

The Use of Connected Masks for Reconstructing the Single Particle Image from X-Ray Diffraction Data. III. Maximum-Likelihood Based Strategies to Select Solution of the Phase Problem

Lunina N.L.¹, Petrova T.E.¹, Urzhumtsev A.G.^{*2,3}, Lunin V.Y.^{**1}

¹*Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Moscow Region, 142290 Russia*

²*Centre for Integrative Biology, IGBMC, CNRS–INSERM–UdS, 1 rue Laurent Fries, BP 10142, Illkirch, 67404, France*

³*Département de Physique, Faculté des Sciences et des Technologies, Université de Lorraine, 54506 Vandoeuvre-lès-Nancy, France*

Abstract. The main experimental limitation of biological crystallography is associated with the need to prepare the object under study in the form of a single crystal. New powerful X-ray sources, namely free-electron X-ray lasers, makes it possible to raise the question of the determination of the structure of isolated biological macromolecules and their complexes in practice. An additional advantage of working with isolated particles is the possibility to obtain information about scattering in all directions, and not only in those limited by the Laue-Bragg diffraction conditions. This significantly facilitates the solution of the phase problem of X-ray diffraction analysis. This paper is devoted to two lines of development of the method for solving the phase problem, proposed earlier by the authors, which is based on the random scanning of the configuration space of potential solutions of the phase problem. The paper suggests a new criterion for the selection of "candidates" for solving the phase problem in the process of scanning. It involves the maximization of statistical likelihood, and its effectiveness is shown in test calculations. The second line concerns the choice of the optimal scanning strategy. It is shown that the gradual expansion of the set of experimental data used in the work allows obtaining solutions of a higher quality than those obtained with all available data included into the work simultaneously from the beginning.

Key words: X-ray crystallography, the phase problem, XFEL, single particle diffraction.

1. THE SOLUTION OF THE PHASE PROBLEM FOR ISOLATED PARTICLES BY SCANNING THE CONFIGURATION SPACE OF PHASE SETS

Putting the new generation X-ray sources, X-ray free-electron lasers (XFEL), into operation allows one to pose new goals in studies of the structure of biological objects [1–7]. One of these tasks is the development of approaches to the study of isolated biological particles (Single Particles) by X-ray diffraction [8–16]. These approaches make it possible to

*sacha@igbmc.fr

**lunin@impb.ru

remove the main limitation of X-ray crystallography, namely, the necessity of preparing the sample in the form of a single crystal. Another important feature of the work with isolated particles is the possibility to measure the intensity of scattered waves for all possible directions of scattering. Note that, in experiments with crystals, the measurement of the intensity of scattered waves is limited by a discrete set of directions determined by the parameters of the crystal lattice.

As in traditional X-ray crystallography, an X-ray experiment with an isolated particle allows one to obtain only the values of the modulus $\{F(\mathbf{s})\}_{\mathbf{s} \in R^3}$ of the complex Fourier transform of the function $\rho(\mathbf{r})$ that describes the distribution in the three-dimensional space of the electrons of the object under study. The value of the phase part $\{\varphi(\mathbf{s})\}_{\mathbf{s} \in R^3}$ is lost in the measurement process, which prevents the direct retrieval of the electron density distribution by calculating the inverse Fourier transform:

$$\rho(\mathbf{r}) = \int_{R^3} F(\mathbf{s}) \exp[i\varphi(\mathbf{s})] \exp[-2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{s}}, \quad (1)$$

where

$$F(\mathbf{s}) \exp[i\varphi(\mathbf{s})] = \int_{R^3} \rho(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{r}}. \quad (2)$$

This leads to the central problem of diffraction methods, namely the phase problem, i.e., to the problem of restoring the phase part of the Fourier transform. A similar problem, the recovery of the function from the moduli of its Fourier transform, is also encountered in optics [17] and is called the phase retrieval problem. Until recently, the development of methods for solving this problem in crystallography and optics was largely independent, and similar approaches in these areas had different names. Here, we will adhere to the crystallographic terminology.

The scheme of an X-ray experiment is shown in Figure 1. In the theoretical analysis, the direction of the X-ray beam and the direction "from the object to the detector" (given by vectors of unit length σ_0, σ) can be considered as variable parameters. In practice, however, the direction σ_0 of the initial beam is fixed, and additional variation of the conditions is accomplished by rotating the object under investigation around the original beam. Within the framework of the kinematic theory of scattering, the intensity of the scattered wave proves to be proportional to the square of the modulus of the Fourier transform (2) calculated at the point \mathbf{s} , which is defined as a combination of vectors σ_0, σ and the wavelength λ of X-rays:

$$\mathbf{s} = \frac{\sigma - \sigma_0}{\lambda}. \quad (3)$$

This vector is called the scattering vector. It should be noted that, theoretically, the experiment makes it possible to measure the values of the function $\{F(\mathbf{s})\}_{\mathbf{s} \in R^3}$ not for all values of the scattering vector \mathbf{s} , but only for a limited range of its values: $|\mathbf{s}| \leq 2/\lambda$. In practice, this zone is further limited by $|\mathbf{s}| \leq s_{\max}$. The value $d_{\min} = 1/s_{\max}$ is called the resolution of the collected data set and is determined by the characteristics of the initial beam, the detector, and the experimental conditions. The approximate distribution of the electron density, calculated from formula (1) with the use of the restricted data set, is called the Fourier synthesis of the electron density of the resolution d_{\min} . It can contain distortions, such as the series of "termination waves" and "blurring" of density peaks. The resolution of the Fourier synthesis determines the size of the minimum details that can be visually discerned when analyzing the synthesis [18].

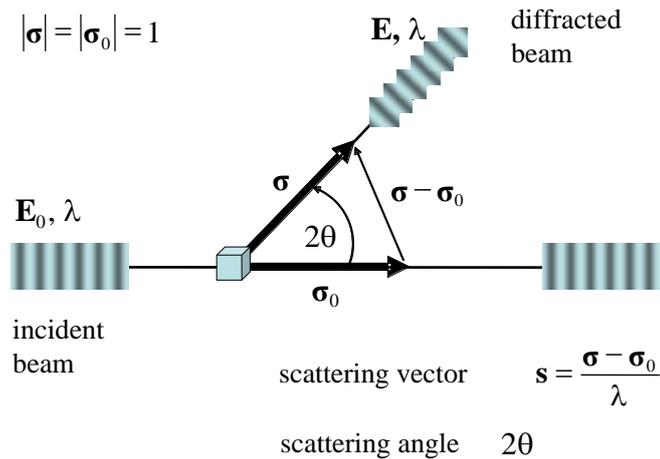


Fig. 1. Scheme of X-ray diffraction experiment.

