Cryo-EM map anisotropy can be attenuated by map post-processing and a new method for its estimation

Authors

Ruben Sanchez-Garcia[a]*, Jose Manuel Cuadra-Troncoso[b] and Javier Vargas[c]*

[a]Department of Statistics, University of Oxford, 24-29 St Giles', Oxford, Oxfordshire, OX1 3LB, United Kingdom

[b]Departamento de Inteligencia Artificial, National University of Distance Education, C. Juan del Rosal, 16, Madrid, Madrid, 28040, Spain

[c]Departamento de Óptica, Complutense University of Madrid, C. de Arcos de Jalón, 118, Madrid, Madrid, 28037, Spain

Correspondence email: ruben.sanchez-garcia@stats.ox.ac.uk; jvargas@ucm.es

Funding information Ministerio de Ciencia e Innovación (grant No. PID2019-108850RA-I00 to Javier Vargas).

Abstract One of the most important challenges in cryogenic electron microscopy (cryo-EM) is the substantial number of samples that exhibit preferred orientations, which leads to an uneven coverage of the projection sphere. As a result, the overall quality of the reconstructed maps can be severely affected, as manifested by the presence of anisotropy in the map resolution. Several methods have been proposed to measure the directional resolution of maps in tandem with experimental protocols to address the problem of preferential orientations in cryo-EM. Following these works, in this manuscript we show that some of the most recently proposed cryo-EM map post-processing algorithms can attenuate map anisotropy, thus offering a computational alternative for cases affected by moderate levels of preferential orientations. In addition, we identified one potential limitation that may affect most of the existing methods for anisotropic resolution estimation and we proposed an alternative method to evaluate the presence of preferential orientations in cryo-EM reconstructions.

Keywords: Cryo-EM, map anisotropy, preferred orientations, algorithms, post-processing.

1. Introduction

Ten years after the beginning of the resolution revolution, cryogenic electron microscopy (cryo-EM) has become one of the most popular approaches in structural biology, delivering thousands of high-resolution structures for a wide range of macromolecules. Despite being used routinely, cryo-EM is

far from being a fully automatic technique and important challenges have yet to be resolved. One of such recurring problems is the preferred orientations that some specimens tend to exhibit when vitrified (D'Imprima and Kühlbrandt, 2021).

Preferred orientations are a consequence of the interactions between the particles and the air-water or substrate-water interfaces (Naydenova & Russo, 2017; Noble, Dandey *et al.*, 2018). This may cause damage or degradation of the macromolecular structure, compositional heterogeneity and nonuniform distributions of angular projections leading to regions of the projection sphere with poor coverage (Chua *et al.*, 2022). As a result, the reconstructed maps may be artefactual because partial to complete sample denaturation and/or tend to exhibit worse resolution in the direction parallel to the preferred orientation axis. For severe cases, important artefacts as map stretching could also appear (Zi Tan *et al.*, 2017). These problems are generally described in terms of (directional) resolution anisotropy and can be explained by the lack of information generated by missing particle projections, as well as by the overabundance of certain projections that can hinder the reconstruction process (Sorzano *et al.*, 2021).

It has been widely accepted that resolution anisotropy is mostly a manifestation of the preferential orientations problem, and given its ubiquity, several metrics based on this idea have been proposed. One of the first proposed metrics was the 3D-FSC, in which the gold standard FSC resolution estimation is computed in a directional manner by splitting the shells into cones sections perpendicular to the radius of the volume (Zi Tan *et al.*, 2017). As a result, an independent FSC curve with an associated resolution estimation is obtained for each of the cones in which the sphere is discretized. As a summary of the per-cone resolution, the sphericity of the thresholded 3D-FSC volume was proposed. Based on a different principle, Naydenova and Russo (Naydenova & Russo, 2017) proposed the orientation distribution efficiency, which is a measurement of the sphericity of the point spread function derived from the distribution of angular assignments. More recent in (Vilas *et al.*, 2020; Vilas & Tagare, 2022) authors proposed a directional local resolution estimator based on the use of the monogenic signal over the EM map along different directions computed from directed cones.

All the methods presented above are based on analysing the directional resolution from a given 3D reconstruction (Vilas *et al.*, 2020; Zi Tan *et al.*, 2017) or a particle orientations distribution (Naydenova & Russo, 2017). Consequently, the estimated directional resolution depends obviously on the particle orientation distribution. However, as it will be shown in the Results Section "Common map anisotropy metrics are affected by the shape of the specimen", this magnitude also depends on the actual geometry of the sample to reconstruct, since the spectral signal to noise ratio (SSNR) at different directions may depend on the actual shape of the macromolecule. While this latter dependency has no impact when the metrics are used to compare different reconstructions of the same sample, it makes direct comparison of different specimens infeasible.

Map anisotropy metrics are invaluable diagnostic tools that can be employed to assess if a given approach can alleviate the preferred orientations problem. Examples of such experimental approaches are the usage of surfactants support films as a barrier (Glaeser & Han, 2017), alternative types of grids (Russo & Passmore, 2014; Han *et al.*, 2012; Wang *et al.*, 2020), reducing the time that the sample dwells in the thin liquid film before vitrification (Noble, Wei *et al.*, 2018) or re-imaging after tilting the sample to introduce missing views (Zi Tan *et al.*, 2017). Although these previously mentioned approaches have been found to be useful, the presence of preferential orientations is still a major hurdle for cryo-EM as it limits the throughput of the technique and could generate artefacts leading to misleading conclusions. From a computational perspective, although it is generally known that different algorithms exhibit different degrees of robustness against this problem, only little research has been carried out about this topic (Sorzano *et al.*, 2021) and, to the best of our knowledge, there are no explicitly designed algorithms for anisotropy correction apart from simple deconvolution strategies (Naydenova & Russo, 2017). Such methods would be especially interesting in the postprocessing step of the cryo-EM pipeline to provide as final map the best possible reconstruction without the need of reprocessing the datasets. Our experience points us to think that some of the recently published map postprocessing algorithms could alleviate the effect of preferential orientations in the post-processed map. However, to the best of our knowledge, no formal study has been carried out to determine if such perception is true.

In this work we have analysed how different map post-processing algorithm can alleviate the presence of anisotropic directional resolution in cryo-EM maps affected by preferential orientations. Our results show that most of these approaches can attenuate this problem for cases exhibiting moderate levels of anisotropy. In addition, we have also found a limitation that may affect all currently employed methods for anisotropy estimation and we have proposed an alternative approach to evaluate the presence of preferred orientation in cryo-EM maps that is not affected by this limitation.

2. Methods

2.1. Datasets

2.1.1. Artificial anisotropy

We processed the EMPIAR-10061 dataset (Bartesaghi *et al.*, 2015) with Relion 3 (Zivanov *et al.*, 2018) to obtain a β-galactosidase map at 3.15 Å (FSC unmasked) from 41,122 particles without applying any masking or post-processing operation. This map shows negligible amounts of anisotropy. Then, we removed 95% of the particles assigned to tilt angles above 70° degrees producing a particle set of 26409 particles which was again processed using Relion 3 to obtain a map at 3.40 Å resolution. Finally, we additionally reconstructed another map after removing 95% of the particles with tilt angles above 80° degrees (32,941 particles in total) resulting in a map at 3.19 Å.

