Journal of Structural Biology 213 (2021) 107771

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Identification of incorrectly oriented particles in cryo-EM single particle analysis

Jeison Méndez [a], Edgar Garduño [d], José María Carazo [c], Carlos Oscar S. Sorzano [b,c,*]

[a] Posgrado en Ingeniería Eléctrica, Universidad Nacional Autónoma de México, Cd. Universitaria, C.P. 04510, Mexico City, Mexico
[b] Univ. San Pablo CEU, Campus Urb. Montepríncipe s/n, 28668, Boadilla del Monte, Madrid, Spain
[c] National Center of Biotechnology, CSIC, Campus Univ. Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain
[d] Department of Computer Science, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

ARTICLE INFO

Keywords:
Cryo-electron microscopy
Cryo-EM
Single particle analysis
Angular assignment validation

ABSTRACT

The quality of a 3D map produced by the single-particle analysis method is highly dependent on an accurate assignment of orientations to the many experimental images. However, the problem's complexity implies the presence of several local minima in the optimized goal functions. Consequently, validation methods to confirm the angular assignment are very useful to yield higher-resolution 3D maps. In this work, we present a graph-signal-processing-based methodology that analyzes the correlation landscape as a function of the orientation, an approach allowing the estimation of the assigned orientations' reliability. Using this method, we may identify low-reliability images that probably incorrectly contribute to the final 3D reconstruction.

1. Introduction

Knowing the three-dimensional (3D) structure of a macromolecule is essential for understanding its function, dynamics and conformational changes. Hence, in the last decades, there have been great advances to obtain the 3D structure of macromolecules, and single-particle analysis (SPA) has become one of the most widespread in recent years because of its flexibility and good results.

The single-particle technique is, at its core, a tomographic imaging modality where multiple copies of a purified macromolecule are imaged with an electron microscope under frozen–hydrated conditions to produce a 3D density map of Coulomb potential. This technique's basic idea is that several copies of the same macromolecule are randomly oriented during image acquisition producing multiple different views of the same macromolecule in a single image known as a micrograph. An important characteristic of the SPA method is that these views are considered individual projections of the macromolecule. Another important peculiarity of this modality is that the projections acquired with a transmission electron microscope, under cryogenic–conditions, are extremely noisy (i.e., signal-to-noise ratios, or SNRs, between 0.1 and 0.01) and, thus, require robust data analysis methods. At this point it is worth noting that the assumption of all the macromolecules in a micrograph being of identical structure and differing only in their orientation is in practice usually off; hence, there are significant efforts to achieve such a goal by identifying heterogeneous sets of projections.

A major requirement of any tomographic modality is the precise knowledge of every projection's orientation; the more exact this information is, the more precise the information is in the final 3D map. However, an undesirable consequence of the SPA method is that the information about the orientation of every copy of the macromolecule under investigation is largely unknown. Therefore, some matching process is necessary to obtain such information that in practice consists of three angles for the orientation and two in-plane displacements (in the recent years a new family of algorithms that do not require an explicit angular assignment has started to appear (Sharon et al., 2020)).

Although there exist many algorithms to estimate these orientation parameters (Scheres, 2012; Sorzano et al., 2015; Punjani et al., 2017; Grant et al., 2018; Reboul et al., 2018; Sorzano et al., 2018), all of them making significant efforts to avoid local minima, the fraction of incorrectly aligned particles any of them bring forth can be non-negligible (Vargas et al., 2016, 2017). The objective function used for the alignment plays an important role (Stewart and Grigorieff, 2004; Jonić et al., 2005; Sorzano et al., 2015) and considering simultaneously several objective functions may be advantageous because it might happen that a local minimum of an objective function is unlikely to coincide with local minima of the others.

* Corresponding author.
E-mail addresses: jmendez@utp.edu.co (J. Méndez), edgargar@ieee.org (E. Garduño), carazo@cnb.csic.es (J.M. Carazo), coss@cnb.csic.es (C.O. S. Sorzano).

https://doi.org/10.1016/j.jsb.2021.107771
Received 5 March 2021; Received in revised form 23 June 2021; Accepted 18 July 2021
Available online 26 July 2021

