

A Review of Methods of Resolution Estimation for 3D Reconstructions of Nanoscale Biological Objects from Experiments Data on Super-Bright X-Ray Free Electron Lasers (XFELs)

Kseniia Ikonnikova^[0000-0002-2412-9680]

National Research Centre "Kurchatov Institute",
1 Akademika Kurchatova pl., Moscow, 123182, Russia
ikonnk8@gmail.com

Abstract. Nowadays the Fourier shell correlation (FSC) is the most common method for estimating the resolution of 3D structures obtained in Single Particle Imaging (SPI) experiments on X-ray free electron lasers (XFELs). In FSC, the resolution is defined as the spatial frequency at which the correlation between two independently reconstructed structures is equal to some given threshold value. There are multiple methods to define the threshold value. In addition, this approach cannot account for the fact that the quality of reconstruction can be non-uniform for different areas of the biomolecule. Thus, the issue of effective resolution estimation methods remains open. This paper considers multiple alternative approaches to the resolution estimation from adjacent scientific field - cryogenic electron microscopy (cryo-EM) and analyzes the applicability of these approaches to the resolution estimation in SPI experiments on XFELs.

Keywords: X-ray Free Electron Laser, Single Particle Imaging, Space Resolution, Fourier Shell Correlation

1 Introduction

The viral worldwide pandemic caused by SARS-CoV-2 has become a serious challenge for the entire scientific community, and the search of effective treatment methods and drugs against COVID-19 is ongoing. Determination of the high-resolution 3D structure of single viruses' particles is one of the key and important points for understanding how viral infection occurs and how we can fight it. Cryogenic electron microscopy (cryo-EM), which has recently been able to obtain a true view of the atomic resolution of a biomolecule (1.2 Å) [1], has consolidated its position as the leading method for imaging biomolecular particles. However, in cryo-EM, the samples are plunge-frozen down to -269 °C and so they are imaged at unphysiological conditions, which prevents the study of biomolecules in their natural state and limits the ability to track conformational changes and dynamic events (for example, how the initial event of cellular recognition occurs between the viral spike (S) protein and the ACE2 receptor [2]).

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

With the invention of super-bright X-ray free electron lasers (e.g. Linac Coherent Light Source (LCLS) and European XFEL)) the Single Particle Imaging (SPI) approach allowed researchers to reconstruct 3D structures from many 2D diffraction images produced in the experiments by X-rays scattered on the single particle exposed in different orientations [3]. Thus, SPI experiments opened new opportunities to study biomolecules in their nature state without previous crystallization or being frozen. Unfortunately, there are still many challenging problems in SPI experiments (weak signal, scattered on single particle, low number of diffraction images), which limit the quality of the obtained 3D structures. Nevertheless, in order to assess the experimental quality and confidence for the interpretation of the obtained 3D structures and to compare results with other structural biology methods, we need to use the resolution estimation.

Nowadays, the standard method for estimating resolution of the obtained 3D structures both in cryo-EM and SPI experiments is the Fourier shell correlation (FSC) method [4]. In the FSC method, the resolution is defined as the spatial frequency at which the correlation between two independently reconstructed structures becomes equal to some given threshold value. There are several criteria to choose threshold value for the resolution estimation, the most popular of which are fixed thresholds of 0.5 and 0.143 [5] (they rely on statistical assumptions on SNR [3]) and also 1/2-bit threshold [4] (based on informational entropy estimations [3]). Even though the FSC is widely accepted by the scientific community, a discussion continues about a threshold value at which the resolution should be defined [4]. For a more detailed description of the FSC method see [4-7].