The sampling of experimental data on a uniform grid with $1/a$ step (where the size a exceeds the linear dimensions of the particle under study) allows us to consider them as a set of the magnitudes of the structure factors (Fourier coefficients) for an imaginary crystal (with the length a of all unit cell sides) and apply the rich toolbox of biological crystallography for solving the phase problem [19, 20]. In more detail, let V be an elementary cell (cube with a side) chosen so that the particle under investigation lies inside it, and $|V|$ is the volume of the unit cell. Let $\rho(\mathbf{r})$ be the function that describes the distribution of the electron density inside the particle under study and is equal to zero outside the particle. Let $\rho^{\text{cryst}}(\mathbf{r})$ be a periodic function which has periods a along the axes of the unit cell and which coincides with $\rho(\mathbf{r})$ inside V . In this case, the function $\rho^{\text{cryst}}(\mathbf{r})$ can be calculated as the sum of the Fourier series

$$\rho^{\text{cryst}}(\mathbf{r}) = \frac{1}{|V|} \sum_{\mathbf{s} \in \mathfrak{R}'} F(\mathbf{s}) \exp[i\phi(\mathbf{s})] \exp[-2\pi i(\mathbf{s}, \mathbf{r})], \quad (4)$$

where

$$F(\mathbf{s}) \exp[i\phi(\mathbf{s})] = \int_V \rho(\mathbf{r}) \exp[2\pi i(\mathbf{s}, \mathbf{r})] dV_{\mathbf{r}}, \quad (5)$$

are the structure factors, and \mathfrak{R}' is the $1/a$ – spaced grid. The nodes of this lattice are called reflexes. The difference from the case with an ordinary crystal is that, by choosing a sufficiently large value a , we obtain the function $\rho^{\text{cryst}}(\mathbf{r})$ equal to zero almost everywhere in the unit cell. This creates a significant redundancy of the experimental data and forms the basis for solving the phase problem [19, 21–26].

The main approach to the practical use of data redundancy created by the presence in the unit cell of a large region of zero electron density values is the implementation of a wide class of iterative methods [27–33]. If the parameters of the methods and the starting point are selected successfully, these methods can lead to fairly accurate solutions. However, these methods work locally, and their stability and convergence to a correct solution may require the choice of parameters that is not obvious. A supplement to the local refinement procedures can be the use of global procedures for scanning the configuration space of possible solutions, which help in choosing the initial approximation. In this paper, we investigate one such procedure for the random scanning of a configuration space, which was originally proposed in [34] and developed further in [13–15, 35–37, 20]. The procedure is shown in Figure 2.

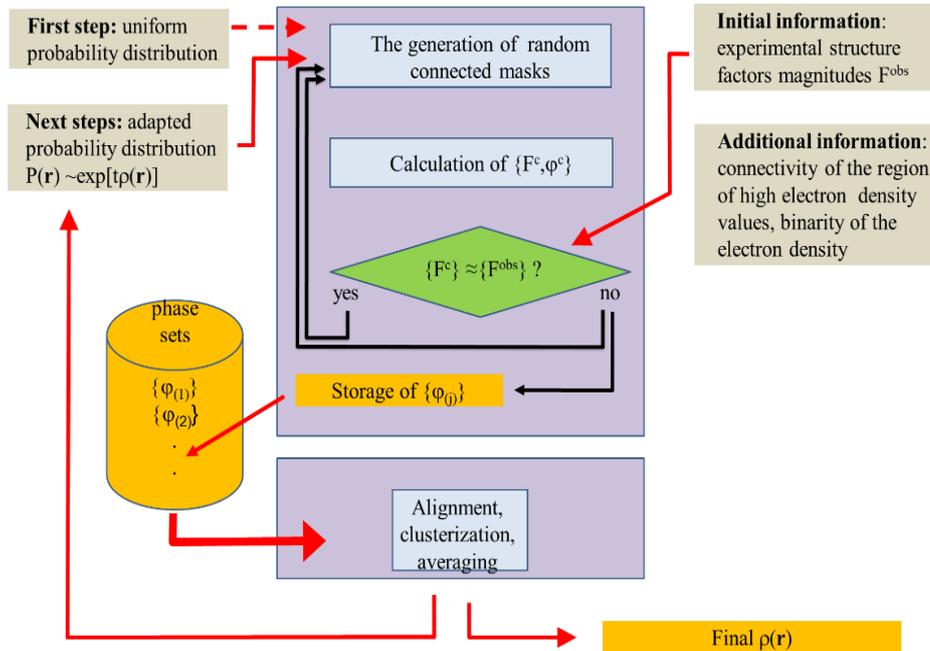


Fig. 2. Scheme of the procedure for random scanning of a configuration space of phase sets.

The main cycle consists in the generation of a large number of random phase sets, the selection of those that lead to reasonable electron density distributions, and the alignment and averaging of the selected variants. At the same time, the choice of the criterion of "suitability" is crucial for the implementation of the approach. The generation of random variants can take place without using additional information on phase values (in the first cycle of the procedure) or with the use of any information on the probability of phase values obtained in previous cycles. When generating random phase sets, the phase values can be generated independently or calculated using models that reflect the expected properties of real electron density distributions in biological objects. In this case, some preliminary models are accidentally generated from which the phases of structural factors are further calculated. In this paper, we consider the procedure in which the generated objects are connected binary masks of the region occupied by the object [20–37]. These masks $M(\mathbf{r})$ are binary functions (with values of 0 or 1) on the grid in the unit cell, which have the connectivity property of the mask region, i.e., of the set of points in which the value is equal to 1. Such masks reflect two important properties of real electron density distributions in biological macromolecules. First, the presence of extended polypeptide chains or DNA chains leads to a connectivity of the region of high electron density values or, at least, to the presence of a small number of connected components [38]. Binarity underlies the interpretation of electron density distributions when points of space are divided into those belonging and not belonging to the object. Earlier it was shown that both of these properties allow the selection of acceptable solutions separately and together. Especially promising was the use of these properties at the level of model generation.

