{mod}}(\mathbf{x})$ and $\psi(\mathbf{x})$ are both known. The least squares solution, using either the full matrix or the diagonal approximation, is obtained using Newton–Raphson technique for FFTs as described by Main (1990).

A flow chart of the protocol for the simultaneous application of constraints is shown in Fig. 6.
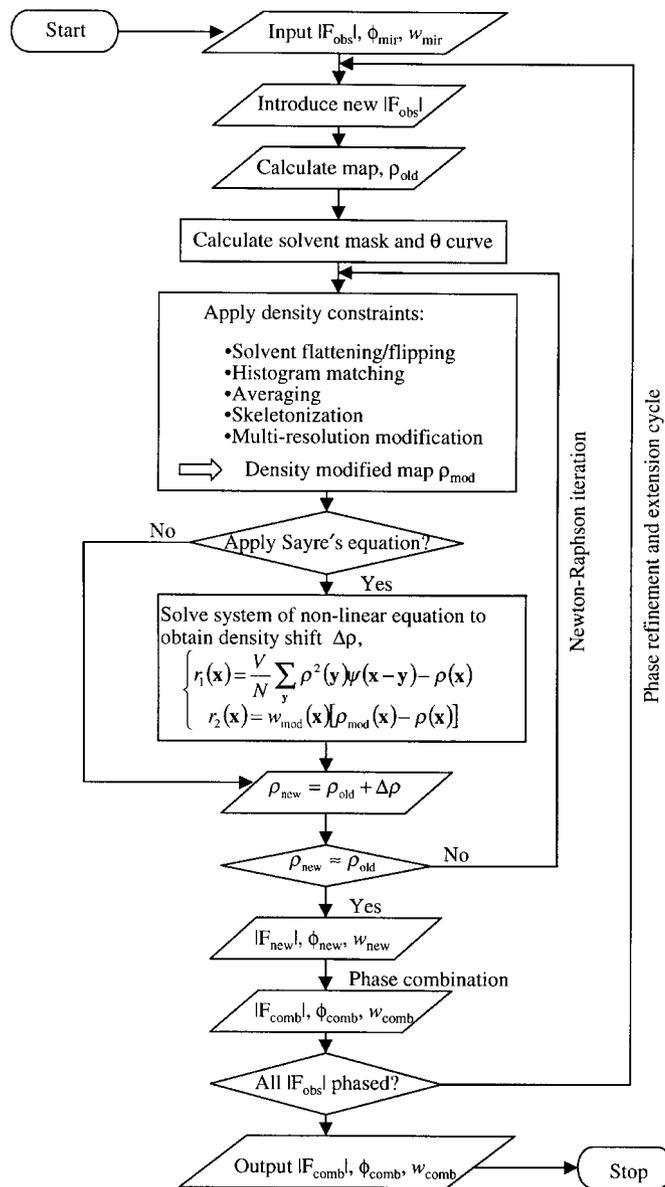
Fig. 6. Flow chart for phase improvement by the simultaneous application of electron density constraints.

4. Example

As an example, density modification was applied to the structure of RNAse from *Streptomyces aureofaciens* (Sevcik et al., 1991). The structure consists of two molecules of 96 amino acids in the asymmetric unit, including one α-helix and a twisted 3-strand anti-parallel β-sheet. The structure has been solved using multiple isomorphous derivatives and refined to
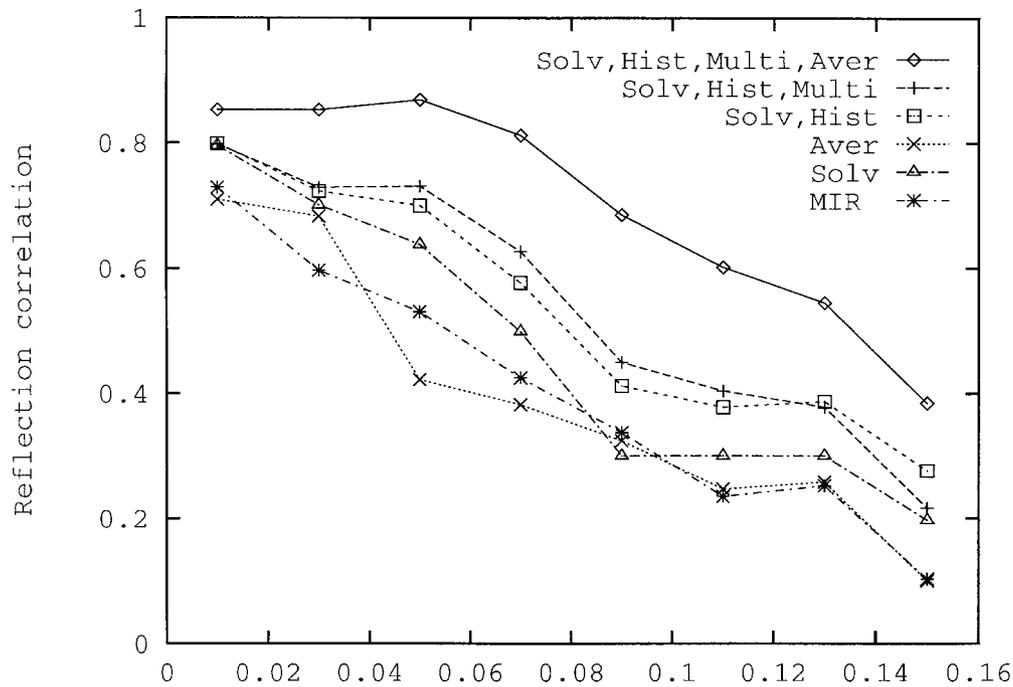
Fig. 7. Density modification results on RNAse-SA using combinations of solvent flattening, histogram matching, multiresolution modification and averaging.

