OPUS-TOMO: Deep Learning Framework for Structural Heterogeneity Analysis in Cryo-electron Tomography

Zhenwei Luo[1,2,3], Qinghua Wang[4], & Jianpeng Ma[1,2,3*]

[1]Multiscale Research Institute of Complex Systems, Fudan University, Shanghai, 200433, China; [2] Zhangjiang Fudan International Innovation Center, Fudan University, Shanghai, 201210, China; [3] Shanghai AI Laboratory, Shanghai, 200030, China; [4]Center for Biomolecular Innovation, Harcam Biomedicines, Shanghai, 201203, China

[*]**Correspondence**: jpma@fudan.edu.cn

Keywords: Convolutional neural network, neural volumes, structural heterogeneity, cryo-electron tomography, molecular sociology

Abstract

Structural heterogeneity due to the dynamic nature of macromolecules *in situ* presents a significant challenge to structural determination by cryo-electron tomography (cryo-ET). In this paper, we present OPUS-TOMO, a deep learning framework for analyzing structural heterogeneity in cryo-ET data. The method adopts a convolutional Encoder-Decoder architecture that adeptly maps real-space subtomograms onto a smooth low-dimensional latent space, which captures the complete landscape of compositional and conformational variations of macromolecules in cryo-ET data. OPUS-TOMO also incorporates algorithms, including a per-particle 3D CTF model and a pose correction network, specifically for handling cryo-ET data. Applications of OPUS-TOMO to multiple real cryo-ET datasets confirm the outstanding capacities of the new method in characterizing structural heterogeneity. The software is available at https://github.com/alncat/opusTOMO.

Introduction

The recent advancements in single-tilt axis holder[1], sample preparation through cryo-focused ion beam (cryo-FIB) instruments[2,3], and the development of image processing algorithms[4-8] have significantly enhanced the capabilities of cryo-electron tomography (cryo-ET) in determining cellular structures at sub-nanometer resolution[5]. In contrast to cryo-electron microscopy (cryo-EM), cryo-ET does not require the isolation of the macromolecules of interest from their native cellular environment[9,10]. Hence, cryo-ET can capture three-dimensional (3D) snapshots of "molecular sociology"[11], and provide valuable insights into the functional states of macromolecules *in situ*[12]. However, structural determination by cryo-ET still faces many challenges.

A major challenge of cryo-ET is the systematic errors arising from the corruption of 3D structural information in cryo-ET image acquisition. In order to prevent radiation damage, minimal doses are used in cryo-ET data collection, leading to subtomograms with a low signal-to-noise ratio (SNR)[9,10,13]. Furthermore, tomograms also have "missing wedges due to the limited rotational angles of specimen holders[10,13]. Moreover, there are substantial defocus variations across the recorded images caused by the tilting of specimen holders and the thickness of specimens[9,13]. Generally, the reconstruction of a 3D structure using cryo-ET data begins with particle localization that selects similar copies of a 3D structure from tomograms. Conventionally, this is done in two ways, template-matching methods or template-free methods. The template-matching method picks particles according to the correlation of structures in cryo-ET tomograms to a known template[14,15]. However, due to the low SNR and "missing wedge" of the cryo-ET data, the template-matching results often contain a large number of false positives, especially for sparsely populated species. On the other hand, the conventional template-free

algorithms are largely limited to picking particles of large cellular structures[16]. Recently, deep learning-based methods have been developed to search structures of interest within cryo-ET tomograms and achieved higher accuracies. For instance, DeePiCt[17] sets an example to identify different structures in cryo-ET data using U-Nets[18] that was trained on a set of expert-annotated tomograms.

Macromolecules undergo complex interactions with each other in the cellular environment, which result in large-scale compositional and conformational dynamics[10,11,19]. These dynamic features are preserved in cryo-ET data. Hence, another major challenge for cryo-ET structural determination[20,21] is to resolve structural heterogeneity, which is crucial for both improving the qualities of subtomogram averages[20] and reconstructing the compositional and conformational dynamics that underpin biological processes. Traditional approaches for characterizing the structural heterogeneity is 3D classification[22], which has been implemented in various frameworks for cryo-ET data analysis such as RELION[9], nextPYP[6], and EmClarity[8], the latter implementing unsupervised machine learning algorithms, such as principal component analysis (PCA) and KMeans clustering, to identify compositionally different macromolecules. However, these traditional approaches can only resolve a few discrete states and very time consuming. Real biological processes happen by continuous compositional and conformational changes of macromolecules. Therefore, a few discrete functional states are insufficient to cover the continuous dynamic processes. Recently, in the cryo-EM community, a number of algorithms leveraging powerful neural networks to reconstruct continuous structural changes have been developed, such as cryoDRGN[23] and OPUS-DSD[24], and have demonstrated great successes in characterizing the structural heterogeneity and reconstructing continuous dynamics.

Continuing on this line of work, TomoDRGN[25] leveraged the neural network architecture proposed in cryoDRGN to resolve structural heterogeneity in subtomograms from cryo-ET.

In this paper, we report the development of a convolutional neural network to effectively address structural heterogeneities in cryo-ET data. Our new method, termed as OPUS-TOMO, is an open-source software that characterizes compositional and conformational variations in subtomograms regardless of particle localization methods. By testing OPUS-TOMO across a range of publicly accessible cryo-ET datasets, encompassing both *in vitro* and *in vivo* samples, we demonstrated the ability of OPUS-TOMO to unravel structural heterogeneity and reconstruct dynamic cellular processes in a highly effective way.

Results

Neural Architecture of OPUS-TOMO

OPUS-TOMO starts with using the subtomogram averaging results from a conventional cryo-ET reconstruction as the input to resolve structural heterogeneity, which ensures that 3D structures reconstructed for each subtomogram are pre-aligned to a common orientation. Then, the reconstruction of the 3D structure for a subtomogram is carried out as following. The 3D subtomogram is first aligned to the canonical orientation of consensus subtomogram average by inversely rotating according to its pose estimated by subtomogram averaging. The Encoder takes the aligned subtomogram as the input, and estimates the distributions of the latent code z and the pose correction latent code z_c. Specifically, it estimates the means $z \in \mathbb{R}^n$ and standard variations $\sigma \in \mathbb{R}^n$ for the latent codes[26] (**Fig.1a**). Next, the latent codes are sampled according to the estimated distributions and supplied to the corresponding components. The sampled latent code is supplied to the Decoder to reconstruct a 3D volume $V(\boldsymbol{x})$ (**Fig.1a**), which is represented

by a discrete grid of voxels. Meanwhile, the sampled pose correction latent code is supplied to the Pose Corrector to estimate pose correction (**Fig.1a**), which mainly serves to correct errors in pose parameters from subtomogram averaging. The pose assignment errors are systematic for subtomogram averaging algorithm since it aligns subtomograms with different structures to a common 3D structure. Finally, the Decoder leverages spatial transformer[27] to produce a 3D reconstruction by applying the rigid transformation with the corrected pose onto the 3D volume $V(\boldsymbol{x})$ (**Fig.1a**). The 3D reconstruction is transformed into a subtomogram with its corresponding 3D-contrast transfer function (3DCTF) following the differentiable model defined in the subsection of **Subtomogram Formation Model** in **Methods**.

The neural architecture of OPUS-TOMO can be trained end-to-end by reconstructing each input subtomogram and minimizing the squared reconstruction errors between the input and the output (**Fig.1a**). During training, to encourage the smoothness of latent spaces, the latent codes and the pose correction latent codes are regularized by the structural disentanglement prior[24] from OPUS-DSD and Kullback-Leibler (KL) divergence with respect to (w.r.t) standard Gaussian[26,28] (**Fig.1a**). The architectures of the Encoder and Decoder are the same as in OPUS-DSD[24]. The Pose Corrector is relatively simple and consists of a series of multi-layer perceptron (MLPs) and nonlinearities (**Fig.1b**). The 4-dimensional pose correction latent code is transformed into a 512-dimensional representation by an MLP followed by LeakyReLU nonlinearity. The 512-dimensional representation is then projected to a 6-dimensional vector by another MLP. This 6-dimensional vector defines the deviation between the ground-truth pose parameters of a subtomogram w.r.t its ground-truth conformation and its estimated pose parameters w.r.t the consensus subtomogram average. The first three dimensions of the 6-dimensional output are used to construct the quaternion that represents the rotational correction

for a subtomogram, while the last three dimensions of the 6-dimensional output represent the translational correction. The construction of pose correction parameters is detailed in the **Training Objective** subsection in **Methods**.

