crystals

IMPACT FACTOR 2.4

CITESCORE 4.2

Article

Editor's Choice

Extending Ab Initio Phasing up to 2.2 Å Resolution: New Superposition Techniques

Maria Cristina Burla, Benedetta Carrozzini, Giovanni Luca Cascarano, Carmelo Giacovazzo and Giampiero Polidori

Special Issue

Nucleic Acid Crystallography Volume II

Edited by

Dr. Blaine Mooers

MDPI

https://doi.org/10.3390/cryst13060874

crystals

MDPI

Article

Extending Ab Initio Phasing up to 2.2 Å Resolution: New Superposition Techniques

Maria Cristina Burla, Benedetta Carrozzini, Giovanni Luca Cascarano, Carmelo Giacovazzo * and Giampiero Polidori

Istituto di Cristallografia, Consiglio Nazionale delle Ricerche (CNR), Via G. Amendola 122/o, 70126 Bari, Italy; mariacristina.burla@unipg.it (M.C.B.); benedetta.carrozzini@ic.cnr.it (B.C.); gianluca.cascarano@ic.cnr.it (G.L.C.); giampiero.polidori@unipg.it (G.P.)

* Correspondence: carmelo.giacovazzo@ic.cnr.it; Tel.: +39-3316202574

Abstract: Patterson superposition techniques are a historical method for solving the structures of small molecules ab initio, provided they contain heavy atoms in the unit cell. In the 1990s, they were combined with effective EDM procedures and succeeded in the crystal structure solution of macromolecular structures with resolution data up to 1.6–1.9 Å. In this paper we enlarge the concept of Patterson superposition by replacing it with the vector superposition concept. We show, indeed, that besides Patterson other Fourier syntheses may also be used for the superposition of the interatomic vectors. Five Fourier syntheses are described and used in the practical applications. We show that even macromolecular structures with 2.2 Å data resolution may be solved via the new approach.

Keywords: ab initio phasing; Patterson techniques; Fourier syntheses; vector superposition techniques

check for updates

Citation: Burla, M.C.; Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C.; Polidori, G. Extending Ab Initio Phasing up to 2.2 Å Resolution: New Superposition Techniques. *Crystals* **2023**, *13*, 874. https://doi.org/10.3390/cryst13060874

Academic Editor: Blaine Mooers

Received: 26 April 2023
Revised: 23 May 2023
Accepted: 25 May 2023
Published: 26 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Patterson superposition techniques (PSTs) were initially proposed by Wrinch [1], Beevers and Robertson [2] and Buerger [3] and were successfully applied to several practical cases. One of their limitations was the necessary presence of heavy atoms in the target structure.

PSTs were soon relegated to a niche by the development of direct methods. The introduction of sophisticated techniques for the estimation of structure invariants and semi-invariants [4–6] and the automation of corresponding programs via the alternate use of reciprocal and real-space refinement enabled direct methods to solve the phase problem for structures of up to 300 non-H atoms in the asymmetric unit. Among the most popular programs, we mention SnB [7], SHELXD [8], ACORN [9] and SIR2002 [10].

The revival of PSTs started with Nordman [11] and continued with Richardson and Jacobson [12], Sheldrick [13] and Pavelčík et al. [14]. A quite general approach was described by Burla et al. [15], who generalized the automatic procedure to high symmetry and centrosymmetric space groups; they combined PSTs with an effective EDM procedure and succeeded in the crystal structure solution of macromolecular structures with resolution data up to 1.6 Å resolution.

To date, vector superposition technique (VST) has essentially been a different term for the Patterson superposition technique. In this paper, we show that the term VST is more appropriate because, besides Patterson, a larger set of Fourier syntheses may be used, each providing an alternative approach for the solution of the phase problem. Let us invoke a scenario to clarify the point. Besides Patterson, the Fourier syntheses traditionally used in phasing procedures are the hybrid Fourier syntheses. The adjective "hybrid" is currently used for syntheses based on the Fourier coefficients $\tau F_o - \omega F_p$, where τ and ω are small integer numbers. According to this definition, observed, difference and calculated syntheses are special hybrid cases. Two new observed syntheses, studied in Paper II [16],

show a special ability in enhancing peaks corresponding to missed target atoms and in depleting peaks corresponding only to model atoms. Their Fourier coefficients are:

$$[m^y|F| - (m^y - m)|F_p|]exp(i\varphi_p), \text{with } y = 0.3 \quad (1)$$

and

$$(|F| - (1 - m)|F_p|)exp(i\varphi_p) \quad (2)$$

respectively. Both the syntheses behave like the observed one for very good models but share some properties of the difference syntheses for poor models.

In this paper, we will be extending the classic PST concept (as described in Section 4). Specifically, we will demonstrate that superposition techniques can be applied to a larger set of Fourier syntheses beyond just the Patterson synthesis. In Sections 3 and 4, we will describe this set of Fourier syntheses that are useful for the success of superposition techniques. As a result, PST might be replaced by the more appropriate term VST.

When applying PST to non-atomic resolution data, rough electron density maps are often produced. This leads to electron density peaks being frequently misplaced, resulting in a high average phase error. Therefore, a phase refinement step is necessary, where the VSTs described in Section 5 can act as a supplementary tool, specifically as a supplementary constraint to hinder phase degradation. In Section 6, we will describe the VST algorithms used for refining model maps, and in Section 7, we will show their application to practical cases. Our results demonstrate that ab initio crystal structure solutions can be extended up to 2.2 Å resolution.

2. Fourier Syntheses for Vector Superposition Techniques

We describe here the main properties of the Fourier syntheses that may be useful for the solution of the phase problem via VSTs.

2.1. The Ideal $|F|$- and F_p-Syntheses

The ideal synthesis $\rho(\boldsymbol{r})$, with Fourier coefficients $F = |F|exp(i\varphi)$, is not available during the phasing procedure because φ is unknown. On the other hand, the model synthesis $\rho_p(\boldsymbol{r})$, with coefficients $F_p = |F_p|exp(i\varphi_p)$, is not informative because it does not provide supplementary information with respect to the model itself.

2.2. The Observed Synthesis

The observed Fourier synthesis $\rho_o(\boldsymbol{r})$ is traditionally calculated via the coefficients $m|F|exp(i\varphi)$ or, according to Paper I [17], via Coefficients (1) and (2). A simple mathematical trick for understanding where the maxima of a Fourier synthesis are expected to lie, and in general for learning its main properties, is to replace the Fourier coefficients by their formal expression in terms of scattering factors and atomic positions. This technique will be applied to all the new syntheses proposed in this paper. For example, for the $\rho_o(\boldsymbol{r})$ synthesis we have

$$\begin{aligned} \rho_o(\boldsymbol{r}) = \tfrac{1}{V}\sum_{\boldsymbol{h}}|F| \quad & exp(i\phi_p - 2\pi i\boldsymbol{hr}) \\ & = \tfrac{1}{V}\sum_{\boldsymbol{h}} exp(i\Delta\varphi_{\boldsymbol{h}}) \sum_{j=1}^{N} f_j exp(2\pi i\boldsymbol{h}(\boldsymbol{r}_j - \boldsymbol{r})) \end{aligned} \quad (3)$$

According to Paper I, the expected amplitude of the observed synthesis at the jth model atomic position is

$$< \rho_o(\boldsymbol{r}_{pj}) >= \frac{2}{V}\sum_{h}[\frac{\pi}{4} + (1 - \frac{\pi}{4})\sigma_A^2]f_j \quad (4)$$

the expected amplitude at the target atomic positions is

$$\langle \rho_o(\boldsymbol{r}_j) \rangle = \frac{2}{V} \sum_{\boldsymbol{h}} f_j cos(\Delta\varphi_{\boldsymbol{h}}). \tag{5}$$

Usually, $\langle \rho_o(\boldsymbol{r}_{pj}) \rangle > \langle \rho_o(\boldsymbol{r}_j) \rangle$, particularly when the model is poor. The use of the Fourier Coefficients (1) and (2) may make such an inequality less critical.

*2.3. The P(**u**) Patterson Synthesis*

It is the convolution of $\rho(\boldsymbol{r})$ and $\rho(-\boldsymbol{r})$. Since

$$P(\boldsymbol{u}) = \frac{1}{V} \sum_{\mathbf{h}} |F|^2 exp(-2\pi i \boldsymbol{h}\boldsymbol{u}) = \frac{1}{V} \sum_{\mathbf{h}} \sum_{i,j} f_i f_j \, exp\left[2\pi i \boldsymbol{h}(\boldsymbol{r}_i - \boldsymbol{r}_j - \boldsymbol{u})\right],$$

the maxima are expected at:

(i) $\boldsymbol{u} = 0$, with intensity $\frac{1}{V} \sum_{\mathbf{h}} \sum_i f_i^2$, or, roughly speaking, proportional to $\sum_{j=1}^{N} Z_j^2$;

(ii) $\boldsymbol{u}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, for i ≠ j, with intensity

$$P(\boldsymbol{u}_{ij}) = \frac{1}{V} \sum_{\boldsymbol{h}} f_i f_j \tag{6}$$

or, roughly speaking, with intensity proportional to $Z_i Z_j$.

