Extending the novel |ρ|-based phasing algorithm to the solution of anomalous scattering substructures from SAD data of protein crystals

Jordi Rius* and Xavier Torrelles

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

Owing to the importance of the single-wavelength anomalous diffraction (SAD) technique, the recently developed |ρ|-based phasing algorithm ($S_{M,|\rho|}$) incorporating the inner-pixel preservation (*ipp*) procedure [Rius & Torrelles (2021). *Acta Cryst* A**77**, 339–347] has been adapted to the determination of anomalous scattering substructures and its applicability tested on a series of 12 representative experimental data sets, mostly retrieved from the Protein Data Bank. To give an idea of the suitability of the data sets, the main indicators measuring their quality are also given. The dominant anomalous scatterers are either SeMet or S atoms, or metals/clusters incorporated by soaking. The resulting SAD-adapted algorithm solves the substructures of the test protein crystals quite efficiently.

1. Introduction

Important present applications of the single-wavelength anomalous diffraction (SAD) technique are the location of SeMet atoms in crystals of multi-site genetically engineered proteins, the determination of the positions and occupancies of the heavy atoms (or clusters) entering the crystal, *e.g.* when soaking it in a solution, or also the direct use of chemical species already present in native crystals as anomalous scatterers (S, Cl, P, ...). Knowledge of the anomalous scattering (AS) substructure provides starting phase values which can be iteratively improved by density modification. Although the substructure can be solved in favourable cases by the direct interpretation of the anomalous Patterson function (Rossmann, 1961), direct methods (DM) often offer the only alternative in complex cases. The application of DM to SAD data takes advantage of the availability of the experimentally accessible absolute values of the anomalous differences ($|D|_{\exp}$) between pairs of acentric reflections (Bijvoet pairs) which follows from the atomic scattering factor definition

$$f_j = f_j^n + f_j' + if_j'', \quad (1)$$

where f_j^n is the normal scattering factor of atom j, and f_j' and f_j'' are the corresponding real and imaginary anomalous dispersion corrections (respective symbols for non-vibrating atoms are f_0, f_0^n, f_0', f_0''). Let us consider a structure composed of N atoms with N_A of them scattering anomalously and with **r** being the atomic position vector. The structure factor of an arbitrary H reflection is then

$$F_H = |F_H'| \exp(i\varphi_H') + i|F_H''| \exp(i\varphi_H'') \quad (2)$$

with
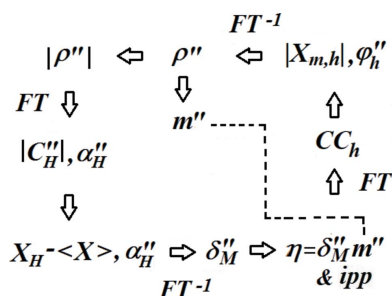
$$\left|F'_H\right|\exp(i\varphi'_H) = \sum_{l=1}^{N} f_{l,H}^{n}\exp(i2\pi\mathbf{H}\mathbf{r}_l) + \sum_{j=1}^{N_A} f'_j\exp(i2\pi\mathbf{H}\mathbf{r}_j) \quad (3)$$

$$\left|F''_H\right|\exp(i\varphi''_H) = \sum_{j=1}^{N_A} f''_j\exp(i2\pi\mathbf{H}\mathbf{r}_j). \quad (4)$$

For two $+H$ and $-H$ reflections constituting a Bijvoet pair (from now on, $F_{+H} = F^+$ and $F_{-H} = F^-$), the absolute value of the anomalous difference D is given by

$$|D| = \left||F^+| - |F^-|\right| \quad (5)$$

which is related to $|F''|$ by the simple relationship (30) (see Appendix A)

$$|D| = 2|F''| \times \left|\sin(\varphi' - \varphi'')\right| \quad (6)$$

if conditions (7*a*) and (7*b*) corresponding to (28) and (29) are met, *i.e.*

$$|F|_{av}^2 \gg |D|^2/4 \quad (7a)$$

and

$$|F|_{av}^2 \gg \left|F''\right|^2 \quad (7b)$$

with

$$|F|_{av}^2 = \left(|F^+|^2 + |F^-|^2\right)/2. \quad (8)$$

Equation (6) constitutes the basis for solving AS substructures by DM. First attempts showing the viability of locating AS in metalloproteins by DM were performed by Mukherjee *et al.* (1989) with the program *MULTAN87* (Debaerdemaeker *et al.*, 1987) following the path previously paved by Wilson (1978) in connection with the isomorphous replacement case and taking advantage of preliminary results on the location of AS using tuneable synchrotron radiation (Einspahr *et al.*, 1985); however, it was the introduction of the dual-space DM that represented a substantial improvement in the determination of AS substructures. This DM strategy refines phases by iteratively alternating structure invariant manipulation (reciprocal space) with Fourier peak optimization (real space). It was first implemented in the *Shake-and-Bake* program (Miller *et al.*, 1994). This philosophy was also incorporated in *SHELX* (Sheldrick & Gould, 1995) which evolved to *SHELXD* by incorporating, among other things, Patterson seeding (Schneider & Sheldrick, 2002). Descriptions of the application of *SHELXD* to the solution of the AS substructures are given by Usón & Sheldrick (2018) and Sheldrick (2010). More recently, the capability of SAD phasing in the presence of only weak AS has increased due to the possibility of extending the SAD experiments to longer wavelengths as well as to the availability of faster and more accurate X-ray detectors (*e.g.* Leonarski *et al.*, 2018), allowing application of lower dose rates and thus increasing data redundancy on a unique crystal (data set scaling from multiple crystals is minimized). A recent promising alternative acquisition mode, especially useful for data collection from small, weakly diffracting and radiation-sensitive crystals, is serial crystallography. This technique is based on taking one single image (containing partial Bragg reflection information) from each microcrystal and completing the diffraction data set by combining the individual indexed images from thousands of crystals. A selection of *de novo* (SAD) phasing serial crystallography studies at synchrotron sources can be found in Nass *et al.* (2020).

Recently, $|\rho|$-based DM in the form of the $S_{M,|\rho|}$ phasing algorithm (Rius, 2020) have been extended to large crystal structures through the introduction of the peakness-enhancing *ipp* (inner-pixel preservation) procedure (Rius & Torrelles, 2021) (hereafter, to simplify its designation, the $S_{M,|\rho|}$ algorithm is specified with the acronym SMAR in which S stands for 'sum function', M for 'modulus function' and AR for 'absolute ρ'). The aim of the present contribution is the adaptation of the *ipp*-improved SMAR to the solution of AS substructures from SAD data (SAD-SMAR). Its feasibility is shown with SAD data sets either kindly supplied by the respective authors or retrieved from the Protein Data Bank (PDB). All calculations have been carried out with a modified version of *XLENS_v1* (Rius, 2011). To help the reader to assess the suitability of the test data, two indicators are given for each data set (extending to all acentric reflections in the corresponding resolution range used in the SAD-SMAR application), namely:

(i) The size of the anomalous signal (Bijvoet ratio), $\langle|D|\rangle/\langle|F|\rangle$ (Hendrickson & Teeter, 1981; Wang, 1985) ranging from 0.012 to 0.070 in the selected test examples.