As an alternative method to select the FSC threshold value, Beckers and Sachse [8,9] have suggested a new adaptive thresholding procedure for identifying the highest resolution shell based on statistical methods of permutation sampling and false discovery rate (FDR) control. Permutation sampling of the FSC for each resolution shell is as follows: firstly, new samples are generated by changing the order of the Fourier coefficients of the second half-map shell and a large series of FSCs are computed [8,9]. Hence a sample of the noise distribution of the FSC for each resolution shell is obtained. When applied to every resolution shell, the distributions together with the original FSC-values can then be statistically tested and conveniently transformed into p-values [8,9]. In order to reduce the risk of false positive errors, p-values are then corrected by means of FDR control and thresholded at 1% [8,9]. The authors demonstrated [8,9], that this method (named FDR-FSC) gives realistic resolution estimates that are similar to most author-reported resolutions in the Electron Microscopy Data Bank (EMDB) [10]. However, the main advantage of this approach is that it makes no assumption about the statistical properties of the signal and noise within the half-maps, and it does not rely on any FSC threshold "criterion".

The main drawback of the FSC method is that it estimates only the global resolution for the whole structure. However, the electron density usually has uneven resolution over the entire volume: to restore the structure, SPI needs to average the diffraction images from a large number of individual biomolecules, thus the more individual biomolecules differ in structure the stronger the heterogeneity of the reconstructed 3D structure [11]. Thus, for a correct interpretation of quality of the reconstruction, it is

important to be able to determine the local resolution for each voxel of volume. Currently, cryo-EM has proposed several approaches to determine local resolution [12]. The first approach to determining the local resolution was blocres [13], where the resolution is locally estimated by means of the FSC, calculated from two independent reconstructions within a moving window. The most-used method to date for the local resolution estimation is ResMap [14]. This approach determines the local resolution by detecting the best 3D sinusoidal wave that fits each map point above the noise level. MonoRes [15] is based on a similar principle of detecting energy at different frequencies above noise. MonoRes has been recently expanded to account for directionality (now named MonoDir) [16]. An important consequence of this work is the introduction into the field of the concept that resolution is simultaneously local and directional. The DeepRes method [17], based on deep learning from filtered atomic models at different frequencies, has also recently been introduced [12]. For a more detailed overview of all local resolution methods, see [12].

2 Analysis workflow

The aim of the present study was to verify which of the currently available alternative approaches for estimating resolution can be successfully applied to evaluate reconstructions in SPI experiments. To estimate the local resolution, we chose the Resmap method (as the most popular method for the local resolution estimation in cryo-EM), and for the global resolution estimation, we opted for the FDR-FSC method. This work is founded on several major steps. In order to evaluate the accuracy of the resolution estimation methods, we need to test the method performance on reconstructions of different quality and resolution. As follows from [3], the resolution value depends on the amount of diffraction images in the dataset. Also, it is important to understand how noise affects the reconstruction result and resolution values. Thus, first we simulated the single particle diffraction experiments with different levels of the Gaussian noise and different number of diffraction images in dataset for structure of hemocyanin of the marine mollusk fissurellia (Keyhole limpet hemocyanin type 1 - KLH1) protein from PDB database [18,19]. For this purpose, we generated one pack of datasets with different numbers of diffraction images ($n = 200, 1000, 10000$ and 20000) without noise. Then, we generated another pack of datasets with different values of noise ($\sigma = 0, 0.5, 0.8, 0.9, 1.0$) with 20000 images in each dataset. Then, we used the workflow for SPI experiments data processing, which was described in detail in [3]. Finally, we estimated the global resolution with the FDR-FSC method and the local resolution with the ResMap method for the obtained reconstructions and compared these results with the FSC estimation.

3 Results

3.1 FDR-FSC method

Figures 1-2 show dependencies of the resolution on the number of diffraction images and on the noise for the FDR-FSC method and the FSC method with a threshold value of 0.143. As expected for both approaches, the resolution value deteriorates as noise increases (Fig.1).

