Journal of Structural Biology 194 (2016) 423–433

Contents lists available at ScienceDirect

Journal of Structural Biology

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

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

S. Jonić [a,*], J. Vargas [b], R. Melero [b], J. Gómez-Blanco [b], J.M. Carazo [b], C.O.S. Sorzano [b]

[a] IMPMC, Sorbonne Universités – CNRS UMR 7590, UPMC Univ Paris 6, MNHN, IRD UMR 206, 75005 Paris, France
[b] Biocomputing Unit, Centro Nacional de Biotecnología – CSIC, Campus de Cantoblanco, Darwin 3, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form 12 April 2016
Accepted 12 April 2016
Available online 13 April 2016

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.

1. 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

* Corresponding author at: IMPMC-UMR 7590, Université Pierre & Marie Curie, Case courrier 115, 4 Place Jussieu, 75005 Paris, France.
E-mail address: Slavica.Jonic@impmc.upmc.fr (S. Jonić).

http://dx.doi.org/10.1016/j.jsb.2016.04.007
1047-8477/

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, analysing 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.

2. Methods

2.1. 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 (manuscript 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.

2.2. 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.

3. 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.