The general equation of δ direct methods and the novel *SMAR* algorithm residuals using the absolute value of ρ and the zero conversion of negative ripples

Jordi Rius*

Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, 08193 Bellaterra, Catalonia, Spain. *Correspondence e-mail: jordi.rius@icmab.es

The general equation $\delta_M(\mathbf{r}) = \rho(\mathbf{r}) + g(\mathbf{r})$ of the δ direct methods (δ-GEQ) is established which, when expressed in the form $\delta_M(\mathbf{r}) - \rho(\mathbf{r}) = g(\mathbf{r})$, is used in the *SMAR* phasing algorithm [Rius (2020). *Acta Cryst* A**76**, 489–493]. It is shown that *SMAR* is based on the alternating minimization of the two residuals $R_\rho(\chi) = \int_V [\rho(\chi) - \rho(\Phi)s_\rho]^2 \, dV$ and $R_\delta(\Phi) = \int_V m_\rho[\delta_M(\chi) - \rho(\Phi)s_\rho]^2 \, dV$ in each iteration of the algorithm by maximizing the respective $S_\rho(\Phi)$ and $S_\delta(\Phi)$ sum functions. While $R_\rho(\chi)$ converges to zero, $R_\delta(\Phi)$ converges, as predicted by the theory, to a positive quantity. These two independent residuals combine δ_M and ρ each with $|\rho|$ while keeping the same unknowns, leading to overdetermination for diffraction data extending to atomic resolution. At the beginning of a *SMAR* phase refinement, the zero part of the m_ρ mask [resulting from the zero conversion of the slightly negative $\rho(\Phi)$ values] occupies ~50% of the unit-cell volume and increases by ~5% when convergence is reached. The effects on the residuals of the two *SMAR* phase refinement modes, *i.e.* only using density functions (slow mode) supplemented by atomic constraints (fast mode), are discussed in detail. Due to its architecture, the *SMAR* algorithm is particularly well suited for Deep Learning. Another way of using δ-GEQ is by solving it in the form $\rho(\mathbf{r}) = \delta_M(\mathbf{r}) - g(\mathbf{r})$, which provides a simple new derivation of the already known δ_M tangent formula, the core of the δ recycling phasing algorithm [Rius (2012). *Acta Cryst.* A**68**, 399–400]. The nomenclature used here is: (i) Φ is the set of φ structure factor phases of ρ to be refined; (ii) $\delta_M(\chi) = \mathrm{FT}^{-1}\{c(|E| - \langle|E|\rangle) \times \exp(i\alpha)\}$ with $\chi = \{\alpha\}$, the set of phases of $|\rho|$ and c = scaling constant; (iii) m_ρ = mask, being either 0 or 1; s_ρ is 1 or −1 depending on whether $\rho(\Phi)$ is positive or negative.

1. Introduction

Historically, direct methods were developed to solve small crystal structures directly from high-resolution single-crystal diffraction data. From their origins in the 1950s (Sayre, 1952; Cochran, 1952; Zachariasen, 1952), they have seen continuous advances over the years, not only driven by the steady increase in computing power, but also by clever algorithms and efficient implementations. Computing power continues to increase but the solution of larger structures at lower resolution is still hindered in the case of equal atoms and requires the use of ubiquitous model fragments and density modification. In their widespread successful application, the time scale is in any case shorter than the experimental effort involved in X-ray structure determination.

One of the most recent advances in direct methods has been the $|\rho|$-based algorithm (*SMAR*) maximizing the $S_{M,|\rho|}$ ($= S_{MAR}$) sum function (Rius, 2020). It corresponds to the latest stage in the evolution of the S_M origin-free modulus sum

$$(|\xi|, \alpha) = FT\{\rho(\Phi) s_\rho\} \leftarrow \rho(\Phi)^{1)} = FT^{-1}\{|E|, \varphi^{new(or\ ini)}\}$$

$$\downarrow \qquad\qquad \uparrow$$

$$\delta_M(\chi) = FT^{-1}\{\Delta E, \alpha\} \rightarrow (|\varepsilon|, \varphi^{new}) = FT\{\delta_M(\chi) s_\rho m_\rho^{2)}\}$$

function (originally named Z_R; Rius, 1993) which still involved triple-phase structure invariants. Even today, Z_R is the simplest and certainly one of the most successful direct methods working exclusively in reciprocal space (Rius *et al.*, 1995). However, it was almost immediately superseded by the *Shake & Bake* strategy alternating between reciprocal- and real-space refinements (dual-space recycling methods) which allowed the solution of larger crystal structures (Weeks *et al.*, 1993; Miller *et al.*, 1993). Dual-space methods do not eliminate phase relationships in reciprocal space but complement them with peak picking in real space as an extreme form of density modification. A comprehensive description at the height of development of dual-space recycling methods can be found in the *International Tables for Crystallography*, Vol. F (Sheldrick *et al.*, 2012).

Following this trend, triple-phase invariants (whose number becomes exceedingly large for large structures) were replaced in S_M by the more efficient and accurate Fourier transforms (FT) (Rius *et al.*, 2007; Rius, 2014). More recently, the ρ^2 function in S_M was replaced by the mathematically simpler $|\rho|$ function and a new mask was introduced that takes the negative values of ρ into account. Both changes have led to the *SMAR* phasing algorithm (Rius, 2020). Important aspects of its application have been covered in two recent publications, the first dealing with its extension to larger crystal structures by introducing the fast inner-pixel preservation procedure (*ipp*) for density modification (Rius & Torrelles, 2021) and using initial phase values derived from the modulus function. In the second publication *SMAR* was adapted to the solution of anomalous scattering substructures in protein crystals (Rius & Torrelles, 2022). The present short introduction is intended to provide a brief overview of the 30-year development of that particular type of direct methods which shares the implicit or explicit use of the $\rho \simeq \delta_M$ (or δ_P) approximation (Rius, 2012a). To distinguish this family of direct methods from the rest – and also to help in their identification – the general term 'δ direct methods' is coined in this publication.

The aim of this article is to complete the theoretical foundations of the *SMAR* phasing algorithm. The algorithm
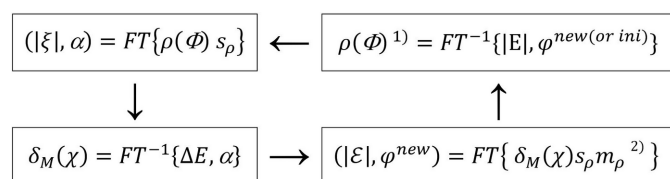

Figure 1
The iterative *SMAR* phasing algorithm in four stages. (Upper right-hand corner) Initial (or updated) φ phase estimates belonging to set Φ are combined with observed $|E|$ values to obtain ρ and $\rho(\Phi)s_\rho = |\rho(\Phi)|$ [the superscript $^{1)}$ indicates that ρ is stored]. (Upper left) The FT of $|\rho(\Phi)|$ is calculated to get the new set χ of α phases as well as the calculated $|\xi|$ values. (Lower left corner) The new α values are combined with the experimental $\Delta E = |E| - \langle|E|\rangle$ and Fourier transformed to obtain $\delta_M(\chi)$. The m_ρ mask and the s_ρ signs are derived from the stored $\rho(\Phi)$, and the $\delta_M(\chi)s_\rho m_\rho = \rho'$ product is carried out. (Lower right) The FT of ρ' supplies the updated φ phases [the superscript $^{2)}$ indicates that when *ipp* is applied, ρ' is further modified to ρ'' before the FT operation] as well as the calculated $|\mathcal{E}|$ values.

Table 1
Values of the $m_\rho(\mathbf{r})$ mask and the $s_\rho(\mathbf{r})$ sign functions obtained from $\rho(\mathbf{r}, \Phi)$ and for $t = 2.5$ with σ_ρ^2 being the (phase-independent) variance of $\rho(\mathbf{r})$ and $m'_\rho = s_\rho m_\rho$ (Rius, 2020).

The meanings of *CORE*s, *SPZ*s and *SNZ*s are explained in Section 2.1. Columns 6 and 7 give the mask compositions in % at the first and last iterations of a *SMAR* phase refinement reaching convergence (slow convergence mode) using the diffraction data for Actinomycin Z3 (Schäfer *et al.*, 1998).

