particle) in the *h*-cell is related to the 'reference' molecule in the *p*-cell. The known *p*-cell crystallographic symmetry then permits the complete construction of the *p*-cell structure from whatever is the current *h*-cell electron-density representation of the molecule.

The *h*-cell is used to represent the density of a molecule in the standard orientation obtained by averaging all the noncrystallographic units in the *p*-cell. While density within a specific molecule will tend to be reinforced by the averaging procedure, the density outside the molecular boundaries will tend to be diminished. Thus, by averaging into the *h*-cell, the molecular envelope is revealed automatically. Indeed, the greater the NCS, the greater the clarity of the molecular boundary. Hence, the averaged molecule in the *h*-cell can be used to define a molecular mask in the *p*-cell automatically.

Averaging into the *h*-cell is also useful for displaying the molecule in a standard orientation (*i.e.* obtaining the electron-density distribution on skew planes). Thus, it is possible to display the molecule, for instance, with sections perpendicular to a molecular twofold axis, and to position the molecular symmetry axes accurately. From this, it is then easy to define the limits of the molecular asymmetric unit (Fig. 13.4.4.1). Hence, it is possible to save a great deal of computing time by evaluating the electron density in the *h*-cell only at those grid points within and immediately surrounding the noncrystallographic asymmetric unit.

## 13.4.5. Combining crystallographic and noncrystallographic symmetry

Transformations will now be described which relate noncrystallographically related positions distributed among several fragmented copies of the molecule in the asymmetric unit of the *p*-cell and between the *p*-cell and the *h*-cell.

### 13.4.5.1. *General considerations*

Let $\mathbf{Y}$ and $\mathbf{X}$ be position vectors in a Cartesian coordinate system whose components have dimensions of length, in the *p*- and *h*-cells, which utilize the same origin as the fractional coordinates, $\mathbf{y}$ and $\mathbf{x}$, respectively. Let $[\beta_p]$ and $[\alpha_h]$ be 'orthogonalization' and 'de-orthogonalization' matrices in the *p*- and *h*-cells, respectively (Rossmann & Blow, 1962). Then

$$\begin{aligned} \mathbf{Y} &= [\beta_p]\mathbf{y} \quad \text{and} \quad \mathbf{x} = [\alpha_h]\mathbf{X}, \\ [\alpha_p] &= [\beta_p]^{-1} \quad \text{and} \quad [\alpha_h] = [\beta_h]^{-1}. \end{aligned} \tag{13.4.5.1}$$

Thus, for instance, $[\alpha_h]$ denotes a matrix that transforms a Cartesian set of unit vectors to fractional distances along the unit-cell vectors $\mathbf{a}_h, \mathbf{b}_h, \mathbf{c}_h$.

Let the Cartesian coordinates $\mathbf{Y}$ and $\mathbf{X}$ be related by the rotation matrix $[\omega]$ and the translation vector $\mathbf{D}$ such that

$$\mathbf{X} = [\omega]\mathbf{Y} + \mathbf{D}. \tag{13.4.5.2}$$

If the molecules are to be averaged among different unit cells, then each *p*-cell must be related to the standard *h*-cell orientation by a different $[\omega]$ and $\mathbf{D}$. Then, from (13.4.5.1) and (13.4.5.2)

$$\mathbf{X} = [\omega][\beta_p]\mathbf{y} + \mathbf{D}. \tag{13.4.5.3}$$

Now, if $[\omega]$ represents the rotational relationship between the 'reference' molecule, $m = 1$, in the *p*-cell with respect to the *h*-cell, then from (13.4.5.3)

$$\mathbf{X} = [\omega][\beta_p]\mathbf{y}_{m=1} + \mathbf{D},$$

where $\mathbf{y}_m$ refers to the fractional coordinates of the *m*th molecule in the *p*-cell.

