Advanced Mathematical Models & Applications
Vol.3, No.2, 2018, pp.117-127

REPAIRING OF THE DIFFRACTION PATTERN IN THE X-RAY FREE-ELECTRON LASER STUDY OF BIOLOGICAL PARTICLES

Vladimir Y. Lunin[1*], **Natalia L. Lunina**[1]

[1]Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of RAS, Moscow, Russia

Abstract. X-ray diffraction structure analysis is the main tool that allows obtaining information on the structure of biological micromolecular objects with atomic resolution. Until now, the main limitations of the method are the need to prepare a sample in the form of a crystal, and the loss in the experiment of a substantial part of the information necessary for the reconstruction of the object under study. The emergence of new powerful X-ray sources, namely X-ray Free Electron Lasers, creates a potential opportunity to get rid of both limitations. Realization of this potential in practice is a challenge for computational methods. This paper discusses possible approaches to solving one of the emerging problems, namely the restoring of lost in the experiment information about the intensity of a part of the scattered waves using the measured intensities of the remaining waves. Theoretically this problem could be solved, as the intensity is an entire holomorphic function, but encounters difficulties in practical implementation.

Keywords: biological crystallography, single particle, X-ray Free Electron Laser.
AMS Subject Classification: 92-08, 92C05.
Corresponding author: Vladimir Y., Lunin, Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, 1 Professor Vitkevich St., Pushchino, Moscow Region, 142290, Russia, Tel.: +7 496 7318535, e-mail: *lunin@impb.ru*
Received: 16 July 2018; Revised: 16 August 2018; *Accepted: 25 August 2018*; *Published: 31 August 2018*

1 Introduction

In the region of molecular biology X-ray diffraction structure analysis is applied to study of proteins, viruses, DNA, RNA, their complexes, and molecular machines (such as the ribosome). In recent years, attempts have been made to apply it to the study of cellular organelles and whole cells (Rodrigues et al., 2015; Lunin et al., 2017). At the final stage of the study, the determination of a structure implies the determination of the Cartesian coordinates of all (sometimes tens of thousands) atoms that compose the molecule and of certain additional parameters of atoms, for example, quantities characterizing the degree of uncertainty (dynamic and static) of each atom. At present, atomic structures of more than one hundred thousand biological macromolecules have been determined and deposited in Protein Data Bank (Berman, 2000; https://www.rcsb.org/). However, at initial stages, the goal is usually the electron density distribution function $\rho(\mathbf{r})$, that describes the spatial distribution of electrons in the molecule. With sufficient accuracy of finding this function, the positions in the space of its maxima can be interpreted as centers of atoms.

The X-ray diffraction analysis is based on the study of X-ray scattering pattern and appears under different names, such as X-ray diffraction analysis, biological crystallography, coherent diffraction imaging, X-ray microscopy, and so on. Within the framework of the kinematic theory of diffraction, it can be described as follows. The object under study is placed in an incident beam of X-rays, that can be considered as a plane electromagnetic wave or as a wave of force act-

ing on the object's electrons. Under the influence of periodically changing forces, the electrons oscillate. Each oscillating electron becomes a source of a secondary spherical electromagnetic wave propagating in all directions of the three-dimensional space. At the detector, the secondary waves superpose, forming the wave scattered by the object toward the detector. This wave is historically, referred to as the reflection. The detector measures the energy of this wave, that is proportional to the square of its amplitude. The amplitude of the scattered wave depends on the phase differences in the secondary waves, which, in turn, are determined by the mutual arrangement in the space of the oscillating electrons. The successive solution of Newton's equations of the motion of an electron under the influence of an electromagnetic wave and Maxwell's equations of the scattering by the oscillating electron allows to represent the complex amplitude $\mathbf{E}$ of the scattered wave as

$$\mathbf{E} = \varepsilon \mathbf{E}_0 \mathbf{F}(\mathbf{s}). \tag{1}$$

Here ε is the combination of physical constants and parameters of the experimental setup, $\mathbf{E}_0$ is the amplitude of the incident wave. The complex value $\mathbf{F}(\mathbf{s}) = F(\mathbf{s}) \exp[i\varphi(\mathbf{s})]]$, called as the structure factor, is determined by the distribution of the electron density in the object under study and the geometry of the experiment by means of the equalities

$$\mathbf{F}(\mathbf{s}) = \int\limits_{\mathbf{R}^3} \rho(\mathbf{r}) \exp[2\pi i \mathbf{s} \cdot \mathbf{r}] \, dV_{\mathbf{r}}, \quad \mathbf{s} \in \mathbf{R}^3, \tag{2}$$

$$\mathbf{s} = \frac{\sigma - \sigma_0}{\lambda}. \tag{3}$$

Here $\rho(\mathbf{r})$ is the function describing the density of the space distribution of the electrons in the sample under study; σ_0, σ are the vectors of unit length defining the direction of propagation of the incident wave and the direction from the sample to the detector; λ is the wavelength of the used X-ray radiation (it is about 1Å in biological crystallography applications); $\mathbf{s} \cdot \mathbf{r}$ denotes the scalar product of the vectors $\mathbf{s}$ and $\mathbf{r}$. The vector $\mathbf{s}$ defined by (3) is called the scattering vector. It identifies the scattered wave, and the term reflection is often applied directly to the scattering vector. Equations (1-3) form the basis for calculations in the kinematic theory of diffraction. Following to crystallographic traditions we call the argument $\varphi(\mathbf{s})$ of the complex number $\mathbf{F}(\mathbf{s})$ the phase.