2.4. The F_oF_o-Synthesis

In Paper I the *FF*-synthesis was considered, with density distribution equal to

$$\begin{aligned} \rho_{FF}(\boldsymbol{r}) = \tfrac{1}{V} \textstyle\sum_{\mathbf{h}} \quad & |F|^2 exp[i(2\varphi - 2\pi \boldsymbol{h}\boldsymbol{r})] \\ & = \tfrac{1}{V} \textstyle\sum_{\boldsymbol{h}} \sum_{i,j} f_i f_j exp\left[2\pi i \boldsymbol{h}(\boldsymbol{r}_i + \boldsymbol{r}_j - \boldsymbol{r})\right] \end{aligned}$$

In practice φ is unknown and therefore ρ_{FF} cannot be calculated. We replace it by the F_oF_o-synthesis with density distribution given by

$$\begin{aligned} \rho_{F_oF_o}(\boldsymbol{r}) = \tfrac{1}{V} \textstyle\sum_{\mathbf{h}} \quad & |F|^2 exp\left[i(2\varphi_p - 2\pi \boldsymbol{h}\boldsymbol{r})\right] \\ & = \tfrac{1}{V} \textstyle\sum_{\boldsymbol{h}} exp(2i\Delta\varphi_{\boldsymbol{h}}) \sum_{i,j} f_i f_j exp[2\pi i \boldsymbol{h}(\boldsymbol{r}_i + \boldsymbol{r}_j - \boldsymbol{r}) \end{aligned} \tag{7}$$

We observe the following:

(i) $\rho_{F_oF_o}$ is the convolution of $\rho_o(\boldsymbol{r})$ with itself, which causes the individual peaks to be broader than in a typically observed electron density map.

(ii) If the quality of the model is sufficiently high (that is, if for a non-negligible percentage of observed reflections $\Delta\varphi_h \sim 0$), then peaks are expected to be located at $\boldsymbol{r} = \boldsymbol{r}_i + \boldsymbol{r}_j$ for i, j = 1, 2, . . . , N, i.e., at the sum of positional vectors rather than at the interatomic vectors $\boldsymbol{r}_i - \boldsymbol{r}_j$ (as in the Patterson synthesis). Thus, the number of peaks in the unit cell is equal to N^2, as in the Patterson map.

(iii) The peak intensity is expected to be proportional to $Z_j Z_j$ as in the Patterson synthesis, but the presence of $2\Delta\varphi_{\boldsymbol{h}}$ modulates the contribution of each reflection to the right-hand side of Equation (7). In analogy with the appendix in Paper I the observed electron density at the target atomic positions is expected to be $\sum_{\boldsymbol{h}} f_i f_j cos(2\Delta\varphi_{\boldsymbol{h}})$. The larger the phase error, the weaker will the peaks be.

(iv) $\rho_o(\boldsymbol{r})$ and $\rho_{F_oF_o}$ show the same rotational symmetry, but different translational symmetries. In particular, the space group symmetry elements of the target electron density transform according to the following rules: $2_1 \rightarrow 2$, $(4_1, 4_3) \rightarrow 4_2$, $3_1 \rightarrow 3_2$, $(6_1, 6_5) \rightarrow 6_2$, $(6_2, 6_4) \rightarrow 6_4$. Glide planes *a*, *b*, *c*, *n* transform into the mirror plane *m* and glide planes *d* into *n*.

(v) $\Delta\varphi_{\boldsymbol{h}} = 0, \pi$ for reflections with symmetry restricted phase. Since $2\Delta\varphi_{\boldsymbol{h}} = 2\pi$, this subset of reflections is not subject to the disturbing effect of $exp(i2\Delta\varphi_{\boldsymbol{h}})$.

(vi) If the quality of the model is sufficiently high, $exp(i2\Delta\varphi_{\boldsymbol{h}})$ may be replaced by its mean value, and the $\rho_{F_oF_o}$-map can be computed.

2.5. The $F_o\overline{F}_p$-Synthesis

The synthesis with coefficients $F\overline{F}_p$ was studied by Carrozzini et al. [18], for brevity indicated as C-function; it was shown that its space group is the symmorphic variant of the space group of the target structure (e.g., if P $2_12_{,1}2_1$ is the target space group, the C-function space group is P 222), The synthesis is equal to

$$\begin{array}{ll}\rho_{F\overline{F}_p}(\boldsymbol{r}) = \frac{1}{V}\sum_{\mathbf{h}} & F\overline{F}_p exp(-2\pi i\boldsymbol{hr}) = \frac{1}{V}\sum_{\mathbf{h}}|F||F_p|exp\left[i\left(\varphi-\varphi_p-2\pi\boldsymbol{hr}\right)\right] \\ & = \frac{1}{V}\sum_{\boldsymbol{h}}\sum_{i,j} f_i f_{pj} exp\left[2\pi i\boldsymbol{h}\left(\boldsymbol{r}_i-\boldsymbol{r}_{pj}-\boldsymbol{r}\right)\right]\end{array} \quad (8)$$

Equation (8) is not computable because the φs are unknown. To overcome this problem, Carrozzini et al. [18] supposed that the quality of the model was sufficiently high to guarantee that, for a non-negligible subset of reflections, the m value is sufficiently large. Then, the relation $exp(i\varphi) \approx m\, exp\left(i\varphi_p\right)$ holds, so that

$$F||\overline{F}_p|exp[i\left(\varphi-\varphi_p-2\pi\boldsymbol{hr}\right) \approx m|F||\overline{F}_p|exp(-2\pi i\boldsymbol{hr})$$

and

$$F\overline{F}_p \approx m|F||\overline{F}_p|. \quad (9)$$

In this case Synthesis (9), denoted as C', is computable but the map is centric. Caliandro et al. [19,20] combined the C'-map with PSTs for solving structures with data at non-atomic resolution.

In this paper we study the computable $F_o\overline{F}_p$-synthesis, with density given by

$$\begin{array}{ll}\rho_{F_o\overline{F}_p}(\boldsymbol{r}) = \frac{1}{V}\sum_{\mathrm{h}} & F_o\overline{F}_p exp(-2\pi i\boldsymbol{hr}) \\ & = \frac{1}{V}\sum_{\boldsymbol{h}} exp(i\Delta\varphi_{\boldsymbol{h}})\sum_{i,j} f_i f_{pj} exp\left[2\pi i\boldsymbol{h}\left(\boldsymbol{r}_i-\boldsymbol{r}_{pj}-\boldsymbol{r}\right)\right]\end{array} \quad (10)$$

We observe:

(i) $\rho_{F_o\overline{F}_p}$ is the convolution between $\rho_o(\boldsymbol{r})$ and $\rho_p(-\boldsymbol{r})$, and therefore peaks are broader than in a typical electron density map.

(ii) The intensity of the peaks is modulated by the factor $exp(i\Delta\varphi_{\boldsymbol{h}})$; for sufficiently good models the maxima are expected at $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_{pj}$, for i = 1, ... , N and j = 1, ... N_p, with intensity proportional to $\sum_{\boldsymbol{h}} f_i f_{pj} cos(\Delta\varphi_{\boldsymbol{h}})$ (approximately proportional to Z_iZ_{pj}). Their number is expected to be equal to NN_p, usually much smaller than the peak number in the Patterson map; peaks' superposition is therefore lower.

(iii) A strong peak is expected at $\boldsymbol{r} = 0$, corresponding to the cases for which $\boldsymbol{r}_i \sim \boldsymbol{r}_{pj}$; its amplitude becomes larger and larger when the model improves. Thus, non-zero peaks correspond only to interatomic distances with $\boldsymbol{r}_i \neq \boldsymbol{r}_{pj}$; they involve atoms which are not in the model.

(iv) Equation (10) is not computable in practical cases unless $exp(i\Delta\varphi_{\boldsymbol{h}})$ is replaced by its expected value $m_{\boldsymbol{h}}$, but it becomes centric, so sharing the Patterson symmetry.

2.6. The $F_p|F_o|^2$-Synthesis

Its electron density is given by

$$\begin{array}{ll}\rho_{F_p|F_o|^2}(\boldsymbol{r}) = \frac{1}{V} & \sum_{\mathbf{h}}|F_p||F|^2 exp\left[i\left(\varphi_p-2\pi\boldsymbol{hr}\right)\right] \\ & = \frac{1}{V}\sum_{\boldsymbol{h}}\sum_{j,l} f_{pi} f_j f_l exp\left[2\pi i\boldsymbol{h}\left(\boldsymbol{r}_{pi}+\boldsymbol{r}_j-\boldsymbol{r}_l-\boldsymbol{r}\right)\right]\end{array} \quad (11)$$

We observe the following:

(i) The map is the convolution between $\rho_p(\boldsymbol{r})$ and the Patterson synthesis $P(\boldsymbol{u})$. Therefore, its peaks will be broader compared to those in $\rho_p(\boldsymbol{r})$ or $\rho_p(\boldsymbol{r})$.

(ii) The maxima are at $\boldsymbol{r} = \boldsymbol{r}_{pi} + \boldsymbol{r}_j - \boldsymbol{r}_l$, for j, l = 1, ... , N and p = 1, ... , N_p, with intensity proportional to $Z_{pi}Z_jZ_l$. Therefore, peaks involving heavy atom positions will be dominant, in the order of peaks involving three, two and one heavy atom. The map is expected to contain N_pN^2 peaks, with large mutual overlap.

(iii) The strongest peaks are expected to lie at $\boldsymbol{r}_{pi}$, i = 1, ... , N_p, and each of them has multiplicity equal to N because j = l for N times. Less dominant peaks lie at $\boldsymbol{r}_j$, for i = 1, ... , N, and each of them will have multiplicity equal to N_p only if, for N_p times, $\boldsymbol{r}_{pi} = \boldsymbol{r}_l$. Unfortunately, this condition is not frequent because the model atoms may be close to or far away from the target atoms. This last characteristic makes $\rho_{F_p|F|^2}$ less appealing; indeed, the usefulness of maps dedicated to the target structure recovery may be estimated from the prominence of the peaks corresponding to target atoms. We will see that more useful maps may correspond to the $|F_p|_2|F_o|$-syntheses.

2.7. The $|F_p|_2|F_o|$-Synthesis

Its electron density is given by

$$\begin{array}{ll}\rho_{|F_p|^2F_o}(\boldsymbol{r}) = \frac{1}{V} & \sum_{\mathbf{h}} |F_p|^2|F| exp\left[i\left(\varphi_p - 2\pi \boldsymbol{hu}\right)\right] \\ & = \frac{1}{V}\sum_{\boldsymbol{h}} exp(i\Delta\varphi_{\boldsymbol{h}}) \sum_{i,j} f_i f_{pj} f_{pl} exp\left[2\pi i \boldsymbol{h}\left(\boldsymbol{r}_i + \boldsymbol{r}_{pj} - \boldsymbol{r}_{pl} - \boldsymbol{r}\right)\right]\end{array} \quad (12)$$

We observe that:

(i) The map is the convolution between $\rho_o(\boldsymbol{r})$ and the Patterson synthesis $P(\boldsymbol{u})$, thus single peaks will be broader than in $\rho_p(\boldsymbol{r})$ or $\rho_p(\boldsymbol{r})$.