(ii) The precision of $|D|$ given by the $\langle|D|/\sigma(|D|)\rangle$ ratio (Schneider & Sheldrick, 2002; Wang, 1985) which should be >1.5 (ideally also for the outermost resolution shell) (Cianci *et al.*, 2008; Giacovazzo, 2014). Logically, the precision of $|D|$ directly depends on the precision of the corresponding $|F^+|$ and $|F^-|$ (more strictly of I^+ and I^-).

In SAD phasing, redundancy of diffraction data is an important data collection parameter, since it affects the variance of the average intensity estimates. As this work is based on published data sets, the cited redundancy values are those given by the respective authors.

2. The composition of the $|D|$ set

Solving AS substructures by DM requires a previous selection of the experimental $|D|$ values, $|D|_{exp}$, since not all of them are appropriate. A preliminary check should ensure that the Bijvoet-pair reflections have $|F|_{av}$ values satisfying conditions (7*a*) and (7*b*). This is accomplished by preserving in the initial set of $|D|$ differences only those reflections with $|F|_{av}$ values (expressed as $|E|$'s) larger than a given ECUT cut-off value. In the test calculations, the used ECUT is $\cong 0.25$ which causes the suppression of approximately 5% of the total of acentric reflections. The selection process continues with two additional rejection criteria which are directly applied to the $|D|$ anomalous differences (to increase their reliability and the absence of outliers). Since $|D|$ is in general much smaller than $|F|_{av}$, random errors inherent to $|F^+|$ and $|F^-|$ seriously affect the precision of $|D|$. Consequently, only those reflections fulfilling the $|D| > DFCUT \times \sigma(|D|)$ criterion are preserved in the $|D|$ set (Hendrickson *et al.*, 1988; Grosse-Kunstleve &

Brunger, 1999). In the test calculations, DFCUT is in general ~0.4 which represents the additional removal of 10–15% of acentric reflections from the $|D|$ set. The selection process ends with the outlier elimination, *i.e.* all reflections with $|D|/$r.m.s.d.$(|D|)$ greater than ~4.0 are filtered out (Hendrickson *et al.*, 1988; Grosse-Kunstleve & Brunger, 1999) [r.m.s.d.$(|D|)$ = root-mean-square deviation of $|D|$]. The surviving reflections in the $|D|$ set are generically denoted by H.

3. The SAD-SMAR algorithm

3.1. The normalized X values

The SAD-SMAR algorithm uses, instead of the experimentally inaccessible quasi-normalized $|E|$ values of the substructure (Main, 1976), the normalized X values based on (6) and defined by the quotient

$$X^2 = \frac{|F''|^2 \sin^2\psi}{\langle |F''|^2 \sin^2\psi \rangle_s} \tag{9}$$

with $\psi = \varphi' - \varphi''$ and where s is the resolution shell corresponding to $|F''|^2$. Since $|F''|^2$ and $\sin^2\psi$ may be assumed uncorrelated, the average term in the denominator can be decomposed into the product of $\langle |F''|^2 \rangle_s$ and $\langle \sin^2\psi \rangle_s$. Furthermore, since φ' predominantly depends on the protein atoms and φ'' only on the anomalous scatterers, both phases can be considered largely uncorrelated and hence $\langle \sin^2\psi \rangle_s$ can be assumed to be 0.5, so that

$$X^2 = \frac{|F''|^2 \sin^2\psi}{0.5\langle |F''|^2 \rangle_s}. \tag{10}$$

On the other hand, according to the $|E|$ definition, $\langle |F''|^2 \rangle_s$ in (10) can be replaced by $|F''|^2/|E''|^2$, so that the expression relating X^2 and $|E''|^2$ reduces to

$$X^2 = |E''|^2 \times \frac{\sin^2\psi}{0.5}. \tag{11}$$

If X^2 is averaged over all reflections in its corresponding s resolution shell, then $\langle X^2 \rangle_s = 1$, since $\langle |E''|^2 \rangle_s$ is 1 by definition and $\langle \sin^2\psi \rangle_s$ is 0.5.
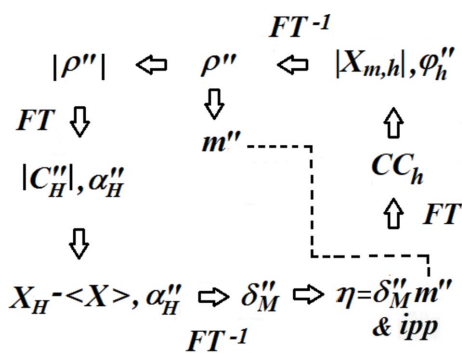

Figure 1
The recursive SAD-SMAR phase refinement algorithm with enhanced peakness (*ipp*). Compared with the unmodified SMAR, the principal differences are the composition of Φ_h as well as the replacement of $|E|$ values either by $X = |E''\sin\psi|$ or by $|X_m|$.

In addition to X values, SAD-SMAR also uses modified X values called $|X_m|$. These are obtained (i) by calculating the M modulus function with X as Fourier coefficients (extending the sum to the H reflections), (ii) by suppressing the negative regions in M, and (iii) by back Fourier transforming the modified M function (Karle, 1980).

3.2. Calculation of X from $|D|_{exp}$

The relation between X and $|D|$ is easily found by introducing the squared (6) into (10)

$$X^2 = \frac{|D|^2}{2\langle |F''|^2 \rangle_s} = \frac{k^2(|D|_{exp})^2}{2\langle |F''|^2 \rangle_s}, \tag{12}$$

where k is the scaling constant putting $|D|_{exp}$ on the same scale as $|D|$. The $\langle |F''|^2 \rangle_s$ quantity in the denominator, *i.e.* the average intensity of the s shell, can be expressed as

$$\langle |F''|^2 \rangle_s = \exp\left[-2B\left(\frac{\sin\theta_s}{\lambda}\right)^2\right] \sum_{j=1}^{N_A} (f''_{0j})^2, \tag{13}$$

where B is the overall atomic displacement parameter including vibrational and disorder effects. At this point, for convenience, each f''_{0j} will be converted to q_j by dividing by f''_{0L} (= the largest f''_{0j}). Replacement of f''_{0j} by $q_j f''_{0L}$ in (13) and subsequent introduction of the modified (13) into (12) leads to the final expression

$$X^2 = K^2 \exp\left[2B\left(\frac{\sin\theta_s}{\lambda}\right)^2\right] \frac{(|D|_{exp})^2}{\sum_{j=1}^{N_A} q_j^2} \tag{14}$$

with

$$K = k/2^{1/2} f''_{0L} \tag{15}$$

which allows the derivation of X^2 from $(|D|_{exp})^2$ provided that the AS composition is known. In view of (14), the estimation of the K constant and the B parameter can be obtained from a Wilson plot, since for each reciprocal-space shell, both $\langle X^2 \rangle_s$ and the $\langle (|D|_{exp})^2 \rangle_s / \sum_{j=1}^{N_A} q_j^2$ quotient are known.

