**Project Report**

# Structure determination of biological particles with X-ray free-electron laser

NKFIH OTKA K-115504
2015-10-01 - 2019-09-30
(extended to 2021-03-29 without further financial support)

## Introduction

The knowledge of the detailed structure of solids, molecules or living cells is essential for many fields of basic science and advanced technology, ranging from materials science to engineering or from biology to biotechnology and the whole pharmaceutical industry. For the last century the vast majority of structural information at atomic resolution was obtained from diffraction experiments on crystals. Until recently, the bottleneck of success was crystallization; not all materials can be crystallized, and these substances are beyond the scope of crystallographic techniques.

This situation has recently changed with the introduction of two methods, which can solve the structure of individual particles without the need for crystallization. One of them, cryogenic electron microscopy (cryoEM) has reached maturity and earned its developers the 2017 Nobel Prize in Chemistry. The other method, still under development, uses the intense and short pulses of the new revolutionary X-ray sources, called X-ray free electron lasers (XFELs). In principle, scattering from a single particle (e.g. an inorganic nanoparticle, biological macromolecule or a virus) could provide even better information on its structure than a crystal, but unfortunately, radiation damage destroys the sample before the scattering pattern could be recorded. This fundamental problem would be avoided if extremely short (10-100 fs) and extremely intense ($>10^{12}$ photons/pulse) X-ray pulses were used. X-ray free electron lasers can produce pulses with just the desired properties, hence a new type of experiment, called single particle imaging (Solem, 1986; Neutze *et al*., 2000), is gradually becoming a reality. In the single particle experiment the sample is also destroyed, but before, it gives valuable information about its original structure. Such a "diffract-and-destroy" approach can detect only a fraction of the scattering space in one measurement, therefore, a large number of measurements is needed on fresh replicas of the same particle with different orientations.

In the vast majority of diffraction experiments on crystals or on single particles, only the diffracted intensity is recorded, the phase of the scattered radiation is lost. Without knowing the phase, reconstruction of the structure from the intensity alone is not possible. In the last century several methods were developed to find the missing phase information. These methods, using extra experimental information or theoretical considerations, work well in many cases, but they are not universal. There are cases when the solution of the phase problem with these methods is difficult or impossible.

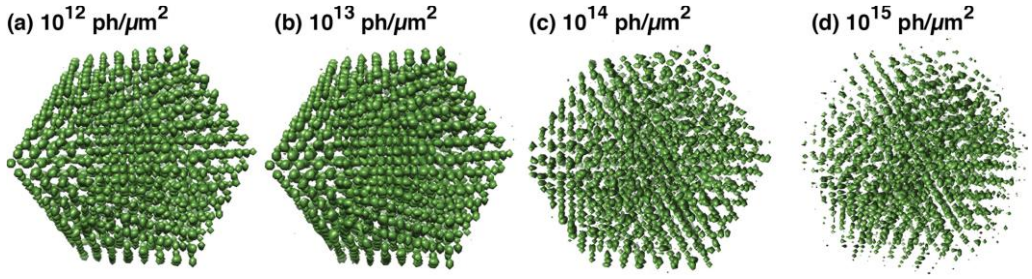
When X-rays from an XFEL, synchrotron or laboratory source hit the sample, a significant fraction of the photons are absorbed by the atoms and reemitted at a longer wavelength. Part of these fluorescent photons will scatter on other atoms of the sample and mix coherently with the fluorescent photons leaving the sample without interaction. The resulting interference leads to the formation of Kossel-lines (Kossel *et al*., 1935). The pattern can be recorded by a 2D

detector and contain information not only on the intensities, but also on the phases of the scattered fluorescent radiation.