(ii) If the model is of sufficiently high quality the maxima are expected at $\boldsymbol{r} = \boldsymbol{r}_i + \boldsymbol{r}_{pj} - \mathbf{r}_{pl}$, for i = 1, ... , N and j, l = 1, ... , N_p, with intensity proportional to the product $Z_iZ_{pj}Z_{pl}$. However, if the model is poor, the factor $exp(i\Delta\varphi_{\boldsymbol{h}})$ may reduce the intensities of all peaks and shift them from the expected positions.

(iii) Peaks involving heavy atom positions will be dominant (in the order of peaks involving three, two and one heavy atom). $\rho_{|F_p|^2F_o}(\boldsymbol{r})$ is expected to contain $NN_p{}^2$ peaks, with large mutual overlap, usually $NN_p{}^2 << N_pN^2$, so that overlap in the $|F_p|_2|F_o|$-synthesis is less severe than in $\rho_{F_p|F|^2}(\boldsymbol{r})$.

(iv) If the model is sufficiently good, large peaks are expected at $\boldsymbol{r}_i$ for i = 1, ... , N, with multiplicity of order N_p (since l = j for N_p times). Peaks are also expected at $\boldsymbol{r}_{pj}$, j = 1, ... , N_p, with each having multiplicity equal to N_p only if $\boldsymbol{r}_i = \boldsymbol{r}_{pl}$ for N_p times. Fortunately, this condition is not common because the model atoms may be close to or far away from the target atoms; consequently, an image of the target structure may be obtained in favorable cases.

(v) Synthesis (12) is not practically computable, so in our calculations $exp(i\Delta\varphi_{\boldsymbol{h}})$ is replaced by its expected value $m_{\boldsymbol{h}}$.

2.8. The Difference Patterson Synthesis

It is defined as

$$\begin{array}{ll}\Delta P(\boldsymbol{u}) = \frac{1}{V}\sum_{\mathbf{h}}(|F|^2 - & |F_p|^2) exp(-2\pi i \boldsymbol{hu}) = \frac{1}{V}\sum_{\boldsymbol{h}}\sum_{i,j} f_i f_j exp 2\pi i \boldsymbol{h}\left(\boldsymbol{r}_i - \boldsymbol{r}_j - \boldsymbol{u}\right) - \\ & \sum_{i,j} f_{pi} f_{pj} exp 2\pi i \boldsymbol{h}\left(\boldsymbol{r}_{pi} - \boldsymbol{r}_{pj} - \boldsymbol{u}\right)\end{array} \quad (13)$$

The right-hand side of Equation (13) suggests the presence of positive maxima at $\boldsymbol{u}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, for i, j = 1, ... , N, and of negative minima at $\boldsymbol{u}_{pij} = \boldsymbol{r}_{pi} - \boldsymbol{r}_{pj}$, for i, j = 1, ... , N_p. However, if $\boldsymbol{u}_{pij}$ and $\boldsymbol{u}_{ij}$ simultaneously belong to the model and to the target (i.e., $\boldsymbol{u}_{pij} \sim \boldsymbol{u}_{ij}$) then the resulting peak will vanish. Positive peaks are expected to correspond to interatomic distances in the target that are not present in the model, and negative peaks are expected to correspond to interatomic distances in the model that are not present in the target.

3. Patterson Superposition Techniques

A brief review of the PSTs may serve as a useful introduction to the more general VSTs. The electron density map $\rho_o(\boldsymbol{r})$, shifted by the vector $\boldsymbol{r}_H$ (the subsript H is usually associated with heavy atoms), is defined by the convolution

$$\rho_o(\boldsymbol{r}+\boldsymbol{r}_H) = \rho_o(\boldsymbol{r}) * \delta(\boldsymbol{r}+\boldsymbol{r}_H)$$

The Fourier transform of this equation gives:

$$T[\rho_o(\boldsymbol{r}) * \delta(\boldsymbol{r}+\boldsymbol{r}_H)] = T[\rho_o(\boldsymbol{r})]T[\delta(\boldsymbol{r}+\boldsymbol{r}_H)] = |F|exp(2\pi i\boldsymbol{h}\boldsymbol{r}_H)$$

The right-hand member is the Fourier coefficient of the shifted Fourier map. Similarly, the shifted Patterson map can be computed by using $|F|^2 exp(2\pi i\boldsymbol{h}\boldsymbol{r}_H)$ as the Fourier coefficient. The equation for this is:

$$\begin{aligned} P(\boldsymbol{u}+\boldsymbol{r}_H) = \tfrac{1}{V} \quad & \textstyle\sum_{\mathbf{h}} |F|^2 exp[-2\pi i\boldsymbol{h}(\boldsymbol{u}-\boldsymbol{r}_H)] \\ & = \tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}} f_H \sum_i f_i exp[2\pi i\boldsymbol{h}(\boldsymbol{r}_i-\boldsymbol{u}) \\ & + \tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}} \sum_{i,j\neq H} f_i f_j \, exp[2\pi i\boldsymbol{h}\big(\boldsymbol{r}_i-\boldsymbol{r}_j+\boldsymbol{r}_H-\boldsymbol{u}\big)] \end{aligned}$$

The maxima of this equation are expected at $\boldsymbol{u} = \boldsymbol{r}_i - \boldsymbol{r}_j + \boldsymbol{r}_H$, for i, j = 1, ... , N. The first term on the right-hand side of the equation corresponds to the case $\boldsymbol{r}_H = \boldsymbol{r}_i$; it may provide an image of the target structure because maxima are expected at $\boldsymbol{r}_i$, for i = 1, ... , N. The last term generates numerous noise peaks located at $\boldsymbol{u} = \boldsymbol{r}_i - \boldsymbol{r}_j + \boldsymbol{r}_H$, for i, j = 1, ... , N.

The noise can be partially eliminated by using PSTs, which can be summarized as follows.

(i) The symmetry minimum function (*SMF*) is calculated by combining symmetry-independent Harker domains according to

$$SMF(\boldsymbol{r}) = Min_{s=1}^{n}[P(\boldsymbol{r}-\boldsymbol{C}_s\,\boldsymbol{r})] \tag{14}$$

where $\boldsymbol{r}-\boldsymbol{C}_s\,\boldsymbol{r}$ is a typical Harker vector. *Min* is the minimum operator to be applied, pixel by pixel, to the n Harker sections.

(ii) A pivot peak in the *SMF* map is selected and used to calculate the minimum superposition function $S(\boldsymbol{r})$ between the *SMF* and the translated Patterson map, according to

$$S(\boldsymbol{r}) = Min[P(\boldsymbol{u}+\boldsymbol{r}_H), SMF(\boldsymbol{r})] \tag{15}$$

where $\boldsymbol{r}_H$ denotes the position of the pivot peak, usually corresponding to a heavy atom.

If more pivot peaks are available, then (15) is replaced by

$$S(\boldsymbol{r}) = Min[P(\boldsymbol{u}+\boldsymbol{r}_{H1}), P(\boldsymbol{u}+\boldsymbol{r}_{H2}), ..., SMF(\boldsymbol{r})] \tag{16}$$

where $\boldsymbol{r}_{H1}, \boldsymbol{r}_{H2}$, ... are the pivot positional vectors, usually corresponding to heavy atom positions.

The above-described techniques frequently provide a poor model structure, containing one or a few correctly located heavy atoms and a small number of light atoms located correctly, mixed with many light atoms located incorrectly. In some space groups $S(\boldsymbol{r})$ shows a residual centrosymmetry, as heritage of the centric nature of the Patterson map. Model refinement is therefore a necessary step, which is traditionally performed by electron density modification (EDM) techniques; we will show that if the data resolution is not sufficiently high, EDM phase refinement may be difficult. When we refer to the effectiveness of the PSTs we will also include the contribution of the EDM techniques, as these are the necessary final step of the method.

In this paper we will show that the use of the VSTs described in the following section makes the crystal structure solution easier. They follow the application of the PSTs and

require that at least one heavy atom has been correctly located via PSTs; for simplicity, its position will be reported as $\boldsymbol{r}_H = \boldsymbol{r}_{pH}$.

4. Vector Superposition Techniques

Let us apply proper origin shifts to the electron syntheses described in Section 3.

4.1. The Shifted F_oF_o-Synthesis

Shifted by the vector $-\boldsymbol{r}_H$, its map may be rewritten as

$$\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H) = \frac{1}{V}\sum_{\boldsymbol{h}}\sum_{i,j} f_i f_j expi\left[2\Delta\varphi_{\boldsymbol{p}} + 2\pi\boldsymbol{h}\left(\boldsymbol{r}_i + \boldsymbol{r}_j - \boldsymbol{r} - \boldsymbol{r}_H\right)\right] \tag{17}$$

When $\boldsymbol{r}_H = \boldsymbol{r}_j$, Equation (17) reduces to

$$\begin{array}{rl}\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H) = & \frac{1}{V}\sum_{\boldsymbol{h}} f_H expi(2\Delta\varphi_{\boldsymbol{h}})\sum_i f_i expi[2\pi\boldsymbol{h}(\boldsymbol{r}_i-\boldsymbol{r})] \\ & +\frac{1}{V}\sum_{\boldsymbol{h}}\sum_{i,j\neq H} f_i f_j expi[2\Delta\varphi_{\boldsymbol{h}} + 2\pi\boldsymbol{h}\left(\boldsymbol{r}_i+\boldsymbol{r}_j-\boldsymbol{r}-\boldsymbol{r}_H\right)\end{array} \tag{18}$$

If $\Delta\varphi_{\boldsymbol{h}} \sim 0$ for a sufficiently large number of reflections the first term on the right-hand side of Equation (18) will produce an image of the target structure; the observed electron density at the target atomic positions is expected to be the sum $\sum_{\boldsymbol{h}} f_i f_H \cos(2\Delta\varphi_{\boldsymbol{h}})$. The second term at the right-hand side of Equation (18) corresponds to noise peaks at $\boldsymbol{r}_i + \boldsymbol{r}_j - \boldsymbol{r}_H$.