Once generated, the mask $M(\mathbf{r})$ can be used to calculate the sets of structure factor magnitudes $\{F_s^{\text{mask}}\}$ and the values of phases $\{\varphi_s^{\text{mask}}\}$ corresponding to the mask structure factors in accordance with (5). The degree of similarity of the calculated and experimental structure factors magnitudes can serve as an indicator of the quality of the mask and, accordingly, of the accuracy of phases corresponding to the mask, when the exact phase values in a real study are unknown. In previous works [20, 37], the correlation coefficient of

these values was used as a criterion for the correspondence of two sets of structure factor magnitudes

$$CM = \frac{\sum_{\mathbf{s}} F_{\mathbf{s}}^{\text{mask}} F_{\mathbf{s}}^{\text{obs}}}{\sqrt{\sum_{\mathbf{s}} (F_{\mathbf{s}}^{\text{mask}})^2 \sum_{\mathbf{s}} (F_{\mathbf{s}}^{\text{obs}})^2}}. \quad (6)$$

In this paper, we investigate the possibility of using a new type of the selection criterion, namely, the value of the mathematical expectation of phase correlation when the generated mask is used as a prior distribution for the coordinates of an atomic model of the object under study.

2. MASK QUALITY INDICATORS BASED ON THE MAXIMIZATION OF THE LIKELYHOOD

The use of statistical likelihood as a criterion for selecting solutions was discussed earlier in [36], where a laborious simulation procedure was used to calculate the likelihood. In this paper, a significantly less time-consuming procedure is proposed, which is based on the use of the diagonal approximation of the likelihood function [39–42].

Suppose that an atomic model is randomly generated inside a mask region. For the particular mask and scattering vector \mathbf{s} , the joint probability distribution for the modulus and phase of the complex structure factor \mathbf{F} calculated from the atomic model, in the framework of the approximation of the central limit theorem of theory of probabilities, can be presented in the form

$$P(F, \varphi) = \frac{F}{\pi\beta} \exp\left[-\frac{F^2 + (\alpha F^{\text{mask}})^2}{\beta}\right] \exp\left[2\frac{\alpha FF^{\text{mask}}}{\beta} \cos(\varphi - \varphi^{\text{mask}})\right]. \quad (7)$$

Here, α and β are the parameters specific for each reflection, which accumulate uncertainties in the choice of the model. For some simple cases of the uncertainty, the parameters α and β can be written out explicitly [41], while in the general case they are to be determined. In many cases, these parameters vary little within groups of reflections $\mathbf{S}_j = \{\mathbf{s} : s_j \leq s \leq s_j + \Delta\}$ where Δ is small ("shells" in the reciprocal space) and, therefore, can be considered to be constant in the shells. With the availability of experimental data $\{F_{\mathbf{s}}^{\text{obs}}\}$, the values of the parameters α and β for the shell \mathbf{S}_j can be determined from the maximum likelihood principle by maximizing the function

$$L(\alpha, \beta) = \prod_{\mathbf{s} \in \mathbf{S}_j} P_F(F_{\mathbf{s}}^{\text{obs}}; \alpha, \beta) \Rightarrow \max. \quad (8)$$

Here, $P_F(F; \alpha, \beta)$ is the distribution of the structure factor magnitude that can be obtained by integrating (7) over the phase

$$P_F(F; \alpha, \beta) = \frac{2F}{\beta} \exp\left[-\frac{F^2 + (\alpha F^{\text{mask}})^2}{\beta}\right] I_0\left(2\frac{FF^{\text{mask}}}{\beta}\right), \quad (9)$$

and distributions corresponding to different reflections \mathbf{s} are supposed to be independent ("the diagonal approximation" of the likelihood function [39–42]), I_0 is the modified Bessel function of the zero order.

After the optimal values of the parameters $\hat{\alpha}_j$ and $\hat{\beta}_j$ have been determined for all shells, it is possible to calculate the likelihood value taking into account all the reflexes and use it (or, rather, its logarithm) as an indicator of the quality of the mask. However, it seems more convenient to use as a selection criterion the derivative characteristic, namely, the mathematical expectation of the coefficient of phase correlation. The distribution (7) makes it possible to write out the conditional distribution of phase values, provided that the value of the modulus is equal to the value observed in the experiment

$$P_{\varphi}(\varphi|F = F^{\text{obs}}) = \frac{1}{2\pi I_0 \left[2(\hat{\alpha}/\hat{\beta}) F^{\text{obs}} F^{\text{mask}} \right]} \exp \left[2 \frac{\hat{\alpha} F^{\text{obs}} F^{\text{mask}}}{\hat{\beta}} \cos(\varphi - \varphi^{\text{mask}}) \right]. \quad (10)$$

The last distribution allows the calculation of

$$c_s = \left\langle \cos(\varphi_s - \varphi_s^{\text{mask}}) \middle| F_s = F_s^{\text{obs}} \right\rangle = \frac{I_1(T_s)}{I_0(T_s)}, \quad (11)$$

$$T_s = 2 \frac{\hat{\alpha}_s F_s^{\text{obs}} F_s^{\text{mask}}}{\hat{\beta}_s} \quad (12)$$

and the conditional expected value of the phase correlation coefficient

$$E_{CP} = \frac{\sum_s (F_s^{\text{obs}})^2 c_s}{\sum_s (F_s^{\text{obs}})^2}. \quad (13)$$

This value can be used as a criterion for the selection of acceptable masks. In this paper, we will discuss the possibility of solving the phase problem with this criterion with the use of different protocols for scanning the configurational space.

The results of testing several work protocols are given below. The first problem being investigated was the choice of the strategy for the inclusion of the data into the work. In the biological crystallography, by convention, the data are included into the work progressively, with a gradual expansion of the data set used. An alternative is the simultaneous inclusion in the work of all available data.

The second problem is related to the construction of the prior probability distribution for the mask generation. Following [20], this probability distribution is built in the form:

$$p(\mathbf{r}) = C \exp \left[\frac{\ln t}{\rho_{\max} - \rho_{\min}} \rho(\mathbf{r}) \right], \quad (14)$$

where C is the normalizing factor, $\rho(\mathbf{r})$ is an approximate electron density distribution obtained at the previous step, t is the parameter that regulates the contrast of the probability distribution $p(\mathbf{r})$:

$$t = \rho_{\max} / \rho_{\min}. \quad (15)$$

We compared two simplest ways of building such distribution which are based on the results of the selection of acceptable masks from randomly generated masks. Let $\{\varphi_1^{\text{mask}}(\mathbf{s})\}, \{\varphi_2^{\text{mask}}(\mathbf{s})\}, \dots, \{\varphi_K^{\text{mask}}(\mathbf{s})\}$ be the phase sets corresponding to the masks that were selected during the generation process and, then, optimally superimposed by shifts and a change of enantiomers (below, this operation will be called optimal alignment [20]). We define “the best phases” and “the figures of merits” by the formula:

$$m(\mathbf{s}) \exp[i\varphi^{\text{best}}(\mathbf{s})] = \frac{1}{K} \sum_{k=1}^K \exp[i\varphi_k^{\text{mask}}(\mathbf{s})] \quad (16)$$

