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, $\overline{\rho}_{\rm solv}$, and the solvent volume of the cell, $V_{\rm solv}$. The $F(000)$ term can then be evaluated from equations (15.1.2.15) and (15.1.2.19):

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

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

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

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 subregion 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*).

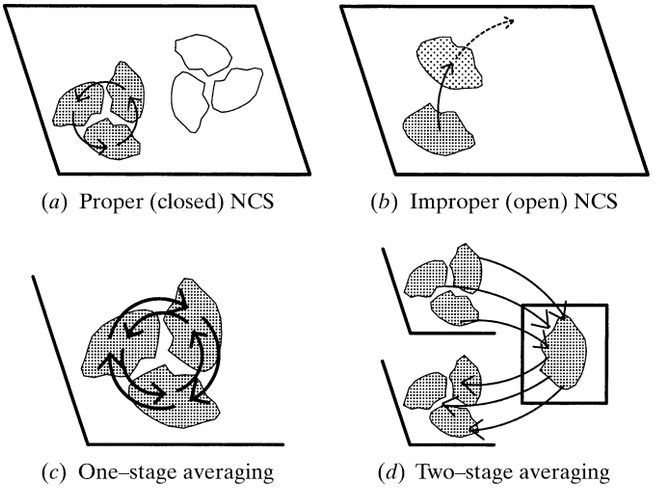

Fig. 15.1.2.4. Types of noncrystallographic symmetry and averaging calculation.

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

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

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 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 $\omega, \Omega = f(\omega)$, 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, 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,

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

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, ω,

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

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