

# A simple and fast approach for missing-wedge invariant classification of subtomograms extracted from filamentous structures

Jagan Mohan Obbineni<sup>1</sup>, Ryosuke Yamamoto<sup>2\*</sup> and Takashi Ishikawa<sup>1</sup>

<sup>1</sup>Laboratory of Biomolecular Research, Paul Scherrer Institute

<sup>2</sup>Shimoda Marine Research Center, University of Tsukuba

\*Present address: Department of Biological Sciences, Graduate School of Science, Osaka University

## Abstract

Unsupervised classification of subtomograms extracted from cryo-electron tomograms is often challenging due to the presence of a missing wedge in tomographic data. Here, we propose a simple new approach to classify subtomograms extracted from cryo-electron tomograms of filamentous objects. This unsupervised classification approach uses the 1D projections of the subtomograms for classification and works independently of the orientations of the missing wedge. We applied this approach to subtomograms from eukaryotic cilia and successfully detected heterogeneity including structural polymorphism of dynein molecules.

## Introduction

Subtomogram averaging enables in vivo molecular structural analysis by cryo-electron tomography despite poor signal-to-noise ratio (S/N) of this technique. In subtomogram averaging, volumes including the target molecule are extracted computationally, aligned to have the same orientation and averaged. However, when there is heterogeneity in the data, it leads to imprecise alignments thereby resulting in a poor average. Image classification of subtomograms is done in order to get rid of heterogeneity of molecular structures due to enzymatic structural polymorphism, thermal fluctuation or coexistence of different molecules which are similar and cannot be distinguished visually. When there is *a priori* knowledge about heterogeneity, template based classification is possible and this has been applied successfully in eukaryotic flagellar tomography to look at conformations of dynein corresponding to the pre- and post-power stroke (Lin et al., 2014; Movassagh et al., 2010; Ueno et al., 2014). However, in general unsupervised (template free) classification is preferable, since no *a priori* knowledge is required and there is a less risk of bias.

Unsupervised classification of images (2D or 3D) by multivariate statistical analysis requires aligned image particles and this is often a challenge because alignment based on cross-correlation is imprecise when heterogeneous structures coexist. Simultaneous alignment and classification provides an

alternative to overcome the problem of imprecise alignments. Another challenge arises due to the presence of missing wedge in tomographic data. Subtomograms having identical structure but different orientations seem to have different structure, because of the different orientations of the missing wedge, resulting in classification primarily based on the orientations of missing wedge, instead of the real structural heterogeneity.

A number of methods have been developed and applied successfully for unsupervised classification of subtomograms independent of the missing wedge (Bartesaghi et al. 2007; Förster et al., 2008; Scheres et al., 2009; Yu et al., 2010; Stölken et al., 2011; Heumann et al. 2011). Förster et al. 2008 and Bartesaghi et al. 2007 address this challenge by using only the non-missing data for computation. Other methods either fill the missing data using imputation and simultaneously align and classify particles (Scheres et al., 2009; Yu et al., 2010; Stölken et al., 2011) or construct variance maps to obtain wedge-masked differences (Heumann et al. 2011). Another possible approach is to reduce the information into a missing wedge invariant form. Alignment invariant transformation of 2D image particles (Schatz and van Heel, 1990) and helical reconstruction (Ishikawa and Wakabayashi, 1995) have been developed for straightforward MSA based classification in the past. In this work, we propose a simple new approach for quick classification of subtomograms extracted from filamentous objects by transformation into data which is invariant of missing wedge orientation. In this approach, a 1D projection of the 3D volume is calculated and classified using principle component analysis and hierarchical clustering using ward's criterion. We applied this approach to successfully detect partial occupancy of proteins and structural polymorphism in subtomograms extracted from tomograms of cilia/flagella.

## Materials and Methods

*Flagella preparation:* Axonemes were prepared from wild type *Tetrahymena thermophila*, wild type (cc-125) and *pf23* mutant (cc-1383) of *Chlamydomonas reinhardtii* as done previously (Maheshwari et al., 2015; Witman, 1986). Briefly, cilia/flagella were isolated by treating the cells with dibucaine and demembrated using a detergent. Membrane free cilia/flagella (axonemes) were spin down and used for tomography.

*Electron tomography:* 3.5 $\mu$ l solution of isolated axonemes was mounted on holey carbon grids (Quantifoil, Germany) and plunge-frozen using Cryoplunge 3 (Gatan, USA). Frozen grids were transferred by Gatan 626 system to JEM2200FS TEM equipped with an FEG and an in-column energy filter. In the search mode, axonemes located nearly parallel to the tilt axis were chosen. Micrographs were recorded by an F416 4096 x 4096 CMOS detector (TVIPS, Germany), using SerialEM software (Mastronarde, 2005).

*Tomography reconstruction and basic subtomogram analysis:* Tomographic reconstruction was carried out using IMOD software (Mastronarde, 1997). Subtomogram averaging was done using in-house programs with the core alignment part based on Spider (Frank et al., 1996) as described elsewhere (Bui and Ishikawa, 2013). Subtomograms with target dynein molecules in the center, were extracted and aligned using pseudo-9 fold symmetry and 96nm periodicity of the axoneme.

*Subtomogram classification:* Subtomograms after alignment were masked in order to extract the molecules of interest. In this work, we prepared masks to classify (1) one outer dynein arm (with three dynein heavy

chains) and (2) inner arm dynein a and b. The positions of these dyneins have been defined based on previous work (Bui et al., 2012). All the masks were prepared with a 2 pixel smoothing. Masked density maps were projected one-dimensionally on the tilt axis. Unsupervised classification was performed by principle component analysis followed by hierarchical clustering using ward's criterion in Spider (Frank et al., 1996). Data was reduced into 10 factors using principle component analysis and a few factors with significant decay of eigenvalues as cutoff were used for hierarchical clustering. If necessary, we visually examine the subaverages to find redundancy among subaverages and merge subclasses which provide visually similar subaverages.

## Results and Discussion

*Classification using 1D projection:* We performed unsupervised classification of subtomograms using one dimensional projection of the subtomograms. We define the y-axis as the tilt axis. The axis for the 1D projection (defined as y'-axis in this work) corresponds to the line in the xy plane overlapping with the axis of the filament as shown in Fig.1A (z is the orientation of the electron beam). Since we collect tomographic datasets of cilia, oriented nearly parallel to the tilt axis, the difference between y and y' is small. In Fig.1A, four subtomograms extracted from one tomogram of cilia are shown. In Fourier space, all the central sections, which have a thickness  $2/d$  (Crowther et al., 1970), include the Y-axis. When subtomograms are projected exactly on the tilt axis, the 1D FFT of the 1D projection is the common line, which is sampled in all the micrographs and thus has high S/N and reliability (Fig.1CD). When subtomograms are not exactly parallel to the tilt axis, the orientation of the 1D FFT of the 1D projection is also rotated from the common line (Fig.1D).

*Classification of simulated data with pre-, post-powerstroke and absent ODA:* In order to assess and validate our approach we performed simulations with model datasets where the truth is already known (Table 1). Specifically, we simulated the effect on classification results under varying levels of noise, extent of missing wedge, difference between the axoneme axis and tilt axis and error in alignment of the subtomograms. To carry out the simulations we took three model data with pre-, post-powerstroke (Ueno et al., 2014) and absent ODA (Fig. S1). Classification of a randomly generated dataset with the model subtomograms without any added noise or wedge yielded a correct classification using our approach and we observed that two eigenfactors described all the variance between the three models. We then performed simulations to assess the effect of varying levels of noise (Fig. S2). At very high noise levels ( $S_3$  in Fig. S2), we could only distinguish between presence and absence of ODA but not subtle change between pre- and post-powerstroke ODA. The simulated data has particles which are  $120^\circ$  rotated around the microtubule axis and hence have different missing wedge orientations. When we plot the distribution of the particles after classification into different classes, we see that the classification was independent of the missing wedge orientation (Fig. S2C). In order to assess the influence of the extent of missing wedge on the classification, we classified data with a missing wedge of  $30^\circ$ ,  $40^\circ$  and  $50^\circ$  (Fig. S3). We observed that the classification is sensitive to the extent of missing wedge and cannot distinguish subtle changes (between pre- and post-powerstroke dynein in our data) if we have a tilt series of  $\pm 50^\circ$  or less. It could still distinguish between presence and absence of ODAs. As this approach uses a 1D projection along the tilt axis of the subtomogram for classification, it would be important to assess also the influence due to the extent of difference between the filament axis and the tilt axis. We checked this with another set of simulations varying the angles between the filament axis and tilt axis (Table 1 and Fig. S4) and found that even at a difference in angle of  $60^\circ$  we could successfully classify the particles.

This is possible because at a large angle between the filament axis and tilt axis the information along the low-resolution range still seems to be enough to classify even subtle conformational changes (Fig. 1D). A prerequisite for this classification approach is the alignment of the subtomograms prior to classification and an error in the alignment could also influence the classification. In order to assess the extent of error tolerable by the approach we performed a series of simulations with varying errors in the three rotation angles and the three translations (Table 1 and Fig. S5). Classification results were very sensitive to the extent of mis-alignment in the subtomograms. Beyond an error of 2 pixels in translation and 2° in rotations it was hard to distinguish between the pre- and post-stroke dynein conformations although the presence and absence was still detectable.

