Denoising of high-resolution single-particle electron-microscopy density maps by their approximation using three-dimensional Gaussian functions

S. Jonić, J. Vargas, R. Melero, J. Gómez-Blanco, J.M. Carazo, C.O.S. Sorzano

Denoising of high-resolution single-particle electron-microscopy density maps by their approximation using three-dimensional Gaussian functions

S. Jonić[1,*], J. Vargas[2], R. Melero[2], J. Gómez-Blanco[2], J.M. Carazo[2], and C. O. S. Sorzano[2]

[1] IMPMC, Sorbonne Universités - CNRS UMR 7590, UPMC Univ Paris 6, MNHN, IRD UMR 206, 75005 Paris, France.

[2]Biocomputing Unit, Centro Nacional de Biotecnología – CSIC, Campus de Cantoblanco, Darwin 3, 28049 Madrid, Spain.

* Corresponding author (Slavica Jonic, IMPMC-UMR 7590, Université Pierre & Marie Curie, Case courrier 115, 4 Place Jussieu, 75005 Paris, France, Phone : +33 1 44 27 72 05, Fax : +33 1 44 27 37 85, E-mail : Slavica.Jonic@impmc.upmc.fr).

Running title: Method for denoising of EM maps

Keywords: Electron microscopy density map, single particle analysis, denoising, macromolecular complexes, structure, Gaussian functions.

Abstract:

Cryo-electron microscopy (cryo-EM) of frozen-hydrated preparations of isolated macromolecular complexes is the method of choice to obtain the structure of complexes that cannot be easily studied by other experimental methods due to their flexibility or large size. An increasing number of macromolecular structures are currently being obtained at subnanometer resolution but the interpretation of structural details in such EM-derived maps is often difficult because of noise at these high-frequency signal components that reduces their contrast. In this paper, we show that the method for EM density-map approximation using Gaussian functions can be used for denoising of single-particle EM maps of high (typically subnanometer) resolution. We show its denoising performance using simulated and experimental EM density maps of several complexes.

INTRODUCTION

Cryo-electron microscopy (cryo-EM) of frozen-hydrated preparations of isolated macromolecular complexes is the method of choice to obtain the structure of complexes that cannot be easily studied by other experimental methods, such as X-ray crystallography (e.g., complexes with a significant degree of flexibility) or nuclear magnetic resonance (e.g., complexes of large size) (Frank, 2006). Recent technological advances, such as the latest generation of electron microscopes, direct electron detectors, software for automated collection of EM images and the availability of increasing computing power, combined with recent advances in image analysis algorithms, have eased the way to subnanometer-resolution structures for a wide range of macromolecular complexes (viruses, ribosomes, membrane proteins) (Allegretti et al., 2014; Amunts et al., 2014; Fischer et al., 2015; Gutsche et al., 2015; Khatter et al., 2015; Liao et al., 2013; Lu et al., 2014; Vinothkumar et al., 2014; Yu et

al., 2008; Zhang et al., 2008). An increasing number of macromolecular structures are being obtained at resolutions better than 4.5 Å (for a review, see (Cheng, 2015)). However, the interpretation of details of EM density maps is often difficult as noise at these high-frequency signal components reduces the contrast that is necessary for their identification. Thus, noise in EM-derived maps is usually reduced by a low-pass filtering. This is sometimes done by setting Fourier coefficients to zero beyond the resolution of the EM density map, which may induce Gibbs oscillations causing artificial features, but adjusting the shape of the low-pass filter to the shape of the Fourier Shell Correlation (FSC) curve has also been proposed (Penczek, 2010). Several other methods have been proposed for EM map denoising, and the majority was conceived in the context of denoising electron tomography reconstructions that usually contain low-resolution features whose interpretation is additionally hindered by strong experimental noise (Bilbao-Castro et al., 2010; Fernandez and Li, 2003; Frangakis and Hegerl, 2001; Jiang et al., 2003; van der Heide et al., 2007; Wei and Yin, 2010). In general, methods developed for one type of data may be used with other data types but their results may be suboptimal in such cases. The reason is that each method is parametrized (a set of parameters is defined and their default or recommended values are chosen) so as to optimally deal with data analysis difficulties linked to the particular experimental technique.

In this paper, we propose a denoising approach that was specifically conceived to deal with single-particle EM maps of high (typically subnanometer) resolution. The key concept is to derive an alternative sparse density-map representation, so that coefficients in this new space have a higher signal-to-noise ratio. To achieve it, we model the map by a linear combination of the same type of “atoms”, more precisely, three-dimensional (3D) Gaussian functions. In the function approximation terminology, these “atoms” are referred to as basis functions. In a simplified manner, we assume that the reconstructed density map should have the same

appearance as the density map that would be derived from the actual structure at atomic resolution. We represent the density map so that it appears similar to the atomic-resolution structure, using no reference atomic model during the approximation and using an optimized number of basis functions (Gaussian functions) that is usually smaller than the number of voxels. Thus, the proposed method is different from the method of convolving the density-map with a Gaussian function, which is known as Gaussian smoothing or blurring of the EM map.

The process of simplifying the density map description is usually referred to as coarse-graining of density maps and the resulting coarse-grain models of EM maps are also referred to as pseudoatomic or bead models. While simplified models of EM density maps have been used for many different purposes such as studying the topology of complexes, analyzing conformational changes, studying hydrodynamic properties of complexes, aligning structures at different resolutions, or density-map visualization improvement (Birmanns and Wriggers, 2007; Chacon et al., 2003; Garcia de la Torre et al., 2001; Jimenez-Lozano et al., 2003; Jin et al., 2014; Kawabata, 2008; Nogales-Cadenas et al., 2013; Spiegel et al., 2015; Wriggers et al., 1998), they are here used for a general task of denoising, as a preliminary step of many possible data analysis workflows. More precisely, the method for EM map coarse-graining based on the control of EM map approximation accuracy, proposed in (Jonic and Sorzano, 2016), is here shown to be useful for EM map denoising. To reach a given target accuracy of EM map approximation (target approximation error), that method adjusts the number, the position and the amplitude of grains represented by 3D Gaussian functions of a given standard deviation (grains are sometimes also referred to as pseudoatoms or beads). The method may not reach the target approximation error when using an inadequately large Gaussian-function standard deviation, but it allows overcoming this situation by suggesting the user to reduce the

Gaussian-function standard deviation or increase the target approximation error (Jonic and Sorzano, 2016). In this paper, we show that one may intentionally specify a very small, unattainable target approximation error and keep the resulting approximation, which can be used to remove noise from the original EM density map. In particular, we show how to choose the Gaussian-function standard deviation and the target approximation error to allow denoising. We use simulated and experimental density maps of several complexes at different resolutions to show the performance of the denoising method.

METHODS

Coarse-graining of EM density maps

We first give a brief description of the method used here for coarse-graining of EM density maps while its full description is given in (Jonic and Sorzano, 2016). This coarse-graining method uses a set of 3D Gaussian functions to approximate the original EM density map, $f(\mathbf{r})$ ($\mathbf{r} \in \mathbb{R}^3$). The approximated map is given by $\hat{f}_N(\mathbf{r}) = \sum_{i=1}^{N} \omega_i K_\sigma(\|\mathbf{r} - \mathbf{r}_i\|)$, where N is the number of Gaussian functions, $K_\sigma(r)$ is the Gaussian function with the standard deviation σ and maximum amplitude of 1, $\mathbf{r}_i$ is the position of the *i*-th Gaussian function, and $\omega_i > 0$ is the weight (amplitude) of the *i*-th Gaussian function. Given a density map, a value of σ, and a target approximation error, ε, the method adjusts N, $\mathbf{r}_i$, and ω_i so that the approximation error, e, satisfies $e = E\{|f(\mathbf{r}) - \hat{f}_N(\mathbf{r})|/\Delta f\} < \varepsilon$, where $E\{\cdot\}$ is the expectation operator, Δf is the effective range of values of f i.e. $\Delta f = F^{-1}(1-\alpha) - F^{-1}(\alpha)$, $F^{-1}(x)$ is the inverse of the cumulative distribution function of the values of f, and α is the statistical confidence on the effective range (typically, $\alpha = 0.025$). Note here that σ is expressed in voxels throughout this article.