3.3. SAD-SMAR recycling

Phasing with the SMAR algorithm was first shown by Rius (2020). Later on, the *ipp* procedure, a simple way of enhancing peakness in Fourier maps, was added (Rius & Torrelles, 2021). To show how the SAD-SMAR modification works, one phase refinement cycle is described in detail in Fig. 1. It has been divided into four stages, each one including one Fourier transform operation. These are:

(i) Calculation of the ρ'' density function. The phase refinement cycle begins with the introduction of Φ_h, the subset of φ'' phases of the h reflections to be refined (either initial or updated estimates). Unlike in non-anomalous SMAR applications where Φ_h contains the phases of all large reflections (*i.e.* those H reflections with $|E| \geq 1.00$), in the case of SAD-SMAR, Φ_h only includes the φ''_h phases of those H reflections with X larger than a given XCUT cut-off (here XCUT = 1.00). Since $X/2^{1/2}$ is equal to $|E''| |\sin\psi|$, the largest possible value of $|E''|$ for a given X is $X/2^{1/2}$ (which is reached for $|\sin\psi| = 1$).

Table 1
Effect of XCUT on the composition of the Φ_h subset of phases.

The central part of the table lists the $|E''| |\sin \psi|$ products for selected $|E''|$ and $|\sin \psi|$ values (numbers in bold refer to XCUT = 1.00). As shown in the rightmost column, Φ_h contains no phases of reflections with $|E''| < 0.70$; however, for $|E''| > 1.00$, the percentage of reflections considered in Φ_h is very high, e.g. 85.56% for $|E''| = 2$.

$|E''|$	$|\sin \psi|$					% in Φ_h
	1.00	0.75	0.50	0.25	0.10	
3.00	**3.00**	**2.25**	**1.50**	**0.75**	0.30	94.28
2.00	**2.00**	**1.50**	**1.00**	0.50	0.20	85.56
1.00	**1.00**	**0.75**	0.50	0.25	0.10	55.56
0.71	**0.71**	0.53	0.36	0.18	0.07	6.40
0.50	0.50	0.38	0.25	0.13	0.05	0.00
0.10	0.10	0.08	0.05	0.03	0.01	0.00

In general, $|\sin \psi|$ will be lower than 1 and therefore $X/2^{1/2}$ is a lower estimate of $|E''|$ (Grosse-Kunstleve & Adams, 2003). How the composition of Φ_h depends on the X values is illustrated in Table 1 for XCUT = 1.00. It can be seen that most phases of reflections with $|E''|$'s > 1.00 are present in Φ_h; however, this number decreases significantly for $|E''|$'s between 1.00 and 0.70 and, finally, for $|E''|$'s < 0.70, it becomes zero. In this work the initial estimates of φ''_h are the phase values corresponding to the Fourier coefficients of M', i.e. the randomly shifted modulus function (Rius & Torrelles, 2021). As can be seen in Fig. 1, the Fourier synthesis with $|X_{m,h}| \exp(i\varphi''_h)$ as Fourier coefficients gives the ρ'' density function from which the m'' mask is derived (and stored). According to Rius (2020), m'' is 1 (for $\rho'' > 0$), 0 (for ρ'' between 0 and $-t\sigma$) and -1 (for $\rho'' < -t\sigma$) with σ^2 being the variance of ρ'' (Φ_h) and $t \sim 2.65$.

(ii) Calculation of the Fourier transform of $|\rho''|$. It gives the $|C''_H| \exp(i\alpha''_H)$ Fourier coefficients and provides the updated α''_H.

(iii) Calculation of δ''_M. The δ''_M density function is the inverse Fourier transform of the $[(X_H - \langle X \rangle) \exp(i\alpha''_H)]$ coefficients formed by the experimental $X_H - \langle X \rangle$ values and the updated α''_H phases. The calculated δ''_M is then multiplied with the previously stored m'' mask to give the η product function.

(iv) Calculation of the Fourier transform of η. Peakness in η is enhanced by applying the *ipp* density modification procedure. Once completed, the modified η is Fourier-transformed to provide the new φ''_h and $|E''_h|$ values, the latter being used in the calculation of the CC_h figure-of-merit to follow the phase refinement convergence,

$$CC_h = \left[\frac{\sum_h (|X_{m,h}| \times |E''_h|_{\mathrm{new}})^2}{\sum_h |X_{m,h}|^2 \times \sum_h |E''_h|^2_{\mathrm{new}}} \right]^{1/2}. \quad (16)$$

If convergence is not achieved, the next cycle begins until the preset maximum number of cycles is reached.

4. Fourier refinement and figure-of-merit

After applying SAD-SMAR, the phases are further refined by Fourier recycling (five to ten cycles). In order not to have to modify the Fourier refinement module of already existing DM programs, e.g. of *XLENS_v1* (Rius, 2011), the F''_n structure factor corresponding to a hypothetical structure with scatterers of $f_{0L}q_j$ strengths is introduced (with f_{0L} being the normal scattering factor corresponding to the largest f''_{0L}). For this purpose, (11) and (14) are equated and both sides of the expression multiplied by f^2_{0L}. After rearranging the resulting expression, we obtain

$$|E''_n|^2 \exp\left[-2B \left(\frac{\sin \theta_s}{\lambda} \right)^2 \right] \left(\sum_{j=1}^{N_A} f^2_{0L} q^2_j \right) \frac{\sin^2 \psi}{0.5}$$
$$= K^2 f^2_{0L} \left(|D|_{\mathrm{exp}} \right)^2. \quad (17)$$

Table 2
Relevant data collection parameters and indicators.

[1]Detailed author references in Section 5; [2]main anomalous scatterers; [3]redundancy of diffraction data taken from the published/deposited data (and later normalized to the point group order); [4]highest resolution (in Å) of SAD data used in the structure refinement with R_{free} values[5] from the respective authors; [6]highest resolution for SAD-SMAR application; Bijvoet ratio[7] estimation; and $\langle |D|/\sigma(|D|) \rangle$[8] calculations (in the whole range and in the outermost reciprocal-space shell).