Assuming there is only one molecule per asymmetric unit in the *p*-cell, let the *m*th molecule in the *p*-cell be related to the reference molecule by the crystallographic rotation $[\mathrm{T}_m]$ and translational operators $\mathbf{t}_m$, such that

$$\mathbf{y}_m = [\mathrm{T}_m]\mathbf{y}_{m=1} + \mathbf{t}_m. \tag{13.4.5.4}$$

For convenience, all translational components will initially be neglected in the further derivations below, but they will be reintroduced in the final stages. Hence, from (13.4.5.3) and (13.4.5.4)

$$\mathbf{X} = \{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}\mathbf{y}_m. \tag{13.4.5.5}$$

Further, if $\mathbf{X}_n$ refers to the *n*th subunit within the molecule in the *h*-cell, and similarly if $\mathbf{y}_{m,n}$ refers to the *n*th subunit within the *m*th molecule of the *p*-cell, then from (13.4.5.5)

$$\mathbf{X}_n = \{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}\mathbf{y}_{m,n}. \tag{13.4.5.6}$$

Finally, the rotation matrix $[\mathrm{R}_n]$ is used to define the relationship among the $N$ ($N = 2$ for a dimer, 4 for a 222 tetramer, 60 for an icosahedral virus *etc.*) noncrystallographic asymmetric units of the molecule within the *h*-cell. Then

$$\mathbf{X}_n = [\mathrm{R}_n]\mathbf{X}_{n=1}. \tag{13.4.5.7}$$

### 13.4.5.2. *Averaging with the p-cell*

Consider averaging the density at $N$ noncrystallographically related points in the *p*-cell and replacing that density into the *p*-cell. By substituting for $\mathbf{X}_n$ and $\mathbf{X}_{n=1}$ in (13.4.5.7) and using (13.4.5.6),

$$\{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}\mathbf{y}_{m,n} = [\mathrm{R}_n]\{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}\mathbf{y}_{m,n=1},$$

or

$$\mathbf{y}_{m,n} = \{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}^{-1} \times [\mathrm{R}_n]\{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}\mathbf{y}_{m,n=1}. \tag{13.4.5.8}$$

Now set

$$\begin{aligned} [\mathrm{E}_{m,n}] &= \{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}^{-1}[\mathrm{R}_n]\{[\omega][\beta_p][\mathrm{T}_m^{-1}]\} \\ &= [\mathrm{T}_m][\alpha_p][\omega^{-1}][\mathrm{R}_n][\omega][\beta_p][\mathrm{T}_m^{-1}], \end{aligned} \tag{13.4.5.9}$$

giving

$$\mathbf{y}_{m,n} = [\mathrm{E}_{m,n}]\mathbf{y}_{m,n=1} + \mathbf{e}_{m,n}, \tag{13.4.5.10}$$

where $\mathbf{e}_{m,n}$ is the corresponding translational element. Note that multiplication by $[\mathrm{E}_{m,n}]$ thus corresponds to the following sequence of transformations: (1) placing all the crystallographically related subunits into the reference orientation with $[\mathrm{T}_m^{-1}]$; (2) 'orthogonalizing' the coordinates with $[\beta_p]$; (3) rotating the coordinates into the *h*-cell with $[\omega]$; (4) rotating from the reference subunit of the molecule of the *h*-cell with $[\mathrm{R}_n]$; (5) rotating these back into the *p*-cell with $[\omega^{-1}]$; (6) 'de-orthogonalizing' in the *p*-cell with $[\alpha_p]$; and (7) placing these back into each of the $M$ crystallographic asymmetric units of the *p*-cell with $[\mathrm{T}_m]$.

The translational elements, $\mathbf{e}_{m,n}$, can now be evaluated. Let $\mathbf{s}_{p,m}$ be the fractional coordinates of the centre (or some arbitrary position) of the *m*th molecule in the *p*-cell; hence, $\mathbf{s}_{p,m=1}$ denotes the molecular centre position of the reference molecule in the *p*-cell. If $\mathbf{s}_{p,m}$ is at the intersection of the molecular rotation axes, then it will be the same for all *n* molecular asymmetric units. Therefore, it follows from (13.4.5.10) that

$$\mathbf{e}_{m,n} = \mathbf{s}_{p,m} - [\mathrm{E}_{m,n}]\mathbf{s}_{p,m=1}, \tag{13.4.5.11$a$}$$

or

$$\mathbf{y}_{m,n} = [\mathrm{E}_{m,n}]\mathbf{y}_{m,n=1} + (\mathbf{s}_{p,m} - [\mathrm{E}_{m,n}]\mathbf{s}_{p,m=1}). \tag{13.4.5.11$b$}$$

Equation (13.4.5.11*b*) can be used to find all the $N$ noncrystallographic asymmetric units within the crystallographic asymmetric

unit of the *p*-cell. Thus, this is the essential equation for averaging the density in the *p*-cell and replacing it into the *p*-cell.