$\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)$ alone cannot provide a high-quality image of the target structure, but the minimum function $Min[P(\boldsymbol{u}+\boldsymbol{r}_H),\ \rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)]$ may help to reduce the number of noise peaks in $\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)$ and consequently the mean phase error. Indeed, the noise peaks of the two maps do not systematically overlap (they are at $\boldsymbol{r}_i - \boldsymbol{r}_j + \boldsymbol{r}_H$ for the shifted Patterson and at $\boldsymbol{r}_i + \boldsymbol{r}_j - \boldsymbol{r}_H$ for the shifted F_oF_o-synthesis). If $Min[P(\boldsymbol{u}+\boldsymbol{r}_H), \rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)]$ reduces the noise in the $\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)$ map, then the average value of $2\Delta\varphi_{\boldsymbol{h}}$ in Equation (7) will also diminish. This property suggests a cyclic algorithm of the above operations.

4.2. The Shifted $F_o\overline{F}_p$-Synthesis

Shifted by the vector $\boldsymbol{r}_H$, it may be rewritten as

$$\begin{array}{rl}\rho_{F\overline{F}_p}(\boldsymbol{r}+\boldsymbol{r}_H) = & \frac{1}{V}\sum_{\boldsymbol{h}} f_{pH} exp(i\Delta\varphi_{\boldsymbol{h}})\sum_i f_i exp[2\pi i\boldsymbol{h}(\boldsymbol{r}_i-\boldsymbol{r})] \\ & +\frac{1}{V}\sum_{\boldsymbol{h}} exp(i\Delta\varphi_{\boldsymbol{h}})\sum_{i,j\neq H} f_i f_{pj} exp\left[2\pi i\boldsymbol{h}\left(\boldsymbol{r}_i-\boldsymbol{r}_{pj}+\boldsymbol{r}_H-\boldsymbol{r}\right)\right]\end{array} \tag{19}$$

However, in practical cases $\Delta\varphi_{\boldsymbol{h}}$ is unknown, and the above synthesis cannot be calculated. If the model is sufficiently good, $exp(i\Delta\varphi_{\boldsymbol{h}})$ may be replaced by its mean value $m_{\mathbf{h}}$ for a large number of reflections, and Equation (19) becomes computable. Then, the first term at the right side of Equation (19) would provide an image of the target structure.

A comparison between $\rho_{F\overline{F}_p}(\boldsymbol{r}+\boldsymbol{r}_H)$ and $\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)$ may be useful:

(i) The noise peaks are located at $\boldsymbol{r}_i - \boldsymbol{r}_{pj} + \boldsymbol{r}_H$; their number is of order NN_p, much smaller than the number of noise peaks (N^2) in $\rho_{F_oF_o}(\boldsymbol{r}-\boldsymbol{r}_H)$.

(ii) The unknown phase error parameter $exp(2i\Delta\varphi_{\boldsymbol{h}})$ on the right-hand side of Equation (18) is replaced by the most convenient parameter $exp(i\Delta\varphi_{\boldsymbol{h}})$ in Equation (19). Both parameters affect the image of the target structure resulting from the two syntheses, but the second one is more manageable.

(iii) $\rho_{F\overline{F}_p}(\boldsymbol{r}+\boldsymbol{r}_H)$ cannot lead to a high-quality image of the target structure, but the minimum function $Min[P(\boldsymbol{u}+\boldsymbol{r}_H),\ \rho_{F\overline{F}_p}(\boldsymbol{r}+\boldsymbol{r}_H)]$ may help to reduce the number of noise peaks and consequently the mean phase error.

4.3. The Shifted $F_p|F_o|^2$-Synthesis

Let $\boldsymbol{C_s}\boldsymbol{r}_H$ be a position symmetry equivalent to $\boldsymbol{r}_H$.

If we shift $\rho_{F_p|F|^2}(\boldsymbol{r})$ by the vector $\boldsymbol{u}_H = (\boldsymbol{I} - \boldsymbol{C_s})\boldsymbol{r}_H$, we have

$$\begin{aligned}\rho_{F_p|F|^2}(\boldsymbol{r}+\boldsymbol{u}_H) &= \tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}}\sum_i f_{pi}f_H^2 exp[2\pi i\boldsymbol{h}(\boldsymbol{r}_i-\boldsymbol{r})] \\ &+\tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}}{\sum_{j,l}}' f_{pi}f_jf_l exp\left[2\pi i\boldsymbol{h}\left(\boldsymbol{r}_{pi}+\boldsymbol{r}_j-\boldsymbol{r}_l+\boldsymbol{u}_H-\boldsymbol{r}\right)\right]\end{aligned} \tag{20}$$

The first term at the right-hand side of Equation (20) corresponds to the case $\boldsymbol{r}_H = \boldsymbol{r}_l$ and $\boldsymbol{C_s}\boldsymbol{r}_H = \boldsymbol{r}_j$, and it is expected to provide an image of the target structure. The second term includes all the other cases (this is the reason for the prime to one summation symbol) and generates noise peaks at $\boldsymbol{r} = \boldsymbol{r}_{pi} + \boldsymbol{r}_j - \boldsymbol{r}_l + (\mathbf{I} - \boldsymbol{C_s})\boldsymbol{r}_H$.

4.4. The Shifted $|F_p|_2|F_o|$-Synthesis

As in Section 4.3, let $\boldsymbol{C_s}\boldsymbol{r}_H$ be a position symmetry equivalent to $\boldsymbol{r}_H$.
Shifting the $|F_p|_2F_o$-synthesis by the vector $\boldsymbol{u}_H = (\boldsymbol{I} - \boldsymbol{C_s})\boldsymbol{r}_H$ gives

$$\begin{aligned}\rho_{|F_p|^2F_o}(\boldsymbol{r}+\boldsymbol{u}_H) &= \tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}} exp(i\Delta\varphi_{\boldsymbol{h}})\sum_i f_i f_H^2 exp[2\pi i\boldsymbol{h}(\boldsymbol{r}_i-\boldsymbol{r})] \\ &+\tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}} exp(i\Delta\varphi_{\boldsymbol{h}}){\sum_{j,l}}' f_i f_{pj} f_{pl} exp\{2\pi i\boldsymbol{h}[\boldsymbol{r}_i+\boldsymbol{r}_{pj}-\boldsymbol{r}_{pl}-\boldsymbol{r} \\ &+(\boldsymbol{I}-\boldsymbol{C_s})\boldsymbol{r}_H]\}\end{aligned} \tag{21}$$

The first term at the right-hand side of Equation (21) corresponds to the case $\boldsymbol{r}_H = \boldsymbol{r}_{pl}$ and $\boldsymbol{C}_s\boldsymbol{r}_H = \boldsymbol{r}_{pj}$; it is expected to provide an image of the target structure if the model structure is sufficiently good. The second term includes all the other cases (see the prime to one summation symbol) and generates noise peaks at $\boldsymbol{r} = \boldsymbol{r}_i + \boldsymbol{r}_{pj} - \boldsymbol{r}_{pl} + (\mathbf{I} - \boldsymbol{Cs})\boldsymbol{r}_H$.

4.5. The Shifted Difference Patterson Synthesis

When shifted by the vector $\boldsymbol{r}_H = \boldsymbol{r}_j$, it may be rewritten as

$$\begin{aligned}\Delta P(\boldsymbol{u}-\boldsymbol{r}_H) = \tfrac{1}{V} &\textstyle\sum_{\boldsymbol{h}} f_H\sum_i f_i exp2\pi i\boldsymbol{h}(\boldsymbol{r}_i-\boldsymbol{u}) \\ &+\tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}}\sum_{i,j\neq H} f_i f_j exp2\pi i\boldsymbol{h}\left(\boldsymbol{r}_i-\boldsymbol{r}_j+\boldsymbol{r}_H-\boldsymbol{u}\right) \\ &-\tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}} f_H\sum_i f_i exp2\pi i\boldsymbol{h}\left(\boldsymbol{r}_{pi}-\boldsymbol{u}\right) \\ &-\tfrac{1}{V}\textstyle\sum_{\boldsymbol{h}}\sum_{i,j} f_i f_{j\neq H} exp2\pi i\boldsymbol{h}\left(\boldsymbol{r}_{pi}-\boldsymbol{r}_{pj}+\boldsymbol{r}_H-\boldsymbol{u}\right)\end{aligned} \tag{22}$$

Equation (22) suggests that the shift results in a positive image of the target crystal structure and a negative image of the model structure. Peaks belonging to both the target and to the model structure are expected to have nearly zero amplitude, while peaks not belonging to the model are expected to have a large positive amplitude. Negative amplitude is expected for peaks belonging to the model but not to the target.

In Table 1 we show the suggested shifting vector ($\mathbf{V}_\mathrm{H}$), the expected positional vectors of the signal peaks (Exp. Signal) and their amplitudes (Exp. Sig. Ampl.) and the expected positional vectors of the noise peaks (Noise Peaks) for each shifted Fourier synthesis. The shifted synthesis $\Delta P(\boldsymbol{u} - \boldsymbol{r}_H)$ shows two lines, the first corresponding to positive peaks and the second to negative peaks. The table suggests that each Fourier synthesis can provide an image of the structure, while different positional vectors are obtained for the noise peaks. In principle, a VSP procedure that combines the structural information provided by the different shifted Fourier syntheses may enhance the image of the structure and reduce the noise peaks.

