15.1.3. Reciprocal-space interpretation of density modification

Density modification, although mostly performed in real space for ease of application, can be understood in terms of reciprocal-space constraints on structure-factor amplitudes and phases.

Main & Rossmann (1966) showed that the NCS-averaging operation in real space can be expressed in reciprocal space as the convolution of the structure factors and the Fourier transform of the molecular envelope and the NCS matrices. Similarly, the solvent-flattening operation can be considered a multiplication of the map by some mask, $g_{\rm sf}(\mathbf{x})$, where $g_{\rm sf}(\mathbf{x}) = 1$ in the protein region and $g_{\rm sf}(\mathbf{x}) = 0$ in the solvent region. Thus

$$\rho_{\rm mod}(\mathbf{x}) = g_{\rm sf}(\mathbf{x}) \times \rho(\mathbf{x}). \qquad (15.1.3.1)$$

This assumes that the solvent level is zero, which can be achieved by suitable adjustment of the $F(000)$ term.

If we transform this equation to reciprocal space, then the product becomes a convolution; thus

$$F_{\rm mod}(\mathbf{h}) = (1/V)\sum_{\mathbf{k}} G_{\rm sf}(\mathbf{k})F(\mathbf{h}-\mathbf{k}), \qquad (15.1.3.2)$$

where $G_{\rm sf}(\mathbf{k})$ is the Fourier transform of the mask $g_{\rm sf}(\mathbf{x})$. The solvent mask $g_{\rm sf}(\mathbf{x})$ shows the outline of the molecule with no internal detail, so must be a low-resolution image. Therefore, all but the lowest-resolution terms of $G_{\rm sf}$ will be negligible.

The convolution expresses the relationship between phases in reciprocal space from the constraint of solvent flatness in real space. Since only the terms near the origin of $G_{\rm sf}$ are nonzero, the convolution can only relate phases that are local to each other in reciprocal space. Thus, it can only provide phase information for structure factors near the current phasing resolution limit.

This reasoning may also be applied to other density modifications. Histogram matching applies a nonlinear rescaling to the current density in the protein region. The equivalent multiplier, $g_{\rm hm}(\mathbf{x})$, shows variations of about 1.0 that are related to the features in the initial map. The function $G_{\rm hm}(\mathbf{h})$ for histogram matching is, therefore, dominated by its origin term, but shows significant features to the same resolution as the current map or further, as the density rescaling becomes more nonlinear. Histogram matching can therefore give phase indications to twice the resolution of the initial map or beyond, although phase indications will be weak and contain errors related to the level of error in the initial map.

$$\rho_{\rm mod}(\mathbf{x}) = g_{\rm ncs}(\mathbf{x})(1/N_{\rm ncs})\sum_i \rho_i(\mathbf{x}). \qquad (15.1.3.3)$$

Averaging may be described as the summation of a number of reoriented copies of the electron density within the region of the averaging mask (Main & Rossmann, 1966), *i.e.* where $\rho_i(\mathbf{x})$ is the initial density, $\rho(\mathbf{x})$, transformed by the ith NCS operator and $g_{\rm ncs}(\mathbf{x})$ is the mask of the molecule to be averaged. This summation is repeated for each copy of the molecule in the whole unit cell. The reciprocal-space averaging function, $G_{\rm ncs}(\mathbf{h})$, is the Fourier transform of a mask, as for solvent flattening, but since the mask covers only a single molecule, rather than the molecular density in the whole unit cell, the extent of $G_{\rm ncs}(\mathbf{h})$ in reciprocal space is greater.

Sayre's equation is already expressed as a convolution, although in this case the function $G(\mathbf{h})$ is given by the structure factors $F(\mathbf{h})$ themselves. It is, therefore, the most powerful method for phase extension. However, as resolution decreases, more of the reflections required to form the convolution are missing, and the error increases.

The functions $g(\mathbf{x})$ and $G(\mathbf{h})$ for these density modifications are illustrated in Fig. 15.1.3.1 for a simple one-dimensional structure.

