Journal of Structural Biology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

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

A Bayesian approach for suppression of limited angular sampling artifacts in single particle 3D reconstruction

Toshio Moriya [a,b,*], Erman Acar [a,b], R. Holland Cheng [c], Ulla Ruotsalainen [a,b]

[a] Department of Signal Processing, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
[b] BioMediTech, Tampere University of Technology, P.O. Box 553, FI-33101 Tampere, Finland
[c] Department of Molecular and Cellular Biology, University of California, Briggs7 (MailCode#0390), Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 3 April 2015
Received in revised form 10 July 2015
Accepted 16 July 2015
Available online xxxx

Keywords:
Cryo-electron microscopy
Sparse angular sampling
Missing cone
Maximum *a posteriori* probability estimation
Median root prior reconstruction
Poisson image formation model

ABSTRACT

In the single particle reconstruction, the initial 3D structure often suffers from the limited angular sampling artifact. Selecting 2D class averages of particle images generally improves the accuracy and efficiency of the reference-free 3D angle estimation, but causes an insufficient angular sampling to fill the information of the target object in the 3D frequency space. Similarly, the initial 3D structure by the random-conical tilt reconstruction has the well-known "missing cone" artifact. Here, we attempted to solve the limited angular sampling problem by sequentially applying maximum *a posteriori* estimate with expectation maximization algorithm (sMAP-EM). Using both simulated and experimental cryo-electron microscope images, the sMAP-EM was compared to the direct Fourier method on the basis of reconstruction error and resolution. To establish selection criteria of the final regularization weight for the sMAP-EM, the effects of noise level and sampling sparseness on the reconstructions were examined with evenly distributed sampling simulations. The frequency information filled in the missing cone of the conical tilt sampling simulations was assessed by developing new quantitative measurements. All the results of visual and numerical evaluations showed the sMAP-EM performed better than the direct Fourier method, regardless of the sampling method, noise level, and sampling sparseness. Furthermore, the frequency domain analysis demonstrated that the sMAP-EM can fill the meaningful information in the unmeasured angular space without detailed *a priori* knowledge of the objects. The current research demonstrated that the sMAP-EM has a high potential to facilitate the determination of 3D protein structures at near atomic-resolution.

1. Introduction

3D structures of proteins at a near-atomic resolution provide valuable information for the understanding of physiological functions, and for the development of medicines. Single particle reconstruction (SPR) is a suitable method for the structural studies of proteins when their crystals are difficult to obtain with sufficient quality for crystallography (Frank, 2006). In general, the SPR assumes free orientations of the protein particles upon the imaging by cryo-electron microscopy (cryo-EM). This flexibility creates the necessity to estimate 3D orientations of the protein particles on the grids solely from their extremely noisy projected 2D images. For this, the traditional SPR process includes "*ab initio* 3D reconstruction" step where the initial estimate of the particle structure is computed using a reference-free 3D angle estimation and a 3D reconstruction method.

The reference-free 3D angle estimation is commonly done using the averaged particle views (Frank, 2006). These averages are calculated beforehand in the "reference-free 2D averaging" step, where particle images are aligned, classified, and averaged to improve signal-to-noise ratio of each 2D particle view. To improve the accuracy of the 3D angle estimation, it is a common practice to

Abbreviations: CTF, phase-contrast transfer function; DF, direct Fourier; EM algorithm, expectation maximization algorithm; FSC, Fourier shell correlation; FSMAR, Fourier shell mean amplitude ratio; FSMSE, Fourier shell mean squared error; MAP, maximum *a posteriori* probability; MAP-EM, maximum *a posteriori* probability expectation maximization; ML, maximum likelihood; ML-EM, maximum likelihood expectation maximization; MRP, median root prior; MSE, mean squared error; MSD2D, mean squared difference in the 2D spatial domain; MSE3D, mean squared error in the 3D spatial domain; OSL, one-step-late; SNR, signal to noise ratio; sMAP-EM, sequential maximum *a posteriori* probability expectation maximization.

* Corresponding author at: Department of Structural Biochemistry, Max-Planck-Institute of Molecular Physiology, Otto-Hahn-Str. 11, 44227 Dortmund, Germany.

E-mail addresses: toshio.moriya@mpi-dortmund.mpg.de (T. Moriya), erman.acar@tut.fi (E. Acar), rhch@ucdavis.edu (R.H. Cheng), ulla.ruotsalainen@tut.fi (U. Ruotsalainen).

http://dx.doi.org/10.1016/j.jsb.2015.07.007
1047-8477/

select a small number of the 2D class averages with the highest qualities (less than 100 classes). However, this class selection causes an insufficient number of angular sampling to fill the information in the 3D frequency space of the interest. Missing gaps in this sparse sampling correspond to the projection angles of the excluded 2D classes. An alternative approach of 3D angle estimation uses physical tilting of the specimen upon imaging with electron microscope. The known angle of the tilt constraint provides the relative angels between the tilt pairs of 2D projection views. This additional information simplifies the computational determination of the 3D angles. For proteins, the most widely-used method of this approach is the random-conical tilt reconstructions (Radermacher et al., 1986). However, this reconstruction method has the well-known drawback of the "missing cone" since the maximum tilting angle is limited to approximately ±60° with conventional electron microscope. Therefore, the initial estimation of structure is highly possible to be contaminated with a significant amount of artifacts. The artifacts can cause errors through the reference bias in the 3D refinement, known as "Einstein from noise" (Henderson, 2013; van Heel, 2013). It is important to suppress the effects of the sparse sampling gaps or missing cone on the quality of the initial 3D reconstruction.