The values immediately obtained in the experiment are the energies of scattered waves that are proportional to the squares of the magnitudes of the corresponding complex structure factors. The value $I(\mathbf{s}) = |\mathbf{F}(\mathbf{s})|^2$ will be referred to as the intensity of the reflection $\mathbf{s}$, and the full set of intensities $\left\{I(\mathbf{s}), \ \mathbf{s} \in \mathbf{R}^3\right\}$ will be referred as the diffraction pattern. The change of the detector position and the direction of the incident X-ray beam relative to the object makes it possible to have different values of the vector $\mathbf{s}$ and thereby determine experimentally intensities for different vectors in the three-dimensional space. In biological crystallography the value inverse to the scattering vector length $d = 1/|\mathbf{s}|$ is referred to as the reflection resolution. It is equal to the distance between the maxima planes of the real and imaginary parts of the three-dimensional Fourier harmonic $\exp[2\pi i \mathbf{s} \cdot \mathbf{r}]$. Let S be the set of vectors $\mathbf{s}$, for which the intensity values $I(\mathbf{s})$ have been observed. Since σ_0, σ are the vectors of unit length, the set S is bounded by the theoretical limit $|\mathbf{s}| \leq 2/\lambda$. In practice, the set of reflections for which it is possible to observe $I(\mathbf{s})$ experimentally can be essentially reduced in comparison with this theoretical limit. It is customarily said that the study is carried out at resolution d_{min}, if all (or almost all) reflection with $|\mathbf{s}| \leq s_{max} = 1/d_{min}$ are present in the set S. The resolution is an important characteristic of the results obtained. It determines the characteristic dimensions of the visible details of the structure and the accuracy of structure determination.

Equation (2) makes it possible to obtain the electron density distribution in the object by means of the inverse Fourier transform provided that both the magnitudes and phases of the

complex structure factors $\mathbf{F}(\mathbf{s})$ are known for all points in the real three-dimensional space $\mathbf{R}^3$

$$\rho(\mathbf{r}) = \int\limits_{\mathbf{R}^3} \mathbf{F}(\mathbf{s}) \exp[-2\pi i \mathbf{s}\cdot\mathbf{r}] dV_{\mathbf{s}} \ , \quad \mathbf{r} \in \mathbf{R}^3 \ . \tag{4}$$

Since only magnitudes of structure factors can be observed in the experiment, this poses the phase problem of X-ray structure analysis, namely, to determine the phase part of the complex Fourier transform provided its magnitude is known. It is clear that in a such general form the problem is meaningless, as any function $\varphi(\mathbf{s})$ can be taken as a solution. In order to make the task reasonable, the class of admissible electron density distributions or the class of admissible sets of phases must be restricted. Two classes of distributions which are of the greatest interest in crystallography are the functions with a compact support (describing individual biological particles) and triple-periodic functions (describing distributions of the electron density in crystals). In this paper we will be interested in the first class of functions describing the distribution of the electron density in a single particle.

The second problem that arises when the electron density distribution is restored by the formula (4) is that even the magnitudes of structure factors are known only for a limited set of reflections S. An exclusion from the calculation of the transform (4) of high-resolution structure factors (the ones corresponding to high-frequency Fourier harmonics) leads to image distortions, manifested as blurring and merging of the peaks of the electron density, as well as the appearance of series termination waves. Therefore, the result obtained in practice is an approximation to the distribution of electron density with a resolution determined by the resolution of the experimental set S. This approximation is referred to in crystallography as Fourier synthesis of the electron density. The approximation found is used to obtain the map that is a graphical representation of iso-surfaces of this approximation. Depending on resolution of the used data set, such maps allow to see the general outlines of the object, the course of the polypeptide chain, the location of the side chains and, in favorable cases, the peaks of the electron density, corresponding to individual atoms. These maps are used to construct the atomic model of the object. The process of constructing a model from electron density maps is referred to as the interpretation of the electron density distribution. Especially significant distortions in the image of an object in electron density maps arise when structure factors of the central zone $|\mathbf{s}| \le s_{min}$ or structure factors for which the values of $F(\mathbf{s})$ are very large are absent. The lack of such information may be caused by the features of experimental equipment and is one of the significant obstacles in the way of obtaining an interpretable map of the electron density distribution. Below we will consider approaches to overcoming this problem.