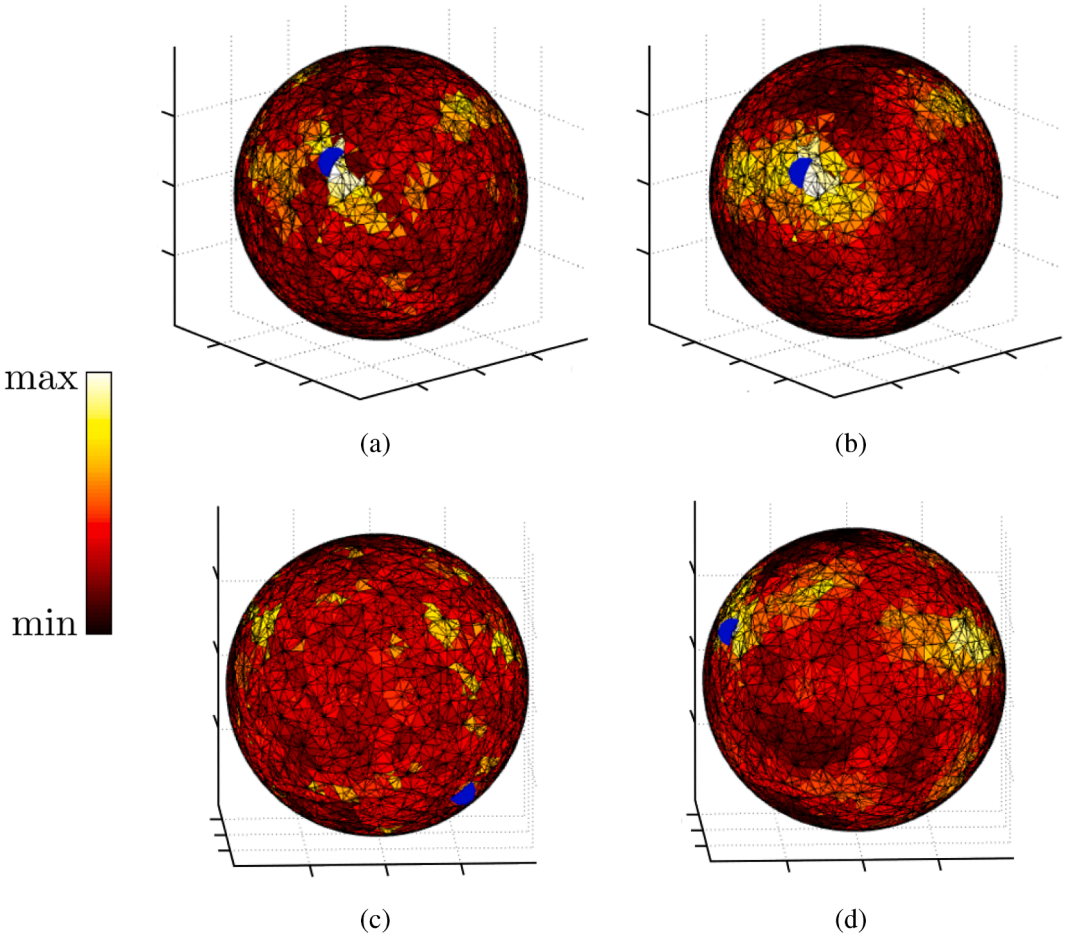

Fig. 1. (a,c) Graphical representations of the similarity measurements between an experimental projection f and all the reference projections in $\mathscr{R}$ computed with the Pearson's correlation coefficient for two different experimental images. (b,d) Graphical representation of the lowpass filtered version of (a,c), respectively. The color bar on the left shows the range of the similarity measures for all the graphs. Blue dots are superimposed where the maximum values occur on every graph. The angular distance between the blue dots on the raw graph and the lowpass filtered version provides a reliability criterion for the orientation assignment; in other words, the occurrence of highest values in the same region of both the raw and low-pass filtered graphs implies more reliable orientation assignments. Hence, the upper row (a,b) shows a case in which we are more confident about the angular assignment than in the case shown in the bottom row (c,d).

In this work, we propose a novel approach to validate the angular assignment performed by any method with the assumption that the orientation assigned to a particular projection should be consistent with the angular assignment it would have obtained had less noise been present in the goal function's landscape. In order to denoise the landscape of a goal function we use signal spectral decomposition based on graph theory because in that way, as it will be clearer below, we can exploit the spatial relationship between neighboring orientations in relation to the registration parameters and the similarity criterion used for matching projections; importantly, this approach will allow us to identify low-reliability projections. Signal processing on images using graph theory (using image-element adjacency graphs) has become common to perform on them tasks of filtering, segmentation, or clustering; whereas, in this work we use image adjacency graphs (images are at graph's vertices with edges determined by their orientation relationship) in a similar way to the approach proposed in (Xie et al., 2020), but here we utilize the image adjacency graphs to analyze the landscape of a goal function.

This paper is organized as follows. The next section presents a detailed description of our proposed validation methodology. Subsequently, in Section 3 we present results to evaluate our approach as a tool for validating previously-obtained angular assignments. Then, in Section 4 we discuss and analyze the results reported in the previous section.

2. Materials and methods

Most methods in SPA assign orientations to the set of experimental projections (the set $\mathscr{P}$) by using a set of reference projections (the set $\mathscr{R}$) obtained from a known 3D map. This process can be described as follows, for every experimental projection $f \in \mathscr{P}$ ($f : \mathbb{R}^2 \rightarrow \mathbb{R}$), an angular assignment method finds

$$\underset{g_i \in \mathscr{R}}{\operatorname{argmax}} \phi(g_i, M_i(f)), \tag{1}$$

where M_i is the optimal rigid transformation (i.e., an in-plane rotation together with a translation) aligning the experimental projection f with a reference projection $g_i \in \mathscr{R}$, and ϕ is a similarity measure between any two projections. This approach assumes that the reference projection's orientation information resulting from the search in Eq. (1) is assigned to the experimental projection f. For our method, we use the standard Pearson's correlation coefficient between the two projections being compared. However, our approach is general and can potentially accommodate any other similarity function.

Given a reference projection g_i, if the orientation assigned to it is correct, one should expect that any of its neighboring reference projections g_j, for $i \neq j$, would also yield a high similarity value (i.e., $\phi\left(g_j, M_i(f)\right)$ should be high). However, in practice there is the possibility of assigning an incorrect orientation to the experimental projection f due to

