UDC: 577.3, 548, 519.6

Single Particle Study by X-Ray Diffraction: Crystallographic Approach

Lunin V.Y., Lunina N.L., Petrova T.E.

Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia

Abstract. An increase in the power of X-ray sources, in particular, the commissioning of X-ray free electron lasers, opens up the possibility of recording the diffraction by single macromolecular biological particles. These opportunities open the way to weakening, and, ideally, removal, of the main limitation of X-ray structure analysis, namely, the need to prepare the sample in the single-crystal form. However, the possibility of the practical recording of diffraction by a single particle is currently limited to a very low-resolution zone, what is one of the main obstacles in the development of this approach. This paper discusses the similarities and differences in the study of crystal samples and single particles. It is shown that the problem of the determination of the structure of a single particle can be formulated as a standard problem of biological crystallography, namely, as the problem of retrieval the electron density distribution in some unit cell from the magnitudes of its Fourier coefficients. This makes it possible to apply the entire range of the methods of biological crystallography to the study of isolated particles. At the same time, the possibility of recording continuous diffraction pattern for a single particle (as opposed to a discrete set of Bragg reflections in the case of a crystal) significantly increases the amount of information derived from the experiment. The analytical properties of the continuous diffraction pattern create a potential opportunity both to restore the structure factors phases (lost in the diffraction experiment), and to extrapolate the experimentally observed pattern onto a wider area, which allows to increase the resolution of Fourier syntheses for the electron density distribution.

Key words: biological macromolecules, single-particles, X-ray scattering, X-ray free electron lasers, phase problem, resolution of Fourier syntheses.

1. INTRODUCTION

X-ray structure analysis (XRSA) is currently the main experimental method to determine the structure of biological macromolecules and their complexes at the atomic level. The available programs for experimental data processing and structure determination [1–6] often allow one to solve a new structure in "almost automatic" mode. The number of macromolecular structures determined by XRSA is estimated to be hundreds of thousands [7]. However, the key limitation of the method is the need to prepare a sample of the object under study in the form of a single crystal. Obtaining the crystals is a complex, poorly predictable procedure. The search for the crystallization conditions can sometimes take years and ends in failure. The development of X-ray experiment techniques, in particular, the commissioning of X-ray free electron lasers, opens up the potential possibility to remove the requirement of the crystallization and use X-ray diffraction methods for determining the structure of a single particle [8–13]. By a single particle we mean the sample that does not have internal periodicity and is a single copy of the biological object being studied. The examples of single particles are macromolecules, macromolecular complexes, viruses, bacterial cells, etc. This paper discusses the similarities and differences of computational problems that arise in studies of crystalline samples and isolated particles.

In this paper we deal with different types of function. To emphasize the type when necessary we use superscript *cr* to mark a periodic function (e.g. $\rho^{cr}(\mathbf{r})$), superscript *sp* to mark a function with a compact support (e.g. $\rho^{sp}(\mathbf{r})$), and subscript *V* to mark a function defined to be zero outside the unit cell *V*, and it's derived functions (e.g. $\rho_V(\mathbf{r})$, $\mathbf{F}_V(\mathbf{s})$, $A_V(\mathbf{r})$).

2. THE BASICS OF THE METHOD

2.1. X-ray diffraction experiment

A scheme of the standard diffraction experiment is shown in Figure 1. A primary X-ray beam irradiates the sample. We specify the direction of the primary beam by a unit length vector $\mathbf{\sigma}_0$. A detector measures the energy $E(\mathbf{\sigma}_0, \mathbf{\sigma})$ or energy-related characteristics of the X-ray beam propagating from the sample to the detector. Here a unit length vector $\mathbf{\sigma}$ specifies the direction from the sample to the detector. The beams, propagated from the sample to detector are usually referred to as reflections. This terminology goes back to the first works on the scattering of X-rays by crystals, where the occurrence of scattered rays was interpreted as a reflection of the primary beam in the "crystal planes". During the experiment, the direction of the primary beam and the detector position change relative to the sample, allowing one to measure the energy $E(\mathbf{\sigma}_0, \mathbf{\sigma})$ for various combinations of vectors $\mathbf{\sigma}_0$ and $\mathbf{\sigma}$. In a real experiment in biological crystallography, the direction of the primary beam relative to the sample changes due to the rotation of the sample, and the energy of reflections is usually recorded by a matrix detector, which represents a matrix (e.g., 1024×1024) composed of closely spaced individual detectors (with a pitch of 0.2 mm, as an example). This allows one to register energy simultaneously for the different directions of the scattering.



Fig. 1. The scheme of X-ray diffraction experiment.

2.2. Kinematic theory of diffraction

The main formulas of the XRSA are derived within the framework of the kinematic theory of diffraction, the main statements of which can be formulated as follows. The primary X-ray beam is a plane monochromatic electromagnetic wave, under the influence of which the

electrons of the sample begin to oscillate. Oscillating electrons become the sources of secondary spherical electromagnetic waves. At the detector, the secondary waves coming from different electrons superpose forming the diffracted beam. The measured energy is proportional to the square of the of the complex amplitude of the summary wave. The solution of Newton's equation for motion of an electron in the field of the primary wave and Maxwell's equations describing the radiation of an oscillating electron [14, 15] allow to calculate the recorded energy as:

$$E(\boldsymbol{\sigma}_{0},\boldsymbol{\sigma}) = \varepsilon E_{0} \left| \mathbf{F}(\mathbf{s}) \right|^{2} \quad . \tag{1}$$

Here, **s** is the combination of σ_0 and σ directions scaled to the wavelength

$$\mathbf{s} = \frac{\mathbf{\sigma} - \mathbf{\sigma}_0}{\lambda} \quad . \tag{2}$$

This vector is referred to as the scattering vector and play a significant role in the theory of diffraction. In crystallography, the space of scattering vectors is referred to as the reciprocal space. The value E_0 is the energy of the primary wave. The complex value $\mathbf{F}(\mathbf{s})$ is the Fourier transform of the electron density distribution function $\rho(\mathbf{r})$, that describes the distribution of electrons in the sample

$$\mathbf{F}(\mathbf{s}) = \int_{\mathbf{R}^3} \rho(\mathbf{r}) \exp[i2\pi \mathbf{s} \cdot \mathbf{r}] dV_{\mathbf{r}} , \ \mathbf{s} \in \mathbf{R}^3 , \qquad (3)$$

where the dot \cdot denotes the scalar (dot) product of two vectors. The value ε is a combination of physical constants and experimental parameters (e.g., the distance from the sample to the detector) and does not depend on the structure of the sample. By changing the directions σ_0 and σ it is possible to obtain the values of the Fourier transform magnitude $F(\mathbf{s})$ for different scattering vectors \mathbf{s} in the reciprocal space. We call as the diffraction pattern the set of squares $I(\mathbf{s}) = |\mathbf{F}(\mathbf{s})|^2$ of these magnitudes.