This last map exhibit anisotropy levels between the raw map and the one obtained removing the particles at tilt angles > 70°. See Supplementary Figure S1 for a visual of the densities and the 3D-FSC plots.

2.1.2. Experimental maps

The first map was the EMD-20794 (Campbell *et al.*, 2020), a 3.3 Å resolution reconstruction of the Integrin αvβ8 in complex with L-TGF-β which exhibits moderate levels of anisotropy. In our analysis. The associated atomic model that we employed for 3D-FSC calculations was the PDB ID 6UJA.

The second specimen was the influenza hemagglutinin (HA) trimer obtained by Tan et al. (Zi Tan *et al.*, 2017). We employed the untitled version, severely affected by preferred orientations (EMPIAR ID 10096), and the tilted version (EMD-8731, EMPIAR ID 10097), in which the anisotropy is not as severe but still quite important. In our analysis, we employed the atomic model with PDB ID 7VDF fitted to the densities.

2.2. FSC-3D calculation

3D-FSC calculations were carried out comparing post-processed densities against the maps simulated from the associated atomic models employing 20° cones and leaving all the remaining parameters set to default. The simulated maps were generated using EMAN2 (Tang *et al.*, 2007) at the FSC global resolution. The employed masks were calculated using the simulated densities and binarized at threshold 0.01. A soft edge of 5 pixels was added to the masks.

2.3. Map post-processing

Four different map postprocessing algorithms were employed: DeepEMhancer (Sanchez-Garcia *et al.*, 2021), LocScale2 (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022), LocSpiral (Kaur *et al.*, 2021), and LocalDeblur (Ramírez-Aportela *et al.*, 2020). DeepEMhancer was executed with default parameters for all cases. LocScale2 was executed with default parameters using the atomic model, the global FSC resolution and the half maps as inputs. LocSpiral was executed with a threshold of 0.9 and a bandwidth manually picked at 4.5. LocSpiral and LocScale were executed using the same masks that were used latter for 3D-FSC estimation. Local resolution maps required for LocalDeblur were computed using Monores (Vilas *et al.*, 2018) using as inputs the half-maps.

2.4. New anisotropy method

As mentioned in the introduction, current approaches to evaluate the presence of anisotropic resolution in reconstructed maps are affected by the actual shape of the macromolecule. Thus, the resolution estimation through different orientations is convoluted with the actual shape of the macromolecule. Consequently, perfectly isotropic 3D reconstructions may be evaluated as anisotropic by current approaches (see Results section “Common map anisotropy metrics are affected by the

shape of the specimen" and Figure 6). A good approach to evaluate the possibility of having preferential orientations and thus anisotropic reconstructions in cryo-EM is to show the distribution of aligned single particles along the different possible orientations. However, this information, which is only an estimation, is not usually provided by the authors, thus it is not possible to calculate its posterior if only the maps are made public.

To robustly evaluate the presence of preferential orientations in cryo-EM reconstructions, we have developed an approach that is not affected by the shape of the reconstructed macromolecule. Our rational is that the noise present in the solvent mask should show isotropic power through different directions for reconstructions not affected by preferential orientations, irrespectively of the shape of the macromolecule. Thus, our approach first defines a spherical soft mask to filter out the macromolecular signal, applying then this mask to the map to analyse the anisotropy of the noise. To do this, the approach obtains the Fourier Transform of the masked map and extracts slices at different orientations given by the elevation and azimuth Euler's angles of the unit vectors perpendicular to the slices. This means, that an elevation angle given by 1.5 rad (or -1.5 rad) corresponds to a top view projection, while an elevation of 0 rad corresponds to a side view projection. For each extracted slide, the average noise power is computed for a defined resolution range between 10 Å and the corresponding Nyquist resolution. Finally, the highest computed noise power along all computed directions is used to normalize a 2D map representing the normalized noise power for different elevation and azimuth Euler's angles.

The reconstruction process is usually done in cryo-EM assuming that each particle projection contributes towards the reconstruction in every possible orientation, but with different weights. The weights are calculated by evaluating the similarity (or likelihood) of that particle with the projection of the map to be refined at every possible orientation. Thus, after every single particle is aligned, we may consider two different sets of particles according to their calculated weights. On one hand, the particles that were aligned with high precision will show a sharpened distribution of weights across the different orientations. Thus, in that case, the particles will show high weights (close to one) for a small set of close orientations, while very low weights (almost zero) for the rest of possible orientations. Note that the sum of all weights for each single particle must be 1. We might model these weights for a given particle with a Gaussian distribution with low variance. On the other hand, the particles that were not aligned with precision will show low and almost uniform weights along all possible orientations. In this case, we may model this particle weight distribution by a uniform distribution. The intensity by which the particle i contributes towards the reconstruction at a given orientation j, can be formalized as a random variable I_{ij}^{+}

$$I_{ij}^{+} = w_{ij}^{\alpha} I_i \qquad (1)$$

where I_i corresponds to the Fourier Transform of the particle i and w_{ij}^{α} the alignment weights that can be modelled as

$$w_{ij}^{\alpha} = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(\psi_j - \overline{\psi})^2}{\sigma^2} \right] & when\ \alpha = 1 \\ \frac{1}{N_{or}} & when\ \alpha = 0 \end{cases} \quad (2)$$

with α equal to 1 if the particle could be aligned with high precision or 0 if not, N_{or} the number of possible particle orientations, ψ_j and $\overline{\psi}$ a potential orientation and the actual orientation of a particle respectively and σ the variance of the Gaussian distribution. If we consider that the particles only contain noise and not macromolecule signal (they were aligned before according to the macromolecular signal, but for this analysis we are filtering out the signal by a soft mask), and that the particle noise is white Gaussian noise, then I_i will correspond also to white Gaussian noise. The resultant reconstructed volume slice in Fourier space will be given by

$$V_j = \frac{\sum_i w_{ij}^{\alpha} I_i}{\sum_i w_{ij}^{\alpha}} \quad (3)$$

Note that orientations where exist many particles that were aligned with high precision and high weights, will have a low volume slice signal as the noise will be effectively reduced by the average. On the other hand, for orientations corresponding to missing views, the particles that were oriented with precision to other orientations will have close to zero weights. Additionally, particles that were not aligned with precision will have small but higher weights in these orientations. As a result, if the number of particles that were not aligned with precision is small, the volume slice signal at these orientations will not be low as the noise will not be effectively reduced. Thus, the proposed approach can provide effectively *post-morten* information of the distribution of particles along projection directions without requiring the orientations of single particles. Additionally, note that typically particle orientation distributions are calculated without considering the particle's weight distributions, thus using only the orientation with highest weight for each particle, while cryo-EM reconstructions are not usually computed in this way. As explained, our proposed approach can evaluate the presence of anisotropic particle distributions taking into consideration the particle´s weight distributions used in the reconstruction and it is insensitive to the shape of the macromolecule.