As an approximation to the electron density values, we will consider two types of the synthesis Fourier, not weighted

$$\rho_u(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{s}} F^{\text{obs}}(\mathbf{s}) \exp[i\varphi^{\text{best}}(\mathbf{s})] \exp[-2\pi i(\mathbf{s}, \mathbf{r})] \quad (17)$$

and weighted

$$\rho_w(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{s}} m(\mathbf{s}) F^{\text{obs}}(\mathbf{s}) \exp[i\varphi^{\text{best}}(\mathbf{s})] \exp[-2\pi i(\mathbf{s}, \mathbf{r})]. \quad (18)$$

As is easily seen, the weighted Fourier synthesis is nothing else but the arithmetic mean of the synthesis calculated using experimental modules and alternative variants of the phase values. In this study, two strategies for constructing the prior distributions (15) used in the mask generation procedure were compared. The strategy with the simultaneous use of all reflections during all calculations was tested in two variants, with two different types of Fourier synthesis (17) and (18).

3. TEST OBJECT AND CONDITIONS OF TESTING

The criteria for the mask selection were tested using the known structure of the monomeric form of Photosystem II (*Thermosynechococcus elongatus* photosystem II, [43], PDB entry 3KZI [44], below PS-II) as a testing object. This complex contains 19 protein subunits and multiple cofactors, which comprises in total about 25 000 non-hydrogen atoms. The molecular weight of PS-II from PDB is approximately equal to 360 KDa. Below, this value was used to estimate of the specific volume of the region of the molecule. The external dimensions of the monomer are approximately equal to 100 Å. An overall structure of the complex is shown in Fig. 3.

The unit cell dimensions of the imaginary crystal were chosen to be $200 \times 200 \times 200$ Å, and, from the atomic coordinates, a set of the structure factors $\{F^{\text{obs}}(\mathbf{s}), \varphi^{\text{true}}(\mathbf{s})\}$ of a resolution 16 Å was calculated (The total number of reflections was 4108). The magnitudes of these structure factors were considered in the tests as the experimental information about the object, and the values of phases $\{\varphi^{\text{true}}(\mathbf{s})\}$ were used only for the control of the results. At every step, random masks were generated until 100 masks satisfying the imposed restrictions were accumulated. The value of the contrast t in the distribution (14) was 10^6 .

The accuracy of the phase determination was characterized by the coefficients of the correlation of the Fourier syntheses with the experimental values of the moduli and the phase sets being compared at different resolution zones [20]. In terms of the structure factors values, these correlation coefficients can be presented in the form

$$CP[d_{\text{max}}, d_{\text{min}}](\{\varphi(\mathbf{s})\}) = \frac{\sum_{\mathbf{s} \in \mathbf{S}} (F^{\text{obs}}(\mathbf{s}))^2 \cos(\varphi^{\text{true}}(\mathbf{s}) - \tilde{\varphi}(\mathbf{s}))}{\sum_{\mathbf{s} \in \mathbf{S}} (F^{\text{obs}}(\mathbf{s}))^2}. \quad (19)$$

Here, the summation is over the structure factors with $\mathbf{s} \in \mathbf{S} = \{\mathbf{s} : 1/d_{\text{max}} < s < 1/d_{\text{min}}\}$, $\{\tilde{\varphi}(\mathbf{s})\}$ is the phase set obtained from the set being evaluated by its optimal alignment relative to $\{\varphi^{\text{true}}(\mathbf{s})\}$. In tests 2 and 3, the alignment was performed over all reflections of the zone $\infty - 16$ Å; in test 1, the upper limit was dependent on the step of the procedure (25, 20, 16 Å, correspondingly).

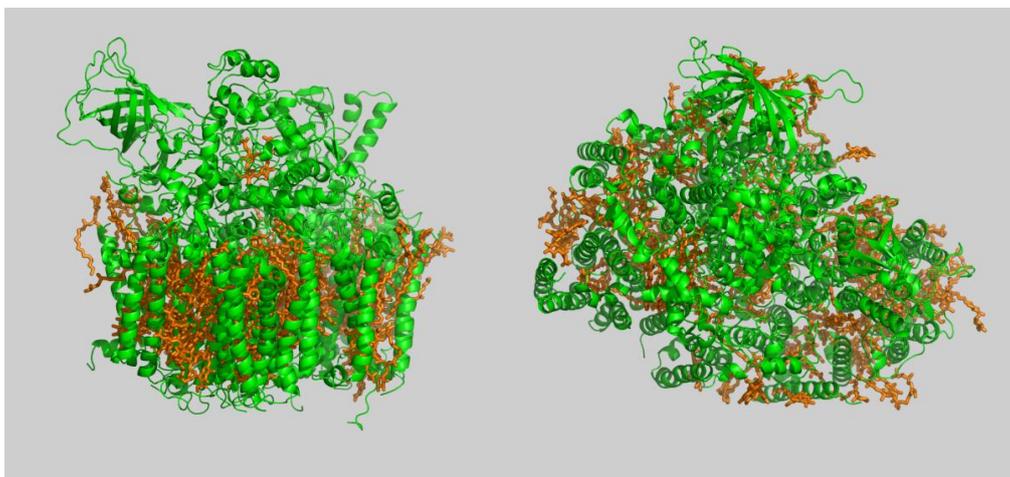


Fig. 3. PS-II. The protein part of the complex is shown in green, the cofactors are shown in orange.

One of the crucial parameters in the method is the size of generated masks or, what is the same, the hypothetical volume of the particle. Here, we characterize the supposed size of the region either by the number of the points inside a mask, or by a specific volume defined as the ratio of the volume of the region (in \AA^3) to the molecular mass of the particle (in Da). In crystallography, upon estimation of the percentage of the solvent in the unit cell, the specific volume of the molecule region is conventionally taken to be $1.23 \text{ \AA}^3/\text{Da}$ [45, 46]. However, this estimate is not optimal for low resolution studies, when the borders of the molecule region are significantly smoothed. Therefore, in this study, the specific volume of the molecule region was taken to be $1.33 \text{ \AA}^3/\text{Da}$. As a result, the grid and the number of the points inside the mask were determined at the resolution of 25, 20, and 16 as $(24 \times 24 \times 24, 831)$, $(30 \times 30 \times 30, 1623)$, and $(40 \times 40 \times 40, 3847)$, correspondingly.