The parameter ε in equation (1) includes the physical constants and parameters of the experimental setup. Under the conditions of a real experiment, it can be estimated to be of the order of 10^{-12}. The registration of such weak waves is a serious experimental problem solved since the discovery of X-ray diffraction at the very beginning of the 20th century. Two obvious ways to solve this problem are the increasing of the power of the X-ray source and sensitivity of the detector. At the present time, such powerful sources of X-ray radiation as synchrotron accelerators are used at the final stages of X-ray study. The main type of detector, currently used in biological crystallography, are two-dimensional matrix detectors, allowing simultaneous recording of a set of intensities for a two-dimensional set of reflections. This two-dimensional set of intensities is referred to as the frame or the image. Rotation of the object allows to get a series of frames, that cover a three-dimensional domain of scattering vectors. The sensitivity of modern detectors is brought to the level of registration of one X-ray photon incident on the pixel of the detector. However, despite all the advances in these areas, until recently the recording of scattering by a separate biological macromolecule remained impossible. The only way to make the intensity measurable was to prepare a crystal of the studied object. In a crystal, a lot of identical and similarly oriented molecules are arranged into three-dimensional regular lattice, defined by the lattice periods $\{\mathbf{a},\ \mathbf{b}, \mathbf{c}\}$. For a middle-size protein and typical for biological crystallography crystal dimensions (0.1 mm) the number of molecules in the crystal

may be estimated as 10^{12}. Depending on the direction of the scattering the waves scattered by different copies of the molecule may come all with the same phase or with different phases. In the first case the amplitude of the resulting wave grows proportionally to the number of copies of the molecule and becomes measurable. In the other case the amplitude of the summary wave remains immeasurable. The condition of a measurable reflection (a Bragg reflection) may be written as

$$\mathbf{s}\cdot\mathbf{a}=h,\ \mathbf{s}\cdot\mathbf{b}=k,\ \mathbf{s}\cdot\mathbf{c}=l,\qquad h,k,l\in\mathbf{Z}, \tag{5}$$

where $\mathbf{Z}$ denotes the integer numbers. The use of crystals makes it possible to measure the intensities for a discrete subset of the scattering vectors, however, the loss of a significant part of the information significantly complicates the solution of the phase problem.

The situation has changed significantly with the advent of new powerful X-ray sources, namely X-ray Free-Electron Lasers (XFEL), in combination with a new scheme for organizing the experiment, referred to as serial crystallography (Ayyer et al., 2015; Spence, 2017). XFEL provides with extremely powerful and short (10-50 femtoseconds) pulses. Under the influence of such a pulse, the biological particle eventually destroys, but during the time of the action of the pulse it is possible to measure the diffraction of yet undisturbed particle (diffraction before destruction). This result in one two-dimensional frame of intensities of scattered waves. In order to assemble a three-dimensional data set, a series of randomly oriented identical particles is used, alternately directed to the point of meeting with the pulse. Unlike traditional crystallography, where the orientation of the object changes in a controlled manner, in the serial crystallography, the mutual orientations of the exposed particles are unknown. This creates a problem when converting a set of two-dimensional frames into a single three-dimensional data set (Loh & Elser, 2009). The success of such a procedure essentially depends on the completeness of the data set collected for each of the frames. Theoretically, the size of this set is limited only by the wavelength λ of the X-rays beam through the restriction $|s|\leq 2/\lambda$, but in practice, due to the rapid decrease in intensity with increasing $|\mathbf{s}|$, this limit is much smaller. In addition, there are a number of additional factors limiting the S-set. First, central-zone reflections (for which scattering directions are close to the direction of the incident beam) may be absent. Second, there may be absent whole bands of reflections that fall on the joints of blocks from which the detector consists. Third, the intensity of some particularly strong reflections can go beyond the sensitivity limits of the detector (oversaturated reflections). The absence of intensity values, and as a consequence, the absence of the corresponding structure factors $\mathbf{F}(\mathbf{s})$ in calculation of the electron density distribution can strongly distort the Fourier synthesis maps and significantly complicate their interpretation. Below we discuss ways to restore, at least in part, intensities lost in the experiment by means of the use of Whittaker-Nyquist-Kotelnikov-Shannon interpolation formula and its extensions. This formula is broadly used in signal processing and it was previously used in crystallographic applications for interpolation of complex structure factors values (Sayre, 1952; Bricogne,1976). We discuss its usefulness for repairing of the missed intensity values. Some other ways to solve this problem were discussed earlier in (Lunin 1988; Urzhumtsev et al., 1989; Urzhumtsev, 1991).

2 Whittaker-Nyquist-Kotelnikov-Shannon interpolation formula

Let the function $\rho(\mathbf{r})$ defined by (4) describes electron density distribution in an isolated particle and so it has a compact support. We define its autocorrelation function

$$P(\mathbf{r})=\int_{\mathbf{R}^3}\rho(\mathbf{u})\,\rho(\mathbf{u}-\mathbf{r})dV_{\mathbf{u}} \tag{6}$$

that has a compact support as well. This function is referred to as Patterson function in crystallography. The Fourier transform of this function is the intensity $I(\mathbf{s})$

$$\int_{\mathbf{R}^3}P(\mathbf{r})\exp\left[2\pi i\mathbf{s}r\right]dV_{\mathbf{r}}=\left|\mathbf{F}(\mathbf{s})\right|^2=I(\mathbf{s}) \tag{7}$$

that is immediately determined in the X-ray diffraction experiment. It should be emphasized that, unlike the complex values $\mathbf{F}(\mathbf{s})$ of the structure factors, for which the experiment provides with the values of magnitudes only, the real nonnegative values $I(\mathbf{s})$ may be completely determined in the experiment. So that, it can be said that the experiment allows to calculate the Patterson function

$$P_S(\mathbf{r}) = \int_S I(\mathbf{s}) \exp[-2\pi i \mathbf{s}\cdot\mathbf{r}] dV_{\mathbf{s}}\ , \quad \mathbf{r}\in\mathbf{R}^3 \tag{8}$$

with an accuracy determined by the completeness of the set S of scattering vectors for which the intensities $I(\mathbf{s})$ have been obtained experimentally.