PDB code[1]	AS[2]	Space group	λ (Å)	Redundancy[3]	RES^4_{ref}	R_{free}[5]	RES^6_{SMAR}	$\langle |D| \rangle^7 / \langle |F| \rangle$	$\langle |D|/\sigma(|D|) \rangle$[8]	
									Whole	Outer
4jiu[(a)]	Zn	$P2_12_12_1$	1.282	1.63	1.60	–	2.50	0.0452	1.41	1.36
5cx8[(b)]	Se	$P2_12_12$	0.979	1.68	2.40	0.208	3.00	0.0568	1.59	0.80
4yu5[(c)]	Se	$P2_12_12_1$	0.979	1.10	2.90	0.207	3.30	0.0693	1.57	0.90
5lac[(d)]	Se	$P2_12_12$	0.918	1.15	1.94	0.207	2.50	0.0696	2.51	1.60
5iqy[(e)]	I	$C222_1$	1.542	6.83	2.40	0.234	3.00	0.0624	3.13	1.69
3k9g[(f)]	I	$P4_32_12$	1.542	1.56	2.25	0.266	2.90	0.0433	2.68	1.41
3km3[(f)]	I	$R3(H)$	1.542	1.87	2.10	0.222	2.80	0.0361	1.60	1.01
3men[(f)]	I	$P2_12_12_1$	1.542	1.70	2.20	0.237	3.00	0.0466	1.71	1.06
2g4h[(g)]	Cd	$F432$	2.000	3.04	2.00	0.218	2.90	0.0141	3.71	1.46
4tno[(h)]	S, Cl	$P4_12_12$	2.066	9.53	2.14	0.305	2.60	0.0132	2.37	1.12
4pgo[(h)]	S, Cl	$P6_522$	2.066	8.58	2.30	0.203	3.00	0.0175	3.03	1.21
2g4s[(g)]	S	$P6_322$	2.000	2.86	2.15	0.323	3.20	0.0116	1.93	1.47

(a) López-Pelegrín et al. (2013); (b) Goulas et al. (2016); (c) Arolas et al. (2016); (d) Kanitz et al. (2019); (e) Krishna Das et al. (2016); (f) Abendroth et al. (2011); (g) Mueller-Dieckmann et al. (2007); (h) Weinert et al. (2013).

Notice that the first three factors of the left-hand side of (17) correspond to $|F''_n|^2$. Replacement of these by $|F''_n|^2$ gives, after taking the square root, the best approximation Γ to the modulus of the structure factor

$$\Gamma = |F''_n| \frac{|\sin\psi|}{(0.5)^{1/2}} = Kf_{0L}|D|_{\exp} \qquad (18)$$

which is used as observational data in the $(2\Gamma - |F''_n|_{\text{calc}})\exp(i\varphi''_{\text{calc}})$ Fourier coefficients during recycling. At the end of the last Fourier refinement cycle, the (correlation coefficient based) residual is calculated

$$R_{\text{CC}} = 1000\left\{1 - \frac{\left[\sum_H \left(\Gamma_H |F''_{n,H}|_{\text{new}}\right)^{1/2}\right]^2}{\left(\sum_H \Gamma_H\right)\left(\sum_H |F''_{n,H}|_{\text{new}}\right)}\right\} \qquad (19)$$

wherein the sums only include the H reflections with $X \geq 0.7$.

5. Results of the test calculations

Relevant experimental information about the data sets used in the test calculations is given in Table 2. To improve the readability of the text, the test compounds are simply referenced with the appropriate PDB code. The verification of the SAD-SMAR tests was greatly facilitated by the availability of the refined model coordinates either kindly provided by the authors or deposited by them in the PDB. In this way, the r.m.s.d.'s between our substructure models and the deposited ones could be calculated. The most relevant results of the test calculations are summarized in Table 3. Table 4 complements this information by giving, for most test examples, the peak heights at the end of the Fourier recycling stage. Peak heights are always given in $\rho_{\text{peak}}/\sigma$ units, where ρ_{peak} is the density at the peak centre and σ^2 is the variance of ρ.

To get a rough idea of the quality of the deposited/supplied SAD refinements, the deposited R_{free} values (listed in Table 2 together with the corresponding upper resolution limits, RES_{ref}) were compared with the median R_{free} values of the PDB which are 0.24, 0.25, 0.26 and 0.28 for upper resolution limits corresponding to the intervals 1.95–2.15, 2.15–2.35, 2.35–2.40 and ~2.90 Å (Read *et al.*, 2011). It is found that the R_{free} values are less than or equal to the corresponding median R_{free} values in all cases, except for 2g4s and 4tno, for which R_{free} is significantly higher.

A preliminary test was the substructure solution of the proenzyme of proabylysin (PDB code 4jiu; $a = 34.679$, $b = 44.896$, $c = 72.233$ Å, $P2_12_12_1$). The data set was measured at ID29 (ESRF) at the Zn absorption edge ($\lambda = 1.282$ Å) (López-Pelegrin *et al.*, 2013). The structure refinement (deposited by the same authors) contains one Zn ion, one macromolecule and 148 water molecules in the asymmetric unit (a.u.), amounting to 1055 atoms. The successful run of this simple case (separation between found and deposited Zn ion positions is ~0.15 Å) confirmed the capability of SAD-SMAR to solve AS substructures at 2.5 Å resolution ($B \cong 25.1$ Å^2).

Table 3
Comparison of the SAD-SMAR phase refinement results for DFCUT = ~0.4 and 0.0.

[1]Completeness as $c_D = N_D/N_{\text{asy}}$ in % (N_D = number of reflections in $|D|$ set; N_{asy} = number of unique reflections); [2]n.c.t. = number of converging (correct) trials out of 25; [3](average) number of cycles to reach convergence; [4,5]final CC_h and R_{CC} values for correct solutions; [6]number of sites found in the a.u. compared with published refined values; [7]sep. = root-mean-square deviation in Å between found and published refined site positions.