4. THE RESULTS OF TESTS

4.1. Test 1

In biological crystallography, the determination of the structure of the model goes usually through several stages with a gradual increase in the amount of the experimental data used [47]. According to this rule, the phase values were sequentially reconstructed in three stages, with the quantity of data at each stage being doubled. At the beginning, we used a set of reflections of the resolution 25 \AA (1054 reflections, steps 1–5). At the next steps, the resolution was increased to 20 \AA (2082 reflections, steps 6–10) and, then, to 16 \AA (4108 reflections, steps 11–16). At the first step, the masks were generated with the uniform prior probability distribution. At the next steps, we used the probability distribution (14) with the value of the contrast $t = 10^6$ and not weighted synthesis (17) being built from the results of the previous step. In Figure 4, a change in the value of phase correlation with the exact phases for the different shells in the reciprocal space during the whole procedure is shown. Table 1 contains the values of the correlation between the final phase set and the exact phases for the extending zones and shells in the reciprocal space.

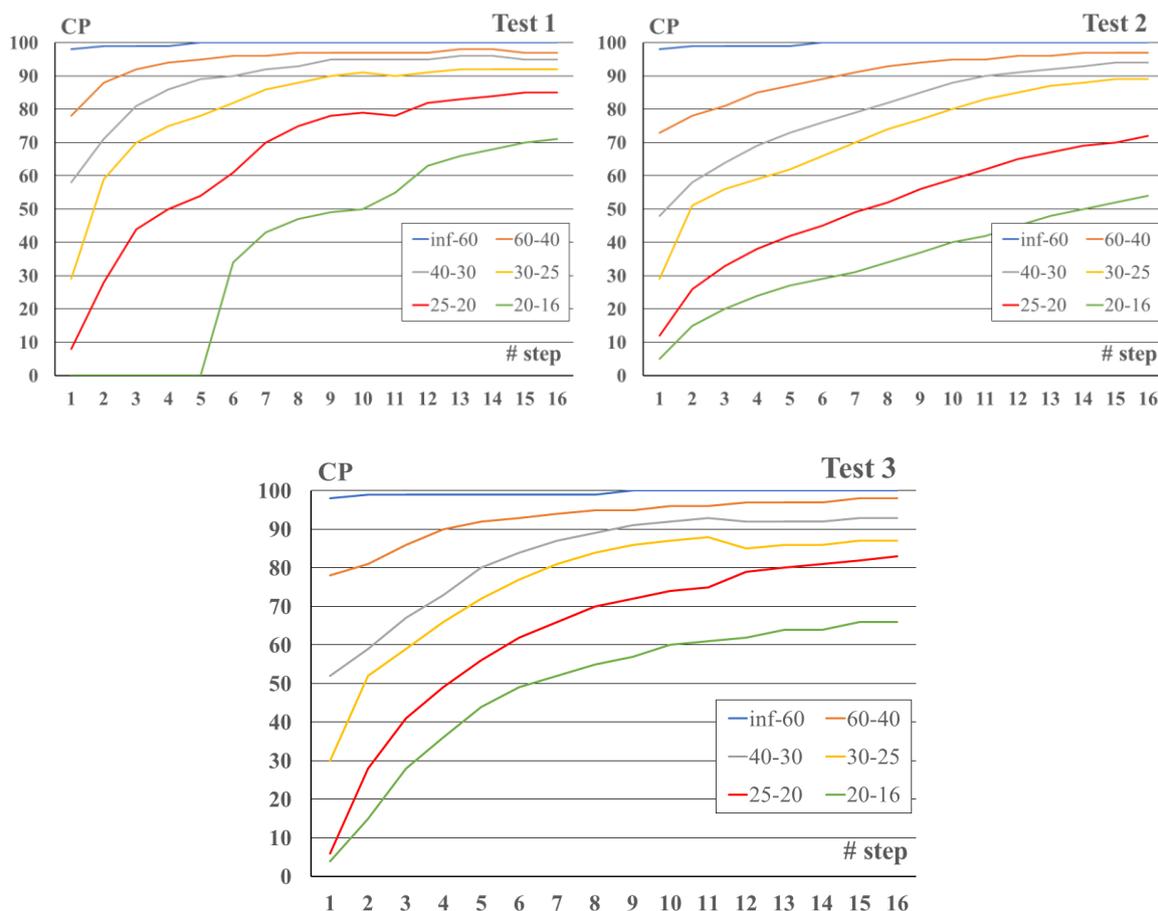


Fig. 4. A change in the value of phase correlation with the exact phases during the solution of the phase problem. The values calculated by the formula (19) are shown for different intervals $d_{\min} - d_{\max}$.

Table 1. The quality of the final data sets. The value of the phase correlation is given (in percent).

	Resolution zone $d_{\max} - d_{\min}$ (Å)										
	$\infty-60$	$\infty-40$	$\infty-30$	$\infty-25$	$\infty-20$	$\infty-16$	60-40	40-30	30-25	25-20	20-16
Test 1	100	100	99	99	99	98	97	95	92	85	71
Test 2	100	100	99	99	98	96	97	94	89	72	54
Test 3	100	100	99	99	98	97	98	93	87	83	66
Number refl in the zone	85	255	618	1054	2082	4108	170	363	436	1028	2026

4.2. Test 2

In the second test, all reflections of the resolution zone $\infty-16$ Å were included in the work simultaneously. Figure 4 shows a change in the value of phase correlation with the exact phases for different shells in the reciprocal space at different steps of the test. The corresponding values of the correlation between the final phase set and the exact phases for extending zones and shells in the reciprocal space are shown in Table 1.

4.3. Test 3

In the third test, an intermediate strategy was used. Formally, all reflections were involved, but for the construction of the prior probability distribution, the weighted Fourier synthesis (18) was used. Because the figures of merit (16) for the shells of the high resolution at initial steps are small, the high-resolution reflections were not involved in the calculation of

prior distributions. At the same time, the selection criteria were calculated using all reflections of the zone 16 \AA . Figure 4 and Table 1, along with the results of the previous tests, give the results of Test 3.