The main aim of the project is related to two types of experiments: single-particle diffraction that was introduced above, and in-flight holography, a completely new experiment exploiting the interference between two different particles in the XFEL beam. We also achieved important new results in the closely related field of Kossel-lines. In the framework of the project we performed experiments using XFEL, synchrotron and laboratory X-ray sources. We developed data analysis algorithms and devised structure reconstruction methods. Part of the work was done in collaboration with other groups working on similar problems in institutions in Europe and in the USA: LCLS Stanford, ESRF Grenoble, ANL Argonne, LBNL Berkeley, ELI Prague, TU Berlin, Uppsala University, XFEL Schenefeld, CFEL Hamburg, and some others. In the following, we describe our results in more detail.

## 1. Modelling and reconstruction of clusters in intense XFEL beams.

For the structure determination of small particles using X-ray free-electron lasers (the main aim of our project), it is very important to understand how these particles behave in the extremely short and intense XFEL beam. On this problem, we worked in collaboration with the group of Linda Young and Christoph Bostedt at Argonne National Laboratory (Argonne, Illinois, USA) on the modelling of the radiation damage of argon clusters. They used a combined Monte Carlo – Molecular Dynamics (MC/MD) method to follow both the ionization, relaxation and recombination processes inside the atoms and the movement of the ions and the electrons. These methods give realistic description of the radiation damage and eventual destruction of the particle. X-ray diffraction snapshots were calculated during the XFEL pulse and averaged to simulate the image obtained in a single-particle scattering experiment. Our part in the collaboration was to solve the phase problem and reconstruct the structure of the cluster. The results of this work were published in *Physical Review A* (Ho *et al*., 2016).

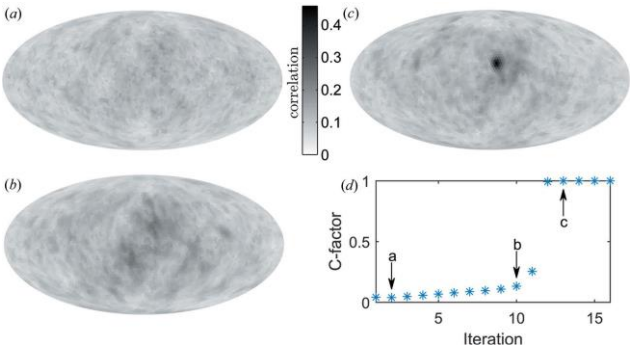



Reconstructed electron densities (green regions) of Ar clusters with different fluence levels.

## 2. Orientation method development for single particle imaging.

### *2.1. Introduction of correlation maps and the C-factor to test the reliability of orientation algorithms.*

In a single-particle diffraction imaging experiment single particles are injected into the XFEL beam. Each diffraction pattern contains only a few thousand photon counts and thousands of diffraction patterns are necessary for a successful structure determination. The individual diffraction patterns belong to particles of different unknown orientations. In our previous project (OTKA K 81348) we developed the correlation maximization (CM) method to find the

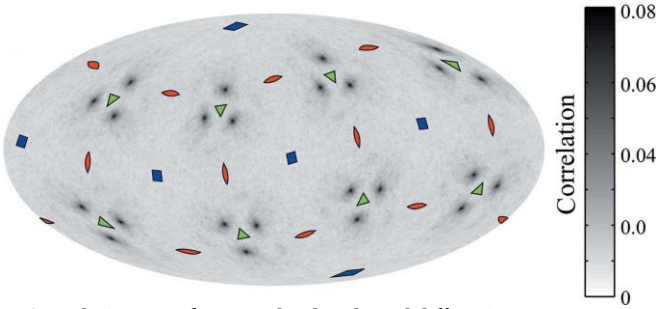



*Correlation maps of a randomly selected diffraction image of lysozyme at various stages of the iteration process by the CM method (a)–(c). C-factor as a function of iteration number (d).*

unknown orientations and merge the patterns into a consistent 3D dataset (Tegze & Bortel, 2012). Although there are some indicators during the orientation process showing that convergence was reached, we found necessary to establish an independent method to test the consistency of the resulting dataset. We developed a method based on correlation maps and introduced a figure of merit, called C-factor, to test the consistency and reliability of the 3D dataset. Our consistency check can be applied to datasets oriented by any method. The results were published in *Acta Crystallographica A* (Tegze & Bortel, 2016) and presented at the workshop *"Modern light sources and their applications"* (ELI-ALPS, Szeged, Hungary, December 7-9, 2017).

