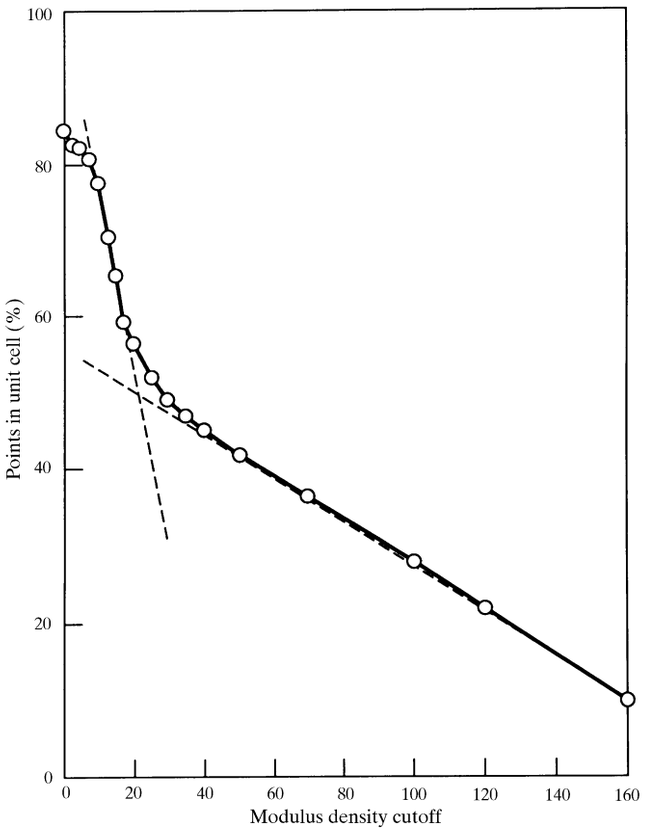

Fig. 13.4.6.1. The volume of the molecular mask expressed as a percentage of the volume of the *p*-cell asymmetric unit, as determined by the density cutoff in the *h*-cell. When the modulus of the density cutoff is decreased to less than the mean smeared electron density within the protein, the mask volume increases rapidly. Intersection of the tangents suggests the most appropriate density cutoff value for mask generation. [Reproduced with permission from McKenna, Xia, Willingmann, Ilag & Rossmann (1992). Copyright (1992) International Union of Crystallography.]

Other improvements to mask generation were discussed by Rossmann *et al.* (1992). In any event, the molecular-envelope definition should be periodically re-examined after a suitable number of electron-density-averaging cycles.

13.4.7. Finding the averaged density

Electron density can be averaged (1) among the N NCS-related molecules in the *p*-cell (the real crystal unit cell), thus creating a new and improved map of the *p*-cell; (2) among the N NCS-related molecules in the *p*-cell and placing the results into a standard orientation in the *h*-cell; or (3) among the N NCS-related molecules in different unit cells and placing the results back into the original different unit cells or into a standard *h*-cell. Before averaging commences, the $M \times N$ matrices $[\mathrm{E}_{m,n}]$ and translation vectors $\mathbf{e}_{m,n}$ must be evaluated [see (13.4.5.9) and (13.4.5.11*a*)]. Here, N is the noncrystallographic redundancy and M is the number of molecules that impinge on the crystallographic asymmetric unit of the *p*-cell. Associated with each grid point in the *p*-cell asymmetric unit will be (1) the value of m designating which molecular centre is to be associated with that grid point (a special value of m is for solvent) and (2) the *p*-cell electron density at that point.

Table 13.4.7.1. *Mean root-mean-square scatter between noncrystallographically related points*

Example taken from φX174 structure determination. $\langle\rho_8\rangle$ is proportional to the mean density (e Å^{-3}) based on eight-point interpolation; n is number of grid points with $\langle\rho_8\rangle$ in a given range; $\langle\sigma(\rho_8)\rangle$ is the root-mean-square deviation from ρ_8 among noncrystallographic asymmetric points averaged over all points in the mask.

	Density derived from an electron microscopy image at 25 Å resolution		Density derived from a 3.3 Å crystal structure	
$\langle\rho_8\rangle$	n	$\langle\sigma(\rho_8)\rangle$	n	$\langle\sigma(\rho_8)\rangle$
−375 to −325	1	44.7	0	0.0
−325 to −275	16	44.4	0	0.0
−275 to −225	22	39.5	41	31.4
−225 to −175	81	34.9	3493	25.5
−175 to −125	299	34.7	65049	20.5
−125 to −75	1119	33.1	290025	17.7
−75 to −25	16617	34.7	661386	15.0
−25 to 25	33818	46.9	1,016274	12.8
25 to 75	6008	31.9	344620	16.3
75 to 125	4512	32.0	215036	18.9
125 to 175	3050	32.1	146690	22.1
175 to 225	1562	32.6	58155	26.3
225 to 275	542	33.4	6032	32.2
275 to 325	213	35.6	227	40.6
325 to 375	33	34.7	9	46.8

The grid points within the asymmetric unit are then examined one at a time. If the grid point is within the mask, it is averaged among the N noncrystallographically related equivalent positions belonging to molecule m. If the grid point is solvent, the density can be set to the average solvent density.

