

A new density-modification procedure extending the application of the recent $|\rho|$ -based phasing algorithm to larger crystal structures

Jordi Rius* and Xavier Torrelles

Institut de Ciència de Materials de Barcelona, CSIC, Campus de la UAB, Bellaterra, Catalonia 08193, Spain.

*Correspondence e-mail: jordi.rius@icmab.es

Received 27 January 2021

Accepted 10 May 2021

Edited by I. Margiolaki, University of Patras, Greece

Keywords: S_M phasing algorithm; *ipp* procedure; $|\rho|$ -based phasing residual; direct methods; origin-free modulus sum function; structure solution.

Supporting information: this article has supporting information at journals.iucr.org/a

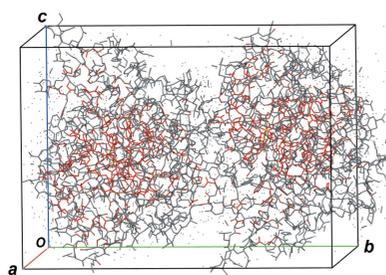
The incorporation of the new peakness-enhancing fast Fourier transform compatible *ipp* procedure (*ipp* = inner-pixel preservation) into the recently published S_M algorithm based on $|\rho|$ [Rius (2020). *Acta Cryst A* **76**, 489–493] improves its phasing efficiency for larger crystal structures with atomic resolution data. Its effectiveness is clearly demonstrated via a collection of test crystal structures (taken from the Protein Data Bank) either starting from random phase values or by using the randomly shifted modulus function (a Patterson-type synthesis) as initial ρ estimate. It has been found that in the presence of medium scatterers (e.g. S or Cl atoms) crystal structures with $1500 \times c$ atoms in the unit cell (c = number of centerings) can be routinely solved. In the presence of strong scatterers like Fe, Cu or Zn atoms this number increases to around $5000 \times c$ atoms. The implementation of this strengthened S_M algorithm is simple, since it only includes a few easy-to-adjust parameters.

1. Introduction

The novel $S_{M,|\rho|}$ phasing function is rooted in the Z_R origin-free modulus sum function, a nearly 30 years-old direct-methods phasing function (Rius, 1993). Both mainly differ in (i) the introduction of ‘Fourier transform’ calculations instead of the complex manipulation of ‘structure invariants’ (Rius *et al.*, 2007); (ii) the replacement of $\rho^2(\mathbf{r})$ by $|\rho(\mathbf{r})|$ at each point \mathbf{r} of the unit cell by using the property that $\rho^2(\mathbf{r})$ and $|\rho(\mathbf{r})|$ are positive-definite functions with similar shape (Rius, 2020). The resulting $S_{M,|\rho|}$ phase refinement function is defined by

$$\begin{aligned} S_{M,|\rho|}(\Phi) &= \frac{1}{V} \sum_K (|E_K| - \langle |E| \rangle) |C_{-K}(\Phi)| \\ &= \frac{1}{V} \sum_K (|E_K| - \langle |E| \rangle) \{ \exp[i\alpha_K(\Phi)] C_{-K}(\Phi) \} \quad (1) \end{aligned}$$

in which the K sum extends over all reflections (*i.e.* strong and weak ones), $|E_K|$ denotes the experimental structure-factor modulus with $\langle |E| \rangle$ being their average value, V is the volume of the unit cell, and Φ denotes the collectivity of φ phases involved in the computation of ρ . The $C_K(\Phi) = |C_K(\Phi)| \exp[i\alpha_K(\Phi)]$ complex quantity is the Fourier transform of the $|\rho(\Phi)|$ density function in terms of the Φ structure-factor phases to be refined. Their refinement is achieved by maximizing $S_{M,|\rho|}(\Phi)$ through the iterative $S_{M,|\rho|}$ fast Fourier transform (FFT) algorithm. This algorithm has been developed in *P1*, since this symmetry is advantageous to *ab initio* phase refinements (Sheldrick & Gould, 1995). (Mathematically, however, nothing prevents its implementation as a full-symmetry algorithm.) As demonstrated by Rius (2020),



OPEN ACCESS

maximizing $S_{M,|\rho|}$ is equivalent to minimizing the phasing residual

$$R_M(\Phi) = \int_V [\delta_M(\Phi) - k|\rho(\Phi)|]^2 dV, \quad (2)$$

which measures the discrepancy between $\delta_M(\Phi)$ and $|\rho(\Phi)|$. In integral (2), $\delta_M(\Phi)$ and k are, respectively, the inverse Fourier transform of $(|E_K| - \langle |E| \rangle) \exp[i\alpha_K(\Phi)]$ and a suitable scaling constant (Rius, 2012). Since integral (2) can be exactly worked out in terms of Φ , its minimum value should correspond (for data reaching atomic resolution) to the true solution or an equivalent, to the maximum of the correlation coefficient

$$CC_M = \left\{ \frac{|\sum_K (|E_K| \times |C_K(\Phi)|)|^2}{\sum_K |E_K|^2 \times \sum_K |C_K(\Phi)|^2} \right\}^{1/2} \quad (3)$$

measuring the agreement between experimental and calculated modulus functions. CC_M rapidly increases at the beginning of the iterative $S_{M,|\rho|}$ phase refinement, gradually stabilizes as it progresses and suddenly increases at the end (normally by 0.035–0.045 in just a few cycles) indicating that convergence has been attained.

One common feature of most iterative phase refinement algorithms working at atomic resolution and alternating between real- and reciprocal-space calculations is the density modification of the intermediate Fourier maps. Peak-picking is the simplest procedure which has been applied in the *Shake-and-Bake* approach (Weeks *et al.*, 1993; Miller *et al.*, 1993), *i.e.* once the centers and heights of the N highest peaks in the map have been determined (N is the expected number of non-H atoms in the unit cell) these are used to calculate the new structure-factor estimates. For large structures, however, application of the FFT algorithm (Cooley & Tukey, 1965) to

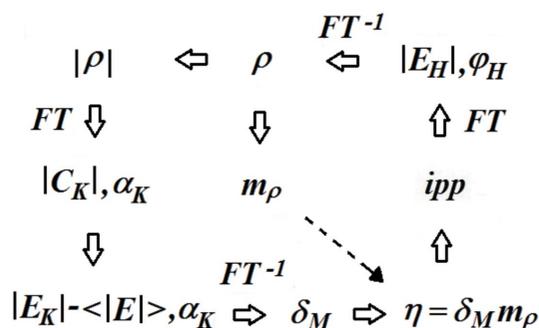


Figure 1
The recursive S_M -*ipp* phase refinement algorithm with enhanced peakness: (upper right corner) φ phase estimates (either initial or updated values) are combined with experimental $|E|$'s to obtain ρ , $|\rho|$ and m_ρ (the latter is stored). Next, the Fourier transform of $|\rho|$ is calculated leading to new $|C|$ and α values, and the former are used in the calculation of CC_M . The new α values are combined with the experimental $(|E| - \langle |E| \rangle)$ (lower left corner), and their inverse Fourier transform, δ_M , is calculated. In the next step, function δ_M is multiplied with the stored m_ρ mask to give the η product function. Peakness in η is enhanced by applying the *ipp* density-modification procedure and, finally, the Fourier transform of the modified η supplies the updated φ phases. [Initial sets of φ estimates investigated in this article are either Φ_{rnd} (random phase values) or Φ_M (phase values corresponding to the Fourier coefficients of M' , *i.e.* the randomly shifted modulus function).]