*2.2. Orientation and structure solution method for symmetric objects*

The CM algorithm (Tegze & Bortel, 2012) and the C-factor was developed and tested on asymmetric objects. We now extended the applicability of these methods for symmetric particles. The presence of symmetry, on one hand, may facilitate the solution of the orientation and the phase problems, but on the other hand, may also cause some complications. For a symmetric object, several orientations are equivalent and all of these should be taken into account. We modified our algorithm to treat the symmetry-equivalent orientations correctly. We successfully tested the modified algorithm on simulated diffraction patterns of the iron-storing protein ferritin. Ferritin is a large protein complex with a high, 24-fold symmetry. We also had to modify the definition of the C-factor for the case of symmetric objects. After finding the orientations of the diffraction patterns, the phase problem was successfully solved by the relaxed averaged alternating reflections (RAAR, Luke, 2005) and error reduction (ER, Fienup, 1982) methods. The resulting electron density shows very good agreement with the structure of ferritin measured by conventional methods. The results of this work were published in *Acta Crystallographica A* (Tegze & Bortel, 2018) and presented at the EUCALL workshop: *Biology at Advanced Laser Light Sources* (Schenefeld, Germany, November 30 - December 1, 2017) and the *7th Annual BioXFEL International Conference* (New Orleans, Louisiana, February 13-15, 2018).



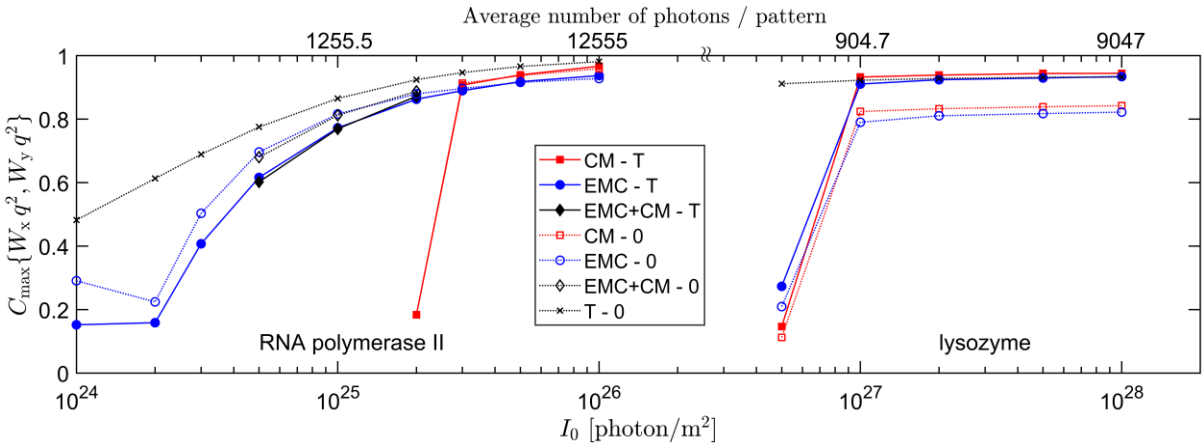

*Correlation map for a randomly selected diffraction pattern of ferritin. The red, green and blue symbols indicate the two-, three- and fourfold rotation axes, respectively.*

*2.3. Comparison and combination of EMC and CM orientation methods for single particle imaging.*