One could also use this approach to classify heterogeneity of proteins sitting on a sheet or membrane. We showed this using a simulation where we assume a hypothetical six-fold symmetric protein sitting on a sheet having an open and closed state (Fig. S6). A random dataset generated from these two model subtomograms was successfully classified thus showing that the application of this approach is not limited to filamentous particles alone. As with other PCA based approaches, this approach will yield the defined number of classes irrespective of the data being heterogeneous or not. To address this issue we plotted the first eigenvector for each subtomogram from a simulation performed earlier (shown in Fig. S2). The plot showed a clear separation of the three states (pre-, post-powerstroke and absent ODAs) (Fig. S7). This can be used as a criterion to determine if there is actual heterogeneity in the data. In order to compare our approach with unsupervised ML-based classification we used the same simulated dataset and performed classification with ML. The results after classification were not as expected (Fig. S4D). However, our approach using PCA with the 1D projection of the subtomograms along the tilt axis yielded the expected result.

*Classification of Tetrahymena ODA (Control):* The 3D structure of *Tetrahymena* cilia was analyzed (Pigino et al., 2012) and is known to contain ODAs at 24nm periodicity along all the nine microtubule doublets (MTDs) in the axoneme. Hence, this dataset serves as a good control experiment as all the extracted subtomograms should contain ODAs. We used aligned subtomograms from *Tetrahymena* and performed unsupervised classification of 1D projection of density of a single outer dynein arm (ODA) masked from subtomograms corresponding to one 96nm periodic unit. Although the box size is 120nm, we masked only the area of interest. In this case we mask to cover one dynein head for each box. In Fig.S1 the mask is indicated. Only density inside the mask (white in the right panels of Fig.S1a) was used for classification, while the area outside the mask was filled with a uniform density. Therefore the size of the box does not influence this classification. Upon classification we obtained four classes ( $C_1-1$ ,  $C_1-2$ ,  $C_1-3$  and  $C_1-4$  in Fig. 2). Almost all of the particles are classified into a single class ( $C_1-1$  in Fig. 2) implying presence of ODAs on all the doublets, which is in line with the previous observation (Pigino et al., 2012). The other classes seem to have distortions in the ODA density due to presence of gold beads nearby or due to deletion of gold beads from the final volumes ( $C_1-2$ ,  $C_1-3$  and  $C_1-4$  in Fig. 2). Additionally, plotting the first eigenvector for each subtomogram revealed no clear separation of particles indicating a very homogenous data with respect to the ODA (middle panel in Fig. S7).

*Classification of Chlamydomonas wt ODA (Control):* Previous observations using EM sections (Hoops and Witman, 1983) and cryo-electron tomography (Bui et al., 2012) of *Chlamydomonas* wild-type flagella have shown that one out of the nine MTDs lack ODAs. Hence, it serves as another control to validate our new approach. In this experiment, we used 338 aligned subtomograms from *Chlamydomonas wt* and performed

unsupervised classification using the major eigenfactors (Fig. 3B), focusing on the ODA. We obtained six classes (Fig. 3B). 41 subtomograms were classified to a class which showed an average without ODA ( $C_{1-3}$  in Fig. 3). These subtomograms are mainly localized to the fifth doublet in the tomogram (Fig. 3C) and is consistent with the observation that *Chlamydomonas* lacks ODAs on one of the nine doublets. However, this observation was clear only when using a loose mask covering four ODAs (shown in Fig. 3A). Moreover, we observed that there are several subtomograms from this doublet classified in other classes (Fig. 3C). We speculate that this misclassification maybe a result of alignment of these subtomograms, which actually lack ODA to a reference with ODA, resulting in over-fitting of noise to ODA.

*Classification of Chlamydomonas pf23 mutant ODA:* *pf23* is a *Chlamydomonas* mutant having paralyzed flagella and hence it is interesting to look at the ultrastructure of the flagella from this mutant. 551 subtomograms were clustered into two classes with 335 and 216 particles (Fig. 4A, class  $C_{1-1}$  and  $C_{1-2}$ ). The main difference between the two classes is located at beta-dyneins (middle panels in Fig. 4A). This suggests two polymorphic structures. While the tails resemble each other (arrowheads in Fig. 5B and C), the AAA rings of  $\beta$ -dynein are oriented differently – in the view angle presented, the  $\beta$ -dynein head shows a clear ring in class  $C_{1-2}$  (arrow in Fig. 5B) while it appears as two vertical edges in class  $C_{1-1}$  (arrow in Fig. 5C). Rotating around y-axis (axis parallel to MTD) by  $-10^\circ$  shows a clear ring in the  $\beta$ -dynein head (arrow in Fig. 5D), similar to class  $C_{1-2}$  (arrow in Fig. 5B). This indicates  $\sim 10^\circ$  rotation of the  $\beta$ -dynein AAA ring relative to its tail. **When we plotted the first eigenvector for each of the particles in the order of subtomogram picking, from the proximal end to the distal end of each MTD, we see a separation of the particles indicating the presence of heterogeneity in the dataset (right panel in Fig. S7). Based on this we could reveal different distribution of ODAs on nine MTDs, although these groups were not separated clearly when only eigenvectors are plotted without positional information, indicating that this algorithm is not a tool to detect subclasses based only on structure.**

When we further classify class  $C_{1-1}$ , some subaverages show lack of ODAs ( $C_{21-1}$  in Fig. 6A), while  $C_{1-2}$  does not have such a subclass without ODA. Moreover, the class without ODA contained almost all the particles from within a doublet as expected. However, the number of particles in this class is slightly larger than the expected number which is 11% (one out of nine MTDs miss ODAs). Nevertheless, it is consistent with biochemical analysis proving lower ODA content in *pf23* mutant than *wt Chlamydomonas* (Yamamoto et al. in preparation).

*Classification of Chlamydomonas pf23 mutant dynein a and b:* Subsequently, we also performed unsupervised classification focusing on inner dyneins a and b. In the total average obtained after subtomogram alignment (Fig. 7A, class  $C_0$ ), there is a dim density at the locus of dynein a, but it is not clear if the density is from dynein. After classification of masked and 1D projected density voxels, we found 55 subtomograms categorized in a class without a clear density for dynein a or b (Fig. 7A, class  $C_{1-1}$ ), while the average of the other class (Fig. 7A, class  $C_{1-2}$ ) showed a clear density for dynein a (arrow in Fig. 7, class  $C_0$ ). The average of sub-class  $C_{1-2}$  shows a typical dynein conformation with a head, and a tail anchored on MTD. This also indicates that dynein a partially occupies this locus in the *pf23* mutant of *Chlamydomonas*. Interestingly, all the classes except  $C_{1-3}$  do not show a density for dynein b indicating that dynein b is relatively affected more when compared to dynein a. Although the class  $C_{1-3}$  shows a density at the position of dynein b, it appears to be morphologically different when compared to the structure in the reference.

In this work, we show that we have successfully used our approach to perform unsupervised classification of heterogeneous data to not only sort data based on presence or absence of a target of interest but also to detect structural polymorphism within the sample. The approach can be used for filamentous objects such as the eukaryotic flagella/cilia which are parallel to the tilt axis. In single-axis tomography, the area close to the tilt axis (which is the common line) in the Fourier space is oversampled. When the filament is parallel to the tilt axis, 1D projection of subtomograms corresponds to data along the tilt axis. Therefore, although the information is limited to one line, the most reliable data in the 3D Fourier space is used for classification. With the help of simulations we see that the approach is robust even at high levels of noise. When the filament is bent or curved, the orientations of the 1D projection with respect to the tilt axis are different between subtomograms. Although the 1D line in 3D Fourier space transformed from the 1D projection is involved in the XY plane, it is involved in the oversampled area only at low resolution. From our analysis we see that this low resolution information was enough to distinguish subtle structural differences even when the filament is 60 degrees rotated from the tilt axis. The proposed approach is independent of the missing wedge orientation and is suitable for initial detection of heterogeneity among molecules in filamentous objects, such as dyneins on the axoneme. However, the approach is sensitive to the extent of missing wedge (more than 30 degrees) and error in alignment (more than 2° rotations and 2pixel shifts). In such cases, it was only possible to classify presence or absence of a target of interest but not small conformational changes. As it is sensitive to error in alignment other approaches which simultaneously align and classify may prove advantageous (Scheres et al., 2008; Stölken et al., 2011) when the error in alignment is more. As some structural information is missing in the 1D projection, there will be structures which cannot be classified correctly with this approach. Such structural variation, for example, motion of density in the x- or z-direction (i.e. direction perpendicular to the filament axis), has to be detected by other available methods. Nevertheless, protein domain movements are often three dimensional and we think it would still be possible to classify the subtle conformational changes with this approach. **Although we could show in simulations a criteria to detect the presence of heterogeneity, this was not easily applicable to real data and hence we could not propose an algorithm for the detection of heterogeneity. This could probably be improved by using data at higher S/N.**

We think that using other well-known methods, for example, PCA using constrained cross-correlation or ML-based approaches will yield similar results. However, one of the advantages of this method when compared to the other methods is that it's relatively simple and fast. Also, this method enables collection of tomographic datasets especially for filamentous objects with target molecules periodically placed on it with relatively low dose, since periodicity helps subtomogram alignment while classification is challenging. This is because in tomograms of periodic objects with extremely low dose, we can align subtomograms, even if we cannot see the feature of subtomograms. We also show using a simulation that it is not limited to the molecules following helical symmetry and one can for example classify particles in closed or open state on a sheet or membrane provided they are in the same orientation along the axis i.e. having the same in-plane rotation. When they have random in-plane orientations, sub-averages must be reclassified to find identical structures with in-plane rotation using other available methods. A combination of the proposed method and other available methods can greatly help in addressing the heterogeneity in the data in a fast and robust manner thereby subsequently aiding in obtaining subtomogram averages with improved S/N and resolution.