3. Results and discussion

3.1. Nonlinear post-processing methods attenuate anisotropy

In order to address the question of whether map post-processing can attenuate map anisotropy, we computed the 3D-FSC (Zi Tan *et al.*, 2017) of post-processed maps using DeepEMhancer (Sanchez-Garcia *et al.*, 2021), LocScale2 (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022), LocSpiral (Kaur *et*

al., 2021), and LocalDeblur (Ramírez-Aportela *et al.*, 2020) against the maps simulated from the corresponding atomic models at the FSC global resolution for three different datasets.

The first dataset is composed of three different versions of the β-galactosidase complex (EMPIAR id 10061, (Bartesaghi *et al.*, 2015)), a standard reconstruction at 3.15 Å and two reconstructions in which we artificially introduced preferred orientations by removing 95% of the particles assigned to tilt angles above 80° or 70° (side views). Figure 1 displays the standard deviation of the per-cone FSC computed on the different postprocessed maps. As expected, the standard deviation for the experimental map (Raw, 0% removed) is small for all frequencies in the reconstructed and post-processed maps, which indicates that the resolution is mostly isotropic. Equally expected, the standard deviation of the per-cone FSC for the reconstructed map increases when we severely reduced the number of particles with high tilts, especially in the high-resolution range. For example, at the FSC global resolution frequency, the standard deviation increases from <0.03 for the Raw map to values above 0.05 when 95% of the particles with tilt angles >80° are removed and above to 0.09 when the particles removed have a till angle >70°.

From direct observation of the standard deviation for the different post-processed maps for the >80° and the >70° cases, we can conclude that all post-processing algorithms are able to reduce the standard deviation in the high-resolution region, thus reducing the map anisotropy. While the effect of the LocalDeblur algorithm is quite small or even negligible in the high-resolution frequencies, the other post-processing methods are able to reduce the standard deviation of the per-cone resolution to levels comparable to the Raw reconstruction.

The second example is the integrin αvβ8 in complex with L-TGF-β (EMD-20794, (Campbell *et al.*, 2020)). This map is a 3.3 Å reconstruction with a poorly defined region in which anisotropy is noticeable. As it can been observed in Figure 2, the per-cone standard deviation of the experimental map is quite high at the frequencies close to the global resolution, with much severe levels than in the previous experiment. Despite that, and in the same manner as in the previous example, all post-processing methods except LocalDeblur are able to reduce the anisotropy of the map as measured by the reduction in per-cone 3D-FSC standard variation at high frequencies, with DeepEMhancer and LocSpiral exhibiting the strongest effect and thus, offering the best quality results, as it can be visually confirmed in Figure 3.

The last example is the influenza hemagglutinin trimer reconstructed by Zi Tan et al. in their seminal work about how to address the preferential orientation problem (Zi Tan et al., 2017). We used both the map reconstructed from untiltled micrographs (EMPIAR ID 10096), and the map reconstructed with micrographs tilted 40° (EMPIAR ID 10097). While the former is an exemplar of a specimen severely affected by the preferential orientation problem, the latter is a corrected version in which although still present, the map anisotropy was massively attenuated.

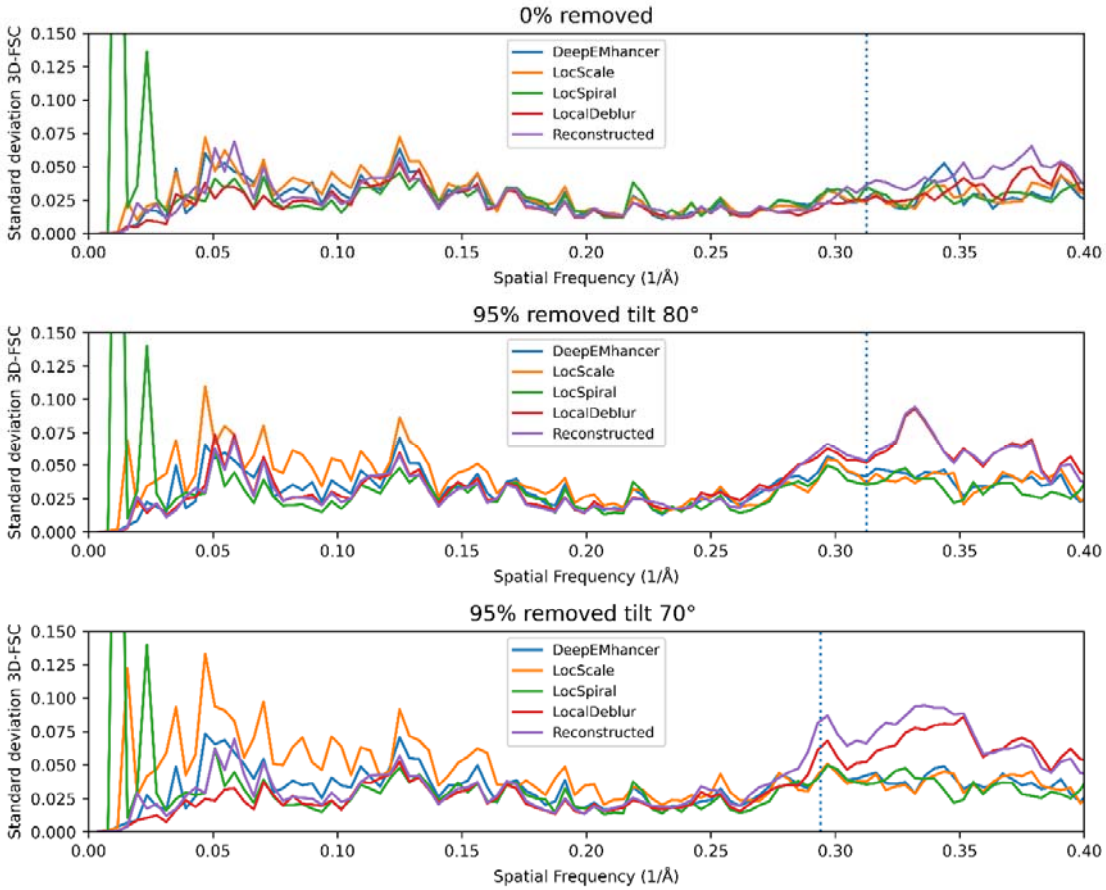

Figure 1 Per-cone FSC standard deviation calculated with 3D-FSC for three different versions of the β-galactosidase complex: reconstructed with all particles (Top) and reconstructed with 95% of the particles with tilt angle >80° (Middle) and >70° (Bottom) removed. The purple line represents the standard deviation for the reconstructed map whereas the blue, orange, green, and red lines represent the standard deviation for the maps obtained when DeepEMhancer (Sanchez-Garcia *et al.*, 2021), LocScale2 (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022), LocSpiral (Kaur *et al.*, 2021), and LocalDeblur (Ramírez-Aportela *et al.*, 2020) are employed on the reconstructed map. The FSC global resolution is displayed as a dotted vertical line.