Fig. 2. Ewald scheme. Scattering vectors **s** corresponding to the existing reflections form the surface of a sphere (Ewald sphere) in reciprocal space (shown in purple). The reflections recorded by the detector matrix correspond to a part of this sphere (shown in green) determined by the physical dimensions of the detector matrix. The rotation of the object is accompanied by the rotation of the basis {**a**^{*}, **b**^{*}, **c**^{*}}, that leads to the change of the position in the reciprocal space of the scattering vectors corresponding to the existing and recorded reflections.

If the position of the object is fixed, the scattering vectors corresponding to all existing at the moment reflections fill the sphere of $1/\lambda$ radius with the center placed to the reciprocal space origin. This sphere is referred to as Ewald sphere (Fig. 2). The scattering vectors corresponding to the reflections that cross the detector area fill a part of the surface of Ewald sphere. If the corresponding scattering angles are small, it may be assumed that the pattern recorded by the detector matrix (the frame or the image) corresponds to a part of a plane in the reciprocal space passing through the origin.

Formulae (1)–(3) provide the solution to the direct problem of the theory of diffraction, namely, the prediction of the diffraction pattern $\{I(\mathbf{s}), \mathbf{s} \in \mathbf{R}^3\}$ supposing that the electron density distribution $\rho(\mathbf{r})$ in the object is known. The purpose of the X-ray diffraction study is to solve the inverse problem, namely, to retrieve the electron density distribution assuming the diffraction pattern is known. At the next stages of the study, the electron density distribution found is "interpreted", i.e., is represented as a sum of the contributions of the atoms producing the atomic model of the object. The parameters of this model are further refined by minimization of the discrepancy between the observed diffraction pattern and one derived from the model [16].

2.3. The phase problem

Equation (3) allows to represent the electron density distribution as an inverse Fourier transform

$$\rho(\mathbf{r}) = \int_{\mathbf{R}^3} \mathbf{F}(\mathbf{s}) \exp\left[-i2\pi \mathbf{s} \cdot \mathbf{r}\right] dV_{\mathbf{s}}, \ \mathbf{r} \in \mathbf{R}^3.$$
(4)

The first obstacle in the calculation of the electron density distribution directly from this formula is that the experiment allows to determine only the values of the magnitude $F(\mathbf{s})$ of the complex Fourier transform

$$\mathbf{F}(\mathbf{s}) = F(\mathbf{s}) \exp[i\varphi(\mathbf{s})], \ \mathbf{s} \in \mathbf{R}^3.$$
(5)

The retrieval of the lost phase values $\varphi(\mathbf{s})$ is a central, the so-called phase problem of XRSA. In biological crystallography, this problem is solved either by conducting additional experiments with chemically modified compounds (isomorphous replacement method) or a series of experiments at different X-rays wavelengths (multiwavelength anomalous diffraction method), or using the known structure of the proposed homologue of the object (molecular replacement method).

2.4. Resolution

The second problem in an attempt to get the electron density distribution by means of (4) is that the integration should be performed over the whole three-dimensional space, while the observed values $F^{obs}(\mathbf{s})$ are available only for a limited area of space, for example, they may be restricted as $|\mathbf{s}| \leq s_{\max}$. In addition, some particular values may not be available due to specific features of the experimental setup, for example, reflections of the central zone with $|\mathbf{s}| \leq s_{\min}$ may not be available. The transformation (4) calculated through a limited region of the reciprocal space is called in crystallography the Fourier synthesis of the electron density. Fourier syntheses contain distortions, which becomes greater, if the region of the reciprocal space included into the calculation of the integral (4) diminishes. The usual types of these distortions are "blurring" and a fusion of the electron density peaks, and the appearance of waves due to a truncation of the Fourier series. Figure 3 shows electron density syntheses calculated for a small protein using data sets of different sizes.

503

The concept of 'resolution' is used to characterize the image quality and amount of data used. In macromolecular crystallography this figure corresponds to the size of the smallest distinguishable details in the synthesis and is specified in Ångströms. Formally, this characteristic is first introduced for an individual Fourier harmonic $H_s(\mathbf{r}) = \exp[-i2\pi \mathbf{s} \cdot \mathbf{r}]$. The real and imaginary parts of this harmonic considered as functions of the vector \mathbf{r} retain constant values on the planes perpendicular to the vector \mathbf{s} and change as sine function along the direction \mathbf{s} . The period of this sine-wave is equal to

$$d = \frac{1}{|\mathbf{s}|} = \frac{\lambda}{2\sin\theta} \quad , \tag{6}$$

and is referred to as the resolution corresponding to the scattering vector **s**. It is said that a data set is assembled or the Fourier synthesis is calculated with the resolution s_{\max} if all (or almost all) reflections with $|\mathbf{s}| \le s_{\max}$ are included in the calculation of (4). In addition to the introduced formal resolution, other concepts of real, optical or effective resolution are used to characterize the quality of the Fourier synthesis, which take into account not only the absence of some reflections, but also distortions introduced by errors in the values of magnitudes and phases [17].



Fig. 3. Fourier syntheses of electron density for a small protein molecule (Protein G). The numbers specify the resolution and the number of Bragg reflections used to calculate the Fourier series.

2.5. The possibility of experimental recording of the diffraction

The main obstacle in the practical use of X-ray diffraction methods is the value of the constant ε included in the expression (1). In the case of scattering by a single electron, this value can be estimated as 10^{-24} . This creates serious problems in attempts to record diffraction; in particular, until recently it was almost impossible to register the scattering performed by individual protein molecules. The main solution remains still to prepare a sample in the form of a single crystal, which plays the role of an amplifier of diffracted waves.