Condition	Corresponds to	m'_ρ	m_ρ	s_ρ	First	Last
$\rho(\mathbf{r}, \Phi) > 0$	*CORE*s and *SPZ*s	1	1	1	50.0	45.3
$0 \leq \rho(\mathbf{r}, \Phi) > -t\sigma_\rho$	*SNZ*s	0	0	−1	49.4	54.5
$\rho(\mathbf{r}, \Phi) \leq -t\sigma_\rho$	Very negative values	−1	1	−1	0.62	0.23

described by Rius (2020) and shown in Fig. 1 essentially consists of the iterative application of the phasing formula

$$\varphi_k^{\text{new}} = \text{phase of} \int_V \delta_M(\mathbf{r})\, m'_{\rho,t}(\mathbf{r}) \exp\left(i2\pi\mathbf{k}\mathbf{r}\right) \mathrm{d}\mathbf{r}, \qquad (1)$$

where (i) φ_k is the phase of the kth structure factor of ρ and belongs to the set Φ of phases to be refined; (ii) $\delta_M(\chi)$ is equal to $\text{FT}^{-1}\left\{c(|E| - \langle|E|\rangle)\exp(i\alpha)\right\}$ with $\chi = \{\alpha\}$ being the set of phases of $|\rho|$ and c a scaling constant; (iii) $\mathbf{r}$ is a position vector inside V, the unit-cell volume; and (iv) $m'_{\rho,t}$ is a mask function defined in Table 1 that can take the values 1, 0 or −1. The representative test case shown in Table 1 ($t = 2.5$) indicates that the zero part in the $m'_{\rho,t}$ mask is around 50% at the beginning of a phase refinement with initially random phase values; as the refinement converges, the zero part increases by up to 5%. Most of the remaining part of the mask is taken up by ones, as the proportion of −1 values is kept very small ($<1.0\%$).

The *SMAR* phasing algorithm was originally derived from the S_{MAR} sum function. A sum function such as S_{MAR} generally corresponds to the mixed integral of a residual as *e.g.* in the case of expression (20) in relation to (21) in this article. Only when the residual is known can the derivation of the phasing algorithm be considered complete, which leads to a better understanding of it and enables, for example, the estimation of the minimum value of the residual. In this context, the use of the $\delta_M \simeq \rho$ approximation in Rius (2020) represented a limitation. To overcome this, the relationship between δ_M and ρ is worked out in Appendix A, resulting in δ-*GEQ* which, when modified accordingly, leads to one of the two desired *SMAR* residuals. The derivation of the second residual is simpler and introduces the phases corresponding to $|\rho(\Phi)|$ in the algorithm. To increase the readability of this article, δ-*GEQ* is derived separately in Appendix A and a summary thereof given in Section 2.2.

To complete this introduction, it is interesting to mention, particularly for newcomers, that density modification in the context of direct methods was efficiently introduced in *ACORN* (Foadi *et al.*, 2000). *ACORN* and *SHELXE* (Sheldrick, 2002) both use density sharpening and negative density elimination. Later, in the *VLD* algorithm, a difference and a flipping term were combined (Burla *et al.*, 2010). Finally, another related and more modern development was *SHELXT*, which is more broadly associated with the charge-

Table 2
Overview of the properties of the two main peak types of the δ_M Fourier synthesis.

The peak strength of a $\rho(\mathbf{r}_l)$ peak corresponds to fN_{ref}/V.

Peak type (function)	Peak positions	Peak strengths	Number in unit cell
$A \rightarrow \rho$	At $\mathbf{r}_l$ atomic positions ($l = 1, N$)	$\rho(\mathbf{r}_l) = \delta_M(\mathbf{r}_l)$	N
$B \rightarrow g$	At $\mathbf{r}_{jlm} = \mathbf{r}_j + \mathbf{r}_l - \mathbf{r}_m$ with $l \neq m$ and $j \neq m$ ($j, l, m = 1, N$)	$g(\mathbf{r}_{jlm}) \simeq$ strength of $\{\rho(\mathbf{r}_l)/(N-1)\}$	$N(N-1)^2$

flipping algorithm but combines it with direct methods and with density modification at part of the peak positions, used to eliminate atoms at random without atoms (Sheldrick, 2015). One distinctive feature of the density modification in *SMAR* is the zero conversion of only slight negative densities and the preservation of the inner peak pixels (Rius & Torrelles, 2021).

All calculations in this article have been performed with a modified version of the *XLENS_v1* code (Rius, 2011). The diffraction data used in the test calculations correspond to:

(i) Actinomycin Z3 with 1228 (C, N, O) + 8 Cl atoms in the unit cell. According to the refinement protocol in the Protein Data Bank (PDB code 1a7z), 4 Cl sites are partially occupied and the other 4 Cl atoms have a rather large B value, so that their scattering powers are considerably reduced. The minimum d spacing ($d_{\min}$) is 0.95 Å; $a = 14.803$, $b = 24.780$ and $c = 65.059$ Å, space group $P2_12_12_1$ (Schäfer et al., 1998).

(ii) Alpha1 peptide with 503 (C, N, O) + 1 Cl. $d_{\min} = 0.90$ Å; $a = 20.846$, $b = 20.909$ and $c = 27.057$ Å, $\alpha = 102.40$, $\beta = 95.33$ and $\gamma = 119.62°$, $P1$ (Privé et al., 1999).

(iii) Pep1 with 344 (C, N, O). $d_{\min} = 1.00$ Å; $a = 13.999$, $b = 21.602$ and $c = 21.615$ Å, $P2_12_12_1$ (Antel et al., 1995).

(iv) Suoa with 188 (C, N, O). $d_{\min} = 1.00$ Å; $a = 18.350$, $b = 21.441$ and $c = 8.350$ Å, $P2_12_12_1$ (Oliver & Strickland, 1984).

2. Basic elements of the *SMAR* algorithm

2.1. The ρ Fourier synthesis: its mask definition and general relationship to $|\rho|$

When solving crystal structures by direct methods from atomic resolution X-ray diffraction intensity data, the electron-density function is normally calculated with the Fourier synthesis

$$\rho(\mathbf{r}, \Phi_T) = \frac{1}{V} \sum_k |E_k| \exp(i\varphi_k) \exp(-i2\pi\mathbf{kr}), \quad (2)$$

where $|E_k|$ is the modulus and φ_k is the phase value of the (quasi)-normalized structure factor of the kth reflection (Main, 1975). The moduli and phases of all reflections form the $\{|E|\}$ and $\Phi = \{\varphi\}$ sets, respectively. For clarity, the phase type (and eventually the modulus type) used in the Fourier synthesis is added to the function name when required to avoid confusion, e.g. $\rho(r, \Phi_T)$ specifies that $\rho(r)$ is calculated with Φ_T (the index T stands for true phase values).

Expression (2) is stated in terms of the normalized scattering factors, i.e. $\hat{f}_j = f_j/\sqrt{\sum_m f_m^2}$ for an arbitrary atom j (with f_j being its normal scattering factor including the Debye–Waller factor). For the sake of simplicity, a crystal structure

with N equal atoms in the unit cell is assumed throughout, so that the normalized scattering factor reduces to $\hat{f} = 1/\sqrt{N}$.

Due to the limited number of Fourier terms (a consequence of the finite number of measured intensities), the $\rho(r, \Phi_T)$ synthesis is affected by Fourier series truncation effects. These termination effects are mainly reflected in the broadening of the atomic peaks, each consisting of a large spherically symmetric positive central part (= *CORE*) surrounded by negative and positive waves (ripples) that decrease as one moves away from the peak center. For point-like atoms (for which broadening due to scattering factors and thermal vibration effects are largely removed), the limiting spherical surface of the *CORE* lies $\sim 0.72 \times d_{\min}$ (Å) from the peak center [it corresponds to the first zero of the T_3 spreading function in Lipson & Cochran (1966)]. For locations lying outside neighboring *CORE*s, the ripple contributions of neighboring atoms add up and result in slightly negative and slightly positive zones (*SNZ*s and *SPZ*s, respectively). The $\rho(r, \Phi)$ values contained in the interval $[0, -t\sigma_\rho]$ make up the *SNZ*s (σ_ρ and t are defined in Table 1). Since the ρ distribution in the crystal is positive definite, the probability that the *SNZ*s accommodate the *CORE*s of atomic peaks is zero, so this information can be introduced in the form of an $m_{\rho,t}$ mask which is 0 for all negative ρ in the $[0, -t\sigma_\rho]$ interval (negative ripple conversion to zero) and 1 elsewhere. As shown in Table 1, the mask value depends on $\rho(r)$ and the t parameter (since t is always 2.5 in this work, it is suppressed in $m_{\rho,t}$ for notation simplicity and we use simply m_ρ). Table 1 also shows that the zero part of the mask extends to at least 50% of the unit-cell volume for $t \simeq 2.5$ (just for comparison, for $t \geq 10$ the threshold is so low that the mask values of all negative ρ values become zero).