the Fourier map is more efficient than direct calculation of the structure factors. In the literature other density-modification procedures can be found, *e.g.* in *SIR2000* the density fraction above a 2.0–2.5% threshold is kept in each map inversion, the rest set to zero [Burla *et al.* (2000) and Shiono & Woolfson (1992) for a related procedure]; Caliendo *et al.* (2008) have later shown the convenience of increasing this threshold when the resolution of the data is poorer than atomic. Also highly effective but more complicated is the density-modification scheme incorporated in *ACORN2* (Dodson & Woolfson, 2009). Alternatively, peakness in the electron-density function can be enhanced by multiplying it with a mask having unit Gaussians only at the previously determined peak positions (the rest being zero). This modification is part of Sheldrick's intrinsic phasing procedure (Sheldrick, 2015) and allows the posterior application of the FFT algorithm. In the present work, the alternative peakness-enhancing *ipp* procedure (*ipp* = inner-pixel preservation) is described. It directly operates on the $\eta = \delta_M m_\rho$ product function of the S_M algorithm wherein m_ρ is the mask relating $|\rho|$ to ρ through the expression

$$|\rho(\Phi)| = \rho(\Phi)m_\rho(\Phi). \quad (4)$$

According to Rius (2020), the values of m_ρ are 1 (for $\rho > 0$), 0 (for ρ between 0 and $-t_\rho\sigma_\rho$) and -1 (for $\rho < -t_\rho\sigma_\rho$) with σ_ρ^2 being the variance of $\rho(\Phi)$ and $t_\rho \simeq 2.5$. Hereafter $S_{M,|\rho|}$ will be shortened to S_M for simplicity.

2. The S_M phasing algorithm with enhanced peakness: the *ipp* procedure

The phasing residual (2) can be minimized with the S_M algorithm (Rius, 2020), *i.e.* by the iterative application of the modified tangent formula

$$\varphi_H^{\text{new}} = \text{phase of } \left\{ \int_V \delta_M(\mathbf{r}, \Phi)m_\rho(\mathbf{r}, \Phi) \exp(i2\pi\mathbf{H}\mathbf{r}) dV \right\} \quad (5)$$

which corresponds to the angular part of the Fourier transform within brackets. One characteristic of the S_M algorithm is the presence of the $\eta = \delta_M m_\rho$ product function. To enhance the peakness of η , the simple *ipp* procedure based on the preservation of the inner-peak pixels has been added to S_M , giving rise to the S_M -*ipp* algorithm (Fig. 1). This procedure consists of two well differentiated parts:

(i) Peak search in the η product function. The lowest value of η which is accepted as a peak is fixed by the $t_\eta \sigma_\eta$ threshold (σ_η^2 is the variance of η , and t_η a parameter allowing tuning of the threshold and normally ranging between 3.5 and 4.0). The η peaks are searched by looking for the density values of all 26 nearest grid points around a given central pixel (satisfying the above threshold criterion). This (x_o, y_o, z_o) central pixel is considered a η peak if its density value is larger than the values of all its 26 nearest neighbor pixels, *i.e.* 8 $(x_o \pm 1, y_o \pm 1, z_o \pm 1)$; 4 $(x_o, y_o \pm 1, z_o \pm 1)$; 4 $(x_o \pm 1, y_o, z_o \pm 1)$; 4 $(x_o \pm 1, y_o \pm 1, z_o)$; 2 $(x_o, y_o, z_o \pm 1)$; 2 $(x_o, y_o \pm 1, z_o)$; 2 $(x_o \pm 1, y_o, z_o)$ (Rollet, 1965). If this is the case, the density value and the pixel coordinates of the central pixel are stored. At the end, the N_η stored peaks are ordered in decreasing strength. (Note, t_η and N_η are inversely related.)

Table 1

Data sets from the Protein Data Bank (PDB) used to compare the S_M -*ipp* and S_M phasing algorithms corresponding to compounds with only weak scatterers (top five) or with weak and medium scatterers (remaining).

Residues = number of residues; c = number of centerings; N = number of non-H atoms in the unit cell (PDB); M and H_2O = number of medium scatterers and refined water molecules; %Sol = solvent volume percentage; d_{\min} = minimum d spacing in Å of used reflection data; T = data collection temperature in K. (1a7y, 1ob4, 1a7z, 1alz, 2erl, 1a0m are rotating-anode or sealed-tube data sets; otherwise synchrotron data.)

| PDB code | Compound | Residues | Space group | N/c | M/c | H_2O/c | %Sol | d_{\min} | T |
|----------|----------------------------------|----------|--------------|-------|-------|----------|------|------------|-----|
| 1a7y | Actino D ⁽¹⁾ | 33 | $P1$ | 314 | – | 44 | 18 | 0.94 | 133 |
| 3sbn | Trichovirin ⁽²⁾ | 30 | $P2_1$ | 444 | – | 32 | 24 | 0.95 | 100 |
| 1ob4 | Cephaibol A ⁽³⁾ | 17 | $P2_12_12$ | 548 | – | 60 | 22 | 1.00 | 100 |
| 1a7z | Actino Z3 ⁽¹⁾ | 22 | $P2_12_12_1$ | 1228 | 8Cl | 4 | 49 | 0.95 | 173 |
| 1alz | Gramicidin A ⁽⁴⁾ | 34 | $P2_12_12_1$ | 1348 | – | 4 | 30 | 1.00 | 120 |
| 1byz | Alpha1-peptide ⁽⁵⁾ | 52 | $P1$ | 479 | 1Cl | 30 | 27 | 0.91 | 100 |
| 2erl | Er-1 pheromone ⁽⁶⁾ | 40 | $C2$ | 656 | 14S | 44 | 20 | 1.00 | 273 |
| 1p9g | Antifungal ⁽⁷⁾ | 41 | $P2_1$ | 702 | 20S | 122 | 23 | 1.00 | 283 |
| 3nir | Crambin ⁽⁸⁾ | 48 | $P2_1$ | 902 | 12S | 196 | 31 | 1.00 | 100 |
| 1a0m | Conotoxin ⁽⁹⁾ | 34 | $I4$ | 1144 | 40S | 168 | 24 | 1.10 | 286 |
| 4lzt | Lysozyme ⁽¹⁰⁾ | 129 | $P1$ | 1183 | 10S | 139 | 32 | 1.00 | 295 |
| 1f94 | Bucandin ⁽¹¹⁾ | 63 | $C2$ | 1232 | 20S | 236 | 35 | 1.02 | 100 |
| 1hhu | Balhimycin ⁽¹²⁾ | 28 | $P2_1$ | 1310 | 16Cl | 250 | 22 | 0.89 | 100 |
| 3odv | Kaliotoxin ⁽¹³⁾ | 76 | $P1$ | 1392 | 32S | 180 | 20 | 1.00 | 100 |
| 3psm | Plant defensin ⁽¹⁴⁾ | 94 | $P2_1$ | 1882 | 16S | 366 | 45 | 0.98 | 100 |
| 3bcj | Aldose reductase ⁽¹⁵⁾ | 316 | $P2_1$ | 7308 | 26S | 1374 | 43 | 0.78 | 15 |
| | | | | | + 3P | | | 0.85 | |

(1) Schäfer *et al.* (1998); (2) Gessmann *et al.* (2012); (3) Bunkóczy *et al.* (2003); (4) Burkhart *et al.* (1998); (5) Privé *et al.* (1999); (6) Anderson *et al.* (1996); (7) Xiang *et al.* (2004); (8) Schmidt *et al.* (2011); (9) Hu *et al.* (1998); (10) Walsh *et al.* (1998); (11) Kuhn *et al.* (2000); (12) Lehmann *et al.* (2002); (13) Pentelute *et al.* (2010); (14) Song *et al.* (2011); (15) Zhao *et al.* (2008).