PDB code	DFCUT	c_D[1]	n.c.t.[2]	ncycle[3]	CC_h[4]	R_{CC}[5]	nsites[6]	Sep.[7]
4jiu	0.375	71.2	25	5	0.91	39	1/1 Zn	0.15
	0.0	85.4	25	5	0.91	41		
5cx8	0.375	71.6	25	12	0.88	59–61	12/12 Se	0.24
	0.0	86.6	25	10	0.88	62–64		
4yu5	0.375	71.0	25	13	0.87	60–62	18/18 Se	0.35
	0.0	87.2	25	14	0.85	68–69		
5lac	0.375	87.6	25	27	0.91	50–51	12/12 Se	0.18
	0.0	90.1	25	14	0.91	49		
5iqy	0.450	76.2	8	<50	0.87	57–59	15/26 I	0.43
	0.0	86.0	13	<50	0.87	54–61		
3k9g	0.375	73.8	21	<55	0.87	59–65	9/12 I	0.35
	0.0	81.8	20	<55	0.87	60–64		
3km3	0.375	78.2	18	<36	0.87	65–69	13/16 I	0.43
	0.0	93.1	24	<45	0.87	68–71		
3men	0.375	74.1	4	<100	0.88	55–58	33/35 I	0.24
	0.0	88.0	11	<55	0.88	57–58		
2g4h	0.750	68.1	25	<50	0.87	57–61	5/5 Cd	0.22
		80.8	25	<50	0.87	59–63		
4tno	0.400	67.3	22	<30	0.88	51–54	4/3 S + 2 Cl	0.48
	0.0	79.2	20	<30	0.88	53–56		
4pgo	0.375	70.9	22	<40	0.88	54–57	4/2 S + 2 Cl	0.56
	0.0	79.9	25	<40	0.87	55–59		
2g4s	0.375	67.6	19	<125	0.88	57–60	3/4 S	0.18
	0.0	79.3	10	<125	0.88	59–62		

Next, it was tested with more challenging cases. To simplify the discussion, the test compounds are divided into three groups.

5.1. SeMet derivatives

Compared with other SAD situations, Se-SAD is particularly favourable due to the large AS strength of Se ($f''_{0Se} \sim 3.9$ and ~ 3.3 e$^-$ for $\lambda = 0.979$ and 0.919 Å, respectively) and because the substitution of S by Se in the methionine amino acids is normally complete. The data sets of the three tested SeMet derivatives correspond to:

5cx8: $a = 56.64$, $b = 184.74$, $c = 144.31$ Å, $P2_12_12$. A major immunodominant outer-membrane surface receptor antigen of *Porphyromonas gingivalis* measured at beamline (BL) XALOC (ALBA, Barcelona) (Goulas *et al.*, 2016; Se derivative refinement deposited in PDB entry 5cx8; SAD data supplied by one of them). There are 12 Se positions, two macromolecules and 509 water molecules in the a.u., amounting to 8119 atoms. Application of SAD-SMAR yields the positions of the 12 Se atoms ($B \cong 2.3$ Å^2) with r.m.s.d. = 0.24 Å compared with the deposited refined model (Table 3).

4yu5: $a = 97.61$, $b = 102.41$, $c = 242.88$ Å, $P2_12_12_1$. Thuringilysin, a variant of zymogenic BaInhA2-E/A measured at BL XALOC (ALBA, Barcelona) (Arolas *et al.*, 2016; Se derivative refinement deposited in PDB entry 4yu5; SAD data supplied by one of them). There are 18 Se, one Zn, two

Table 4
Heights of peaks in the final map of Fourier recycling for most test examples expressed in ρ_{peak}/σ units (ρ_{peak} = maximum peak density; σ^2 = variance of ρ).

The peaks in the a.u., ordered in decreasing height, are divided into two sets: A containing all correct signal peaks down to the first uninterpreted peak (only the heights of the first and last peaks are given, followed by the corresponding number of AS in brackets); B with mixed correct and uninterpreted peaks (with the heights of the latter in italics). According to these results, cut-off values of $\rho_{peak}/\sigma(\rho)$ for considering Fourier peaks as part of the substructure model can be set at around 5.0–7.0 (for soaked native crystals, they are slightly higher).

Code	A	B
5cx8	43.2 → 20.6 [12 Se]	5.6
4yu5	21.1 → 16.3 [18 Se +1 Zn]	15.0, *12.5, 7.0*
5lac	48.0 → 12.4 [12 Se]	5.5
5iqy	17.8 → 7.0 [14 I]	6.8, 6.6, 6.3, 5.3, *5.3, 5.1, 5.0*
3k9g	35.4 → 10.8 [8 I]	9.9, 9.2, *8.8*
3km3	35.6 → 14.4 [9 I]	*13.5*, 12.5, 10.9, 10.7, *7.4, 7.2, 7.1*
3men	36.3 → 8.7 [32 I]	8.1, 8.0, *7.6*
4tno	17.0 → 11.0 [2 S, 2 Cl]	5.3
4pgo	23.9 → 9.6 [2S, 1 Cl]	8.1, 6.4, *6.3*
2g4s	17.2 → 14.0 [3 S]	6.4, 5.9, *5.6*

macromolecules and 104 water molecules in the a.u., amounting to 10 942 atoms. Application of SAD-SMAR supplies the positions of the 18 Se atoms ($B \cong 9.8$ Å^2) with r.m.s.d. = 0.35 Å. Regarding the Zn ion, it shows up in the Fourier map 1.06 Å apart from the deposited refined position. Its strength is similar to that of the two Se atoms with higher B values.

5lac: $a = 94.144$, $b = 111.353$, $c = 58.191$ Å, $P2_12_12$. A 3C-like protease of Cavalli virus collected at BL 14.2 (BESSY II, Berlin) (Kanitz *et al.*, 2019; SAD and refinement data deposited in PDB entry 5lac). There are 12 Se positions (one of them split in the refinement), one macromolecule and 303 water molecules in the a.u., amounting to 4875 atoms. Application of SAD-SMAR yields the positions of the 12 Se atoms ($B \cong 4.9$ Å^2) with r.m.s.d. = 0.18 Å compared with the deposited model.

5.2. Native crystals soaked in heavy metal/metal cluster solutions

The first four cases of this subsection are native crystals soaked in a solution containing iodide ions and with their diffraction data being collected in-house on rotating anodes (Cu $K\alpha$ radiation) where the anomalous signal for I is large (f''_{0I} ~6.9 e$^-$). The fifth case corresponds to crystals soaked in a Cd^{2+}-containing solution.

5iqy: $a = 40.89$, $b = 132.08$, $c = 97.57$ Å, $C222_1$. An apo-dehydroascorbate reductase from *Pennisetum glaucum* (Krishna Das *et al.*, 2016; SAD and refinement data deposited in PDB entry 5iqy). According to the deposited data, there are 26 sites occupied by a total of 13.3 I^{1-}, one macromolecule and 95 water molecules in the a.u. (1719 atoms). Application of SAD-SMAR yields 15 sites ($B \cong 45.5$ Å^2) containing 9.74 I^{1-} which show a good agreement with the deposited data (r.m.s.d.

= 0.43 Å) as shown in Fig. 2. Table 5 compares the resulting site occupancies with the deposited ones.