1.8 Å. A poor set of starting phases was generated using the two weakest derivatives, giving MIR phases to 3.1 Å and very weak SIR phasing to 2.5 Å. The MIR map was uninterpretable.

The MIR and density modified phases are compared by plotting the mean of the cosine of the phase error, weighted by the figure-of-merit and structure factor magnitude, as a function of resolution (over all reflections this is equivalent to map correlation). The results of density modification by various techniques and their combinations, using the perturbation gamma correction for bias reduction over five cycles, are shown in Fig. 7.

Solvent flattening or averaging alone improves the phases slightly at low resolution, but does not lead to significant phase extension. The combination of histogram matching and solvent flattening improves the low-resolution phases and gives significant phase extension to higher resolutions. Adding multiresolution modification gives a further slight improvement across the resolution range. Adding averaging as well gives a dramatic improvement over the whole resolution range. These results have demonstrated that each constraint contains some degree of independent phasing information and the constrains are synergistic in phase improvement. The simultaneous application of all available constraints gave the best result in phase improvement.

Acknowledgements

material used in this paper was originally prepared for International Tables for Crystallography, Volume F (Kluwer Academic Press; Dodrecht, Boston, London).

References

Abrahams, J.P., 1997. Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. Acta Cryst. D53, 371–376.

Abrahams, J.P., Leslie, A.G.W., 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Cryst. D52 (Part 1), 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. D49, 429–439.

Bricogne, G., 1974. Geometric redundancy in intensity data and their use for phase determination. Acta Cryst. A30, 395–405.

Bricogne, G., 1976. Methods and Programs for Direct-Space Exploitation of Geometric Redundancies. Acta Cryst. A32, 832–847.

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

Chapman, M.S., 1998. Introduction to the use of non-crystallographic symmetry in phasing. In: Fortier, S. (Ed.), Direct Methods for Solving Macromolecular Structures. Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 99–108.

Cowtan, K.D., 1999. Error Estimation and Bias Correction in Phase Improvement Calculations. Acta Cryst D, in pres.˜.

Cowtan, K.D., Main, P., 1996. Phase combination and cross validation in iterated density modification calculations. Acta Cryst. D52, 43–48.

Cowtan, K.D., Main, P., 1998. Miscellaneous algorithms for density modification. Acta Cryst. D54, 487–493.

Debaerdemaeker, T., Woolfson, M.M., 1983. On the application of phase relationships to complex structures. XXII. Techniques for random phase refinement. Acta Cryst. A39, 193–196.

Greer, J., 1985. Computer skeletonization and automatic electron-density map analysis. In: Wyckoff, H.W., Hirs, C.H.W., Timasheff, S.N. (Eds.), Diffraction Methods for Biological Macromolecules, Vol. 115. Academic Press, Orlando, pp. 206–224.

Gu, Y.X., Woolfson, M.M., Yao, J.X., 1996. Direct-space methods in phase extension and phase refinement. V. The histogram moments method. Acta Cryst. D52 (Part 6), 1114–1118.

Harrison, R.W., 1988. Histogram Specification as a Method of Density Modification. Appl. Cryst. 21, 949–952.

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

Hauptman, H., 1991. Phasing Methods for Protein Crystallography. In: Moras, D., Podjarny, A.D., Thierry, J.C. (Eds.), Crystallographic Computing 5: From Chemistry to Biology. Oxford University Press, Oxford, pp. 324–332.