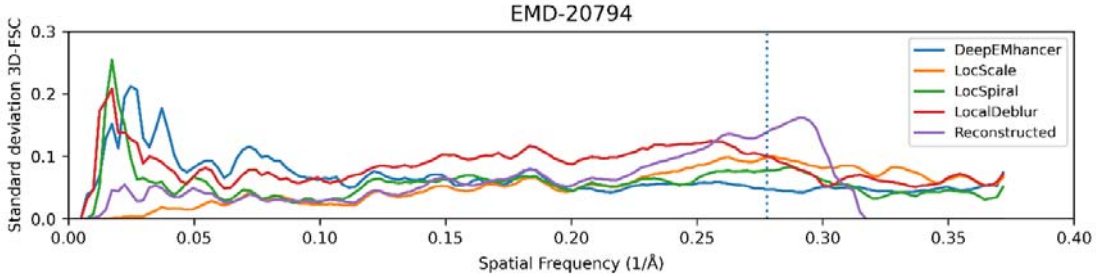

Figure 2 Per-cone FSC standard deviation calculated with 3D-FSC for the EMD-20794 raw map (purple) and maps post-processed with DeepEMhancer (blue, (Sanchez-Garcia *et al.*, 2021)), LocScale2 (orange, (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022)), LocSpiral (green, (Kaur *et al.*, 2021)), and LocalDeblur (red, (Ramírez-Aportela *et al.*, 2020)). The FSC global resolution of the reconstructed is displayed as a dotted vertical line.

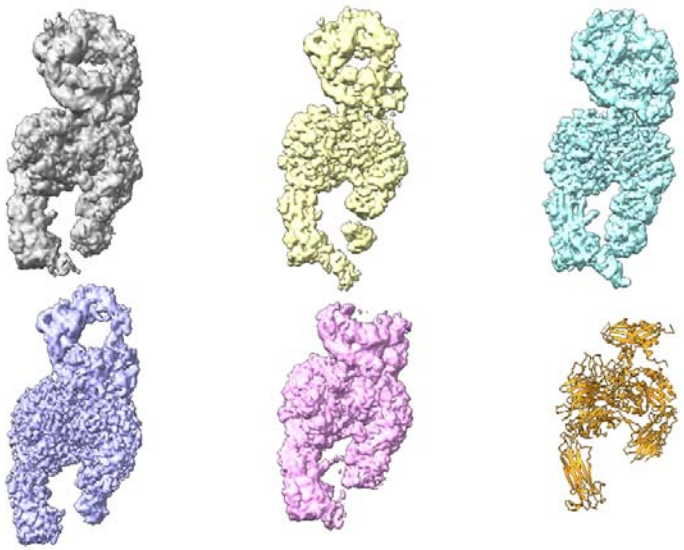

Figure 3 Reconstructed map (grey), associated atomic model (orange), and the results of applying DeepEMhancer (yellow, (Sanchez-Garcia *et al.*, 2021)), LocSpiral (cyan, (Kaur *et al.*, 2021)), LocalDeblur (purple, (Ramírez-Aportela *et al.*, 2020)), and LocScale2 (pink, (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022)) to the reconstructed map.

When we repeated the same experiment than in the other examples with the untilted hemagglutinin (see Figure 4), the results showed that none of the post-processing methods was able to reduce the standard deviation of the directional resolution. Indeed, some of them made the situation even worse (see Supplementary Figure S2), which is in line with what Sorzano et al. observed (Sorzano *et al.*, 2021). However, when the post-processing methods were employed on the map reconstructed on the titled micrographs, in which the standard deviation decreased by a factor 2 compared to the untilted, we were able to observe again a reduction in standard deviation, thus reducing the map anisotropy. The improvement is visually noticeable, as it is shown in Figure 5.

From these three experiments, we can confirm the capability of the map post-processing methods to correct moderate amounts of map anisotropy. It can also be appreciated that the anisotropy attenuation effect is also dependant on the amount of anisotropy of the original map. Thus, for the highly anisotropic untilted hemagglutinin map, no reduction in anisotropy was measured, whereas for the

tilted hemagglutinin and the EMD-20974 maps, the standard deviation of the directional resolution decreased from values around 0.15 to 0.07 at frequencies close to the global resolution while for the β-galactosidase with artificial preferred orientations, a less anisotropic map, the values were reduced from 0.10 to 0.03. Finally, for these three experiments, DeepEMhancer was the technique that exhibited better attenuation capabilities.

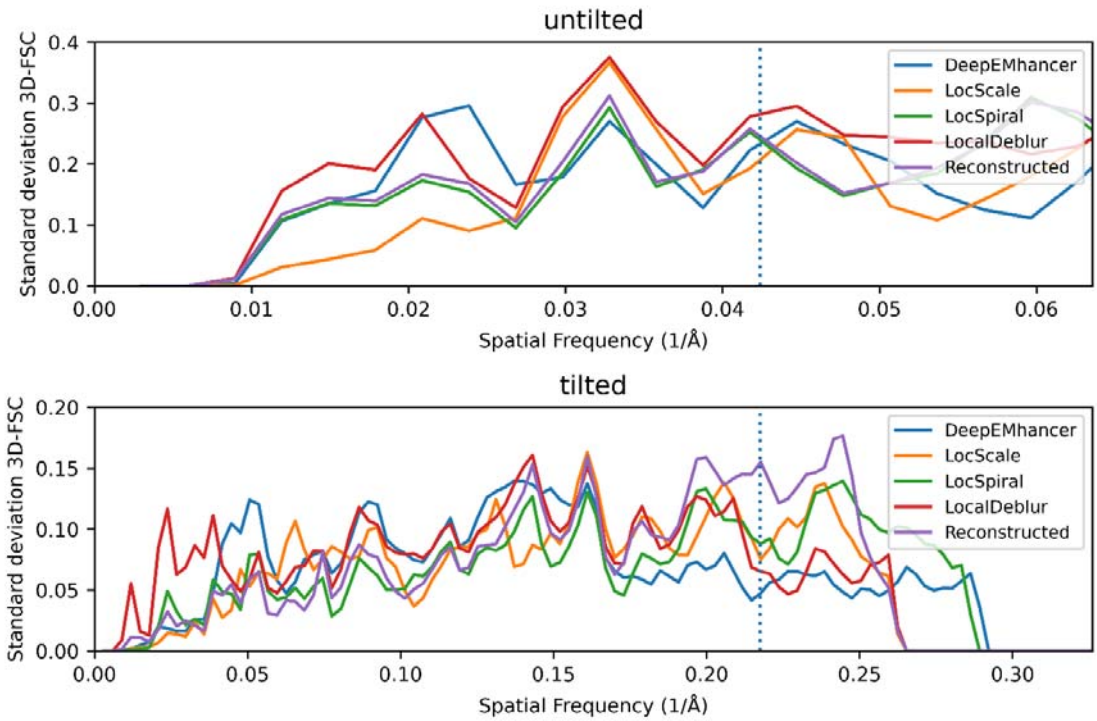

Figure 4 Per-cone FSC standard deviation calculated with 3D-FSC for two different versions of the influenza hemagglutinin trimer: reconstructed with untilted micrographs (severe anisotropy problems), and reconstructed with micrographs tilted 40°. The purple line represents the standard deviation for the reconstructed map whereas the blue, orange, green, and red lines represent the standard deviation for the maps obtained when DeepEMhancer (Sanchez-Garcia *et al.*, 2021), LocScale2 (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022), LocSpiral (Kaur *et al.*, 2021), and LocalDeblur (Ramírez-Aportela *et al.*, 2020) are employed on the reconstructed map. The FSC global resolution against the atomic model of the reconstructed map is displayed as a dotted vertical line.

