

15.1. PHASE IMPROVEMENT BY ITERATIVE DENSITY MODIFICATION

If we use two threshold values, ρ_1 and ρ_2 , to divide the three density regions, the complete formula can be expressed as

$$P(\rho) = \begin{cases} N \exp\left[-(\rho - \bar{\rho})^2/2\sigma^2\right] & \text{for } 2\rho \leq \rho_2 \\ N(a\rho^3 + b\rho^2 + c\rho + d) & \text{for } 2\rho_2 < \rho \leq \rho_1 \\ N(A/\rho)[\ln(\rho_0/\rho)]^{1/2} & \text{for } 2\rho_1 < \rho \leq \rho_0 \end{cases} \quad (15.1.2.11)$$

The parameters a, b, c, d in the cubic polynomial are calculated by matching function values and gradients at ρ_1 and ρ_2 . The parameters in the histogram formula, $\bar{\rho}, \sigma, A, \rho_0, \rho_1, \rho_2$, can be obtained from histograms of known structures.

15.1.2.2.3. The process of histogram matching

Zhang & Main (1990a) demonstrated that, at better than 4 Å resolution, the histogram for an MIR map is generally significantly different from the ideal distribution calculated from atomic coordinates. The obvious course is therefore to alter the map in such a way as to make its density histogram equal to the ideal distribution. Unfortunately, there are an infinite number of maps corresponding to any chosen density distribution, so we must choose a systematic method of altering the map.

The conventional method of performing such a modification is to retain the ordering of the density values in the map. The highest point in the original map will be the highest point in the modified map, the second highest points will correspond in the same way, and so on.

Mathematically, this transformation is represented as follows. Let $P(\rho)$ be the current density histogram and $P'(\rho)$ be the desired distribution, normalized such that their sums are equal to 1. The cumulative distribution functions, $N(\rho)$ and $N'(\rho)$, may then be calculated:

$$N(\rho) = \int_{\rho_{\min}}^{\rho} P(\rho) \, d\rho, \quad (15.1.2.12)$$

$$N'(\rho') = \int_{\rho_{\min}}^{\rho'} P'(\rho) \, d\rho.$$

The cumulative distribution function of a variable transforms a value chosen from the distribution into a number between 0 and 1, representing the position of that value in an ordered list of values chosen from the distribution.

The transformation may, therefore, be performed in two stages. A density value is taken from the initial distribution and the cumulative distribution function of the initial distribution is applied to obtain the position of that value in the distribution. The inverse of the cumulative distribution function for the desired distribution is applied to this value to obtain the density value for the corresponding point in the desired distribution. Thus, given a density value, ρ , from the initial distribution, the modified value, ρ' , is obtained by

$$\rho' = N'^{-1}[N(\rho)]. \quad (15.1.2.13)$$

The distribution of ρ' will then match the desired distribution after the above transformation. The transformation of an electron-density value by this method is illustrated in Fig. 15.1.2.3. The transformation in equation (15.1.2.13) can be achieved through a linear transform represented by

$$\rho'_i = a_i \rho_i + b_i, \quad (15.1.2.14)$$

where $i = \{1, \dots, n\}$ and n is the number of density bins. The above linear transform is sufficient if the number of density bins is large enough. An n value of about 200 is usually quite satisfactory.

Various properties of the electron density are specified in the density histogram, such as the minimum, maximum and mean density, the density variance, and the entropy of the map. The mean density of the ideal map can be obtained by

$$\bar{\rho} = \int_{\rho_{\min}}^{\rho_{\max}} \rho P(\rho) \, d\rho. \quad (15.1.2.15)$$

The variance of the density in the ideal map can be obtained by

$$\sigma(\rho) = \left(\overline{\rho^2} - \bar{\rho}^2\right)^{1/2}, \quad (15.1.2.16)$$

where

$$\overline{\rho^2} = \int_{\rho_{\min}}^{\rho_{\max}} \rho^2 P(\rho) \, d\rho. \quad (15.1.2.17)$$

The entropy of the ideal map can be calculated by

$$S = - \int_{\rho_{\min}}^{\rho_{\max}} P(\rho) \rho \ln(\rho) \, d\rho. \quad (15.1.2.18)$$

Therefore, the process of histogram matching applies a minimum and a maximum value to the electron density, imposes the correct mean and variance, and defines the entropy of the new map. The order of electron-density values remains unchanged after histogram matching.

Histogram matching is complementary to solvent flattening since it is applied to the protein region of a map, whereas

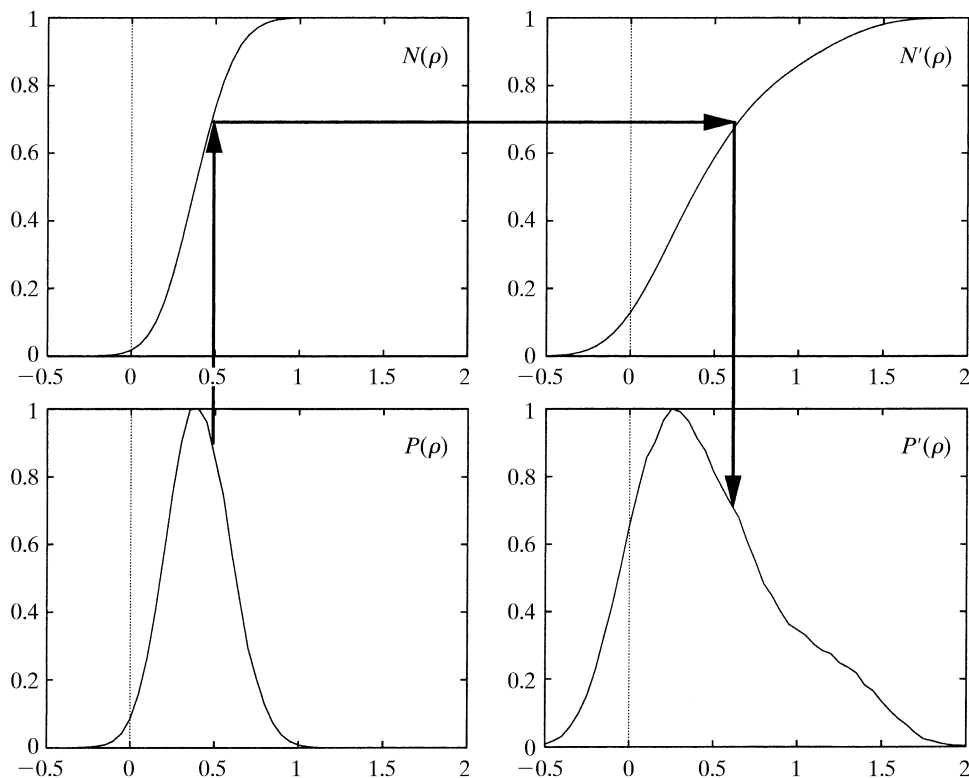


Fig. 15.1.2.3. Transformation of density ρ to ρ'_{mod} by histogram matching.