Reconstruction of 80S Ribosome Structures from Tomograms of Wild-type *Schizosaccharomyces pombe* (*S. pombe*)

We tested the performance of OPUS-TOMO in structural heterogeneity analysis on *in vivo* 80S ribosomes in tomograms acquired from cryo-FIB-milled lamellae of wild-type *S. pombe*[17]. The tomograms were first reconstructed using the tilt series and their corresponding parameters of tilt series alignment deposited in EMPIAR-10988[17]. Then ribosome subtomograms were extracted from tomograms using the particle coordinates picked by DeePiCt and deposited in EMPIAR-10988[17]. Structural heterogeneity analysis was performed using OPUS-TOMO with the pose parameters determined by subtomogram averaging using RELION 3.0.8[9].

Clustering the latent space learned by OPUS-TOMO using KMeans[29] results in 24 well-separated clusters (**Extended Data Fig.1a**). The 3D structures for these classes are reconstructed by supplying the classes' centroids into the trained Decoder (**Extended Data Fig.1b**). Within these 24 classes, Class 17 shows non-ribosome densities, Class 0 presents only the 60S subunit, Class 10 shows 60S subunit with a partial 40S subunit, Classes 2, 12 and 14 show identifiable additional cofactors bound to 80S ribosomes (**Fig.2a**), and many classes correspond to transient states of the translation elongation cycle performed by 80S ribosome, such as Classes 1, 3, 5, 12 and 20 (**Fig.2b**). In particular, Class 2 from OPUS-TOMO reveals the ES27L RNA, the large subunit rRNA expansion segment, at the exit tunnel, as seen in the human Ebp1-ribosome

structure in which Ebp1 recruits ES27L to the tunnel exit (EM Data Bank (EMDB) accession code: EMD-10608[30]) (**Fig.2a, Left panel; Extended Data Fig.2a**). Rigid body fitting of the ES27L model based on a 6.1 Å cryo-EM map of *Saccharomyces cerevisiae* (*S. cerevisiae*) translating 80S ribosome (Protein Data Bank (PDB) accession code: 3IZD[31]) into Class 2 map achieves a correlation coefficient (CC) of 0.73 (**Extended Data Fig.2a, Middle panel**). In addition, Class 12 exhibits an additional density that binds with the head of 40S subunit (**Fig.2a, Middle panel**). This additional density overlaps with the density for translation elongation factor eEF3 in eEF3-bound *S. cerevisiae* 80S ribosome map (EMD-12064[32]) (**Extended Data Fig.2b**), while rigid body fitting of atomic structure of eEF3 (PDB accession code: 2IX8[33]) results in a CC of 0.7 (**Extended Data Fig.2b, Middle panel**). Furthermore, the additional density in Class 14 bound to 40S subunit is mapped to elongation factor 1-alpha-A (eEF1A) (**Fig.2a, Right panel**), supported by the cryo-EM map of mammalian 80S ribosome with eEF1A cofactor reconstructed by nextPYP[6] (EMDB accession code: EMD-41205) (**Extended Data Fig.2c**). Rigid body fitting of the predicted structure of eEF1A from AlphaFold Protein Structure Database (AF-P0CT53) [34,35] validated the additional density in Class 14 having the same overall triangular arrangement as eEF1A, with a cc of 0.6 (**Extended Data Fig.2c, Middle panel**). DeePiCt also captured the conformations seen in our Classes 2 and 12, after labor-consuming, multi-round hierarchical focused classifications[17].

In addition to revealing compositional differences of 80S ribosome in various classes, OPUS-TOMO unraveled a series of transient states of 80S ribosome throughout the translation elongation cycle. 80S ribosome has three tRNA binding sites, A, P and E sites. In the elongation cycle, these sites are sequentially occupied and released by tRNAs, facilitating the translation of mRNA to peptide chains. As outlined by Ranjan *et. al.*[32], the elongation cycle initiates with

tRNAs at the A- and P-site (Pre-1 state), transitions through various configurations including Pre-3 state with A- and hybrid P/E-site tRNAs, Pre-4 state with hybrid A/P- and P/E-site tRNAs, Post-1 state with P- and E-site tRNAs, and ends with Post-3 state with a sole occupancy at the P-site. Impressively, these distinct dynamic states are faithfully captured in OPUS-TOMO's reconstructions: Class 1 shows A- and P-site tRNAs as in Pre-1 state; Class 5 shows A- and hybrid P/E-site tRNAs as in Pre-3 state; Class 12 shows hybrid A/P- and hybrid P/E-site tRNAs as in Pre-4 state; Class 20 shows P- and E-site tRNAs as in Post-1 state; Class 3 shows only P-site tRNA as in Post-3 state (**Fig.2b** and **Extend Data Fig.3**). The occupied sites in density maps are confirmed by the rigid-body fitting of the atomic structure of yeast 80S ribosome with A/A- and P/P- tRNAs (PDB accession code: 7B7D[32]) (**Fig.2b**). Furthermore, by analyzing the principal components (PCs) of the latent space learned by OPUS-TOMO, we can trace the progression of these states. For example, traversing along the PC8 reveals a transition from a state with P-site tRNA and A/T tRNA bound to eEF1A cofactor (-5PC8), through a state with A- and P-site tRNAs (1.6PC8), and culminating in a state with hybrid A/P- and P/E-site tRNAs (5PC8) (**Fig.2c, Extended Data Fig.4a, Extended Data Fig.5**). The densities for eEF1A and A/T tRNA in "-5PC8" overlaps with the corresponding densities in EMD-41205 showing an eEF1A cofactor delivering the A/T tRNA[6] (**Extended Data Fig.5a~b**) The AlphaFold predicted structure for eEF1A, AF-P0CT53, has a correlation coefficient of 0.8 with the corresponding density of "-5PC8" (**Extended Data Fig.5c**). The transition along PC8 clearly demonstrates the translocation dynamics of tRNAs inside the 80S ribosome: the translocation of A/T tRNA to the A-site, and the P-site tRNA to P/E-site (**Supplementary Video 1**). Noticeably, the CC between "-5PC8" and AF-P0CT53 is higher than that between "Class 14" and AF-P0CT53. This might be caused by the difference between PCA and KMeans clustering in latent space. PC8 captures the

compositional changes in higher qualities by filtering out latent dimensions that are uncorrelated with the occupancy changes of eEF1A cofactor. In contrast, "Class 14" was reconstructed as the average of a set of closely distributed latent encodings, which may include noisy latent encodings from particles without eEF1A. Additionally, PC8 also reveals the binding and dissociation of elongation factors, eEF1A and eEF3, with the 40S subunit, while the ES27L RNA gradually appearing at the exit tunnel, as well as the rotation of 40S subunit (**Supplementary Video 2**). Traversing along the PC9 shows the transition beginning with the state with A- and P-site tRNA, progressing to a state where one end of the A-site tRNA detaches from 60S subunit, and ending with a state with A/T tRNA and P-site tRNA (**Extended Data Fig.4b** and **Supplementary Video 3**). Traversing along the PC12 depicts the movement of L1 stalk from an "open" conformation, where it is disengaged from subunits, leaving the E-site accessible, to a "closed" conformation where it occupies the inter-subunit space between the head of the 40S subunit and the 60S subunit, closing the E-site (**Supplementary Video 4**).