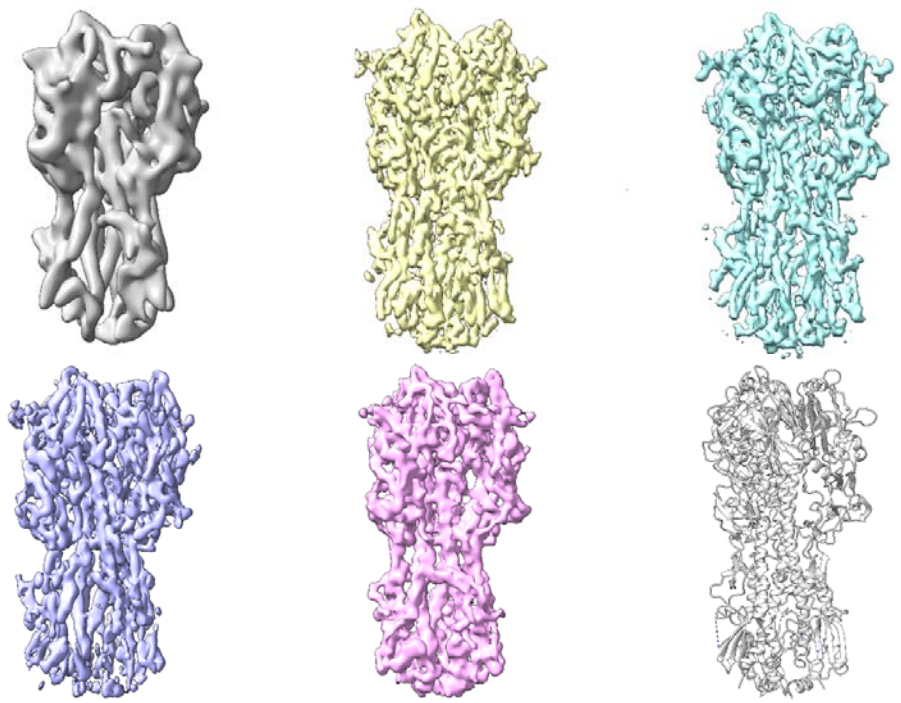

Figure 5 Influenza hemagglutinin trimer map reconstructed from tilted micrographs (grey), and the results of applying DeepEMhancer (yellow, (Sanchez-Garcia *et al.*, 2021)), LocSpiral (cyan, (Kaur *et al.*, 2021)), LocalDeblur (purple, (Ramírez-Aportela *et al.*, 2020)), and LocScale2 (pink, (Jakobi *et al.*, 2017; Bharadwaj & Jakobi, 2022)) to the reconstructed map. The atomic model used for evaluation is depicted in pale grey.

3.2. Common map anisotropy metrics are affected by the shape of the specimen

All the map anisotropy estimation methods described in the introduction are based on analysing the directional resolution from a given 3D reconstruction (Vilas *et al.*, 2020; Zi Tan *et al.*, 2017) or a particle orientations distribution (Naydenova & Russo, 2017). The obtained directional resolution depends obviously on the particle orientation distribution. However, this magnitude also depends on the actual geometry of the sample to reconstruct. Note that the spectral signal to noise ratio (SSNR) at different directions may depend on the actual shape of the macromolecule. Thus, based on the relation between the FSC and SSNR (Rosenthal & Henderson, 2003), we can conclude that the 3D-FSC (or the efficiency) may also depend on the actual shape of the macromolecule. As example, consider a narrow and long filament along the z-axis. This structure will provide high Fourier amplitudes mostly localized close to the x-y plane and to the z axis. Consequently, in the presence of noise, the SSNR will show an anisotropic shape related with the macromolecular geometry. For the sake of clarity, we have included in Figure 6 an example showing this case. In the first figure row, we show slices at z=0, y=0 and x=0 of a filament aligned with the z axis. In the second row, we show corresponding slices of the Fourier amplitudes of this map. Finally, in the third row, we show respective Fourier amplitude

slices but when the filament map was corrupted with additive Gaussian noise. As can be seen from this figure, the SSNR at different directions (and equivalently the 3D-FSC or the efficiency) may depend on the actual shape of the macromolecule.

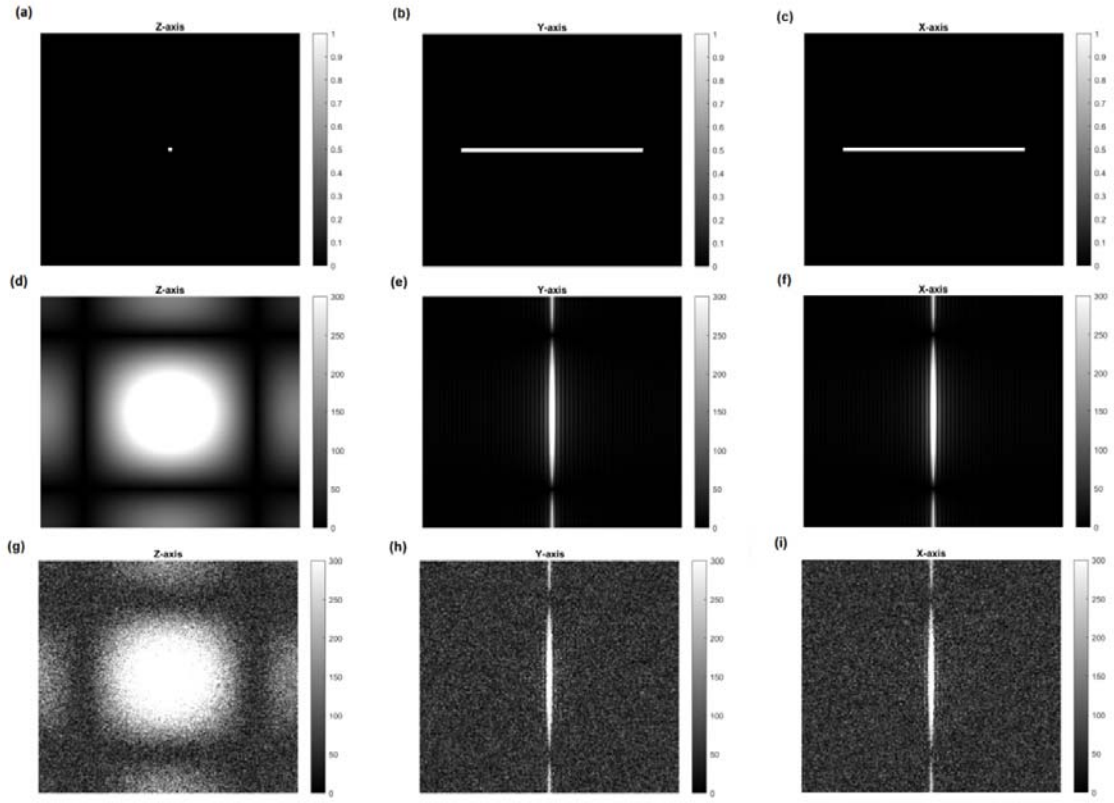

Figure 6 Filament map of size 201x201x201 pixels aligned with the z axis and its FFT. The first row displays slices of the filament map at a) z=0, b) y=0 and c) x=0. The second row displays the corresponding FFT amplitude slices of this map for d) z=0, e) y=0 and f) x=0. The last row shows the FFT amplitude slices for this map after adding gaussian noise with a standard deviation of 0.025 for g) z=0, h) y=0 and i) x=0. Fourier transform images are displayed using the FFTW convention.