The approximation error e is minimized iteratively until it reaches ε and, in optimization terminology, it is referred to as objective function. To make this optimization process more robust to local minima, Gaussian functions are added progressively, using a given initial number of Gaussian functions and a given speed of adding Gaussian functions. For the given current number of Gaussian functions, their amplitudes and positions are computed by gradient descent minimization of the approximation error. This coarse-graining method is available in the software package Xmipp (de la Rosa-Trevin et al., 2013; Scheres et al., 2008; Sorzano et al., 2004) and Scipion (de la Rosa Trevín et al, in preparation). The initial number of Gaussian functions of 300 and the speed of adding Gaussian functions of 30% usually produce good results and, thus, these values were used as default values in the available software as well as in all experiments in this article.

Use of coarse-graining of EM maps for their denoising

As said above, the coarse-grain representation is adjusted by minimizing the objective function until it reaches ε. However, in some cases, particularly for very small values of ε (around 1%), the objective-function minimization results in the density map whose error of approximation of the original EM density map is larger than ε, which indicates that the objective function cannot reach ε for a given value of σ. Such cases correspond to local minima of the objective function that, for a given value of ε, can be escaped by reducing σ because smaller values of σ produce larger numbers of Gaussian functions that, in their turn, can better approximate fine details including noise (Jonic and Sorzano, 2016). Keeping the current approximation (based on Gaussian functions of larger σ) instead of reducing σ results in smoother density maps in which noise is less represented. Given a small value of ε such as $\varepsilon = 1\%$, the question is then how to choose the value of σ to allow denoising. As we show below, we have found that optimal results can be obtained by adjusting σ (usually between 1

and 2) to suit the original (input) density map though $\sigma = 1.5$ could potentially be used as a default value that generally produces good results.

RESULTS

This section consists of three parts. In the first part, we show the performance of the method proposed here by fully evaluating the results of denoising of a synthetic and two EM density maps (**Experiments 1-3**) in terms of their FSC correlation with a non-filtered density map obtained from the corresponding atomic structure. In the synthetic case, this atomic structure is the exact ground-truth solution and, in the experimental case, it is considered to be close enough to the exact ground-truth solution. Density maps from atomic-resolution structures were computed using a method based on electronic-form atomic factors (Sorzano et al., 2015). The non-filtered density maps from atomic structures are referred to as reference density maps and were used only for the FSC computations. More importantly, neither the reference density map nor the corresponding atomic structure was used in the density map approximation process.

In the case of effective noise attenuation, the correlation (FSC) between the output and reference density maps is expected to be higher than the correlation between the input and reference density maps over a range of frequency shells. Thus, the effective noise attenuation is expected to result in an extension of the output-map FSC curve to higher frequencies than the input-map FSC curve, which was here evaluated by computing the difference between the frequencies at which the input and output density maps have the same correlation (FSC) with the reference density map. This frequency shift was computed for four FSC values selected to cover equidistantly the range [0.1, 0.7] in which the FSC could change significantly after processing the input density map. We would like to stress that the frequency shift was

computed for the multiple FSC values (more precisely, 0.7, 0.5, 0.3, and 0.1) to avoid reducing the entire FSC curve to a single value that is not carrying enough information in the denoising context that is considered in this paper. Single FSC values are commonly used in EM in the context of measuring the resolution of reconstructed EM density maps (the FSC curve is computed between two so-called half-maps independently computed from two subsets of the same set of images, which is known as gold-standard FSC, and the resolution is obtained as the frequency at which the FSC curve crosses a selected threshold value). However, a single FSC value does not carry enough information when using the FSC to evaluate, across frequencies, the amount of noise that is removed from a density map by its denoising, which is in the scope of this paper. Each of the different FSC values is related to a different value of the spectral signal-to-noise ratio (SSNR), but we are not particularly more interested in the frequencies at which SSNR=1 than in those at which SSNR>1 or SSNR<1, as we aim at increasing the SSNR in the entire frequency range. Indeed, it does not mean that the signal is useless at the frequencies where SSNR<1 but just that it is low, and the increase of the SSNR at these frequencies may thus be beneficial for the analysis of the entire signal.

Experiments 1-3 show results of the proposed method using input density maps that were not masked. Also, FSC curves in **Experiments 1-3** were computed with no masking of density maps. In the second part of this section (right after presenting **Experiments 1-3**), we compare, in FSC terms, the results obtained by the proposed method with those that could be obtained by masking the input density map as well as with those that could be obtained by several other available denoising methods. Also, we present a potential approach for tuning the value of σ, and we show the use of a measure of signal-to-noise ratio (SNR) as an alternative to FSC-based denoising evaluation that can be useful when high-resolution structures close enough to the exact ground-truth solution are unavailable. In the last part of this section, we show that

the proposed method can be used for denoising subtomogram average density maps of subnanometer resolution. Also, we demonstrate that the method does denoising inside the particle, by showing an example of denoising an empty virus particle (requiring an important denoising of the central region of the particle) and by comparing this result with the result of denoising a full virus particle.

Experiment 1: Synthetic EM density map of anthrax toxin protective antigen

In this experiment, the input density map for the denoising method was a synthetic EM density map computed from the atomic structure of anthrax toxin protective antigen with the PDB code 1ACC (Petosa et al., 1997). The atomic structure was first converted into a density map of size $220 \times 220 \times 220$ voxels with the voxel size of $1\ \text{Å} \times 1\ \text{Å} \times 1\ \text{Å}$ (reference density map). Then, 10,000 random projections of the reference density map were generated, and noise (signal-to-noise ratio of 0.1) and contrast transfer function (a defocus between 2.1 and 2.2 μm) were applied onto the projections to simulate EM images. Finally, the synthetic density map was reconstructed from the simulated EM images using their ground-truth orientations (known orientations of the generated random projections). Note here that the reference density map is the exact ground-truth solution in this experiment. Indeed, as the reference map was used to compute the input map for denoising, this reference map can be considered as a truly error- and noise-free solution.

The input density map was processed with the proposed method using σ values between 1 and 2 and $\varepsilon = 1\%$, which resulted in an output density map for each value of σ. FSC results show that the reference density map correlates better with the output density map than with the input density map for all tested values of σ. Indeed, at almost all frequencies, the FSC is larger for the output map than for the input map i.e., the output-map FSC curve extends more

to higher frequencies than the input-map FSC curve (**Fig. 1A and Fig**. **S1A**). In these terms, the best noise attenuation results were obtained for $\sigma = 1.6$ producing the frequency shift of 2.7 Å, 1.5 Å, 1.2 Å, and 1.5 Å for the FSC of 0.7, 0.5, 0.3, and 0.1, respectively (**Table 1, Fig. 1A, Fig**. **S1A**). The number of Gaussian functions obtained using $\sigma = 1.6$ was 6,280 and the corresponding error of the input density map approximation was 3.08% (**Table 2**). An arbitrary slice extracted from the reference map, the input map, and the output map for $\sigma = 1.6$ is shown in **Fig. 1B-D**, respectively. The isosurface representations of these density maps are provided in **Fig. S1B-D**.

It is tempting to think that the proposed denoising method introduces just a masking-type effect onto the EM density map, but this is not the case, which we show in a separate subsection.

Experiment 2: EM density map of 50S-ObgE

In this experiment, the EM density map of 50S ribosome subunit bound to ObgE (50S-ObgE) was obtained from the EMDB data bank and used as the input density map by the proposed denoising method (EMDB:EMD-2605; map size: $256 \times 256 \times 256$ voxels; voxel size: 1.5 Å $\times$ 1.5 Å $\times$ 1.5 Å (Feng et al., 2014)). The nominal resolution of the map is 5.5 Å, based on the gold-standard FSC with the 0.143 FSC threshold (Feng et al., 2014). The reference density map for evaluating the denoising results was the density map computed from the corresponding atomic model (PDB:4CSU) derived by fitting modelled and crystal structures of different parts of the complex into the EMD-2605 map and model refinement (Feng et al., 2014).

As in **Experiment 1**, the input density map was processed with the proposed method using σ values between 1 and 2 and $\varepsilon = 1\%$ and, for all tested values of σ, we could observe that the output-map FSC curve extends more to higher frequencies than the input-map FSC curve (**Fig. 1E and Fig**. **S2A**). According to the FSC computations, the best denoising results were obtained for $\sigma = 1.7$ producing the frequency shift of 1.5 Å, 0.3 Å, 0.7 Å, and 2.7 Å for the FSC of 0.7, 0.5, 0.3, and 0.1, respectively (**Table 1, Fig. 1E and Fig**. **S2A**). The number of Gaussian functions produced using $\sigma = 1.7$ was 50,356 and the corresponding error of the input density map approximation using these Gaussian functions was 2.29% (**Table 2**). An arbitrary slice of the reference, noisy input, and denoised 1.7-σ output maps is shown in **Fig. 1F-H**, respectively (for isosurface representations of these density maps, see **Fig. S2B-D**).