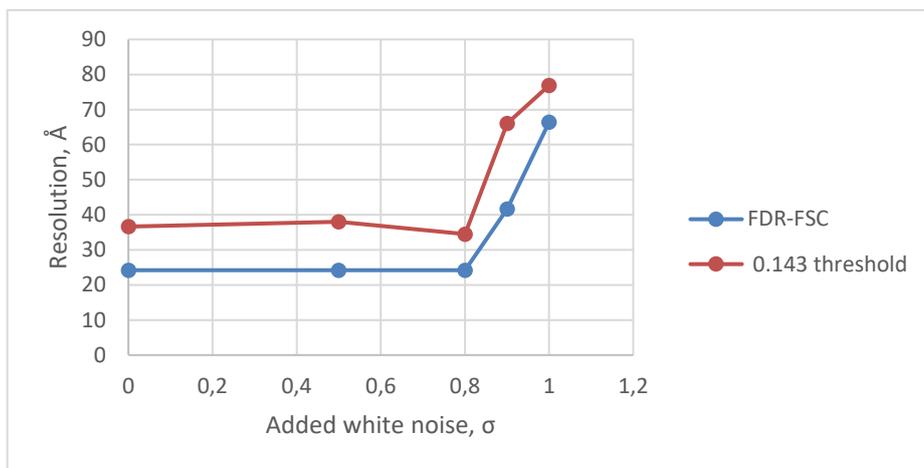


Fig. 1. Resolution estimates for 0.143 FSC and FDR-FSC thresholds with different levels of added noise.

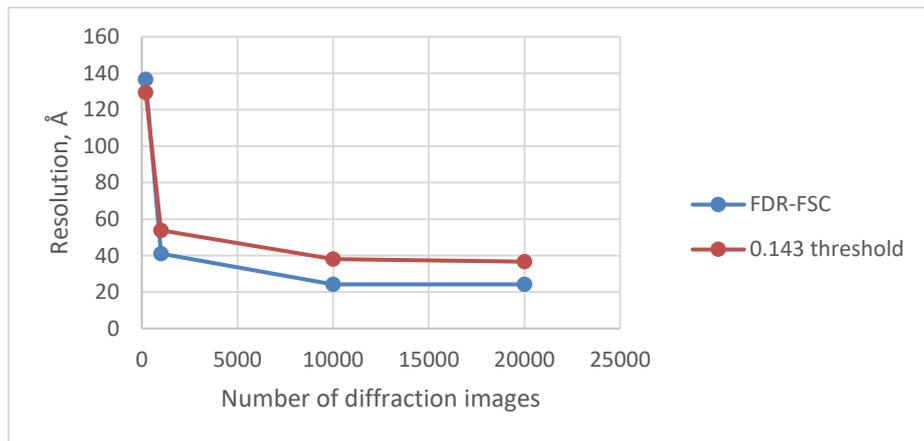


Fig. 2. Resolution estimates for 0.143 FSC and FDR-FSC thresholds with different number of diffraction images in dataset.

Comparison of the 0.143 cutoff threshold values with the FDR-FSC values demonstrated a good agreement between both estimations, but the FDR-FSC method shows a slightly more optimistic estimation. It is worth noting that this result is consistent with the results obtained for cryo-EM [8,9], which proves the universality of the FDR-FSC approach. One main advantage of the FDR-FSC is that inference of statistically significant signal in the resolution shells only requires the distribution of random noise correlations determined by permutation. Thus, this method avoids the consideration of complicated correlations between signal and noise [4-7] – one of the most controversial issues that arise in determining any threshold "criterion"[6]. Thus, the FDR-FSC method has a good chance to become a new "gold standard" for estimating resolution in SPI.

3.2 ResMap Method

Figures 3-4 show the result of estimating local resolution using the ResMap method.