We also carried out tests in which we generated a simulated electron density map from the atomic model of the influenza hemagglutinin (HA) trimer and we corrupted it by adding isotropic gaussian noise. Consequently, the map is perfectly isotropic and an ideal metric should not report any level of anisotropy. In Figure 7, we show the results obtained by the 3D-FSC (Zi Tan *et al.*, 2017) a), the efficiency (Naydenova & Russo, 2017) b), MonoDir (Vilas *et al.*, 2020) c-e) and our proposed approach for the simulated map. The computed sphericity for the FSC-3D method was 0.67 and 0.85 at FSC thresholds 0.14 and 0.5 respectively. The reported efficiency calculated with the second technique was 0.81 with mean and standard deviation PSF resolution of 3.40 Å and 0.32 Å

respectively. In Figure 7 c-e), we show the lowest, highest and half interquartile resolutions at some slices as computed by MonoDir, which shown different values depending on the plane.

As can be concluded from these results, all methods evaluated a perfectly isotropic simulated map as anisotropic. This limitation, far from being exclusively theoretical, may have an important consequence as anisotropy estimation across different specimens are not comparable. Note however, that these metrics are perfectly valid to analyse the anisotropy of different reconstructions of the same sample since they share the same shape.

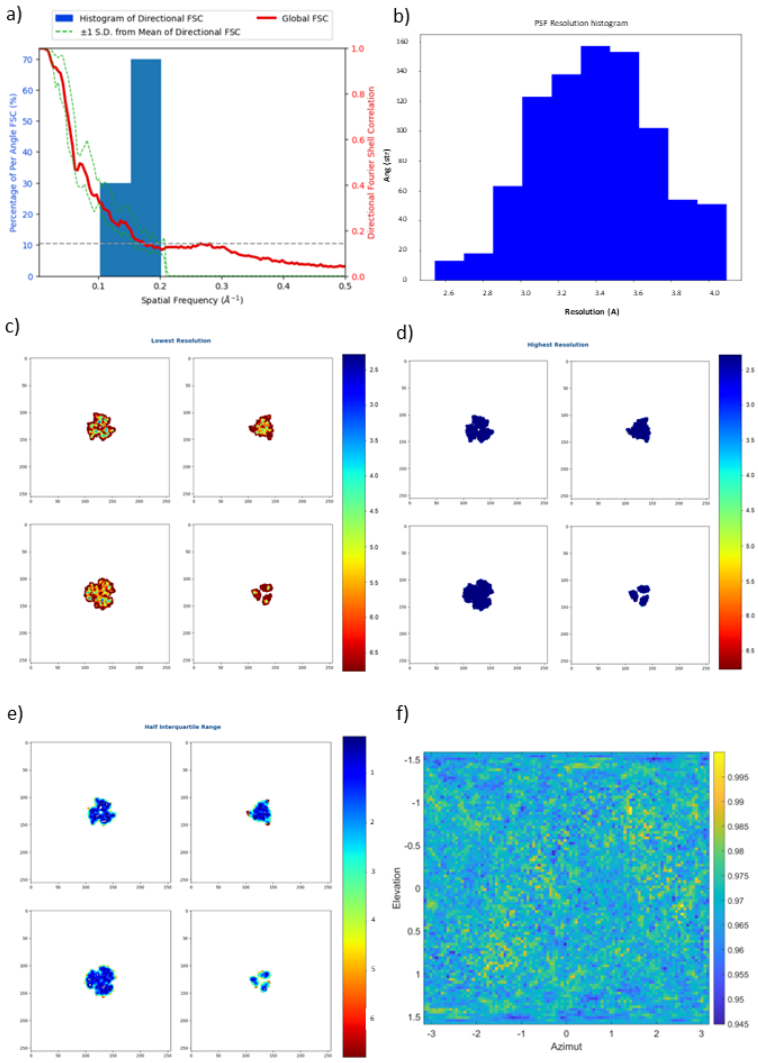

Figure 7 Results obtained for the simulation of the soluble portion of the small influenza hemagglutinin (HA) trimer by the 3D-FSC (Zi Tan *et al.*, 2017) a), the Efficiency (Naydenova & Russo, 2017) b), MonoDir (Vilas *et al.*, 2020) c-e) and the approach proposed in this manuscript f). All the studied methods except our proposed approach consider that the resolution of the ideal simulated volume is anisotropic.

3.3. New anisotropy method results

We have also processed the case shown in the previous section with our proposed method. Figure 7 f) shows the result obtained by our approach on the isotropic simulated map, where the map is evaluated as isotropic because the noise power does not change significantly along different directions.

We have also evaluated our approach with the experimental maps of the influenza hemagglutinin (HA) trimer. Figure 8 a) and b) show the results provided by our method for the reconstruction using untilted micrographs and the reconstruction using micrographs tilted at 40° respectively. As can be seen from these figures, for the untilted case, the 2D heatmap shows high values for elevations between [-1, 1] rad approximately (note that an elevation of 0 rad corresponds to side views), while for the 40° tilted case, this range is reduced to elevations between [-0.5, 0.5] rad. Thus, in the 40° case, particles with higher tilts were included in the reconstruction producing a better-quality map. Also note that the 2D map corresponding to the 40° case shows high values al elevations close to -1.5 and 1.5 rad. Note that these elevations correspond to top views (particles with tilts close to 0°), which do not introduce much information in the reconstruction.

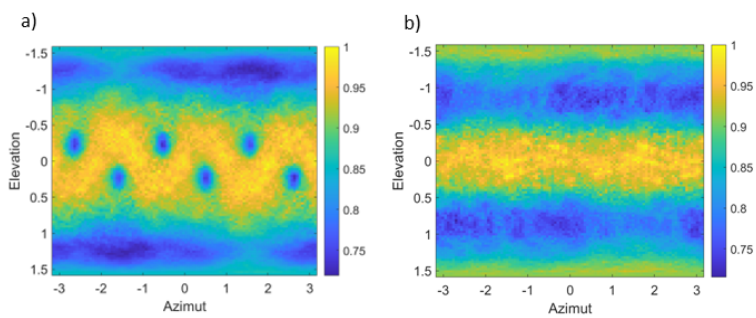

Figure 8 Results obtained by the proposed method for experimental maps obtained from untilted (a) and 40° tilted single particles of the influenza hemagglutinin (HA) trimer (b). The elevations and azimuths are given in radians

Finally, we also evaluated our newly proposed method on the β-galactosidase reconstructions used in the previous section in which anisotropy was artificially induced by orientations by removing 95% of the particles assigned to tilt angles above 80° or 70° (side views). The results obtained by the proposed approach are shown in Figure 9. As can be seen from this figure, the case where no particle was removed (0% removed) shows low values at two orientations corresponding to side views approximately (elevation close to 0 rad). Oppositely, the cases with tilt angles >80° or >70° removed show high values at elevation close to 0 rad, revealing the lack of side view (or close-to-side view) particles in these cases. Notice also that the case with particles removed at tilt angles >80° exhibit larger values for higher tilt angles that the case with particles removed at tilt angles >70°. Indeed, the dark stipe that can be appreciated in the latter case appears at values above 1.4 radians, corresponding

to 70º, which suggests that our method is able to detect missing views without any information about the particles or the angular assignments.