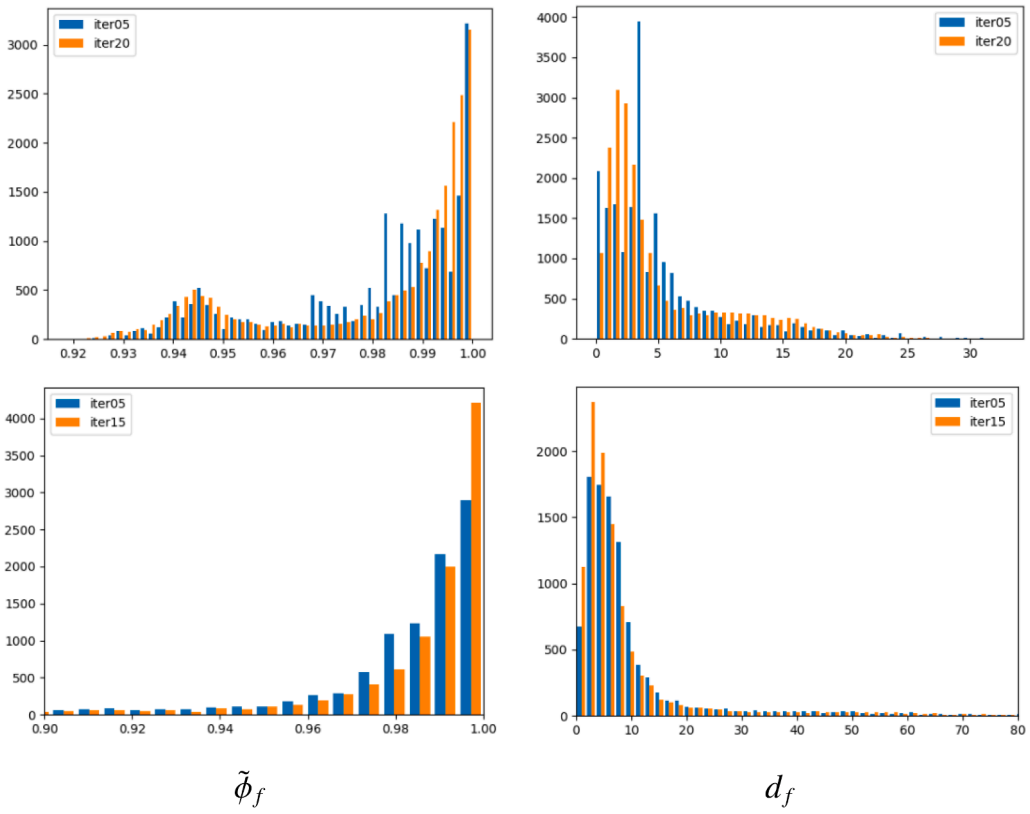

Fig. 2. Histograms for the measurements (*left*) $\tilde{\phi}_f$, and (*right*) d_f produced at different iterations during the Relion's assignment of orientations to the experimental projections of (*upper row*) the BMV and (*bottom row*) the β-galactosidase. The blue histograms were produced at the 5th iteration for both the BMV and β-galactosidase, whereas the orange histograms were produced at the 15th and 20th iterations for the β-galactosidase and BMV, respectively.

the presence of spurious similarity values for some of the neighboring reference projections g_j (i.e., high values of $\phi\left(g_j, M_i(f)\right)$ for some g_js in an otherwise low-valued-ϕ neighborhood).

In order to remove such spurious similarity values from the objective-function's landscape, we use a graph signal processing (GSP) approach (Zhu and Rabbat, 2012; Manoj et al., 2018; Waheed et al., 2020; Sandryhaila and Moura, 2013). This is a recently-developed branch of signal processing where its classical tools developed in the Euclidean domain have been generalized to irregular domains, including graphs; importantly, in this approach it is possible to incorporate the relationship among data samples as part of the analysis process (this is especially useful for the method we propose here because we compare any experimental projection against a set of synthetically generated reference projections from known orientations; projections that keep a spatial relationship among them). While classic signal processing deals with functions defined over a set of discretized values typically arranged over a regular grid and have some spatial relationship among themselves, graph signal processing deals with functions defined over the nodes of a graph, nodes that likely have an irregular distribution in space and might not have spatial interpretation.

A graph is defined by a set of vertices V and a set of edges E connecting the vertices. Additionally, every edge $e = (v, v') \in E$ may have a weight $w(v, v') \in \mathbb{R}$ associated to it. Given a graph defined by (V,E,w), we say that a graph signal is a function $\phi : V \rightarrow \mathbb{R}$ whose values are known for the graph vertices ($\phi(v)$ for all $v \in V$). The reader interested in the general topic of GSP is referred to (Zhu and Rabbat, 2012; Manoj et al., 2018; Waheed et al., 2020; Sandryhaila and Moura, 2013).

In our approach, every reference projection g_i is a vertex of the graph. An edge connects two reference projections g_i and g_j if their angular distance is closer than a maximum angle $D_{\max}$ (at this point, the angular sampling of the projection sphere determines the number of neighbors of a node). A decaying exponential gives the weight of any edge depending on the angular distance d

$$w_{ij} = \begin{cases} e^{-\frac{d\left(g_i, g_j\right)}{D_{\max}}}, & \text{if } d\left(g_i, g_j\right) < D_{\max}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Finally, the function value at each one of the vertices is given by $\phi(g_i, M_i(f))$. In our algorithm, we set $D_{\max}$ to be 3 times the angular sampling distance used to generate the set of reference projections, $\mathscr{R}$.

The Fourier transform of a function is the representation of such a function as a superposition of complex exponentials (Wang, 2012). Importantly, the eigenfunctions of the Laplace operator ∇^2 are precisely the Fourier basis (i.e., the Laplacian of the plane wave $e^{j2\pi x \cdot \xi}$ is equal to $-4\pi^2 \|\xi\|^2 e^{j2\pi x \cdot \xi}$). Thus, the Fourier transform is simply a method of expressing a function in terms of a sum of its projections onto the set of eigenfunctions (i.e., basis functions) while diagonalizing the Laplacian operator ∇^2.

To extend the concept of Fourier transform to a graph signal we take advantage of an interesting connection between the Laplace differential operator of a function and the Laplacian of a graph. Both, the Laplacian operator of a function and the Laplacian operator for graph signals are similar, except for a negative factor, because both of them provide information about how much a function (respectively, a graph signal) differs in a point from its average value taken over the neighboring points. Thus, to obtain the Fourier transform of a graph signal is necessary to calculate first the Laplacian of the graph signal (i.e., $\nabla^2\phi = \nabla \cdot \nabla\phi$). The Laplacian matrix $\mathbf{L}$ of the graph is a $|V| \times |V|$ symmetric matrix whose elements are defined as