Availability and Implementation: The tools for preparing the data for classification have been written in Perl. Classification has been carried out using Spider scripts. All the tools and scripts are available upon request.

## Acknowledgement

This work was supported by Swiss National Science Foundation (NF31003A\_144035/1) and National Center of Competence of Biomedical Imaging to TI, Fellowship from Japan Society of the Promotion of Science and Young Researchers Exchange Programme between Japan and Switzerland (Japanese-Swiss Science and Technology Programme) to RY.

## References

- Bui, K.H., Ishikawa, T., 2013. 3D structural analysis of flagella/cilia by cryo-electron tomography. *Methods Enzymol.* 524, 305–323. doi:10.1016/B978-0-12-397945-2.00017-2
- Bui, K.H., Yagi, T., Yamamoto, R., Kamiya, R., Ishikawa, T., 2012. Polarity and asymmetry in the arrangement of dynein and related structures in the Chlamydomonas axoneme. *J. Cell Biol.* 198, 913–925. doi:10.1083/jcb.201201120
- Crowther, R., Derosier, D., Klug, A., 1970. Reconstruction of 3 Dimensional Structure from Projections and Its Application to Electron Microscopy. *Proc. R. Soc. Lond. Ser. -Math. Phys. Sci.* 317, 319-. doi:10.1098/rspa.1970.0119
- Förster, F., Pruggnaller, S., Seybert, A., Frangakis, A.S., 2008. Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* 161, 276–286. doi:10.1016/j.jsb.2007.07.006
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A., 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116, 190–199. doi:10.1006/jsbi.1996.0030
- Hoops, H.J., Witman, G.B., 1983. Outer doublet heterogeneity reveals structural polarity related to beat direction in Chlamydomonas flagella. *J. Cell Biol.* 97, 902–908.
- Ishikawa, T., Wakabayashi, T., 1995. Proposal of alignment-independent classification of electron microscopic images with helical symmetry and its application to reconstituted thin filaments os skeletal muscle. *Ultramicroscopy* 57, 91–101.
- Lin, J., Okada, K., Raytchev, M., Smith, M.C., Nicastro, D., 2014. Structural mechanism of the dynein power stroke. *Nat. Cell Biol.* doi:10.1038/ncb2939
- Maheshwari, A., Obbineni, J.M., Bui, K.H., Shibata, K., Toyoshima, Y.Y., Ishikawa, T., 2015.  $\alpha$ - and  $\beta$ -Tubulin Lattice of the Axonemal Microtubule Doublet and Binding Proteins Revealed by Single Particle Cryo-Electron Microscopy and Tomography. *Struct. Lond. Engl.* 1993 23, 1584–1595. doi:10.1016/j.str.2015.06.017
- Mastrorarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* 152, 36–51. doi:10.1016/j.jsb.2005.07.007
- Mastrorarde, D.N., 1997. Dual-axis tomography: an approach with alignment methods that preserve resolution. *J. Struct. Biol.* 120, 343–352. doi:10.1006/jsbi.1997.3919
- Movassagh, T., Bui, K.H., Sakakibara, H., Oiwa, K., Ishikawa, T., 2010. Nucleotide-induced global conformational changes of flagellar dynein arms revealed by in situ analysis. *Nat. Struct. Mol. Biol.* 17, 761–767. doi:10.1038/nsmb.1832
- Pigino, G., Maheshwari, A., Bui, K.H., Shingyoji, C., Kamimura, S., Ishikawa, T., 2012. Comparative structural analysis of eukaryotic flagella and cilia from Chlamydomonas, Tetrahymena, and sea urchins. *J. Struct. Biol.* 178, 199–206. doi:10.1016/j.jsb.2012.02.012
- Schatz, M., van Heel, M., 1990. Invariant classification of molecular views in electron micrographs. *Ultramicroscopy* 32, 255–264.

- Scheres, S.H.W., Melero, R., Valle, M., Carazo, J.-M., 2009. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Struct. Lond. Engl.* 1993 17, 1563–1572. doi:10.1016/j.str.2009.10.009
- Scheres, S.H.W., Núñez-Ramírez, R., Sorzano, C.O.S., Carazo, J.M., Marabini, R., 2008. Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* 3, 977–990. doi:10.1038/nprot.2008.62
- Stölken, M., Beck, F., Haller, T., Hegerl, R., Gutsche, I., Carazo, J.-M., Baumeister, W., Scheres, S.H.W., Nickell, S., 2011. Maximum likelihood based classification of electron tomographic data. *J. Struct. Biol.* 173, 77–85. doi:10.1016/j.jsb.2010.08.005
- Ueno, H., Bui, K.H., Ishikawa, T., Imai, Y., Yamaguchi, T., Ishikawa, T., 2014. Structure of dimeric axonemal dynein in cilia suggests an alternative mechanism of force generation. *Cytoskelet. Hoboken NJ.* doi:10.1002/cm.21180
- Witman, G.B., 1986. Isolation of *Chlamydomonas* flagella and flagellar axonemes. *Methods Enzymol.* 134, 280–290.

## Figure captions

### Figure 1: Scheme of classification

(a) Subtomograms (shown in yellow dashed boxes) are extracted and aligned. The  $y'$ -axis is the axis of the filament and the density map is projected on the  $y'$ -axis. The  $y$ -axis is the tilt axis. (b) Tomographic sampling scheme in 3D Fourier space. Each micrograph corresponds to a central section with the thickness of  $2/d$  ( $d$ : dimension of the object). (c) Projection on the XY plane showing the area close to the Y-axis which is oversampled. (d)  $y'$ -axes juxtaposed on the projection on the XY plane in (c). The orientations of the four  $y'$ -axes (shown as red dotted lines) corresponds to the four subtomograms extracted in (a).

### Figure 2: Classification of subtomograms from *Tetrahymena wt* cilia based on ODA.

(A) Longitudinal sections (parallel to the microtubule) at three different heights including  $\alpha$ -,  $\beta$ -,  $\gamma$ -dyneins of averages of four sub-classes are shown. "C<sub>1-1</sub>" and "C<sub>1-2</sub>" denotes the first and second class obtained during the first round of classification, respectively. Number of particles in each class is indicated. Mask used for classification is also shown. (B) Left: Histogram showing percentage of eigenvalue variance accounted for by each factor. Factors used for classification are enclosed. Cutoff was set where the percentage decreases significantly. Right: Dendrogram for this classification.

### Figure 3: Classification of subtomograms from *Chlamydomonas wt* flagella based on ODA.

(A) Longitudinal sections (parallel to the microtubule) at three different heights including  $\alpha$ -,  $\beta$ -,  $\gamma$ -dyneins of averages of six sub-classes are shown. C<sub>1-3</sub> is the class which contains no ODA. Number of particles in each class is indicated. The mask is also shown. (B) Left: Histogram showing percentage of eigenvalue variance accounted for by each factor. The first two eigenfactors are considered for classification. Right: Dendrogram. (C) Plot showing the distribution of subtomograms in different sub-classes. C<sub>1-3</sub> (black squares) is localized at the fifth doublet.

### Figure 4: Classification of subtomograms from *Chlamydomonas pf23* mutant flagella based on ODA.

(A) Longitudinal sections (parallel to the microtubule) at three different heights including  $\alpha$ -,  $\beta$ -,  $\gamma$ -dyneins of averages of two sub-classes are shown. C<sub>1-1</sub> and C<sub>1-2</sub> show slight difference in dynein structure which is shown in detail in Fig. 5. Number of particles in each class is indicated. The mask is also shown. (B) Left: Histogram showing percentage of eigenvalue variance accounted for by each factor. Factors used for classification are enclosed. Right: Dendrogram.

### Figure 5: Longitudinal sections from classified and averaged subtomograms of flagella from *Chlamydomonas pf23*.

(a) Subaverage of C<sub>1-1</sub> as defined in Fig.4 indicating the AAA-ring and tail of the  $\beta$ -dynein (in red). (b) Subaverage of C<sub>1-1</sub>. The AAA-ring and the tail of beta-dynein are indicated by arrows and arrowheads, respectively (c) Subaverage of C<sub>1-2</sub> seen at the same view angle as (b). (d) Subaverage of C<sub>1-1</sub>, rotated -10° around the  $y$ -axis. The AAA-ring seems similar to (b), indicating that the AAA-ring of beta-dynein in C<sub>1-1</sub> is ~-10° rotated with respect to that in C<sub>1-2</sub>.

