ELSEVIER

Crystallization Notes

Methods for aligning and for averaging 3D volumes with missing data

Michael F. Schmid *, Christopher R. Booth [1]

*National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology,
Baylor College of Medicine, Houston, TX 77030, USA*

Abstract

The visibility and resolution of a tomographic reconstruction containing multiple copies of discrete particles can be enhanced by averaging subtomograms after they are corrected aligned. However, the "missing wedge" in electron tomography can easily lead to erroneous alignment. We have explored a Fourier space cross-correlation method with a proper weighting scheme to align and average different sets of volumetric data, each of which has different missing data due to the limited specimen tilts. This approach depends neither on a preexisting template, nor an exact knowledge of the geometry, orientation, or amount of the missing data. This paper introduces a procedure where the missing data might be gradually "filled in" by consecutively aligning and averaging volumes with different orientations of their missing data. We have validated these techniques by a set of simulated data with various symmetries and extent of missing data. We have also successfully applied these procedures to experimental cryo-electron tomographic data [Chang, J.T., Schmid, M.F., Rixon, F.J., and Chiu, W., 2007. Electron cryotomography reveals the portal in the herpesvirus capsid. J. Virol. 81, 2065–2068; Schmid, M.F., Paredes, A.M., Khant, H.A., Soyer, F., Aldrich, H.C., Chiu, W., and Shively, J.M., 2006. Structure of Halothiobacillus neapolitanus carboxysomes by cryo-electron tomography. J. Mol. Biol. 364, 526–535].
© 2007 Elsevier Inc. All rights reserved.

Keywords: Cryo-electron tomography; Computational methods; Aligning and averaging; Electron cryotomography

1. Introduction

After completing a tomographic reconstruction, it is often feasible and desirable to align and average similar objects that occur in the reconstructed volume. Cryo-electron tomograms that contain multiple copies of discrete macromolecular assemblies can be used to develop such a 3D average. For the case where the composition is known and 3D models are created or are available, as described below, for use as templates, procedures have been developed and applied to do this. However, there are circumstances where the size, shape and/or symmetry of the objects are not known *a priori*, and it would be advantageous to have a method that does not need a starting template. Such a method would involve all-vs.-all comparisons of the 3D objects against each other, the criterion for choosing both the correct orientation and the best-matching pairs of particles usually being the cross-correlation peak height. Therefore the critical requirement for such a method to work is that 3D volumes extracted from the tomogram must be able to be aligned properly to each other. This study investigates the effect of symmetry, relative orientations and completeness of the data in Fourier space on the feasibility of aligning objects to each other. This method has been applied to classify, align and average subvolumes that turned out to be icosahedral particles which varied in size (Schmid et al., 2006).

Because a tilt angle range of ±90° cannot be achieved in the electron microscope, central sections in Fourier space for tilts higher than 60–70° are missing, leading to the problem commonly referred to as the "missing wedge" in Fourier space (Hoppe and Hegerl, 1980). After extraction of 3D subvolumes from the tomogram and before applying any rotations, the orientations of the missing wedges in

---

* Corresponding author. Fax: +1 713 798 1625.
 E-mail address: mschmid@bcm.tmc.edu (M.F. Schmid).
[1] Present address: Gatan, Inc., Pleasanton, CA 94588, USA.

Fourier space are identical for each particle/volume. However, the orientations of the particles themselves are unknown with respect to each other and must be searched for in 3D rotation space. Because of the effect of the missing wedge, the cross-correlation map may merely lead to the alignment of the two missing wedges to each other, which is certainly not correct. This problem is not as severe in cases where an isotropic starting template is used for 3D alignment (Walz et al., 1997), although accounting for the missing wedge in the data to be aligned is still important (Frangakis et al., 2002). Here "isotropic" is used to refer to a template whose resolution is similar in all directions. The template may be based on a higher resolution structure (Bohm et al., 2000; Frangakis et al., 2002), a rotational (usually cylindrical) average of the particles (Beck et al., 2004; Zanetti et al., 2006; Zhu et al., 2006), a simple object like a cylinder (Nickell et al., 2007), or on an arbitrarily chosen example from the data set (Murphy et al., 2006). The starting orientation may also be estimated (Forster et al., 2005). However, in the case of the carboxysome, for instance, it was not possible to postulate any starting model, because the size and symmetry of the particle were not known *a priori*. The method described here is applicable without the use of a preexisting model of the computationally extracted subvolume from the tomogram and thus can be a general solution to post-tomographic averaging.