3k9g: $a = 55.81$, $c = 200.90$ Å, $P4_312$. A plasmid partition protein (Abendroth *et al.*, 2011; SAD and refinement data deposited in PDB entry 3k9g). According to the structure refinement deposited in the PDB, there are 12 I^{1-} sites, one macromolecule and 91 water molecules in the a.u. (1858 atoms) with 6.6 I^{1-} in the 12 sites. Application of SAD-SMAR yields nine coincident I^{1-} sites ($B \cong 19.8$ Å^2) (r.m.s.d. = 0.35 Å) which justify a total of 5.3 I^{1-}, *i.e.* 81% of the refined I^{1-}content. By normalizing the sum of the heights of the nine strongest Fourier peaks to 5.3, the respective found and deposited site occupancies (using the original site labelling) are I1: 0.91, 0.99; I2: 0.91, 1.00*; I3: 0.60, 0.61; I4: 0.50, 0.53; I5: 0.59, 0.38; I6: 0.54, 0.38; I7: 0.60, 0.53; I10: 0.32, 0.36; I12: 0.33, 0.29 (* truncated to 1.00).

3km3: $a = 84.66$, $c = 140.74$ Å, $R3(H)$. A deoxycytidine triphosphate deaminase (Abendroth *et al.*, 2011; SAD and refinement data deposited in PDB entry 3km3). The refinement in the PDB includes 16 I^{1-} sites, two macromolecules and 516 water molecules in the a.u. (10 752 atoms) with 10.9 I^{1-} in the 16 sites. SAD-SMAR gives 13 coincident I^{1-} sites ($B \cong 6.2$ Å^2) (r.m.s.d. = 0.43 Å) which justify the 87% of the refined I^{1-}content. (Due to the large variability of the individual isotropic B values affecting the metal sites, no attempt to estimate the site occupancies from the corresponding peak heights was made.)

3men: $a = 45.70$, $b = 162.12$, $c = 173.07$ Å, $P2_12_12_1$. An acetylpolyamine aminohydrolase (Abendroth *et al.*, 2011; SAD and refinement data deposited in PDB entry 3men). According to the deposited data, there are 35 sites occupied by ~23.1 I^{1-}, four macromolecules and 516 water molecules in the a.u. (10 825 atoms). Application of SAD-SMAR yields 33 coincident I^{1-} sites ($B \cong 10.5$ Å^2) (r.m.s.d. = 0.24 Å) which justify ~92% of the refined I^{1-}content. The r.m.s.d. between the 33 found and corresponding deposited site occupancies is 0.172.

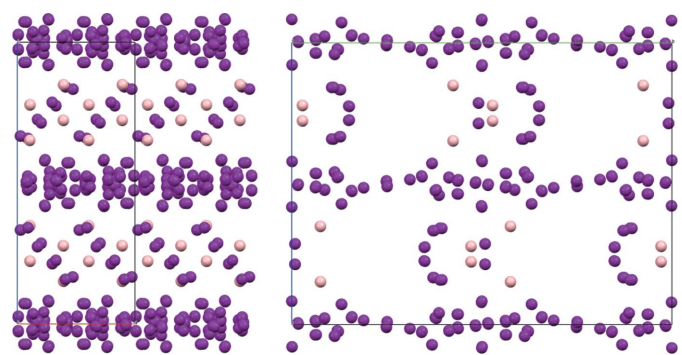**Figure 2**
5iqy: (010) and (100) projections of the I^{1-} site arrangement in the unit cell (only sites with occupancies ≥ 0.40): (violet) 120 (= 15 × 8) sites obtained by SAD-SMAR and Fourier recycling (which are also present in the deposited refinement; r.m.s.d. between found and deposited sites is 0.43 Å); (pink) 16 (= 2 × 8) additional sites only present in the deposited refinement (0.53 and 0.40 occupancies) (see Table 5).

Table 5
5iqy: list of top-ranked iodide site occupancies (≥ 0.40) obtained by applying the SAD-SMAR algorithm compared with those in the deposited refinement (Krishna Das *et al.*, 2016) (see Fig. 2).

Only two peaks are missing. (Sep = separation between corresponding sites.)

Site No.	Occ. SAD-SMAR	Occ. LS	Sep. (Å)
1	1.00	1.00	0.11
2	0.90	0.76	0.12
3	0.82	0.86	0.14
4	0.76	0.78	0.57
5	0.72	0.62	0.34
6	0.71	0.65	0.25
7	0.67	0.62	0.52
8	0.62	0.66	0.34
9	0.60	0.63	0.45
10	–	0.53	–
11	0.48	0.44	0.81
12	0.48	0.60	0.46
13	0.44	0.54	0.24
14	0.40	0.40	0.36
15	0.40	0.42	0.55
16	–	0.40	–
17	0.40	0.40	0.55

2g4h: $a = 182.16$ Å, $F432$. A Cd-containing apoferritin measured at BL X12 (EMBL/DESY, Hamburg) (Mueller-Dieckmann *et al.*, 2007; SAD and refinement data deposited in PDB entry 2g4h). Anomalous signal for Cd^{2+} at $\lambda = 2.00$ Å is large ($f''_{0Cd} \sim 7.2$ e$^-$). According to the deposited refinement, the a.u. contains five Cd^{2+} sites (with occupancies > 0.10), two Cl^{1-} sites, 101 water molecules and one apoferritin subunit (a macromolecule with 1374 atoms). Apoferritin is made up of 24 such protein subunits which assemble to form a roughly spherical hollow shell, with an external diameter of ~ 120 Å and an internal diameter of ~ 80 Å (Chrichton, 2019). The shell is placed at the nodes of the F lattice complex. Application of SAD-SMAR yields the five Cd^{2+} sites ($B \cong 33.7$ Å^2) with the found positions and occupancies close to the deposited values (r.m.s.d. between corresponding sites is 0.32 Å). The respective found and deposited occupancies (using the original site labelling) are Cd1: 0.50, 0.50; Cd2: 0.25, 0.25; Cd3: 0.14, 0.20; Cd4: 0.20, 0.18; Cd5: 0.14, 0.16. The Cd1 sites are located pairwise (~ 8 Å separation) at the 12 vertices of a cubo-octahedron centred at (0, 0, 0) (with opposite vertices separated by ~ 129 Å), *i.e.* close to the external diameter of the hollow shell. The same applies for Cd2 but with a somewhat longer intra-pair distance (~ 13 Å) and a separation between opposite vertices of ~ 75 Å which roughly corresponds to the internal diameter of the hollow shell.

5.3. S-SAD phasing

The data sets of Pf1117 and Pf0907, two hypothetical proteins from *Pyrococcus furiosus*, were collected at BL X06DA at the Swiss Light Source (Weinert *et al.*, 2015; the corresponding SAD and refinement information deposited with respective PDB codes 4tno and 4pgo).