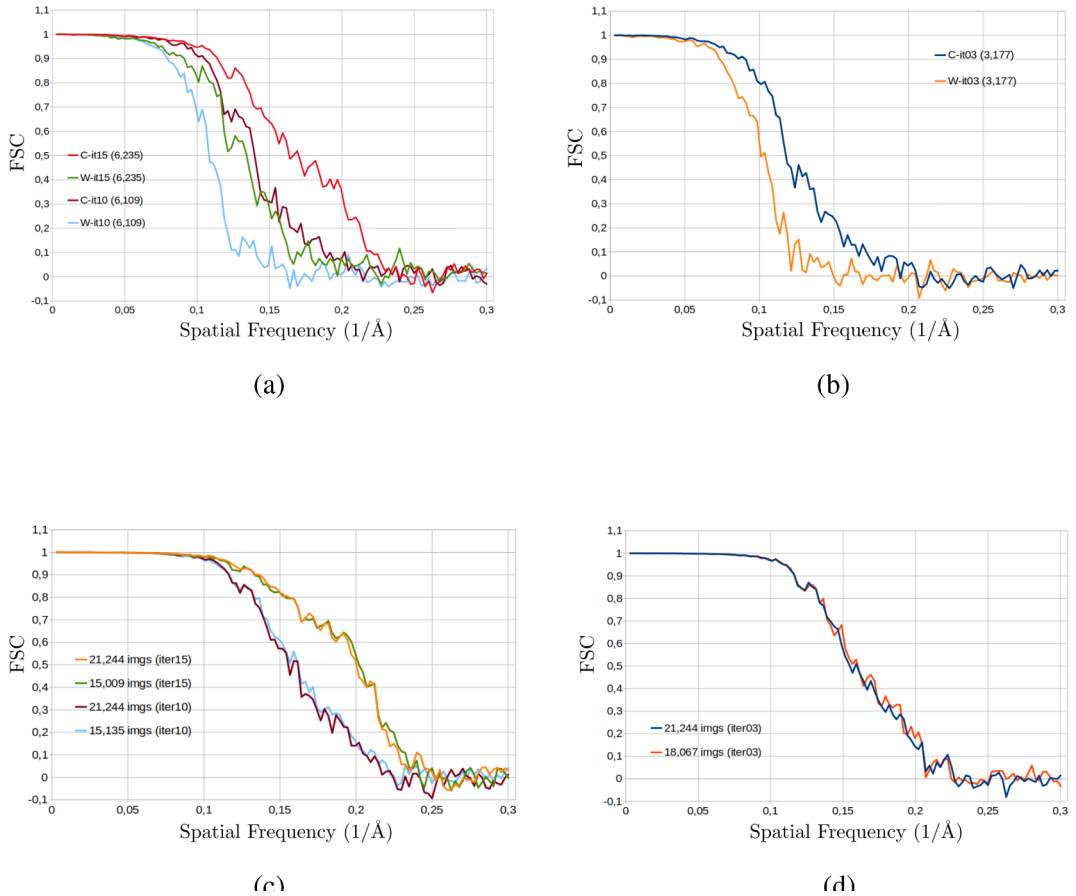

Fig. 3. Different FSCs corresponding to 3D maps produced with several methods using the experimental projections of the BMV. Each image in the upper row shows the FSCs corresponding to 3D maps obtained with projections in the C and the W sets using (a) Relion after the 10th and 15th iteration, respectively, and (b) Xmipp's highres after the 3rd iteration; for all the curves, the number of projections in the W set is presented in parentheses. Additionally, each image in the bottom row presents the FSCs corresponding to 3D maps obtained with projections in the C set and the whole dataset using (c) Relion stopping at different iterations (shown in parentheses) and (d) Xmipp's highres after the 3rd iteration.

$$l_{ij} = \begin{cases} \sum_{k \neq i} w_{i,k}, & \text{if } i = j, \\ -w_{ij}, & \text{if } i \neq j, \end{cases} \tag{3}$$

where w_{ij} is the edge weight between the ith and jth nodes obtained with (2). A direct consequence of $\mathbf{L}$ being symmetric is that it has associated to it a set $\{\lambda_1, \ldots, \lambda_{|V|}\}$ of real, non-negative eigenvalues (these eigenvalues are sorted in nondecreasing order, $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{|V|}$), and a set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{|V|}\}$ of corresponding real eigenvectors. The eigenvalues of $\mathbf{L}$ play the role of frequency and its eigenvectors play the role of the Fourier basis (complex exponentials in the case of standard signal or image processing). Hence, the Graph Fourier Transform can be defined as

$$\widehat{\phi}_m = \sum_{i=1}^{|V|} \phi\left(g_i\right) u_{mi}, \tag{4}$$

where u_{mi} is the i-th component of the m-th eigenvector (note that m plays the role of Fourier index, as in the Discrete Fourier Transform). Conversely, the Inverse Fourier Transform is given by

$$\phi\left(g_i\right) = \sum_{m=1}^{|V|} \widehat{\phi}_m u_{mi}. \tag{5}$$

As with the standard Fourier transform, we have the following Parseval's identity

$$\sum_{i=1}^{|V|} (\phi(g_i))^2 = \sum_{m=1}^{|V|} \left(\widehat{\phi}_m\right)^2. \tag{6}$$

It is worth mentioning that in this equation $(\widehat{\phi}_m)^2$ is the energy of the m-th Fourier component.

At this point is worth mentioning the following interesting fact, a standard discrete image defined over a regular grid can be represented with a graph whose vertices are the image pixels and its edges are determined by the adjacency relation between pixels sharing an edge (this representation includes the standard periodic boundary conditions used for Fourier analysis). The Graph Fourier Transform of such a representation is exactly the same as the standard 2D Discrete Fourier Transform of the image (Gray, 2006). However, the Graph Fourier Transform is much more general as it can deal with arbitrary adjacencies, irregular sampling grids, and, even, graphs whose vertices are not associated to any spatial domain (i.e., they may be words, objects, or any other entity).