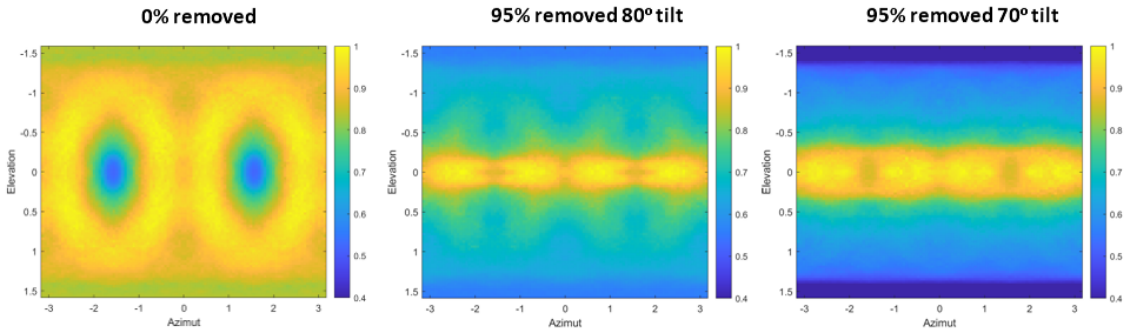

Figure 9 Results obtained by the proposed method for reconstructions of the β-galactosidase complex (EMPIAR-10061) consisting of a standard reconstruction of the B-gal at 3.15 Å and two reconstructions in which anisotropy was artificially induced by removing 95% of the particles assigned to tilt angles above 80° or 70° (side views).

4. Conclusions

The preferred orientations that some specimens tend to exhibit remain as a major challenge in cryo-EM. Preferred orientations are the main cause of the resolution anisotropy and due to their relevance, several directional resolution methods have been proposed as proxy metrics for the preferential orientation problem. Some of these metrics have been used to evaluate newly proposed experimental approaches aimed to minimise the problem of preferred orientations, but all them present an important limitation: they are sensitive to the shape of the specimen and consequently, inappropriate for large scale experiments comparing different samples. Our newly proposed method for anisotropy estimation has been designed to explicitly account for this shortcoming.

Although, this limitation may not be regarded as an important issue, if new approaches to tackle the problem of preferred orientations are to be proposed, robust metrics for its estimation are a requirement to assess their effectiveness across multi-specimen benchmarks. This is especially important if computational approaches are developed for this problem. While currently there are no computational approaches explicitly designed to tackle the problem of preferred orientations, in this manuscript we have shown that some of the most recently published map post-processing algorithms are capable of attenuating map anisotropy. According to our results, these algorithms are able to restore maps affected with moderate amounts of anisotropy with negligible additional computing cost, making them ideal to rescue some projects in which the preferred orientations problem is not severe enough to prevent the reconstruction process but it is strong enough to affect the quality of the map. To the best of our knowledge, this is the first time that algorithmic approaches have been shown to be

useful for the preferred orientations problem and we can envision that other new methods will follow this path in the future.

Acknowledgements We would like to thank our colleagues for interesting discussions and feedback. Ruben Sanchez Garcia is a post-doctoral fellow funded via an Astex Pharmaceuticals Sustaining Innovation Post-Doctoral Fellowship. The authors have no conflicts of interest to declare.

References

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S. & Subramaniam, S. (2015). *Science*. **348**, 1147–1151.

Bharadwaj, A. & Jakobi, A. J. (2022). *Faraday Discuss.*

Campbell, M. G., Cormier, A., Ito, S., Seed, R. I., Bondesson, A. J., Lou, J., Marks, J. D., Baron, J. L., Cheng, Y. & Nishimura, S. L. (2020). *Cell*. **180**, 490-501.e16.

Chua, E. Y. D., Mendez, J. H., Rapp, M., Ilca, S. L., Tan, Y. Z., Maruthi, K., Kuang, H., Zimanyi, C. M., Cheng, A., Eng, E. T., Noble, A. J., Potter, C. S. & Carragher, B. (2022). *Https://Doi.Org/10.1146/Annurev-Biochem-032620-110705*. **91**, 1–32.

D'Imprima, E. & Kühlbrandt, W. (2021). *Q Rev Biophys*. **54**, e4.

Glaeser, R. M. & Han, B.-G. (2017). *Biophys Rep*. **3**, 1–7.

Han, B. G., Walton, R. W., Song, A., Hwu, P., Stubbs, M. T., Yannone, S. M., Arbeláez, P., Dong, M. & Glaeser, R. M. (2012). *J Struct Biol*. **180**, 249–253.

Jakobi, A. J., Wilmanns, M. & Sachse, C. (2017). *Elife*. **6**,.

Kaur, S., Gomez-Blanco, J., Khalifa, A. A. Z., Adinarayanan, S., Sanchez-Garcia, R., Wrapp, D., McLellan, J. S., Bui, K. H. & Vargas, J. (2021). *Nat Commun*. **12**, 1240.

Naydenova, K. & Russo, C. J. (2017). *Nature Communications 2017 8:1*. **8**, 1–5.

Noble, A. J., Dandey, V. P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y. Z., Zhang, Z., Kim, L. Y., Scapin, G., Rapp, M., Eng, E. T., Rice, W. J., Cheng, A., Negro, C. J., Shapiro, L., Kwong, P. D., Jeruzalmi, D., Georges, A. des, Potter, C. S. & Carragher, B. (2018). *Elife*. **7**,.

Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S. & Carragher, B. (2018). *Nature Methods 2018 15:10*. **15**, 793–795.

Ramírez-Aportela, E., Vilas, J. L., Glukhova, A., Melero, R., Conesa, P., Martínez, M., Maluenda, D., Mota, J., Jiménez, A., Vargas, J., Marabini, R., Sexton, P. M., Carazo, J. M. & Sorzano, C. O. S. (2020). *Bioinformatics*. **36**, 765–772.

Rosenthal, P. B. & Henderson, R. (2003). *J Mol Biol*. **333**, 721–745.

Russo, C. J. & Passmore, L. A. (2014). *Nature Methods 2014 11:6*. **11**, 649–652.

Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S. & Vargas, J. (2021). *Commun Biol*. **4**, 1–8.

Sorzano, C. O. S., Semchonok, D., Lin, S. C., Lo, Y. C., Vilas, J. L., Jiménez-Moreno, A., Gragera, M., Vacca, S., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Melero, R., Cuervo, A., Conesa, J. J., Conesa, P., Losana, P., Caño, L. del, de la Morena, J. J., Fonseca, Y. C., Sánchez-García, R., Strelak, D., Fernández-Giménez, E., de Isidro, F., Herreros, D., Kastritis, P. L., Marabini, R., Bruce, B. D. & Carazo, J. M. (2021). *J Struct Biol*. **213**, 107695.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. (2007). *J Struct Biol*. **157**, 38–46.