In coherent diffractive imaging, a critical step of data analysis is finding the orientation of individual single particle diffraction patterns. For this purpose, we have already developed the correlation maximization (CM) method (Tegze & Bortel, 2012). The EMC method is another orientation algorithm (Loh & Elser, 2009) that may work better for diffraction data with very

low photon counts. Unfortunately, EMC is computationally rather wasteful, its estimated computing time scales with the sixth(!) power of the particle-size/resolution ratio. By combining the best parts of the computationally effective CM and the statistically solid EMC algorithms, we developed a computer program which keeps the advantages of the EMC method but is faster and scales better with the size of the particle. We have successfully modified the original EMC algorithm using the correlation theorem and the Fast Fourier Transform to calculate the costliest parts of the computational tasks. This modification makes the solution of the orientation problem significantly faster. It has also changed the computer time scaling properties of the algorithm from the sixth power to the fifth power of the particle-size/resolution ratio. This improvement makes possible the solution of the orientation problem for large biological particles in real-time during the XFEL experiment. For these computation-intensive tasks, we have purchased and installed state-of-the-art graphics processors.

We performed extensive testing of the modified EMC algorithm and compared its efficiency to the CM algorithm. Our test molecules for the comparison were the lysozyme, a small protein molecule and the much larger multiprotein complex RNA polymerase II. Many thousands (20,000 for lysozyme and 100,000 for RNA polymerase II) of noisy diffraction patterns of the molecules in random orientations were calculated at various incident XFEL intensities. Then orientation of the patterns was attempted by both methods. For the relatively small lysozyme molecule the two methods performed similarly, although the CM method gave more accurate results. Both methods were able to solve the orientation problem for XFEL intensities $\geq 10^{27}$ photon/m$^2$, corresponding to an average 904.7 photons in a pattern. For the more realistic case of the RNA polymerase II, the EMC method performed better at low XFEL intensities (below $3 \cdot 10^{25}$ photon/m$^2$, corresponding to an average 927.4 photons in a pattern). However, at higher




*Correlation between intensity volumes $W_x$ and $W_y$ for lysozyme (right) and RNA polymerase II (left) at various incident XFEL fluence values. $W_x$ denotes the intensity distribution reconstructed by either EMC (blue circles), CM (red squares), a combination of them (black diamonds) or using the true orientations (black crosses). $W_y$ refers to the reference intensity distributions $W_T$ (full symbols and solid lines) or $W_0$ (empty symbols and dotted lines).*

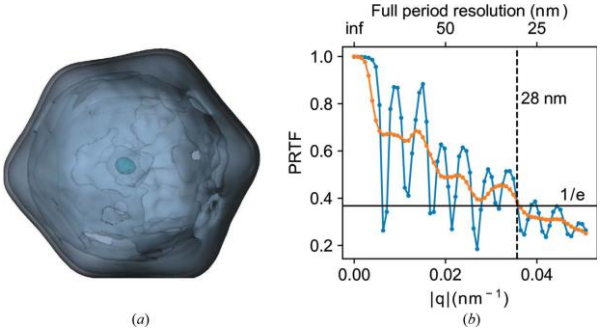
intensities the CM method gave again more accurate results. After orienting the patterns and recovering the 3D intensity distribution we solved the phase problem using Fienup's hybrid input-output algorithm. At high XFEL intensities the resulting electron density distributions showed very good agreement with the known atomic structure of the molecule. However, as expected, for lower intensities the resolution was strongly degraded. The results are under review for publication at *IUCrJ*. Part of the results were also presented at the *7th Annual BioXFEL International Conference* (San Juan, Puerto Rico, January 28-30, 2020).

### 3. Volumic omit maps in ab initio dual-space phasing

We studied a special class of projection-type algorithms that alternate between dual spaces and prescribe only weak constraints on the electron density and the diffraction data. While conceptually appealing, these basic algorithms are usually avoided in practice due to their stagnation properties. To cure this issue, we introduced new perturbations called "volumic omit maps", and found that their periodic application improves the solution speed by orders of magnitude. Our findings were published in *Acta Crystallographica A* (Oszlányi & Sütő, 2016).