The relationship between the syntheses $|\rho(\Phi)|$ and $\rho(\Phi)$ is given by the equality

$$|\rho(\mathbf{r}, \Phi)| = \rho(\mathbf{r}, \Phi) s_\rho(\mathbf{r}) \quad \forall \mathbf{r} \subset V, \quad (3)$$

in which $s_\rho(r)$ is 1 or -1 depending on whether $\rho(r, \Phi)$ is positive or negative. This equality is completely general, so the product $\rho(\mathbf{r}, \Phi) s_\rho(\mathbf{r})$ can always replace $|\rho(r, \Phi)|$ in the derivation of the residuals. Finally, since ρ is positive definite and diffraction data are assumed to reach atomic resolution, it is clear for $\Phi = \Phi_T$ that the negative $s_\rho(\mathbf{r})$ values in (3) are always associated with small $\rho(r, \Phi_T)$, so that (3) becomes $\rho(\mathbf{r}, \Phi_T) = |\rho(\mathbf{r}, \Phi_T)| \, \forall \mathbf{r} \subset V$. If we denote the phase set of the Fourier coefficients of the $|\rho(r, \Phi_T)|$ synthesis by χ_T, then $\Phi_T \cong \chi_T$.

2.2. The general equation of δ direct methods and its different forms

The δ_M Fourier synthesis defined by

$$\delta_M(\mathbf{r}, \Phi_T) = \frac{c}{V}\sum_k (|E_k| - \langle|E|\rangle)\exp(i\varphi_k)\exp(-i2\pi\mathbf{kr}), \quad (4)$$

with $c = 2/(|E| - 1/\sqrt{N})$, is studied in detail in Appendix A, showing that it contains two types of positive peak (A and B). Table 2 lists their peak strengths and positions. The stronger peaks of type A correspond to ρ (which are resolved in the Fourier map). The $N - 1$ times weaker peaks of type B are located at positions other than the atomic positions (coincidence is accidental). These are the main constituents of the function $g(r)$ and, due to their large number, must be severely overlapped in the unit cell. The standard form of the general equation of the δ direct methods (δ-GEQ) corresponds to the sum of both contributions $\rho(r)$ and $g(r)$ [equation (43)]. However, δ-GEQ can be used in other forms. For example it can be solved for ρ, so δ-GEQ then takes the form

$$\rho(\mathbf{r}) = \delta_M(\mathbf{r}) - g(\mathbf{r}). \quad (5)$$

One obvious difficulty here is how to handle the unknown g function. This difficulty can be circumvented by introducing the mask $m_{\Delta\delta}$ (being either 0 or 1), which results from the realizations that (i) δ_M and ρ have their strong peaks at the atomic positions; and (ii) g is formed by the positive strongly overlapped peaks of type B which are much more numerous but also much smaller than the peaks in δ_M and ρ. As shown in Fig. 2, the $m_{\Delta\delta}$ mask is obtained by expressing the threshold $\Delta\delta$ in terms of the computable $\sigma(\delta_M)$ by means of $\Delta\delta = t_1\sigma(\delta_M)$, with $t_1 \simeq 2.5$, such that $m_{\Delta\delta}(\mathbf{r}) = 1$ for $\delta_M(r) \geq \Delta\delta$ and $m_{\Delta\delta}(\mathbf{r}) = 0$ otherwise. This can be mathematically expressed by
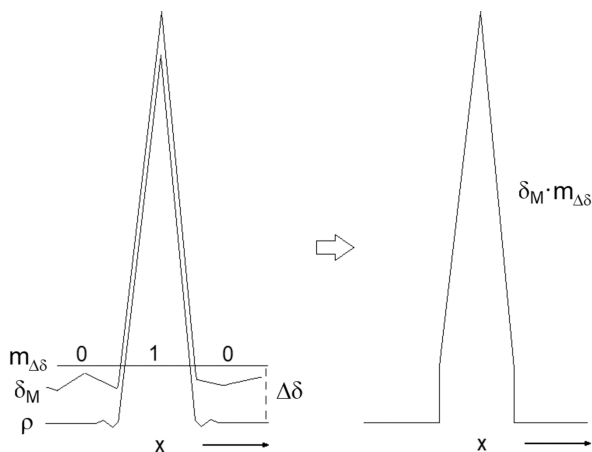

Figure 2
(Left) Part of a hypothetical one-dimensional unit cell corresponding to an atomic position, with schematic representation of the associated $\delta_M(x)$ and $\rho(x)$ [equation (5)]. The binary mask $m_{\Delta\delta}(x)$ is 1 if $\delta_M(x)$ is larger than the $\Delta\delta$ threshold and 0 otherwise. (Right) The same part of the unit cell shows the product function $\delta_M(x)m_{\Delta\delta}(x)$, which is proportional to $\rho(x)$ (for equiatomic structures). If the number N of expected atoms in the unit cell is known, then the δ_M tangent formula reduces to a structure factor calculation over the N largest product-function peaks greater than $\Delta\delta$, i.e. the integral of the FT in expression (7) reduces to a sum.

$$\rho(\mathbf{r}) = K\delta_M(\mathbf{r})\, m_{\Delta\delta}(\mathbf{r}), \quad (6)$$

with K being a suitable scaling constant. Fourier transforming both sides of (6), and since E and ρ are linked by the Fourier transform $E = \text{FT}(\rho)$, the formula

$$E = K\,\text{FT}\{\delta_M m_{\Delta\delta}\} \quad (7)$$

results. The angular part of (7) corresponds to the δ_M tangent formula which forms the core of the δ recycling algorithm (Rius, 2012a,b). It has been successfully applied to X-ray diffraction data from small crystal structures, to 3D electron diffraction data (Rius et al., 2013; Capitani et al., 2014) and, due to its robustness, to synchrotron tts (tts = through the substrate) microdiffraction data (Rius et al., 2017). Some considerations regarding the implementation of the δ_M tangent formula are given in Section A3.

SMAR uses another form of δ-GEQ in which g is isolated, namely

$$\delta_M(\mathbf{r}, \Phi_T) - \rho(\mathbf{r}, \Phi_T) = g(\mathbf{r}) \quad \forall\, \mathbf{r} \subset V. \quad (8)$$

According to $\Phi_T \cong \chi_T$ at the end of Section 2.1, (8) can be expressed in terms of χ_T (= the set of phases corresponding to $|\rho(\Phi_T)|$) so that

$$\delta_M(\mathbf{r}, \chi_T)\, m_\rho(\mathbf{r}) - \rho(\mathbf{r}, \chi_T)\, m_\rho(\mathbf{r}) = g(\mathbf{r})\, m_\rho(\mathbf{r}) \quad \forall\, \mathbf{r} \subset V, \quad (9)$$

where both sides are multiplied by m_ρ [which is derived from $\rho(\Phi_T)$]. Expression (9) is the basic equation for one of the two SMAR residuals (R_δ). Note the positive effect of introducing the mask m_ρ. Since the zero part of the mask is $\geq 50\%$ of the unit-cell volume, the unwanted contribution of g in (9) is suppressed for at least half of the unit cell.

3. The SMAR residuals

Each iteration of the SMAR algorithm consists of two differentiated parts, ending each part with the application of the corresponding phasing formula (upper-left and lower-right corners of Fig. 1). In this section, the two SMAR residuals leading to these phasing formulae are determined. In the following $\mathbf{r}$ is omitted unless absolutely necessary.