Table 1. For each Fourier synthesis applied in this paper we show: the shifting vector ($\mathbf{V_H}$), the expected image of the target structure (Exp. Signal), the expected amplitude of the signal peaks (Exp. Sig. Amp.), the location of the noise peaks (Noise Peaks).

Synthesis	$\mathbf{V_H}$	Exp. Signal	Exp. Sig. Ampl.	Noise Peaks
P	$\mathbf{r}_H$	$\mathbf{r_i}$, for i = 1, . . . , N	$Z_H Z_i$	$\boldsymbol{r}_i - \boldsymbol{r}_j + \boldsymbol{r}_H$
$\boldsymbol{F_o F_o}$	$-\mathbf{r}_H$	$\mathbf{r_i}$, for i = 1, . . . , N	$Z_H Z_i \langle exp(2i\Delta\varphi)\rangle$	$\boldsymbol{r}_i + \boldsymbol{r}_j - \boldsymbol{r}_H$
$\boldsymbol{F_o \overline{F}_p}$	$\mathbf{r}_H$	$\mathbf{r_i}$, for i = 1, . . . , N	$Z_H Z_i \langle exp(i\Delta\varphi)\rangle$	$\boldsymbol{r}_i - \boldsymbol{r}_{pj} + \boldsymbol{r}_H$
$\boldsymbol{F_p \left\vert F_o \right\vert^2}$	$(\mathbf{I} - \boldsymbol{C_s})\boldsymbol{r}_H$	$\mathbf{r_i}$, for i = 1, . . . , N	$Z_H^2 Z_i$	$\boldsymbol{r}_{pi} + \boldsymbol{r}_j - \boldsymbol{r}_l + (\mathbf{I} - \boldsymbol{Cs})\boldsymbol{r}_H$
$\boldsymbol{\left\vert F_p \right\vert_2 F_o}$	$(\mathbf{I} - \boldsymbol{C_s})\boldsymbol{r}_H$	$\mathbf{r_i}$, for i = 1, . . . , N	$Z_H^2 Z_i \langle exp(i\Delta\varphi)\rangle$	$\boldsymbol{r}_i + \boldsymbol{r}_{pj} - \boldsymbol{r}_{pl} + (\mathbf{I} - \boldsymbol{Cs})\boldsymbol{r}_H$
$\boldsymbol{\Delta P}$	$\mathbf{r}_H$	$\mathbf{r_i}$, for i = 1, . . . , N	$Z_H Z_i$	$\boldsymbol{r}_i - \boldsymbol{r}_j + \boldsymbol{r}_H$
		$\mathbf{r_{pi}}$, for i = 1, . . . , $\mathrm{N_p}$	$Z_H Z_i$	$\boldsymbol{r}_{pi} - \boldsymbol{r}_{pj} + \boldsymbol{r}_H$

5. Discussion

We selected 12 test structures with data resolution (RES) between 1.8 Å and 2.2 Å, each containing one or more heavy atoms in the asymmetric unit. Their PDB codes and RES are shown in Table 2: 10 are proteins, 2 are nucleic acids. For each test structure we calculated the symmetry minimum function (SMF) using Equation (14). On the corresponding map we selected the six peaks with the largest intensities to use them as PST pivots. A number of six was suggested by two conflicting needs: first, taking a low number of trials, second, considering the fact that the largest peak in the Harker section is not always really a Harker peak.

Table 2. For the 12 test structures we show: PDB codes (PDB), data resolution (RES), minimum average phase error among the 6 $S(\boldsymbol{r})$ maps (MPE_S), minimum average phase error (MPE_{EDM}) found when the 6 $S(\boldsymbol{r})$ sets of phases are submitted to 30 EDM cycles. In the columns $\mathrm{MPE}_{F_oF_o}$, $\mathrm{MPE}_{F_o\overline{F}_p}$, $\mathrm{MPE}_{F_p|F_o|^2}$, $\mathrm{MPE}_{|F_p|^2F_o}$, we show the minimum average phase errors attained when each of the 6 $S(\boldsymbol{r})$ sets of phases are submitted to 30 VST cycles based on the F_oF_o-, $F_o\overline{F}_p$-, $F_p\left|F_o\right|^2$- and $\left|F_p\right|_2F_o$ -synthesis, respectively.

PDB	RES	MPE_s	MPE_{EDM}	$\mathrm{MPE}_{F_oF_o}$	$\mathrm{MPE}_{F_o\overline{F}_p}$	$\mathrm{MPE}_{F_p\vert F_o\vert^2}$	$\mathrm{MPE}_{\vert F_p\vert^2F_o}$
4ms5	2.23	59.92	69.05	51.77	51.08	56.21	55.48
3ajw	2.10	69.40	75.35	60.71	61.91	60.92	61.18
4ltl	2.07	57.35	67.05	43.28	42.99	52.35	49.88
1crm	2.02	68.77	71.28	68.15	68.53	73.44	73.57
1z1y	2.00	72.59	73.95	63.11	63.73	74.88	73.62
1buu	1.93	66.64	64.45	56.73	55.93	48.46	47.70
1yfd	1.90	73.17	74.70	71.99	71.54	70.01	69.73
1jpr	1.89	70.87	74.00	64.67	64.58	65.73	65.59
1naq	1.86	68.44	70.75	59.33	59.12	68.85	69.02
1ytt	1.80	61.06	63.08	50.21	50.07	52.37	53.01
1pm2	1.80	77.40	77.96	69.93	69.94	78.81	78.53
1arm	1.80	69.42	70.34	68.34	69.08	74.33	70.99

For each pivot peak a $S(\boldsymbol{r})$ map was calculated using Equation (15). The six maps provided the first model structures: one EDM cycle, with map modification and subsequent Fourier inversion, yields the corresponding six sets of phases. Our computing tool is DM [21], a well-known and effective EDM program. To simplify the results, for each test structure we report in Table 2 the minimum MPE value among the six trials (MPE_s). We did not use a figure of merit to recognize MPE_s among the six trials; we simply identified it a posteriori. The high MPE_s values suggest that PSTs provide rough model structures, a usual result when the data resolution is low.

The first point to assess now is whether the PST phases are refinable via a typical EDM procedure or if the mean phase errors are too high for a successful refinement. If

EDM techniques succeed, VST refinement is no longer necessary. Accordingly, for each test structure we submitted the corresponding six PST phase sets (one set per pivot) to thirty EDM cycles. In the MPE_{EDM} column of Table 2 we show, for each test structure, the minimum MPE values (over the six trials) obtained via EDM techniques. Once again, we did not use a figure of merit for recognizing the solutions with the minimum MPE value, we identified them a posteriori.

For most of the test structures the phase refinement procedure diverged or remained stationary. In practice, none of the test structure could be solved by standard EDM approaches.

To show the failure of the EDM techniques in more detail, for each test structure we selected the pivot (among six) leading to the minimum value of the average phase error (MPE_{EDM} in Table 2). In Figure 1a we show the progress, throughout the 30 EDM refinement cycles, for the selected pivots. The judgement is clear: EDM refinement fails.

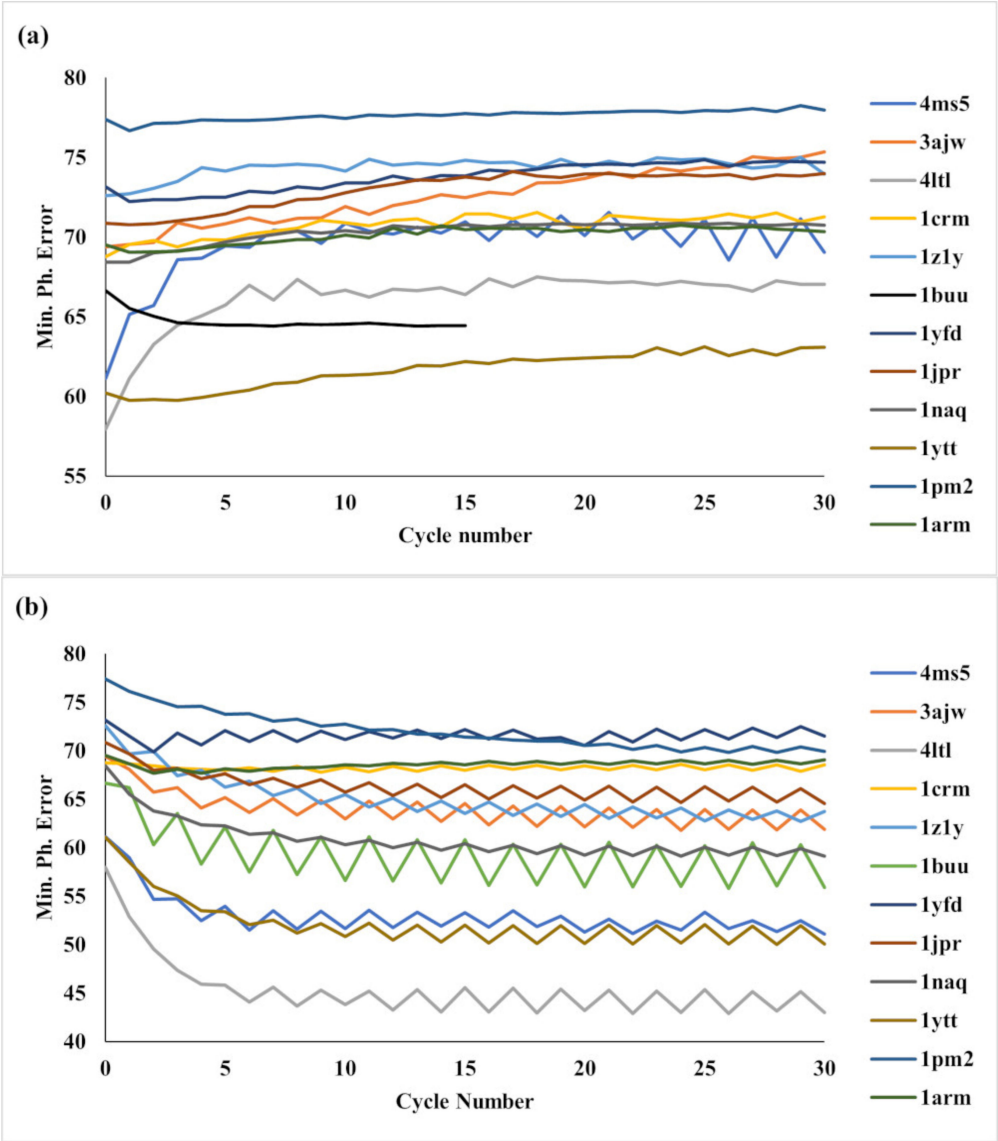

Figure 1. (**a**) Trend of the average phase error throughout the 30 cycles of the EDM refinement. Each line corresponds to the pivot (among six) leading to the minimum value of the average phase error. (**b**) Trend of the average phase error throughout the 30 cycles of the VST refinement. Each line corresponds to the pivot (among six) leading to the minimum value of the average phase error.