If comparing results obtained in **Experiments 1 and 2**, one should consider different sources of imperfection of the input and reference maps in these two experiments. While the input density map in **Experiment 1** is a 3D reconstruction from images at their ground-truth orientations and the map imperfections come from the added noise, the added contrast transfer function effects, and the use of a finite number of images, the input-map imperfections in **Experiment 2** also come from image alignment errors as well as other real-experiment-related imaging issues. Also, while the reference map is the exact ground-truth solution in **Experiment 1**, it is a solution close to the exact ground truth in **Experiment 2**.

Experiment 3: EM density map of beta-galactosidase

In this experiment, the EM density map of beta-galactosidase from the EMDB data bank was used as the input density map by the proposed denoising method (EMDB:EMD-5995; map size: $340 \times 340 \times 340$ voxels; voxel size: 0.64 Å $\times 0.64$ Å $\times 0.64$ Å (Bartesaghi et al., 2014)). The nominal resolution of the map is 3.2 Å, based on the gold-standard FSC with the 0.143

FSC threshold (Bartesaghi et al., 2014). The reference density map for evaluating the denoising results was the density map computed from the corresponding atomic model (PDB: 3J7H) that has been derived from a 1.7-Å resolution crystal structure of the entire complex, involving fitting of the crystal structure into the EMD-5995 map and model refinement (Bartesaghi et al., 2014).

As in the previous experiments, the input density map was processed with the proposed method using σ values between 1 and 2 and $\varepsilon = 1\%$, and we could again notice a shift to higher frequencies of the output-map FSC curve with respect to the input-map FSC curve for all tested values of σ (**Fig. 1I and Fig**. **S3A**). According to the FSC computations, the best denoising results were obtained for $\sigma = 1.4$ producing the frequency shift of 0.5 Å, 0.5 Å, 0.5 Å, and 1.1 Å for the FSC of 0.7, 0.5, 0.3, and 0.1, respectively (**Table 1**, **Fig. 1I, Fig**. **S3A**). The number of Gaussian functions produced using $\sigma = 1.4$ was 120,827 and the corresponding error of the input density map approximation using these Gaussian functions was 9.75% (**Table 2**). An arbitrary slice of the reference, noisy input, and denoised 1.4-σ output maps is shown in **Fig. 1J-L**, respectively (for isosurface representations of these density maps, see **Fig. S3B-D**).

Experiments 2 and 3 show that the proposed method can denoise experimental EM density maps as well as they show that the denoising results can be evaluated based on the FSC with an atomic reference structure if such a structure is available. Also, they show that the input and output density-map slices can be displayed to visually check the denoising results (i.e., the noise attenuation on the output density-map slices with respect to the input density-map slices). The exact ground-truth solution is not available in **Experiments 2 and 3**. Thus, the reference density map in these experiments comes from the best available atomic model for

the given EM map, and can be considered as noise-free but not as completely error-free. The reference density map is closer to the exact ground-truth solution in **Experiment 3** than in **Experiment 2** because of the higher resolution of the EM map yielding the atomic model for **Experiment 3** than for **Experiment 2**. Interestingly enough, **Experiment 2** shows that atomic models obtained from EM maps of nominal resolution in the range 5-6 Å (gold-standard FSC 0.143) are still reliable for evaluation of denoising results.

Comparison with input-map masking and with several available denoising methods

In this subsection, the best results obtained for denoising unmasked input maps in **Experiments 1-3** using the proposed method are compared with the denoising results obtained by masking the input density map and with the results obtained by other publicly available denoising methods (**Fig. 2, Figs. S4-9**). The output (density map) of the proposed method was compared with unmasked and masked output of each of these methods. Density-map masks adjusted to the shape of each particular complex were created using the standard masking procedure in single particle analysis, based on low-pass filtering, thresholding, and binarization (Frank, 2006)). The results are shown for iterative median filtering (van der Heide et al., 2007), bilateral filtering (Jiang et al., 2003), and nonlinear anisotropic diffusion filtering (Fernandez and Li, 2003; Frangakis and Hegerl, 2001) available in Bsoft software (programs *bmedian*, *bbif*, and *bnad*, respectively) (Heymann, 2001; Heymann et al., 2008) and, regarding nonlinear anisotropic diffusion filtering, also available in TOMOAND software (programs *tomoeed* and *tomoand*) (Fernandez and Li, 2003; Fernández et al., 2007). While *bnad* proposes edge-enhancing diffusion as a default diffusion mode and coherence-enhancing diffusion as an optional mode, *tomoeed* only allows edge-enhancing diffusion and *tomoand* combines the two diffusion modes. Regarding *bnad*, only edge-enhancing diffusion results are shown here as those are the best results obtained for this method with the data used

in this article. Furthermore, the methods were run using different values of their parameters and only the best results obtained with each method are shown here.

The FSC results in **Figure 2** show that the reference density map correlates better with the masked input density map than with the unmasked input density map. They also show that the reference density map correlates better with the output of the proposed method than with the masked input density map or with the output of other methods (**Figs. 2A-C, S4-6**), even when the output of these methods is masked (**Figs. 2D-F, S7-9**). This is particularly the case at higher frequencies i.e. a higher FSC of the output of the proposed method is observed where the FSC of the output of other methods (including input-map masking) is below 0.7 in the simulated case and below 0.3 in the experimental case (**Fig. 2**). One can also note that the FSC differences between the output of the proposed method and the output of other methods (including input-map masking) are larger in the simulated case than in the experimental case (**Fig. 2**), which is interesting because the FSC-based results evaluation should be the most correct in the simulated case, taking into account that the reference density map used for computing the FSC curves in that case is the exact ground-truth solution.

Identifying "optimal" σ based on input-map approximation error

In **Experiments 1-3**, we could identify the "optimal" value of σ using the FSC-based approach as we had at our disposal atomic-resolution models that were close enough to the corresponding exact ground-truth solutions. When this is not the case, the FSC-based approach cannot be used, but σ=1.5 could be used in such cases as it usually gives satisfactory results (**Figs. S1A, S2A, S3A**). We refer to this value (σ=1.5) as "default" value of σ. Alternatively, one could chose σ by analysing the actual dependence of the final input-map approximation error on the given value of σ. More precisely, we have found interesting to

compare the input-map approximation errors achieved when reducing σ and to identify the σ value for which the approximation error starts to decrease slowly or for which there is an unchanged amount of the approximation error decrease (a kind of saturation in the decrease). A decrease in the σ value produces a decrease in the speed of the global approximation-error convergence towards the target approximation error (**Table 2**). For instance, in the case of synthetic density map, the approximation error decreases by around 0.005 when σ decreases from 1.7 to 1.6, and by around 0.002 when σ decreases from 1.6 to 1.5 or from 1.5 to 1.4 (**Table 2**), which means a saturation in the decrease of the approximation error when reducing σ below the value that was identified as the optimal value by the FSC analysis (σ=1.6). Recall here that the recommended target approximation error of 1% was used in all three experiments and that the approximation error achieved for each σ corresponds to a local minimum of the objective function from which one can escape by reducing σ (Jonic and Sorzano, 2016). Indeed, the decrease in σ produces a larger number of Gaussian functions to better approximate fine details including noise (**Table 2**). Thus, stopping to decrease σ below the value identified as the optimal value by the FSC analysis suggests that this σ produces the approximation where noise is optimally removed. Here, we could show (**Table 2**) that this value of σ can be identified as the value for which the approximation error starts to decrease slowly or for which there is a saturation in the decrease.