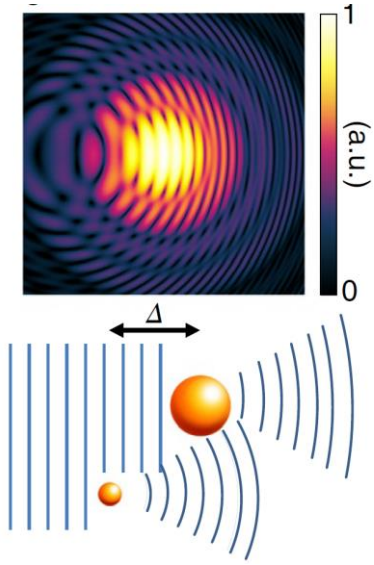
### 4. Investigation of the effect of experimental conditions on single particle imaging

In a broad international cooperation, we have explored the effect of background, sample heterogeneity and blurring on the reconstructed model quality in coherent diffractive imaging using both experimental data and simulated data. The results emphasize the importance of background reduction and removal to improve the quality of the reconstructed 3D structure. For this study we used the experimental data recorded earlier at the Linac Coherent Light Source (LCLS) in Stanford, USA. The results were published in *IUCrJ* (Lundholm *et al*., 2018).




*(a) Three-dimensional representation of the reconstruction from experimental data on the icosahedral Melbournevirus. (b) PRTF for the reconstruction in (a).*

### 5. In-flight X-ray holography

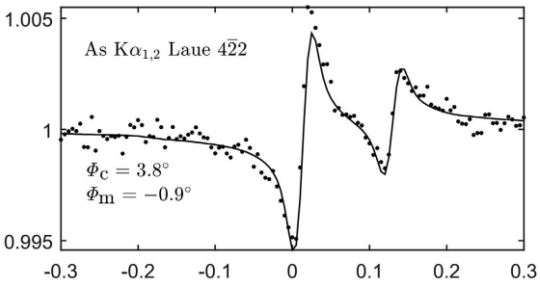



We participated in another experiment at the LCLS in a broad cooperation with research groups in the US, Sweden and Germany. In this experiment virus particles and xenon clusters were injected into the XFEL beam in two separate particle jets. Those diffraction patterns were selected where the X-ray pulse had hit a single virus and a single xenon cluster simultaneously. The interference between two different particles can be interpreted as a hologram. The X-rays scattered on the xenon cluster provide the reference beam of the hologram. Our main part of the evaluation work was the development of a method to find the relative positions of the particles which makes possible the determination of the phase of the X-rays diffracted on the virus particle. The results of this work were published in *Nature Photonics* (Gorkhover *et al*., 2018).

### 6. Experimental determination of the phase from Kossel-lines

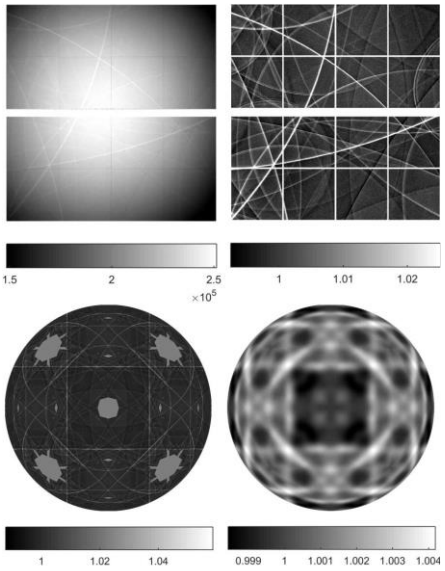
If a coherent reference beam is mixed to the scattered radiation, then the resulting interference pattern contains information on the phase. This is the basic principle of holography and this was applied in our earlier work on atomic resolution X-ray holography using internal source (Tegze & Faigel, 1996). If the source of the radiation is inside the crystal, then sharp lines appear in the radiation image recorded by a 2D detector. The profile of these Kossel-lines

depends on the phase of the scattered radiation. In an experiment at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) we measured Kossel-line patterns in great detail on a GaAs single crystal sample. Using these patterns, we were able to determine the phases of the scattered radiation. The results were published in *Scientific Reports* (Faigel *et al*., 2016).