Vilas, J. L., Gómez-Blanco, J., Conesa, P., Melero, R., Miguel de la Rosa-Trevín, J., Otón, J., Cuenca, J., Marabini, R., Carazo, J. M., Vargas, J. & Sorzano, C. O. S. (2018). *Structure*. **26**, 337-344.e4.

Vilas, J. L., Tagare, H. D., Vargas, J., Carazo, J. M. & Sorzano, C. O. S. (2020). *Nature Communications 2020 11:1*. **11**, 1–7.

Vilas, J. & Tagare, H. D. (2022). 0–14.

Wang, F., Liu, Y., Yu, Z., Li, S., Feng, S., Cheng, Y. & Agard, D. A. (2020). *Proc Natl Acad Sci U S A*. **117**, 24269–24273.

Zi Tan, Y., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B. & Lyumkis, D. (2017). *Nature Methods 2017 14:8*. **14**, 793–796.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E. & Scheres, S. H. W. (2018). *Elife*. **7**,.

Supporting information

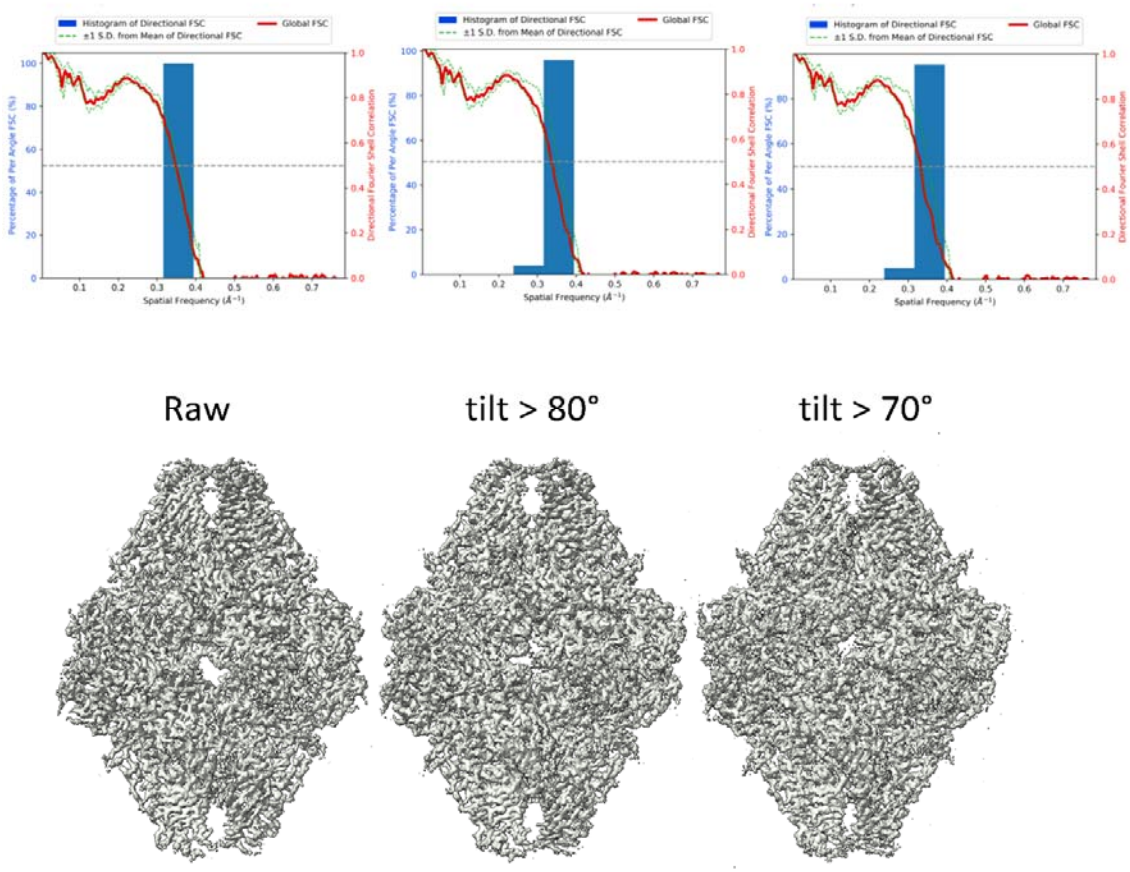

Figure S1 Bottom, cryo-EM maps for the β -galactosidase reconstructed from the EMPIAR-10061 dataset, using particles from all the projection directions (Left), and removing 95% of the particles with tilt angle >80° (Middle) and >70° (Right). Top, 3D-FSC displaying the global FSC resolution (red) and standard deviation of the directional resolution over 20-degree cones (green) as well as the histogram of the per angle FSC for the raw β-galactosidase map (Left), and the maps reconstructed with 95% of the particles with tilt angle >80° (Middle) and >70° (Right) removed.

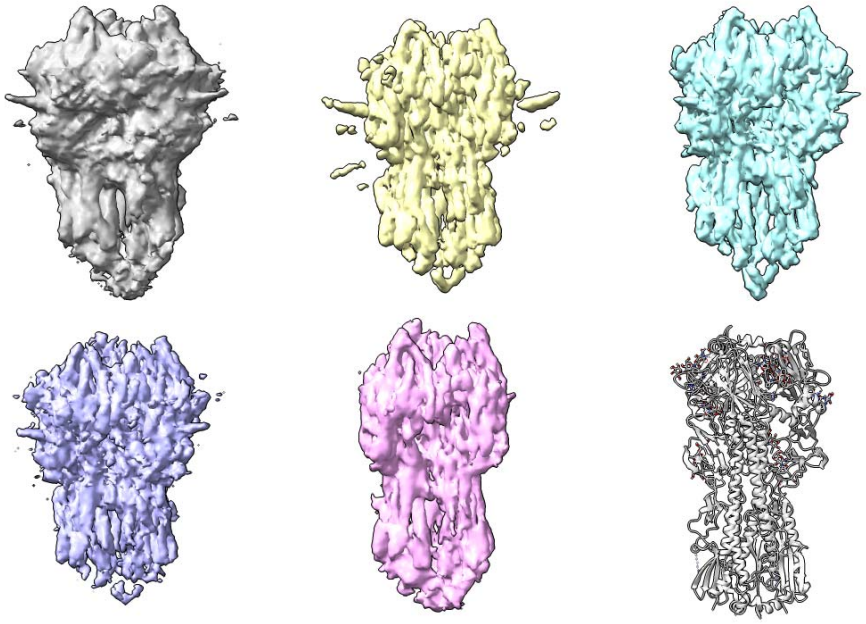

Figure S2 Hemagglutinin trimer map reconstructed from untilted micrographs (grey), and the results of applying DeepEMhancer (yellow, (Sanchez-Garcia *et al.*, 2021)), LocSpiral (cyan, (Kaur *et al.*, 2021)), LocalDeblur (purple, (Ramírez-Aportela *et al.*, 2020)), and LocScale2 (pink, (Jakobi *et al.*, 2017; Bharadwaj and Jakobi, 2022)) to the reconstructed map. The atomic model used for evaluation is depicted in pale grey.