Reconstruction of Fatty Acid Synthase Structures from Tomograms of Wild-type *S. pombe*

Following the same procedure, we performed structural heterogeneity analysis on *in vivo* fatty acid synthase (FAS) in tomograms acquired from cryo-FIB-milled lamellae of wild-type *S. pombe*, using the tilt series and their corresponding parameters of tilt series alignment deposited in EMPIAR-10988[17]. A total of 4800 particles for FAS were picked using template-matching method in pyTOM[14] with a published *S. cerevisiae* FAS map[36] (EMDB accession code: EMD-1623) as the template. The picked subtomograms were extracted from the tomograms using

RELION 3.0.8[9]. We then performed 3D refinement using picked subtomograms and the low-pass filtered *S.cerevisiae* FAS map as the starting model, and obtained a low-resolution and not-so-continuous subtomogram average. Given the D_3 symmetry inherent to the FAS structure, we applied symmetry expansion to the refined pose parameters of subtomograms using RELION 3.0.8 as a preparatory step for structural heterogeneity analysis. Following this symmetry expansion, OPUS-TOMO was trained on subtomograms with the assumption that their projection angles correspond to one of the six equivalent angles satisfying D_3 symmetry.

The UMAP visualization of the 10-dimensional latent space learned by OPUS-TOMO shows distinct clusters, which we clustered into 20 classes by KMeans algorithm, and reconstructed 3D structure for each class by supplying cluster centroids into the trained Decoder of OPUS-TOMO (**Extended Data Fig.6**). Only Class 19 shows complete densities for FAS, and contains 221 particles.

The map of Class 19 displays higher-resolution features, for instance in regions marked by red box and red dashed box (**Fig.3a, Left panel**), compared to the subtomogram average obtained by DeePiCt[17] (EMDB accession code: EMD-14422) (**Fig.3a, Right panel**), when both maps are contoured at the same level. This is further supported by fitting the atomic structure of FAS (PDB accession code: 2VKZ[37]) into density maps (**Fig.3b**). Judging by the threshold Fourier Shell Correlations (FSC)=0.143 between the atomic model 2VKZ and the density maps, Class 19 reconstructed by OPUS-TOMO reaches a resolution of 20.0 Å, while the subtomogram average EMD-14422 from DeePiCt has a resolution of 27.4 Å (**Fig.3c**).

Among the total number of 4800 subtomograms from our template matching results, there were 115 subtomograms that overlapped with DeePiCt's result, among which 89 subtomograms were shared with Class 19 of our OPUS-TOMO analysis (**Extended Data Table**

1, Extended Data Fig.7a). The distribution of these subtomograms in each tilt series is shown in **Extended Data Fig.7b** and **Extended Data Table 1**. Although the portion of subtomograms that were included in both OPUS-TOMO and DeePiCt was relatively small, both methods faithfully re-captured the essential features of FAS. This is exemplified by the similarity of subtomograms from tilt series numbered as TS_045, where the 16 subtomograms in Class 19 of our analysis and the 18 subtomograms in DeePiCt's results, in spite of no overlaps, exhibited similar densities corresponding to the FAS complex (**Extended Data Fig.7c**).

Reconstruction of *E. coli* 70S Ribosome Structures

We similarly performed structural heterogeneity analysis on *in vitro E. coli* 70S ribosome with an EMPIAR accession code of 10304[38]. The tilt series were automatically aligned and reconstructed as tomograms using AreTOMO[39]. 168,000 subtomograms were picked in the 4x binned tomograms by template matching method in pyTOM[14] using an *E. coli* ribosome structure[40] (EMD accession code: EMD-6311) as the template, and subtomogram averaging were performed by RELION 3.0.8[9]. Structural heterogeneity analysis was performed by training OPUS-TOMO using the 2x binned subtomograms with pose parameters from subtomogram averaging.

The UMAP visualization[41] of the 12-dimensional latent spaced learned by OPUS-TOMO exhibits well-separated clusters, which we clustered into 16 classes using KMeans algorithm (**Extended Data Fig.8a**). The density maps for 16 classes were reconstructed by supplying the centroids of classes into the trained Decoder of OPUS-TOMO (**Extended Data Fig.8b**). Among these reconstructions, Classes 7 and 12~15 were obvious noises, Classes 0 and 8 correspond to the single 50S subunit, while Classes 1~6 and 9 show complete densities for 70S ribosome

(**Extended Data Fig.8b; Extended Data Fig.9**). There are visible compositional and conformational differences among those classes. As an example, the relative displacement between the 30S subunits of Class 5 and Class 6 is highlighted with red arrows, while the relative displacement of L1 stalks is indicated by a yellow dashed circle between the two classes (**Fig.4a**). Furthermore, the density maps reconstructed by OPUS-TOMO unveiled the translocation of tRNAs in the three binding sites of ribosome, *i.e*., A, P and E sites. For reference, we selected the density map of aminoacyl-tRNA·EF-Tu·GDP·kirromycin-bound *E. coli* ribosome[42] (EMDB accession code: EMD-5036) and its corresponding atomic model (PDB accession code: 4V69[42]), which exhibits tRNAs occupying the P and E sites. Compared to EMD-5036 and 4V69, we observed that the A, P and E tRNA binding sites in Class 4 are empty, whereas the E site in Class 5, and the P site in Class 6 are occupied (**Fig.4b**).

We performed PC analysis to explore the latent space learned by OPUS-TOMO, and reconstructed dynamical changes for 70S ribosome. Traversing the PC5 revealed significant motions: the L1 stalk is moving towards the inter-subunit region between the head of 30S subunit and the 50S subunit, while the L11 stalk transits from visible to invisible, and concurrently, that the head of the 30S subunit rotates towards the L1 stalk (**Supplementary Video 5)**. Traversing the PC4 uncovers that the head of 30 subunit moves away from its initial association with the 50S subunit, accompanied by a partial loss of head density (**Supplementary Video 6)**, indicating a complex dynamical behavior during the dissociation process.

Discussion

In their native cellular environment, macromolecules undergo constant compositional and conformational changes to perform their biological functions, leading to substantial structural

heterogeneity that presents enormous challenges to many existing methods for structural studies. In particular, the fulfillment of cryo-ET's promise to resolve biological processes *in situ* at molecular resolution requires joint efforts. In this paper, we introduced OPUS-TOMO, a framework for resolving structural heterogeneity in cryo-ET data. The method trains a convolutional neural network to encode subtomograms into a low-dimensional latent space, from which the per-particle 3D structures for subtomograms can be decoded. Hence, it can retrieve the complete landscape of compositional and conformational variations preserved in cryo-ET data.

A major advantage of OPUS-TOMO is that, by employing an Encoder-Decoder architecture, it distills the structural information from subtomograms into a smooth, low-dimensional latent space, within which structurally heterogeneous subpopulations can be clustered, and compositional and conformational dynamics can be explored by traversing along the PC axes. The smoothness of the latent space is critically important for structural heterogeneity analysis, as it guarantees the linear interpolation in latent space corresponding to the continuous structural changes in 3D space. In OPUS-TOMO, the smoothness of the latent space with respect to structural variations is prompted by both KL divergence from β-VAE[23,28] and the structural disentanglement prior from OPUS-DSD[24]. Such enhancements have not only strengthened OPUS-TOMO's classification precision, as evidenced by the substantial improvement of the resolution from 27.4 Å to 20.0 Å in reconstructing FAS (**Fig.3c**), but also allowed smooth modeling of complexed dynamical processes through linear interpolations across its PCs, as exemplified by visualizations of the translocation of tRNAs during the elongation cycle and the binding of eEF3 cofactor in the 80S ribosome (**Supplementary Videos 1** and **2**),

OPUS-TOMO is successful in reconstructing both compositional and conformational dynamics using latent codes sampled along PCs.