We may exploit the Graph Fourier Transform of the landscape of our proposed similarity function ϕ for the reference g_i (i.e., $\widehat{\phi}(g_i)$) to remove noise from its associated graph signal. To achieve this, we can apply an ideal lowpass filter to the graph signal by simply truncating the sum in the Inverse Graph Fourier Transform to a smaller set with only M components ($M < |V|$)

$$\widetilde{\phi}\left(g_i\right) = \sum_{m=1}^{M} \widehat{\phi}_m u_{mi}. \tag{7}$$

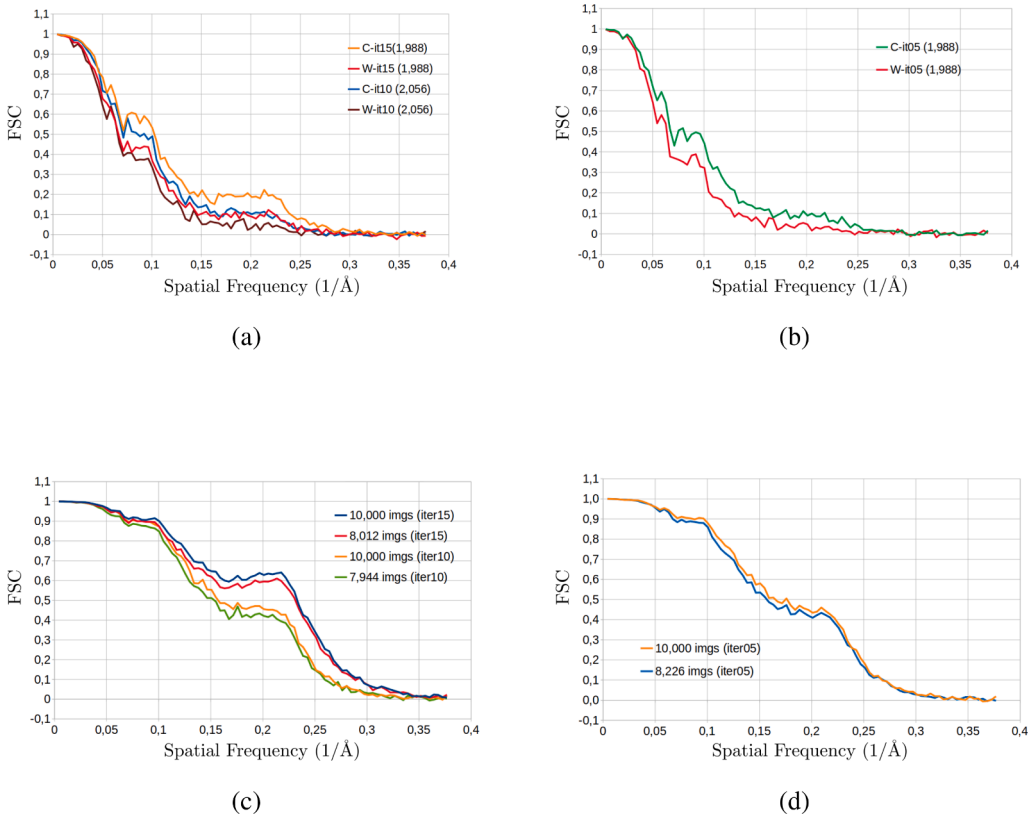

Fig. 4. Different FSCs corresponding to 3D maps produced with several methods using the experimental projections of the β-galactosidase. Each image in the upper row shows the FSCs corresponding to 3D maps obtained with projections in the C and the W sets using (a) Relion after the 10th and 15th iteration, respectively, and (b) Xmipp's highres after the 5th iteration; for all the curves, the number of projections in the W set is presented in parentheses. Additionally, each image in the bottom row presents the FSCs corresponding to 3D maps obtained with projections in the C set and the whole dataset using (c) Relion stopping at different iterations (shown in parentheses) and (d) Xmipp's highres after the 5th iteration.

In our algorithm, we choose the subset of M Fourier components to be those with the largest energies and whose combined energies amount for 95% of the total spectral energy. Hence, the filtered landscape $\widetilde{\phi}(g_i)$ is a smoother version of the landscape of similarities (see an example in Fig. 1).

Once we have the raw landscape of similarities and its lowpass filtered version, we look for the projection direction with maximum similarity in both graphs. Let us refer to them as $\mathbf{v}_f$ and $\widetilde{\mathbf{v}}_f$ (both vectors having unit norm). We then define their similarity as

$$d_f = \mathbf{v}_f, \widetilde{\mathbf{v}}_f, \tag{8}$$

where $.,.$ is the standard inner product in $\mathbb{R}^3$.

Another interesting quality criterion is the correlation between the two reference projections associated with both directions $\mathbf{v}_f$ (best matching reference in the unfiltered graph) and $\widetilde{\mathbf{v}}_f$ (best matching reference in the filtered graph), let us refer to them as g_i and $\widetilde{g}_i$. Although the two best directions may be quite far apart, the corresponding references may be similar. We try to detect such a situation by aligning projections $\widetilde{g}_i$ to g_i and comparing them

$$\widetilde{\phi}_f = \phi\left(g_i, \widetilde{M}_i\left(\widetilde{g}_i\right)\right). \tag{9}$$

The quantities d_f and $\widetilde{\phi}_f$ are two qualifiers of the quality of the angular assignment of the image f.

3. Results

We implemented our method within the Scipion image processing framework for cryo-EM (de la Rosa-Trevín et al., 2016) using the Xmipp image processing suite for microscopy (Sorzano et al., 2004).

We illustrate the applicability of our method with three experimental examples. The first dataset comes from the Brome Mosaic Virus - BMV (EMPIAR-10010, 2014; Wang et al., 2014), the second from a β-galactosidase (EMPIAR-10061, 2016; Bartesaghi et al., 2015), and the third from an in–house acquired apoferritin. The 21,244 projections of the BMV were acquired with a voltage of 300 kV, a 50,000× magnification, a spherical aberration of 4.1 mm, an amplitude contrast of 0.1, and pixel size equal to 0.99 Å/pix. The 10,000 projections for the β-galactosidase were acquired with a voltage of 300 kV, a 50,000× magnification, a spherical aberration of 2.7 mm, an amplitude contrast of 0.1, and pixel size equal to 0.32 Å/pix. Lastly, the 46,182 projections for the apoferritin were acquired with a voltage of 300 kV, a 50,000× magnification, a spherical aberration of 2.7 mm, an amplitude contrast of 0.1, and pixel size equal to 0.94 Å/pix. We followed a standard image analysis pipeline, such as the one described in (Sorzano et al., 2020), and used Relion's autoreffne and Xmipp's highres to produce the final maps of both datasets.