At the same time, since the function $P(\mathbf{r})$ has the compact support, then, by the Paley-Wiener theorems, its Fourier transform $I(\mathbf{s})$ can be extended to entire function in the complex three-dimensional space $\mathbf{C}^3$. This means a strong interconnection of the values of this function taken at different points in $\mathbf{R}^3$ and, in theory, gives the possibility of restoring the missing values, on the basis of the set of available intensities. For example, if the values of this function are known in a certain sphere in $\mathbf{R}^3$, then they determine uniquely the values of this functions in all points in $\mathbf{C}^3$, for example, by means of the Taylor series. Unfortunately, such an opportunity while exists theoretically, cannot be easily implemented in practice. Such disbalance between theoretical expectations and practical possibilities is a challenge for computational methods. Below we will discuss some ways of partially solving this problem. We discuss the problem of repairing the diffraction pattern using the Whittaker-Nyquist- Kotelnikov-Shannon interpolation formula.

Let the basis $\{\mathbf{a},\ \mathbf{b},\ \mathbf{c}\ \}$ in $\mathbf{R}^3$ and the origin are chosen so that the support of $P(\mathbf{r})$ is contained inside parallelepiped V (the unit cell), build on these vectors, and the origin is placed to the center of V

$$V = \left\{ x_1\mathbf{a} + x_2\mathbf{b} + x_3\mathbf{c} :\ -\frac{1}{2} < x_1, x_2, x_3 \le \frac{1}{2} \right\}\ , \quad \mathrm{supp}P(\mathbf{r}) \subset V. \tag{9}$$

Let $\{\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*\}$ be the conjugate basis, and $\mathfrak{R}'$ be the lattice, build on this basis

$$\mathfrak{R}' = \{k_1\mathbf{a}^* + k_2\mathbf{b}^* + k_3\mathbf{c}^* : k_1, k_2, k_3 \in \mathbf{Z}\}, \tag{10}$$

where $\mathbf{Z}$ denotes the integers. The coordinates of the scattering vector in the conjugate basis will be called reflection indexes. The lattice vectors have integer indexes. Inside the cell V the function $P(\mathbf{r})$ can be represented as the Fourier series

$$\mathrm{P}(\mathbf{r}) = \frac{1}{|\mathrm{V}|} \sum_{\mathbf{k}\in\mathfrak{R}'} I(\mathbf{k}) \exp[-2\pi i \mathbf{k}\mathbf{r}], \quad \mathbf{r}\in V, \tag{11}$$

where

$$I(\mathbf{k}) = \int_V P(\mathbf{r}) \exp[2\pi i \mathbf{k} r]\, dV_{\mathbf{r}} = \int_{\mathbf{R}^3} P(\mathbf{r}) \exp[2\pi i \mathbf{k} r]\, dV_{\mathbf{r}}\ , \quad \mathbf{k}\in\mathfrak{R}'. \tag{12}$$

A discrete set of intensities $\{I(\mathbf{k}) : \mathbf{k}\in\mathfrak{R}'\}$ with integer indexes will be called cardinal intensities. This set completely defines the function $P(\mathbf{r})$, and hence the values of the remaining (inferior) intensities $I(\mathbf{s})$ for all other vectors $\mathbf{s}\in\mathbf{R}^3$:

$$I(\mathbf{s}) = \int_{\mathbf{R}^3} P(\mathbf{r}) \exp[2\pi i \mathbf{s} r]\, dV_{\mathbf{r}} = \int_V P(\mathbf{r}) \exp[2\pi i \mathbf{s} r]\, dV_{\mathbf{r}}\ , \quad \mathbf{s}\in R^3. \tag{13}$$

Substitution of $P(\mathbf{r})$ in (13) by its value defined in (11) results in Whittaker-Nyquist-Kotelnikov-Shannon interpolation formula:

$$\mathrm{I}(\mathbf{s}) = \sum_{\mathbf{k}\in\mathfrak{R}'} \gamma(\mathbf{s}-\mathbf{k}) I(\mathbf{k}), \quad \mathbf{s}\in\mathbf{R}^3, \tag{14}$$

where

$$\gamma\left(\mathbf{h}\right)=\frac{1}{|V|}\int_{V}\exp\left[2\pi i\mathbf{h}r\right]dV_{\mathbf{r}},\quad \mathbf{h}\in\mathbf{R}^{3}. \tag{15}$$

The calculation of the integral, taking into account the condition (9), leads to

$$\begin{gathered}\gamma(\mathrm{h}_{1}\mathbf{a}^{*}+h_{2}\mathbf{b}^{*}+h_{3}\mathbf{c}^{*})=sinc(h_{1})sinc(h_{2})sinc(h_{3}),\\ sinc(x)=\tfrac{\sin(\pi x)}{\pi x}.\end{gathered} \tag{16}$$