Signal-to-noise ratio as an alternative to FSC-based denoising evaluation

In **Experiments 1-3**, we could evaluate the denoising results numerically using the FSC-based approach. When the FSC-based approach cannot be used, an alternative could be to measure the SNR. Here, we use the SNR definition already introduced in EM (Bilbao-Castro et al., 2010), namely $SNR=(I_s-I_b)/\sigma_b$, where I_s and I_b denote the average intensity in the meaningful signal region (structure of interest) and in the unwanted signal region (noise),

respectively, and σ_b is the standard deviation in the noise region. The noise region is defined as the region outside a mask suited to the shape of the complex such as the one used in **Experiments 1-3**. Although the average intensity in the noise region, I_b, is small and could be neglected in many cases, we here take it into account by considering a general case where this may not be true (e.g., subomogram average data). **Table 3** summarizes the SNRs before and after denoising by the proposed method and the other methods tested in **Experiments 1-3** as well as the ratio between the input and output SNRs. Though the information measured by the SNR and the FSC is not the same but related, one can notice that the proposed method achieves the best results over all tested methods according to both the FSC and the SNR (**Fig. 2** and **Table 3**). In the next two subsections, aiming at answering particular additional questions regarding the performance of the proposed method, we only report the SNR because the particular experiments described in these subsections were done using density maps for which atomic models are either unavailable or may not be sufficiently reliable for the purpose of denoising evaluation.

Denoising of density maps obtained by subtomogram averaging

The proposed method was specifically conceived to deal with single-particle EM density maps. It cannot be used for denoising 2D images and it is not efficient in denoising electron tomogram reconstructions that are large and very noisy. However, it can be used to denoise subnanometer-resolution subtomogram averages, as we show in this subsection. **Figure 3** shows the use of the proposed method (with σ=0.5 and $\varepsilon = 5\%$), *bmedian*, *bbif*, *bnad*, and *tomoeed* for denoising a density map of immature HIV-1 capsid of intact virus particles obtained by subtomogram averaging at 8.8 Å resolution (FSC 0.143; EMDB:EMD-2706; size: 140×140×140voxels, voxel size: 2.02 Å) (Schur et al., 2015). Note that the value of ε was set larger than 1% in this case to avoid noise overfitting that here refers to as an

important noise reproduction in the approximation by Gaussian functions. The SNR of the input density map (SNR=7.3) was increased to SNR=42.7 (output map) using the proposed method, while it was increased only to 8.0, 9.4, 9.3, 7.9, and 8.4 using *bmedian*, *bbif*, *bnad*, *tomoand*, and *tomoeed*, respectively. This means that the proposed method improved the SNR of the input density map 5.8 times, whereas the other methods improved it only 1.1-1.3 times.

Interior denoising

In this subsection, we show that the proposed method denoises the interior of a complex and not only the background. For this purpose, we show in **Figure 4** denoising of a density map of genogroup II genotype 10 norovirus virus-like particle obtained by single-particle analysis at 10 Å resolution (FSC 0.5; EMDB: EMD-5374; size: 250×250×250 voxels, voxel size: 2.40 Å) (Hansman et al., 2012). This density map shows an empty particle, meaning that there is no "signal" in the middle of the particle but only noise. **Figure 4** clearly shows that the proposed method denoises not only the background but also the interior of the virus. While the SNR of the input density map is 10.4, the SNR of the map obtained by the proposed method (using σ=1.5 and $\varepsilon = 15\%$) is 30.3, which makes an improvement of 2.9 times. These results can be compared with results of denoising of a full virus particle where the middle part of the particle is not occupied by noise but by "signal" (meaningful information) and should thus not be suppressed by denoising (**Fig. 5**). **Figure 5** shows denoising of a density map of human rhinovirus 2 135S full particle obtained by single-particle analysis at 8.8 Å resolution (FSC 0.5; EMDB: EMD-2109; size: 244×244×244voxels, voxel size: 1.89 Å) (Pickl-Herk et al., 2013). In this case, the increase in the SNR obtained by the proposed method (using σ=1.3 and $\varepsilon = 5\%$) is from 13.6 (input map) to 121.4 (output map), which makes an improvement of 8.9 times. In these two data cases, as in the case of subtomogram average data shown in the

previous subsection, the value of ε had to be set larger than in **Experiments 1-3** to avoid noise overfitting.

DISCUSSION AND CONCLUSION

In this paper, the method of representing an EM density map by a linear combination of 3D Gaussian functions, whose positions, amplitudes, and number are computed through an optimization (approximation) procedure (Jonic and Sorzano, 2016), was shown to be useful for denoising single-particle EM maps of high (typically subnanometer) resolution. This was shown by a successful denoising of one synthetic and several experimental input density maps without masking. The proposed method is a denoising method, meaning that it aims at attenuating noise while preserving structural features as they are. For instance, in lower resolution EM maps, the structural features are not fine and they will remain so after the use of the method. Depending on the density map size and the combination of σ (the standard deviation of Gaussian functions) and ε (the target approximation error of the density map using Gaussian functions), the method takes from a few minutes to a few hours to perform on a single processor of a laptop computer, but multithreading is also allowed.

Choice of σ and ε

In the context of denoising, ideal values of σ and ε are those that result in the complete noise removal and signal reproduction. However, such σ and ε values are difficult to determine, meaning that some noise is usually reproduced together with the signal. In the case of higher-resolution density maps (resolutions higher than 6 Å, gold-standard FSC 0.143), we have shown that σ=1.5 and $\varepsilon = 1\%$ produce satisfactory results in most cases, but σ can be additionally adjusted (for the given ε value) to achieve optimal results in terms of the FSC with an atomic model of the density map, if such a model is available (**Figs. S1A, S2A, S3A**).

We have also shown an approach to adjust σ that does not rely on the FSC computations but produces consistent results with the FSC-based ones. It compares the values of the input-map approximation error achieved when reducing σ to identify the σ value for which the approximation error starts to decrease slowly or the approximation error decrease starts to saturate. Note here that $\varepsilon = 1\%$ is a quite small value, and as such, it may not be achieved (for the given value of σ). In the case of higher-resolution density maps, a failure in reaching ε of 1% usually means a "failure" in noise overfitting. This was the case with the data in **Experiments 1-3**, where the target approximation error of 1% could not be reached and adding of Gaussian functions to the approximation ended as soon as the approximation error achieved the values of 3.08%, 2.29%, and 9.75%, respectively (for the given σ value of 1.6, 1.7, and 1.4, respectively). Target approximation errors smaller than 1% are often unattainable and the resulting approximations (for the achieved approximation error) suffer from noise overfitting. Thus, setting the value of ε to 1% is likely a good compromise for the majority of higher-resolution density maps. In the case of lower-resolution density maps (resolutions lower than 6 Å, gold-standard FSC 0.143), the target approximation error of 1% is usually unreachable and this choice usually produces approximations with noise overfitted. As we have shown, noise overfitting in such cases can be avoided by setting ε to larger values (often 5%, and in some cases, 15%) and by readjusting the value of σ around the suggested value of 1.5 (usually slightly, but a more important readjustment may be needed in some cases of noise such as subtomogram noise).

Evaluation of denoising results

In **Experiments 1-3**, denoising results were evaluated quantitatively based on the frequency shift between the FSC curves of the input and output density maps with respect to the reference density map, where the reference density map was derived from an atomic-

resolution structure of the complex that was available in these three cases (**Table 1**). Here, we decided to examine the frequency shifts of four FSC values, taken at equal distances (0.2) from FSC=0.1 to FSC=0.7, which allows a more complete numerical evaluation of the noise reduction across different frequency shells than it would be the case if the frequency shift of a single FSC value was used for this purpose. Though only the entire FSC curve carries the full information and the four selected FSC values are only one of many possible reduced representations of the entire FSC curve, the results presented here show that the four selected values can be considered as representative enough of the FSC curve shape changes due to denoising (**Figs. 1A**,**E,I**).

When atomic-resolution structures are unavailable or perhaps not enough reliable for the purpose of denoising evaluation (e.g., when atomic models are obtained from EM maps of resolution lower than 6-7 Å (gold-standard FSC 0.143)), numerical evaluation of denoising results can be done by measuring the signal-to-noise ratio, as shown in this article. Additionally, the denoising results can be inspected visually, by comparing the input and output density-map slices.

Comparison with masking and with other available methods

We have shown that the denoising results cannot be reproduced by a simple masking of the input density map, even using a mask suited to the shape of the complex. This suggests that the masks remove the background noise only, whereas the method proposed here removes noise additionally from non-background voxels (voxels occupied by the complex). The removal of noise from non-background voxels was illustrated by denoising two (empty and full) virus particles (**Figs. 4-5**). We have also shown that the proposed method provides better results than other available methods tested using the same data (**Fig. 2**). The majority of these

methods have been conceived for denoising raw electron tomography images or 3D reconstructions from these images (electron tomograms). The method proposed here has been conceived for single particle 3D applications and it is not efficient in tomogram denoising, but it can be used for denoising subtomogram averages (**Fig. 3**).

