

1. GENERAL RELATIONSHIPS AND TECHNIQUES

1.3.4.4.7.9. Discussion of macromolecular refinement techniques

The impossibility of carrying out a full-matrix least-squares refinement of a macromolecular crystal structure, caused by excessive computational cost and by the paucity of observations, led Diamond (1971) to propose a real-space refinement method in which stereochemical knowledge was used to keep the number of free parameters to a minimum. Refinement took place by a least-squares fit between the 'observed' electron-density map and a model density consisting of Gaussian atoms. This procedure, coupled to iterative recalculation of the phases, led to the first highly refined protein structures obtained without using full-matrix least squares (Huber *et al.*, 1974; Bode & Schwager, 1975; Deisenhofer & Steigemann, 1975; Takano, 1977*a,b*).

Real-space refinement takes advantage of the localization of atoms (each parameter interacts only with the density near the atom to which it belongs) and gives the most immediate description of stereochemical constraints. A disadvantage is that fitting the 'observed' electron density amounts to treating the phases of the structure factors as observed quantities, and to ignoring the experimental error estimates on their moduli. The method is also much more vulnerable to series-termination errors and accidentally missing data than the least-squares method. These objections led to the progressive disuse of Diamond's method, and to a switch towards reciprocal-space least squares following Agarwal's work.

The connection established above between the Cruickshank–Agarwal modified Fourier method and the simple use of the chain rule affords a partial refutation to both the premises of Diamond's method and to the objections made against it:

(i) it shows that refinement *can* be performed through localized computations in real space *without* having to treat the phases as observed quantities;

(ii) at the same time, it shows that measurement errors on the moduli *can* be fully utilized in real space, *via* the Fourier synthesis of the functional derivative $\partial R / \partial \rho^{\text{calc}}(\mathbf{x})$ or by means of the coefficients of a quadratic model of R in a search subspace.

1.3.4.4.7.10. Sampling considerations

The calculation of the inner products $(D, \partial \rho^{\text{calc}} / \partial u_p)$ from a sampled gradient map D requires even more caution than that of structure factors *via* electron-density maps described in Section 1.3.4.4.5, because the functions $\partial \sigma_j / \partial u_p$ have transforms which extend even further in reciprocal space than the σ_j themselves. Analytically, if the σ_j are Gaussians, the $\partial \sigma_j / \partial u_p$ are finite sums of multivariate Hermite functions (Section 1.3.2.4.4.2) and hence the same is true of their transforms. The difference map D must therefore be finely sampled and the relation between error and sampling rate may be investigated as in Section 1.3.4.4.5. An examination of the sampling rates commonly used (*e.g.* one third of the resolution) shows that they are insufficient. Tronrud *et al.* (1987) propose to relax this requirement by applying an artificial temperature factor to σ_j (*cf.* Section 1.3.4.4.5) and the negative of that temperature factor to D , a procedure of questionable validity because the latter 'sharpening' operation is ill defined [the function $\exp(-\|\mathbf{x}\|^2)$ does not define a tempered distribution, so the associativity properties of convolution may be lost]. A more robust procedure would be to compute the scalar product by means of a more sophisticated numerical quadrature formula than a mere grid sum.

1.3.4.4.8. Miscellaneous correlation functions

Certain correlation functions can be useful to detect the presence of multiple copies of the same molecule (known or unknown) in the asymmetric unit of a crystal of unknown structure.

Suppose that a crystal contains one or several copies of a molecule \mathcal{M} in its asymmetric unit. If $\mu(\mathbf{x})$ is the electron density of that molecule in some reference position and orientation, then

$$\rho^0 = \sum_{j \in J} \left[\sum_{g \in G} S_g^{\#}(T_j^{\#} \mu) \right],$$

where $T_j : \mathbf{x} \mapsto \mathbf{C}_j \mathbf{x} + \mathbf{d}_j$ describes the placement of the j th copy of the molecule with respect to the reference copy. It is assumed that each such copy is in a general position, so that there is no isotropy subgroup.

The methods of Section 1.3.4.2.2.9 (with p_j replaced by $C_j^{\#} \mu$, and \mathbf{x}_j by \mathbf{d}_j) lead to the following expression for the auto-correlation of ρ^0 :

$$\check{\rho}^0 * \rho^0 = \sum_{j_1} \sum_{j_2} \sum_{g_1} \sum_{g_2} \tau_{S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})} \times [(R_{g_1}^{\#} C_{j_1}^{\#} \check{\mu}) * (R_{g_2}^{\#} C_{j_2}^{\#} \mu)].$$

If μ is unknown, consider the subfamily σ of terms with $j_1 = j_2 = j$ and $g_1 = g_2 = g$:

$$\sigma = \sum_j \sum_g R_g^{\#} C_j^{\#} (\check{\mu} * \mu).$$

The scalar product $(\sigma, R^{\#} \sigma)$ in which R is a variable rotation will have a peak whenever

$$R = (R_{g_1} C_{j_1})^{-1} (R_{g_2} C_{j_2})$$

since two copies of the 'self-Patterson' $\check{\mu} * \mu$ of the molecule will be brought into coincidence. If the interference from terms in the Patterson $\pi = r * \check{\rho}^0 * \rho^0$ other than those present in σ is not too serious, the 'self-rotation function' $(\pi, R^{\#} \pi)$ (Rossmann & Blow, 1962; Crowther, 1972) will show the same peaks, from which the rotations $\{C_j\}_{j \in J}$ may be determined, either individually or jointly if for instance they form a group.

If μ is known, then its self-Patterson $\check{\mu} * \mu$ may be calculated, and the C_j may be found by examining the 'cross-rotation function' $[\pi, R^{\#}(\check{\mu} * \mu)]$ which will have peaks at $R = R_g C_j, g \in G, j \in J$. Once the C_j are known, then the various copies $C_j^{\#} \mu$ of \mathcal{M} may be Fourier-analysed into structure factors:

$$M_j(\mathbf{h}) = \mathcal{F}[C_j^{\#} \mu](\mathbf{h}).$$

The cross terms with $j_1 \neq j_2, g_1 \neq g_2$ in $\check{\rho}^0 * \rho^0$ then contain 'motifs'

$$(R_{g_1}^{\#} C_{j_1}^{\#} \check{\mu}) * (R_{g_2}^{\#} C_{j_2}^{\#} \mu),$$

with Fourier coefficients

$$\overline{M_{j_1}(\mathbf{R}_{g_1}^T \mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T \mathbf{h}),$$

translated by $S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$. Therefore the 'translation functions' (Crowther & Blow, 1967)

$$\mathcal{T}_{j_1 g_1, j_2 g_2}(\mathbf{s}) = \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \overline{M_{j_1}(\mathbf{R}_{g_1}^T \mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T \mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{s})$$

will have peaks at $\mathbf{s} = S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$ corresponding to the detection of these motifs.