(ii) Density modification of η . If $N_\eta > N$, then for each one of the N highest-ranked η peaks, the density values of the 26+1 inner-peak pixels are preserved. The density-modification procedure finishes by setting to zero all pixels of η not having preserved density values. For $N_\eta \leq N$, the inner pixels of all N_η peaks will have preserved density values. The Fourier transform of the modified η yields the new φ estimates.

Notice that accurate peak center positions are not necessary for the application of the *ipp* procedure; consequently, no peak interpolation is needed. Notice, also, that it is compatible with the ‘random omit maps’ strategy introduced in direct methods by Sheldrick (Usón & Sheldrick, 1999). For illustrative purposes, a successful S_M -*ipp* phase refinement obtained with starting random (rnd) phases and with $t_\eta = 3.7$ is reproduced in Fig. 2. It is interesting to note that only $N_\eta(1)$ is

smaller than N (the number 1 in parentheses indicates the iteration number).

Compared with the S_M algorithm in Rius (2020) in which all reflections participate in the computation of the ρ synthesis, S_M -*ipp* works better if ρ is calculated with only those H reflections for which $|E| \geq |E|_{\min}$ with $|E|_{\min} \simeq 1.0$, *i.e.* Φ only includes the large and moderate $|E|$ values [however, the calculation of the δ synthesis remains unchanged, *i.e.* it extends to all K reflections (Fig. 1)]. Notice that the faster calculation of ρ in S_M -*ipp* counteracts the extra computing time due to *ipp*. Concerning this point, a test performed on data set 1pwl showed that the duration of one iteration in S_M -*ipp* and in S_M is very similar. The S_M -*ipp* algorithm has been programmed in a modified version of the *XLENs_v1* code (Rius, 2011). In the test calculations, N always includes, besides the number of protein atoms, the number of solvent ones, *i.e.* water molecules.

3. The modulus function as initial estimate of ρ

It is clear that the phasing process not only depends on the phasing algorithm but also on the starting phase values. In Rius (2020), the S_M algorithm was only tested by assigning random values to the initial phases, $\Phi_{\text{rnd}} = \{\varphi_{\text{rnd}}\}$. However, the ideal situation for any phasing algorithm is to start with phase values derived from initial ρ estimates (ρ_{ini}) containing structural information. Since the M modulus synthesis is a Patterson-type synthesis (Ramachandran & Raman, 1959), it can be regarded as the sum of N weighted shifted images of the crystal structure (or its enantiomorph) (Wrinch, 1939; Buerger, 1950). Consequently, it contains valuable structural information and can be taken as ρ_{ini} . The success of the

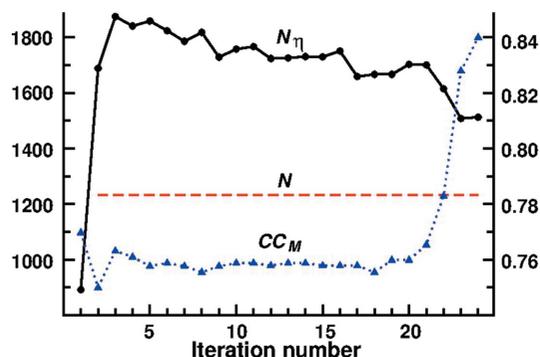


Figure 2

S_M -*ipp* phasing with initial random phases (Φ_{rnd}): variation of N_η and CC_M with the iteration number for data set 1a7z ($t_\eta = 3.7$). N is the number of non-H atoms in the unit cell.

Table 2

PDB data sets used to test the S_M -*ipp* and S_M phasing algorithms corresponding to compounds with strong scatterers.

Residues, space group, c , %Sol and d_{\min} as in Table 1. N = number of non-H atoms in the unit cell (PDB); M and S = number of medium and strong scatterers; H_2O = number of refined water molecules. Data sets 2bf9, 8rxn and 1c7k measured at room temperature; otherwise at 100 K.

| PDB code | Compound | Residues | Space group | N/c | $(M+S)/c$ | H_2O/c | %Sol | d_{\min} |
|----------|-------------------------------------|----------|--------------|-------|-------------|----------|------|------------|
| 2bf9 | aPP ⁽¹⁾ | 36 | $C2$ | 768 | 2Zn | 164 | 31 | 1.00 |
| 8rxn | Rubredoxin ⁽²⁾ | 52 | $P2_1$ | 1010 | 12S+2Fe | 204 | 35 | 1.00 |
| 1w3m | Tsushimycin ⁽³⁾ | 132 | $P1$ | 1276 | 10Cl+24Ca | 191 | 35 | 1.00 |
| 2ov0 | Amicyanin ⁽⁴⁾ | 105 | $P2_1$ | 2060 | 6S+2P+2Cu | 432 | 34 | 0.95 |
| 1c75 | Cytochrome 553 ⁽⁵⁾ | 71 | $P2_12_12_1$ | 2660 | 12S+4Fe | 500 | 38 | 0.97 |
| 3d1p | Transferase ⁽⁶⁾ | 120 | $C2$ | 2702 | 2S+2Cl+4Se | 498 | 29 | 0.95 |
| 1pwl | Aldose reductase Br ⁽⁷⁾ | 316 | $P1$ | 3030 | 14S+3P+1Br | 429 | 25 | 1.10 |
| 1a6m | Myoglobin ⁽⁸⁾ | 151 | $P2_1$ | 3154 | 8S+2Fe | 372 | 36 | 1.00 |
| 41au | Geodin ⁽⁹⁾ | 161 | $P2_1$ | 3278 | 2Ca+ 6Se† | 740 | 40 | 0.99 |
| 1eb6 | Deuterolysin ⁽¹⁰⁾ | 177 | $P2_1$ | 3300 | 12S+2Zn | 518 | 39 | 1.00 |
| 1b0y | H42Q ⁽¹¹⁾ | 85 | $P2_12_12_1$ | 3348 | 36S+16Fe | 824 | 30 | 0.90 |
| 1x8q | Nitrophorin 4C ⁽¹²⁾ | 184 | $C2$ | 3662 | 10S+2Fe | 720 | 24 | 0.90 |
| 2fdn | Ferredoxin ⁽¹³⁾ | 55 | $P4_322$ | 3964 | 128S+64Fe | 768 | 35 | 1.00 |
| 3fsa | Azurin ⁽¹⁴⁾ | 125 | $P2_12_12_1$ | 4488 | 36S+4Cu | 856 | 38 | 1.00 |
| 1c7k | Endoprotease Zn ⁽¹⁵⁾ | 132 | $P2_12_12_1$ | 4532 | 12S+4Ca+4Zn | 464 | 37 | 1.00 |
| 3ks3 | H. C. anhydrase II ⁽¹⁶⁾ | 260 | $P2_1$ | 5626 | 6S+2Zn | 962 | 41 | 0.95 |
| 1heu | L. A. dehydrogenase ⁽¹⁷⁾ | 748 | $P1$ | 7618 | 58S+4Cd | 1297 | 50 | 1.15 |