4tno: $a = 47.21$, $c = 82.28$ Å; $P4_12_12$. According to the deposited data, its a.u. contains one macromolecule, three methionine S atoms and two Cl^{1-} (709 atoms; $f''_{0Cl} \sim 1.20$ and $f''_{0S} \sim 0.95$ e$^-$). Application of SAD-SMAR yields the two Cl^{1-} and two S atoms ($B \cong 47.9$ Å^2). The third (more disordered) S atom could not be located. The r.m.s.d. between found and deposited positions is 0.48 Å.

4pgo: $a = 88.50$, $c = 73.12$ Å; $P6_522$. The deposited data indicate that besides the macromolecule and water molecules, there are two methionine S atoms and two Cl^{1-} in the a.u. (~ 689 atoms). Application of SAD-SMAR leads to the same AS model ($B \cong 57.3$ Å^2) with r.m.s.d. = 0.56 Å.

The third and last example is the PB1 domain of the human scaffold protein NBR1 (Müller *et al.*, 2006):

2g4s: $a = 101.40$, $c = 42.59$ Å; $P6_322$. The data set was collected at BL X12 (EMBL/DESY, Hamburg) (Mueller-Dieckmann *et al.*, 2007; SAD and refinement data deposited in PDB entry 2g4s). According to the deposited refinement, the a.u. contains, besides the macromolecule and the refined water molecules, four methionine S atoms (one of them with a higher B value) (689 atoms; $f''_{0Cl} \sim 1.11$ and $f''_{0S} \sim 0.91$ e$^-$). Application of SAD-SMAR shows the four expected S atoms ($B \cong 42.6$ Å^2), three of them as the three strongest Fourier peaks with a r.m.s.d. of only 0.18 Å compared with the deposited model. The fifth-ranked Fourier peak corresponds to the fourth S atom (the one with the higher B value in the refinement) and is shifted by 1.1 Å from the deposited position. The fourth-ranked Fourier peak could not be assigned (perhaps corresponding to some missing Cl^{1-}).

6. Conclusions

Based on the experimental conditions covered by the test examples, it may be concluded that SAD-SMAR can solve efficiently AS substructures from SAD data (i) with upper resolution limits (RES$_{SMAR}$) between 2.50 and 3.3 Å; (ii) with average Bijvoet ratios of 0.065 (for SeMet derivatives), 0.014 (for S-SAD phasing) and 0.041 (for soaked native crystals); (iii) with $\langle |D|/\sigma(|D|) \rangle$ values greater than 1.5; and (iv) with $\langle |D|/\sigma(|D|) \rangle$ values for the outermost resolution shell ranging from 0.90 to 1.69 (the average being 1.25). The cut-off values of the various rejection criteria used in the tests have been ECUT $\cong 0.25$, r.s.m.d.($|D|$) = 4 and DFCUT = ~ 0.4. The introduction of DFCUT ensures the suppression of the less reliable $|D|$'s while keeping enough observations for a satisfactory DM run. It can be clearly seen that the corresponding CC$_h$ values are close to 0.88 for converging trials (with the corresponding R_{CC} values lying between 51 and 69). Since for non-converging trials CC$_h$ values are normally smaller by 0.02–0.03 (and R_{CC} values are in general 1.3–1.4 times larger), identification of the correct trials should not be a problem. Notable is how quickly convergence is reached, especially for SeMet derivatives and for soaked native crystals. For native crystals with only S and/or Cl as AS, the test results clearly indicate that SAD-SMAR can be successfully applied to them. In the three test structures, the S atoms belong to methionine amino acids and no disulfide bridges are present. Since SAD-SMAR only considers the lattice symmetry operations, it processes the initial phase estimates derived from the

randomly shifted M' modulus function quite efficiently (Rius & Torrelles, 2021). As shown in Table 4, the ρ_{peak}/σ limit for considering the peaks at the end of Fourier recycling as part of the structure model can usually be set between 5.0 and 7.0.

To evaluate the influence of the DFCUT value in the phase refinement results, the test calculations were repeated with DFCUT = 0.0 and the results included in Table 3 for comparison. It can be seen that, for converging trials, the CC_h values are similar (~ 0.88) and the R_{CC} values are 2 or 3 units larger (an increase which is otherwise logical since the less reliable $|D|$ values enter in the calculation). The comparison of the number of converging trials (n.c.t.) for both series of calculations indicates that DFCUT = ~ 0.4 gives significantly higher n.c.t. values only for 2g4s and 4tno (by factors 1.89 and 1.10, respectively). This is surely related to their higher R_{free} values (0.323 and 0.305, respectively) when compared with the median R_{free} value of the PDB (0.265).

One characteristic of SAD-SMAR is the delivery of almost complete models when it converges. Most probable causes of non-convergence are, besides the poor quality of the experimental data, some functional limitations of the model description, e.g. when the resolution of the data is not enough to resolve the AS peaks in the Fourier map. Fortunately, due to the large separation among anomalous scatterers, this limitation is generally not a problem. However, at intermediate resolutions (>2.0 Å), the presence of disulfide bridges in proteins, e.g. between cysteine residues, represents a limitation of the otherwise highly effective ipp procedure (the approximate spherical symmetry of individual S Fourier peaks is lost in the overlapped S–S peak). This problem has already been addressed in *SHELXD* (Usón & Sheldrick, 2018; Sheldrick, 2010). It is clear that adapting the ipp philosophy to the treatment of disulfide bridges would considerably expand the scope of SAD-SMAR in S-SAD phasing.

APPENDIX A

By considering $i = \exp[i(\pi/2)]$ in expression (2), this can be written as

$$F_H = |F'_H| \exp(i\varphi'_H) + |F''_H| \exp\left[i\left(\varphi''_H + \frac{\pi}{2}\right)\right]. \quad (20)$$

Multiplication of F_H and F_{-H} by their respective complex conjugates gives, after some algebraic manipulation (Ramachandran & Srinivasan, 1970),

$$|F_{\pm H}|^2 = |F'_H|^2 + |F''_H|^2 \pm 2|F'_H||F''_H|\sin(\varphi'_H - \varphi''_H). \quad (21)$$

