relates objects within one crystal lattice or between crystal lattices. The NCS rotational relationship in real space is exactly mimicked in reciprocal space. Local symmetry in real space has the equivalent effect of rotating a reciprocal lattice onto itself or another (with origins coincident), such that the integral reciprocal-lattice points of one reciprocal space coincide with non-integral reciprocal-lattice positions in the other. As the reciprocal lattice samples the Fourier transform of a molecule only at finite and integral reciprocal-lattice points, the effect of an NCS operation is to permit sampling of the molecular transform at intermediate non-integral reciprocal-lattice positions. If such sampling occurs frequently enough, it will constitute a plot of the continuous transform of the molecule and, hence, amount to a structure determination.

Whenever a molecule exists more than once either in the same unit cell or in different unit cells, then error in the molecular electron-density distribution due to error in phasing can be reduced by averaging the various molecular copies. The number of such copies, N, is referred to as the noncrystallographic redundancy. As the NCS is, by definition, only local (often pertaining to a particular molecular centre), there are holes and gaps between the averaged density, which presumably are solvent space between molecules. Thus, the electron density can be improved both by averaging electron density and by setting the density between molecules to a low, constant value ('solvent flattening'). Phases calculated by Fourier back-transforming the improved density should be more accurate than the original phases. Hence, the observed structure amplitudes (suitably weighted) can be associated with the improved phases, and a new and improved map can be calculated. This, in turn, can again be averaged until convergence has been reached and the phases no longer change. In addition, the back-transformed map can be used to compute phases just beyond the extremity of the resolution of the terms used in the original map. The resultant amplitudes will not be zero because the map had been modified by averaging and solvent flattening. Thus, phases can be gradually extended and improved, starting from a very low resolution approximation to the molecular structure. This procedure was first implemented in reciprocal space (Rossmann & Blow, 1963; Main, 1967; Crowther, 1969) and then, more recently, in real space (Bricogne, 1974, 1976; Johnson, 1978; Jones, 1992; Rossmann *et al.*, 1992). More recently still, there has been an attempt to reproduce the very successful real-space procedure in reciprocal space (Tong & Rossmann, 1995).

Early examples of such a procedure for phase improvement are the structure determinations of deoxyhaemoglobin (Muirhead *et al.*, 1967), α-chymotrypsin (Matthews *et al.*, 1967), lobster glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al.*, 1974), hexokinase (Fletterick & Steitz, 1976), tobacco mosaic virus disk protein (Champness *et al.*, 1976; Bloomer *et al.*, 1978), the influenza virus haemagglutinin spike (Wilson *et al.*, 1981), tomato bushy stunt virus (Harrison *et al.*, 1978) and southern bean mosaic virus (Abad-Zapatero *et al.*, 1980). Early examples of phase extension, using real-space electron-density averaging, were the study of glyceraldehyde-3-phosphate dehydrogenase (Argos *et al.*, 1975), satellite tobacco necrosis virus (Nordman, 1980), haemocyanin (Gaykema *et al.*, 1984), human rhinovirus 14 (Rossmann *et al.*, 1985) and poliovirus (Hogle *et al.*, 1985). Since then, this method has been used in numerous virus structure determinations, with the phase extension being initiated from ever lower resolution.

A once-popular computer program for real-space averaging was written by Gerard Bricogne (1976). Another program has been described by Johnson (1978). Both programs were based on a double-sorting procedure. Bricogne (1976) had suggested that, with interpolation between grid points using linear polynomials, it was necessary to sample electron density at grid intervals finer than one-sixth of the resolution limit of the Fourier terms that were used in calculating the map. With the availability of more computer memory, it was possible to store much of the electron density, thus avoiding time-consuming sorting operations (Hogle *et al.*, 1985; Luo *et al.*, 1989). Simultaneously, the storage requirements could be drastically reduced by using interpolation with quadratic polynomials. While the latter required a little extra computation time, this was far less than what would have been needed for sorting. Furthermore, it was found that Bricogne's estimate for the fineness of the map storage grid was too pessimistic, even for linear interpolation, which works well to about 1/2.5 of the resolution limit of the map.

In addition to changes in strategy brought about by computers with much larger memories, experience has been gained in program requirements for real-space averaging for phase determination (Dodson *et al.*, 1992). Here we give a general procedure for electron-density averaging.

13.4.4. The *p*- and *h*-cells

It is useful to define two types of unit cells.

(1) The '*p*-cell' is the unit cell of the unknown crystal structure and is associated with fractional coordinates $\mathbf{y}$ and unit-cell vectors $\mathbf{a}_p, \mathbf{b}_p, \mathbf{c}_p$.

(2) The '*h*-cell' is the unit cell with respect to which the noncrystallographic axes of the molecule (or particle) are to be defined in a standard orientation and is associated with fractional coordinates $\mathbf{x}$ and unit-cell vectors $\mathbf{a}_h, \mathbf{b}_h, \mathbf{c}_h$.

Since the averaged molecule is to be placed into all crystallographically related positions in the *p*-cell, it is essential to know the envelope that encloses a single molecule. Care must be taken that the envelopes from neighbouring molecules in the *p*-cell do not overlap. The remaining space between the limits of the envelopes of the variously placed molecules in the *p*-cell can be taken to be solvent and, hence, flattened, a useful physical assumption for helping phase determination.

The *h*-cell must be chosen to be at least as large as the largest dimension of the molecule. In general, it is convenient to define the *h*-cell with $a_h = b_h = c_h$ and $\alpha = \beta = \gamma = 90°$, while placing the molecular centre at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. For example, if the molecule is a viral particle with icosahedral symmetry, the standard orientation can be defined by placing the twofold axes to correspond to the *h*-cell unit-cell axes, a procedure which can be done in one of two ways (Fig. 13.4.4.1). It will be necessary to know how the molecule (or

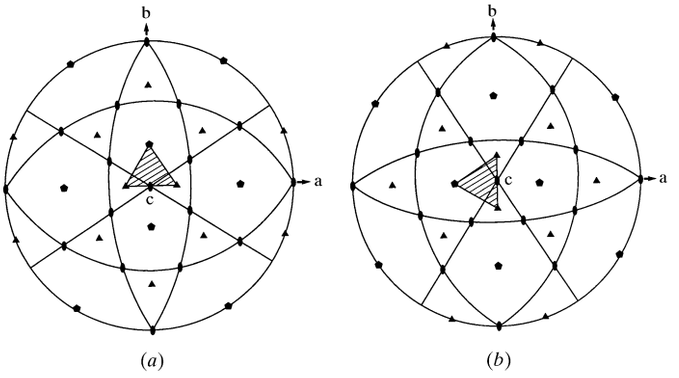

Fig. 13.4.4.1. Stereographic projections showing alternative definitions of the 'standard orientation' of an icosahedron in the *h*-cell. Icosahedral axes are placed parallel to the cell axes. Limits of a noncrystallographic asymmetric unit are shaded, representing 1/60th of the volume of an object with icosahedral symmetry. [Reproduced with permission from Rossmann *et al.* (1992). Copyright (1992) International Union of Crystallography.]

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

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

or

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

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