This function is equal to zero for non-zero integer values of the argument. For the vector $\mathbf{h}=h_{1}\mathbf{a}^{*}+h_{2}\mathbf{b}^{*}+h_{3}\mathbf{c}^{*}$ let denote

$$\begin{gathered}\mathbf{H}=(\mathrm{h}_{1},h_{2},h_{3})^{T},\quad I_{\mathbf{H}}=I(\mathbf{h}),\\ \mathbf{sinc}(\mathbf{H})=sinc(h_{1})sinc(h_{2})sinc(h_{3}).\end{gathered} \tag{17}$$

The interpolation formula can now be represented in the form

$$I_{\mathbf{H}}=\sum_{\mathbf{K}\in\mathbf{Z}^{3}}\mathbf{sinc}\left(\mathbf{H}-\mathbf{K}\right)I_{\mathbf{K}},\quad \mathbf{H}\in\mathbf{R}^{3}. \tag{18}$$

The interpolation formula makes it possible to reconstruct the inferior intensities, assuming that the cardinal intensities are known. Since inferior intensities are expressed through the cardinal ones, they are sometimes treated as redundant, not carrying additional information. However, such an interpretation is valid only in the ideal case, when the full infinite set of cardinal intensities is known. In real cases, this set is limited, and the formula (18) is not exact more and turns into an equation restricting the values of unknown cardinal intensities, provided that inferior ones are known.

3 Repairing of cardinal intensities

In the case when one of the cardinal intensities is unknown, formula (18) does not allow obtaining it value, since for integer indexes $\mathbf{H}$ the formula degenerates into the tautology $I_{\mathbf{H}}=I_{\mathbf{H}}$. We now derive a modification of the interpolation formula that allows to reconstruct the missing cardinal intensities from the known values of inferior ones. For this we will use the available freedom in choosing the sampling basis $\{\mathbf{a}^{*},\mathbf{b}^{*},\mathbf{c}^{*}\}$. Some restriction on the choice of this basis is imposed by the condition (9). The cell V build on the conjugate basis $\{\mathbf{a},\ \mathbf{b},\ \mathbf{c}\ \}$ must be large enough to contain the support of the function $P\left(\mathbf{r}\right)$. In the one-dimensional case, this limits the minimal cell size by the diameter of the support of $P\left(\mathbf{r}\right)$, and the discretization step by the inverse of this value. This maximal possible discretization step, with which interpolation formula (18) is still valid is called the Nyquist limit. Sometimes this limit is interpreted as a rigid recommendation on the choice of the step of discretization, motivating it by the fact that for a larger step the cardinal intensities become dependent. However, all this is true only if the full (infinite) series of cardinal values is present. In the absence of a part of such values, the freedom in choosing of discretization basis can serve as a tool for restoring the lost values of cardinal intensities. When the sampling basis changes, the nodes of the lattice $\mathfrak{R}'$ can receive non-integer indexes in the new basis and their intensities can be restored by interpolating from the new cardinal values, which were inferior ones in the old basis.

Let $\{\mathbf{a}^{*}_{new},\mathbf{b}^{*}_{new},\mathbf{c}^{*}_{new}\}$ be a new sampling basis and $\mathbf{A}$ is 3×3 matrix whose columns are the coordinates of the new basis vectors in the old basis. Let $\mathbf{H}$ and $\mathbf{H}^{new}$ denote the columns of the coordinates (indexes) of the vector $\mathbf{h}\in\mathbf{R}^{3}$ in the old and new bases correspondingly. These coordinates are related by $\mathbf{H}=\mathbf{A}\mathbf{H}^{new}$. We denote by

$$I_{\mathbf{H}}=I\left(h_{1}\mathbf{a}^{*}+h_{2}\mathbf{b}^{*}+h_{3}\mathbf{c}^{*}\right),\quad J_{\mathbf{H}}=I\left(h_{1}\mathbf{a}^{*}_{new}+h_{2}\mathbf{b}^{*}_{new}+h_{3}\mathbf{c}^{*}_{new}\right) \tag{19}$$

the intensities indexed in the old and new bases, respectively, which are related as

$$J_{\mathbf{H}} = I_{\mathbf{AH}}, \quad \mathbf{H} \in \mathbf{R}^3. \tag{20}$$

Writing out in the new basis the interpolation formula for the vector $\mathbf{A}^{-1}\mathbf{h}$

$$J_{\mathbf{A}^{-1}\mathbf{H}} = \sum_{\mathbf{K}\in\mathbf{Z}^3} \mathbf{sinc}\left(\mathbf{A}^{-1}\mathbf{H} - \mathbf{K}\right) J_{\mathbf{K}} \ , \quad H \in \mathbf{R}^3, \tag{21}$$

and returning to the coordinates of the original basis using (20), we obtain a repairing interpolation formula in the form

$$I_{\mathbf{H}} = \sum_{\mathbf{K}\in\mathbf{Z}^3} \mathbf{sinc}\left(\mathbf{A}^{-1}\mathbf{H} - \mathbf{K}\right) I_{\mathbf{AK}}. \tag{22}$$

If $\mathbf{H} \in \mathbf{Z}^3$, but $\mathbf{A}^{-1}\mathbf{H} \notin \mathbf{Z}^3$, then this formula allows one to obtain the intensity value from the cardinal set as a linear combination of intensities with, generally speaking, non-integer indexes. Thus, the possession of intensity values with non-integer indexes may allow one to restore the cardinal intensity values. This formula is complementary to the formula (18), where inferior intensities were restored through the cardinal ones. The matrix $\mathbf{A}$ in the formula (22) can be any non-singular matrix ensuring the condition $\mathrm{supp}P(\mathbf{r}) \subset V_{new}$.