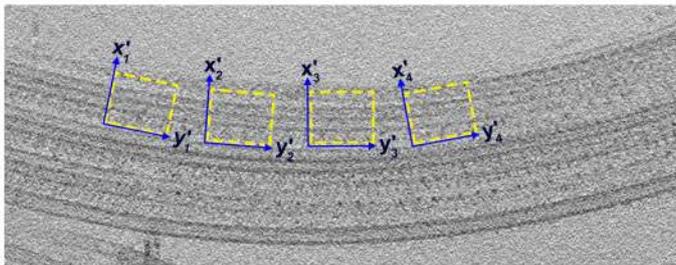
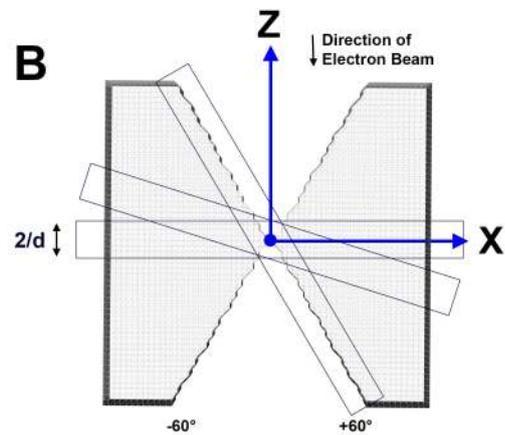
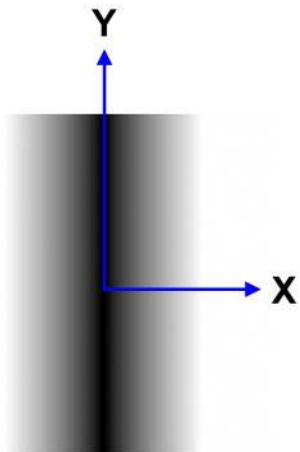
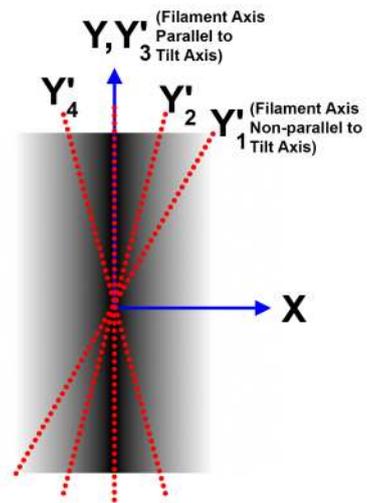
### Figure 6: Further classification of subtomograms from *Chlamydomonas pf23* mutant flagella based on ODA.

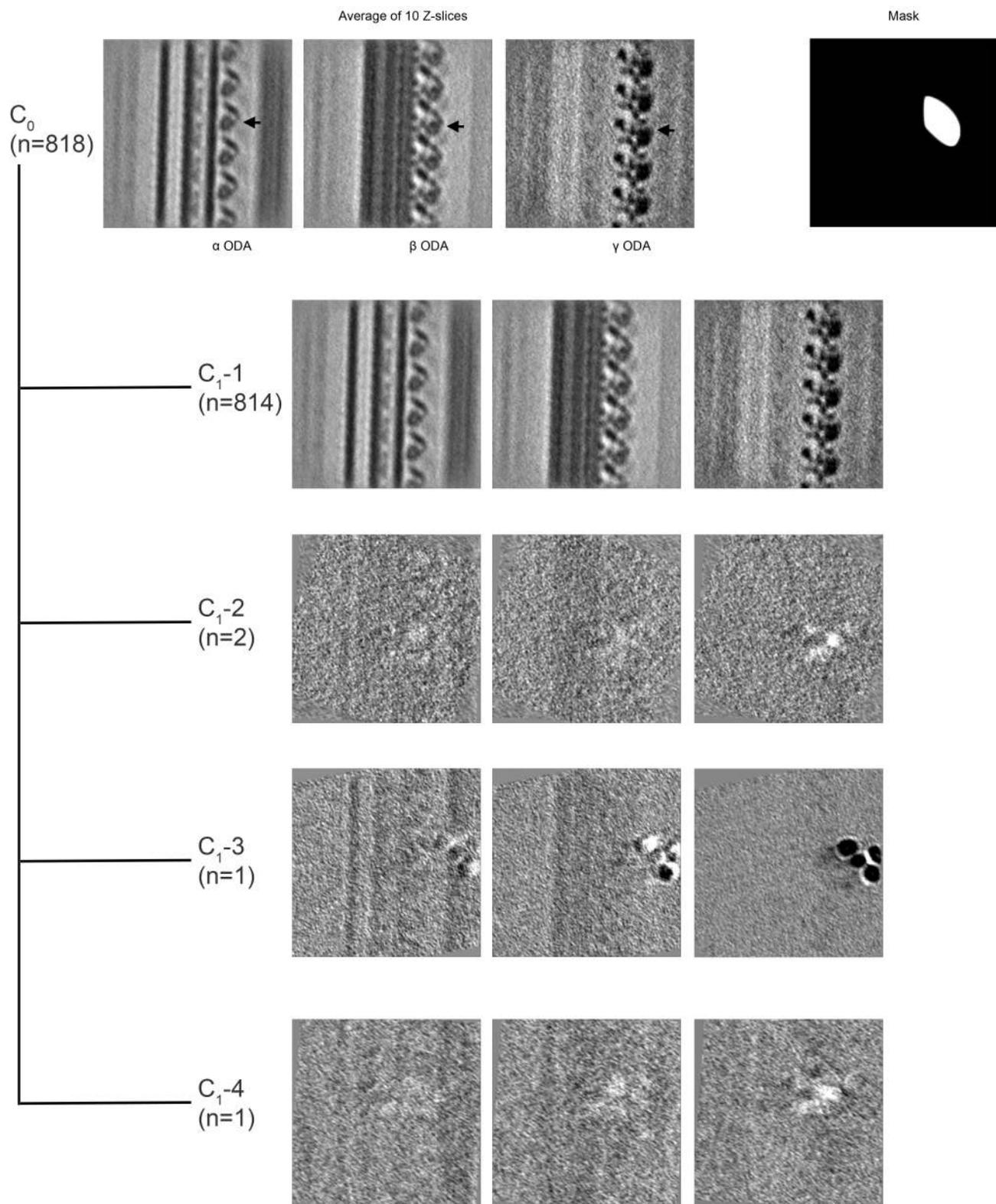
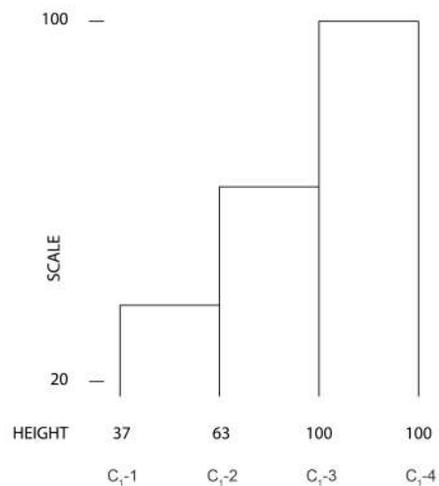
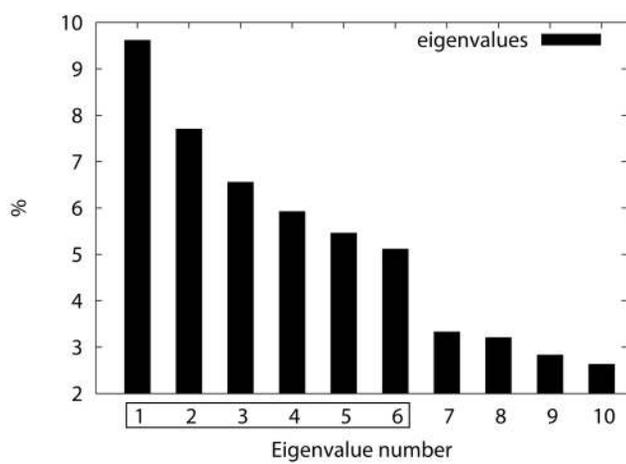
(A) 335 subtomograms belonging to C<sub>1-1</sub> of the first step of classification (Fig.4) were classified into three subclasses. C<sub>21-1</sub> is dominated with particles having no ODA. Number of particles in each class is indicated. Mask used for classification is shown. (B) Left: Histogram showing percentage of eigenvalue variance

accounted for by each factor and Right: the dendrogram. (C, D) Classification of 218 subtomograms from C<sub>1</sub>-2 into two classes, the corresponding eigenfactor plot and the dendrogram. Both classes showed presence of ODAs. Number of particles in each class is indicated. Mask used for classification is also shown.