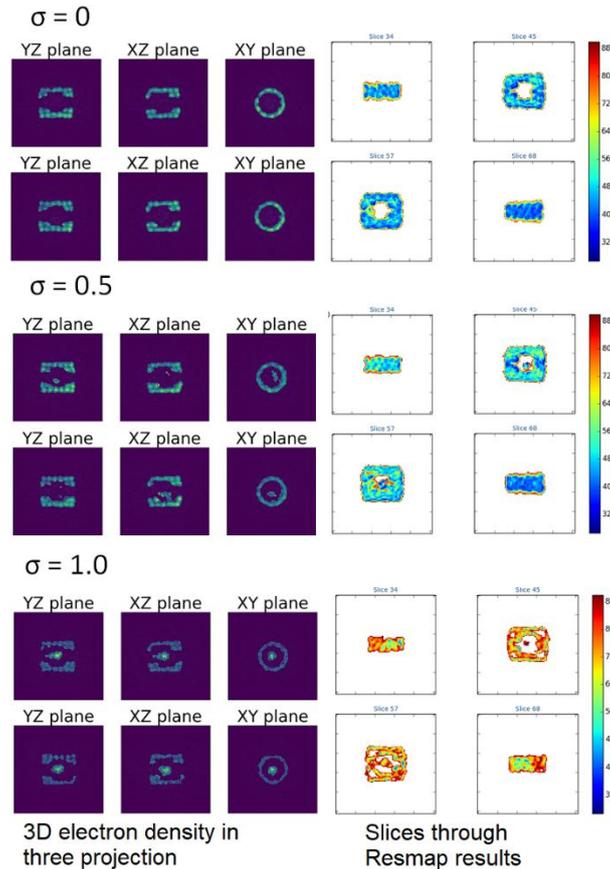


Fig. 3. Result of estimating local resolution for datasets with different level of noise σ (number of images = const=20000).

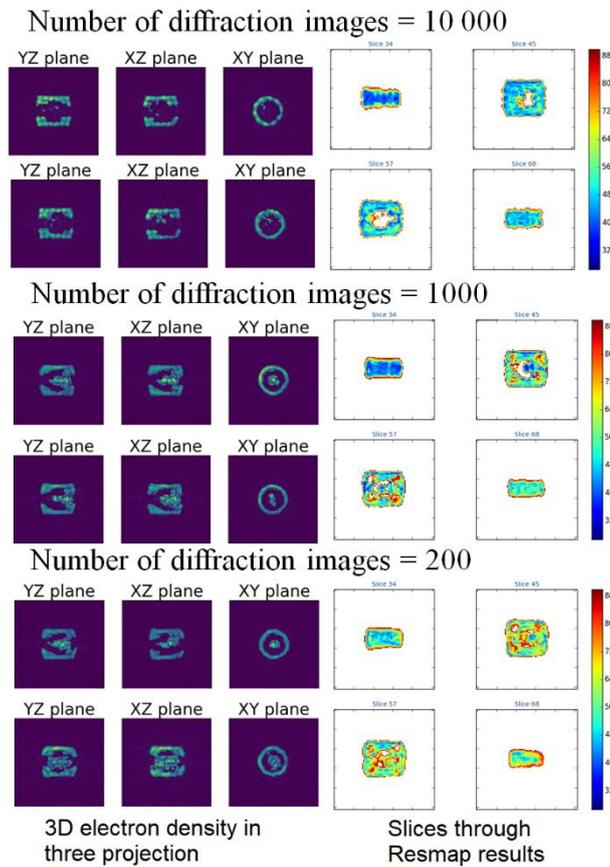


Fig. 4. Results of estimating local resolution for datasets with different number of images (level of noise $\sigma = \text{const} = 0$)

It can be seen that the resolution over the entire structure of the biomolecule is, in fact, not uniform, and as the parameters for reconstruction deteriorate (increased noise or a decrease in the number of diffraction images), this unevenness only increases. Additionally, we can observe that local resolution decreases near the edges of the biomolecule, which may be due to errors of reconstruction algorithms [14].

4 Conclusions

In order to evaluate the quality of the experiment and reliably interpret the results of reconstructing the spatial structure from diffraction data in SPI, it is important to have effective methods for the resolution estimation of these structures. In this research, we have demonstrated that the ResMap and FDR-FSC methods can be used to estimate resolution in SPI experiments and show reasonable results for the model data of KLH1 particle. However, more research is needed in this field: one should test more

particles with various spatial features as well as data from real experiments. Future work implies testing other local resolution estimation methods [12,13,15-17] and comparison of the obtained results with the results for ResMap.

