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

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

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 $\Omega$ 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, $\Omega$,

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

Here, $\Omega$ is a function of $\omega, \Omega = f(\omega)$, 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, $\omega$, 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,

$$\begin{aligned}
\frac{\partial\rho}{\partial x} &= -\frac{2\pi i}{V} \sum_{hkl} h F_{hkl} \exp[-2\pi i(hx + ky + lz)] \\
\frac{\partial\rho}{\partial y} &= -\frac{2\pi i}{V} \sum_{hkl} k F_{hkl} \exp[-2\pi i(hx + ky + lz)] \qquad (5.1.2.29) \\
\frac{\partial\rho}{\partial z} &= -\frac{2\pi i}{V} \sum_{hkl} l F_{hkl} \exp[-2\pi i(hx + ky + lz)],
\end{aligned}$$

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, $\omega$,

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

#### 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. 15.1.2.4*c*). 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. 15.1.2.4*d*). 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

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.



references