3.1. The $R_\rho(\chi)$ residual

The $|\rho(\Phi)|$ density function results from applying the absolute value operator to the $\rho(\Phi)$ Fourier synthesis. The structure factors of $|\rho(\Phi)|$ correspond to its Fourier transform,

$$|\xi|\exp(i\alpha) = \text{FT}[|\rho(\Phi)|], \quad (10)$$

with the moduli and phase values of the structure factors being globally denoted $\{|\xi|\}$ and χ, respectively. The inverse Fourier transform of both sides of (10) yields

$$\rho(\{|\xi|\}, \chi) = |\rho(\Phi)|. \quad (11)$$

For $\chi = \chi_T$ it can be assumed that $\{|\xi|\} \cong \{|E|\}$. Consequently, the integral

$$R_\rho(\chi) = \int_V [\rho(\{|E|\}, \chi) - \rho(\{|\xi|\}, \chi)]^2 \mathrm{d}V \quad (12)$$

must be close to zero for $R_\rho(\chi_T)$ which corresponds to the minimum of R_ρ. Simplifying the notation of $\rho(\{|E|\}, \chi)$ to $\rho(\chi)$ and replacing $\rho(\{|\xi|\}, \chi)$ first by $|\rho(\Phi)|$ according to (11) and then by $\rho(\Phi)s_\rho$ according to (3), integral (12) takes the simpler form

$$R_\rho(\chi) = \int_V \left[\rho(\chi) - \rho(\Phi)s_\rho\right]^2 dV. \tag{13}$$

During the refinement the function $\rho(\Phi)s_\rho$ is always positive. To find the new χ set minimizing $R_\rho(\chi)$, the integrand of (13) is developed into three integrals. The two integrals with integrands $|\rho(\Phi)|^2$ and $\rho^2(\chi)$ are both equal to $\frac{1}{V}\sum_k |E_k|^2$ and hence phase independent; however, the third one,

$$-2S_\rho(\chi) = -2\int_V \rho(\chi)\rho(\Phi)s_\rho \, dV, \tag{14}$$

is phase dependent. The maximum of a functional like $S_\rho(\chi)$ [which is equivalent to the minimum of $R_\rho(\chi)$ due to the minus sign in (14)] can be found by solving the condition for an extremum, $\partial S_\rho/\partial\alpha = 0$, $\forall\, \alpha \in \chi$, which, in parallel to Rius *et al.* (2007), yields the χ phasing formula,

$$\alpha^{new} = \text{phase of } FT\{\rho(\Phi)s_\rho\}. \tag{15}$$

3.2. The $R_\delta(\Phi)$ residual

The residual R_δ is obtained from the left side of (9) after generalizing χ_T to χ. This generalization entails two changes: (i) $\delta_M(\mathbf{r}, \chi_T)$ is simply changed to $\delta_M(\mathbf{r}, \chi)$, since in both cases the Fourier coefficients of δ_M contain the observed $|E| - \langle|E|\rangle$ values; and (ii) $\rho(\mathbf{r}, \chi_T)$ is changed to $\rho(\mathbf{r}, \{|\xi|\}, \chi)$. However, since $\rho(\mathbf{r}, \{|\xi|\}, \chi) = |\rho(\mathbf{r}, \Phi)| = \rho(\mathbf{r}, \Phi)s_\rho(\mathbf{r})$ because of (11) and then (3), the selected generalized form is $\rho(\mathbf{r}, \Phi)s_\rho(\mathbf{r})$ [this selection ensures that $\rho(\Phi)$ enters the residual expression (17)]. By applying these two changes to the left-hand side of (9), it becomes

$$\delta_M(\mathbf{r}, \chi)m_\rho(\mathbf{r}) - \rho(\mathbf{r}, \Phi)s_\rho(\mathbf{r})m_\rho(\mathbf{r}) \quad \forall\, \mathbf{r} \subset V. \tag{16}$$

Integration of (16) after squaring gives the R_δ residual,

$$R_\delta(\Phi) = \int_V \left[\delta_M(\chi)m_\rho - \rho(\Phi)m_\rho s_\rho\right]^2 dV, \tag{17}$$

where $\rho(\Phi)m_\rho s_\rho$ is always positive $[= |\rho(\Phi)|m_\rho]$. Of methodological importance is the minimum value of R_δ which occurs for $R_\delta(\Phi_T)$. An estimate of this value is obtained by squaring and integrating the right-hand side of (9), namely

$$R_\delta(\Phi_T) = \int_V m_\rho g^2 \, dV. \tag{18}$$

In this integral, $g = \delta_M(\Phi_T) - \rho(\Phi_T)$ given in (8) and $\rho(\Phi_T)$ used to calculate m_ρ are both different functions with different peak distributions. Consequently, the samples of g^2 at the points where $m_\rho = 1$ can be assumed to be random, allowing the factorization of $\langle m_\rho\rangle$ from the integral

$$R_\delta(\Phi_T) \cong \langle m_\rho\rangle \int_V g^2 \, dV = \langle m_\rho\rangle I_{g^2}. \tag{19}$$

The value of the normalized I_{g^2} is given by equation (49) in Appendix B, *i.e.*

$$I_{g^2} = (c-1)^2\langle|E|^2\rangle_k - c(c-2)(\langle|E|\rangle_k)^2 \simeq 1.12.$$

According to (19), R_δ does not converge to zero but to the positive value $R_\delta(\Phi_T) \simeq 1.12 \times \langle m_\rho\rangle$. Since it is known from Table 1 that $\langle m_\rho\rangle$ is ~0.45 (at the end of a converging refinement), the approximated value of $R_\delta(\Phi_T)$ should be 0.45 $\times$ 1.12 = 0.50.

Once the residual R_δ is defined and its minimum value known, the last step is to find the Φ phase set that minimizes R_δ. For this purpose, (17) is transformed into the sum

$$R_\delta = P + Q - 2S_\delta, \tag{20}$$

with P, Q and S_δ being the following integrals:

$$S_\delta = \int_V \delta_M(\chi)\,\rho(\Phi)s_\rho m_\rho \, dV, \tag{21}$$

$$P = \int_V \rho^2(\Phi)\,m_\rho \, dV, \tag{22}$$

$$Q = \int_V \delta_M^2(\chi)\,m_\rho \, dV. \tag{23}$$

The presence of the mask m_ρ complicates the solution of these integrals. The interested reader can find in Section 3.2.1 their evaluations with the help of experimental information. The principal conclusion of Section 3.2.1 is that minimizing R_δ is essentially equivalent to maximizing S_δ. Knowing this, one only needs to find the desired maximum of $S_\delta(\Phi)$ by solving the condition for an extremum, $\partial S_\delta/\partial\varphi = 0$, $\forall\, \varphi \in \Phi$. By expressing $\rho(\Phi)$ in (21) as a Fourier synthesis, then

$$S_\delta(\Phi) = \frac{1}{V}\sum_k |E_{-k}| \exp(i\varphi_{-k})$$
$$\times \int_V \delta_M(\mathbf{r}, \chi)\,m_\rho(\mathbf{r})\,s_\rho(\mathbf{r}) \exp(i2\pi\mathbf{kr})\,d\mathbf{r} \tag{24}$$

and, in parallel to Rius *et al.* (2007) and Rius (2020), the application of the condition for an extremum to (24) gives the Φ (or also *SMAR*) phasing formula,

$$\varphi^{new} = \text{phase of } FT\{\delta_M(\chi)\,m_\rho s_\rho\}. \tag{25}$$

For simplicity, the Fourier-transformed function $\delta_M m_\rho s_\rho$ is denoted ρ' in the slow convergence mode and ρ'' in the fast convergence mode (see Section 4).

3.2.1. Evolution of R_δ, S_δ, P and Q during the *SMAR* phase refinements

First, the values of the integrals (21), (22) and (23) are normalized by division with

$$\mathrm{SRO2} = \int_V \rho^2 \, \mathrm{d}V, \tag{26}$$

which only depends on the $|E|$ values and is therefore computable. The values of $-2S_\delta$, P and Q during the phase refinement progress were determined with the *SMAR* phasing algorithm already implemented in *XLENS_v1* (Rius, 2020). The density function values used in the estimation of (21), (22) and (23) are those available before applying *ipp* (Fig. 1). Tables 3 and 4 show the evolutions for Actinomycin Z3, Suoa, Pep1 and Alpha1 peptide (each grouping of four numbers given in this subsection always refers to this order of test structures; the output files of the test calculations are available in the supporting information). For Actinomycin Z3, the evolutions of $2S_\delta$, ΔP and ΔQ are also represented in Fig. 3. The evolution of the different integrals can be summarized as follows:

(i) Integral S_δ: When starting from random phase values, the initial $-2S_\delta$ values are always close to zero for all test structures and become -1.17, $-1,18$, -1.16 and -1.09 at the end of the respective convergent refinements.