Avoiding local density overrepresentation or subrepresentation

The proposed denoising method approximates an input density map by a linear combination of Gaussian functions whose number is usually smaller than the number of voxels of the input density map (see **Table 2** for the numbers of Gaussian functions used for different complexes and different values of σ). Also, amplitudes and locations of Gaussian functions are adjusted so as to achieve a uniform representation of densities of the complex while avoiding local overrepresentations or subrepresentations (Jonic and Sorzano, 2016). In this respect, the proposed method is original and has advantages. It may be tempting to think that some non-linear transformations of the input density map such as raising input densities to the power of 2 could produce similar results as the method proposed here. However, such transformations produce a non-uniform representation of input densities (high intensities are strongly amplified and low intensities are in a way attenuated, which makes that some very bright voxels "appear" and some dark voxels "disappear"), which in turn makes that the reference map correlates less with the map transformed using this transformation than with the map processed using the proposed method (results not shown).

No use of a priori knowledge

Recently, a visualization improvement of EM density maps has been proposed using a method based on bead models (Spiegel et al., 2015). Though the method presented here has been conceived for a general task of denoising of EM maps, one may consider its use also for the

EM map visualisation improvement. The method proposed in (Spiegel et al., 2015) places the beads randomly in the density regions that are above a certain threshold. The coordinates of the beads are then used to place real atoms inside the EM map (randomly assigned atom types to the beads inside the particle boundaries defined by the density threshold such that the atomic composition is the same as the average atomic composition observed for proteins in the PDB database). Thus, the method proposed in (Spiegel et al., 2015) requires some *a priori* knowledge about protein structures and an estimate of the number of atoms inside the particle boundaries defined by a density threshold. This implies that it works best for resolutions better than 5 Å because, at lower resolutions, the boundaries of the protein are not well defined and a radial mass distribution cannot be compared with the average one obtained from PDB structures (Spiegel et al., 2015). In contrast to that method, the approach proposed here uses no *a priori* knowledge about proteins or from PDB, which allows analysing any type of macromolecular assemblies.

Basis functions

Gaussian functions are used for the purpose of approximating the input density map (in real space) and for the purpose of building the reference density map from the atomic structure (in Fourier space). However, the two tasks rely on completely different real-space basis functions i.e., Gaussian functions for approximating the input map (Jonic and Sorzano, 2016) and Low-pass filtered Electron Atomic Scattering Factors for building the reference map (Sorzano et al., 2015). Thus, while the "extension" of the FSC curve for the denoised map to higher frequencies than the FSC curve for the input map could be attributed to the denoising effect of real-space Gaussian functions used for the input map approximation, one cannot say that this FSC "extension" was facilitated by the use of Fourier-space Gaussian functions for building

the reference map as the corresponding real-space basis functions are not Gaussian functions (Sorzano et al., 2015).

In the future work, we plan to extend this method to use other basis functions and evaluate the performance of different basis functions for a detailed analysis of cryo EM maps.

SUPPORTING MATERIAL

Supporting Material includes nine supplementary figures (**Figs. S1-9**) with their legends.

AUTHOR CONTRIBUTIONS

COSS and SJ designed the method and the experiments, performed the experiments, and analyzed and discussed the results. JV, RM, JGB and JMC helped to analyze the results. SJ wrote the manuscript. All authors reviewed and discussed the manuscript.

ACKNOWLEDGMENTS

The work was partially funded by the CNRS (France) and the CSIC (Spain) [Projet International de Coopération Scientifique **-** PICS 2011]; the French National Research Agency ANR [ANR-11-BSV8-010-04]; the European Social Fund and the Ministerio de Educación y Ciencia ["Ramón y Cajal" fellowship to COSS]; the Spanish Ministry of Economy and Competitiveness [AIC–A–2011–0638 and BIO2013-44647-R]; and the Comunidad de Madrid [CAM S2010/BMD-2305]. We thank GENCI-CINES/IDRIS (France) for HPC resources [x2014072174, x2015072174, x2016072174 to S.J.].

REFERENCES

Allegretti, M., Mills, D.J., McMullan, G., Kuhlbrandt, W., Vonck, J., 2014. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. Elife 3, e01963.

Amunts, A., Brown, A., Bai, X.C., Llacer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H., Ramakrishnan, V., 2014. Structure of the yeast mitochondrial large ribosomal subunit. Science 343, 1485-1489.

Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A., Subramaniam, S., 2014. Structure of beta-galactosidase at 3.2-A resolution obtained by cryo-electron microscopy. Proc Natl Acad Sci U S A 111, 11709-11714.

Bilbao-Castro, J.R., Sorzano, C.O., Garcia, I., Fernandez, J.J., 2010. XMSF: Structure-preserving noise reduction and pre-segmentation in microscope tomography. Bioinformatics 26, 2786-2787.

Birmanns, S., Wriggers, W., 2007. Multi-resolution anchor-point registration of biomolecular assemblies and their components. J Struct Biol 157, 271-280.

Chacon, P., Tama, F., Wriggers, W., 2003. Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. J Mol Biol 326, 485-492.

Cheng, Y., 2015. Single-Particle Cryo-EM at Crystallographic Resolution. Cell 161, 450-457.

de la Rosa-Trevin, J.M., Oton, J., Marabini, R., Zaldivar, A., Vargas, J., Carazo, J.M., Sorzano, C.O., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J Struct Biol 184, 321-328.

Feng, B., Mandava, C.S., Guo, Q., Wang, J., Cao, W., Li, N., Zhang, Y., Zhang, Y., Wang, Z., Wu, J., Sanyal, S., Lei, J., Gao, N., 2014. Structural and functional insights into the mode of action of a universally conserved Obg GTPase. PLoS Biol 12, e1001866.

Fernandez, J.J., Li, S., 2003. An improved algorithm for anisotropic nonlinear diffusion for denoising cryo-tomograms. J Struct Biol 144, 152-161.

Fernández, J.J., Li, S., Lucic, V., 2007. Three-Dimensional Anisotropic Noise Reduction with Automated Parameter Tuning: Application to Electron Cryotomography, p. 60-69, Current Topics in Artificial Intelligence: 12th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2007, Salamanca, Spain, November 12-16, 2007. Selected Papers, Springer Berlin Heidelberg, Berlin, Heidelberg.

Fischer, N., Neumann, P., Konevega, A.L., Bock, L.V., Ficner, R., Rodnina, M.V., Stark, H., 2015. Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. Nature 520, 567-570.

Frangakis, A.S., Hegerl, R., 2001. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. J Struct Biol 135, 239-250.

Frank, J., 2006. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State Oxford University Press, New York.

Garcia de la Torre, J., Llorca, O., Carrascosa, J.L., Valpuesta, J.M., 2001. HYDROMIC: prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy images. Eur Biophys J 30, 457-462.

Gutsche, I., Desfosses, A., Effantin, G., Ling, W.L., Haupt, M., Ruigrok, R.W., Sachse, C., Schoehn, G., 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. Science 348, 704-707.

Hansman, G.S., Taylor, D.W., McLellan, J.S., Smith, T.J., Georgiev, I., Tame, J.R., Park, S.Y., Yamazaki, M., Gondaira, F., Miki, M., Katayama, K., Murata, K., Kwong, P.D., 2012. Structural basis for broad detection of genogroup II noroviruses by a monoclonal antibody that binds to a site occluded in the viral particle. J Virol 86, 3635-3646.

Heymann, J.B., 2001. Bsoft: image and molecular processing in electron microscopy. J Struct Biol 133, 156-169.

Heymann, J.B., Cardone, G., Winkler, D.C., Steven, A.C., 2008. Computational resources for cryo-electron tomography in Bsoft. J Struct Biol 161, 232-242.

Jiang, W., Baker, M.L., Wu, Q., Bajaj, C., Chiu, W., 2003. Applications of a bilateral denoising filter in biological electron microscopy. J Struct Biol 144, 114-122.

Jimenez-Lozano, N., Chagoyen, M., Cuenca-Alba, J., Carazo, J.M., 2003. FEMME database: topologic and geometric information of macromolecules. J Struct Biol 144, 104-113.