#### 13.4.5.3. *Averaging the p-cell and placing the results into the h-cell*

Consider averaging the density at $N$ noncrystallographically related points in the *p*-cell and placing that result into the *h*-cell. From (13.4.5.7), multiplying by $[\alpha_h]$,

$$[\alpha_h][\mathbf{X}_{n=1}] = [\alpha_h][\mathrm{R}_n^{-1}]\mathbf{X}_n.$$

From (13.4.5.1) and (13.4.5.6),

$$\mathbf{x}_{n=1} = [\alpha_h][\mathrm{R}_n^{-1}]\{[\omega][\beta_p][\mathrm{T}_m^{-1}]\}\mathbf{y}_{m,\,n}. \qquad (13.4.5.12)$$

Since it is only necessary to place the reference molecule of the *p*-cell into the *h*-cell, it is sufficient to consider the case when $m = 1$, in which case $[\mathrm{T}_m^{-1}]$ is the identity matrix [I]. It then follows, by inversion, that

$$\begin{aligned}\mathbf{y}_{m=1,\,n} &= \{[\omega][\beta_p]\}^{-1}[\mathrm{R}_n][\beta_h]\mathbf{x}_{n=1}\\ &= [\alpha_p][\omega^{-1}][\mathrm{R}_n][\beta_h]\mathbf{x}_{n=1},\end{aligned}$$

which corresponds to: (1) 'orthogonalizing' the *h*-cell fractional coordinates with $[\beta_h]$; (2) rotating into the *n*th noncrystallographic unit within the molecule using $[\mathrm{R}_n]$; (3) rotating into the *p*-cell with $[\omega^{-1}]$; and (4) 'de-orthogonalizing' into fractional *p*-cell coordinates with $[\alpha_p]$.

Now, if $\mathbf{s}_h$ is the molecular centre in the *h*-cell (usually $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$), then

$$\mathbf{y}_{m=1,\,n} = [\mathrm{E}'_{m=1,\,n}]\mathbf{x} + (\mathbf{s}_{p,\,m=1} - [\mathrm{E}'_{m=1,\,n}]\mathbf{s}_h),$$

and

$$[\mathrm{E}'_{m=1,\,n}] = [\alpha_p][\omega^{-1}][\mathrm{R}_n][\beta_h]. \qquad (13.4.5.13)$$

Equation (13.4.5.13) determines the position of the $N$ noncrystallographically related points $\mathbf{y}_{m=1,\,n}$ in the *p*-cell whose average value is to be placed at $\mathbf{x}$ in the *h*-cell.

### 13.4.6. Determining the molecular envelope

Various techniques are available for determining the molecular envelope within which density can be averaged and outside of which the solvent can be flattened.

(1) By assumption of a simple geometric shape, such as a sphere. This is frequently used for icosahedral viruses.

(2) By manual inspection of a poor electron-density map which, nevertheless, gives some guidance as to the molecular boundaries. A variety of interactive graphical programs are available to help define the molecular boundary.

(3) By use of a homologous structure or other information, such as a cryo-electron-microscopy (cryo-EM) reconstruction at low resolution. The information about a homologous structure may be either in the form of an electron-density grid or, often more conveniently, as an atomic model.

(4) By inspection of an averaged map which should have weaker density beyond the limits of the molecular boundary where the NCS is no longer true.

Procedures (2) and (3) are advisable when the NCS redundancy is low. Procedure (4) works well when the NCS redundancy is four or higher. The crystallographic asymmetric unit is likely to contain bits and pieces of molecules centred at various positions in the unit cell and neighbouring unit cells. Therefore, it is necessary to associate each grid point within the *p*-cell crystallographic asymmetric unit to a specific molecular centre or to solvent.