(ii) Integral P: The initial P_0 value of integral P is 0.50 for all test structures and, as the phase refinement progresses, the difference $\Delta P = P - P_0$ increases. ΔP is approximately proportional to $2S_\delta$ with the slopes $\langle k_{\Delta P}\rangle$ equal to 0.171 (6), 0.155 (6), 0.159 (6) and 0.162 (5) for the four test structures (Tables 3 and 4). Consequently, the following empirical linear relationship between P and $2S_\delta$ can be established,

$$P = 0.50 + \langle k_{\Delta P}\rangle 2S_\delta. \tag{27}$$

(iii) Integral Q: To understand the significance of integral Q, the integral

$$\mathrm{SDEL} = \int_V \delta_M^2 \, \mathrm{d}V \tag{28}$$

is also calculated for each test structure [it only depends on the $(|E| - \langle|E|\rangle)^2$ quantities]. The corresponding SDEL values are 1.762, 1.790, 1.756 and 1.728. According to Tables 3 and 4, the initial Q values are $Q_0 = 0.89$, 0.89, 0.88 and 0.89, *i.e.* $Q_0 \simeq 0.50 \times \mathrm{SDEL}$. In addition, during the respective phase refinements, the largest $Q - Q_0$ differences are only 0.05, 0.01, 0.02 and 0.01. Consequently, it can be assumed that $Q \simeq Q_0$.

By taking all these results into account, R_δ can be simplified to

$$R_\delta \simeq (P_0 + Q_0) - (1 - \langle k_{\Delta P}\rangle)2S_\delta. \tag{29}$$

Since P_0, Q_0 and $\langle k_{\Delta P}\rangle$ can be considered nearly constant during the phase refinement, it follows from (29) that minimizing R_δ is essentially equivalent to maximizing S_δ.

4. The phase refinement modes in *SMAR*

Since most experimental results of *SMAR* have already been discussed by Rius & Torrelles (2021, 2022) and in Section 3.2.1 of this contribution, only a selection of points directly related to the topic of this article are treated here, grouped according to the convergence mode.

4.1. The slow convergence mode

This mode only works with density functions, that is, the positions of the atomic peaks are not used. This mode requires the inclusion of the Fourier terms of all reflections (strong + weak) in the calculation of ρ and δ_M. The $\rho' = \delta_M m_\rho s_\rho$ values entered in the *SMAR* phasing formula are obtained as follows:

(i) For slightly negative ρ values (amounting to 50–55% of the unit cell), the ρ' values are 0.

(ii) For positive ρ values (regardless of their strength and representing 45–50% of the unit cell), ρ' is equal to δ_M.

(iii) Only for very negative ρ values (<1% of the unit cell for $t \simeq 2.5$), ρ' is equated to $-\delta_M$ (the minus sign multiplying δ_M tends to restore the sign of the very negative ρ value).

Table 3
Evolution of $-2S_\delta$ [equation (21)], P [equation (22)], Q [equation (23)] and R_δ [equation (20)] during a converging default *SMAR* refinement from random starting phases for Actinomycin Z3 (Schäfer *et al.*, 1998), normalized by division by SRO2 [equation (26)] ($t = 2.5$).

$\langle k_{\Delta P}\rangle$ in (27) is the proportionality constant $k_{\Delta P}$ averaged over the number of refinement cycles. The columns headed 0 and -1 list the number of zero and negative voxels, respectively, in the ρ' map in %. The columns headed $CC_{\rho'}$ and $CC_{\rho''}$ give the correlation coefficients before and after the *ipp* application.

Iteration	$-2S_\delta$	P	Q	R_δ	$k_{\Delta P}$	0	-1	$CC_{\rho'}$	$CC_{\rho''}$
1	0.000	0.518	0.888	1.41	–	49.75	0.17	0.002	0.515
2	−0.436	0.574	0.880	1.02	0.169	51.60	0.09	0.307	0.652
5	−0.622	0.601	0.890	0.87	0.162	52.32	0.08	0.425	0.710
10	−0.678	0.614	0.898	0.83	0.168	52.39	0.10	0.457	0.729
15	−0.700	0.621	0.905	0.82	0.172	52.48	0.10	0.467	0.745
20	−0.724	0.624	0.913	0.81	0.172	52.53	0.11	0.479	0.746
21	−0.744	0.628	0.915	0.80	0.172	52.62	0.10	0.491	0.767
22	−0.826	0.642	0.936	0.75	0.172	52.91	0.09	0.533	0.800
23	−1.070	0.687	0.944	0.56	0.175	55.00	0.07	0.664	0.876
24	−1.164	0.710	0.936	0.48	0.172	54.65	0.08	0.714	0.906
25	−1.169	0.710	0.934	0.48	0.172	54.70	0.07	0.717	0.908
29	−1.166	0.710	0.938	0.48†	0.180	54.68	0.08	0.715	0.903
					$\langle k_{\Delta P}\rangle = 0.171$ (6) (28×)				

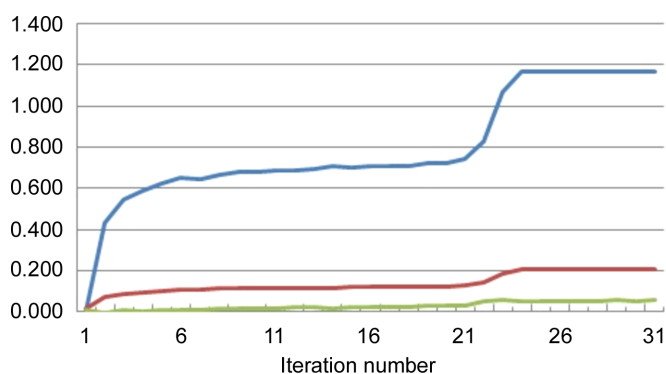† R_δ value at the end of the converging refinement.

Figure 3
Evolution of the normalized $2S_{MAR}$ (top, in blue), $(P - P_0)$ (middle, in red) and $(Q - Q_0)$ (bottom, in green) for a converging *SMAR* refinement using Actinomycin Z3 data (Schäfer *et al.*, 1998) ($t = 2.5$). See the heading of Table 3 for further details.

Table 4
As for Table 3 but for three additional test examples.

Only three stages of each phase refinement have been selected (at the beginning, when convergence begins and when it ends). In all three examples $Q - Q_0 \leq 0.02$ during the phase refinement.

Data set	Iteration	$-2S_\delta$	P	Q	R_δ	$k_{\Delta P}$	0	-1	$CC_{\rho'}$	$CC_{\rho''}$
Suoa	1	−0.033	0.517	0.893	1.38	–	49.71	0.17	0.024	0.602
(Oliver & Strickland, 1984)	13	−0.747	0.617	0.890	0.76	0.157	53.62	0.04	0.504	0.753
	27	−1.180	0.702	0.900	0.42	0.171	56.44	0.01	0.742	0.917
						$\langle k_{\Delta P}\rangle = 0.155$ (6) (26×)				
Pep1	1	−0.025	0.520	0.880	1.38	–	49.86	0.18	0.018	0.602
(Antel *et al.*, 1995)	14	−0.712	0.611	0.875	0.77	0.154	53.13	0.05	0.487	0.740
	28	−1.169	0.702	0.899	0.43	0.173	55.76	0.02	0.736	0.907
						$\langle k_{\Delta P}\rangle = 0.159$ (6) (27×)				
Alpha1 peptide	1	−0.013	0.519	0.866	1.37	–	49.90	0.18	0.000	0.575
(Privé *et al.*, 1999)	21	−0.651	0.605	0.861	0.82	0.156	52.40	0.09	0.451	0.711
	42	−1.089	0.687	0.875	0.47	0.167	54.18	0.05	0.703	0.906
						$\langle k_{\Delta P}\rangle = 0.162$ (5) (41×)				

In summary, ρ' corresponds either to unrestricted δ_M and $-\delta_M$ values or to fixed $\delta_M = 0$ ones. The mask used by the *SMAR* algorithm results in smooth phase refinements based only on density functions. On the other hand, the experimental $R_\delta(\Phi_T)$ values calculated at the end of converging phase refinements using (20) are 0.48, 0.42, 0.43 and 0.47 for Actinomycin Z3, Suoa, Pep1 and Alpha1 peptide, respectively (Tables 3 and 4). These values agree with the theoretical estimation of $R_\delta(\Phi_T) \simeq 0.50$ found in Section 3.2.