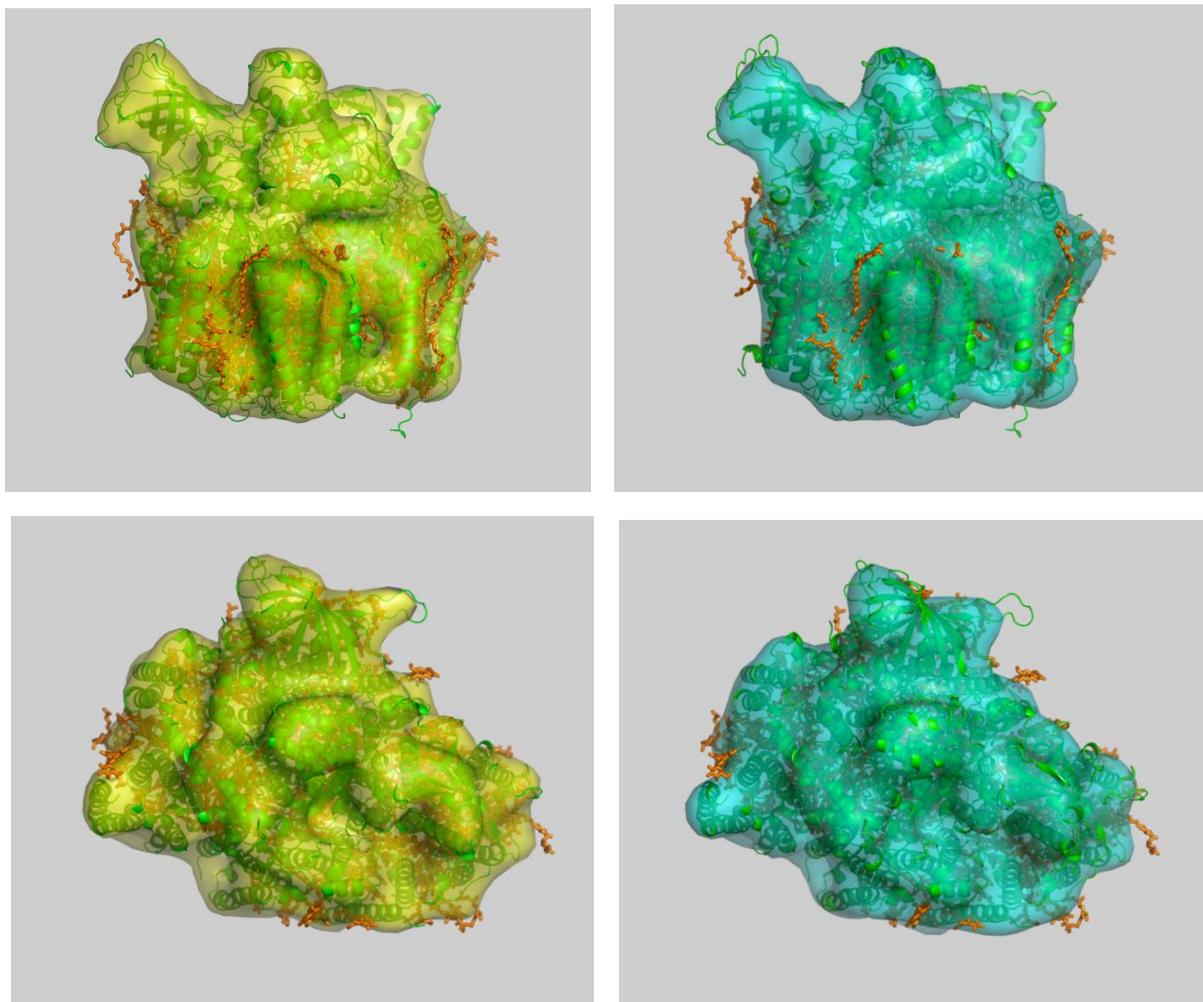


Fig. 5. "An ideal" Fourier synthesis at a resolution of 16 \AA (on the left) and the synthesis calculated with the phase values obtained at the 16th step in test 3 (on the right). The syntheses are superimposed with the model of PS-II. The specific volume of the object region shown is $1.23 \text{ \AA}^3/\text{Da}$.

The final Fourier synthesis obtained in Test 3 and an "ideal" Fourier synthesis calculated using the exact phase values at a resolution of 16 \AA are shown in Figure 5.

5. CONCLUSIONS

Our results have shown that the use of the new type of the selection criterion allows one to solve the phase problem in X-ray diffraction studies of single particles. The number of the determined phase values increased two times in comparison with the tests performed in the earlier studies based on the previous criterion, the correlation coefficient of magnitudes [20–37].

A gradual increase in resolution in calculations of prior distributions and the selection criterion gives better results compared with the simultaneous use of all reflections.

The study has shown that in the case of all reflection included into the work simultaneously, the use of weighed syntheses in the calculation of the prior distribution leads to the convergence to the solution for highest-resolution reflections that is faster than the convergence achieved with the use of non-weighted syntheses. Another advantage of this approach is that it automatically takes the reflections of the highest resolution shell into the

work, in accordance with the increasing reliability of the corresponding phase information and does not demand a manual intervention and an expert evaluation.

This work was supported by the Russian Foundation for Basic Research (project 16-04-01037a). The calculations were performed at the multiprocessor computing complex K100 of Keldysh Institute of Applied Mathematics, RAS. AU acknowledges the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC.