A number of challenges remained for cryo-ET. One major challenge is from the spatially varying CTF of subtomograms due to tilting in data acquisition[13]. OPUS-TOMO accounts for the defocus gradient by implementing a 3DCTF model[9]. The CTF parameters of a subtomogram are calculated according to its coordinate in the tomogram and stored into files in Self-defining Text Archiving and Retrieval (STAR) format[43]. OPUS-TOMO then reads the defocus parameters from STAR files and reconstructs the 3DCTF volume for each subtomogram *ad hoc* during training. This implementation is flexible enough to accept per-particle defocus parameters estimated by other programs, such as TomoAlign[44], thereby allowing a wide interoperability with other software for 3DCTF correction. Moreover, since OPUS-TOMO operates on subtomograms, it can benefit from advancements of preprocessing algorithms that improve the quality of subtomograms, such as Warp[7].

Another major challenge is pose assignment error in recovering the per-particle ground-truth conformation for subtomograms. Theoretically, the ground-truth conformation can be reconstructed from subtomograms with poses defined in relative to their ground-truth structures. However, the ground-truth pose of a subtomogram is unattainable. In OPUS-TOMO, the projection pose of a subtomogram in training data is determined by comparing them to a consensus model at different projection poses exhaustively. This comparison process introduces pose assignment errors to accommodate the subtomograms with incomplete and heterogeneous structure into a single average. Neural network could reconstruct distorted conformations when trained by subtomograms with erroneous pose parameters. To reduce the impact of pose assignment error, OPUS-TOMO introduces an end-to-end pose correction neural network to

estimate the deviation between the estimated pose of a subtomogram and its ground-truth by minimizing the reconstruction loss. This network serves to minimize the deviation between the estimated pose and the ground-truth during training, thus reducing the distortion of reconstructed structures.

The process of 3D reconstruction in cryo-ET using subtomograms shares lots of similarities to that of cryo-EM using 2D images. They both intend to recover 3D structures using inputs with incomplete structural information, *i.e.*, subtomograms or 2D images. Hence, practices with established structural heterogeneity resolved in cryo-EM can help resolve structural heterogeneity in cryo-ET. Furthermore, structural disentanglement prior[24] from OPUS-DSD is leveraged to alleviate the 'missing wedge' problem in OPUS-TOMO. This prior can aggregate subtomograms from the same structure with different orientations to the same latent encoding, thus increasing the coverage of the spectrum of 3D structural information at each latent encoding during training. In addition, to reduce the projection pose errors in inputs, subtomograms are aligned with the canonical orientation of consensus subtomogram average before supplying into the Encoder[24]. Lastly, data augmentation is implemented in OPUS-TOMO to reduce its sensitivity to image contrast variations in subtomograms[24].

Methods

Subtomogram Formation Model

During data acquisition in cryo-ET, the specimen is usually tilted around a single axis to collect a series of images with different tilt angles. Each tilt image corresponds to a 2D projection of the specimen. According to the Fourier-slice theorem, the Fourier transform of the 2D projection is a central slice through the 3D Fourier transform of the specimen, while the

orientation of the 2D slice corresponds the projection angle. The Fourier transform of the tilt series can fill up a 3D volume, *i.e.*, to allow reconstruction of the 3D structure of the specimen, which is often referred to as a tomogram. However, due to the limitation of the geometry of specimen holders, the attainable tilt angles are limited to a certain range. The tilt series from cryo-ET lack high tilt images, thus creating wedge shaped empty region in the Fourier transform of a tomogram, which is often referred to as 'missing wedge'. Similar as single particle analysis of cryo-EM data, there are multiple copies of the same structure within a tomogram. Cryo-ET can determine structures to sub-nanometer resolution by aligning and averaging these copies, *i.e.*, subtomograms.

The image formation model of cryo-ET can be defined in the Fourier domain. For a subtomogram X_i, denote the Fourier transform of X_i as $\mathcal{X}_i$, let the Fourier transform of the reference volume V be $\mathcal{V}$, suppose X_i is rotated from V by rotation matrix R_i, the Fourier transform of a subtomogram has the following relation w.r.t the Fourier transform of the reference volume,

$$\mathcal{X}_i(\boldsymbol{x}) = \mathrm{CTF}_i(\boldsymbol{x})V(R_i\boldsymbol{x}) + \epsilon_i(\boldsymbol{x}), \qquad \text{(Equation 1)}$$

where $\boldsymbol{x}$ is the 3D coordinate of a voxel, $\mathrm{CTF}_i(\boldsymbol{x})$ is a 3D model in the Fourier domain that describes the effects of the CTF on the 2D images in the tilt series, and the effects of the missing wedge, and ϵ_i is the 3D Gaussian noise, which is assumed to be independent and zero-mean for each voxel. As in cryo-EM, the image collected in cryo-ET is the 2D projection of a 3D structure which is convoluted with a point spread function, whose Fourier transform is the CTF, to model the effects of underfocus and the aberration of the objective lens. The subtomogram is of three dimensionality and consists of 2D images with different tilt angles. The CTF model for the subtomogram in cryo-ET hence becomes 3D and accounts for the CTF of each tilt image. It

can be obtained by backprojecting the 2D CTFs of all images in the tilt series into an empty 3D volume in Fourier domain. Each 2D CTF is placed as a central slice in the 3D CTF, whose orientation is determined by its tilt angle. Reconstructing the CTF model for subtomograms by backprojection naturally handles the missing wedges as the region which is not covered by tilt series remains empty. However, the CTF is spatially variant because the tilting and the thickness of specimen[20]. Due to the tilting of the specimen holder in cryo-ET, the defocus values vary significantly along the axis in the image plane that is perpendicular to the tilt axis. In addition, particles in the specimen are distributed in different heights along Z axis because of the thickness of the specimen. Hence, tilt images in subtomograms extracted from different locations of the image have different defocus. Various methods have been proposed to correct the CTF gradient in cryo-ET [20]. In this paper, we followed the approach proposed in RELION[9] to correct CTF for each subtomogram individually. Using this approach, for a particle, the defocus of each tilt image is calculated according to its coordinate in the tomogram and the tilt angle[9,44]. Formally, let the tilt axis be y axis, suppose the coordinate of a particle in relative to the center of tomogram is $[x, y, z]$, and the tilt angle of the image is θ, following the convention of rotation in AreTomo[39], the coordinate of the particle in XZ plane after tilting is

$$\begin{pmatrix} x' \\ z' \end{pmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} x \\ z \end{pmatrix}. \qquad \text{(Equation 2)}$$

If the average defocus of the tilt image is D, the defocus of the particle D_i is

$$D_i = D - Az', \qquad \text{(Equation 3)}$$

where A is the angstrom per pixel, and the defocus difference is proportional to the negative of tilted z coordinate, where the minus sign is taken since higher underfocus corresponds to more negative z.

To capture the structural heterogeneity, the reference volume in (Equation 1) is defined on a per-particle basis. We assume that the structures of subtomograms reside in a low-dimensional latent space as in OPUS-DSD[24]. Hence, the structure of a particle can be associated with a low-dimensional vector $\boldsymbol{z}$. For a particle whose latent encoding is $\boldsymbol{z}$, given a neural network V that transforms $\boldsymbol{z}$ into a 3D density map, the Fourier transform of the subtomogram of the particle can be expressed as

$$\mathcal{X}_i(\boldsymbol{x}|\boldsymbol{z}, R_i) = \mathrm{CTF}_i(\boldsymbol{x})V(R_i\boldsymbol{x}|\boldsymbol{z}) + \epsilon_i(\boldsymbol{x}). \qquad \text{(Equation 4)}$$

(Equation 4) defines the subtomogram formation model for the particle in cryo-ET which describes the effects of both spatial CTF variations and structural heterogeneity.