Let consider one more method of using inferior intensities to restore the values of the cardinal ones. Let some basis of discretization be chosen and a corresponding set of cardinal intensities be assigned. We divide this set into two parts: the cardinal intensities known from the experiment (we will designate this set B_K) and the cardinal intensities, the value of which are unknown (the set B_U). For each inferior intensity $I^{obs}_{\mathbf{H}}$ measured in the experiment, the interpolation formula allows to set the equation

$$\sum_{\mathbf{K}\in B_U} \mathbf{s}inc\left(\mathbf{H} - \mathbf{K}\right) I_{\mathbf{K}} = I^{obs}_H - \sum_{\mathbf{K}\in B_K} \mathbf{s}inc\left(\mathbf{H} - \mathbf{K}\right) I^{obs}_{\mathbf{K}}, \tag{23}$$

which restricts the values of unknown cardinal intensities $\{I_{\mathbf{K}},\ \mathbf{K} \in B_U\}$. Writing out such an equation for each measured inferior intensity, we get a system of linear equations that makes it possible to determine unknown base intensities in a limited resolution zone $|\mathbf{k}| \le 1/d_{min}$.

4 Test results

As the test object, the aldose reductase structure, previously determined with a sub-atomic resolution of 0.8Å was selected (Petrova et al., 2006; PDB ID 2*i*16;). The dimensions of the minimal orthogonal cell enclosing the molecule were defined as 70. × 50. × 70.Å, and the dimensions of the expanded cell containing the support of the Patterson function (6) were set as 140×100×140Å. To simplify the visualization the presented tests were carried out for one dimensional set of data, which corresponds to projection of the aldose molecule on O_x axis.

In one-dimensional case the interpolation formula may be presented as

$$I_h = \sum_{k\in\mathbf{Z}} \mathrm{sinc}\left(h - k\right) I_k \quad , \quad h \in \mathbf{R}, \tag{24}$$

where $I_h = I\left(h/a\right)$, $a = 140\,\mathring{\mathrm{A}}$. The intensities have Hermitian symmetry $I_{-h} = I_h$, so that (24) may be reduced to

$$I_h = sinc\left(h\right) I_0 + \sum_{k=1}^{\infty} \left[sinc\left(h - k\right) + sinc\left(h + k\right)\right] I_k \quad , \quad h \in \mathbf{R}, \tag{25}$$

Two sets of exact intensity values were calculated starting from atomic coordinates in the resolution range $\infty = 1\,\mathring{\mathrm{A}}$, that means up to $h_{max} = 140$. The first set was composed by cardinal intensities $\{I_k : k = 0, \pm 1, \pm 2, \ldots, \pm 140\}$, the second one contained inferior intensities $\{I_h : h \in \mathbf{R}\}$ calculated with the step $\triangle h = 5 \times 10^{-4}\,\mathring{\mathrm{A}}$.

Fig. 1 shows the values of exact intensities and reveals two important features of crystallographic data. First, the scale of the data may change a billion times from the central zone to periphery. Second there exist a lot of very deep and narrow dips in intensity values, which are very difficult for the interpolation.

Fig. 2 shows the quality of the interpolation with the use of the interpolation formula (18) with the cardinal set restricted by resolution 1Å ($|h| \leq 140$). The quality of interpolation is quite good. It is worthy to note especially, the quality of interpolation between cardinal intensities I_6 and I_7 , where a deep pit is reproduced quite well.

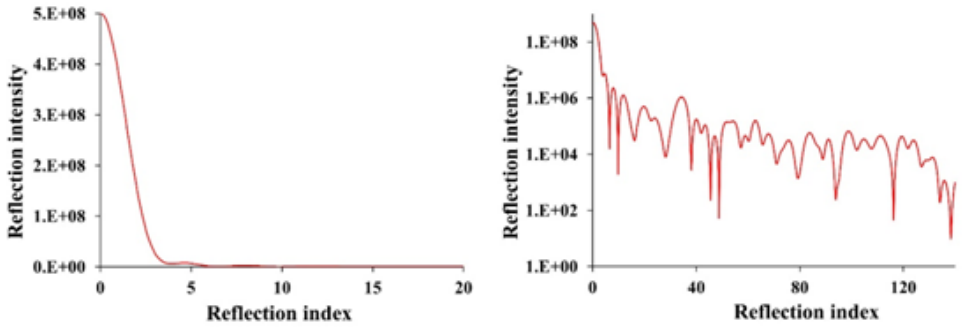

Figure 1. The exact intensities values in usual and logarithmic scales.