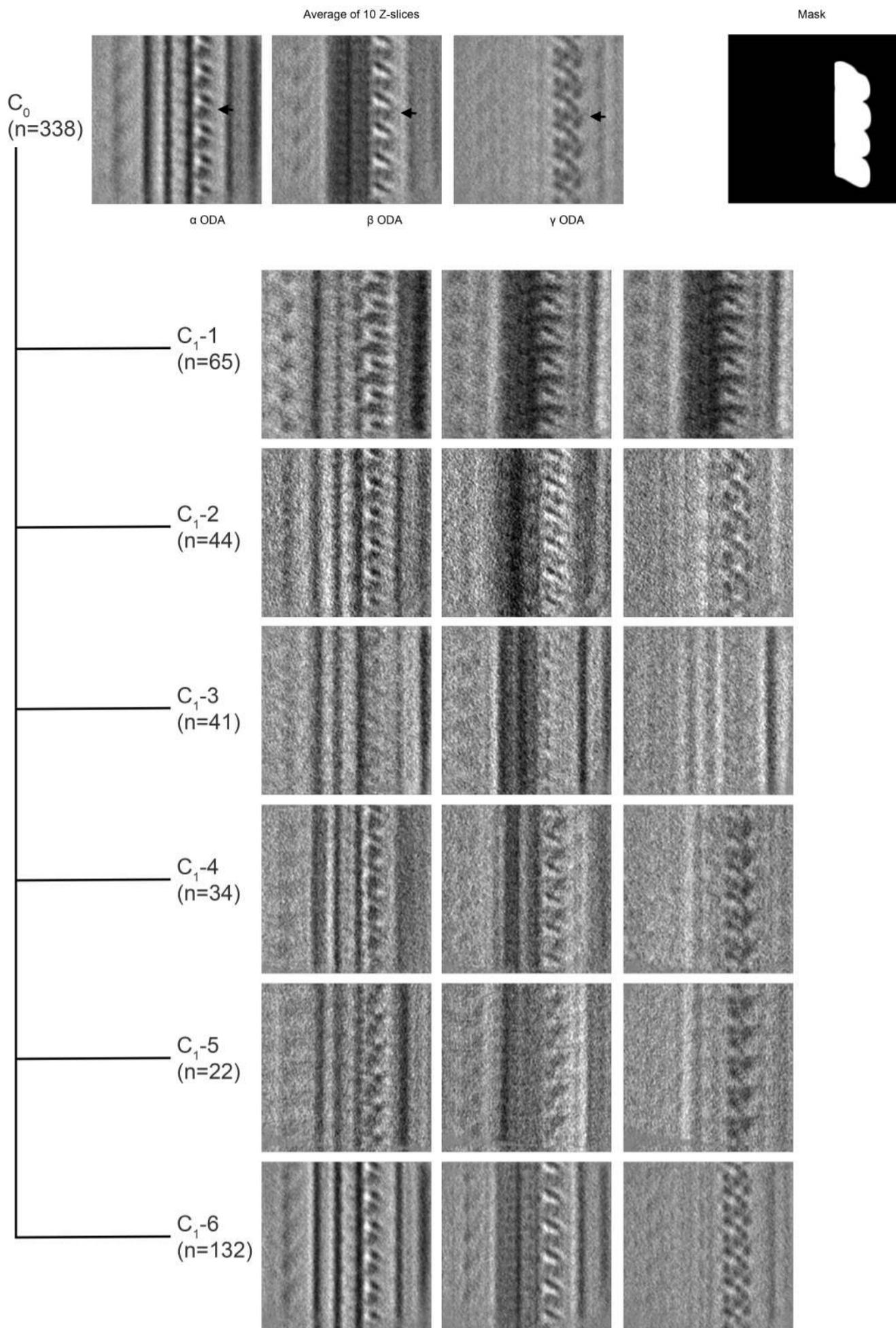
**Figure 7:** Classification of *Chlamydomonas pf23* flagella based on dynein a and b.

(A) Longitudinal section of averages of five sub-classes are shown. Positions of dyneins a and b are indicated in the center top panel. Number of particles in each class is indicated. Mask used for classification is shown. The position of dynein a is also indicated by an arrow in the total average. C<sub>1</sub>-2 subclass shows clear density of dynein a. Dynein b seems to be relatively more affected in this mutant and can be only seen in C<sub>1</sub>-3 although it is morphologically different when compared to the reference. (B) Eigenfactor plot and dendrogram. The first five eigenfactors were used for classification.

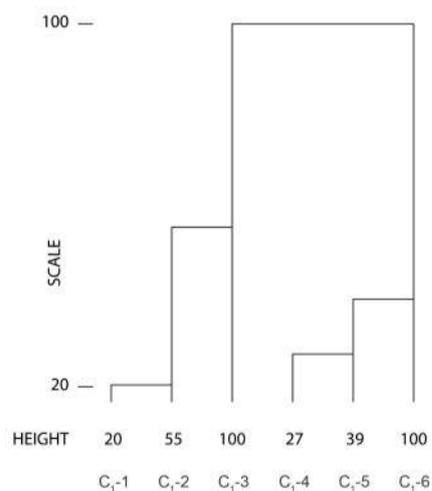
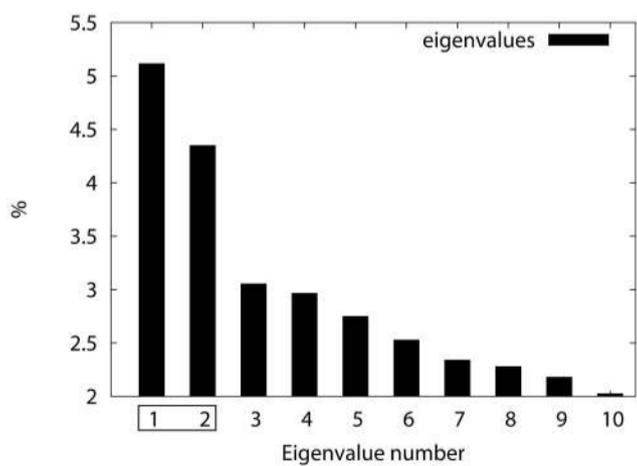
**A****B****C****D**

**A****B**

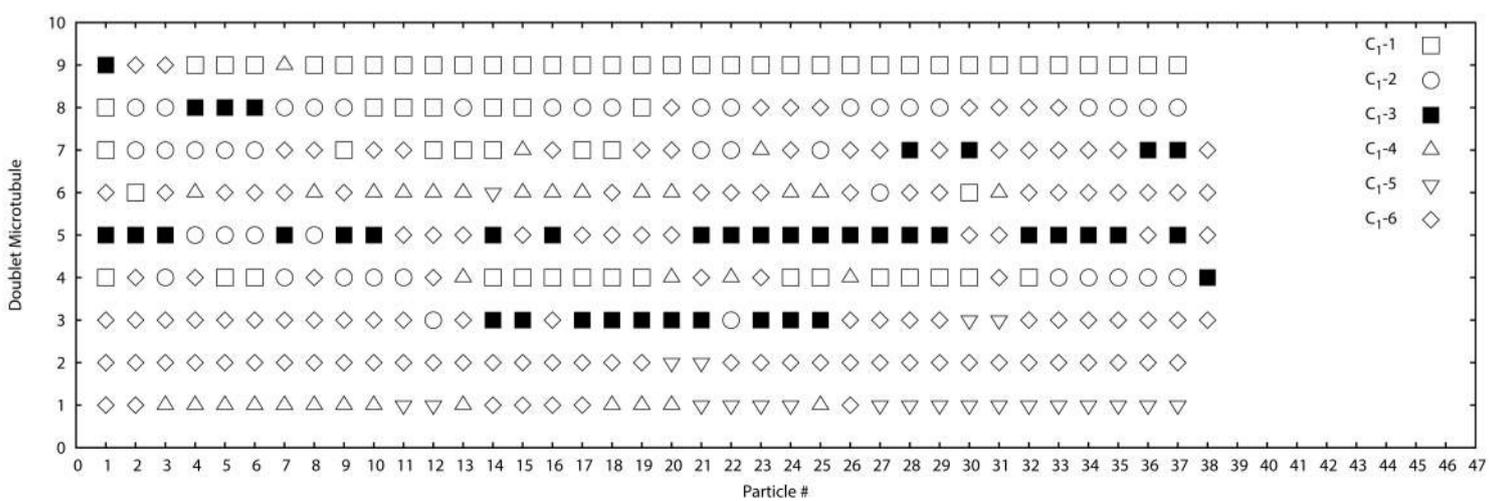
**A**



**B**



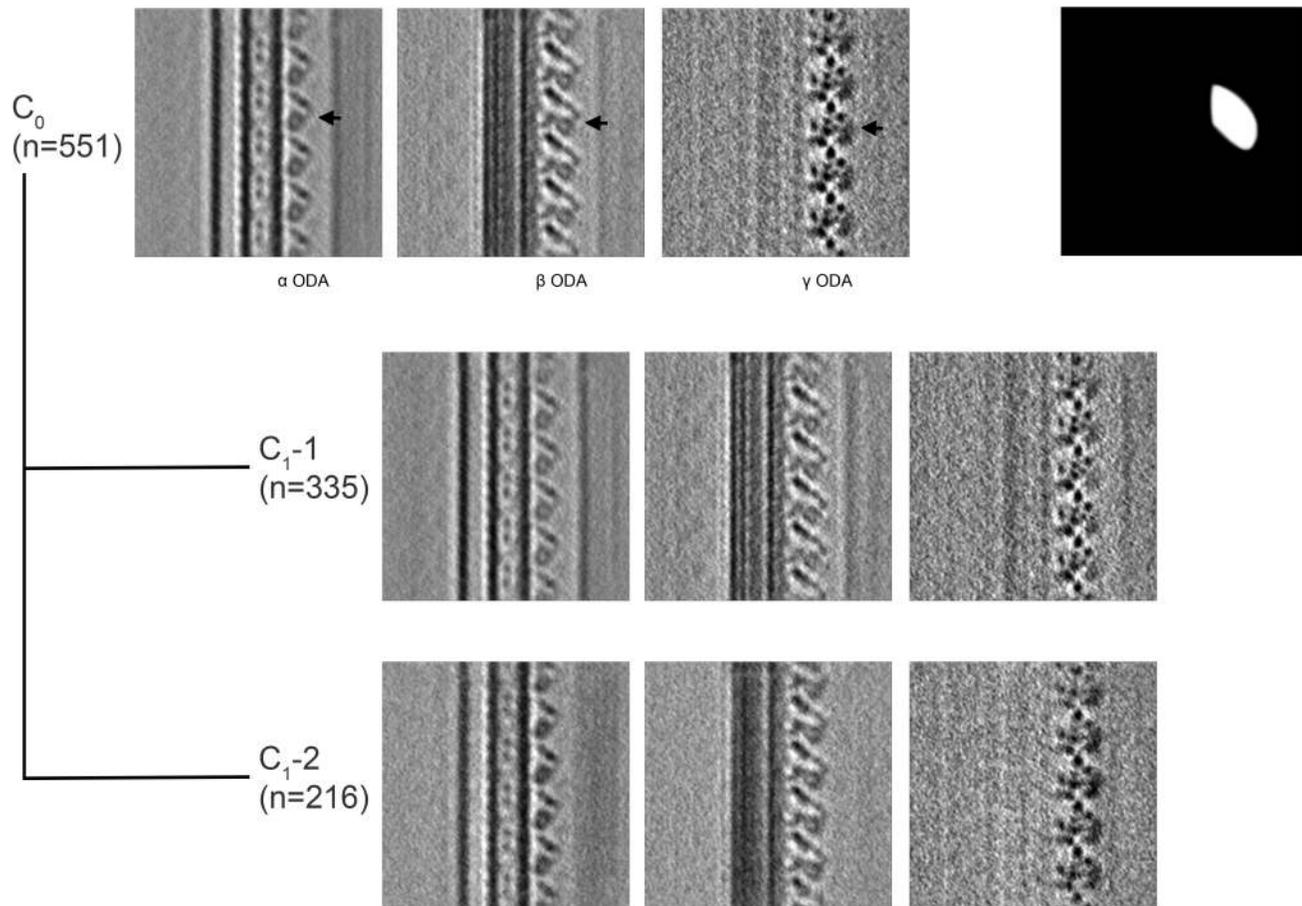
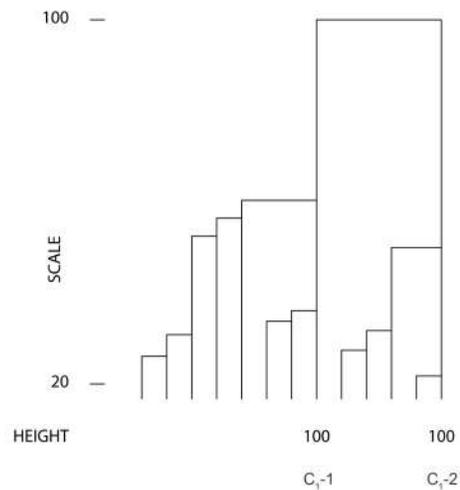
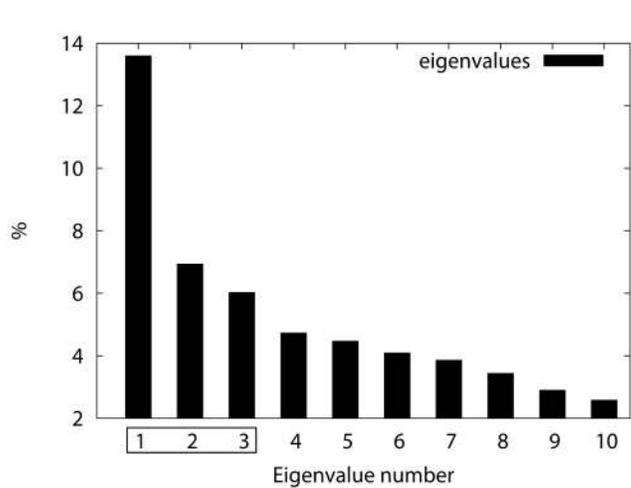
**C**