3. SCATTERING BY A CRYSTAL

3.1. Mathematical (ideal infinite) crystal

3.1.1. Crystallographic notations

For an ideal (mathematical) infinite three-dimensional crystal, the electron density distribution $\rho^{cr}(\mathbf{r})$ has three non-coplanar periods {**a**, **b**, **c**}, i.e.,

$$\rho^{cr}(\mathbf{r}) = \rho^{cr}(\mathbf{r} + \mathbf{a}) = \rho^{cr}(\mathbf{r} + \mathbf{b}) = \rho^{cr}(\mathbf{r} + \mathbf{c}), \ \mathbf{r} \in \mathbf{R}^{3}.$$
(7)

We call the vectors $\{a, b, c\}$ the basis of the direct space; the coordinates of the vector in this (in general, oblique) basis, the relative coordinates; and the lattice \Re composed from the points with the integer coordinates:

$$\mathfrak{R} = \{ x\mathbf{a} + y\mathbf{b} + z\mathbf{c}, x, y, z - \text{integers} \}, \qquad (8)$$

the lattice of the direct space. We define the basis $\{a^*, b^*, c^*\}$ of the reciprocal space as the basis that is conjugate to the basis $\{a, b, c\}$, i.e.

$$\mathbf{a} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1,$$

$$\mathbf{a} \cdot \mathbf{b}^* = \mathbf{a} \cdot \mathbf{c}^* = \mathbf{b} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{c}^* = \mathbf{c} \cdot \mathbf{a}^* = \mathbf{c} \cdot \mathbf{b}^* = 0.$$
(9)

We call the coordinates of a point in the reciprocal space basis the Miller indices, and the integer lattice

$$\mathfrak{R}' = \{ha^* + kb^* + lc^*, h, k, l - \text{integers}\}$$
(10)

the reciprocal space lattice, or the Bragg lattice. The reflections corresponding to the scattering vectors $\mathbf{s} \in \mathfrak{R}'$ are called the Bragg reflections. If (x, y, z) are the relative coordinates of the vector \mathbf{r} and (h, k, l) are the Miller indices of the vector \mathbf{s} , then, the scalar product of two vectors can be calculated as $\mathbf{s} \cdot \mathbf{r} = hx + ky + lz$. In crystallography, it is customary to speak of vectors represented in the coordinates of the basis $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ as the vectors of direct space, while vectors represented by coordinates in the basis $\{\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*\}$ are called vectors of the reciprocal space.

We call the parallelepiped constructed on the vectors {**a**, **b**, **c**} and centered at the origin the unit cell *V*. Let denote as $\rho_V(\mathbf{r})$ the restriction of the function $\rho(\mathbf{r})$ to the unit cell *V*, i.e.

$$\rho_{V}\left(\mathbf{r}\right) = \begin{cases} \rho^{cr}\left(\mathbf{r}\right), \ \mathbf{r} \in V\\ 0, \ \mathbf{r} \notin V \end{cases}.$$
(11)

The set of the Fourier harmonics $H_{\mathbf{s}}(\mathbf{r}) = \exp[-i2\pi\mathbf{s}\cdot\mathbf{r}]$, with $\mathbf{s} \in \mathfrak{R}'$ forms a complete orthogonal basis in the space $L^2(V)$ of functions defined in the region *V*, therefore, the functions $\rho^{cr}(\mathbf{r})$ and $\rho_V(\mathbf{r})$ can be represented in *V* as Fourier series

$$\rho^{cr}\left(\mathbf{r}\right) = \rho_{V}\left(\mathbf{r}\right) = \frac{1}{|V|} \sum_{\mathbf{s} \in \Re'} \mathbf{F}_{V}\left(\mathbf{s}\right) \exp\left[-i2\pi\mathbf{s} \cdot \mathbf{r}\right], \ \mathbf{r} \in V,$$
(12)

where |V| is the unit cell volume, and

$$\mathbf{F}_{V}(\mathbf{s}) = \int_{V} \rho(\mathbf{r}) \exp[i2\pi\mathbf{s}\cdot\mathbf{r}] dV_{\mathbf{r}} = \int_{\mathbf{R}^{3}} \rho_{V}(\mathbf{r}) \exp[i2\pi\mathbf{s}\cdot\mathbf{r}] dV_{\mathbf{r}} \quad .$$
(13)

For the function $\rho^{cr}(\mathbf{r})$, the presentation (12) is valid for all $\mathbf{r} \in \mathbf{R}^3$, while for $\rho_v(\mathbf{r})$ it is valid in the unit cell only. In crystallography, the complex Fourier coefficients $\mathbf{F}_v(\mathbf{s})$ are referred to as structure factors. It follows from (12) that to determine the crystal structure it is sufficient to know structure factors corresponding to Bragg reflections.

3.1.2. Fourier transform of a periodic function

For a periodic function the Fourier transform does not exist in the framework of classical functions, but may be defined in term of generalized functions (distributions) [18]. The periodic electron density distribution $\rho^{cr}(\mathbf{r})$ may be represented as

$$\rho^{cr}\left(\mathbf{r}\right) = \sum_{\mathbf{u}\in\mathfrak{R}} \rho_{V}\left(\mathbf{r}-\mathbf{u}\right) = \rho_{V} * \sum_{\mathbf{u}\in\mathfrak{R}} \delta_{\mathbf{u}}, \ \mathbf{r}\in\mathbf{R}^{3},$$
(14)

where $\delta_{\mathbf{u}}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{u})$ is the three-dimensional Dirac δ -function localized in the point \mathbf{u} , and the sign * denotes the convolution of two functions. Let apply Fourier transform to both sides of identity (14). Using the property of Fourier transform to transform convolution into product and the property to transform \Re -grid of δ -functions into \Re' -grid, we get

$$\mathbf{F}(\mathbf{s}) = \mathbf{F}_{V}(\mathbf{s}) \sum_{\mathbf{u} \in \mathfrak{R}'} \delta_{\mathbf{u}} = \sum_{\mathbf{u} \in \mathfrak{R}'} \mathbf{F}_{V}(\mathbf{u}) \delta_{\mathbf{u}}, \ \mathbf{s} \in \mathbf{R}^{3}.$$
(15)

It follows from the last formula that, in the case of the crystal structure, the Fourier transform exists only as a generalized function and is equal to zero at points corresponding to non-Bragg reflections. At the points corresponding to Bragg reflections, it is the delta-functions modulated by the values of the corresponding structure factors. Thus, the crystal significantly enhances the magnitudes of the Fourier transform of $\rho_v(\mathbf{r})$ function for Bragg reflections. At the same time the values corresponding to non-Bragg reflections are totally lost.