Training Objective

The main objective of OPUS-TOMO is learning the per-particle structure for a subtomogram using subtomogram averaging result. Due to the presence of pose assignment error in consensus refinement resulted from subtomogram averaging, the pose of each particle is corrected by a pose $R_{i,c}$, which can be estimated by neural network. The neural networks are trained by minimizing the difference between the subtomogram reconstructed from the structure and the experimental subtomogram X_i, that is,

$$\min_{V, R_{i,c}} \left\| \mathrm{CTF}_i V\left(R_{i,c} R_{i,t}, \boldsymbol{z}_i\right) - \mathcal{X}_i \right\|^2 \qquad \text{(Equation 5)}$$

where $\boldsymbol{z}_i$ is the latent code for the template volume of image i. To estimate the pose correction $R_{i,c}$, we propose using an additional 4-dimensional latent space to predict it. For each particle, its pose correction can be estimated by a neural network R which maps a 4-dimensional latent vector $\boldsymbol{z}_i^c$ to the pose parameter. The rotation of pose correction $R_{i,c}$ is represented by quaternion.

To limit the range of rotation correction, the neural network only estimates the last three dimensions of the quaternion, while the first dimension of quaternion is fixed to be a constant 8. The initial weights of the MLP for pose correction are set to zeros. The training objective in (Equation 5) is not complete yet since the latent codes $\boldsymbol{z}_i$ and $\boldsymbol{z}_i^c$ of each particle remains to be inferred. We adopt the variational autoencoder (VAE) architecture where the distributions of latent codes $\boldsymbol{z}$ and $\boldsymbol{z}^c$ are inferred by an Encoder network f. Specifically, OPUS-TOMO leverages a combination of the training objective of β-VAE[28] and the structural disentanglement prior proposed in OPUS-DSD[24]. Let the dimension of 3D volume be N^3, the training objective of OPUS-TOMO is of the form,

$$\min_{f,V} E_{f(\boldsymbol{z}_i,\boldsymbol{z}_i^c|X_i)} - \log \sum_{\theta' \in NN(\theta)} \exp -\frac{1}{2}\left\|\mathcal{X}_i - \mathrm{CTF}_i V\left(R(\boldsymbol{z}_i^c)R_{i,t}, \boldsymbol{z}_i\right)\right\|^2 \qquad \text{(Equation 6)}$$

$$+ \lambda_{KL} D_{KL}\left(f(\boldsymbol{z}_i, \boldsymbol{z}_i^c|X_i)\middle|\middle|N(0,I)\right) + \lambda J\left(f(\boldsymbol{z}_i, \boldsymbol{z}_i^c|X_i)\right)$$

$$+ \lambda_{tv} \sum_{\boldsymbol{x}} \left\|\frac{\partial V(\boldsymbol{x}|\boldsymbol{z}_i)}{\partial \boldsymbol{x}}\right\|_2 + \lambda_{l_1} \sum_{\boldsymbol{x}} \|V(\boldsymbol{x}|\boldsymbol{z}_i)\|_1,$$

where the first term is expectation of error between the experimental subtomogram X_i and the reconstruction over the distribution of latent codes, $R(\boldsymbol{z}_i^c)$ represents the pose correction estimated by the neural network using the latent code $\boldsymbol{z}_i^c$ as input, $N(0,I)$ is the standard Gaussian distribution, D_{KL} is the KL divergence between the distributions of latent codes estimated by the Encoder network f and the standard Gaussian distribution, λ_{KL} is the restraint strength for the KL divergence, λ is the restraint strength of structural disentanglement prior for the latent code of structure $\boldsymbol{z}$, λ_{tv} is the restraint strength for the total variation of structure, and λ_{l_1} represents the restraint strength of the L_1 norm of the structure. It is worth noting that the

expectation term in (Equation 6) is intractable, and approximated by the reparameterization trick during training [26].

Experimental Protocol

We followed the conventional approaches to prepare subtomograms and performed subtomogram averaging before the structural heterogeneity analysis by OPUS-TOMO[20]. Due to inaccuracies in image tracking during data collection, the tilt series was first aligned to make y axis as the tilt axis, and the center of image being the center of tilting before tomogram reconstruction. The alignment of tilt series can be performed either by IMOD[45] or AreTOMO[39]. The tomogram can then be reconstructed by backprojecting the aligned tilt series into a 3D volume according to their tilt angles.

Template matching from pyTOM was used to select subtomograms from target species. For each tilt series, the subtomograms were sorted according to their correlation coefficients with respect to template calculated by pyTOM, and the subtomograms with correlation coefficients that were higher than a threshold was selected. For FAS, the threshold was set to 0.2, and the number of subtomograms was rounded to multiple of hundred for simplicity. For 70S ribosome, the threshold was set at 0.3 since the corresponding dataset had a higher signal to noise ratio. To speed up template matching, the tomograms were 4x binned. The subtomograms for picked particles were extracted by *relion_preprocess* routine in RELION 3.0.8[9]. The average defocus parameters for each tilt image in the tilt series were determined by CTFPLOTTER[46,47] in IMOD. We used the data preparation scripts from RELION[4] to compute the per-particle variations of CTF parameters for each subtomogram according to its coordinates in a tomogram. We reconstructed the 3D CTF for each subtomogram using *relion_reconstruct* routine in RELION

3.0.8[9]. With subtomograms and their 3D CTFs, we performed subtomogram averaging using *relion_refine_mpi* routine in RELION 3.0.8[9] to determine the pose parameters for subtomograms in relative to the subtomogram average. Finally, we performed structural heterogeneity analysis by training OPUS-TOMO using the subtomograms with their poses estimated by subtomogram averaging.

Training

The training process of OPUS-TOMO is similar to OPUS-DSD[24]. The subtomogram undergoes data augmentation processes before being supplied into the Encoder. The data augmentation consists of several steps. Firstly, the Fourier transform of the subtomogram is blurred by a randomly sampled B-factor in the range $[-0.25, 0.75] \times 4\pi^2$. The contrast and brightness of the subtomogram is also randomly altered according to the SNR of the subtomogram. Next, the subtomogram is multiplied with a randomly chosen constant in the range $1 + [-0.15, 0.15] \times \mathrm{SNR}$ to change its overall contrast, and the subtomogram is added by a randomly chosen constant in the range $[-0.2, 0.2] \times \mathrm{SNR}$ to change its overall brightness. The Encoder takes the randomly augmented subtomogram and outputs its latent codes. The Decoder takes the latent codes to reconstruct the corresponding 3D volume and predict the pose correction of the input subtomogram. OPUS-TOMO reads the CTF parameters for the input subtomogram from STAR file and computes the corresponding 3D CTF. The Eecoder can then reconstruct the subtomogram using the 3D volume, pose, and 3D CTF according the subtomogram formation model in (Equation 4). The reconstructed subtomogram and the original subtomogram for input together forms the reconstruction loss. We construct the training objective by combining the reconstruction loss, the priors for latent codes, and the total variation and L_1 norms[48] for the

template volumes with specified weights as it is given in (Equation 6). During training, OPUS-TOMO uses the same batching process as in OPUS-DSD[24] to sample subtomograms with similar projection directions as a batch.

OPUS-TOMO requires a moderate amount of training time. It only takes 1 hour to train 25900 subtomograms of size 128x128x128 on 4 NVIDIA V100 GPUs for one epoch. This is comparable to the traditional 3D classification methods in RELION using the same hardware for the same number of data when the number of class is set to 4.

Hyperparameters' Settings

The training losses of OPUS-TOMO were minimized by the Adam optimizer[49]. The learning rate of Adam was set to 5×10^{-5}. Let the SNR of dataset be SNR, the restraint strength λ for structural disentanglement prior was set to $0.5 \times \mathrm{SNR}$, and the restraint strength λ_{KL} for KL divergence between latent codes and standard Gaussian was set to $0.5 \times \mathrm{SNR}$. The SNR was estimated *ad hoc* during training by treating the reconstruction from OPUS-TOMO as signal and the input subtomogram with reconstruction subtracted as noise. Formally, using a batch of N images, the SNR was estimated as,

$$\mathrm{SNR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left\|\tilde{X}_i\right\|^2}{\left\|X_i - \tilde{X}_i\right\|^2}, \qquad \text{(Equation 7)}$$

where $\tilde{X}_i$ was the subtomogram reconstructed by OPUS-TOMO, and X_i was the experimental subtomogram. $\left\|X_i - \tilde{X}_i\right\|^2$ estimated the norm of noises in the experimental subtomogram, while $\left\|\tilde{X}_i\right\|^2$ estimated the norm of signals. The restraint strength λ_{tv} for the total variation of 3D reconstruction was set to 0.1, and the restraint strength λ_{l_1} for the L_1 norm of 3D reconstruction was set to 0.3.