Jin, Q., Sorzano, C.O., de la Rosa-Trevin, J.M., Bilbao-Castro, J.R., Nunez-Ramirez, R., Llorca, O., Tama, F., Jonic, S., 2014. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. Structure 22, 496-506.

Jonic, S., Sorzano, C.O.S., 2016. Coarse-graining of volumes for modeling of structure and dynamics in electron microscopy: Algorithm to automatically control accuracy of approximation. IEEE Journal of Selected Topics in Signal Processing 10, 161 - 173.

Kawabata, T., 2008. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. Biophys J 95, 4643-4658.

Khatter, H., Myasnikov, A.G., Natchiar, S.K., Klaholz, B.P., 2015. Structure of the human 80S ribosome. Nature 520, 640-645.

Liao, M., Cao, E., Julius, D., Cheng, Y., 2013. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature 504, 107-112.

Lu, P., Bai, X.C., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H., Shi, Y., 2014. Three-dimensional structure of human gamma-secretase. Nature 512, 166-170.

Nogales-Cadenas, R., Jonic, S., Tama, F., Arteni, A.A., Tabas-Madrid, D., Vazquez, M., Pascual-Montano, A., Sorzano, C.O., 2013. 3DEM Loupe: Analysis of macromolecular dynamics using structures from electron microscopy. Nucleic Acids Res 41, W363-367.

Penczek, P.A., 2010. Image restoration in cryo-electron microscopy. Methods Enzymol 482, 35-72.

Petosa, C., Collier, R.J., Klimpel, K.R., Leppla, S.H., Liddington, R.C., 1997. Crystal structure of the anthrax toxin protective antigen. Nature 385, 833-838.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612.

Pickl-Herk, A., Luque, D., Vives-Adrian, L., Querol-Audi, J., Garriga, D., Trus, B.L., Verdaguer, N., Blaas, D., Caston, J.R., 2013. Uncoating of common cold virus is preceded by RNA switching as determined by X-ray and cryo-EM analyses of the subviral A-particle. Proc Natl Acad Sci U S A 110, 20063-20068.

Scheres, S.H., Nunez-Ramirez, R., Sorzano, C.O., Carazo, J.M., Marabini, R., 2008. Image processing for electron microscopy single-particle analysis using XMIPP. Nat Protoc 3, 977-990.

Schur, F.K., Hagen, W.J., Rumlova, M., Ruml, T., Muller, B., Krausslich, H.G., Briggs, J.A., 2015. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. Nature 517, 505-508.

Sorzano, C.O., Marabini, R., Velazquez-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. J Struct Biol 148, 194-204.

Sorzano, C.O.S., Vargas, J., Otón, J., Abrishami, V., de la Rosa-Trevín, J.M., del Riego, S., Fernández-Alderete, A., Martínez-Rey, C., Marabini, R., Carazo, J.M., 2015. Fast and accurate conversion of atomic models into electron density maps. AIMS Biophysics 2, 8-20.

Spiegel, M., Duraisamy, A.K., Schroder, G.F., 2015. Improving the Visualisation of Cryo-EM Density Reconstructions. J Struct Biol 191, 207-213.

van der Heide, P., Xu, X.P., Marsh, B.J., Hanein, D., Volkmann, N., 2007. Efficient automatic noise reduction of electron tomographic reconstructions based on iterative median filtering. J Struct Biol 158, 196-204.

Vinothkumar, K.R., Zhu, J., Hirst, J., 2014. Architecture of mammalian respiratory complex I. Nature 515, 80-84.

Wei, D.Y., Yin, C.C., 2010. An optimized locally adaptive non-local means denoising filter for cryo-electron microscopy data. J Struct Biol 172, 211-218.

Wriggers, W., Milligan, R.A., Schulten, K., McCammon, J.A., 1998. Self-organizing neural networks bridge the biomolecular resolution gap. J Mol Biol 284, 1247-1254.

Yu, X., Jin, L., Zhou, Z.H., 2008. 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. Nature 453, 415-419.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., Grigorieff, N., 2008. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc Natl Acad Sci U S A 105, 1867-1872.

FIGURE LEGENDS

Figure 1: Denoising of one synthetic and two experimental EM maps (**Experiments 1-3**). (A-D) Denoising of the synthetic density map of anthrax toxin protective antigen using σ=1.6 (the optimal σ regarding the FSC results shown in **Figure S1A**) and $\varepsilon = 1\%$ i.e., the FSC of the input and output density maps with respect to the reference density map (A) and an arbitrary slice (slice 104 along Z axis) of the reference, input, and output density maps ((B)-(D), respectively). (E-H) Denoising of the EM density map of 50S-ObgE complex (EMDB:EMD-2605) using σ=1.7 (the optimal σ regarding the FSC results shown in **Figure S2A**) and $\varepsilon = 1\%$ i.e., the FSC of the input and output density maps with respect to the reference density map (E) and an arbitrary slice (slice 116 along Y axis) of the reference, input, and output density maps ((F)-(H), respectively). (I-L) Denoising of the EM density map of beta-galactosidase (EMDB:EMD-5995) using σ=1.4 (the optimal σ regarding the FSC results shown in **Figure S3A**) and $\varepsilon = 1\%$ i.e., the FSC of the input and output density maps with respect to the reference density map (I) and an arbitrary slice (slice 144 along Y axis) of the reference, input, and output density maps ((J)-(L), respectively). In (A), (E), (I), a reduced range of spatial frequencies is shown for a better visibility. The FSC obtained for other values of σ and the reference-, input-, and output-map isosurface representations are provided in **Figures S1-3**.

Figure 2: Comparison of the FSC curves obtained by the proposed method for one synthetic and two experimental EM maps (**Experiments 1-3, Figs. 1A,E,I**) with the FSC curves obtained for the output of *bmedian*, *bbif*, *bnad*, *tomoand*, and *tomoeed* methods. (A-C) Comparison of different methods using synthetic anthrax toxin protective antigen data (A), experimental 50S-ObgE data (B), and experimental beta-galactosidase data (C), using non-masked output density maps obtained by *bmedian*, *bbif*, *bnad*, *tomoand*, and *tomoeed*. (D-F) Comparison of different methods using synthetic anthrax toxin protective antigen data (D), experimental 50S-ObgE data (E), and experimental beta-galactosidase data (F), using masked output density maps obtained by *bmedian*, *bbif*, *bnad*, *tomoand*, and *tomoeed*. The FSC curves obtained for the non-masked and masked input density maps are also shown. Slices of the different non-masked and masked maps for the three data cases are provided in **Figures S4-9**.

Figure 3: Denoising of a density map of immature HIV-1 capsid in intact virus particles (EMDB:EMD-2706) by the proposed method (using σ=0.5 and $\varepsilon = 5\%$) and *bmedian*, *bbif*, *bnad*, and *tomoeed* methods. (A) Input density map (transparent gray) and the output of the proposed method (solid orange) superposed and cut with the same plane using Chimera (Pettersen et al., 2004). (B) Another view of the overlap shown in (A). (C-H) Input density map (C), output of the proposed method (D), output of *bmedian*, *bbif*, *bnad*, and *tomoeed* ((E)-(H), respectively), cut with the same cutting plane as in (A) and shown in the same orientation as in (A) but using a higher density isosurface level than in (A). Note that the input density map is masked as deposited at EMDB.

Figure 4: Denoising of a density map of genogroup II genotype 10 norovirus virus-like particle (EMDB:EMD-5374) by the proposed method (using σ=1.5 and $\varepsilon = 15\%$). (A) Input density map (transparent gray) and the output of the proposed method (solid cyan) superposed

and cut with the same plane using Chimera. (B-E) Cut of the input density map for two different density isosurface levels, a lower one in (B) and a higher one in (C) (both higher than the isosurface level in (A)), and the corresponding cut of the output of the proposed method ((D) and (E), respectively). (F-G) Arbitrary slice (slice 108 along Z axis) extracted from the input density map (F) and from the output of the proposed method (G). In (B)-(E), a half of the density map, radially colored, is shown using Chimera (bluish densities are those on the medial slice or close to it that are among the most far away from the center of the slice while less distant ones are greenish; yellowish and reddish densities correspond to the slices that are further away from the medial slice and the most distant ones are in dark red).