Hauptman, H., 1997. A Minimal Principle in the Phase Problem. Current Opinion in Structural Biology 7 (5), 672–680.

Hendrickson, W.A., Lattman, E.E., 1970. Representation of Phase Probability Distributions for Simplified Combination of Independent Phase Information. Acta Cryst. B26, 136–143.

Hoppe, W., Gassmann, J., 1968. Phase Correction, a New Method to Solve Partially Known Structures. Acta Cryst. B24, 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 in Enzymology 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. A43, 134–136.

Lunin, V.Y., 1988. Use of the Information on Electron Density Distribution in Macromolecules. Acta Cryst. A44, 144–150.

Lunin, V.Y., Lunina, N.L., Petrova, T.E., Urzhumtsev, A.G., Podjarny, A.D., 1998. On the ab initio solution of the phase problem for macromolecules at very low resolution. II. generalized likelihood based approach to cluster discrimination. Acta Cryst. D54, 726–734.

Lunin, V.Y., Lunina, N.L., Petrova, T.E., Vernoslova, E.A., Urzhumtsev, A.G., Podjarny, A.D., 1995. On the ab initio solution of the phase problem for macromolecules at very low resolution: the few atoms model method. Acta Cryst. D51, 896–903.

Lunin, V.Y., Urzhumtsev, A.G., Skovoroda, T.P., 1990. Direct low-resolution phasing from electron-density histograms in protein crystallography. Acta Cryst. A46, 540–544.

Main, P., 1990. The Use of Sayre's Equation with Constraints for the Direct Determination of Phases. Acta Cryst. A46, 372–377.

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 (5100), 1430–1433.

Navaza, J., 1994. AMoRe: an Automated Package for Molecular Replacement. Acta Cryst. A50, 157–163.

Read, R.J., 1986. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. Acta Cryst. A42, 140–149.

Read, R.J., Schierbeek, A.J., 1988. A Phased Translation Function. J. Appl. Cryst. 21, 490–495.

Refaat, L.S., Tate, C., Woolfson, M.M., 1996. Direct-Space Methods in Phase Extension and Phase Refinement. IV. The Double-Histogram Method. Acta Cryst. D52, 252–256.

Refaat, L.S., Woolfson, M.M., 1993. Direct-space methods in phase extension and phase determination. II. Developments of low-density elimination. Acta Cryst. D49, 367–371.

Roberts, A.L.U., Brünger, A.T., 1995. Phase improvement by cross-validated density modification. Acta Cryst. D51, 990–1002.

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

Sayre, D., 1974. Least-Squares Phase Refinement. II. High-Resolution Phasing of a Small Protein. Acta Cryst. A30, 180–184.

Schuller, D.J., 1996. MAGICSQUASH – More Versatile Non-Crystallographic Averaging With Multiple Constraints. Acta Cryst. D52 (Part 3), 425–434.

Sevcik, J., Dodson, E., Dodson, G.G., 1991. Determination and restrained least-squares refinement of the crystal structures of ribonuclease Sa and its eomplex with 3′guanylic acid at 1.8A resolution. Acta Cryst. B47, 240–253.

Sheldrick, G.M., Gould, R.O., 1995. Structure solution by iterative peaklist optimization and tangent expansion in space group P1. Acta Cryst. B51, 423–431.

Shiono, M., Woolfson, M.M., 1992. Direct-space methods in phase extension and phase determination. I. Low-density elimination. Acta Cryst. A48, 451–456.

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. D50, 695–708.

Vellieux, F.M.D., 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: Wyckoff, H.W., Hirs, C.H.W., Timasheff, S.N. (Eds.), Diffraction Methods for Biological Macromolecules, Vol. 115. Academic Press, Orlando, pp. 90–113.

Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P., Miller, R., 1994. Application of the Minimum Principle to peptide structures. Acta Crystallographica—Section a—Foundations of Crystallography 50 (Part 2), 210–220.

Weeks, C.M., Hauptman, H.A., Chang, C.S., Blessing, R.H., Teeter, M.M., Miller, R., 1995. R. Crambin: a direct solution for a 400-atom structure. Acta Cryst. D51, 33–38.

Wilson, A.J.C., 1949. The Probability Distribution of X-ray Intensities. Acta Cryst. 2, 318–321.

Woolfson, M.M., 1987. Direct Methods – from Birth to Maturity. Acta Cryst. A43, 593–612.

Zhang, K.Y.J., 1993. SQUASH – Combining Constraints for Macromolecular Phase Refinement and Extension. Acta Cryst. D49, 213–222.

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

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. A46, 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. A46, 377–381.