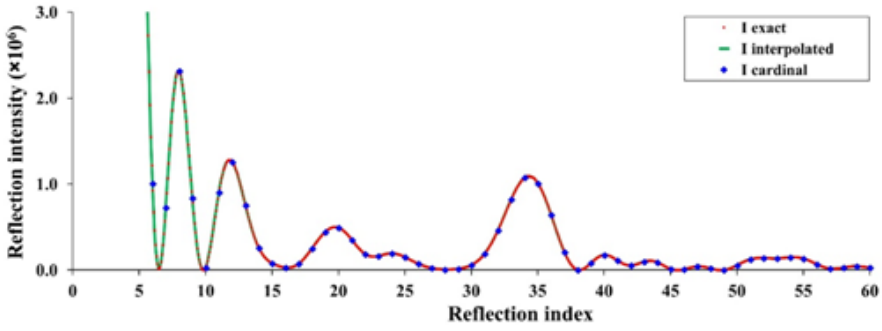

Figure 2. Cardinal, interpolated and exact intensities. Cardinal set (∞-1.Å) is used.

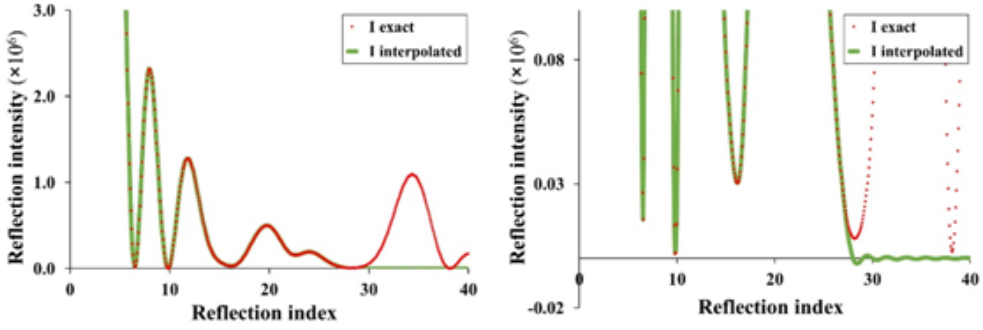

Figure 3. Interpolation using (∞-5.Å) cardinal set.

Fig. 3 demonstrates the effect of further truncation of cardinal series by resolution. Now the resolution of the cardinal set was restricted to 5Å ($|h| \leq 28$). We see that the interpolation is still good for resolution lower than 5Å, but totally fails to represent the signal beyond this limit. This is not surprising, as the function $\mathrm{sinc}\,(x)$ is rather well localized, so that the value of the interpolating line at some point is defined, mainly, by a small number of neighboring cardinal intensities.

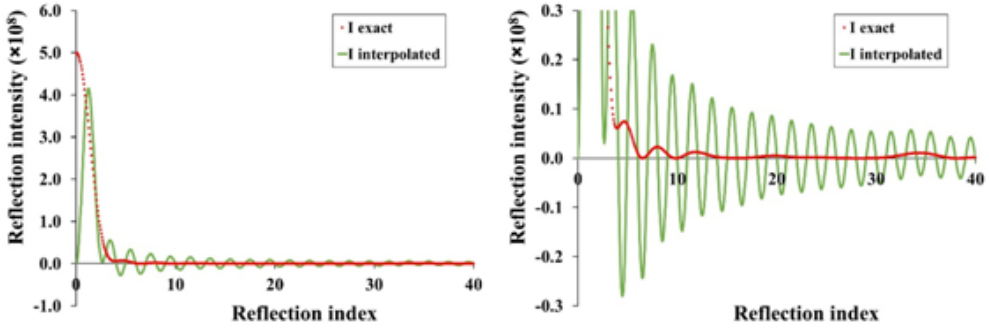

Figure 4. Interpolation using (∞-1.Å) cardinal set with the only one reflection I_0 excluded.

The situation changes drastically, if a small number of lowest resolution reflections are absent in the cardinal set. Fig. 4 shows the effect of exclusion of only one intensity I_0 from the cardinal set. It is interestingly to note, that the extension of the zone of lost low-resolution cardinal intensities improves the quality of interpolation in the middle resolution zone (Fig.5). If 29 reflections of the 10Å resolution central zone ($|h| \leq 14$) are excluded from the cardinal set, then the interpolation becomes not ideal, but reasonable in the resolution zone 10-1Å. Such behavior of the interpolation curve can be explained by the existence of extremely large central peak of intensity in the low-resolution zone. When being excluded these reflections influence the interpolated values even at large distances.

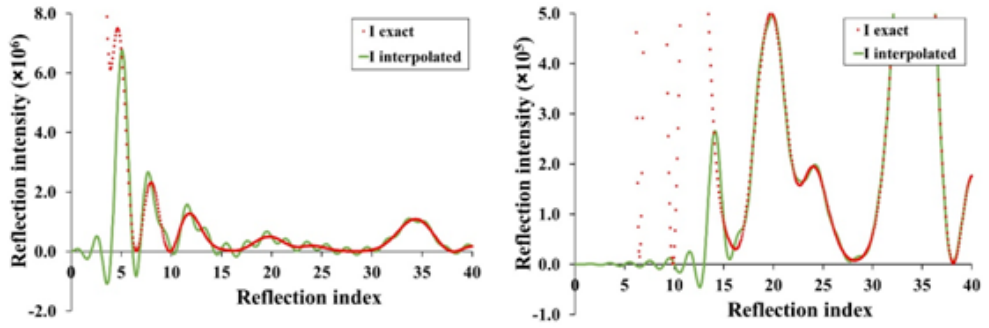

Figure 5. Interpolation using restricted cardinal sets: 30.-1.Å (left) and 10.-1.Å (right).