Figure 5: Denoising of a density maps of human rhinovirus 2 135S full particle obtained by single-particle analysis (EMDB:EMD-2109) by the proposed method (using σ=1.3 and $\varepsilon = 5\%$). (A-D) Cut of the input density map for two different density isosurface levels, a lower one in (A) and a higher one in (B), and the corresponding cut of the output of the proposed method ((C) and (D), respectively). (E-F) Arbitrary slice (slice 132 along Z axis) extracted from the input density map (E) and from the output of the proposed method (F). In (A)-(D), a half of the density map, radially colored, is shown using Chimera (the same color code is used as in **Figure 4**).

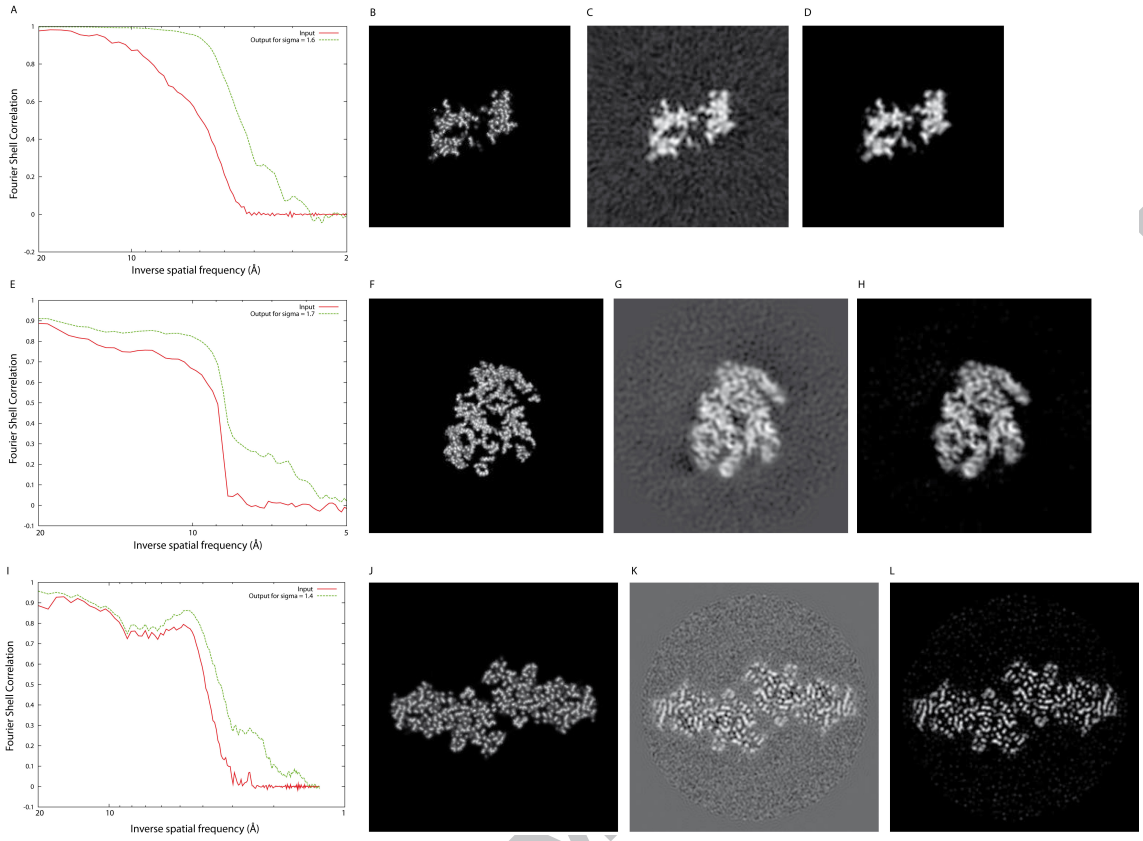

Fig. 1

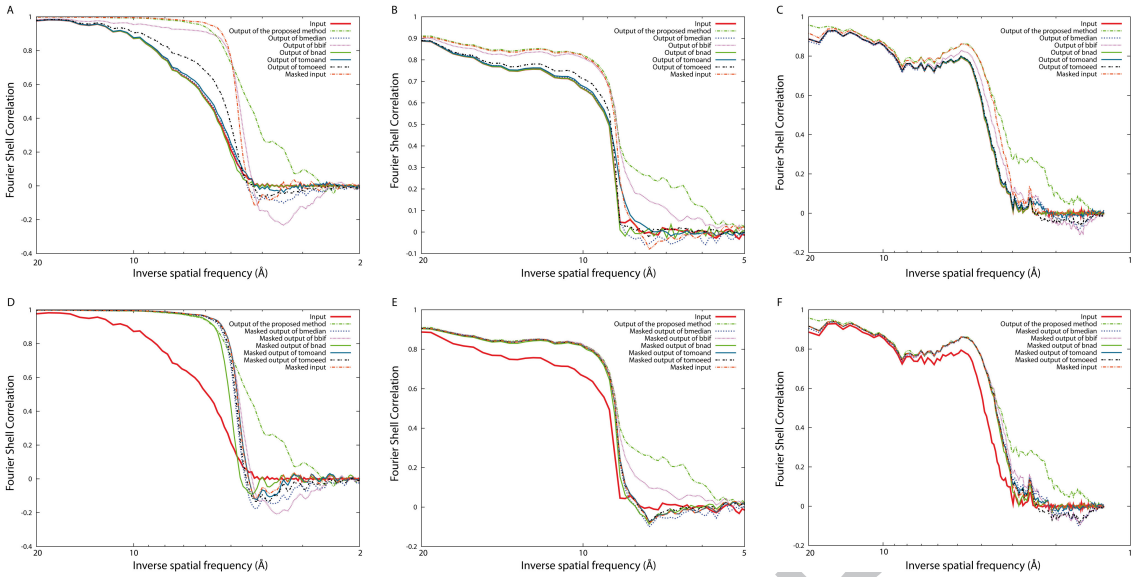

Fig. 2

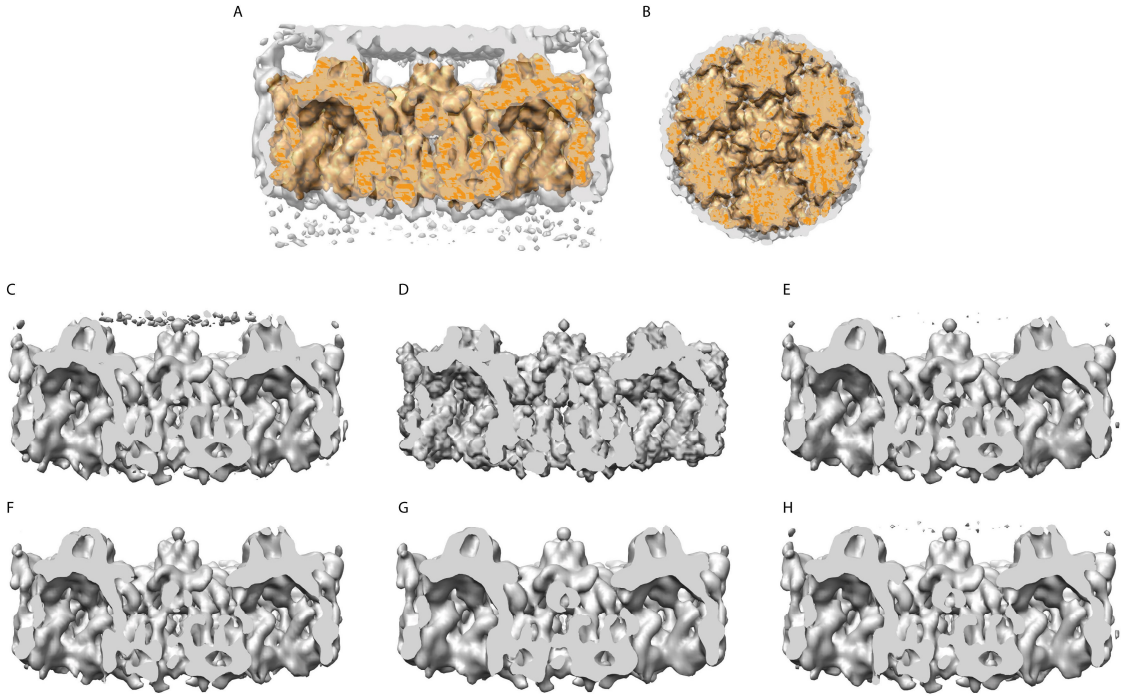

Fig. 3

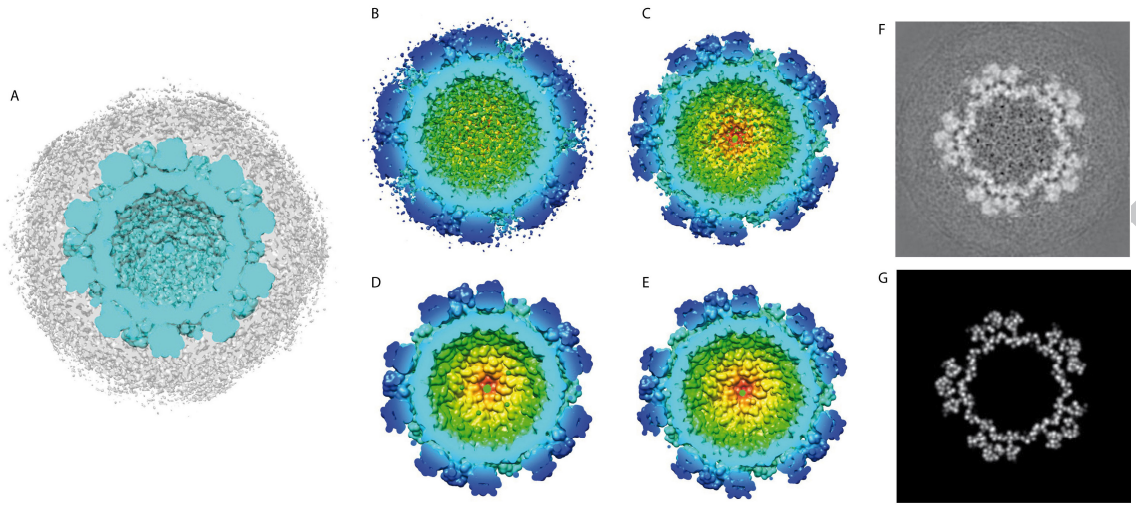

Fig. 4

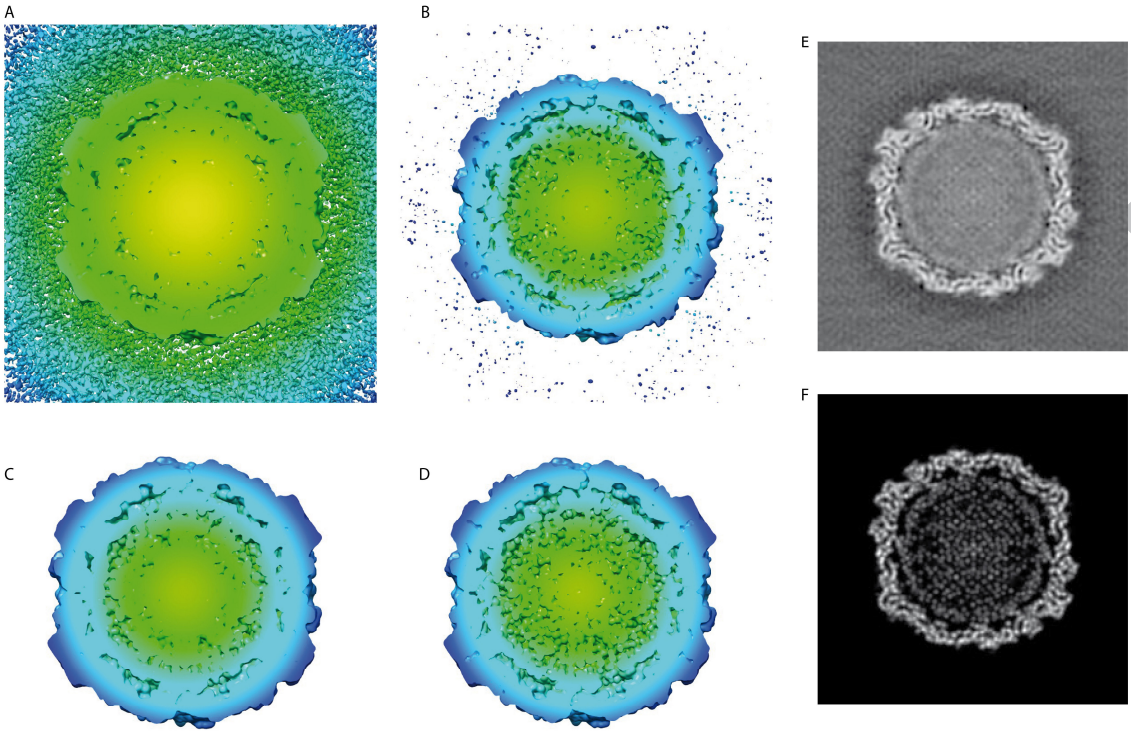

Fig. 5

Fourier Shell Correlation	Inverse spatial frequencies and differences								
	Synthetic density map			EMD-2605			EMD-5995		
	Input [Å]	Ouput [Å]	Difference [Å]	Input [Å]	Ouput [Å]	Difference [Å]	Input [Å]	Ouput [Å]	Difference [Å]
0.7	7.6	4.9	2.7	10.4	8.9	1.5	4.3	3.8	0.5
0.5	5.9	4.4	1.5	8.9	8.6	0.3	3.9	3.4	0.5
0.3	5.2	4.0	1.2	8.7	8.0	0.7	3.5	3.0	0.5
0.1	4.7	3.2	1.5	8.5	5.8	2.7	3.1	2.0	1.1

Table 1: Noise attenuation obtained by the proposed method and evaluated by measuring the output-map FSC curve shift to higher frequencies with respect to the input-map FSC curve, for synthetic (anthrax toxin protective antigen) and experimental (EMD-2605 and EMD-5995) input density maps (the columns denoted by different shades of transparent grey). The FSC curves for the input and output density maps of a complex were computed with respect to the same reference density map of the complex (see the text for more details on reference maps). The frequencies corresponding to the input- and output-map FSC curve values of 0.7, 0.5, 0.3, and 0.1 are shown in the first two columns of the part of the table related to each map and the differences among these frequency pairs are shown in the third column of the same part of the table.