3.2. A finite crystal

Let consider now a real crystal of finite dimensions. Let Q be a finite set of nodes of the lattice \Re , and the crystal is composed of copies of the unit cell V shifted by different vectors **u** from the set Q. The electron density distribution in the crystal can be represented as

$$\rho(\mathbf{r}) = \sum_{\mathbf{u} \in \mathcal{Q}} \rho_V(\mathbf{r} - \mathbf{u}) , \mathbf{r} \in \mathbf{R}^3 , \qquad (16)$$

instead of (8). In this case, the formula (15) for the Fourier transform of $\rho(\mathbf{r})$ takes the form

$$\mathbf{F}(\mathbf{s}) = \mathbf{F}_{V}(\mathbf{s}) \ \Omega_{\mathcal{Q}}(\mathbf{s}), \mathbf{s} \in \mathbf{R}^{3},$$
(17)

where

$$\Omega_{Q}(\mathbf{s}) = \sum_{\mathbf{u} \in Q} \exp(i2\pi \mathbf{s} \cdot \mathbf{u}), \, \mathbf{s} \in \mathbf{R}^{3} .$$
(18)

Accordingly, the intensity of the reflections can be calculated as

$$\left|\mathbf{F}(\mathbf{s})\right|^{2} = \left|\mathbf{F}_{V}(\mathbf{s})\right|^{2} \left|\Omega_{Q}(\mathbf{s})\right|^{2} .$$
⁽¹⁹⁾

The function $\Omega_{\varrho}(s)$ is the periodical function with the same periods $\{a^*, b^*, c^*\}$ as the lattice of δ -functions $\sum_{u \in \Re'} \delta_u$, but is a 'classical' function, and depends on the size and shape of

the sample. For all Bragg reflections, the value of this function is the same and equal to the number of unit cells in the crystal. If the crystal is a parallelepiped composed from $N_x \times N_y \times N_z$ cells, the computation yields

$$\left|\Omega_{Q}\left(h\mathbf{a}^{*}+k\mathbf{b}^{*}+l\mathbf{c}^{*}\right)\right|^{2}=\frac{\sin^{2}\left(\pi N_{x}h\right)}{\sin^{2}\left(\pi h\right)}\cdot\frac{\sin^{2}\left(\pi N_{y}k\right)}{\sin^{2}\left(\pi k\right)}\cdot\frac{\sin^{2}\left(\pi N_{z}l\right)}{\sin^{2}\left(\pi l\right)}.$$
(20)

For integer *h*, *k*, *l* we have $|\Omega_Q|^2 = (N_x N_y N_z)^2$, so that the energy of diffracted waves of the Bragg reflections increases $(N_x N_y N_z)^2$ times in comparison with the single unit cell diffraction. For example, if the crystal has the size 0.1 MM and the unit cell has the size of 100 Å, then, $N_x = N_y = N_z = 10^4$, i.e. the intensity gain is 10^{24} , which makes possible the practical registration of scattered rays.

The identity (19) leads to the several important conclusions:

• for a finite crystal, Fourier transform of the electron density distribution is proportional to the structure factors with a multiplier that depends only on the size and the external shape of the sample;

• for each Bragg reflection, the proportionality coefficient is the same and is equal to the number of the cells involved in the scattering; this makes it possible to observe the intensities of reflections $I(\mathbf{s}) = |\mathbf{F}_{V}(\mathbf{s})|^{2}$ in the same, relative scale;

• for the intermediate (non-Bragg) scattering vectors, the proportionality coefficient $Q(\mathbf{s})$ is small and is practically random; this prevents the measurement of the intensities of non-Bragg reflections.

4. SCATTERING BY A SINGLE PARTICLE

Let consider now the case when the sample is a particle with no strong periodicity. In this case, the Fourier transform does not have the shape of δ -functions, and reflections intensities are distributed more smoothly. At the same time, the effect of the strengthening of reflections is absent, which makes the task of their measurement extremely difficult. Only in the last decade, progress in the development of experimental techniques has made it possible to practically obtain X-ray images of isolated particles, both when working with a synchrotron accelerator [19] and when using X-ray lasers [20, 21].

Electron density distribution for the isolated particle $\rho^{sp}(\mathbf{r})$ is not equal to zero only in the limited region *M* of the space, which we call the molecule region:

$$M = \operatorname{supp} \rho^{sp} \left(\mathbf{r} \right) = \left\{ \mathbf{r} \in \mathbf{R}^{3} : \rho^{sp} \left(\mathbf{r} \right) \neq 0 \right\} .$$
(21)

For an arbitrarily chosen basis {**a**, **b**, **c**}, we define, as before , the unit cell *V*, the restriction $\rho_V(\mathbf{r})$ of the function $\rho^{sp}(\mathbf{r})$ to the region *V*, the conjugate basis and the lattices of the direct and reciprocal spaces.

Suppose that the cell V is chosen large enough so that (with the proper selection of the origin and the placement of the center of the cell in it) the entire region of the molecule is contained inside the cell

$$M \subset V. \tag{22}$$

In this case, the functions $\rho_V(\mathbf{r})$ and $\rho^{sp}(\mathbf{r})$ coincide, and the structure factors corresponding to $\rho_V(\mathbf{r})$ coincide with the Fourier transform of the function $\rho^{sp}(\mathbf{r})$

$$\mathbf{F}_{V}(\mathbf{s}) = \int_{V} \rho_{V}(\mathbf{r}) \exp(i2\pi\mathbf{s}\cdot\mathbf{r}) dV_{\mathbf{r}} = \int_{\mathbf{R}^{3}} \rho(\mathbf{r}) \exp(i2\pi\mathbf{s}\cdot\mathbf{r}) dV_{\mathbf{r}} = \mathbf{F}(\mathbf{s}).$$
(23)

As before, the function $\rho^{sp}(\mathbf{r})$ can be represented by Fourier series (12) inside the unit cell. However, this representation is not valid outside the region *V*, where the function $\rho^{sp}(\mathbf{r})$ is equal to zero. We define the periodic function $\rho^{cr}(\mathbf{r})$ by the equation (12) for all $\mathbf{r} \in \mathbf{R}^3$. We shall call this function the electron density distribution in a virtual crystal with the unit cell *V*. As before, the task is to determine the distribution $\rho^{cr}(\mathbf{r})$ based on the magnitudes of structure factors of Bragg reflections. From this point of view, the problem of the determination of the spatial structure of a sample from the diffraction pattern of an isolated particle is equivalent to the problem of the determination of the crystal structure [22, 23].