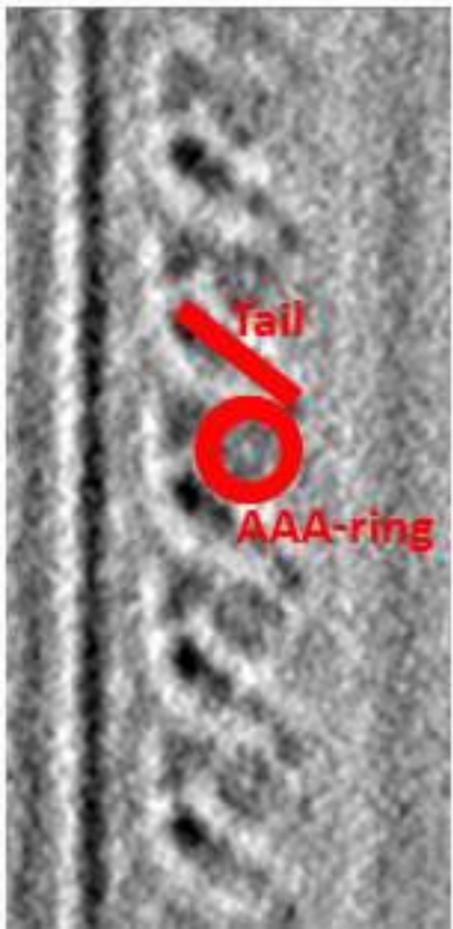
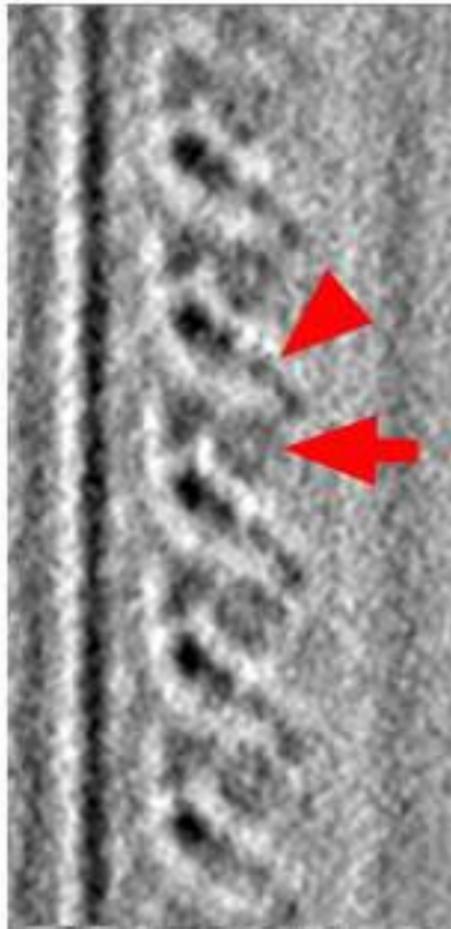
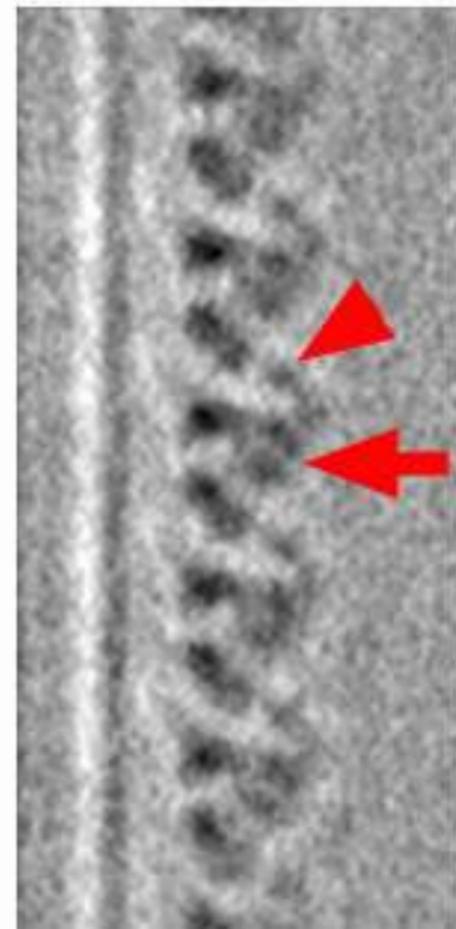
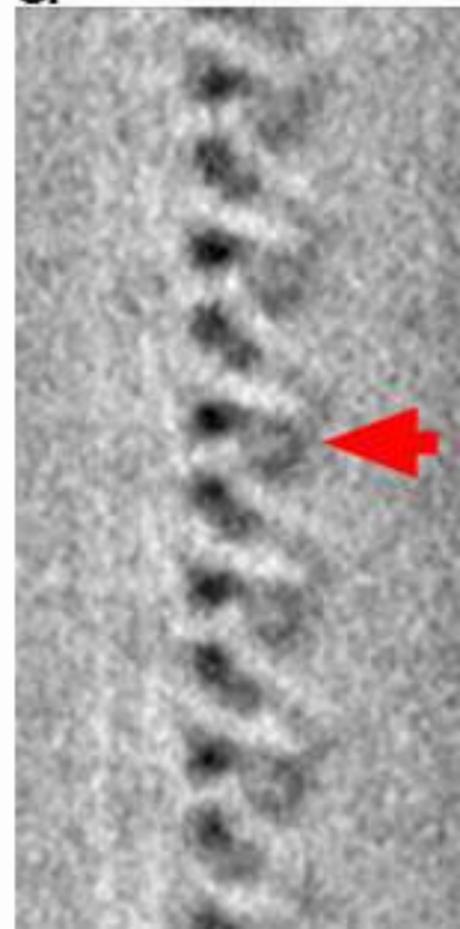