We now describe the VST phase refinement approach. To make our algorithm more general we denoted by G a generic Fourier synthesis as described in Section 3 (one of the

F_oF_o, $F_o\overline{F}_p$, $F_p|F_o|^2$ and $|F_p|_2F_o$) and by $\mathbf{V}_\mathrm{H}$ the generic shift vector associated with it (one of those listed in Table 1). After the application of the PST algorithm described in Section 4 a set of $S(\boldsymbol{r})$ maps (six in our procedure) were available, each corresponding to a specific shift vector $\mathbf{V}_H$. The Fourier inversion of each $S(\boldsymbol{r})$ generates a set of phases, which, together with the corresponding $\mathbf{V}_H$ vector, are the input for the application of our VST algorithm. Thirty super-cycles were then applied, each consisting of:

(a) Calculation of the G synthesis;
(b) Calculation of the minimum function $\rho_{min} = Min[P(\boldsymbol{u}+\boldsymbol{r}_H), G(\boldsymbol{r}-\boldsymbol{V}_H)]$;
(c) Calculation of the ρ_{min} map and its Fourier inversion, followed by a few EDM cycles.

We used 30 cycles of the F_oF_o-, $F_o\overline{F}_p$-, $F_p|F_o|^2$- and $|F_p|_2F_o$-syntheses, one at a time. After each VST refinement we selected, for each synthesis, the trial (among six) with the smallest MPE value; as before, it was identified a posteriori. These minimum values are denoted in Table 2 by $\mathrm{MPE}_{F_oF_o}$, $\mathrm{MPE}_{F_o\overline{F}_p}$, $\mathrm{MPE}_{F_p|F_o|^2}$, $\mathrm{MPE}_{|F_p|^2F_o}$, respectively.

For most of the test structures, the VST refinement is able to significantly reduce the MPE_s values, whereas EDM techniques, generally, degraded the quality of the *S*-phases. The VST syntheses may therefore be considered as constraints to which phases obtained by EDM inversion are subject. Such constraints are often able to prevent stationarity or divergence of the phase refinement process. VST refinement leads to different MPE values, depending on the synthesis involved in the procedure. There is no synthesis which systematically provides phases significantly better than the others.

To provide a more detailed illustration of the success of the VST refinement, we show the trend of the average phase error when the $F_o\overline{F}_p$-synthesis is used. For each test structure we selected the pivot (among six) that resulted in the minimum value of the phase error ($\mathrm{MPE}_{F_o\overline{F}_p}$ in Table 2). In Figure 1b, we present the progress according to the cycles of the EDM refinement for the selected pivots. It can be observed that in most cases the phase error continuously decreases with the cycle number. The reader will also notice an oscillatory behavior in the VST phase refinement process. Oscillations are essential for the success of the procedure as they generate new starting points for the EDM cycles. Any attempt to reduce these oscillations would result in a stationary phase refinement process.

Even though the quality of the VST phases is generally higher than that of the corresponding minimum superposition function $S(\boldsymbol{r})$, the corresponding MPE values are, however, too high for defining a model structure of sufficient quality. Therefore, the VST phases were automatically subjected to a further refinement procedure (direct space refinement, DSR). The DSR procedure [15] is based on various criteria and consists of four modules that are repeatedly applied: sets of EDM cycles according to the VLD algorithm [22,23] HAFR (the heaviest atomic species present in the crystal is associated with a number of peaks selected from the current electron density map; their site occupancies are suitably modified), LSQH (the isotropic displacement parameters of the heavy atoms are refined via least-squares), EXALT (the electron density map is periodically enhanced, to avoid a possible weakening of the heavy atom peaks).

To make the procedure suitable for practical cases, each set of DSR phases is submitted to a figure of merit (FFOM) which selects the most promising solution. FFOM is defined as follows [20]:

$$\mathrm{FFOM} = \frac{RAT_{Current}\ CC(all)_{Current}\ CC(large)_{Current}}{RAT_{initial}\ CC(all)_{initial}\ CC(large)_{initial}}$$

CC is the correlation factor between the observed normalized structure factors (R_{obs}) and the calculated normalized structure factor (R_{calc}). We notice that R_{calc} is obtained by inverting a small percentage (3.5%, corresponding to the pixels with highest intensity) of the current electron density map. $CC(all)$ is calculated for all observed reflections, while $CC(large)$ is calculated over the subset of 70% of the largest $|F_{obs}|$ values.

In $RAT = CC_w / < R^2_{calc} >_{weak}$, CC_w is the correlation coefficient between the largest R_{obs} amplitudes (about 70% of the total) and the corresponding σ_A weights, and the average

$< R^2_{calc} >_{weak}$ is calculated over 30% of the measured reflections (those with the weakest |Fobs| values). Basically, the FFOM selects the model for which the numerator at the right-hand side attains its largest value with respect to its initial value, while the denominator is essentially a normalizing constant.

Based on our experience, an FFOM value larger than 2.9 typically indicates a good set of phases. This result allowed us to modify the software into a program capable of solving unknown crystal structures. The following procedure is adopted: if an FFOM larger than 2.9 is found, the procedure does not check any other pivot and jumps from the refinement process to the AMB and graphical steps (see Section 7). Otherwise, all six trials are explored, but now only the RAT component is used to identify the best solution.

The average phase errors of the trials selected by our figures of merit are reported in Table 3 for each Fourier synthesis, and the corresponding FFOM (or RAT) are in parentheses.

Table 3. For each test structure the average phase errors obtained at the end of the DSR phase refinement for the syntheses F_oF_o, $F_o\overline{F}_p$, $F_p|F_o|^2$ and $|F_p|_2F_o$ are shown. The corresponding trials are selected by our figures of merit (FFOM) or RAT (in parentheses).

PDB	$\mathbf{DSR}_{F_oF_o}$	$\mathbf{DSR}_{F_o\overline{F}_p}$	$\mathbf{DSR}_{F_p\lvert F_o\rvert^2}$	$\mathbf{DSR}_{\lvert F_p\rvert^2F_o}$
4ms5	50 (2.93)	49 (2.92)	48 (2.98)	49 (2.99)
3ajw	49 (3.08)	49 (3.11)	49 (3.23)	49 (3.20)
4ltl	42 (3.01)	42 (3.16)	42 (3.21)	41 (3.17)
1crm	71 (2.29)	69 (2.31)	89 (2.15)	70 (2.23)
1z1y	69 (2.28)	68 (2.43)	76 (2.29)	89 (2.24)
1buu	31 (4.85)	32 (3.95)	32 (3.78)	30 (4.87)
1yfd	52 (3.12)	52 (3.01)	89 (2.93)	50 (3.03)
1jpr	47 (2.99)	47 (3.13)	51 (2.92)	46 (3.22)
1naq	51 (2.26)	51 (2.10)	51 (2.14)	51 (2.13)
1ytt	43 (3.08)	43 (2.91)	43 (3.08)	43 (3.09)
1pm2	48 (2.93)	48 (3.01)	47 (2.91)	48 (2.96)
1arm	88 (2.26)	58 (2.49)	89 (2.42)	88 (2.32)

We observe:

(i) The quality of the DSR phases is higher than that of the VST phases.
(ii) For most of the test structures good sets of phases are obtained and FFOM (or RAT) is able to recognize them. That is particularly true for the three structures with RES > 2.05.
(iii) The final average phase errors for 1crm and 1z1y remain high for all the syntheses. We do not know why these test structures resist the phasing procedure.
(iv) None of the four columns in Table 2 provides a set of phases systematically better than the others; nevertheless, a minor effectiveness may be associated with the synthesis $F_p|F_o|^2$ and a major effectiveness with the synthesis $F_o\overline{F}_p$.

The DSR phases so obtained were subjected to an additional refinement procedure based on the shifted difference Patterson synthesis $\Delta P(\boldsymbol{u} - \boldsymbol{r}_H)$. Positive peaks were expected to introduce a new electron density into the model structure, while negative peaks were expected to eliminate incorrect density from the model. However, no significant improvement was observed, and therefore we consider the DSR phase refinement to be the conclusive step of our phasing procedure.

6. The CAB Application

To fully automatize the ab initio crystal structure solution via VSTs, a further step is necessary: submitting the DSR phases to an appropriate AMB program. The DSR step (see Section 5) was able to substantially reduce the average phase error of 10 over 12 test structures up to a level suitable for succeeding in the automatic model building process (1crm and 1z1y are excluded). The AMB procedure of our pipeline is CAB [24,25], an AMB approach which cyclically applies BUCCANEER [26] and NAUTILUS [27] algorithms,

respectively. The high quality of CAB has been shown in a recent paper [28], but, unfortunately, it is a program specifically written for molecular replacement studies. The standard CAB needs a molecular model for the MR step, and also needs to know how many models should be located in the asymmetric unit of the target. Therefore, heavy modifications are necessary for optimizing CAB when phases are obtained by VSTs. Despite the above limitations, we submitted to CAB the phases obtained by the $F_o\overline{F}_p$ refinement (see Table 3) as selected by the FFOM (or RAT) criterion; we excluded 1crm and 1z1y from the tests because of their large average phase value. For each model structure CAB calculated the MA parameter (MA is the percentage of the atoms that lie within 0.6 Å from the published positions).