For a single particle (in contrast to the crystal case), it exists a considerable freedom in the choice of the unit cell basis {**a**, **b**, **c**} and, correspondingly, the conjugate basis {**a**^{*}, **b**^{*}, **c**^{*}}. The latter determines the steps of sampling of the experimental data, and correspondingly, the choice of a set of Bragg structure factors $\mathbf{F}(\mathbf{s}), \mathbf{s} \in \mathfrak{R}'$. It is only necessary that the condition (22) is valid. Let the diameter of the particle is estimated as D_{sp} , then it is enough to choose a cubic unit cell with the edge lengths not less than this diameter. This corresponds to the sampling of the continuous diffraction pattern $I(\mathbf{s}), \mathbf{s} \in \mathbf{R}^3$ on an orthogonal three-dimensional grid \mathfrak{R}' with a step of no greater than $1/D_{sp}$ along each axis. Note that there is also an experimental restriction on the minimum value of this step determined by the detector pixel size. The value of this minimum possible sampling step can be estimated as $\Delta s_{\min} \approx h/L\lambda$, where *h* is the pixel size, *L* is the distance from the sample to the detector, λ is the wavelength of the X-ray beam, and it is assumed that $h/L \ll 1$.

5. PATTERSON FUNCTION AND AUTOCORRELATION FUNCTION

5.1. Autocorrelation function

In contrast to electron density distribution, which requires the knowledge of the both magnitude and phase values, the associated autocorrelation function and Patterson function require the structure factors magnitudes only and can be calculated immediately from the observed information.

For any function $\rho = \rho(\mathbf{r})$, we denote $\breve{\rho}$ its enantiomer

$$\tilde{\rho}(\mathbf{r}) = \rho(-\mathbf{r}), \mathbf{r} \in \mathbf{R}^3$$
, (24)

and define the associated autocorrelation function by

$$A(\mathbf{r}) = \rho * \breve{\rho} = \int_{\mathbf{R}^3} \rho(\mathbf{u}) \rho(\mathbf{u} - \mathbf{r}) dV_{\mathbf{u}}, \ \mathbf{r} \in \mathbf{R}^3.$$
(25)

The Fourier transform of an autocorrelation function is the square of the magnitude of the Fourier transform of the distribution $\rho(\mathbf{r})$

$$\int_{\mathbf{R}^3} A(\mathbf{r}) \exp[i2\pi \mathbf{s} \cdot \mathbf{r}] dV_{\mathbf{r}} = \left| \mathbf{F}(\mathbf{s}) \right|^2 , \mathbf{s} \in \mathbf{R}^3.$$
(26)

To illustrate the relation between the autocorrelation function and the corresponding electron density distribution, consider the special case, namely an atomized density distribution, which is the sum of the contributions of identical atoms placed at different points in space.

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} g_0(\mathbf{r} - \mathbf{r}_j).$$
(27)

Here, $\{\mathbf{r}_j, j=1,...,N\}$ are atomic centers, $g_0(\mathbf{r})$ is a peak-like function describing the electron density distribution in a single atom placed at the origin, *N* is the number of atoms in the object. In this case, the autocorrelation function is expressed as

$$A(\mathbf{r}) = \sum_{j,k=1}^{N} g_2 \left(\mathbf{r} - \left(\mathbf{r}_j - \mathbf{r}_k \right) \right), \qquad (28)$$

where $g_2 = g_0 * \breve{g}_0$, so that the autocorrelation function is the sum of peak-like functions centered at the points corresponding to interatomic vectors $\mathbf{r}_j - \mathbf{r}_k$. Theoretically, the complete interpretation of the autocorrelation function, which is the determination of all N^2 interatomic vectors $\{\mathbf{r}_j - \mathbf{r}_k, j, k = 1, ..., N\}$, allows one to recover the values of the coordinates of the atoms $\{\mathbf{r}_j, j = 1, ..., N\}$. In the whole, this feature of the autocorrelation function remains also in the general case; the maxima of $A(\mathbf{r})$ correspond to the vectors between the maxima of the function $\rho(\mathbf{r})$.

In the case of a single particle, which means that $\rho^{sp}(\mathbf{r})$ function has a bounded support the autocorrelation function also has a bounded support. Suppose that the virtual unit cell *W* is chosen to be so large that it contains the support of the autocorrelation function, i.e. $\operatorname{supp} A \subset W$. Then, $I(\mathbf{s}) = |\mathbf{F}_{v}(\mathbf{s})|^{2} = |\mathbf{F}(\mathbf{s})|^{2}$, and the autocorrelation function can be represented inside the *W* as a Fourier series

$$A(\mathbf{r}) = \frac{1}{|W|} \sum_{\mathbf{s} \in \mathfrak{R}'} I^{obs}(\mathbf{s}) \exp\left[-i2\pi \mathbf{s} \cdot \mathbf{r}\right], \ \mathbf{r} \in V ,$$
(29)

where the quantities $I^{obs}(\mathbf{s})$ are directly measured in the diffraction experiment. A sufficient condition to satisfy supp $A \subset W$ is the choice of the unit cell dimensions twice large as the object size.

5.2. Patterson function

In the case of a crystal sample with the density distribution $\rho^{cr}(\mathbf{r})$, the unit cell cannot be chosen arbitrarily large, but is predetermined by the periods $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$. We define Patterson function as the sum of Fourier series

$$P_{V}(\mathbf{r}) = \frac{1}{|V|} \sum_{\mathbf{s} \in \mathfrak{R}'} |\mathbf{F}_{V}(\mathbf{s})|^{2} \exp[-i2\pi \mathbf{s} \cdot \mathbf{r}], \ \mathbf{r} \in \mathbf{R}^{3},$$
(30)

with $\mathbf{F}_{V}(\mathbf{s})$ values defined by (13). The Patterson function can be represented as

$$P_{V}(\mathbf{r}) = \rho^{cr} *_{V} \breve{\rho}^{cr} = \int_{V} \rho^{cr}(\mathbf{u}) \rho^{cr}(\mathbf{u} - \mathbf{r}) dV_{\mathbf{u}} , \ \mathbf{r} \in \mathbf{R}^{3},$$
(31)

where $*_{v}$ stands for period-based convolution of periodic functions defined as (31). In contrary to convolution (25), the integration extends here to the unit cell only.

509

If the support of function $\rho_v(\mathbf{r})$ is so small that the cell contains not only it, but also the support of the autocorrelation function

$$A_{V}(\mathbf{r}) = \rho_{V} * \breve{\rho}_{V} = \int_{\mathbf{R}^{3}} \rho_{V}(\mathbf{u}) \rho_{V}(\mathbf{u}-\mathbf{r}) dV_{\mathbf{u}}, \ \mathbf{r} \in \mathbf{R}^{3},$$
(32)

then the Patterson function and autocorrelation functions coincide in the unit cell. This situation is possible when working with isolated particles, but it is impossible in a real crystal due to the dense packing of the molecules.



