The Alignment and Classification of 3D Reconstructions of Rod-Like Molecules Obtained by Electron Tomography

Claire J. Brooks[1], Teresa Ruiz[1], and Michael Radermacher[1]

[1]Department of Molecular Physiology & Biophysics, University of Vermont, Burlington, VT

Electron tomography, a 3D reconstruction technique for cellular organelles and macromolecular complexes, is particularly useful for the structural characterization of heterogeneous samples. Averaging techniques have been developed to increase the signal to noise ratio and eliminate artifacts induced by the missing wedge, thus resulting in higher resolution 3D reconstructions. However, accurate subvolume alignment and classification methods are essential for effective volume averaging. The missing wedge artifact that originates from the limited angular range of tomographic tilt series can interfere with the accuracy of alignment and averaging. Using the functional domain of the collagen adhesin EmaA of *Aggregatibacter actinomycetemcomitans* as a model, we developed a workflow of iterative multireference alignment and classification of 3D reconstructions of rod-liked molecules using processes that are cognizant of and minimize the effects of the missing data, while maintaining the subprojections from which subvolumes were calculated.

Elongated molecules were selected from tomograms generated using IMOD by marking two points on their long-axis [1]. For each selected molecule, a tilt series of subprojections was extracted and the angles recalculated to orient the molecules' long-axis approximately parallel to the Y-axis, which represents the rotation axis for Theta, using algorithms in EMIRA [2]. Subvolumes were calculated by averaging the 2D Radon transformed projections into a 3D Radon volume, where every line of the 3D transform contains an occupancy index that keeps track of the number of 2D transforms averaged into it. This index can be used to determine the location of the missing data, where the index value is zero [3]. The subvolumes were aligned using projection-based 3D volume alignment (PB3A). This method aligns a sample volume to a reference by comparing the Radon transforms of a set of reference projections to the 3D Radon transform of a volume resulting in a rotational cross-correlation function [4]. The occupancy index serves to exclude missing data from the cross-correlation and to maintain accurate normalization.

The reference volume for the EmaA adhesin was projected into 13 directions over a ±60° angular range to cover all rotationally sensitive features of the molecule. The subvolumes and reference projections were Fourier transformed with Fermi high pass and low pass filters, as well as a sqrt(ρ^*) filter (where ρ^* is the Fourier radius) to obtain a sharper cross-correlation maximum. As the rod-shaped molecules are already oriented along Y-axis, Theta was given a search range of +/- 180° to achieve complete angular coverage around Y, with Phi and Psi given a much smaller range such as +/- 10° to provide a margin of freedom for the pivoting of the rod-like molecule. In addition, the search included a 5-10 voxel translational search to account for small subvolume shifts.

The aligned volumes were grouped using Probabilistic Principle Component Analysis with Expectation Maximization (PPCA-EM) and a clustering method, such as Diday's method of moving centers [5,6] (Fig 1). PPCA-EM is a method of classification that treats missing data and principle components as hidden variables and, not only compensates for the missing data, but estimates them for each subvolume. The method relies on the occupancy index in the 3D Radon transform to identify the missing data. In

contrast to earlier methods, with PPCA-EM the missing data can be recovered (provided that data is present in at least one subvolume), even in cases where continuous variations are present without relying on class averages. Representative averages for the clusters were calculated by *de novo* reconstruction from the combined 2D Radon transforms of the subprojections and were used as references in a subsequent multireference alignment step (Fig 2). For each cluster, the same number of volumes were averaged to avoid possible bias caused by signal to noise ratio differences between the references. Iterations of multireference alignment using PB3A and classification with PPCA-EM and Diday's method were performed until clusters were stable. Final 3D structures were calculated by *de novo* reconstruction from the 2D Radon transforms of the subprojections of all subvolumes in each cluster.

References
[1] Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., J. Struct. Biol., **116** (1996), p.71-76.
[2] Radermacher, M., Microscopy & Microanalysis, **19** (2013), Supplement S2 p.762-763.
[3] Radermacher, M. et al, Ultramicroscopy, **53** (1994), p.121-136.
[4] Yu, L. et al. Journal of Structural Biology **182** (2013), p. 93–105
[5] Yu, L. et al. Journal of Structural Biology, **171** (2010), p.18-30
[6] Diday E., Rev Stat Appl., **19** (1971), p.19–34.
[7] This work was supported by NIH grant DE024554, (T.R. & K.P.M) and GM078202 (M.R.).
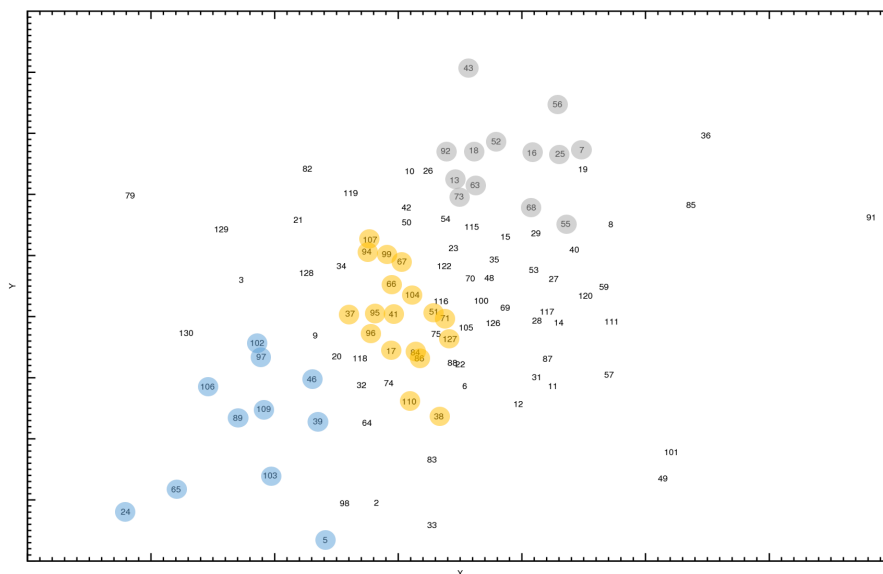
Figure 1. Non-linear map of PPCA-EM factors 2 through 8, with three largest clusters calculated with Diday's method of moving centers highlighted in yellow, grey, and blue respectively.
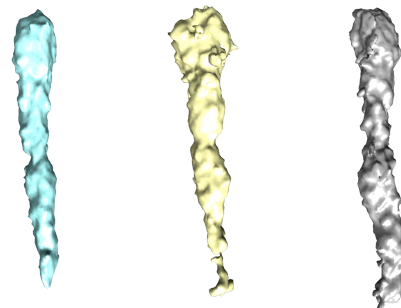
Figure 2. Surface representation of representative average volumes for the clusters calculated for multireference alignment from PPCA-EM clusters. Length represented: 30 nm.