After obtaining the angular assignments, and 3D maps, for the two datasets we apply the methodology of Section 2 to determine the quality of these angular assignments by obtaining their corresponding measurements $\widetilde{\phi}_f$ and d_f with our graph-based method. In Fig. 2 we present histograms exemplifying how these measurements behave when validating the assigned orientations; concretely, they are produced at two different iterations of Relion during its assignment of orientations to the experimental projections of both the BMV and the β-galactosidase. These histograms demonstrate that the higher the values for $\widetilde{\phi}_f$ and the lower the values for d_f the better the indication is for a good quality of orientation assignment.

To automate the classification into reliably (C) and non-reliably (W) assigned orientations, we have segmented both histograms using the method proposed by Otsu (Otsu, 1979). The assumption being that images whose orientations are described as reliable will produce better 3D reconstructions and to verify this assumption we reconstruct a 3D map using the same number of projections either from the smaller set, normally the non-reliably assigned, or a random subset of a predetermined-size chosen from the larger set. Finally, we compare the

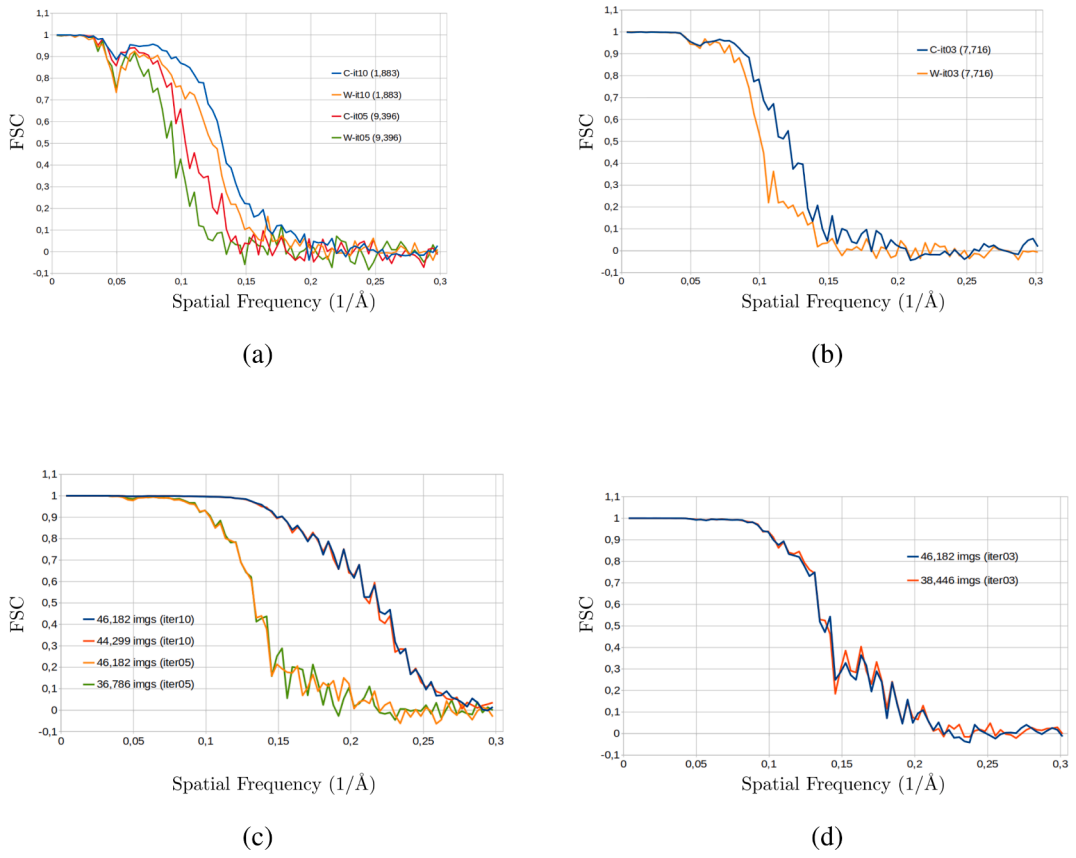

Fig. 5. Different FSCs corresponding to 3D maps produced with several methods using the experimental projections of the apoferritin. Each image in the upper row shows the FSCs corresponding to 3D maps obtained with projections in the C and the W sets using (a) Relion after the 5th and 10th iteration, respectively, and (b) Xmipp's highres after the 3rd iteration; for all the curves, the number of projections in the W set is presented in parentheses. Additionally, each image in the bottom row presents the FSCs corresponding to 3D maps obtained with projections in the C set and the whole dataset using (c) Relion stopping at different iterations (shown in parentheses) and (d) Xmipp's highres after the 3rd iteration.

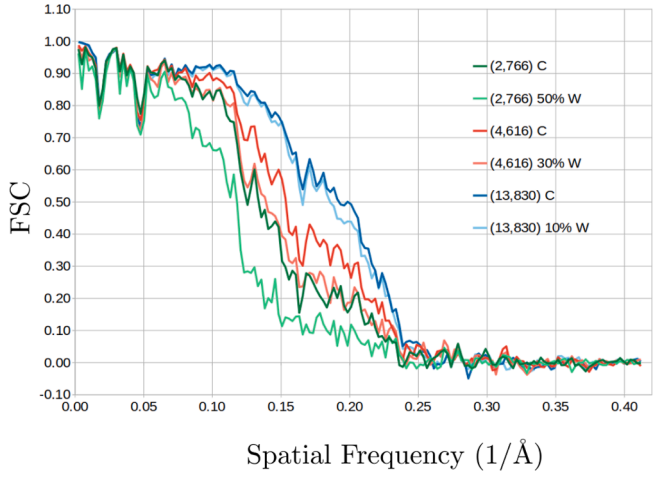

Fig. 6. Comparison of the FSCs produced when comparing the high-resolution 3D map of BMV (EMD-6000 from the Electron Microscopy Data Bank) against the 3D maps obtained from the two datasets when the W contains (green) 50%, (red) 30%, and (blue) 10% of misaligned projections, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

resolution of both sets, reliable and non-reliable, with the FSC.