*Measured and calculated Kossel-line profiles*

## 7. Fast X-ray fluorescent holography




The measured images of the Kossel-line experiment also contain holographic information. We recorded X-ray fluorescent holograms using a 2D detector on a NiO single crystal sample at the ESRF. We have shown that it is possible to collect sufficient structural information to image the 3D structure of the crystal in 1 second at a synchrotron. This opens a series of new possibilities in structural studies at X-ray free electron lasers, which have sufficient intensities to measure a full hologram with a single femtosecond pulse. The results were published in the *Journal of Synchrotron Radiation* (Bortel *et al*., 2019).

*Steps of the data preparation. Top left: raw detector image. Top right: normalized image. Bottom left: hologram expanded to a sphere. Bottom right: low-pass filtered hologram.*

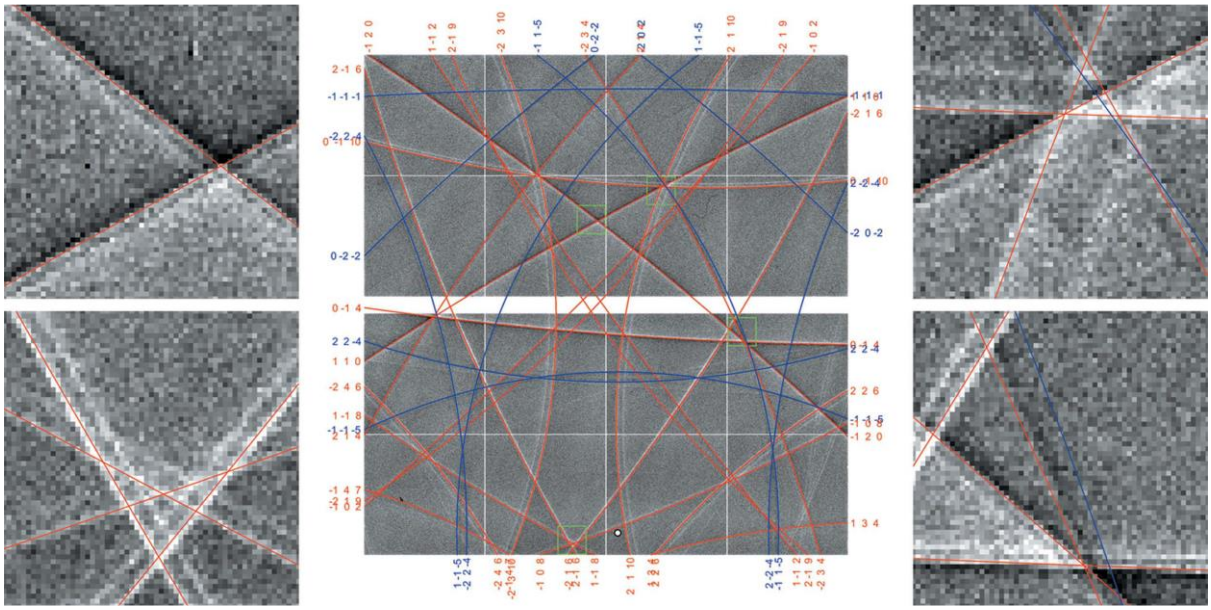
## 8. Experimental determination of crystal structure from Kossel lines.