REFERENCES

1. Barends T.R.M., Foucar L., Botha S., Doak R.B., Shoeman R. L., Nass K., Koglin J.E., Williams G.J., Boutet S., Messerschmidt M., Schlichting I. De novo protein crystal structure determination from X-ray free-electron laser data. *Nature*. 2014. V. 505. P. 244–247. doi: [10.1038/nature12773](https://doi.org/10.1038/nature12773)
2. Chapman H.N., Fromme P., Barty A., White T.A., Kirian R.A., Aquila A., Hunter M.S., Schulz J., DePonte D.P., Weierstall U., et al. Femtosecond X-ray protein nanocrystallography. *Nature*. 2011. V. 470. P. 73–77. doi: [10.1038/nature09750](https://doi.org/10.1038/nature09750)
3. Boutet S., Lomb L., Williams G.J., Barends T.R., Aquila A., Doak R.B., Weierstall U., DePonte D.P., Steinbrener J., Shoeman R.L., et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science*. 2012. V. 337. P. 362–364. doi: [10.1126/science.1217737](https://doi.org/10.1126/science.1217737)
4. Redecke L., Nass K., DePonte D.P., White T.A., Rehders D., Barty A., Stellato F., Liang M., Barends T.R.M., Boutet S., et al. Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser. *Science*. 2013. V. 339. P. 227–230. doi: [10.1126/science.1229663](https://doi.org/10.1126/science.1229663)
5. Lomb L., Barends T.R.M., Kassemeyer S., Aquila A., Epp S.W., Erk B., Foucar L., Hartmann R., Rudek B., Rolles D., et al. Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Physical Review B*. 2011. V. 84. 214111. doi: [10.1103/PhysRevB.84.214111](https://doi.org/10.1103/PhysRevB.84.214111)
6. Johansson L.C., Arnlund D., White T.A., Katona G., DePonte D.P., Weierstall U., Doak R.B., Shoeman R.L., Lomb L., Malmerberg E., et al. Lipidic phase membrane protein serial femtosecond crystallography. *Nature Methods*. 2012. V. 9. P. 263–265. doi: [10.1038/nmeth.1867](https://doi.org/10.1038/nmeth.1867)
7. Kern J., Alonso-Mori R., Hellmich J., Tran R., Hattne J., Laksmono H., Glöckner C., Echols N., Sierra R.G., Sellberg J., et al. Room temperature femtosecond X-ray diffraction of photosystem II microcrystals. *Proceedings of the National Academy of Sciences of the USA*. 2012. V. 109. P. 9721–9726. doi: [10.1073/pnas.1204598109](https://doi.org/10.1073/pnas.1204598109)
8. Lunin V.Y., Lunina N.L., Petrova T.E. The biological crystallography without crystals. *Mathematical Biology and Bioinformatics*. 2017. V. 12. No. 1. P. 55–72. doi: [10.17537/2017.12.55](https://doi.org/10.17537/2017.12.55)
9. <https://doi.org/10.17537/2017.12.55>Thibault P., Elser V., Jacobsen C., Shapiro D., Sayre D. Reconstruction of a yeast cell from X-ray diffraction data. *Acta Crystallographica Section A: Foundations of Crystallography*. 2006. V. 62. P. 248–261. doi: [10.1107/S0108767306016515](https://doi.org/10.1107/S0108767306016515)
10. Rodriguez J.A., Xu R., Chen C.C., Huang Z., Jiang H., Chen A.L., Raines K.S., Pryor A. Jr, Nam D., Wiegart L., Song C., Madsen A., Chushkin Y., Zontone F., Bradley P.J., Miao J. Three-dimensional coherent X-ray diffractive imaging of whole frozen-hydrated cells. *IUCr Journal*. 2015. V. 2. P. 575–583. doi: [10.1107/S205225251501235X](https://doi.org/10.1107/S205225251501235X)
11. Takayama Y., Yonekura K. Cryogenic coherent X-ray diffraction imaging of biological samples at SACLA: a correlative approach with cryo-electron and light microscopy. *Acta Crystallographica Section A: Foundations and Advances*. 2016. V. 72. P. 179–189. [10.1107/S2053273315023980](https://doi.org/10.1107/S2053273315023980)

12. Munke A., Andreasson J., Aquila A., Awel S., Ayyer K., Barty A., Bean R.J., Berntsen P., Bielecki J., Boutet S. *et al.* Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source. *Sci. Data*. 2016. V. 3. Article No. 160064. doi: [10.1038/sdata.2016.64](https://doi.org/10.1038/sdata.2016.64)
13. Ekeberg T., Svenda M., Abergel C., Maia F.R.N.C., Seltzer V., Claverie J.M., Hantke M., Jönsson O., Nettelblad C., van der Schot G. *et al.* Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. *Physical Review Letters*. 2015. V. 114. Article No. 098102. doi: [10.1103/PhysRevLett.114.098102](https://doi.org/10.1103/PhysRevLett.114.098102)
14. Song C., Jiang H., Mancuso A., Amirbekian B., Peng L., Sun R., Shah S.S., Zhou Z.H., Ishikawa T., Miao J. Quantitative Imaging of Single, Unstained Viruses with Coherent X Rays. *Physical Review Letters*. 2008. V. 101. Article No. 158101. doi: [10.1103/PhysRevLett.101.158101](https://doi.org/10.1103/PhysRevLett.101.158101)
15. Seibert M.M., Ekeberg T., Maia F.R.N.C., Svenda M., Andreasson J., Jönsson O., Odic D., Iwan B., Rucker A., Westphal D. Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature*. 2011. V. 470. P. 78–82. doi: [10.1038/nature09748](https://doi.org/10.1038/nature09748)
16. Van der Schot G., Svenda M., Maia F.R.N.C., Hantke M., DePonte D., Seibert M.M., Aquila A., Schulz J., Kirian R., Liang M. *et al.* Imaging single cells in a beam of live cyanobacteria with an X-ray laser. *Nature Communication*. 2015. V. 6. Article No. 5704. doi: [10.1038/ncomms6704](https://doi.org/10.1038/ncomms6704)
17. Fienup J.R. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*. 1978. V. 3. No. 1. P. 27–29. doi: [10.1364/OL.3.000027](https://doi.org/10.1364/OL.3.000027)
18. Urzhumtseva L., Klaholz B., Urzhumtsev A. On effective and optical resolutions of diffraction data sets. *Acta Crystallographica Section D: Biological Crystallography*. 2013. V. 69. P. 625–634. doi: [10.1107/S09074444913016673](https://doi.org/10.1107/S09074444913016673)
19. Sayre D. Some implications of a theorem due to Shannon. *Acta Crystallographica*. 1952. V. 5. P. 843. doi: [10.1107/S0365110X52002276](https://doi.org/10.1107/S0365110X52002276)
20. Lunin V.Y., Lunina N.L., Petrova T.E., Baumstark M.W., Urzhumtsev A.G. Mask-based approach to phasing of single-particle diffraction data. *Acta Crystallographica Section D: Structural Biology*. 2016. V. 72. P. 147–157. doi: [10.1107/S2059798315022652](https://doi.org/10.1107/S2059798315022652)
21. Marchesini S., He H., Chapman H.N., Hau-Riege S.P., Noy A., Howells M.R., Weierstall U., Spence J.H.C. X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B*. 2003. V. 68. Article No. 140101(R). doi: [10.1103/PhysRevB.68.140101](https://doi.org/10.1103/PhysRevB.68.140101)
22. Maia F.R.N.C., Ekeberg T., Spoel D., Hajdu J. Hawk: the image reconstruction package for coherent X-ray diffractive imaging. *J. Applied Crystallography*. 2010. V. 43. P. 1535–1539. doi: [10.1107/S0021889810036083](https://doi.org/10.1107/S0021889810036083)
23. Bricogne G. Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallographica Section A: Foundations of Crystallography*. 1974. V. 30. P. 349–405. doi: [10.1107/S0567739474010722](https://doi.org/10.1107/S0567739474010722)
24. Rodriguez J.A., Xu R., Chen C.-C., Zou Y., Miao J. Oversampling smoothness: an effective algorithm for phase retrieval of noisy diffraction intensities. *J. Applied Crystallography*. 2013. V. 46. P. 312–318. doi: [10.1107/S0021889813002471](https://doi.org/10.1107/S0021889813002471)
25. Miao J., Kirz J., Sayre D. The oversampling phasing method. *Acta Crystallographica Section D: Biological Crystallography*. 2000. V. 56. P. 1312–1315. doi: [10.1107/S0907444900008970](https://doi.org/10.1107/S0907444900008970)
26. He H., Su, W.P. Direct phasing of protein crystals with high solvent content. *Acta Crystallographica Section A: Foundations of Crystallography*. 2015. V. 71. P. 92–98. doi: [10.1107/S2053273314024097](https://doi.org/10.1107/S2053273314024097)