In Fig. 3 we present results obtained with the experimental projections from the BMV. Fig. 3a shows the FSCs, with different line colors, corresponding to the 3D maps produced with Relion when using the projections classified in sets C and W for the orientation-assignment at the 10th and 15th iteration, respectively. In turn, Fig. 3b shows the FSCs, also with different line colors, corresponding to the 3D maps produced with Xmipp's highres using sets C and W. These figures show that in every case, the 3D maps produced with orientation-assigned projections in the C sets have better resolution than their counterparts produced with projections in the W sets. For additional comparison, we present in Figs. 3c and 3d the FSCs corresponding to the 3D maps reconstructed with all the experimental projections in the dataset and the 3D maps reconstructed only with the C sets. Fig. 3c corresponds to the FSCs associated with the 3D maps produced with Relion using the experimental projections whose orientations have been assigned at the 10th and 15th iteration, respectively. On the other hand, Fig. 3d corresponds to the FSCs associated with the 3D maps produced with Xmipp's highres.

For the β-galactosidase and the apoferritin, we reproduce the experiments of Fig. 3 and present the corresponding results in Figs. 4 and 5, respectively. Similar to the experiments with the BMV, the results with the β-galactosidase and the apoferritin also suggest that in every case, the 3D maps produced with orientation-assigned projections in the C sets have better resolution than their counterparts produced with projections in the W sets.

Finally, in order to make clearer the usefulness of the proposed validation method we include the results for an experiment with the BMV in which we study the effect of including projections identified as potentially misaligned during the reconstruction of the 3D map. For this, we used the high-resolution map EMD-6000 of the virus, obtained from the Electron Microscopy Data Bank (EMDB-6000, 2014; Wang et al., 2014), as a reference for FSC comparison. Our validation method identified 1,383 projections as potentially misaligned for the BMV

dataset; based on this result, two equal-size sets of projections were formed, the set C containing only projections identified as correctly assigned and the set W containing complementary percentages of correctly-assigned and misaligned projections. Then, these sets of projections were used to reconstruct two 3D maps that were compared against the reference 3D map. We performed this experiment with the second set of projections containing 50%, 30%, and 10% of projections identified as misaligned, respectively. We report the results of these experiments in Fig. 6 where we can see that the higher the percentage of potentially misaligned projections in a set, the greater the difference between the FSC curves.

4. Discussion and conclusions

In the course of reconstructing a 3D map, there is the possibility of some errors occurring in the refinement process when using the projection-matching approach and when they occur they negatively affect the quality of the resulting map. These errors are mainly due to suboptimal precision in the estimation of the projections' orientations (usually stemming from errors in the registration among projections) and heterogeneity (either structural or compositional) of the specimens sampled in the experimental projections and undetected during the classification stage. Furthermore, the problem of suboptimal precision in the assignment of orientations to experimental projections can become persistent as a consequence of reference bias, a phenomenon resulting from the incorrect registration of reference projections to data or noise in experimental projections, (Henderson, 2013; van Heel, 2013) or overfitting, a phenomenon occurring when registering high-frequency noise wrongly interpreted as high-resolution structural features, (Chen et al., 2013; Stewart and Grigorieff, 2004).

Currently, the established practice in cryo-EM is to validate the final reconstructed 3D maps. Although there have been several attempts to validate the 3D angular assignment (Vargas et al., 2016, 2017), these techniques are not yet widely adopted and the problem is sufficiently open to new explorations.

We have proposed a methodology to validate such assignments and the results presented in Section 3 show that the assignment validation approach allows identifying projections that contribute little to improving the map's resolution because their angular assignment is probably wrong according to our proposed quality criteria. Notably, such a validation procedure can be used in conjunction with any orientation-assigning method as part of any refinement process, where it can assist in the exclusion of poorly registered projections (that otherwise could continue having their orientations wrongly assigned during the refinement process) and, thus, facilitating the reconstruction of higher-resolution 3D maps. Based on the growing interest in the cryo-EM community to provide tests to validate reported 3D maps, the orientation-assignment validation method we are proposing could complement those existing approaches, either at the refinement-process level or when cross-validating reconstructed 3D maps.

The quality of structures determined by cryo-EM SPA can continue to improve by introducing novel image processing methods hand in hand with more robust validation tools that can identify small differences in high-resolution 3D maps and permit the application of SPA to more complex or heterogeneous specimens (Rosenthal and Rubinstein, 2015; Rosenthal, 2016). Certainly, these computational tools need to work conjointly with other SPA processes such as specimen preparation and projection acquisition.

CRediT authorship contribution statement

Jeison Méndez: Conceptualization, Methodology, Software, Investigation, Validation, Writing - original draft. **Edgar Garduño:** Conceptualization, Formal analysis, Writing - review & editing, Supervision, Funding acquisition. **José María Carazo:** Writing - review & editing, Supervision, Funding acquisition. **Carlos Oscar S. Sorzano:** Conceptualization, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

A CONACyT-Mexico graduate program grant supports the work of Jeison Méndez No. 487646. A DGAPA-UNAM sabbatical research grant supported the work of Edgar Garduño.

References

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L., Subramaniam, S., 2015. 2.2 Å resolution cryo-EM structure of β)galactosidase in complex with a cell-permeant inhibitor. Science 348, 1147–1151.

Chen, S., McMullan, G., Faruqi, A.R., Murshudov, G.N., Short, J.M., Scheres, S.H.W., Henderson, R., 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy 135, 24–35.

de la Rosa-Trevín, J.M., Quintana, A., del Cano, L., Zaldívar-Peraza, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castells, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C.O.S., Carazo, J., 2016. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. Journal of Structural Biology 195, 93–99.

EMDB-6000, 2014. ID: EMD-6000. Full virus map of Brome Mosaic Virus. https://pdbj.org/emnavi/quick.php?id=EMD-6000.

EMPIAR-10010, 2014. Entry 10010 of the Electron Microscopy Public Image Archive (EMPIAR). https://www.ebi.ac.uk/pdbe/emdb/empiar/entry/10010/. doi: 10.6019/EMPIAR-10010.