If the molecular-boundary assignments are to be made automatically, then the following procedure can be used. The number, $M$, of such molecules can be estimated by generating all centres, derived from the given position of the centre for the reference molecule, $\mathbf{s}_{p,\,n=1}$, and then determining whether a molecule of radius $R_{\rm out}$ would impinge on the crystallographic asymmetric unit within the defined boundaries. Here, $R_{\rm out}$ is a liberal estimate of the molecular radius. The corresponding rotation matrices $[\mathrm{E}_{m,\,n}]$ and translation vectors $\mathbf{e}_{m,\,n}$ can then be computed from (13.4.5.9) and (13.4.5.11*a*).

Any grid point whose distance from all $M$ centres is greater than $R_{\rm out}$ can immediately be designated as being in the solvent region. For other grid points, it is necessary to examine the corresponding *h*-cell density. From (13.4.5.12), it follows that (setting $n = 1$)

$$\mathbf{x} = [\mathrm{E}''_{m,\,n=1}]\mathbf{y}_m + (\mathbf{s}_h - [\mathrm{E}''_{m,\,n=1}]\mathbf{s}_{p,\,m}),$$

where

$$[\mathrm{E}''_{m,\,n=1}] = [\alpha_h][\mathrm{R}_n^{-1}][\omega][\beta_p][\mathrm{T}_m^{-1}] \qquad (13.4.6.1)$$

($n$ can be set to 1, since the *h*-cell presumably contains an averaged molecular electron density, in which case it does not matter which molecular asymmetric unit is referenced). Thus, (13.4.6.1) can be used to determine the electron density at $\mathbf{y}_m$ by inspecting the corresponding interpolated density, $\rho(\mathbf{x})$, at $\mathbf{x}$ in the *h*-cell. Transfer of the electron density, $\rho(\mathbf{x})$, from the *h*-cell to the *p*-cell using (13.4.6.1) is often useful to obtain an initial structure. However, to determine a suitable mask, it is useful to evaluate a modified electron density, $\langle\rho(\mathbf{x})\rangle$, (see below) for the grid points immediately around $\mathbf{x}$ in the *h*-cell.

A variable parameter 'CRIT' can be specified to establish the distribution of grid points that are within the molecular envelope. When the modified electron density, $\langle\rho(\mathbf{x})\rangle$, is less than CRIT, the corresponding grid point at $\mathbf{y}$ is assumed to be in solvent. Otherwise, when $\langle\rho(\mathbf{x})\rangle$ exceeds CRIT, the grid point at $\mathbf{y}$ is assigned to that molecule which has the largest $\langle\rho(\mathbf{x})\rangle$. If the percentage of grid points which might be assigned to more than one molecule is large (say, greater than 1% of the total number of grid points), it probably signifies that the value of CRIT is too low, that the molecular boundary is far from clear, or that the function used to define $\langle\rho(\mathbf{x})\rangle$ was badly chosen (Fig. 13.4.6.1). Grid points outside the molecular envelope can be set to the average solvent density.

An essential criterion for the molecular envelope is that it obeys the noncrystallographic point-group symmetry. If the original *h*-cell electron density already possesses the molecular symmetry (*e.g.* icosahedral 532, 222 *etc.*), then the *p*-cell mask should also have that symmetry. However, if the mask boundaries were chosen manually, masks from different molecular centres might be in conflict and have local errors in the correct molecular symmetry. Such errors can be corrected by reimposing the noncrystallographic point-group symmetry on the *p*-cell mask. This can be conveniently achieved by setting the density at each grid point that was considered within the molecular envelope to a value of 100, and all other grid points to a density of zero. If the resultant density is averaged using the same routine as is used for averaging the actual electron density of the molecule, then the average density will remain 100 if the interpolated density is 100 at all noncrystallographically related points. However, if the original grid point is near the edge of the mask, finding the density at symmetry-related points may involve interpolation between density at level 100 and at level 0, giving an averaged density of less than 100. Hence, any grid point whose averaged density is below some criterion should be attributed to solvent.



references