† Four disordered Se positions. (1) Glover *et al.* (1983); (2) Dauter *et al.* (1992); (3) Bunkóczy *et al.* (2005); (4) Carrell *et al.* (unpublished); (5) Benini *et al.* (2000); (6) Nocek *et al.* (unpublished); (7) El-Kabbani *et al.* (2004); (8) Vojtěchovský *et al.* (1999); (9) Fanfrlik *et al.* (2013); (10) McAuley *et al.* (2001); (11) Parisini *et al.* (1999); (12) Kondrashov *et al.* (2004); (13) Dauter *et al.* (1997); (14) Sato *et al.* (2009); (15) Kurisu *et al.* (2000); (16) Avvaru *et al.* (2010); (17) Meijers *et al.* (2001).

phasing process will obviously depend on the capability of the phasing algorithm to develop one incomplete shifted image of the crystal structure while (gradually) suppressing the rest (working in $P1$ allows selection of one arbitrary image). The phasing process is greatly facilitated by the presence of a reduced number of strong scatterers in the unit cell with their corresponding images standing out from the rest [this justifies the separate treatment in the test calculations of compounds with weak, medium (atoms with $Z < 19$) and strong scatterers ($Z \geq 19$)]. In multisolution phasing methods, each phase refinement trial requires a different ρ_{ini} . This can be achieved by shifting the experimental M by a randomly generated $\mathbf{u} = \mathbf{OO}'$ vector to obtain the correspondingly shifted M' function (O and O' are the respective origins). The Fourier coefficients of M' are $|E_H| \exp(\phi'_H)$ with $\phi'_H = -i2\pi\mathbf{H}\mathbf{u}$ and $\Phi_{M'} = \{\phi'_H\}$. In this way each trial follows a different refinement path (in the test calculations, the sequence of \mathbf{u} vectors is the same for all data sets). The number of selected phase refinement trials (N_{trials}) is either 5, 25 or 50 depending on the success rate; the maximum number of allowed iterations per trial is always $N_{\text{iter(max)}} = 1000$ (excepting 3bcj with 200).

4. Comparison of the phasing efficiencies of the S_M -*ipp* and S_M algorithms

The efficiencies of the S_M -*ipp* and S_M algorithms have been calculated for both Φ_{rnd} and $\Phi_{M'}$. For simplicity, the various phase refinement strategies are specified by A1, A2, B1, B2, *i.e.* A1: Φ_{rnd} with S_M -*ipp*; A2: Φ_{rnd} with S_M ; B1: $\Phi_{M'}$ with S_M -*ipp*; B2: $\Phi_{M'}$ with S_M .

The compounds participating in the test calculations are listed in Tables 1 and 2. For those compounds in Table 1 only containing weak scatterers, the checked strategies are A1, A2 and B1 (Table 3). In the case of compounds with medium/

strong scatterers (Tables 1 and 2), the investigated strategies are B1 and B2 (Tables 4, 5 and 6). To make comparisons between strategies stricter, corresponding refinement trials started with the same set of randomly generated phase values.

4.1. Compounds with only weak scatterers

The data sets used in the tests of crystal structures with only weak scatterers are 1a7y, 3sbn, 1ob4, 1a7z and 1alz (Table 1). The first three data sets belong to small crystal structures and the last two to relatively large ones. Of these, 1a7z corresponds to a Cl-containing compound with 1228 atoms in the unit cell. In spite of the presence of Cl, it has been included in this section because the refinement protocol deposited in the Protein Data Bank (PDB) indicates that one Cl is partially occupied and the other has a rather large B value, so that their scattering powers are considerably reduced. The last data set (1alz) corresponds to the notoriously difficult crystal structure of gramicidin with 1348 C, N and O atoms in the unit cell and with nearly 25% of the atoms showing positional disorder.

Of the two A1 and A2 phasing strategies, the best one is A1 (Table 3). Compared with A2, A1 yields the smallest $\langle N_{\text{iter}} \rangle$ values and the largest number of successful trials for all five tested data sets, *i.e.* the correct solutions are found much faster when *ipp* is applied. The faster convergence of A1 is illustrated in Fig. 3 for data sets 3sbn and 1a7z. In the case of gramicidin, two correct solutions are obtained with A1 (trial 21 with $N_{\text{iter}} = 136$ and trial 45 with $N_{\text{iter}} = 520$) which represents one solution every 2.5 h using a desk computer (3.4 GHz); however, with A2 no correct solution was found. Regarding the A1 and B1 strategies, inspection of Table 3 indicates that A1 converges somewhat faster than B1 and is superior in the case of gramicidin (B1 gives no correct solutions).

Table 3

Application of the S_M -*ipp* and S_M algorithms to crystal structures only containing weak scatterers (A1, A2 and B1 phasing strategies).

The t_p parameter controlling the threshold of the m_p mask is always 2.50. N/c as in Table 1; N_p = number of peaks showing up in the final E map above the $n\sigma_p$ threshold; CC_M = correlation coefficient between experimental and calculated modulus function; N_{iter} = number of iterations to achieve convergence (n.c. = no convergence in 1000 iterations); t_η is the parameter controlling the number N_η of strongest η peaks; $Q = N_\eta(2)/N$.

| PDB code | N/c | Phasing strategy | N_p/c (n) | CC_M | $\langle N_{iter} \rangle$ for 5, 25 or 50 trials | t_η | Q |
|----------|-------|------------------|-----------------|--------|--|----------|-----|
| 1a7y | 314 | A1 | 376 (1.1) | 0.86 | $\langle N_{iter} \rangle = 37$ (25 \times) | 4.0 | 1.0 |
| | | A2 | 363 (1.1) | 0.86 | $\langle N_{iter} \rangle = 105$ (25 \times) | – | – |
| | | B1 | 370 (1.1) | 0.85 | $\langle N_{iter} \rangle = 100$ (25 \times) | 4.0 | 0.9 |
| 3sbn | 444 | A1 | 449 (1.1) | 0.82 | $\langle N_{iter} \rangle = 112$ (23 \times); n.c. (2 \times) | 3.7 | 1.4 |
| | | A2 | 456 (1.1) | 0.85 | $\langle N_{iter} \rangle = 558$ (13 \times); n.c. (12 \times) | – | – |
| | | B1 | 456 (1.1) | 0.82 | $\langle N_{iter} \rangle = 139$ (21 \times); n.c. (4 \times) | 3.7 | 1.2 |
| 1ob4 | 548 | A1 | 564 (1.1) | 0.86 | $\langle N_{iter} \rangle = 308$ (10 \times); n.c. (15 \times) | 3.7 | 1.1 |
| | | A2 | 573 (1.1) | 0.87 | $\langle N_{iter} \rangle = 392$ (3 \times); n.c. (22 \times) | – | – |
| | | B1 | 569 (1.1) | 0.86 | $\langle N_{iter} \rangle = 446$ (8 \times); n.c. (17 \times) | 3.7 | 1.0 |
| 1a7z | 1228 | A1 | 1279 (1.1) | 0.83 | $\langle N_{iter} \rangle = 133$ (22 \times); n.c. (3 \times) | 3.7 | 1.4 |
| | | A2 | 1372 (1.1) | 0.85 | $\langle N_{iter} \rangle = 334$ (16 \times); n.c. (9 \times) | – | – |
| | | B1 | 1281 (1.1) | 0.83 | $\langle N_{iter} \rangle = 338$ (23 \times); n.c. (2 \times) | 3.7 | 1.0 |
| 1alz | 1348 | A1 | 1308 (1.5) | 0.84 | 136, 520; n.c. (48 \times) | 3.8 | 1.2 |
| | | A2 | – | – | n.c. (50 \times) | – | – |
| | | B1 | – | – | n.c. (50 \times) | 3.8 | 0.9 |