Fig. 4. The autocorrelation function $A_V(\mathbf{r})$ and Patterson function $P_V(\mathbf{r})$ for one-dimensional crystal structure consisting of two atoms. The Patterson function in the unit cell V is the superposition of three shifted autocorrelation functions.

Since $\rho^{cr}(\mathbf{r}) = \rho_V(\mathbf{r})$ in *V*, and $\rho^{cr}(\mathbf{r})$ may be presented as

$$\rho^{cr}\left(\mathbf{r}\right) = \sum_{\mathbf{w}\in\mathfrak{R}} \rho_{V}\left(\mathbf{r} - \mathbf{w}\right), \qquad (33)$$

it follows from (31) and (33) that

$$P_{V}(\mathbf{r}) = \int_{\mathbf{R}^{3}} \rho_{V}(\mathbf{u}) \sum_{\mathbf{w} \in \Re} \rho_{V}(\mathbf{u} - \mathbf{r} - \mathbf{w}) dV_{\mathbf{u}} = \sum_{\mathbf{w} \in \Re} A_{V}(\mathbf{r} + \mathbf{w}) .$$
(34)

Taking into account that the size of the support supp A_V does not exceed twice the cell size, the latter identity has only a small number of members corresponding to the "tails" of the autocorrelation function $A_V(\mathbf{r})$ in the cells adjacent to V

$$P_{V}\left(\mathbf{r}\right) = \sum_{j,k,l=-1}^{1} A_{V}\left(\mathbf{r} - \left(j\mathbf{a} + k\mathbf{b} + l\mathbf{c}\right)\right), \ \mathbf{r} \in V \ .$$
(35)

Therefore, in the case of a crystal structure, the Patterson function is, generally speaking, an overlay of 27 shifted copies of the autocorrelation function (32) of the unit cell content. As a consequence, the presence of a peak of the Patterson function at the point **u** does not necessarily mean that the function $\rho_V(\mathbf{r})$ has this interatomic vector, but may mean the presence, as an interatomic vector, the vector **u** shifted by a period of the crystal (Fig. 4).

6. INTERPOLATION AND EXTRAPOLATION OF THE OBSERVED DATA

6.1. Interpolation of observed intensities

Let the virtual unit cell W with the basis $\{a, b, c\}$ be chosen so large that it contains the support of the autocorrelation function

$$\operatorname{supp} A \subset W . \tag{36}$$

In this case, the autocorrelation function inside the unit cell can be represented as the Fourier series (30), and is equal to zero outside the cell *W*. Therefore, by restricting the integration volume in (26) to the unit cell, substituting expression (30), and calculating the integrals, we obtain (taking into account that the center of the cell is the origin) the interpolation formula for an arbitrary point $\mathbf{u} \in \mathbf{R}^3$

$$I(\mathbf{u}) = \sum_{\mathbf{s}\in\mathfrak{R}'} \operatorname{sinc}(\mathbf{u}-\mathbf{s})I(\mathbf{s}), \ \mathbf{u}\in\mathbf{R}^3,$$
(37)

where

$$\operatorname{sinc}\left(h\mathbf{a}^{*}+k\mathbf{b}^{*}+l\mathbf{c}^{*}\right)=\frac{\sin\pi h}{\pi h}\cdot\frac{\sin\pi k}{\pi k}\cdot\frac{\sin\pi l}{\pi l}\,.$$
(38)

It follows from this formula that, under condition (36), the intensity value at any point in three-dimensional space is a linear combination of the intensities of the Bragg reflections with the fixed coefficients (which are independent on the function I(s)).

Formula (38) plays an important role in the theory of information and is usually associated with the names of V. Kotelnikov, E. Whittaker, C. Shannon, and H. Nyquist, although its early mention can be found in the mathematical works of E. Borel [24, 25]. The perspectives of using this formula in biological crystallography were discussed in the works of D. Sayre and G. Bricogne [26–28]. A sufficient condition for (36) is the choice of the cell edges to be twice the diameter of the particle, or, equivalently, the sampling step of the experimental data should be less than $1/2D_{sp}$. This step value is often referred to as the Nyquist limit.

It follows from formula (38) that the knowledge of Bragg reflection intensities with a sampling step of $1/2D_{sp}$ uniquely determines the intensities for all other reflections. However, it should be noted that the summation in (38) occurs throughout the complete infinite set of reflections. If the intensities of a part of Bragg reflections are unknown or are known with an error, the formula becomes approximate only. Then, each observed intensity value $I^{obs}(\mathbf{u})$ of the non-Bragg reflections results in the equation linking the unknown values of the Bragg reflections

$$\sum_{\mathbf{s}\in\mathfrak{R}'}\operatorname{sinc}(\mathbf{u}-\mathbf{s})I(\mathbf{s}) = I^{obs}(\mathbf{u}), \, \mathbf{u}\notin\mathfrak{R}',$$
(39)

and can be used for their determination or refinement [29].

6.2. Interpolation of structure factors

An interpolation formula similar to (37) can be written for structure factors as well. Let the unit cell satisfies the condition $\sup \rho \subset V$. Representing the function $\rho_V(\mathbf{r})$ inside the unit cell V as the Fourier series and then calculating the integral (3) for an arbitrary (non-Bragg) vector **u**, we obtain an interpolation formula for the complex structural factors

$$\mathbf{F}(\mathbf{u}) = \sum_{\mathbf{s}\in\mathfrak{R}'} \operatorname{sinc}(\mathbf{u} - \mathbf{s}) \mathbf{F}(\mathbf{s}), \, \mathbf{u} \in \mathbf{R}^3.$$
(40)

Suppose now that the unit cell is chosen to be the smallest possible cell V_{\min} among the cells containing inside the support of the function $\rho^{sp}(\mathbf{r})$. Given the measured structure factor magnitudes for the grid of the reciprocal space $\{F^{obs}(\mathbf{s}), \mathbf{s} \in \mathfrak{R}_{\min}^{\mathsf{r}}\}$ corresponding to this cell, the function $\rho^{sp}(\mathbf{r})$ is completely determined in V_{\min} , and, hence, in the whole space, by the values of the phases $\{\varphi(\mathbf{s}), \mathbf{s} \in \mathfrak{R}_{\min}^{\mathsf{r}}\}$. The presence of the measured structure factor magnitude $F^{obs}(\mathbf{u})$ with $\mathbf{u} \notin \mathfrak{R}_{\min}^{\mathsf{r}}$ leads to the equation that links the phase values

$$\left|\sum_{\mathbf{s}\in\mathfrak{H}_{\min}}\operatorname{sinc}(\mathbf{u}-\mathbf{s})F^{obs}(\mathbf{s})\exp[i\varphi(\mathbf{s})]\right| = F^{obs}(\mathbf{u}).$$
(41)

Considering similar equations for other non-Bragg reflections, we get a system of equations that links the unknown values of structure factor phases. These equations can be used to find the phase values of structure factors [30]. It should be noted that, as in the case of intensities, reducing in equations (41) the sampling step of non-Bragg vectors to a value less than half the edge of the cell $V_{\rm min}$ leads to a formal complex linear dependence of the coefficients of the equations. However, even using only vectors **u** with half-integer values of at least one of the coordinates gives a sevenfold excess of the number of equations (41) over the number of phases to be determined.