The N noncrystallographically equivalent non-integral grid points can be computed from (13.4.5.11*a*). Some of these will lie outside the crystallographic asymmetric unit. These will, therefore, have to be operated on by unit-cell translations and crystallographic symmetry operations to bring them back into the asymmetric unit before the corresponding interpolated density can be calculated.

Averaging into the *h*-cell can be done by a procedure similar to averaging in the *p*-cell, except that the rotation and translation matrices are given by (13.4.5.13). Furthermore, no mask is required as all the averaging into the *h*-cell (from *p*-cell electron density) can be done with respect to the reference molecule centred at $\mathbf{s}_{p,m=1}$ in the *p*-cell. Each grid point is taken in turn in the *h*-cell. The electron density at any grid point that is further away from $\mathbf{s}_\mathrm{h}$ than from R_out is set to zero. Other grid-point positions are expanded into the N equivalent positions in the *p*-cell surrounding $\mathbf{s}_{p,m=1}$. The interpolated density is then found, averaged over the N equivalent positions, and stored at the original *h*-cell grid point in successive sections, in the same way as in the *p*-cell averaging. As in averaging within the *p*-cell, a record is kept of $\langle\sigma(\rho)\rangle$ as a function of $\langle\rho(\mathbf{x})\rangle$ (Table 13.4.7.1). In general, the local NCS is valid only within the molecule. Hence, the *h*-cell density will show the molecular envelope and can be used to recompute an improved *p*-cell density mask. The rate of build up of signal within the molecule should be roughly proportional to N, while the rate outside the molecule should be proportional to about $N^{1/2}$.

Copyright © 2006 International Union of Crystallography

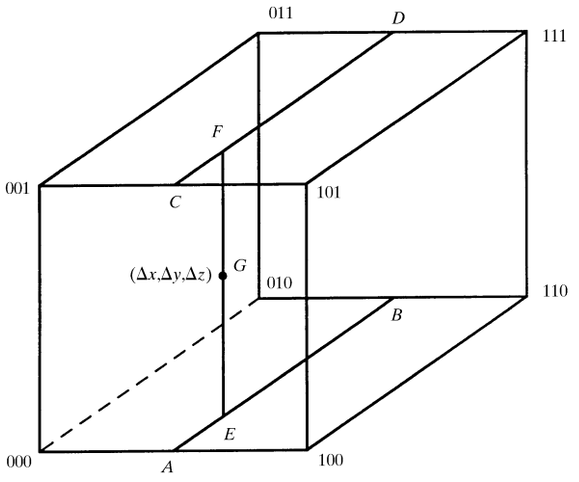

Fig. 13.4.8.1. Interpolation box for finding the approximate electron density at $G(\Delta x, \Delta y, \Delta z)$, given the eight densities at the corners of the box. The interpolated value can be built up by first using interpolations to determine the densities at *A*, *B*, *C* and *D*. A second linear interpolation then determines the density at *E* (from densities at *A* and *B*) and at *F* (from densities at *C* and *D*). The third linear interpolation determines the density at *G* from the densities at *E* and *F*. [Reproduced with permission from Rossmann *et al.* (1992). Copyright (1992) International Union of Crystallography.]

13.4.8. Interpolation

Some thought must go into defining the size of the grid interval. Shannon's sampling theorem shows that the grid interval must never be greater than half the limiting resolution of the data. Thus, for instance, if the limiting resolution is 3 Å, the grid intervals must be smaller than 1.5 Å. Clearly, the finer the grid interval, the more accurate the interpolated density, but the computing time will increase with the inverse cube of the size of the grid step. Similarly, if the grid interval is fine, less care and fewer points can be used for interpolation, thus balancing the effect of the finer grid in terms of computing time. In practice, it has been found that an eight-point interpolation (as described below) can be used, provided the grid interval is less than 1/2.5 of the resolution (Rossmann *et al.*, 1992). Other interpolation schemes have also been used (*e.g.* Bricogne, 1976; Nordman, 1980; Hogle *et al.*, 1985; Bolin *et al.*, 1993).