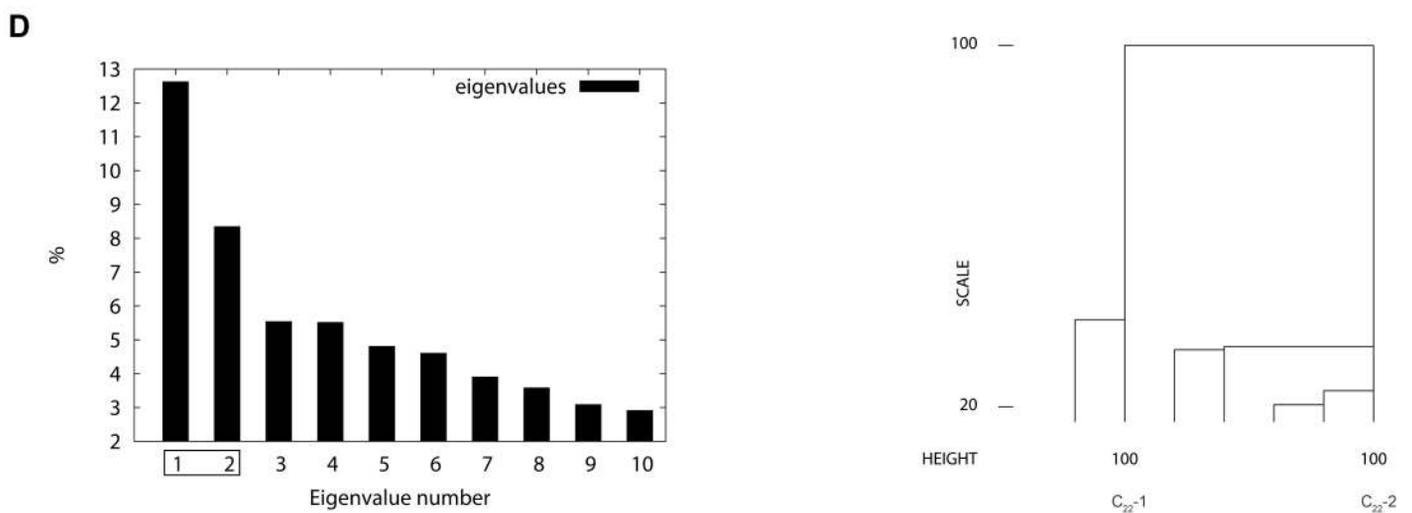
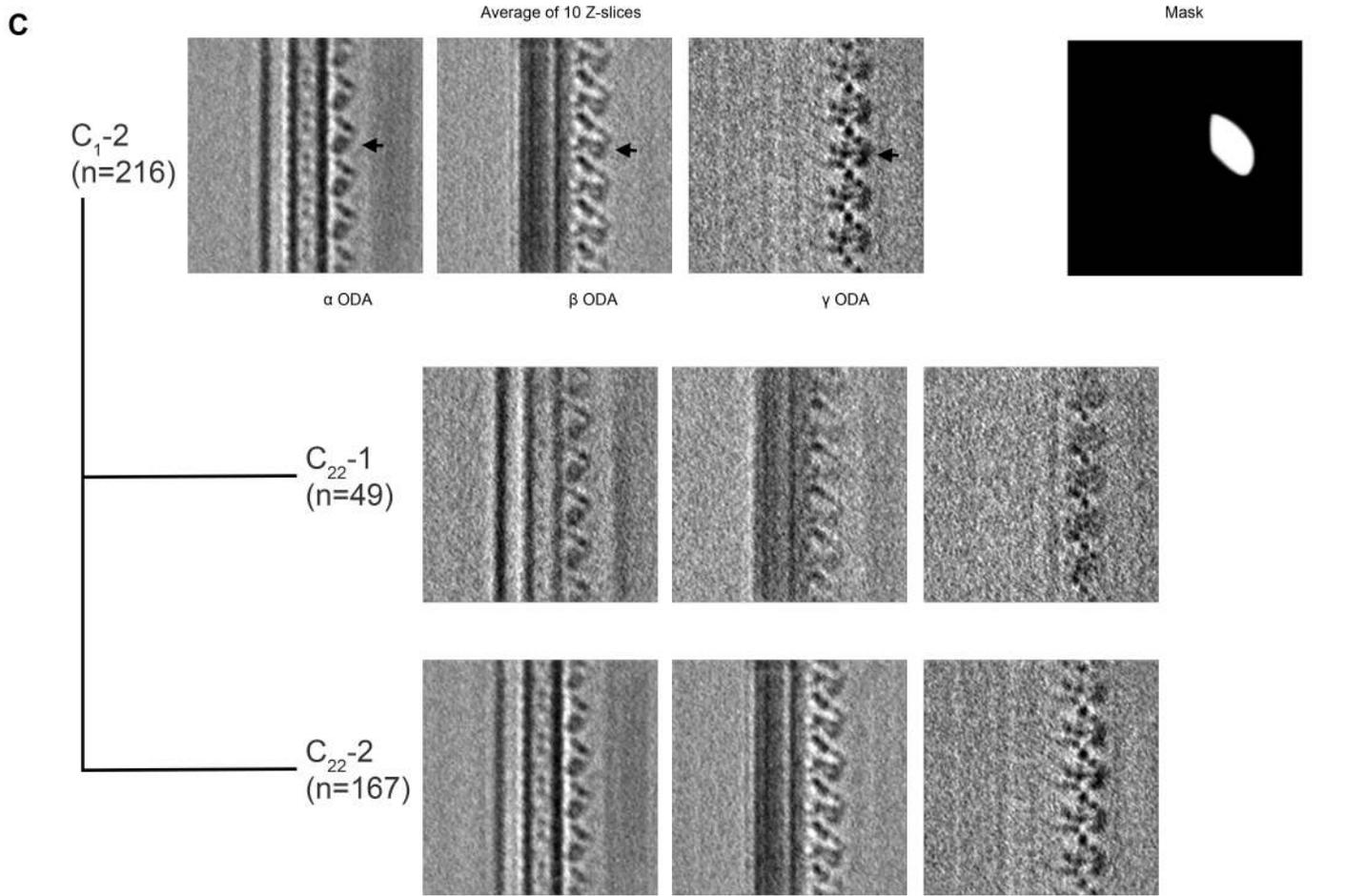
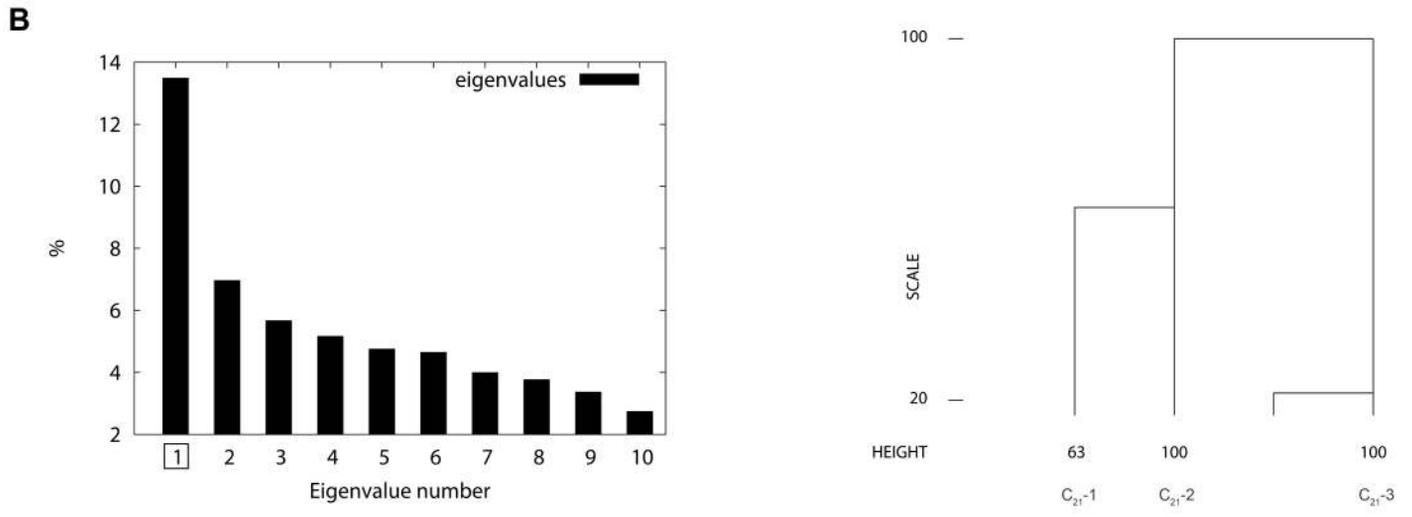
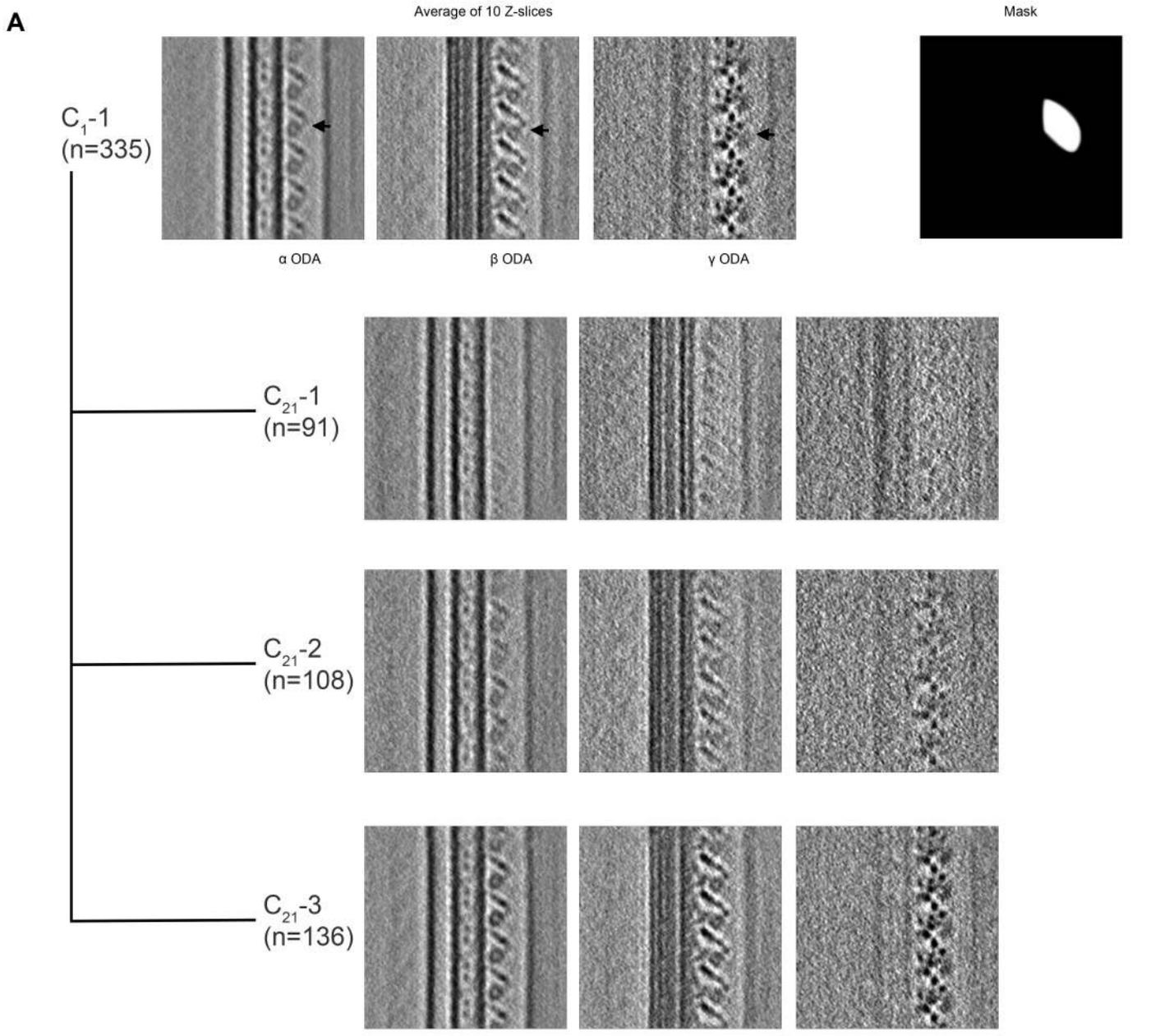
**A**