Regarding the very negative densities, some of them are due to the wrong input model (generated by the random starting phases) and disappear during the convergence process. The reductions observed in ρ' are 0.17% → 0.08% for Actinomycin Z3, 0.18% → 0.06% for Alpha1 peptide, 0.18% → 0.02% for Pep1 and 0.17% → 0.01% for Sucrose (Tables 3 and 4). To get an idea of the effect of t on the phasing of the intensity data of the four test structures, the sums of their total number of successful trials (out of 25) were determined for t = 1.5, 2.0, 2.5 and 10.0 (in the last case, the negative densities are all zero). The respective sums are 34, 76, 69 and 59 (Table S1 in the supporting information). The largest sums are obtained for t = 2.0 and 2.5 and the smallest for t = 1.5. For t = 10.0, the resulting sum is slightly worse than for t = 2.0 or 2.5. These results suggest that (i) the best t values are between 2.0 or 2.5, and (ii) although they represent only a small percentage of the unit cell, setting the very negative densities equal to zero is not beneficial for phase refinement. A possible explanation of the physical meaning of the very negative densities can be found in Section 3.3 of Rius (2020). In any case, a future comprehensive study focusing on this point would be useful.

4.2. The fast convergence mode

In this mode, which is the default operating mode of the *SMAR* algorithm, the part working only with density functions is supplemented by an additional step in which ρ' is modified by the *ipp* method to give ρ'' (Rius & Torrelles, 2021), which in turn replaces ρ' in the Φ phasing formula (25). The *ipp* method is an effective way of accelerating the phase refinement and

assumes that the approximate number N of expected atoms is known (which is normally the case). Briefly explained, *ipp* identifies in the ρ' Fourier map those grid points closest to the centers of the N largest peaks. The ρ' values of the 27 inner-peak grid points are then preserved for each peak and the remaining grid points of the ρ' Fourier map are set to zero, giving rise to the new ρ''. In this way no interpolation is required to find peak centers and, at the same time, the large ρ' values are preserved. Additionally, if the grid size Δ_{grid} is ~0.33 Å, *ipp* implicitly applies the minimum interpeak separation (*mips*) constraint. The criterion for considering a positive maximum of the ρ' Fourier synthesis a peak is that the voxel closest to the peak center must be surrounded by 26 smaller nearest-neighbor voxels (some can even be negative). Consequently, none of the nearest neighboring voxels can become the center of another ρ' peak. This means that for an isometric grid element with Δ_{grid} = 0.33 Å the average *mips* value is 0.955 Å [the minimum, intermediate and maximum separations are $2 \times 0.33 \times 1 = 0.67$ Å (6×), $2 \times 0.33 \times \sqrt{2} = 0.96$ Å (12×) and $2 \times 0.33 \times \sqrt{3} = 1.16$ Å (8×), respectively]. According to (19), the value of $R_\delta(\Phi_T)$ is proportional to $\langle m_\rho \rangle$. In the fast convergence mode, due to the application of *ipp*, $\langle m_\rho \rangle = 27N/N_{vox}$, where N_{vox} is the total number of voxels in the unit cell. For Actinomycin Z3, $27N = 33372$ and $N_{vox} = 664875$, and hence $\langle m_\rho \rangle = 0.05$, much smaller than the typical $\langle m_\rho \rangle$ values for the slow convergence mode (~0.45). Accordingly, $R_\delta(\Phi_T) = 1.12 \times 0.05 = 0.06$ is also much smaller than in the slow convergence mode.

A characteristic of this mode is that the calculation of $\rho(\Phi)$ only includes Fourier terms of those reflections satisfying the $|E| \geq |E|_{lim}$ condition with $|E|_{lim} = 1.0$, while the calculation of $\delta_M(\chi)$ (and thus of ρ') is always done with the Fourier terms of all k reflections (Fig. 1). That only the large $|E|$ values should participate in the ρ update is certainly related to the fact that only the 27 inner voxels close to the peak center are preserved (the rest of the peak voxels become part of the zero mask). This is supported by the fact that, in the slow convergence mode, the Fourier terms of all reflections must be included in the calculation of $\rho(\Phi)$.

4.3. The correlation coefficient

In addition to estimating R_δ, the agreement of minuends and subtrahends in (17) can also be estimated using the correlation coefficient

$$\text{CC}_{\rho'} = \frac{S_{MAR}}{(P\,Q)^{1/2}} \tag{30}$$

(using the density values before *ipp* for the slow convergence case). For refinements reaching convergence, the found $\text{CC}_{\rho'}$ values are 0.715 for Actinomycin Z3 (Table 3), 0.742 for Suoa, 0.736 for Pep1 and 0.705 for Alpha1 peptide (Table 4). These moderately high correlation coefficients also confirm the small discrepancy introduced in the $\text{CC}_{\rho'}$ calculation by the g contributions of those voxels with $m_\rho = 1$. As expected, the corresponding $\text{CC}_{\rho''}$ values (also listed in Tables 3 and 4) are much higher due to the smaller number of voxels with $m_\rho = 1$, *e.g.* $\text{CC}_{\rho'} = 0.72$ and $\text{CC}_{\rho''} = 0.90$ for Actinomycin Z3.

5. Conclusions

The main objective of this research was to complete the theoretical aspects of the *SMAR* phasing algorithm. For this purpose, the connection between δ_M and ρ has been examined in detail. This leads to the general equation of δ direct methods (δ-*GEQ*) which, written in its standard form, is $\delta_M = \rho + g$, where the density function g is mainly formed by a large number of small positive B-type peaks. Two ways of using δ-*GEQ* have been investigated. In *SMAR*, δ-*GEQ* is used in its difference form, $\delta_M - \rho = g$, while in δ recycling it is used solved for ρ, so $\rho = \delta_M - g$. In this second case, it has been shown that the δ_M tangent formula can be derived directly from it by including a suitable mask.

Regarding the *SMAR* residuals it can be concluded that:

(i) $R_\rho(\chi)$ measures the $[\rho(\chi) - \rho(\Phi)s_\rho]^2$ differences in the entire unit cell. The minimum value of R_ρ is $R_\rho(\chi_T) \simeq 0$.

(ii) The $R_\delta(\Phi)$ residual is based on its basic equation (9), *i.e.* $\delta_M(\chi_T)m_\rho - \rho(\chi_T)m_\rho = gm_\rho$, where χ_T are the α phases corresponding to $|\rho(\Phi_T)|$.

(iii) $R_\delta(\Phi)$ measures the $[\delta_M(\chi)m_\rho - \rho(\Phi)m_\rho s_\rho]^2$ differences in the entire unit cell after considering that $\rho(\chi) \cong \rho(\Phi)s_\rho$. The minimum value corresponds to $R_\delta(\Phi_T) \simeq \langle m_\rho \rangle I_{g^2}$ where $I_{g^2} \simeq 1.12$. It is shown that minimizing $R_\delta(\Phi)$ is essentially equivalent to maximizing the sum function $S_\delta(\Phi)$ [equation (21)], despite the presence of the m_ρ mask in the $R_\delta(\Phi)$ definition. In all the examples calculated by the author, the S_δ maximum (characterized by a sudden S_δ increase) is always the true solution Φ_T which stands out clearly from the false solutions.

(iv) The convergence of *SMAR* is achieved by alternately applying the χ and Φ (or *SMAR*) phasing formulas in each iteration. These are $\alpha^{\text{new}} = $ phase of $\text{FT}\{\rho(\Phi)s_\rho\}$ and $\varphi^{\text{new}} = $ phase of $\text{FT}\{\delta_M(\chi)m_\rho s_\rho\}$, respectively.

