

15. DENSITY MODIFICATION AND PHASE COMBINATION

efficient alternative algorithm for tracing density ridges is given by Swanson (1994).

15.1.2.5. Sayre's equation

Sayre's equation constrains the local shape of electron density. It provides a link between all structure-factor amplitudes and phases. It is an exact equation at atomic resolution in an equal-atom system. It is, therefore, very powerful for phase refinement and extension for small molecules at atomic resolution (Sayre, 1952, 1972, 1974). However, its power diminishes as resolution decreases. It can still be an effective tool for macromolecular phase refinement and extension if the shape function can be modified to accommodate the overlap of atoms at non-atomic resolution (Zhang & Main, 1990b).

15.1.2.5.1. Sayre's equation in real and reciprocal space

Sayre's equation (Sayre, 1952, 1972, 1974) expresses the constraint on structure factors when the atoms in a structure are equal and resolved, and the equation has formed the foundation of direct methods. In protein calculations, the resolution is generally too poor for atoms to be resolved, and this is reflected in the bulk of the terms required to calculate the equation for any particular missing structure factor.

For equal and resolved atoms, squaring the electron density changes only the shape of the atomic peaks and not their positions. The original density may therefore be restored by convoluting with some smoothing function, $\psi(\mathbf{x})$, which is a function of atomic shape,

$$\rho(\mathbf{x}) = (V/N) \sum_{\mathbf{y}} \rho^2(\mathbf{y}) \psi(\mathbf{x} - \mathbf{y}), \quad (15.1.2.31)$$

where

$$\psi(\mathbf{x} - \mathbf{y}) = (1/V) \sum_{\mathbf{h}} \theta(\mathbf{h}) \exp[2\pi i \mathbf{h} \cdot (\mathbf{x} - \mathbf{y})]. \quad (15.1.2.32)$$

Here, $\theta(\mathbf{h})$ is the ratio of scattering factors of real, $f(\mathbf{h})$, and 'squared', $g(\mathbf{h})$, atoms, and V is the unit-cell volume, *i.e.*,

$$\theta(\mathbf{h}) = f(\mathbf{h})/g(\mathbf{h}). \quad (15.1.2.33)$$

Sayre's equation states that the convolution of the squared electron density with a shape function restores the original electron density. It can be seen from equation (15.1.2.31) that Sayre's equation puts constraints on the local shape of electron density. The local shape function is the Fourier transform of the ratio of scattering factors of the real and 'squared' atoms.

Sayre's equation is more frequently expressed in reciprocal space as a system of equations relating structure factors in amplitude and phase:

$$F(\mathbf{h}) = [\theta(\mathbf{h})/V] \sum_{\mathbf{k}} F(\mathbf{k}) F(\mathbf{h} - \mathbf{k}). \quad (15.1.2.34)$$

The reciprocal-space expression of Sayre's equation can be obtained directly from a Fourier transformation of both sides of equation (15.1.2.31) and the application of the convolution theorem.

15.1.2.5.2. The application of Sayre's equation to macromolecules at non-atomic resolution – the $\theta(\mathbf{h})$ curve

Sayre's equation is exact for an equal-atom structure at atomic resolution. The reciprocal-space shape function, $\theta(\mathbf{h})$, can be calculated analytically from the ratio of the scattering factors of real and 'squared' atoms, which can both be represented by a Gaussian function. At infinite resolution, we expect $\theta(\mathbf{h})$ to be a spherically symmetric function that decreases smoothly with increased \mathbf{h} . However, for data at non-atomic resolution, the $\theta(\mathbf{h})$

curve will behave differently because atomic overlap changes the peak shapes. Therefore, a spherical-averaging method is adopted to obtain an estimate of the shape function empirically from the ratio of the observed structure factors and the structure factors from the squared electron density using the formula

$$\theta(s) = V \left\langle \frac{F(\mathbf{h})}{\sum_{\mathbf{k}} F(\mathbf{k}) F(\mathbf{h} - \mathbf{k})} \right\rangle_{|\mathbf{h}|}, \quad (15.1.2.35)$$

where the averaging is carried out over ranges of $|\mathbf{h}|$, *i.e.*, over spherical shells, each covering a narrow resolution range. Here, s represents the modulus of \mathbf{h} .

The empirically derived shape function only extends to the resolution of the experimentally observed phases. This is sufficient for phase refinement. However, there are no experimentally observed phases to give the empirical $\theta(s)$ for phase extension. Therefore, a Gaussian function of the form

$$\theta(s) = K \exp(-Bs^2) \quad (15.1.2.36)$$

is fitted to the available values of $\theta(s)$, and the parameters K and B are obtained using a least-squares method. The shape function $\theta(s)$ for the resolution beyond that of the observed phases is extrapolated using the fitted Gaussian function. The derivation of the shape function $\theta(s)$ from a combination of spherical averaging and Gaussian extrapolation is the key to the successful application of Sayre's equation for phase improvement at non-atomic resolution (Zhang & Main, 1990b).

15.1.2.6. Atomization

The atomization method uses the fact that the structure underlying the map consists of discrete atoms. It attempts to interpret the map by automatically placing atoms and refining their positions.

Agarwal & Isaacs (1977) proposed a method for the extension of phases to higher resolutions by interpreting an electron-density map in terms of 'dummy' atoms. These are so called because at the initial resolution of 3.0 Å, true atom peaks could not be resolved. The placement of 'dummy atoms' is subject to constraints of bonding distance and the number of neighbours. The coordinates and temperature factors of these dummy atoms may then be refined against all the available diffraction amplitudes. Structure factors may then be calculated from the refined coordinates to provide phases for the high-resolution reflections and to improve the phases of the starting set.

The atomization approach has been extended in the ARP program (Lamzin & Wilson, 1997) by the use of difference-map criteria to test dummy-atom assignments, with the aim of removing wrong atoms and introducing missing atoms. With modern refinement algorithms, this technique has become very effective for the solution of structures at high resolution from a poor molecular-replacement model, or even directly from an MIR/MAD map.

Map improvement has also been demonstrated at intermediate resolutions by Perrakis *et al.* (1997) using a multi-solution variant of the ARP method, and by Vellieux (1998).

The interpretation of an approximately phased map has also been applied very successfully as part of the 'Shake n' Bake' direct-methods procedure (Miller *et al.*, 1993; Weeks *et al.*, 1993). The alternating application of phase refinement by the minimum principle in reciprocal space ('Shake') and atomization in real space ('Bake') has proved to be a very powerful method for solving small protein structures at atomic resolution using only structure-factor amplitudes.