Acknowledgements

J.M. wants to thank the support from the National Key Research and Development Program of China (no. 2021YFF1200400), Science and Technology Innovation Plan of Shanghai Science and Technology Commission (no.23JS1400200), Shanghai Municipal Science and Technology Major Project (no.2018SHZDZX01), and ZJ Lab.

Author Contributions

Z.L., and J.M. conceived the work. Z.L. designed the algorithm and implemented the software. All authors wrote and approved the final manuscript.

Competing Interests

Q.W. is an employee of Harcam Biomedicines who is bound by confidentiality agreements that prevents her from disclosing the competing interests in this work. All remaining authors declare no competing interest.

Figure Legends

Figure 1. Architecture of OPUS-TOMO. a). **Overall architecture.** Pose refers to the projection direction of the subtomogram with respect to the consensus subtomogram average. 3DCTF refers to its 3D contrast transfer function. z is the mean vector of a latent code. σ is the standard deviation vector of a latent code. z_c is the mean vector of pose correction latent code. σ_c is the standard deviation vector of pose correction latent code. $\mathrm{kNN}(z, z_c)$ refers to the K nearest neighbor of the concatenation of z and z_c. $\mathrm{kFP}(z, z_c)$ refers to the K farthest points of

the concatenation of z and z_c. For better visualization, the link between the pose correction latent code and the memory is omitted. The standard Gaussian prior for latent distribution as well as the smoothness and sparseness priors for the 3D volume are also omitted. **b)**. **Architecture of the pose corrector of OPUS-TOMO.** This diagram shows the pose corrector that translates the latent encoding z_c into a pose correction. MLP denotes the multi-layer perceptron. All MLPs except the last one were with LeakyReLU nonlinearities with negative slope 0.2.

Figure 2. Heterogeneity analysis on the *S. pombe* 80S ribosome (data from EMPIAR-10988). a). Comparison between classes with different cofactors. Class 2 shows the structure of 80S ribosome with an ES27L RNA at the exit tunnel, Class 12 shows the 80S ribosome with the cofactor eEF3 bound to 40S subunit, and Class 14 shows the 80S ribosome with eEF1A bound to 40S subunit. The densities for additional components are marked by red dashed ellipses. **b)**. The transient states for tRNA translocation found in Classes 1, 5, 12, 20, and 3 from the KMeans clustering of latent space learned by OPUS-TOMO. The sites occupied by tRNAs are marked by red dashed ellipses and annotated with corresponding labels. The atomic structures of tRNAs at A- and P- sites are superimposed into the density maps for comparison. The atomic structure of yeast 80S ribosome with A/A- and P/P- tRNAs (PDB accession code: 7B7D) resolved by Ranjan *et. al*[32] is fitted into these classes. The A-site tRNA is colored in red, and the P-site tRNA is colored in cyan. **c)**. The transient states for tRNA translocation reconstructed by the Decoder of OPUS-TOMO using latent codes at different locations along PC8. The sites occupied by tRNAs are marked by red dashed ellipses and annotated with corresponding labels. The A/T-site consists of A-site at 40S subunit and T-site at 60S subunit, the hybrid A/P-site consists of A-site

at 40S subunit and P-site at 60S subunit, and the hybrid P/E-site consists of P-site at 40S subunit and E-site at 60S subunit. The locations of latent codes are specified as xPC8. All maps and models are visualized by ChimeraX[50].

Figure 3. Heterogeneity analysis on the *S. pombe* FAS (data from EMPIAR-10988). a). Comparison between the FAS reconstructed by OPUS-TOMO (Class 19) and the deposited density maps for FAS (EMD-14422). Side and top views of these EM maps are presented. Class 19 reconstructed by the Decoder of OPUS-TOMO shows complete densities for FAS, and EMD-14422 shows the subtomogram average of subtomograms picked by DeePiCt[17]. The regions where Class 19 presents structural features at a higher resolution than EMD-14422 are marked by red boxes. **b).** The atomic structure of FAS with PDB accession code 2VKZ[37] is fitted into density maps. The fitted atomic structures and density maps in marked regions are shown in zoomed views, respectively. Distinct subunits are annotated by labels. **c**) The model-map FSCs between the atomic structure of FAS with PDB accession code 2VKZ and the density maps from OPUS-TOMO and DeePiCt. All maps and models are visualized by ChimeraX[50].

Figure 4. Heterogeneity analysis on the *E. coli* 70S ribosome (data from EMPIAR-10304). a). Comparison between Class 5 and Class 6. The head of 30S subunit marked by the red dashed circle in these two classes undergoes a large rotation. Red arrows indicate the directions of relative displacements between the corresponding structures in these two classes. **b)**. The occupancy differences of the tRNA binding sites in different classes. The atomic structures of tRNAs at A/T-, P- and E- sites are superimposed into the density maps for comparison. The atomic structure of *E.coli* 70S ribosome with PDB accession code 4V69[42] is fitted into these

maps. The A/T-site tRNA is colored in green, the P-site tRNA is colored in red, and the E-site tRNA is colored in cyan. EMD-5036[42] is a 70S ribosome map with A/T-, P- and E-site tRNAs. A/T-site consisting of A-site at 30S subunit and T-site at 50S subunit. Class 4 is a 70S ribosome map with no tRNAs in binding sites. Class 5 is a 70S ribosome map with E-site tRNA. Class 6 is a 70S ribosome map with P-site tRNA. Zoomed views are shown for regions marked by red boxes. The tRNAs in binding sites are marked by red dashed circles and annotated with corresponding labels. All maps and models are visualized by ChimeraX[50].

Extended Data Figure Legends

Extended Data Figure 1. Heterogeneity analysis on the *S. pombe* 80S ribosome (data from EMPIAR-10988). a). UMAP visualization of the 12-dimensional latent space learned by OPUS-TOMO and the distribution of centroids for clusters found by KMeans algorithm. Solid black dot represents the cluster center for labelled class. **b)**. Twenty-four density maps at class centers reconstructed by the Decoder of OPUS-TOMO. All maps are contoured at the same level and visualized by ChimeraX[50].

Extended Data Figure 2. Comparison between density maps from different classes reconstructed by the Decoder of OPUS-TOMO and the atomic structures of cofactors. Atomic structures are fitted into maps by rigid-body fitting. Zoomed views are shown for cofactors marked by red dashed ellipses. "CC" represents the correlation coefficient between the atomic structure and the density map, which was calculated by *measure* in ChimeraX[50] and for densities from atomic structure above level 0.4. The deposited maps with the same cofactors are shown for reference. **a)**. Comparison between the density map of Class 2 and the additional

ES27L RNA at the exit tunnel. The deposited density map EMD-10608[30] with ES27L RNA is shown for reference. The structure with PDB accession code 3IZD[31] is fitted into maps. **b)**. Comparison between the density map of Class 12 and the cofactor eEF3 bound to the 40S subunit. The deposited density map EMD-12064[32] with eEF3 is shown in reference. The structure with PDB accession code 2IX8[33] is fitted into maps. **c)**. Comparison between the density map of Class 14 and the cofactor eEF1A bound to the 40S subunit. The deposited density map EMD-41205[6] with eEF1A is shown for reference. The predicted structure of AF-P0CT53[34] is fitted into maps. All maps and models are visualized by ChimeraX[50].