It has been found that at the start of a *SMAR* refinement, the zero part of the mask (created by converting the slightly negative density function values to zero) occupies 50% of V and increases by $\sim$5% after convergence. According to

$R_\delta(\Phi_T) \simeq 1.12\langle m_\rho \rangle$ the presence of the zero part of the mask leads to a drop in the R_δ value when convergence begins, since the volume of the regions with only a g contribution is reduced. When the number N of expected atoms is actively used (fast convergence mode), each *SMAR* iteration is supplemented with the *ipp* application, which increases significantly the volume of the zero part of the mask.

Finally, a brief reflection is in order. It is known that non-crystalline materials have a continuous diffraction pattern and that oversampling of the intensity data (Shannon, 1949) results in an overdetermined system of equations from which the phases can be solved (even at non-atomic resolution) (Miao *et al.*, 2000). It is also known that oversampling cannot be applied to crystals due to their 3D periodicity. Here it is shown that, in the case of crystals, the combination of δ_M and ρ each with $|\rho|$ produces two independent residuals while keeping the same unknowns. This also leads to overdetermination which should explain the observed efficiency of *SMAR*. It is also interesting to note that the *SMAR* algorithm is particularly well suited for Deep Learning due to its architecture.

APPENDIX *A*
Derivation of the general equation of δ direct methods

A1. Definition and peak analysis of the δ_M Fourier synthesis

The term δ_M is defined in the unit cell by

$$\delta_M(\mathbf{r}, \Phi_T) = \frac{c}{V}\sum_k \left(|E_k| - \langle|E|\rangle\right)\exp\left(i\varphi_k\right)\exp\left(-i2\pi\mathbf{kr}\right), \tag{31}$$

with $\{\varphi\} = \Phi_T$ and with c being an appropriate scaling constant. It can be reformulated by making use of the $\delta_M = \delta_P/2$ equality (Rius, 2012a), wherein δ_P is the Fourier synthesis which differs from (31) only in that $(|E_k| - \langle|E|\rangle)$ is replaced by $(|E_k|^2 - \langle|E|^2\rangle)$. Consequently, (31) can be written as

$$\delta_M(\mathbf{r}) = \frac{c}{2V}\sum_k \frac{\left(|E_k|^2 - \langle|E|^2\rangle\right)}{|E_k|}|E_k|\exp\left(i\varphi_k\right)\exp\left(-i2\pi\mathbf{kr}\right). \tag{32}$$

By expressing $(|E_k|^2 - |E|^2)$ and $|E_k|\exp(i\varphi_k)$ in terms of atom positions, it follows that

$$\delta_M(\mathbf{r}) =$$
$$\frac{c}{2V}\sum_k \frac{1}{|E_k|}\sum_j \hat{f}_j \sum_l \sum_{m(\neq l)}\hat{f}_l\hat{f}_m\exp\left[i2\pi\mathbf{k}(\mathbf{r}_j + \mathbf{r}_l - \mathbf{r}_m - \mathbf{r})\right], \tag{33}$$

with $\hat{f}_j$ denoting the normalized scattering factor of atom j. As can be derived from (33), δ_M has $N^3 - N^2$ peaks at $\mathbf{r} = \mathbf{r}_j + \mathbf{r}_l - \mathbf{r}_m$ ($l \neq m$), since $\exp\left[i2\pi\mathbf{k}(\mathbf{r}_j + \mathbf{r}_l - \mathbf{r}_m - \mathbf{r})\right]$ is the unit for all k. Of interest are the values of δ_M at the N atomic positions, *i.e.* at $\mathbf{r} = \mathbf{r}_l$ with $l = 1, N$. It can easily be verified that there are $N - 1$ superposed peaks contributing to $\delta_M(\mathbf{r}_l)$, *i.e.* those that satisfy the equation $\mathbf{r}_l = \mathbf{r}_j + \mathbf{r}_l - \mathbf{r}_m$ with $l \neq m$ and $j = m$, *e.g.* for $N = 3$ and $\mathbf{r}_2$ these are $\mathbf{r}_1 + \mathbf{r}_2 - \mathbf{r}_1$ and $\mathbf{r}_3 + \mathbf{r}_2 - \mathbf{r}_3$. To

estimate the total strength of the δ_M peak at $\mathbf{r} = \mathbf{r}_l$, expression (33) is rearranged into

$$\delta_M(\mathbf{r}) = \frac{c}{2V} \sum_l \hat{f}_l \sum_k \exp\left[i2\pi\mathbf{k}(\mathbf{r}_l - \mathbf{r})\right]$$
$$\times \sum_{m(\neq l)} \hat{f}_m \exp\left(-i2\pi\mathbf{k}\mathbf{r}_m\right) \sum_j \frac{1}{|E_k|}\hat{f}_j \exp\left(i2\pi\mathbf{k}\mathbf{r}_j\right),$$
$$(34)$$

with

$$\sum_j \frac{1}{|E_k|}\hat{f}_j \exp\left(i2\pi\mathbf{k}\mathbf{r}_j\right) = \exp\left(i\varphi_k\right), \qquad (35)$$

$$\sum_{m(\neq l)} \hat{f}_m \exp\left(-i2\pi\mathbf{k}\mathbf{r}_m\right) = |E_k| \exp\left(-i\varphi_k\right) - \hat{f}_l \exp\left(-i2\pi\mathbf{k}\mathbf{r}_l\right). \qquad (36)$$

For $r = r_l$, the $\exp\left[i2\pi\mathbf{k}(\mathbf{r}_l - \mathbf{r})\right]$ term in (34) becomes unity. In addition, if (35) and (36) are considered, expression (34) can be further simplified to

$$\delta_M(\mathbf{r}_l) = \frac{c}{2V}\hat{f}_l \sum_k\left[|E_k| - \hat{f}_l \exp\left(i\varphi_k\right)\exp\left(-i2\pi\mathbf{k}\mathbf{r}_l\right)\right] \quad (37)$$

$$= \frac{c}{2}\left[\left(\frac{\hat{f}_l N_k}{V}\right)\langle|E|\rangle_k - \hat{f}_l\rho(\mathbf{r}_l)\right], \qquad (38)$$

since $\sum_k |E_k| = \langle|E|\rangle_k N_k$ and, for an equal atom structure, it holds that

$$\frac{1}{V}\sum_k \hat{f}_l \exp\left(i\varphi_k\right)\exp\left(-i2\pi\mathbf{k}\mathbf{r}_l\right) = \rho(\mathbf{r}_l). \qquad (39)$$

The strength of an atomic peak in ρ is $\hat{f}_l N_k/V$, so (38) can be simplified to

$$\delta_M(\mathbf{r}_l) = c\frac{\langle|E|\rangle_k - (1/\sqrt{N})}{2}\rho(\mathbf{r}_l) \qquad (40)$$

by considering $\hat{f}_l = 1/\sqrt{N}$. Finally, by making

$$c = 2/\left[\langle|E|\rangle_k - (1/\sqrt{N})\right], \qquad (41)$$

the strength of $\delta_M(\mathbf{r}_l)$ is equal to $\rho(\mathbf{r}_l)$ in (40). The δ_M peaks placed at atomic positions compose the set of type A peaks.

Let us consider the remaining δ_M peaks at $\mathbf{r} = \mathbf{r}_j + \mathbf{r}_l - \mathbf{r}_m$ with $l \neq m$ and $j \neq m$, which form the set of type B peaks. For a given B peak, the corresponding $\mathbf{r}$ position is obtained by adding the $\mathbf{r}_j - \mathbf{r}_m$, $j \neq m$, interatomic vector to the atomic position vector $\mathbf{r}_l$, so the superposition of $\mathbf{r}$ with an atomic position is accidental. Consequently, the strength of a (single) B-type peak, e.g. at $\mathbf{r}_{jlm}$, is weaker than that of a $\delta_M(\mathbf{r}_l)$ peak (formed by the superposition of $N - 1$ single peaks). According to (40), it holds that

$$\text{strength of } \delta_M\left(\mathbf{r}_{jlm}\right) \approx \text{strength of }\left[\rho(\mathbf{r}_l)/(N - 1)\right]. \qquad (42)$$

A2. The general equation for $\delta_{(M)}$ direct methods

According to the above results, $\delta_M(\Phi_T)$ contains two types of positive peaks (A and B). Table 2 lists their peak strengths

and positions. The stronger peaks of type A correspond to ρ and are resolved in the Fourier map. The $N - 1$ times weaker peaks of type B are located at positions other than the atomic positions (coincidence is accidental). Due to their large number, i.e. $N(N - 1)^2$ [for comparison, there are only $N(N - 1)$ non-origin peaks in the modulus function], these peaks must have a strong overlap in the unit cell. The peaks of type B are the main constituents of the g function. Consequently, δ_M can be considered as the sum of both contributions, i.e.

$$\delta_M(\mathbf{r}) = \rho(\mathbf{r}) + g(\mathbf{r}) \quad \forall\, \mathbf{r} \subset V, \qquad (43)$$

which is called the general equation of $\delta_{(M)}$ direct methods. In the following, the subscript M is left out of the equation name for the sake of generality. (Note that the g function could also contain contributions from even weaker unconsidered peaks.)