Average of 10 Z-slices

Mask

**B**

**a****b****c****d**



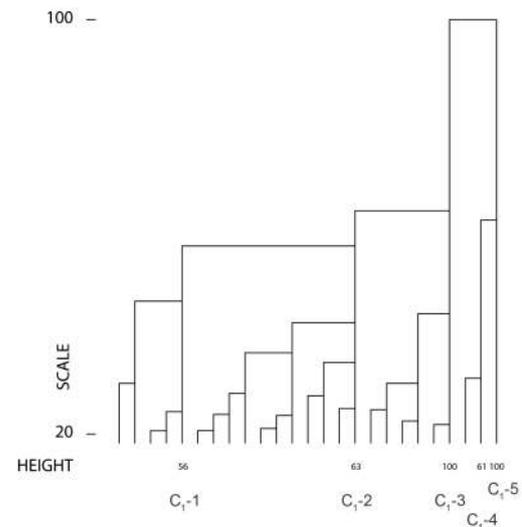
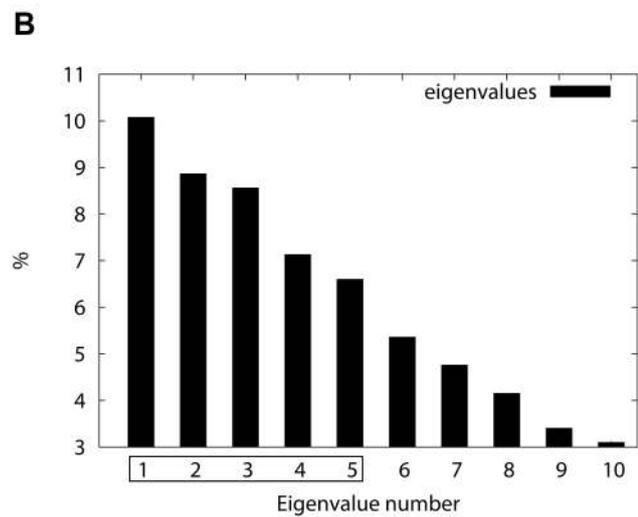
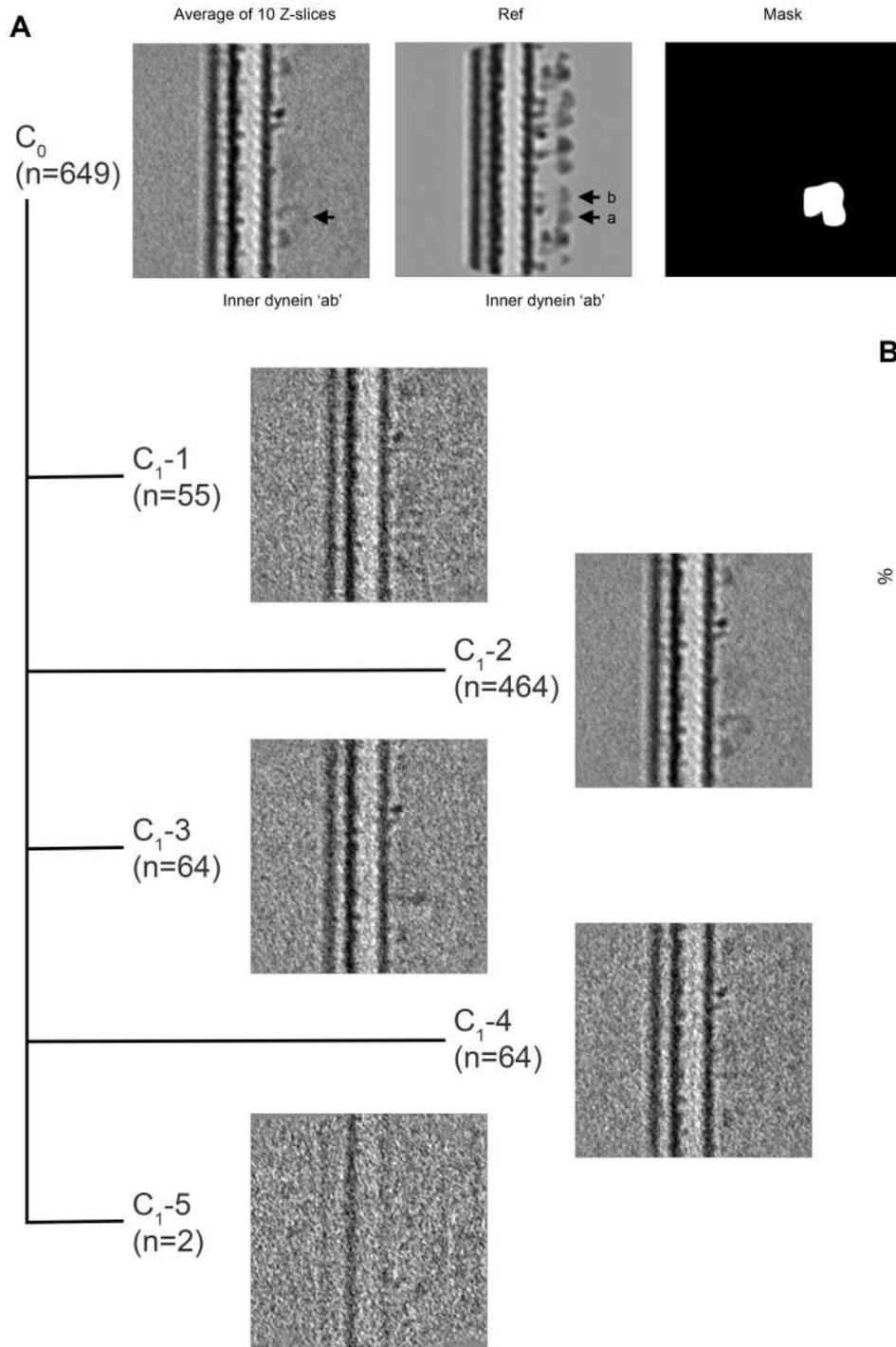


Table 1: Summary of systems and parameters for model simulated data

Parameters Simulation	$n_{bw}$	w	$n_{aw}$	$\theta_{at}$	e	$C_{err}$	
1	10	30°	5	0°	0°, 0pix	0	No error in classification
2	10	30°	10	0°	0°, 0pix	0	No error in classification
3	10	30°	20	0°	0°, 0pix	12	Only presence and absence of dynein classified
4	10	40°	10	0°	0°, 0pix	10	Only presence and absence of dynein classified
5	10	50°	10	0°	0°, 0pix	12.7	Only presence and absence of dynein classified
6	10	30°	10	0°, ±15°	0°, 0pix	0	No error in classification
7	10	30°	10	0°, ±15°, ±30°	0°, 0pix	0	No error in classification
8	10	30°	10	0°, ±15°, ±30°, ±60°	0°, 0pix	0	No error in classification
9	10	30°	10	0°	±1°, ±1pix	0	No error in classification
10	10	30°	10	0°	±2°, ±2pix	3.3	Only presence and absence of dynein classified
11	10	30°	10	0°	±3°, ±3pix	4	Only presence and absence of dynein classified

$n_{bw}$  – noise before wedge, w – wedge,  $n_{aw}$  – noise after wedge,  $\theta_{at}$  – difference between axonemal axis and tilt axis, e – error in alignment,  $C_{err}$  – percentage error in classification given by the formula  $\frac{1}{n} \sum_i^n \frac{FP_i + FN_i}{T} * 100$  where n is the number of classes, FP is the number of false positives, FN is the number of false negatives, T is total number of particles.