In five of the ten suitable cases (3ajw, 1buu, 1jpr, 1ytt, 1pm2) CAB provides MA values larger than 0.6. For them, Table 4 displays the pivot order (PIV) corresponding to the selected solution, the average phase error of the CAB structural model (MPE_{CAB}) and the corresponding MA value. The MA values are often sufficiently high to suggest that the CAB model is substantially correct.

Table 4. For five test structures for which MA > 0.60 we show the pivot order (PIV) corresponding to the selected solution, the average phase error of the CAB structural model (MPE_{CAB}) and the corresponding MA value.

PDB	PIV	MPE_{CAB}	MA
3ajw	1	36	0.63
1buu	1	30	0.87
1jpr	1	35	0.89
1ytt	1	55	0.60
1pm2	1	33	0.89

The CAB coordinates were processed using the graphical software JAV [29], providing an instant visualization of the structural model. JAV is capable of reading coordinates in different formats such as pdb, ent, cif, ins and res. In Figures 2–4 we show for 1ytt [30] (the case with the smallest MA value), 1pm2 [31] and 1jpr [32] the CAB backbones in red and the published backbones in blue. The figures demonstrate that the CAB models have good chain overlap with the published structures.

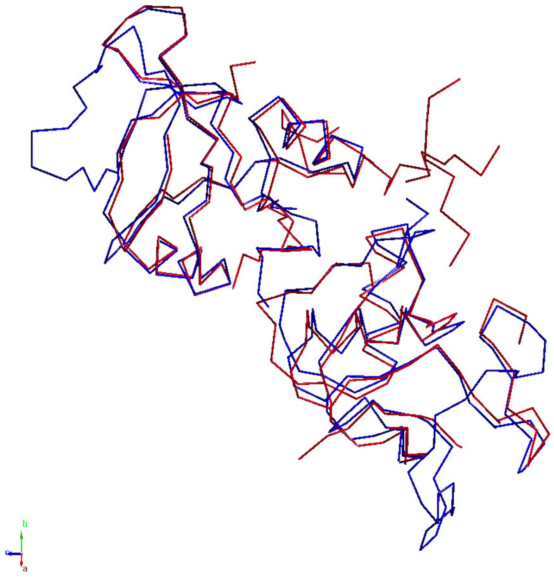

Figure 2. 1ytt: CAB chain-trace in red, published chain-trace in blue.

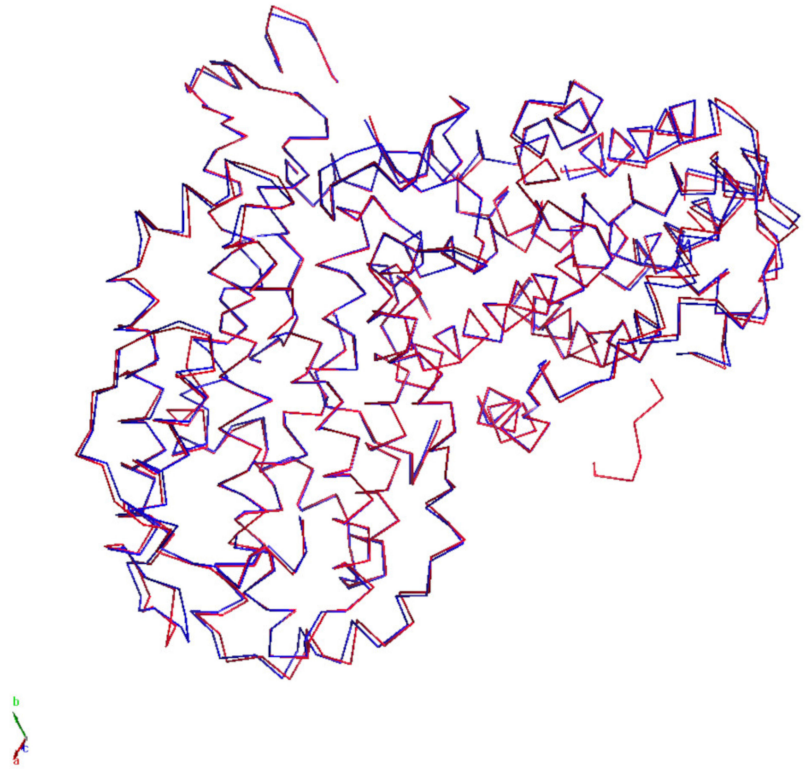

Figure 3. 1pm2: CAB chain-trace in red, published chain-trace in blue.

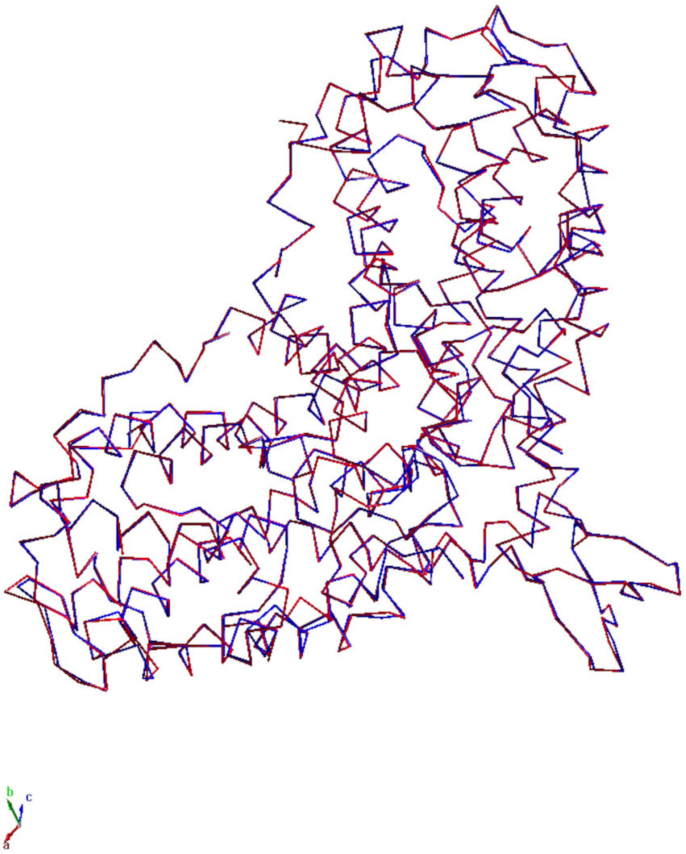

Figure 4. 1jpr: CAB chain-trace in red, published chain-trace in blue.

7. Conclusions

When the data resolution is low, Patterson superposition techniques are unable to solve macromolecular structures. They usually provide sets of phases that are too rough to be refined by standard EDM techniques; the refinement procedure is stationary or divergent. This paper extends the concept of Patterson superposition techniques, by including them in the larger concept of vector superposition techniques. In fact, we showed that the Patterson map is only one of a variety of Fourier syntheses to which superposition techniques can be applied. In this context we described the properties of the F_oF_o-, $F_o\overline{F}_p$-, $F_p|F_o|^2$- and $|F_p|_2F_o$-syntheses; the usefulness of the latter three has always been overlooked and, as a result, they have never been applied in a phase refinement procedure. However, our theoretical analysis of their properties indicates that they can offer images of the structure that are accurate enough to serve as a valuable starting point for phase refinement.

In this paper, VSTs are used as a tool for refining PST phases; they may be considered as a constraint hindering EDM techniques from being stationary or diverging. The VST phases are a better starting point for a further phase refinement procedure, performed by our DSR program, which adopts an atypic EDM approach which leads to a solution for most of the test structures. The two failure cases (1crm and 1z1y) can be attributed more to the limited effectiveness of the refinement procedure rather than to the limited utility of the new Fourier syntheses. The problem will be the object of a new investigation.

The procedure described above shows that 2.2 Å resolution is an accessible threshold for ab initio crystal structure solutions of macromolecular structures, provided heavy atoms are in the unit cell. The new resolution threshold is an important factor as a non-negligible percentage of macromolecular structures currently exhibit a resolution better than 2.2 Å. Furthermore, the use of ligands containing heavy atoms is a current technique for improving the crystallization process, especially for nucleic acids.

Author Contributions: Conceptualization, C.G.; Methodology, M.C.B., C.G. and G.P.; Software, M.C.B., B.C., G.L.C. and G.P.; Validation, M.C.B., B.C., G.L.C. and G.P.; Formal Analysis, M.C.B., B.C., G.L.C. and G.P.; Writing—Original Draft Preparation, C.G.; Writing—Review and Editing, M.C.B., B.C., G.L.C., C.G. and G.P.; Visualization, M.C.B., B.C., G.L.C. and G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are present within the article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

$\boldsymbol{F} = \|\boldsymbol{F}\|\mathbf{exp}(i\boldsymbol{\varphi})$	$\|F\|$ is the observed structure factor of the target structure; φ is its phase value.
$\boldsymbol{F_p} = \|\boldsymbol{F_p}\|\mathbf{exp}(i\boldsymbol{\varphi_p})$	structure factor of the current structure model.
$\Delta\boldsymbol{\varphi} = \boldsymbol{\varphi_p} - \boldsymbol{\varphi}$	
$\boldsymbol{\rho}(\boldsymbol{r})$	ideal electron density map, calculated via observed amplitudes $\|\boldsymbol{F}\|$ and true phases $\boldsymbol{\varphi}$.
$\boldsymbol{\rho_p}(\boldsymbol{r})$	electron density map of the current structure model, calculated by using $\|\boldsymbol{F_p}\|$ amplitudes $\|\boldsymbol{F}\|$ and estimated phases $\boldsymbol{\varphi_p}$. Usually, it contains heavy atoms.
$\boldsymbol{C_s}$, **for** $\boldsymbol{s} = 1, \ldots, \boldsymbol{m}$	symmetry operators of the space group.
N	number of atoms in the target unit cell.