4.2. Crystal structures with only medium scatterers

The application of strategy B1 to ten compounds containing medium scatterers (1byz, 2erl, 1p9g, 3nir, 1a0m, 4lzt, 1f94, 1hhu, 3odv and 3psm) is summarized in Table 4. In most cases (nine out of ten) phase refinements performed smoothly, *i.e.* all five trials converged. Of these nine cases, only 1a0m (conotoxin) required more iterations. The acquisition of the conotoxin data with a Cu rotating anode at room temperature (outermost shell is 1.10–1.14 Å) surely contributes to the different behavior of this data set. In contrast to the nine preceding cases, application of S_M -*ipp* to 1f94 (bucandin) was less successful. Consequently, N_{trials} was increased to 25 to estimate more reliably the success percentage (32%). This structure has large atomic disorder ($B_{Wilson} = 14.3 \text{ \AA}^2$) which is

reflected in the large fraction of unobserved data in the 1.06–1.02 Å interval, *i.e.* 0.50 with $I > 2\sigma(I)$. The influence of *ipp* on the phase refinement accuracy can be estimated with ΔCC_M , *i.e.* the difference between CC_M values for S_M -*ipp* and for S_M . As can be clearly seen in Tables 3 and 4, ΔCC_M is only slightly negative, generally between -0.02 and -0.03 , which suggests that truncation of the outer-peak regions during the application of the *ipp* procedure is not critical.

To estimate the influence of *ipp* on the convergence of the phase refinement, the same tests carried out with strategy B1 were repeated with B2 (Table 4). Comparison of both sets of N_{iter} values confirms the much faster convergence of B1.

4.3. Crystal structures with strong scatterers

From Table 5 it follows that for compounds with heavy atoms of the first transition series, application of the B1 strategy allows the routine determination (in a reduced number of iterations) of crystal structures with N up to $\approx 5000 \times c$ (c = number of centerings) provided that the data are of good quality and that at least the scattering power of one of the heaviest atoms is not weakened. The resulting $\langle N_{iter} \rangle$ values go from 10 to 60 except for data sets 41au, 1pwl, 1heu and 1c7k for which it is larger. In the case of 41au the increase can be related to two of the three symmetry-independent selenomethionine Se atoms showing partial occupancies, *i.e.* (0.52, 0.48) and (0.31 and 0.69) (Fanfrlik *et al.*, 2013). For 1pwl and 1heu, the larger $\langle N_{iter} \rangle$ values could be ascribed to the larger d_{min} values (Table 2). For comparison purposes, the results obtained with strategies B1 and B2 are summarized in Table 6. Its inspection confirms the clear superiority of B1 over B2, especially for the larger test crystal structures.

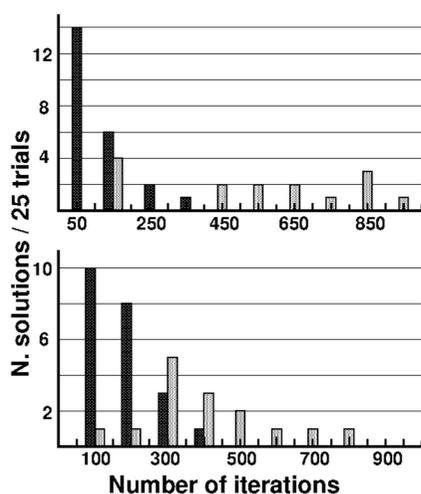


Figure 3

Effect of the *ipp* procedure on the phasing efficiency of the S_M algorithm with Φ_{rnd} . The two selected data sets belong to: (top) 3sbn (trichovirin) with 444 atoms in the unit cell; (bottom) 1a7z (Actino Z3) with 1228. True solutions obtained with/without the *ipp* procedure in black/gray (same starting random phase values for each pair of trials).

5. Discussion

One characteristic of the S_M algorithm is its mathematical simplicity, a consequence of the straightforward implementa-

Table 4

Application of the S_M -*ipp* and S_M algorithms to crystal structures with medium scatterers.

Upper and lower lines refer to phasing strategies B1 and B2, respectively (except for 3bcj). N , M , c as in Table 1; N_p = number of peaks showing up in the final E map above the $n\sigma_\rho$ threshold; CC_M = correlation coefficient between experimental and calculated modulus function; N_{iter} = number of iterations to achieve convergence (n.c. = no convergence in 1000 iterations); t_ρ , t_η = parameters controlling, respectively, the threshold of the m_ρ mask and the number N_η of strongest η peaks; $Q = N_\eta(2)/N$.

| PDB code | N/c (M/c) | N_p/c (n) | CC_M | N_{iter} for 5 or 25 trials | t_ρ | t_η | Q |
|----------|-----------------|-----------------|--------|--|----------|----------|-----|
| 1byz | 479 (1Cl) | 520 (1.1) | 0.86 | 46, 56, 102, 134, 146 | 2.65 | 4.0 | 0.9 |
| | | | 0.86 | 234, 356, 535; n.c. (2×) | | | |
| 2erl | 656 (14S) | 610 (1.1) | 0.78 | 32, 77, 112, 251, 270 | 2.50 | 4.0 | 0.8 |
| | | | 0.83 | 330, 475, 731; n.c. (2×) | | | |
| | | | 0.85 | 24, 29, 30, 31, 43 | | | |
| 1p9g | 702 (20S) | 695 (1.4) | 0.86 | 156, 174, 279, 296, 409 | 2.60 | 3.8 | 1.1 |
| | | | 0.86 | 24, 38, 46, 98, 127 | | | |
| 3nir | 902 (12S) | 934 (1.1) | 0.86 | 211, 333, 666, 690; n.c. (1×) | 2.70 | 4.0 | 0.7 |
| | | | 0.88 | 217, 344, 366, 510, 844 | | | |
| | | | 0.83 | 43, 47, 48, 49, 51 | | | |
| 1a0m | 1144 (40S) | 1124 (1.3) | 0.83 | n.c. (5×) | 2.65 | 4.0 | 1.0 |
| | | | 0.86 | 108, 110, 111, 171, 189, 353, 834, 897; n.c. (17×) | | | |
| 4lzt | 1183 (10S) | 1134 (1.5) | 0.81 | 342, 681; n.c. (23×) | 2.50 | 3.8 | 1.1 |
| | | | 0.83 | 108, 110, 111, 171, 189, 353, 834, 897; n.c. (17×) | | | |
| 1f94 | 1232 (20S) | 1160 (1.1) | 0.81 | 108, 110, 111, 171, 189, 353, 834, 897; n.c. (17×) | 2.50 | 3.8 | 1.1 |
| | | | 0.83 | 342, 681; n.c. (23×) | | | |
| | | | 0.85 | 108, 110, 111, 171, 189, 353, 834, 897; n.c. (17×) | | | |
| 1hhu | 1310 (16Cl) | 1360 (1.4) | 0.82 | $\langle N_{iter} \rangle = 117$ (17×); n.c. (8×) | 2.60 | 3.9 | 1.1 |
| | | | 0.85 | $\langle N_{iter} \rangle = 426$ (12×); n.c. (13×) | | | |
| | | | 0.72 | 18, 23, 23, 27, 36 | | | |
| 3odv | 1392 (32S) | 1480 (1.0) | 0.72 | 176, 249, 256, 290, 583 | 2.50 | 3.5 | 1.2 |
| | | | 0.77 | 176, 249, 256, 290, 583 | | | |
| 3psm | 1882 (16S) | 1854 (1.4) | 0.78 | 23, 24, 27, 33, 45 | 2.60 | 4.0 | 0.9 |
| | | | 0.82 | 309; n.c. (4×) | | | |
| | | | 0.81 | $\langle N_{iter} \rangle = 73$ (20×); n.c. (5×) | | | |
| 3bcj† | 7308 (26S) | 7222 (1.3) | 0.81 | $\langle N_{iter} \rangle = 73$ (20×); n.c. (5×) | 2.65 | 3.8 | 1.5 |
| | | | 0.82 | $\langle N_{iter} \rangle = 114$ (9×); n.c. (16×) | | | |