Fig. 6 shows the result of restoring of 9 low-resolution reflections in (∞-30.Å) resolution zone ($|h| \leq 4$) by solving of the system of linear equations (23). The exclusion of these 9 reflections from the cardinal set leads to a poor quality of the interpolation (Fig. 5, left). To restore these values, the system of linear equations (23) was formed using 1929 inferior intensities ($|h| > 4$), which were supposed to be known. The linear equations were solved by Singular Value Decomposition method. The found values are listed in the Table 1. The found values for low-resolution cardinal intensities were added to the cardinal set to restore the excluded ones. Fig. 6 shows the result of application of the interpolation formula when using the restored low-resolution cardinal intensities. In contrast to Fig. 5 (left) here the quality interpolation is quite good.

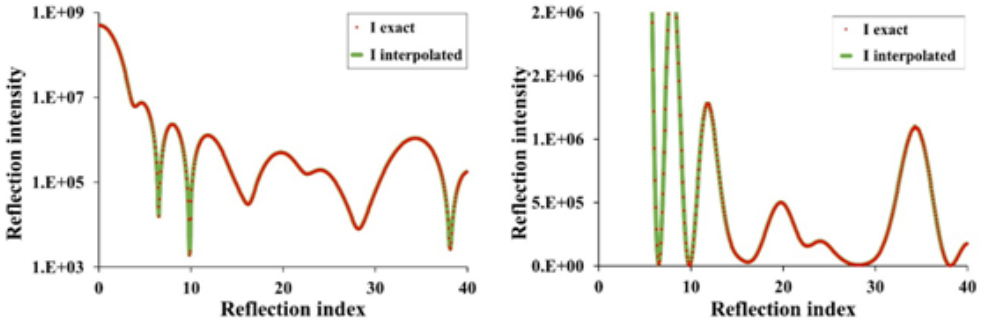

Figure 6. Interpolation using (∞-1.Å) cardinal set with restored intensities in (∞-30.Å) zone.

Table 1. The exact and restored values of the excluded cardinal intensities.

h	I exact	I restored
0	0.500396E+09	0.504212E+09
1	0.380557E+09	0.383364E+09
2	0.156493E+09	0.157573E+09
3	0.272373E+08	0.274242E+08
4	0.620644E+07	0.621577E+07

5 Conclusion

Despite the tremendous successes achieved in determining the structure of biological macromolecules by the X-ray diffraction methods, some theoretical and practical questions remain open. Advance in this area could be extremely useful for the crystallographic practice and biological applications. This especially applies to the case of the study of isolated particles, where the gap between the theoretical possibility of recovering intensities and phases of structure factors and the practical possibilities of doing this is still large.

Acknowledgement

This work was supported by the Russian Foundation for Basic Research (project 16-04-01037a).

References

Ayyer, K., Geloni, G., Kocharyan, V., Saldin, E., Serkez, S., Yefanov, O., Zagorodnov, I. (2015). Perspectives for imaging single protein molecules with the present design of the European XFEL. *Structural Dynamics*, 2, Article number 041702.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Cryst.*, A32, 832-847.

Loh N.-T.D., Elser V. (2009). Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. E.*, 80, Article number 026705.

Lunin, V.Y. (1988). Use of Information on Electron Density Distribution in Macromolecules. *Acta Cryst.* A44, 144-150.

Lunin, V.Y., Lunina, N.L., Petrova, T.E. (2017). The biological crystallography without crystals. *Math. Biol. Bioinf., 12*(1), 55-72 (in Russian).

Petrova, T., Ginell, S., Mitschler, A., Hazemann, I., Schneider, T., Cousido, A., Lunin, V.Y., Joachimiak, A., Podjarny, A. (2006). Ultrahigh-resolution study of protein atomic displacement parameters at cryotemperatures obtained with a helium cryostat. *Acta Cryst.*, D62, 1535–1544.

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., Song, C., Madsen, A., Chushkin, Y., Zontone, F., Bradley, P. J., Miao, J. (2015). Three-dimensional coherent X-ray diffractive imaging of whole frozen-hydrated cells. *IUCrJ*, 2, 575-583.

Sayre, D. (1952). Some implications of a theorem due to Shannon. *Acta Cryst.*, 5, 843.

Spence, J.C.H. (2017). XFELs for structure and dynamics in biology. *IUCrJ*, 4, 322-339.

Urzhumtsev, A.G. (1991). Low-resolution phases: influence on SIR syntheses and retrieval with double-step filtration. *Acta Cryst.*, A47, 794-801.

Urzhumtsev, A.G., Lunin, V.Y., Luzyanina, T.B. (1989). Bounding a Molecule in a Noisy Synthesis. *Acta Cryst.*, A45, 34-39.