Extended Data Figure 3. Reference density maps for transient states in translation elongation cycle resolved by Ranjan *et. al*[32]. The atomic structures of tRNAs at A- and P- sites are superimposed into the density maps for comparison. The atomic structure of yeast 80S ribosome with A- and P- site tRNAs (PDB accession code: 7B7D) resolved by Ranjan *et. al*[32] is fitted into these states. The A-site tRNA is colored in red, and the P-site tRNA is colored in cyan. Maps show densities overlapping with tRNA structures at the occupied sites. **a)**. The density map for Pre-1 state. The red dashed ellipses highlight the A- and P- site tRNAs. **b)**. The density map for Pre-3 state. The red dashed ellipses highlight the A- and P/E- site tRNAs. The P/E site tRNA extends from P-site in 40S subunit to E-site in 60S subunit. **c)**. The density map for Pre-4 state. The red dashed ellipses highlight the A/P- and P/E- site tRNAs. The A/P site tRNA extends from A-site in 40S subunit to P-site in 60S subunit. **d)**. The density map for Post-1 state. The red dashed ellipses highlight the P- and E- site tRNAs. The E-site tRNA extends from E-site in 40S subunit to E-site in 60S subunit. **e)**. The density map for Post-3 state. The red dashed ellipse highlights the P-site tRNA. All maps and models are visualized by ChimeraX[50].

Extended Data Figure 4. The transient states for tRNA translocation reconstructed by the Decoder of OPUS-TOMO using latent codes at different locations along different PCs. The locations of latent codes are specified as xPCn. The atomic structures of tRNAs at A- and P-sites are superimposed into the density maps for comparison. The atomic structure of yeast 80S ribosome with A- and P- site tRNAs (PDB accession code: 7B7D[32]) is fitted into these maps. The tRNA at A-site is colored in red, and the tRNA at P-site is colored in cyan. Zoomed views are shown for regions marked by red boxes. **a)**. The transient states for tRNA translocation reconstructed by the Decoder of OPUS-TOMO using latent codes at different locations on PC8. A/T tRNA in "-5PC8", and A/P and P/E tRNAs in "5PC8" are marked by red dashed ellipses. A/T tRNA extends from A-site in 40S subunit to T-site in 60S subunit, A/P tRNA extends from A-site in 40S subunit to P-site in 60S subunit, and P/E tRNA extends from P-site in 40S subunit to E-site in 60S subunit. **b)**. The transient states for tRNA translocation reconstructed by the Decoder of OPUS-TOMO using latent codes at different locations on PC9. The tRNA marked by red dashed ellipse in "0.8PC9" detached from the A-site in 60S subunit. Red dashed ellipse in "4PC9" marks the A/T tRNA. All maps and models are visualized by ChimeraX[50].

Extended Data Figure 5. The densities for eEF1A and A/T tRNA in the density map "-5PC8" reconstructed by the Decoder of OPUS-TOMO. a). The densities for eEF1A and A/T tRNA in "-5PC8". The density for eEF1A is highlighted by red ellipse, and the density for A/T tRNA is highlighted by red dashed ellipse. The predicted structure for eEF1A under AF-P0CT53[34] is fitted into the density map of "-5PC8" to highlight the density for eEF1A. The atomic structure of yeast 80S ribosome with A- and P- site tRNAs (PDB accession code: 7B7D[32])

is fitted into "-5PC8" to highlight the A- and P- sites. **b)**. The densities for eEF1A and A/T tRNA in the deposited density map EMD-41205. **c)**. The fitting between the density of eEF1A from "-5PC8" and the predicted structure for eEF1A under AF-P0CT53[34]. "CC" represents the correlation coefficient between map and model. All maps and models are visualized by ChimeraX[50].

Extended Data Figure 6. Heterogeneity analysis on the *S. pombe* FAS (EMPIAR-10988). a). UMAP visualization of the 12-dimensional latent space learned by OPUS-TOMO and the distribution of centroids for clusters found by KMeans algorithm. Solid black dot represents the cluster center for labelled class. **b)**. Density maps reconstructed by the Decoder of OPUS-TOMO using twenty cluster centers as inputs. All maps are contoured at the same level and visualized by ChimeraX[50].

Extended Data Figure 7. Comparison of subtomograms for *S. pombe* FAS from different methods. a). The Venn diagram for the subtomograms for FAS from different methods. The blue ellipse represents the subtomograms from DeePiCt[17], which is 370 in total. The yellow ellipse represents the subtomograms from template matching[51], which is 4800 in total. The red ellipse represents the Class 19 from OPUS-TOMO's clustering result, which is 221 in total. The number of subtomograms in the intersection between DeePiCt and Class 19 is 89. The number of subtomograms in the intersection between DeePiCt and template matching is 115. **b)**. Distribution of the number of subtomograms from Class 19 given by OPUS-TOMO in each tilt series which overlap with DeePiCt's result. "TM overlapping with DeePiCt" refers to the number of subtomograms from the template matching (TM) result that overlap with DeePiCt's

result in each tilt series. "OPUS-TOMO" refers to the number of subtomograms from Class 19 in OPUS-TOMO's result that overlap with DeePiCt's result in each tilt series. **c)**. The projections of subtomograms from the tilt series TS_045 for Class 19 in OPUS-TOMO, and the projections of subtomograms from DeePiCt. The subtomograms are 4x binned and blurred by a Gaussian with standard deviation 1 before projecting along the Z axis.

Extended Data Figure 8. Heterogeneity analysis on the *E. coli* 70S ribosome (data from EMPIAR-10304). a). UMAP visualization of the 12-dimensional latent space learned by OPUS-TOMO and the distribution of centroids for clusters found by KMeans algorithm. Solid black dot represents the cluster center for labelled class. **b)**. Sixteen density maps reconstructed by the Decoder of OPUS-TOMO at the class centers. All maps are contoured at the same level and visualized by ChimeraX[50].

Extended Data Figure 9. Projections of subtomograms for different classes. a). The projections of subtomograms from the tomogram numbered Tilt1 for Class 0 from OPUS-TOMO, which show densities correspond to 50S subunit. **b)**. The projections of subtomograms from the tomogram numbered Tilt1 for Class 1 from OPUS-TOMO, which show densities correspond to 70S ribosome. The subtomograms are 8x binned and blurred by a Gaussian with standard deviation 0.75 before projecting along the Z axis.

Extended Data Table

Extended Data Table 1. Distributions of the number of subtomograms in each tilt series. "Template matching" refers to the number of subtomograms localized by template matching;

"DeePiCt" refers to the number of subtomograms localized by DeePiCt; "Template matching overlapping with DeePiCt" refers to the number of subtomograms that in both template matching and DeePiCt's localization result; "Class 19" refers to the number of subtomograms from Class 19 in each tilt series; and "Class 19 overlapping with DeePiCt" refers to the number of subtomograms that in both class 19 and DeePiCt's localization result.

Supplementary Video Legends

Supplementary Video 1. Traversal of PC8 of the latent space of the *S. pombe* 80S ribosome reveals the transition from the state with A/T site tRNA and P-site tRNA to the state with hybrid A/P and P/E site tRNAs, corresponding to the transition from Pre-1 state to Pre-4 state in the elongation cycle. The atomic model for A-site tRNA is colored in red, and the atomic model for P-site tRNA is colored in cyan. The T- and E- sites on the 60S subunit are annotated by corresponding labels. The video is created by traversing PC8 forward and backward twice. The dynamics of tRNAs can be identified by observing the occupancy changes at the A- and P- sites with atomic models.

Supplementary Video 2. Traversal of PC8 of the latent space of the *S. pombe* 80S ribosome reveals the binding and dissociation of the eEF1A and eEF3, and the occupancy transition of the ES27L RNA at the exit tunnel. The atomic model for eEF1A is colored in light blue, the atomic model for eEF3 is colored in blue, and the atomic model for ES27L RNA is colored in green. The video is created by traversing PC8 forward and backward twice.