5 Acknowledgments

This research was supported by the Helmholtz Association's Initiative and Networking Fund and the Russian Science Foundation (Project No. 18-41-06001). This work was carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-Science Facilities at the NRC "Kurchatov Institute", <http://ckp.nrcki.ru/>.

References

1. Nakane, T., Kotecha, A., Sente, A., et al: Single-particle cryo-EM at atomic resolution. *bioRxiv: the preprint server for biology* (2020).
2. Melero, R., Sorzano, C., Foster, B., et al: Continuous flexibility analysis of SARS-CoV-2 Spike prefusion structures. *bioRxiv: the preprint server for biology* (2020).
3. Ikonnikova, K.A., Teslyuk, A.B., Bobkov, S.A., Zolotarev, S.I., Ilyin, V.A.: Reconstruction of 3D structure for nanoscale biological objects from experiments data on super-bright X-ray free electron lasers (XFELs): dependence of the 3D resolution on the experiment parameters, *Procedia Computer Science* 156, 49-58 (2019).
4. Van Heel, M., Schatz, M.: Fourier shell correlation threshold criteria. *Journal of Structural Biology* 151(3), 250-262 (2005).
5. Rosenthal, P.B., Henderson, R.: Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of molecular biology* 333(4), 721-745(2003).
6. Van Heel, M., Schatz, M.: Reassessing the revolutions resolutions. *bioRxiv: the preprint server for biology* (2017).
7. Sorzano, CO, Vargas J, Otón J, et al.: A review of resolution measures and related aspects in 3D Electron Microscopy. *Progress in Biophysics and Molecular Biology* 124, 1-30 (2017).
8. Beckers, M.: Statistical Inference of cryo-EM Maps. *European Molecular Biology Laboratory (EMBL), Heidelberg* (2020).
9. Beckers, M., Sachse, C.: Permutation testing of Fourier shell correlation for resolution estimation of cryo-EM maps. *Journal of Structural Biology* (2020), doi: <https://doi.org/10.1016/j.jsb.2020.107579>
10. Protein Data Bank in Europe. <https://www.ebi.ac.uk/pdbe/emdb/>, last accessed 2020/07/12
11. Mandl, T., Östlin, Ch., Dawod, I. E., et al: Structural Heterogeneity in Single Particle Imaging Using X-ray Lasers. *J. Phys. Chem. Lett.* 11, 6077–6083 (2020)
12. Vilas, J.L., Heymann J.B., Tagare, H.D, Ramirez-Aportela, E, Carazo JM, Sorzano COS: Local resolution estimates of cryo-EM reconstructions, *Current Opinion in Structural Biology* 64, 74-78 (2020).
13. Cardone, G., Heymann, J. B., Steven, A. C.: One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *Journal of structural biology*, 184(2), 226–236 (2013).

14. Kucukelbir, A., Sigworth, F.J., Tagare, H.D.: Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65 (2014).
15. Vilas, J.L., Gómez-Blanco, J., Conesa, P., et al.: MonoRes: Automatic and accurate estimation of local resolution for electron microscopy maps. *Structure* 26, 337–344 (2018).
16. Vilas, J.L., Tagare, H.D., Vargas, J. et al.: Measuring local-directional resolution and local anisotropy in cryo-EM maps. *Nat Commun* 11, 55 (2020).
17. Ramírez-Aportela, E., Mota, J., Conesa, P., Carazo, J. M., Sorzano, C.: DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ*, 6(6), 1054–1063 (2019).
18. Gatsogiannis C, Markl J.: Keyhole limpet hemocyanin: 9-A CryoEM structure and molecular model of the KLH1 didecamer reveal the interfaces and intricate topology of the 160 functional units. *Journal of Molecular Biology* 385(3), 963-983(2009).
19. The Protein Data Bank. <https://www.rcsb.org/structure/4BED>, last accessed 2020/07/12.