	Standard deviation of Gaussian functions, σ	Number of Gaussian functions	Approximation error [%]
Synthetic density map	1.7	4,399	3.54
	1.6	**6,280**	**3.08**
	1.5	8,546	2.86
	1.4	11,519	2.66
EMD-2605	1.8	31,046	2.68
	1.7	**50,356**	**2.29**
	1.6	76,748	2.01
	1.5	98,666	1.84
EMD-5995	1.5	55,043	11.48
	1.4	**120,827**	**9.75**
	1.3	196,998	8.64
	1.2	269,189	7.80

Table 2: Numbers of Gaussian functions produced using different values of the Gaussian-function standard deviation (σ) and the target approximation error of the input density map of 1%, together with the achieved approximation error using these Gaussian functions, for synthetic (anthrax toxin protective antigen) and experimental (EMD-2605 and EMD-5995) input density maps. The values in bold denote the optimal solution obtained by analysing the FSC between the output and reference density maps (see the text for more details on reference maps). The results are shown for the optimal value of σ, two smaller values than the optimal value (smaller by 0.1 and 0.2) and one larger value than the optimal value (larger by 0.1).

	SNR			SNR(output)/SNR(input)		
	Synthetic density map	EMD-2605	EMD-5995	Synthetic density map	EMD-2605	EMD-5995
Input	25.7	8.2	4.2	N/A	N/A	N/A
Output of the proposed method	138.2	34.8	13.1	5.4	4.2	3.1
Output of bmedian	30.9	9.4	5.4	1.2	1.1	1.3
Output of bbif	60.2	13.5	6.6	2.3	1.6	1.6
Output of bnad	39.7	12.0	7.2	1.5	1.5	1.7
Output of tomoand	27.0	8.4	4.4	1.1	1.0	1.0
Output of tomoeed	37.4	9.9	5.5	1.5	1.2	1.3

Table 3: Noise attenuation evaluated by measuring the signal-to-noise ratio (SNR) of the input density map (synthetic anthrax toxin protective antigen map and experimental EMD-2605 and EMD-5995 maps) and the SNR of the corresponding output density map obtained by each tested method (the proposed method, bmedian, bbif, bnad, tomoand, and tomoeed). The ratio of the SNRs of the output and input density maps is also provided.