In addition, the proper scaling of these aligned 3D volumes during averaging should also take into account the fact that some data from the aligned particles is missing. In real space, the missing data is convoluted with the genuine data, but in Fourier space, they are separable. A weighting scheme is proposed for 3D volume averaging.

2. Examples used

To illustrate and test this method, we are using simulated data at known orientations with varying amounts of missing data. Our purpose is to specifically investigate the effect that the size of the missing data wedge, the starting orientations of the particles and the symmetry have on structure alignment for tomography.

Our data includes an icosahedral herpesvirus pentonless nucleocapsid map (EMD-1305) (Chang et al., 2007), a d7 symmetry GroEL map (EMD-1080) (Ludtke et al., 2001), and a c1 symmetry ribosome map (EMD-1003) (Gabashvili et al., 2000). These test cases represent macromolecular complexes with different symmetries. All were low-pass filtered to 40 Å nominal resolution. The herpes capsid map was 90^3 pixels at 16.2 Å/pixel. The GroEL map was resampled to 44^3 pixels at 6.1 Å/pixel, and the ribosome map was resampled to 32^3 pixels at 5.9 Å/pixel. Four orientations of the GroEL and ribosome were used (three mutually perpendicular orientations, and one at 50°). For the herpes case, two views, a 5-fold along *z* and a 3-fold along *z*, were tested.

We then created missing wedges in the Fourier transforms of these 3D maps, centered about the *z* axis for each orientation of the maps, varying from 0 to 50% missing data, equivalent to tilt series of ±90° (0% missing data) down to ±45° (50% missing data).

3. The alignment problem and our solution

As one of a pair of 3D volumes represented in Fourier space, both containing a missing wedge, is rotated against the other in the orientation cross-correlation search, and one volume is multiplied by the complex conjugate of the other, zeros are generated. They occur when the missing-data region for one of the particles is multiplied by data or zeros in the other particle and vice versa, and the number of such zeros changes at each rotation angle in the search. This is shown schematically in Fig. 1. The fraction of zero data has a profound effect on the peak height of a correlation map as shown in Fig. 2. If we renormalize the cross-correlation peak at each orientation by a factor of 1/(fraction of the non-zero data), we restore the relative peak height to approximately the value it had with no missing data (Fig. 2). Of course, the result of the correction depends on what data was zeroed, which will vary depending on the starting orientations of the particles. This is illustrated in Fig. 3 and analyzed further in the Discussion, but Fig. 2 suggests that the volume-fraction of non-zero data appears to be an acceptable approximation. It also has the advantage that it is a quick and easy calculation to perform.

4. Averaging in Fourier space

Our approach was simply to amplitude-weight the amplitude sum for the real and imaginary parts of the Fourier voxels when averaging two or more rotationally and translationally aligned 3D volumes together.

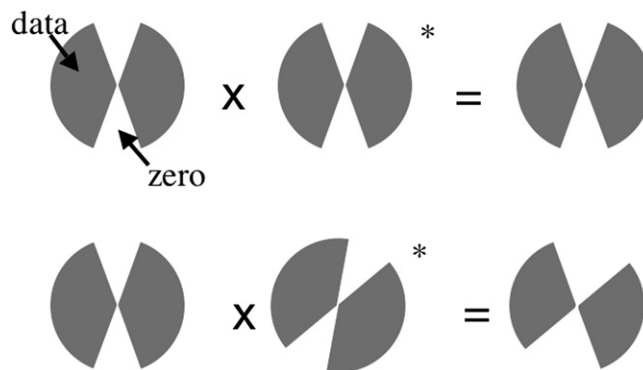$$A_t = \frac{1}{\mathrm{amp}_t} \Sigma_i A_i * \mathrm{amp}_i$$

Fig. 1. Cartoon showing the relationships of the missing wedges of two particles in Fourier space. Before the orientation search (top row), the missing wedges are aligned with maximum overlap. As one particle is rotated with respect to the other (bottom row), zeros in either Fourier transform will yield zeros in the complex product.