† Upper and lower lines correspond to B1 at $d_{min} = 0.78$ and 0.85 Å, respectively. $N_{iter(max)} = 200$.

tion of the modified tangent formula (5). One relevant parameter of S_M is t_ρ which modifies the threshold value in the calculation of $|\rho|$ through expression (4). The value of t_ρ mainly depends on the scattering power of the strongest scatterer present in the crystal structure. In Rius (2020), t_ρ was found to be close to 2.5. In the current work, the test examples extend to a larger variety of structures in which the strongest scatterer can be weak, medium or strong. Respective t_ρ values giving satisfactory results have been found to be ≈ 2.5 , ≈ 2.6 and ≈ 2.8 .

Regarding the *ipp* procedure, its application requires the approximate knowledge of N and the estimation of t_η . The N value used in the test calculations is the sum of both protein and solvent atoms (taken from the PDB), *i.e.* $N_{Prot} + N_{H_2O}$. An idea of $\langle N_{H_2O} \rangle$ can be obtained by averaging $(N_{Prot} + N_{H_2O})/N_{Prot}$ over all structures with more than 700 atoms listed in Tables 1 and 2 which gives 1.22 (5), *i.e.* $\langle N_{H_2O} \rangle \approx 0.22 \times N_{Prot}$. The second parameter, t_η , controls the number of η peaks above the $t_\eta \sigma_\eta$ threshold. It can be estimated from $Q = N_\eta(2)/N$. Suitable t_η values are those for which Q is close to 1 or not much smaller (the *ipp* procedure does not use N_η peaks exceeding N). According to Tables 3, 4 and 5, values of t_η from 3.5 to 4.0 give Q values ranging from 1.5 to 0.7. Whatever the initial phase values may be, a successful refinement ends with a sudden increase of CC_M concomitant with a marked N_η decrease.

Of interest is the comparison of the $N_\eta(1)$ values obtained with strategies A1 (Φ_{rnd}) and B1 (Φ_M) by using similar t_η values. As was already shown in Section 2, $N_\eta(1)$ is smaller than N for Φ_{rnd} (Fig. 2). However, for Φ_M (Fig. 4), $N_\eta(1)$ is

much larger than N , since here η essentially corresponds to the shifted modulus function with weakened origin peak. In the test calculations, the Φ_M set at the end of the first iteration is always calculated with the N largest η peaks. The only exception is 1b0y. Since the unit cell of this compound contains four dominant scattering units (Fe_4S_4 clusters), only the 240 ($= 16^2 - 16$) strongest η peaks (mostly corresponding to Fe–Fe interactions) were used.

For the compounds in Table 1 (except for 3bcj), the average strength of the S/Cl peaks in the Fourier map is 30 (5) a.u. (a.u. = arbitrary units). For 3bcj, however, the strength increases to 59 a.u. The explanation for the much larger peak strength has to be sought in the ultra-high resolution of the experimental data favored by its lower measurement temperature (15 K

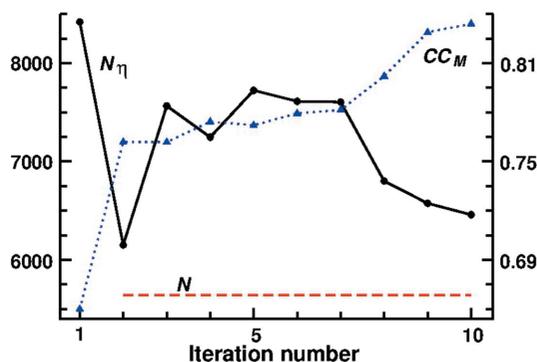


Figure 4
 S_M -*ipp* phasing with Φ_M : variation of N_η and CC_M with the iteration number for data set 3ks3 ($t_\eta = 3.9$). N = number of non-H atoms in the unit cell.

Table 5

Application of S_M -*ipp* to crystal structures containing strong scatterers (S) (strategy B1).

N = number of non-H atoms in the unit cell (PDB); c = number of centerings; N_p , CC_M , N_{iter} , n.c., t_p , t_η and Q as in Table 3. [t_p = 2.80 except for 1w3m (2.60), 3d1p (2.70), 1a6m (2.75), 41au (2.70) and 3fsa (2.70); $\langle B_{Wilson} \rangle$ is 6.8 (1.1) Å² with the extrema being 5.3 for 2fdn and 9.1 for 1eb6.]

| PDB code | N/c (S/c) | N_p/c (n) | CC_M | N_{iter} for 5 trials | t_η | Q |
|----------|-------------|-------------|--------|-------------------------|----------|-----|
| 2bf9 | 768 (2Zn) | 709 (1.1) | 0.81 | 10, 11, 12, 13, 15 | 3.5 | 2.2 |
| 8rxn | 1010 (2Fe) | 905 (1.1) | 0.83 | 14, 15, 17, 18, 22 | 3.5 | 1.4 |
| 1w3m | 1276 (24Ca) | 1275 (1.4) | 0.81 | 30, 33, 37, 42, 80 | 4.0 | 1.1 |
| 2ov0 | 2060 (2Cu) | 1990 (1.5) | 0.84 | 14, 15, 16, 16, 29 | 4.0 | 0.9 |
| 1c75 | 2660 (4Fe) | 2541 (1.4) | 0.82 | 12, 13, 14, 16,16 | 4.0 | 1.1 |
| 3d1p | 2702 (4Se) | 2642 (1.1) | 0.83 | 12, 12, 14, 15, 16 | 3.8 | 1.1 |
| 1pwl | 3030 (1Br) | 3123 (1.1) | 0.83 | 32, 54, 57, 62, 149 | 3.8 | 1.2 |
| 1a6m | 3154 (2Fe) | 3203 (1.1) | 0.83 | 28, 30, 31, 37, 48 | 4.0 | 0.8 |
| 41au | 3278 (6Se†) | 3440 (1.1) | 0.85 | 41, 69, 79; n.c. (2×) | 3.8 | 1.0 |
| 1eb6 | 3300 (2Zn) | 3406 (1.1) | 0.82 | 16, 18, 23, 24, 40 | 4.0 | 0.9 |
| 1b0y | 3348 (16Fe) | 3360 (1.3) | 0.76 | 38, 39, 41, 53, 79 | 3.5 | 1.5 |
| 1x8q | 3662 (2Fe) | 3510 (1.5) | 0.83 | 34, 36, 58, 64, 92 | 4.0 | 1.0 |
| 2fdn | 3944 (64Fe) | 3832 (1.3) | 0.81 | 21, 21, 22, 23, 26 | 3.8 | 1.0 |
| 3fsa | 4488 (4Cu) | 4580 (1.5) | 0.83 | 31, 39, 40, 44, 56 | 4.0 | 1.0 |
| 1c7k | 4532 (4Zn) | 4548 (1.3) | 0.84 | 80, 96, 128, 202, 399 | 4.0 | 0.9 |
| 3ks3 | 5626 (2Zn) | 5588 (1.2) | 0.83 | 9, 10, 10, 10, 10 | 3.9 | 1.1 |
| 1heu | 7618 (4Cd) | 7603 (1.1) | 0.82 | 35, 40, 42, 45, 176 | 3.9 | 1.1 |