Addition of $|F_{+H}|^2$ and $|F_{-H}|^2$ leads to

$$2|F'_H|^2 = |F_{+H}|^2 + |F_{-H}|^2 - 2|F''_H|^2. \quad (22)$$

Likewise, by calculating their difference, the expression

$$|F_{+H}|^2 - |F_{-H}|^2 = 4|F'_H||F''_H|\sin(\varphi'_H - \varphi''_H) \quad (23)$$

is obtained which when squared yields

$$8|F''_H|^2\sin^2(\varphi'_H - \varphi''_H) = \frac{\left(|F_{+H}|^2 - |F_{-H}|^2\right)^2}{2|F'_H|^2} \quad (24)$$

$$= |D_H|^2\frac{|F_{+H}|^2 + |F_{-H}|^2 + 2|F_{+H}||F_{-H}|}{|F_{+H}|^2 + |F_{-H}|^2 - 2|F''_H|^2} \quad (25)$$

wherein

$$D_H = |F_{+H}| - |F_{-H}|. \quad (26)$$

By squaring (26), it follows, after rearranging terms, that $2|F_{+H}||F_{-H}| = |F_{+H}|^2 + |F_{-H}|^2 - |D_H|^2$. Replacement of $2|F_{+H}||F_{-H}|$ in (25) gives

$$= 2|D_H|^2\frac{|F_{+H}|^2 + |F_{-H}|^2 - |D_H|^2/2}{|F_{+H}|^2 + |F_{-H}|^2 - 2|F''_H|^2}. \quad (27)$$

Finally, for Bijvoet pairs satisfying the conditions

$$|F_{+H}|^2 + |F_{-H}|^2 \gg |D_H|^2/2 \quad (28)$$

$$|F_{+H}|^2 + |F_{-H}|^2 \gg 2|F''_H|^2 \quad (29)$$

the fractional term in (27) tends to 1, so that

$$4|F''_H|^2\sin^2(\varphi'_H - \varphi''_H) \simeq |D_H|^2 \quad (30)$$

holds.

This expression is also valid for structures containing anomalous scatterers of different type.

References

Abendroth, J., Gardberg, A. S., Robinson, J. I., Christensen, J. S., Staker, B. L., Myler, P. J., Stewart, L. J. & Edwards, T. E. (2011). *J. Struct. Funct. Genomics*, **12**, 83–95.

Arolas, J. L., Goulas, T., Pomerantsev, A. P., Leppla, S. H. & Gomis-Rüth, F. X. (2016). *Structure*, **24**, 25–36.

Chrichton, R. (2019). *Biological Inorganic Chemistry: a New Introduction to Molecular Structure and Function*, 3rd ed. Amsterdam: Elsevier B. V.

Cianci, M., Helliwell, J. R. & Suzuki, A. (2008). *Acta Cryst.* D**64**, 1196–1209.

Debaerdemaeker, T., Germain, G., Main, P., Tate, C. & Woolfson, M. M. (1987). *MULTAN87. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*. University of York, England.

Einspahr, H., Suguna, K., Suddath, F. L., Ellis, G., Helliwell, J. R. & Papiz, M. Z. (1985). *Acta Cryst.* B**41**, 336–341.

Giacovazzo, C. (2014). *Phasing in Crystallography: a Modern Perspective*. Oxford University Press.

Goulas, T., Garcia-Ferrer, I., Hutcherson, J. A., Potempa, B. A., Potempa, J., Scott, D. A. & Gomis-Rüth (2016). *Mol. Oral Microbiol.* **31**, 472–485.

Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.

Grosse-Kunstleve, R. W. & Brunger, A. T. (1999). *Acta Cryst.* D**55**, 1568–1577.

Hendrickson, W. A., Smith, J. L., Phizackerley, R. P. & Merritt, E. A. (1988). *Proteins*, **4**, 77–88.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature*, **290**, 107–113.

Kanitz, M., Blanck, S., Heine, A., Gulyaeva, A. A., Gorbalenya, A. E., Ziebuhr, J. & Diederich, N. E. (2019). *Virology*, **533**, 21–33.

Karle, J. (1980). *Int. J. Quantum Chem.: Quantum Biol. Symp.* **7**, 357–367.

Krishna Das, B., Kumar, A., Maindola, P., Mahanty, S., Jain, S. K., Reddy, M. K. & Arockiasamy, A. (2016). *Biochem. Biophys. Res. Commun.* **473**, 1152–1157.

Leonarski, F., Redford, S., Mozzanica, A., Lopez-Cuenca, C., Panepucci, E., Nass, K., Ozerov, D., Vera, L., Olieric, V., Buntschu, D., Schneider, R., Tinti, G., Froejdh, E., Diederichs, K., Bunk, O., Schmitt, B. & Wang, M. (2018). *Nat. Methods*, **15**, 799–804.

López-Pelegrín, M., Cerdà-Costa, N., Martínez-Jiménez, F., Cintas-Pedrola, A., Canals, A., Peinado, J. R., Marti-Renom, M. A., López-Otín, C., Arolas, J. L. & Gomis-Rüth, F. X. (2013). *J. Biol. Chem.* **288**, 21279–21294.

Main, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlácek, pp. 97–105. Copenhagen: Munksgaard.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.

Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. & Weiss, M. S. (2007). *Acta Cryst.* D**63**, 366–380.

Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* A**45**, 715–718.

Müller, S., Kursula, I., Zou, P. & Wilmanns, M. (2006). *FEBS Lett.* **580**, 341–344.

Nass, K., Cheng, R., Vera, L., Mozzanica, A., Redford, S., Ozerov, D., Basu, S., James, D., Knopp, G., Cirelli, C., Martiel, I., Casadei, C., Weinert, T., Nogly, P., Skopintsev, P., Usov, I., Leonarski, F., Geng, T., Rappas, M., Doré, A. S., Cooke, R., Nasrollahi Shirazi, S., Dworkowski, F., Sharpe, M., Olieric, N., Bacellar, C., Bohinc, R., Steinmetz, M. O., Schertler, G., Abela, R., Patthey, L., Schmitt, B., Hennig, M., Standfuss, J., Wang, M. & Milne, C. J. (2020). *IUCrJ*, **7**, 965–975.

Ramachandran, G. N. & Srinivasan, R. (1970). *Fourier Methods in Crystallography*. New York: John Wiley & Sons Inc.

Read, R. J., Adams, P. D., Arendall, W. B., Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lütteke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., Vriend, G. & Zwart, P. H. (2011). *Structure*, **19**, 1395–1412.

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, https://crystallography.icmab.es/software.

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

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

Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.

Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.

Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.

Usón, I. & Sheldrick, G. M. (2018). *Acta Cryst.* D**74**, 106–116.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., Li, D., Howe, N., Schnapp, G., Pautsch, A., Bargsten, K., Prota, A. E., Surana, P., Kottur, J., Nair, D. T., Basilico, F., Cecatiello, V., Pasqualato, S., Boland, A., Weichenrieder, O., Wang, B. C., Steinmetz, M. O., Caffrey, M. & Wang, M. (2015). *Nat. Methods*, **12**, 131–133.

Wilson, K. S. (1978). *Acta Cryst.* B**34**, 1599–1608.