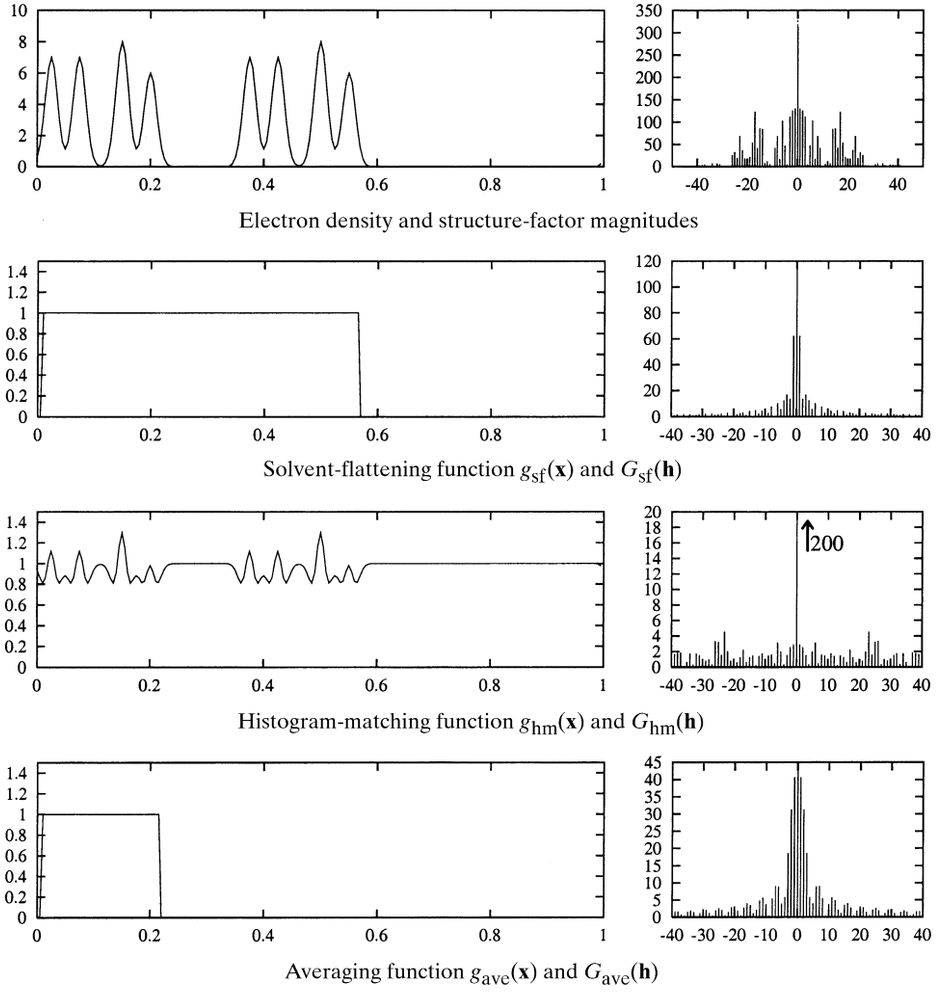

Fig. 15.1.3.1. The functions $g(\mathbf{x})$ and $G(\mathbf{h})$ for solvent flattening, histogram matching and averaging.

15.1.4. Phase combination

Phase combination is used to filter the noise in the modified phases and eliminate the incorrect component of the modified phases through a statistical process. The observed structure-factor amplitudes are used to estimate the reliability of the phases after density modification. The estimated probability of the modified phases is combined with the probability of observed phases to produce a more reliable phase estimate,

$$P_{\rm new}[\varphi(\mathbf{h})] = P_{\rm obs}[\varphi(\mathbf{h})]P_{\rm mod}[\varphi(\mathbf{h})]. \qquad (15.1.4.1)$$

Once a modified map has been obtained, modified phases and amplitudes may be derived from an inverse Fourier transform. The modified phases are normally combined with the initial phases by multiplication of their probability distributions. The probability distribution for the experimentally observed phases is usually de-

references