An alternative way to use the redundancy of the set of observed magnitudes values for solution of the phase problem is an *ab initio* mask-based approach developed by the authors in a series of papers [22, 23, 31–33].

6.3. Extrapolation of observed intensities

The importance of intensities of non-Bragg reflections is emphasized by the fundamental Schwartz's–Paley–Wiener theorem, which states that the Fourier transform of a function possessing of compact support is an entire holomorphic function. In our case, this, in particular, means that setting the function I(s) in an arbitrarily small ball B of the space \mathbb{R}^3 completely determines its values at all other points. Theoretically, such extrapolation can be performed, for example, by a Taylor series constructed at an inner point of the ball. Its construction requires partial derivative values that are fully determined by an arbitrarily small neighborhood of the decomposition center. The Taylor series converges at any other point due to the fact that the function I(s) is the entire function. However, this procedure is difficult to implement in practice. The construction of algorithms that allow to effectively extrapolate the values of the intensity of reflections to a wider area is a challenge for the methods of computational mathematics.

Two observations should be made. First, the theoretical possibility of an extrapolation of the values extends to all vectors $\mathbf{s} \in \mathbf{R}^3$ including vectors beyond the experimentally possible resolution $\lambda/2$. This removes the experimental restriction on the potential resolution of the results. On the other hand, the possible sampling step of the experimental data is limited from the bottom by detector pixel size. Therefore, the extrapolation task could be formulated in a narrower sense, namely, to expand the set of known intensities as far as possible, assuming that the intensities are known first, for a sufficiently finely sampled limited region *B* in the reciprocal space.

7. CONCLUSIONS

An important stage in the diffraction study of a biological macromolecule is the determination of the function that describes the density of the distribution of the scattering electrons in the object under study. The study of the full distribution in the three-dimensional space may be reduced to the analysis of the 'unit cell distribution' that is equal to zero outside a specified parallelepiped referred to as the unit cell. The content of the unit cell is repeated periodically in the case of a crystal sample and coincides with the full distribution in the case of a single particle. A standard diffraction experiment allows one to determine in a relative scale the intensity of reflections, which are the squared magnitudes of the complex coefficients (structure factors) in the expansion of the unit cell electron density distribution into the Fourier series. Both in a crystal and a single particle cases, the problem of the structure determination can be reformulated as the problem of restoring the values of the phases of structure factors under the assumption that their magnitudes are known from the experiment. The difference is that, in the case of a crystal structure, the unit cell basis (density distribution periods) is rigidly defined by the physical characteristics of the crystal and the experiment allows one to determine the values of the structure factor magnitudes only for reflections that have integer coordinates in the basis conjugated to the cell basis (Bragg reflections). In the absence of an additional information about the electron density distribution in the cell, all structure factors are mutually independent. The phase values of structure factors lost in the experiment are irreplaceable in the sense that they cannot be recovered without additional information (experimental or theoretical) about the object.

In the idealized case of an infinite crystal, the Fourier transform of the periodic electron density distribution in the crystal, which determines the diffraction pattern, does not exist within the framework of classical functions. But it can be defined in terms of the theory of generalized functions and is a Bragg lattice of delta functions modulated by structure factors. The situation changes radically for a single particle. In this case, the Fourier transform is an entire holomorphic function, which opens up opportunities for recovering the phase part of the Fourier transform and expanding the set of the experimentally recorded reflection intensities to a wider region.

The choice of unit cell, in the case of a single particle, is sufficiently arbitrary and must only satisfy the requirement that the virtual cell has to be large enough to accommodate the object. The set of Bragg reflection phases corresponding to the minimum possible unit cell, that still allows to place the object inside, is sufficient to restore the electron density distribution in the object. The use of an information about the Fourier transform magnitudes of non-Bragg reflections provides an additional information limiting the values of the phases of the structure factors of Bragg reflections. This information can be the basis for the recovery of unknown phase values.

An increase in the size of the unit cell means a reduction in the sampling step in the continuous diffraction pattern, which leads to an increase in the amount of the experimental data involved in the work. This makes it possible to determine the unknown values of the phases of the structure factors due to the excess of the set of magnitudes. If the size of the unit cell is larger than twice the object size, then the non-Bragg intensities are the linear combinations of the Bragg intensities with the predefined (and non-depending on the object) coefficients. Nevertheless, if a part of Brag reflections is unknown, but the intensity of some non-Brag reflection has been measured, then it leads to equation that links unknown Bragg intensities. Such equations may be used to restore the lost Bragg intensities. The choice of the unit cell dimensions that exceed twice or more the object size allows to calculate autocorrelation function in the unit cell by means of Fourier series, where the observed intensities are used as the series' coefficients.

The analytical properties of the Fourier transform in the case of an isolated particle give the theoretical possibility of an unlimited extrapolation of the intensity values from the central

zone of the three-dimensional space to the whole space. However, this possibility remains purely theoretical for biological crystallography and is currently not supported by working algorithms. The development of such algorithms is a challenge for specialists in the field of computational mathematics.