Before solving the phase problem by using the line shapes of Kossel lines, the structure of the crystal must be known. In many cases, not only the crystal lattice but also its orientation is unknown. We have developed an algorithm and computer program for fitting the Kossel lines. The program can identify the Kossel lines semi-automatically and determine the corresponding reciprocal lattice vectors. These vectors are then used as input for a single-crystal indexing program (DIRAX).

In order to test our indexing and line shape fitting algorithms on other than synchrotron data, we have built a laboratory setup to produce Kossel-line patterns. We used a scanning electron microscope (Jeol JSM-5800LV) to excite the atoms in a crystal by accelerated electrons. We recorded the emitted fluorescent radiation through a Kapton foil window by a 2D X-ray detector (Dectris Eiger R 1M). The first experiment on a single crystal hematite sample has shown sharp Kossel lines.

We have developed an algorithm and a computer program which can determine the unknown lattice, its orientation and indices of the lines simultaneously with extraction of the precise experimental geometry from the measured Kossel- or Kikuchi-line pattern. Using only a single recorded image, we were able to index the lines and determine the unit cell of the crystal. The

procedure was successfully demonstrated on the experimental Kossel line patterns. The results were published in the *Journal of Applied Crystallography* (Bortel *et al*., 2021).



*Final lattice-constrained fit of the Kossel lines (in red) to the experimental pattern (in the middle), and four magnified regions (in the corners) indicated by green boxes.*

## References


Bortel, G., Faigel, G., Tegze M. & Angelov B.: Fast inside-source X-ray fluorescent holography. *J. Synchrotron Rad.* **26**, 170 (2019).

Bortel, G., Tegze, M. & Faigel, G.: Constrained geometrical analysis of complete K-line patterns for calibrationless auto-indexing. *J. Appl. Cryst.* **54**, 123 (2021).

Faigel, G., Bortel, G. & Tegze, M.: Experimental phase determination of the structure factor from Kossel line profile. *Sci. Rep.* **6**, 22904 (2016).

Fienup, J. R.: Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758 (1982).

Gorkhover, T. *et al.*: Femtosecond X-ray Fourier holography imaging of free-flying nanoparticles. *Nature Photonics*, **12**, 150 (2018).

Ho, P. J., Knight, C., Tegze, M., Faigel, G., Bostedt, C. & Young, L.: Atomistic three-dimensional coherent x-ray imaging of nonbiological systems. *Phys. Rev. A* **94**, 063823 (2016).

Kossel, W., Loeck, V. & Voges, H.: Die Richtungsverteilung der in einem Kristall entstandenen charakteristischen Röntgenstrahlung. *Z. Phys.* **94**, 139 (1935).

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

Luke, D. R.: Relaxed averaged alternating reflections for diffraction imaging. *Inverse Probl.* **21**, 37 (2005).

Lundholm I. V. *et al*.: Considerations for three-dimensional image reconstruction from experimental data in coherent diffractive imaging. *IUCrJ* **5**, 531(2018).

Neutze, R., Wouts, R., van der Spoel, D., Weckert & E., Hajdu, J.: Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752 (2000).

Oszlányi, G. & Sütő, A.: Volumic omit maps in *ab initio* dual-space phasing. *Acta Cryst.* A**72**, 480 (2016).

Solem J.C.: Imaging biological specimens with high-intensity soft x rays. *J. Opt. Soc. Am. B* **3**, 1551 (1986).

Tegze M. & Bortel G.: Atomic structure of a single large biomolecule from diffraction patterns of random orientations. *J. Struct. Biol.* **179**, 41 (2012).

Tegze, M. & Bortel, G.: Coherent diffraction imaging: consistency of the assembled three-dimensional distribution. *Acta Cryst.* A**72**, 459 (2016).

Tegze, M. & Bortel, G.: Incorporating particle symmetry into orientation determination in single-particle imaging. *Acta Cryst.* A**74**, 512 (2018).

Tegze, M. & Faigel, G.: X-ray holography with atomic resolution. *Nature* **380,** 49 (1996).