27. Bricogne G. Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallographica Section A: Foundations of Crystallography*. 1976. V. 32. P. 832–847. doi: [10.1107/S0567739476001691](https://doi.org/10.1107/S0567739476001691)
28. Marchesini S. A unified evaluation of iterative projection algorithms for phase retrieval. *Rev. Sci. Instrum.* 2007. V. 78. Article No. 011301. doi: [10.1063/1.2403783](https://doi.org/10.1063/1.2403783)
29. Zhang K.Y.J., Cowtan K.D., Main P. Phase improvement by iterative density modification. In: *International Tables for Crystallography*. Vol. F. Second Edition. Eds. Arnold E., Himmel D.M., Rossmann M.G. Chichester: John Wiley and Sons, 2012. P. 385–400. doi: [10.1107/97809553602060000847](https://doi.org/10.1107/97809553602060000847)
30. Millane R., Lo V.L. Iterative projection algorithms in protein crystallography. I. Theory. *Acta Crystallographica Section A: Foundations of Crystallography*. 2013. V. 69. P. 517–527. doi: [10.1107/S0108767313015249](https://doi.org/10.1107/S0108767313015249)
31. Wang B.C. Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology*. 1985. V. 115. P. 90–112.
32. Loh N.-T.D., Elser V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Physical Review E*. 2009. V. 80. Article No. 026705. doi: [10.1103/PhysRevE.80.026705](https://doi.org/10.1103/PhysRevE.80.026705)
33. Elser V. Solution of the crystallographic phase problem by iterated projections. *Acta Crystallographica Section A: Foundations of Crystallography*. 2003. V. 59. P. 201–209. doi: [10.1107/S0108767303002812](https://doi.org/10.1107/S0108767303002812)
34. Lunin V.Y., Urzhumtsev A.G., Skovoroda T.P. Direct low-resolution phasing from electron-density histograms in protein crystallography. *Acta Crystallographica Section A: Foundations of Crystallography*. 1990. V. 46. P. 540–544. doi: [10.1107/S0108767390003464](https://doi.org/10.1107/S0108767390003464)
35. Lunin V.Y., Lunina N.L., Urzhumtsev A.G. Connectivity properties of high-density regions and ab initio phasing at low resolution. *Acta Crystallographica Section A: Foundations of Crystallography*. 2000. V. 56. P. 375–382. doi: [10.1107/S0108767300004633](https://doi.org/10.1107/S0108767300004633)
36. Petrova T.E., Lunin V.Y., Podjarny A.D. Ab initio low-resolution phasing in crystallography of macromolecules by maximization of likelihood. *Acta Crystallographica Section D: Biological Crystallography*. 2000. V. 56. P. 1245–1252. doi: [10.1107/S0907444900009343](https://doi.org/10.1107/S0907444900009343)
37. Lunin V.Y., Lunina N.L., Petrova T.E. The use of connected masks for reconstructing the single particle image from X-ray diffraction data. *Mathematical Biology and Bioinformatics*. 2014. V. 10. No. S. P. t1–t19. doi: [10.17537/2015.10.t1](https://doi.org/10.17537/2015.10.t1)
38. Baker D., Krukowski A.E., Agard D.A. Uniqueness and the ab initio phase problem in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*. 1993. V. 49. P. 186–192. doi: [10.1107/S0907444992008801](https://doi.org/10.1107/S0907444992008801)
39. Lunin V.Y., Urzhumtsev A.G. Improvement of protein phases by coarse model modification. *Acta Crystallographica Section A: Foundations of Crystallography*. 1984. V. 40. P. 269–277. doi: [10.1107/S0108767384000544](https://doi.org/10.1107/S0108767384000544)
40. Lunin V.Y., Skovoroda T.P. R-free likelihood-based estimates of errors for phases calculated from atomic models. *Acta Crystallographica Section A: Foundations of Crystallography*. 1995. V. 51. P. 880–887. doi: [10.1107/S010876739500688X](https://doi.org/10.1107/S010876739500688X)
41. Urzhumtsev A.G., Skovoroda T.P., Lunin V.Y. A procedure compatible with X-PLOR for the calculation of electron-density maps weighted using an R-free-likelihood approach. *J. Applied Crystallography*. 1996. V. 29. P. 741–744. doi: [10.1107/S0021889896007194](https://doi.org/10.1107/S0021889896007194)
42. Read R.J. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Crystallographica Section A: Foundations of Crystallography*. 1986. V. 42. P. 140–190. doi: [10.1107/S0108767386099622](https://doi.org/10.1107/S0108767386099622)

43. Broser M., Gabdulkhakov A., Kern J., Guskov A., Müh F., Saenger W., Zouni A. Crystal structure of monomeric Photosystem II from *Thermosynechococcus elongatus* at 3.6 Å resolution. *J. Biol. Chem.* 2010. V. 285. P. 26255–26262. doi: [10.1074/jbc.M110.127589](https://doi.org/10.1074/jbc.M110.127589)
44. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. *Nucleic Acids Research.* 2000. V. 28. P. 235–242. <https://doi.org/10.1093/nar/28.1.235>
45. Matthews B.M. Solvent content of protein crystals. *Journal of Molecular Biology.* 1968. V. 33. P. 491–497. doi: [10.1016/0022-2836\(68\)90205-2](https://doi.org/10.1016/0022-2836(68)90205-2)
46. Weichenberger C.X., Rupp B. Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallographica Section D: Biological Crystallography.* 2014. V. 70. P. 1579–1588. doi: [10.1107/S1399004714005550](https://doi.org/10.1107/S1399004714005550)
47. Urzhumtsev A., Afonine P.V., Adams P.D. On the use of logarithmic scales for analysis of diffraction data. *Acta Crystallographica Section D: Biological Crystallography.* 2009. V. 65. P. 1283–1291. doi: [10.1107/S09074444090039638](https://doi.org/10.1107/S09074444090039638)

Received 25 June 2018.
Published 29 June 2018.