† Four Se atoms are partially disordered.

compared with the usual 100 K). This test structure was selected to check the phasing capability of S_M -*ipp* with ultra high resolution data. With 5934 atoms in the unit cell (solvent atoms excluded) this crystal structure is in the same order of magnitude as those listed in Table 2 containing strong scatterers. Application of S_M -*ipp* with Φ_M (strategy B1) yields success percentages of 80%, 36% and 0% for $d_{min} = 0.78, 0.85$ and 0.90 Å, respectively (Fig. 5 reproduces the E map of one arbitrary successful refinement). Notice that S_M -*ipp* solves

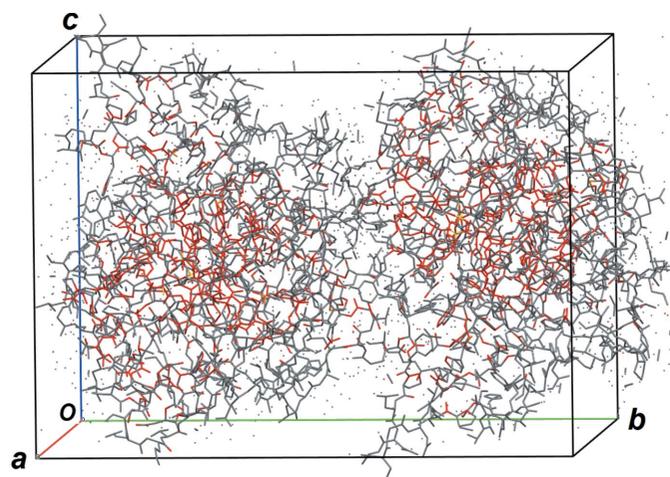


Figure 5

Unit-cell content of aldose reductase (Zhuo *et al.*, 2008; data set 3bcj) showing the two unique protein chains related by the screw axis along b as obtained with the S_M -*ipp* phasing algorithm directly from the experimental modulus synthesis (Φ_M) by assuming $P1$ symmetry (S and light atoms are found simultaneously). Atoms with higher refined peak strength are shown in red.

Table 6

Comparison of strategies B1 and B2 when applied to crystal structures with strong scatterers (S).

For B2, the individual N_{iter} values are given; for B1, $\langle N_{iter} \rangle$ corresponds to N_{iter} values in Table 5. It is evident that B1 (using *ipp*) performs better than B2 in all cases. N = number of non-H atoms in the unit cell (PDB); c = number of centerings; $\langle N_{iter} \rangle$ = average number of iterations to achieve convergence (n.c. = no convergence in 1000 iterations).

| PDB code | N/c (S/c) | B1 strategy $\langle N_{iter} \rangle$ | B2 strategy N_{iter} for 5 trials |
|----------|-------------|---|--|
| 2bf9 | 768 (2Zn) | 12.2 (5×) | 29, 29, 36, 39, 55 |
| 8rxn | 1010 (2Fe) | 17.2 (5×) | 44, 53, 54, 56, 61 |
| 1w3m | 1276 (24Ca) | 44.4 (5×) | 97, 109, 118, 124, 134 |
| 2ov0 | 2060 (2Cu) | 18.0 (5×) | 52, 54, 61, 62, 86 |
| 1c75 | 2660 (4Fe) | 14.2 (5×) | 41, 51, 52, 57, 73 |
| 3d1p | 2702 (4Se) | 13.8 (5×) | 35, 38, 44, 45, 47 |
| 1pwl | 3030 (1Br) | 70.8 (5×) | n.c. (5×) |
| 1a6m | 3154 (2Fe) | 34.8 (5×) | n.c. (5×) |
| 41au | 3278 (6Se†) | 63.0 (3×) | 400, n.c. (4×) |
| 1eb6 | 3300 (2Zn) | 24.2 (5×) | 69, 86, 109, 156, 289 |
| 1b0y | 3348 (16Fe) | 50.0 (5×) | 210, 225, 243, 254,259 |
| 1x8q | 3662 (2Fe) | 56.8 (5×) | 261, 317, 404, 961, n.c. |
| 2fdn | 3944 (64Fe) | 22.6 (5×) | 62, 77, 82, 93, 97 |
| 3fsa | 4488 (4Cu) | 42.0 (5×) | 163, 288, 324, 413, n.c. |
| 1c7k | 4532 (4Zn) | 181.0 (5×) | n.c. (5×) |
| 3ks3 | 5626 (2Zn) | 9.8 (5×) | 36, 42, 45, 46, 48 |
| 1heu | 7618 (4Cd) | 67.6 (5×) | 226, 259, 273, 534, n.c. |

† Four Se atoms are partially disordered.

here the protein structure in one stage, *i.e.* it is not necessary to first locate single S atoms as, *e.g.*, done by McCoy *et al.* (2017).

A limitation of S_M -*ipp* (when used as an *ab initio* phasing algorithm) arises for crystal structures belonging to high-symmetry point groups and having large asymmetric units, since then N becomes exceedingly large. Normally, the usual way to cope with such situations is to derive the initial Φ from a larger structure model by using, among others, molecular replacement or anomalous dispersion techniques. In such cases S_M -*ipp* will become the phase refinement stage of a more general two-stage strategy.

6. Conclusions

It has been shown that the introduction of the new peakness-enhancing *ipp* procedure in the S_M phase refinement algorithm significantly improves the algorithm efficiency for diffraction data at atomic resolution and, consequently, has been incorporated as the default option. For *ab initio* structure determinations with S_M -*ipp*, the proper choice of the type of starting phases is important. Regarding this point, the following rules could be established on the basis of the test calculations:

(a) For very small light-atom crystal structures either Φ_{rnd} or Φ_M phases can be used (peak overlap in the modulus function can still be managed by S_M -*ipp*).

(b) Starting with Φ_{rnd} is appropriate for crystal structures containing only weak scatterers (the largest N value tested is around 1500 atoms).

(c) Starting with Φ_M is the best option for crystal structures with medium scatterers like S or Cl (largest N for routine determinations is $1500 \times c$). If no trial converges in $N_{iter(max)}$

iterations, then phase refinement with Φ_{rnd} should be tried (with a larger $N_{\text{iter(max)}}$); however, $\Phi_{M'}$ should always be the first choice.

(d) Use of $\Phi_{M'}$ is the best choice for crystal structures with strong scatterers. For metals belonging to the first transition series like Fe, Cu and Zn, the largest N value for routine determinations has been estimated to be about $5000 \times c$ atoms (tests performed on data sets collected at ≈ 100 K). One characteristic of successful phase refinements starting with $\Phi_{M'}$ is their fast convergence. This allows one to reduce $N_{\text{iter(max)}}$ and, consequently, increase the number of explored trials.

Finally, some words regarding data completeness are in order. As already mentioned in Section 1, the S_M algorithm relies on the validity of the R_M residual (2) which assumes that δ and ρ are proportional (which is satisfied for data sets reaching atomic resolution as is the case with the test calculations described in this work). If the intensities of the outer reflection shells are unobserved (a common situation for protein crystals), R_M is no longer strictly fulfilled. Extrapolating the structure factors of unobserved reflections beyond the experimental resolution limit, e.g. by Fourier inversion of a suitably modified map, could be a solution for extending the applicability range of R_M to moderate-resolution data sets. This 'structure-factor extrapolation' technique (Caliandro *et al.*, 2005a,b, 2007; see also Jia-xing *et al.*, 2005) is particularly effective for crystal structures containing heavy atoms (Caliandro *et al.*, 2008; Burla *et al.*, 2012). The combination of S_M with the extrapolation technique could represent a further source of progress.