A3. The δ_M Fourier synthesis expressed with the signs of the $|E| - \langle|E|\rangle$ differences in the phase terms

In general, the Fourier coefficients of δ_M and ρ have different moduli and may differ in phase values. This is not obvious when comparing (2) and (4) because in both expressions the corresponding Fourier terms have the same φ phase values (which is quite useful for programming). However, when the absolute values $||E| - \langle|E|\rangle|$ are used in the synthesis, a phase shift $\Delta\varphi$ must be added to each φ to take into account the sign. Consequently, while the Fourier coefficients of ρ are $|E|\exp(i\varphi)$, those of δ_M are $||E| - \langle|E|\rangle|\exp[i(\varphi + \Delta\varphi)]$, with $\Delta\varphi = 0$ for $|E| > \langle|E|\rangle$ (strong reflections) and $\Delta\varphi = \pi$ for $|E| < \langle|E|\rangle$ (weak reflections).

APPENDIX B
Solution of integral $I_{g^2} = \int_V g^2\, dV$

Assuming that the integrand of I_{g^2} is $g^2 = \left[\delta_M(\Phi_T) - \rho(\Phi_T)\right]^2$, squaring the binomial gives

$$I_{g^2} = \int_V \left[\delta_M^2(\Phi_T) + \rho^2(\Phi_T) - 2\delta_M(\Phi_T)\,\rho(\Phi_T)\right] dV. \qquad (44)$$

The integral of the first two summands in (44) when expressed in terms of the respective Fourier coefficients is

$$\int_V \left(\delta_M^2 + \rho^2\right) dV = \frac{1}{V}\sum_k\left\{\left[c\left(|E_k| - \langle|E|\rangle\right)\right]^2 + |E_k|^2\right\}$$

$$= \frac{1}{V}\left[\left(c^2 + 1\right)\sum_k|E_k|^2 - c^2\langle|E|\rangle^2\right]. \qquad (45)$$

If one proceeds analogously with the third summand in (44), it follows that

$$-2\int_V \delta_M(\Phi_T)\rho(\Phi_T)\,dV =$$

$$\frac{-2c}{V^2}\sum_k\sum_{k'}\left(|E_k| - \langle|E|\rangle\right)|E_{k'}|\exp\left[i(\varphi_k + \varphi_{k'})\right]$$

$$\times \int_V \exp\left[-i2\pi(k+k')\mathbf{r}\right]d\mathbf{r}. \tag{46}$$

This means that the integral is zero except for $k' = -k$ where it becomes V, so that the final term in (46) can be approximated by

$$\frac{-2c}{V}\sum_k\left(|E_k| - \langle|E|\rangle\right)|E_k|\exp\left[i(\varphi_k - \varphi_k)\right]. \tag{47}$$

The values of the exponentials in (47) are 1, so that it reduces to

$$\frac{-2c}{V}\sum_k\left(|E_k| - \langle|E|\rangle\right)|E_k|. \tag{48}$$

By adding (45) and (48) and normalizing by $\int_V \rho^2\,dV = N_k\langle|E|^2\rangle_k/V = N_k/V$, the final expression of the integral I_{g^2} is obtained, namely

$$I_{g^2} = (c-1)^2\langle|E|^2\rangle_k - c(c-2)\left(\langle|E|\rangle_k\right)^2. \tag{49}$$

From the theory of the distribution of $|E|$ values, the values of $\langle|E|^2\rangle$, $\langle|E|\rangle$ (acentric case) and $c \simeq 2/\langle|E|\rangle$ can be derived, *i.e.* 1.00, 0.89 and 2.25, respectively, so that (49) gives $I_{g^2} \simeq 1.12$. Note that I_{g^2} only depends on $|E|$ and $|E| - \langle|E|\rangle$, since the phases cancel out. Consequently, the values of I_{g^2} for g^2 equal to $[\delta_M(\Phi_T) - \rho(\Phi_T)]^2$ and to $[\delta_M(\chi_T) - \rho(\chi_T)]^2$ are identical.

References

Antel, J., Shedrick, G. M., Bats, J. W., Kessler, H. & Müller, A. (1995). Unpublished work.

Burla, M. C., Caliandro, R., Giacovazzo, C. & Polidori, G. (2010). *Acta Cryst.* A**66**, 347–361.

Capitani, G. C., Mugnaioli, E., Rius, J., Gentile, P., Catelani, T., Lucotti, A. & Kolb, U. (2014). *Am. Mineral.* **99**, 500–510.

Cochran, W. (1952). *Acta Cryst.* **5**, 65–67.

Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.

Lipson, H. & Cochran, W. (1966). *The Determination of Crystalline Structures. The Crystalline State*, Vol. III, edited by W. L. Bragg, pp. 324–325. London: G. Bell and Sons Ltd.

Main, P. (1975). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, p. 99. Copenhagen: Munksgaard.

Miao, J., Kirz, J. & Sayre, D. (2000). *Acta Cryst.* D**56**, 1312–1315.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.

Oliver, J. D. & Strickland, L. C. (1984). *Acta Cryst.* C**40**, 820–824.

Privé, G., Anderson, D. H., Wesson, L., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1400–1409.

Rius, J. (1993). *Acta Cryst.* A**49**, 406–409.

Rius, J. (2011). *XLENS_v1: A Computer Program for Solving Crystal Structures from Diffraction Data by Direct Methods*. Institut de Ciència de Materials de Barcelona, CSIC, Spain.

Rius, J. (2012*a*). *Acta Cryst.* A**68**, 77–81.

Rius, J. (2012*b*). *Acta Cryst.* A**68**, 399–400.

Rius, J. (2014). *IUCrJ*, **1**, 291–304.

Rius, J. (2020). *Acta Cryst.* A**76**, 489–493.

Rius, J., Crespi, A. & Torrelles, X. (2007). *Acta Cryst.* A**63**, 131–134.

Rius, J., Mugnaioli, E., Vallcorba, O. & Kolb, U. (2013). *Acta Cryst.* A**69**, 396–407.

Rius, J., Sañé, J., Miravitlles, C., Amigó, J. M. & Reventós, M. M. (1995). *Acta Cryst.* A**51**, 268–270.

Rius, J. & Torrelles, X. (2021). *Acta Cryst.* A**77**, 339–347.

Rius, J. & Torrelles, X. (2022). *Acta Cryst.* A**78**, 473–481.

Rius, J., Vallcorba, O., Crespi, A. & Colombo, F. (2017). *Z. Kristallogr.* **232**, 827–834.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Schäfer, M., Sheldrick, G. M., Bahner, I. & Lackner, H. (1998). *Angew. Chem. Int. Ed.* **37**, 2381–2384.

Shannon, C. E. (1949). *Proc. Inst. Radio Eng.* **37**, 10–21.

Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.

Sheldrick, G. M. (2015). *Acta Cryst.* A**71**, 3–8.

Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller & Usón, I. (2012). *International Tables for Crystallography*, Vol. F, *Crystallography of Biological Macromolecules*, edited by E. Arnold, D. M. Himmel & M. G. Rossmann, ch. 16.1, pp. 413–442. Chichester: Wiley.

Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Acta Cryst.* D**49**, 179–181.

Zachariasen, W. H. (1952). *Acta Cryst.* **5**, 68–73.