REFERENCES

- Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.-W., Kapral G.J., Grosse-Kunstleve R.W. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica D*. 2010. V. 66. P. 213–221. doi: 10.1107/S0907444909052925.
- Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G.W., McCoy A. et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica D*. 2011. V. 67. P. 235–242. doi: <u>10.1107/S0907444910045749</u>.
- 3. Sheldrick G.M. A short history of SHELX. *Acta Crystallographica A*. 2008. V. 64. P. 112–122. doi: <u>10.1107/S0108767307043930</u>.
- Bricogne G., Vonrhein C., Flensburg C., Schiltz M., Paciorek W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallographica D*. 2003. V. 59. P. 2023–2030. doi: <u>10.1107/S0907444903017694</u>.
- 5. Blanc E., Roversi P., Vonrhein C., Flensburg C., Lea S.M., Bricogne G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica D*. 2004. V. 60. P. 2210–2221. doi: <u>10.1107/S0907444904016427</u>.
- Minor W., Cymborowski M., Otwinowski Z., Chruszcz M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallographica D.* 2006. V. 62. P. 859–866. doi: <u>10.1107/S0907444906019949</u>.
- 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. doi: <u>10.1093/nar/28.1.235</u>.
- 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.
- 9. Spence J.C.H. XFELs for structure and dynamics in biology. *IUCrJ*. 2017. V. 4. P. 322–339. doi: <u>10.1107/S2052252517005760</u>.
- Standfuss J., Spence J. Serial crystallography at synchrotrons and X-ray lasers. *IUCrJ*. 2017. V. 4. P. 100–101. doi: <u>10.1107/S2052252517001877</u>.
- Aquila A., Barty A., Bostedt C., Boutet S., Carini G., dePonte D., Drell P., Doniach S., Downing K.H., Earnest T. et al. The linac coherent light source single particle imaging road map. *Structural Dynamics*. 2015. V. 2. doi: <u>10.1063/1.4918726</u>.
- 12. Ayyer K., Geloni G., Kocharyan V., Saldin E., Serkez S., Yefanov O., Zagorodnov I. Perspectives for imaging single protein molecules with the present design of the European XFEL. *Structural Dynamics*. 2015. V. 2. doi: <u>10.1063/1.4919301</u>.
- Daurer B.J., Okamoto K., Bielecki J., Maia F.R.N.C., Muhlig K., Seibert M.M., Hantke M.F., Nettelblad C., Benner W.H., Svenda M. et al. Experimental strategies for imaging bioparticles with femtosecond hard X-ray pulses. *IUCrJ*. 2017. V. 4. P. 251– 262. doi: <u>10.1107/S2052252517003591</u>.
- 14. Landau L.D., Lifshitz E.M. *Mechanics*, 3d edition. Butterworth-Heinemann; 1976. 224 p.
- 15. Landau L.D., Lifshitz E.M. *The Classical Theory of Fields*, 4th edition. Butterworth-Heinemann; 1980. 402 p.

- Urzhumtsev A.G., Lunin V.Y. Introduction to crystallographic refinement of macromolecular atomic models. *Crystallography Reviews*. 2019. V. 25. P. 164–262. doi: <u>10.1080/0889311X.2019.1631817</u>.
- Urzhumtseva L., Klaholz B., Urzhumtsev A. On effective and optical resolutions of diffraction data sets. *Acta Crystallographica D*. 2013. V. 69. P. 1921–1934. doi: <u>10.1107/S0907444913016673</u>.
- 18. Rudin W Functional analysis. McGRAW-HILL BOOK COMPANY, 1973.
- Rodriguez J.A., Xu R., Chen C.-C., Huang Z., Jiang H., Chen A.L., Raines K.S., Pryor Jr.A., Nam D., Wiegart L. et al. Three-dimensional coherent X-ray diffractive imaging of whole frozen-hydrated cells. *IUCrJ*. 2015. V. 2. P. 575–583. doi: 10.1107/S205225251501235X.
- 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. doi: <u>10.1103/PhysRevLett.114.098102</u>.
- 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. *Scientific Data*. 2016. V. 3. doi: <u>10.1038/sdata.2016.64</u>.
- 22. Lunin V.Y., Lunina N.L., Petrova T.E., Baumstark M.W., Urzhumtsev A.G. Maskbased approach to phasing of single-particle diffraction data. *Acta Crystallographica D*. 2016. V. 72. P. 147–157. doi: <u>10.1107/S2059798315022652</u>.
- Lunin V.Y., Lunina N.L., Petrova T.E., Baumstark M.W., Urzhumtsev A.G. Maskbased approach to phasing of single-particle diffraction data. II. Likelihood-based selection criteria. *Acta Crystallographica D*. 2019. V. 75. P. 79–89. doi: <u>10.1107/S2059798318016959</u>.
- 24. Meijering E. A chronology of interpolation: from ancient astronomy to modern signal and image processing. *Proceedings of the IEEE*. 2002. V. 90. P. 319–342. doi: 10.1109/5.993400.
- 25. Kotel'nikov V.A. On the transmission capacity of 'ether' and wire in electric communications. *Physics–Uspekhi*. 2006. V. 49. № 7. P. 736–744. doi: 10.1070/PU2006v049n07ABEH006160.
- 26. Sayre D. Some implications of a theorem due to Shannon. *Acta Crystallographica*. 1952. V. 5. P. 843. doi: <u>10.1107/S0365110X52002276</u>.
- Bricogne G. Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallographica A*. 1974. V. 30. P. 395–405. doi: 10.1107/S0567739474010722.
- Bricogne G. Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallographica A*. 1976. V. 32. P. 832–847. doi: <u>10.1107/S0567739476001691</u>.
- 29. Lunin V.Y., Lunina N.L. Repairing of the diffraction pattern in the X-ray freeelectron laser study of biological particles. *Advanced Mathematical Models & Applications*. 2018. V. 3. P. 117–127.
- Misnovs A., Mishnev A. On phasing of oversampled diffraction data. In: 32-nd European Crystallographic Meeting (Vienna, Austria, 18.-23.08): Book of abstracts. 2019. P. 706. URL: <u>https://ecm2019.org/fileadmin/user_upload/k_ecm2019/images/</u> <u>Programm/ECM32AbstractBooklet_18.08.2019.pdf</u> (accessed 06.11.2019).
- 31. 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. № Suppl. P. t1–t19. doi: 10.17537/2015.10.t1.

- 32. Lunina N.L., Petrova T.E., Urzhumtsev A.G., Lunin V.Y. The use of connected masks for reconstructing the single particle image from X-ray diffraction data. II. The dependence of the accuracy of the solution on the sampling step of experimental data. *Mathematical Biology and Bioinformatics*. 2015. V. 10. № Suppl. P. t56–t72. doi: 10.17537/2015.10.t56.
- Lunina N.L., Petrova T.E., Urzhumtsev A.G., Lunin V.Y. 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. *Mathematical Biology and Bioinformatics*. 2018. V. 13. № Suppl. P. t70–t83. doi: <u>10.17537/2018.13.t70</u>.

Received 25.09.2019. Revised 27.10.2019. Published 07.11.2019.