A straightforward 'linear' interpolation can be discussed with reference to Fig. 13.4.8.1 (in mathematical literature, this is called a trilinear approximation or a tensor product of three one-dimensional linear interpolants). Let *G* be the position at which the density is to be interpolated, and let this point have the fractional grid coordinates Δx, Δy, Δz within the box of surrounding grid points. Let 000 be the point at $\Delta x = 0,\ \Delta y = 0, \Delta z = 0$. Other grid points will then be at 100, 010, 001 *etc.*, with the point diagonally opposite the origin at 111.

The density at *A* (between 000 and 100) can then be approximated as the value of the linear interpolant of ρ_{000} and ρ_{100}:

$$\rho(A) \cong \rho_A = \rho_{000} + (\rho_{100} - \rho_{000})\Delta x.$$

Similar expressions for $\rho(B)$, $\rho(C)$ and $\rho(D)$ can also be written. Then, it is possible to calculate an approximate density at *E* from

$$\rho(E) \cong \rho_E = \rho_A + (\rho_B - \rho_A)\Delta y,$$

with a similar expression for $\rho(F)$. Finally, the interpolated density at *G* between *E* and *F* is given by

$$\rho(G) \cong \rho_G = \rho_E + (\rho_F - \rho_E)\Delta z.$$

Putting all these together, it is easy to show that

$$\begin{aligned}\rho_G = {} & \rho_{000} + \Delta x(\rho_{100} - \rho_{000}) + \Delta y(\rho_{010} - \rho_{000}) + \Delta z(\rho_{001} - \rho_{000}) \\ & + \Delta x\Delta y(\rho_{000} + \rho_{110} - \rho_{100} - \rho_{010}) \\ & + \Delta y\Delta z(\rho_{000} + \rho_{011} - \rho_{010} - \rho_{001}) \\ & + \Delta z\Delta x(\rho_{000} + \rho_{101} - \rho_{001} - \rho_{100}) \\ & + \Delta x\Delta y\Delta z(\rho_{100} + \rho_{010} + \rho_{001} + \rho_{111} - \rho_{000} - \rho_{101} \\ & - \rho_{011} - \rho_{110}).\end{aligned}$$

13.4.9. Combining different crystal forms

Frequently, a molecule crystallizes in a variety of different crystal forms [*e.g.* hexokinase (Fletterick & Steitz, 1976), the influenza virus neuraminidase spike (Varghese *et al.*, 1983), the histocompatibility antigen HLA (Bjorkman *et al.*, 1987) and the CD4 receptor (Wang *et al.*, 1990)]. It is then advantageous to average between the different crystal forms. This can be achieved by averaging each crystal form independently into a standard orientation in the *h*-cell (if the redundancy is $N = 1$ for a given crystal form, then this simply amounts to producing a skewed representation of the *p*-cell in the *h*-cell environment). The different results, now all in the same *h*-cell orientation, can be averaged. However, care must be taken to put equal weight on each molecular copy. If the *i*th cell contains N_i noncrystallographic copies, then the average of the densities, $\rho_i(\mathbf{x})$ $(i = 1, 2, \ldots, I)$, is

$$\sum_i N_i \rho_i(\mathbf{x}) \Big/ \sum_i N_i$$

at each grid point, $\mathbf{x}$, in the *h*-cell. Additional weights can be added to account for the subjective assessment of the quality of the electron densities in the different crystal cells.

With the *h*-cell density improved by averaging among different crystal forms, it can now be replaced into the different *p*-cells. These *p*-cells can then be back-transformed in the usual manner to obtain a better set of phases. These, in turn, can be associated with the observed structure amplitudes for each *p*-cell structure, and the cycle can be repeated.

13.4.10. Phase extension and refinement of the NCS parameters

Fourier back-transformation of the modified (averaged and solvent-flattened) map leads to poor phase information immediately outside the previously used resolution limit. If no density modification had been made, the Fourier transform would have yielded exactly the same structure factors as had been used for the original map. However, the modifications result in small structure amplitudes just beyond the previous resolution limit. The resultant phases can then be used in combination with the observed amplitudes in the next map calculation, thus extending the limit of resolution.

If the cell edge of an approximately cubic unit cell is *a*, and the approximate radius of the molecule is $\mathcal{R}$ (therefore, $\mathcal{R} < a$), then the first node of a spherical diffraction function will occur when $H\mathcal{R} = 0.7$, where *H* is the length of the reciprocal-lattice vector between the closest previously known structure factor and the structure factor just outside the resolution limit. Let $H = n(1/a)$, and let it be assumed that the diffraction-function amplitude is negligible when $H\mathcal{R} > 0.7$. Thus, for successful extension, $n = a/\mathcal{R}$. In general, that means that phase extension should be less than two reciprocal-lattice units in one step.

As phase extension proceeds, the accuracy of the NCS elements and the boundaries of the envelope must be constantly improved and updated to match the improved resolution. Arnold & Rossmann

references