EMPIAR-10061, 2016. Entry 10061 of the Electron Microscopy Public Image Archive (EMPIAR). https://www.ebi.ac.uk/pdbe/emdb/empiar/entry/10061/. doi: 10.1126/science.aab1576.

Grant, T., Rohou, A., Grigorieff, N., 2018. cisTEM, user-friendly software for single-particle image processing. eLife, 7. doi: 10.7554/eLife.35383.

Gray, R.M., 2006. Toeplitz and circulant matrices: a review. Foundations and Trends in Communications and Information Theory, 2, 155–239. Originally published as Information Systems Laboratory Technical Report, Stanford University, 1971.

Henderson, R., 2013. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 110, pp. 18037–18041.

Jonić, S., Sorzano, C.O.S., Thévenaz, P., El-Bez, C., De Carlo, S., Unser, M., 2005. Spline-based image-to-volume registration for three-dimensional electron microscopy. Ultramicroscopy 103, 303–317.

Manoj, B.S., Chakraborty, A., Singh, R., 2018. Complex Networks: A Networking and Signal Processing Perspective. Prentice Hall.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66.

Punjani, A., Brubaker, M.A., Fleet, D.J., 2017. Building proteins in a day: Efficient 3D molecular structure estimation with electron cryomicroscopy. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 706–718.

Reboul, C.F., Eager, M., Elmlund, D., Elmlund, H., 2018. Single-particle cryo-EM-improved ab initio 3D reconstruction with simple/prime. Protein Science 27, 51–61.

Rosenthal, P.B., 2016. Chapter nine – testing the validity of single-particle maps at low and high resolution, In: R.A. Crowther (Ed.), The resolution revolution: recent advances in cryoEM (pp. 227–253). Academic Press volume 579 of Methods in Enzymology.

Rosenthal, P.B., Rubinstein, J.L., 2015. Validating maps from single particle electron cryomicroscopy. Current Opinion in Structural Biology 34, 135–144.

Sandryhaila, A., Moura, J.M.F., 2013. Discrete signal processing on graphs: Graph Fourier Transform. In: ICASSP 2013–2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6167–6170.

Scheres, S.H.W., 2012. A Bayesian view on Cryo-EM structure determination. Journal of Molecular Biology 415, 406–418.

Sharon, N., Kileel, J., Khoo, Y., Landa, B., Singer, A., 2020. Method of moments for 3-D single particle ab initio modeling with non-uniform distribution of viewing angles. Inverse Problems 36, 044003.

Sorzano, C.O.S., Marabini, R., Velázquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., Pascual-Montano, A., 2004. Xmipp: A new generation of an open-source image processing package for electron microscopy. Journal of Structural Biology 148, 194–204.

Sorzano, C.O.S., Vargas, J., de la Rosa-Trevín, J.M., Otón, J., Álvarez Cabrera, A., Abrishami, V., Sesmero, E., Marabini, R., Carazo, J.M., 2015. A statistical approach to the initial volume problem in Single Particle Analysis by Electron Microscopy. Journal of Structural Biology 189, 213–219.

Sorzano, C.O.S., Vargas, J., de la Rosa-Trevín, J.M., Jiménez, A., Maluenda, D., Melero, R., Martínez, M., Ramírez-Aportela, E., Conesa, P., Vilas, J.L., Marabini, R., Carazo, J.M., 2018. A new algorithm for high-resolution reconstruction of single particles by electron microscopy. Journal of Structural Biology.

Sorzano, C.O.S., Jimenez-Moreno, A., Maluenda, D., Ramirez-Aportela, E., Martinez, M., Cuervo, A., Melero, R., Conesa, P., Sanchez-Garcia, R., Strelak, D., 2020. Image processing in cryo-electron microscopy of single particles: the power of combining methods. In: Filipovic, J., Fernandez-Gimenez, E., de Isidro-Gomez, F., Herreros, D., Conesa, P., del Caño, L., Fonseca, Y., Jimenez de la Morena, J., Macias, J.R., Losada, P., Marabini, R., Carazo, J.M. (Eds.), Methods in Molecular Biology: Structural Proteomics. (in press).

Stewart, A., Grigorieff, N., 2004. Noise bias in the refinement of structures derived from single particles. Ultramicroscopy 102, 67–84.

van Heel, M., 2013. Finding trimeric HIV-1 envelope glycoproteins in random noise. In: Proceedings of the National Academy of Sciences of the United States of America, vol. 110, pp. E4175–E4177.

Vargas, J., Oton, J., Marabini, R., Carazo, J.M., Sorzano, C.O.S., 2016. Particle alignment reliability in single particle electron cryomicroscopy: A general approach. Scientific Reports 6, 21626.

Vargas, J., Melero, R., Gomez-Blanco, J., Carazo, J.M., Sorzano, C.O.S., 2017. Quantitative analysis of 3D alignment quality: Its impact on soft-validation, particle pruning and homogeneity analysis. Scientific Reports 7, 6307.

Waheed, W., Deng, G., Liu, B., 2020. Discrete Laplacian operator and its applications in Signal Processing. IEEE Access 8, 89692–89707.

Wang, R., 2012. Introduction to Orthogonal Transforms: With Applications in Data Processing and Analysis. Cambridge University Press, Cambridge.

Wang, Z., Hryc, C.F., Bammes, B., Afonine, P.V., Jakana, J., Chen, D.H., Liu, X., Baker, M. L., Kao, C., Ludtke, S.J., Schmid, M.F., Adams, P.D., Chiu, W., 2014. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. Nature Communications 5, 4808 – 4808.

Xie, R., Chen, Y.-X., Cai, J.-M., Yang, Y., Shen, H.-B., 2020. SPREAD: A fully automated toolkit for single-particle cryogenic electron microscopy data 3D reconstruction with image-network-aided orientation assignment. Journal of Chemical Information and Modeling 60, 2614–2625. https://doi.org/10.1021/acs.jcim.9b01099.

Zhu, X., Rabbat, M., 2012. Approximating signals supported on graphs. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3921–3924.