N_p	number of atoms in the model unit cell.
N_L	number of atoms in the $\boldsymbol{\rho}(\boldsymbol{r}) - \boldsymbol{\rho_p}(\boldsymbol{r})$ substructure.
Z_i	atomic number of the ith atom of the target structure.
Z_{pi}	atomic number of the ith atom of the model structure.
$\boldsymbol{r_i}, \boldsymbol{r_{pi}}, \boldsymbol{r_{Hi}}$	positional vector of the ith atom of the target structure, of the ith atom of the model structure and of the ith heavy atom, respectively. Frequently, a heavy atom is part of the model structure.
$\overline{\boldsymbol{F}}, \overline{\boldsymbol{F}}_{\boldsymbol{p}}$	complex conjugates of F and F_p.
$\boldsymbol{E} = \boldsymbol{R}\mathbf{exp}(\boldsymbol{i\varphi})$, $\boldsymbol{E_p} = \boldsymbol{R_p}\mathbf{exp}(\boldsymbol{i\varphi_p})$ $\boldsymbol{\Sigma_N} = \sum_{j=1}^{N} \boldsymbol{f_j^2}$ $\boldsymbol{\Sigma_p} = \sum_{j=1}^{p} \boldsymbol{f_j^2}$	normalized structure factors of F, F_p, respectively.
$\boldsymbol{D} = \langle \mathbf{cos}(2\boldsymbol{\pi h \Delta r}) \rangle$	the average is found per resolution shell.
$\sigma_A = D\sqrt{\Sigma_p/\Sigma_N}$	
$I_i(X)$	modified Bessel function of order i.
$m = \langle \cos(\varphi - \varphi_p) \rangle$	$I_1(X)/I_0(X)$, where $X = 2\sigma_A R R_p/(1-\sigma_A^2)$.
asu	asymmetric unit.
EDM	electron density modification.
PST	Patterson superposition technique.
VST	vector superposition technique.
*	convolution symbol.
Paper I [17]	https://doi.org/10.1107/S0021889806017894.
Paper II [16]	https://doi.org/10.3390/cryst10060538.

References

1. Wrinch, D.M. X. The Geometry of Discrete Vector Maps. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1939**, *27*, 98–122. [CrossRef]
2. Beevers, C.A.; Robertson, J.H. Interpretation of the Patterson Synthesis. *Acta Cryst.* **1950**, *3*, 164. [CrossRef]
3. Buerger, M.J. Phase Determination with the Aid of Implication Theory. *Phys. Rev.* **1948**, *73*, 927–928. [CrossRef]
4. Hauptman, H. A New Method in the Probabilistic Theory of the Structure Invariants. *Acta Crystallogr. A Cryst. Phys. Diffr. Theor. Gen. Crystallogr.* **1975**, *31*, 680–687. [CrossRef]
5. Giacovazzo, C. *Direct Methods in Crystallography*; Academic Press: London, UK; New York, NY, USA, 1980; ISBN 978-0-12-282450-0.
6. Giacovazzo, C. The Method of Representations of Structure Seminvariants. II. New Theoretical and Practical Aspects. *Acta Crystalogr. A* **1980**, *36*, 362–372. [CrossRef]
7. Weeks, C.M.; DeTitta, G.T.; Hauptman, H.A.; Thuman, P.; Miller, R. Structure Solution by Minimal-Function Phase Refinement and Fourier Filtering. II. Implementation and Applications. *Acta Crystallogr. A* **1994**, *50*, 210–220. [CrossRef]
8. Sheldrick, G.M. SHELX Applications to Macromolecules. In *Direct Methods for Solving Macromolecular Structures*; Fortier, S., Ed.; Springer: Dordrecht, The Netherlands, 1998; pp. 401–411, ISBN 978-94-015-9093-8.
9. Foadi, J.; Woolfson, M.M.; Dodson, E.J.; Wilson, K.S.; Jia-xing, Y.; Chao-de, Z. A Flexible and Efficient Procedure for the Solution and Phase Refinement of Protein Structures. *Acta Crystallogr. D* **2000**, *56*, 1137–1147. [CrossRef]
10. Burla, M.C.; Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C.; Polidori, G. More Power for Direct Methods: *SIR2002*. *Z. Krist. Cryst. Mater.* **2002**, *217*, 629–635. [CrossRef]
11. Nordman, C.E. Vector Space Search and Refinement Procedures. *Trans. Am. Crystallogr. Assoc.* **1966**, *2*, 29–38.
12. Richardson, J.W.; Jacobson, R.A. *Patterson and Pattersons: Fifty Years of the Patterson Function: Proceedings of a Symposium Held at the Institute for Cancer Research, the Fox Chase Cancer Center, Philadelphia, PA, USA, 13–15 November 1984*; Glusker, J.P., Patterson, B.K., Rossi, M., Eds.; International Union of Crystallography Crystallographic Symposia; International Union of Crystallography; Oxford University Press: New York, NY, USA, 1987; ISBN 978-0-19-855230-7.
13. Sheldrick, G.M. Tutorial on Automated Patterson Interpretation to Find Heavy Atoms. In *Crystallographic Computing 5: From Chemistry to Biology*; Moras, D., Podjarny, A.D., Thierry, C., Eds.; IUCr Crystallographic Symposia; International Union of Crystallography; Oxford University Press: New York, NY, USA, 1991; pp. 145–157, ISBN 978-0-19-855384-7.
14. Pavelčík, F.; Kuchta, L.; Sivý, J. Patterson-Oriented Automatic Structure Determination. Utilizing Patterson Peaks. *Acta Crystallogr. A* **1992**, *48*, 791–796. [CrossRef]
15. Burla, M.C.; Caliandro, R.; Carrozzini, B.; Cascarano, G.L.; De Caro, L.; Giacovazzo, C.; Polidori, G. *Ab Initio* Protein Phasing: The Patterson Deconvolution Method in *SIR2002*. *J. Appl. Crystallogr.* **2004**, *37*, 258–264. [CrossRef]
16. Burla, M.C.; Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C.; Polidori, G. Properties of Fourier Syntheses and New Syntheses. *Crystals* **2020**, *10*, 538. [CrossRef]
17. Burla, M.C.; Caliandro, R.; Carrozzini, B.; Cascarano, G.L.; De Caro, L.; Giacovazzo, C.; Polidori, G.; Siliqi, D. The Revenge of the Patterson Methods. I. Protein *Ab Initio* Phasing. *J. Appl. Crystallogr.* **2006**, *39*, 527–535. [CrossRef]

18. Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C. The Cross-Correlation Function: Main Properties and First Applications. *J. Appl. Crystallogr.* **2010**, *43*, 221–226. [CrossRef]
19. Caliandro, R.; Carrozzini, B.; Cascarano, G.L.; Comunale, G.; Giacovazzo, C. On the Use of the *C* Map in Patterson Deconvolution Procedures. *Acta Crystallogr. A Found. Crystallogr.* **2013**, *69*, 98–107. [CrossRef]
20. Caliandro, R.; Carrozzini, B.; Cascarano, G.L.; Comunale, G.; Giacovazzo, C.; Mazzone, A. Protein Phasing at Non-Atomic Resolution by Combining Patterson and *VLD* Techniques. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 1994–2006. [CrossRef]
21. Cowtan, K. Error Estimation and Bias Correction in Phase-Improvement Calculations. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55*, 1555–1567. [CrossRef]
22. Burla, M.C.; Giacovazzo, C.; Polidori, G. From a Random to the Correct Structure: The VLD Algorithm. *J. Appl. Crystallogr.* **2010**, *43*, 825–836. [CrossRef]
23. Burla, M.C.; Caliandro, R.; Giacovazzo, C.; Polidori, G. The Difference Electron Density: A Probabilistic Reformulation. *Acta Crystallogr. A Found. Crystallogr.* **2010**, *66*, 347–361. [CrossRef]
24. Burla, M.C.; Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C.; Polidori, G. Cyclic Automated Model Building (CAB) Applied to Nucleic Acids. *Crystals* **2020**, *10*, 280. [CrossRef]
25. Cascarano, G.L.; Giacovazzo, C. Towards the Automatic Crystal Structure Solution of Nucleic Acids: Automated Model Building Using the New *CAB* Program. *Acta Crystallogr. D Struct. Biol.* **2021**, *77*, 1602–1613. [CrossRef] [PubMed]
26. Cowtan, K. The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains. *Acta Crystallogr. D* **2006**, *62*, 1002–1011. [CrossRef] [PubMed]
27. Cowtan, K. Automated Nucleic Acid Chain Tracing in Real Time. *IUCrJ* **2014**, *1*, 387–392. [CrossRef] [PubMed]
28. Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C. The Automatic Solution of Macromolecular Crystal Structures via Molecular Replacement Techniques: REMO22 and Its Pipeline. *Int. J. Mol. Sci.* **2023**, *24*, 6070. [CrossRef]
29. Cascarano, G.L.; Cuocci, C.; Mallamo, M.; Carrozzini, B.; Moliterni, A. JAV (Just Another Viewer)—Unpublished Software. 2021.
30. Burling, F.T.; Weis, W.I.; Flaherty, K.M.; Brünger, A.T. Direct Observation of Protein Solvation and Discrete Disorder with Experimental Crystallographic Phases. *Science* **1996**, *271*, 72–77. [CrossRef]
31. Voegtli, W.C.; Sommerhalter, M.; Saleh, L.; Baldwin, J.; Bollinger, J.M.; Rosenzweig, A.C. Variable Coordination Geometries at the Diiron(II) Active Site of Ribonucleotide Reductase R2. *J. Am. Chem. Soc.* **2003**, *125*, 15822–15830. [CrossRef]
32. Högbom, M.; Andersson, M.E.; Nordlund, P. Crystal Structures of Oxidized Dinuclear Manganese Centres in Mn-Substituted Class I Ribonucleotide Reductase from Escherichia Coli: Carboxylate Shifts with Implications for O2 Activation and Radical Generation. *J. Biol. Inorg. Chem.* **2001**, *6*, 315–323. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.