In the SPR field, the artifact contaminations in 3D reconstruction with limited angular sampling have been well known (Penczek, 2010a) but no well-established solutions have been proposed. Currently, many varieties of back-projection and algebraic 3D reconstruction methods are available through the freely-distributed SPR software packages (Hohn et al., 2007; Scheres, 2012a; Scheres et al., 2008; Shaikh et al., 2008; Tang et al., 2007; Van Heel et al., 2011). These methods are simple and efficient. The packages also support methods that iteratively refine 3D volume by minimizing errors between the measured and estimated projections to minimize noise-induced inconsistency. The same approaches have been well studied for 2D image reconstruction from measured 2D projections with the missing information in the field of tomography (e.g. emission, transmission, and electron tomography). These studies have shown that back-projection methods are sensitive to the missing information and can result in severe artifacts (Delaney and Bresler, 1998) and that algebraic methods also do not fill the information gaps (Penczek, 2010a).

In contrast, there are tomography studies showing that some statistical reconstruction methods are capable of suppressing the artifacts yielded from the missing information, even when detailed *a priori* knowledge of the objects is not available (Paavolainen et al., 2014; Tuna et al., 2013). These studies used the maximum *a posteriori* probability estimates with expectation maximization algorithms (MAP-EM). An inspiring study is the application in electron tomography of biological samples, where the authors applied the sequential MAP-EM (sMAP-EM) method to compensate for missing wedges of 2D image reconstructions (Paavolainen et al., 2014). This sequential method was originally developed for the suppression of the artifacts yielded from the missing angles in positron emission tomography and single-photon emission computed tomography (Tuna et al., 2013). It was consistently superior to a weighted back-projection and a simultaneous iterative reconstruction technique. The sMAP-EM showed robustness against noise, and suppressed artifacts by filling meaningful information in the missing wedge (Paavolainen et al., 2014). The MAP-EM based approaches have been used also in the 3D refinement and 3D classification of SPR (Kucukelbir et al., 2012; Lyumkis et al., 2013; Scheres, 2012a, 2012b). The recent growth of statistical reconstruction methods is prominent in the SPR field (Scheres, 2010; Sigworth et al., 2010), and lead to "the resolution revolution" as described in a recent review (Kühlbrandt, 2014). All the three near-atomic resolution SPR structures referred in this review used the statistical reconstruction approach implemented in RELION (Scheres, 2012a). It is possible now to achieve the 3 Å-level resolution even for small membrane protein with low symmetry, such as TRPV1 ion channel (300 kDa and C4 symmetry) (Liao et al., 2013). However, the objectives of these statistical 3D reconstruction methods have not been in the compensation for limited angular sampling. It is because a reasonably dense sampling can be usually assumed at the 3D classification and 3D refinement stages if the number of the observed particle images is sufficient (conventionally at least 5000 images) and the target protein does not have strong "preferred orientations". However, it is frequently not reasonable to assume the same for the initial 3D reconstruction right after the reference-free 3D angle estimation.

The current research aims to develop a statistical 3D reconstruction method which performs well even with sparse angular sampling or conical tilt sampling used in the *ab initio* 3D reconstruction step of SPR. Since it is not necessary to estimate the 3D projection angles simultaneously in the initial 3D reconstruction, the task for the 3D reconstruction is much simpler than the ones for the previous MAP-EM based algorithms in SPR. This allowed us to focus on the issue of the limited angular sampling cases. Here, we propose to expand the concepts of the sMAP-EM to the 3D domain and develop the implementation for SPR. The evenly distributed and the conical tilt angular samplings of simulated datasets were examined with various noise levels. The three different levels of sparseness were also examined with the evenly distributed angles. To understand the characteristic and behavior of the sMAP-EM reconstruction process, the Fourier shell correlation (FSC) (van Heel and Schatz, 2005) and the mean squared error (MSE) against the ground truth were used as the numerical measurements. Following the convention of the SPR field, the 3D resolutions were evaluated using the FSC between two independent reconstructions, along with the visual inspection. To numerically assess the gap filling ability of sMAP-EM in the frequency domain, the Fourier shell mean amplitude ratio (FSMAR) and Fourier shell MSE (FSMSE) were developed in this study. Finally, the sMAP-EM reconstruction from the experimental cryo-EM particle images was conducted. We expect that the sMAP-EM would be able to suppress the artifacts in the *ab initio* 3D reconstruction of SPR with limited angular sampling by estimating meaningful values for the information missing gaps.

2. Theoretical background

2.1. Bayesian inference

The Bayesian inference can be expressed as:

$$P(\Theta|X) \propto P(X|\Theta)P(\Theta),$$

where $P(\Theta|X)$ is *a posterior* probability density function of a parameter set Θ when a set of particle images X has been observed, $P(X|\Theta)$ is the likelihood function of X when Θ is known already, and $P(\Theta)$ is the *a priori* probability density function of Θ. The most important parameter in Θ is a voxel value set of a 3D volume for the 3D reconstruction in the SPR, since this is the objective. After we observe a set of particle images, the best estimate of 3D structure is the particular one having the maximum *a posterior* probability among all the other possible reconstructions in a given 3D space. This probability is not directly accessible but it is same as the right-hand terms which can be calculated using an estimate of Θ including 3D volume. This estimate can be refined using an iterative procedure. The expectation maximization (EM) algorithm (Dempster et al., 1977) is a popular choice to compute the maximum *a posteriori* probability (MAP) estimation using this approach, and so called MAP-EM.

Please cite this article in press as: Moriya, T., et al. A Bayesian approach for suppression of limited angular sampling artifacts in single particle 3D reconstruction. J. Struct. Biol. (2015), http://dx.doi.org/10.1016/j.jsb.2015.07.007