Funding information

The following funding is acknowledged: MINECO and FEDER (grant No. RTI2018-098537-B-C21); Severo Ochoa Programme for Centres of Excellence in R&D (grant No. SEV-2015-0496).

References

Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* **D52**, 469–480.

Avvaru, B. S., Kim, C. U., Sippel, K. H., Gruner, S. M., Agbandje-McKenna, M., Silverman, D. N. & McKenna, R. (2010). *Biochemistry*, **49**, 249–251.

Benini, S., González, A., Rypniewski, W. R., Wilson, K. S., Van Beeumen, J. J. & Ciurli, S. (2000). *Biochemistry*, **39**, 13115–13126.

Buerger, M. J. (1950). *Acta Cryst.* **3**, 87–97.

Bunkóczi, G., Schiell, M., Vértesy, L. & Sheldrick, G. M. (2003). *J. Pept. Sci.* **9**, 745–752.

Bunkóczi, G., Vértesy, L. & Sheldrick, G. M. (2005). *Acta Cryst.* **D61**, 1160–1164.

Burkhardt, B. M., Gassman, R. M., Langa, D. A., Pangborn, W. A. & Duax, W. L. (1998). *Biophys. J.* **75**, 2135–2146.

Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mallamo, M., Mazzone, A., Polidori, G. & Spagna, R. (2012). *J. Appl. Cryst.* **45**, 357–361.

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2000). *Acta Cryst.* **A56**, 451–457.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Mazzone, A. & Siliqi, D. (2008). *J. Appl. Cryst.* **41**, 548–553.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005a). *Acta Cryst.* **D61**, 556–565.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005b). *Acta Cryst.* **D61**, 1080–1087.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2007). *J. Appl. Cryst.* **40**, 931–937.

Cooley, J. W. & Tukey, J. W. (1965). *Math. C.* **19**, 297–301.

Dauter, Z., Sieker, L. C. & Wilson, K. S. (1992). *Acta Cryst.* **B48**, 42–59.

Dauter, Z., Wilson, K. S., Sieker, L. C., Meyer, J. & Moulis, J. M. (1997). *Biochemistry*, **36**, 16065–16073.

Dodson, E. J. & Woolfson, M. M. (2009). *Acta Cryst.* **D65**, 881–891.

El-Kabbani, O., Darmanin, C., Schneider, T. R., Hazemann, I., Ruiz, F., Oka, M., Joachimiak, A., Schulze-Briese, C., Tomizaki, T., Mitschler, A. & Podjarny, A. (2004). *Proteins*, **55**, 805–813.

Fanfrlik, J., Kolar, M., Kamlar, M., Hurny, D., Ruiz, F. X., Cousido-Siah, A., Mitschler, A., Rezac, J., Munusamy, E., Lepsik, M., Matejíček, P., Veselý, J., Podjarny, A. & Hobza, P. (2013). *ACS Chem. Biol.* **8**, 2484–2492.

Gessmann, R., Axford, D., Owen, R. L., Brückner, H. & Petratos, K. (2012). *Acta Cryst.* **D68**, 109–116.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I. & Blundell, T. L. (1983). *Biopolymers*, **22**, 293–304.

Hu, S.-H., Loughnan, M., Miller, R., Weeks, C. M., Blessing, R. H., Alewood, P. F., Lewis, R. J. & Martin, J. L. (1998). *Biochemistry*, **37**, 11425–11433.

Jia-xing, Y., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2005). *Acta Cryst.* **D61**, 1465–1475.

Kondrashov, D. A., Roberts, S. A., Weichsel, A. & Montfort, W. R. (2004). *Biochemistry*, **43**, 13637–13647.

Kuhn, P., Deacon, A. M., Comoso, S., Rajaseger, G., Kini, R. M., Usón, I. & Kolatkar, P. R. (2000). *Acta Cryst.* **D56**, 1401–1407.

Kurisu, G., Kai, Y. & Harada, S. (2000). *J. Inorg. Biochem.* **82**, 225–228.

Lehmann, C., Bunkóczi, G., Vértesy, L. & Sheldrick, G. M. (2002). *J. Mol. Biol.* **318**, 723–732.

McAuley, K. E., Jia-Xing, Y., Dodson, E. J., Lehmebeck, J., Østergaard, P. & Wilson, K. S. (2001). *Acta Cryst.* **D57**, 1571–1578.

McCoy, A. J., Oeffner, R. D., Wrobel, A. G., Ojala, J. R. M., Tryggvason, K., Lohkamp, B. & Read, R. J. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 3637–3641.

Meijers, R., Morris, R. J., Adolph, H. W., Merli, A., Lamzin, V. S. & Cedergren-Zeppezauer, E. S. (2001). *J. Biol. Chem.* **276**, 9316–9321.

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

Parisini, E., Capozzi, F., Lubini, P., Lamzin, V., Luchinat, C. & Sheldrick, G. M. (1999). *Acta Cryst.* **D55**, 1773–1784.

Pentelute, B. L., Mandal, K., Gates, Z. P., Sawaya, M. R., Yeates, T. O. & Kent, S. B. (2010). *Chem. Commun.* **46**, 8174–8176.

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

Ramachandran, G. N. & Raman, S. (1959). *Acta Cryst.* **12**, 957–964.

Rius, J. (1993). *Acta Cryst.* **A49**, 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 (downloadable from <https://departments.icmab.es/crystallography/software>).

Rius, J. (2012). *Acta Cryst.* **A68**, 77–81.

Rius, J. (2020). *Acta Cryst.* **A76**, 489–493.

Rius, J., Crespi, A. & Torrelles, X. (2007). *Acta Cryst.* **A63**, 131–134.

Rollet, J. S. (1965). Editor. *Computing Methods in Crystallography*, pp. 35–37. Oxford: Pergamon Press.

Sato, K., Li, C., Salard, I., Thompson, A. J., Banfield, M. J. & Dennison, C. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 5616–5621.

- Schäfer, M., Sheldrick, G. M., Bahner, I. & Lackner, H. (1998). *Angew. Chem. Int. Ed.* **37**, 2381–2384.
- Schmidt, A., Teeter, M., Weckert, E. & Lamzin, V. S. (2011). *Acta Cryst.* **F67**, 424–428.
- Sheldrick, G. M. (2015). *Acta Cryst.* **A71**, 3–8.
- Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* **B51**, 423–431.
- Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* **A48**, 451–456.
- Song, X., Zhang, M., Zhou, Z. & Gong, W. (2011). *FEBS Lett.* **585**, 300–306.
- Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
- Vojtěchovský, J., Chu, K., Berendzen, J., Sweet, R. M. & Schlichting, I. (1999). *Biophys. J.* **77**, 2153–2174.
- Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* **D54**, 522–546.
- Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Acta Cryst.* **D49**, 179–181.
- Wrinch, D. M. (1939). *London, Edinb. Dubl. Philos. Mag. J. Sci.* **27**, 98–122.
- Xiang, Y., Huang, R. H., Liu, X. Z., Zhang, Y. & Wang, D. C. (2004). *J. Struct. Biol.* **148**, 86–97.
- Zhao, H. T., Hazemann, I., Mitschler, A., Carbone, V., Joachimiak, A., Ginell, S., Podjarny, A. & El-Kabbani, O. (2008). *J. Med. Chem.* **51**, 1478–1481.