Supplementary Video 3. Traversal of PC9 of the latent space of the *S. pombe* 80S ribosome reveals the transition from the state with A/T site tRNA and P-site tRNA to the state with A- and P-site tRNAs, demonstrating the delivery of tRNA into A-site. The video is created by traversing PC9 forward and backward twice.

Supplementary Video 4. Traversal of PC12 of the latent space of the *S. pombe* 80S ribosome reveals the dynamics of L1 stalk that opens and closes the E-site. The video is created by traversing PC12 forward and backward twice.

Supplementary Video 5. Traversal of PC5 of the latent space of the *E. coli* 70S ribosome reveals the movement of L1 stalk, and the compositional changes of L11 stalk. The video is created by traversing PC5 forward and backward twice.

Supplementary Video 6. Traversal of PC4 of the latent space of the *E. coli* 70S ribosome reveals the dynamics and compositional changes of the 30S subunit. The video is created by traversing PC4 forward and backward twice.

References

1 Chreifi, G., Chen, S., Metskas, L. A., Kaplan, M. & Jensen, G. J. Rapid tilt-series acquisition for electron cryotomography. *J. Struct. Biol.* **205**, 163-169 (2019).

2 Klumpe, S. *et al.* A modular platform for automated cryo-FIB workflows. *Elife* **10** (2021).

3 Zachs, T. *et al.* Fully automated, sequential focused ion beam milling for cryo-electron tomography. *Elife* **9**, e52286 (2020).

4 Bharat, T. A. & Scheres, S. H. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature protocols* **11**, 2054-2065 (2016).

5 Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat. Methods* **18**, 186-193 (2021).

6 Liu, H.-F. *et al.* nextPYP: a comprehensive and scalable platform for characterizing protein variability in situ using single-particle cryo-electron tomography. *Nat. Methods* **20**, 1909-1919 (2023).
7 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146-1152 (2019). https://doi.org:10.1038/s41592-019-0580-y
8 Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods* **15**, 955-961 (2018). https://doi.org:10.1038/s41592-018-0167-z
9 Bharat, T. A. M., Russo, C. J., Löwe, J., Passmore, L. A. & Scheres, S. H. W. Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure* **23**, 1743-1753 (2015). https://doi.org:10.1016/j.str.2015.06.026
10 Turk, M. & Baumeister, W. The promise and the challenges of cryo‐electron tomography. *FEBS letters* **594**, 3243-3261 (2020).
11 Robinson, C. V., Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450**, 973-982 (2007).
12 Beck, M. & Baumeister, W. Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? *Trends in cell biology* **26**, 825-837 (2016).
13 Schur, F. K. Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging. *Current opinion in structural biology* **58**, 1-9 (2019).
14 Chaillet, M. L. *et al.* Extensive Angular Sampling Enables the Sensitive Localization of Macromolecules in Electron Tomograms. *International Journal of Molecular Sciences* **24**, 13375 (2023).
15 Böhm, J. *et al.* Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14245-14250 (2000).
16 Martinez-Sanchez, A. *et al.* Template-free detection and classification of membrane-bound complexes in cryo-electron tomograms. *Nat. Methods* **17**, 209-216 (2020).
17 de Teresa-Trueba, I. *et al.* Convolutional networks for supervised mining of molecular patterns within cellular context. *Nat. Methods* **20**, 284-294 (2023).
18 Ronneberger, O., Fischer, P. & Brox, T. 234-241 (Springer).
19 Ben-Harush, K., Maimon, T., Patla, I., Villa, E. & Medalia, O. Visualizing cellular processes at the molecular level by cryo-electron tomography. *Journal of Cell Science* **123**, 7-12 (2010).
20 Pyle, E. & Zanetti, G. Current data processing strategies for cryo-electron tomography and subtomogram averaging. *Biochemical Journal* **478**, 1827-1845 (2021). https://doi.org:10.1042/bcj20200715
21 Wan, W. & Briggs, J. A. G. Cryo-Electron Tomography and Subtomogram Averaging. *Methods in Enzymology* **579**, 329-367 (2016). https://doi.org:10.1016/bs.mie.2016.04.014
22 Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Current Opinion in Structural Biology* **58**, 249-258 (2019). https://doi.org:10.1016/j.sbi.2019.05.021
23 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176-185 (2021). https://doi.org:10.1038/s41592-020-01049-4
24 Luo, Z., Ni, F., Wang, Q. & Ma, J. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nat. Methods*, 1729-1738 (2023).
25 Powell, B. M. & Davis, J. H. Learning structural heterogeneity from cryo-electron sub-tomograms with tomoDRGN. *Nat. Methods*, 1-12 (2024).
26 Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. *International Conference for Learning Representations* (ICLR, 2014).
27 Jaderberg, M., Simonyan, K., Zisserman, A. & Kavukcuoglu, K. Spatial transformer networks. *Advances in neural information processing systems* **28** (NeurIPS, 2015).

28 Higgins, I. *et al.* β-VAE: Learning basic visual concepts with a constrained variational framework. *International Conference on Learning Representations* (ICLR, 2017).
29 Lloyd, S. P. Least squares quantization in pcm. *IEEE Trans. Inf. Theory* **28**, 129-137 (1982). https://doi.org:10.1109/TIT.1982.1056489
30 Wild, K. *et al.* MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nature Communications* **11**, 776 (2020).
31 Armache, J.-P. *et al.* Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19748-19753 (2010).
32 Ranjan, N. *et al.* Yeast translation elongation factor eEF3 promotes late stages of tRNA translocation. *The EMBO journal* **40**, e106449 (2021).
33 Andersen, C. B. *et al.* Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature* **443**, 663-668 (2006).
34 Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **50**, D439-D444 (2022).
35 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
36 Gipson, P. *et al.* Direct structural insight into the substrate-shuttling mechanism of yeast fatty acid synthase by electron cryomicroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9164-9169 (2010).
37 Johansson, P. *et al.* Inhibition of the fungal fatty acid synthase type I multienzyme complex. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12803-12808 (2008).
38 Eisenstein, F., Danev, R. & Pilhofer, M. Improved applicability and robustness of fast cryo-electron tomography data acquisition. *J. Struct. Biol.* **208**, 107-114 (2019).
39 Zheng, S. *et al.* AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *Journal of Structural Biology: X* **6**, 100068 (2022).
40 Arenz, S., Nguyen, F., Beckmann, R. & Wilson, D. N. Cryo-EM structure of the tetracycline resistance protein TetM in complex with a translating ribosome at 3.9-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5401-5406 (2015).
41 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://arxiv.org/abs/1802.03426. (2018).
42 Villa, E. *et al.* Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1063-1068 (2009).
43 Hall, S. R., Allen, F. H. & Brown, I. D. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A: Foundations of Crystallography* **47**, 655-685 (1991).
44 Fernandez, J.-J. & Li, S. TomoAlign: A novel approach to correcting sample motion and 3D CTF in CryoET. *J. Struct. Biol.* **213**, 107778 (2021).
45 Mastronarde, D. N. & Held, S. R. Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102-113 (2017).
46 Xiong, Q., Morphew, M. K., Schwartz, C. L., Hoenger, A. H. & Mastronarde, D. N. CTF determination and correction for low dose tomographic tilt series. *J. Struct. Biol.* **168**, 378-387 (2009).
47 Mastronarde, D. N. Accurate, automatic determination of astigmatism and phase with Ctfplotter in IMOD. *J. Struct. Biol.*, 108057 (2024).
48 Luo, Z., Campos-Acevedo, A. A., Lv, L., Wang, Q. & Ma, J. Sparseness and Smoothness Regularized Imaging for improving the resolution of Cryo-EM single-particle reconstruction. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013756118 (2021). https://doi.org:10.1073/pnas.2013756118
49 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *International Conference for Learning Representations* (ICLR, 2015).

50 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* **27**, 14-25 (2018).
51 Hrabe, T. *et al.* PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* **178**, 177-188 (2012).

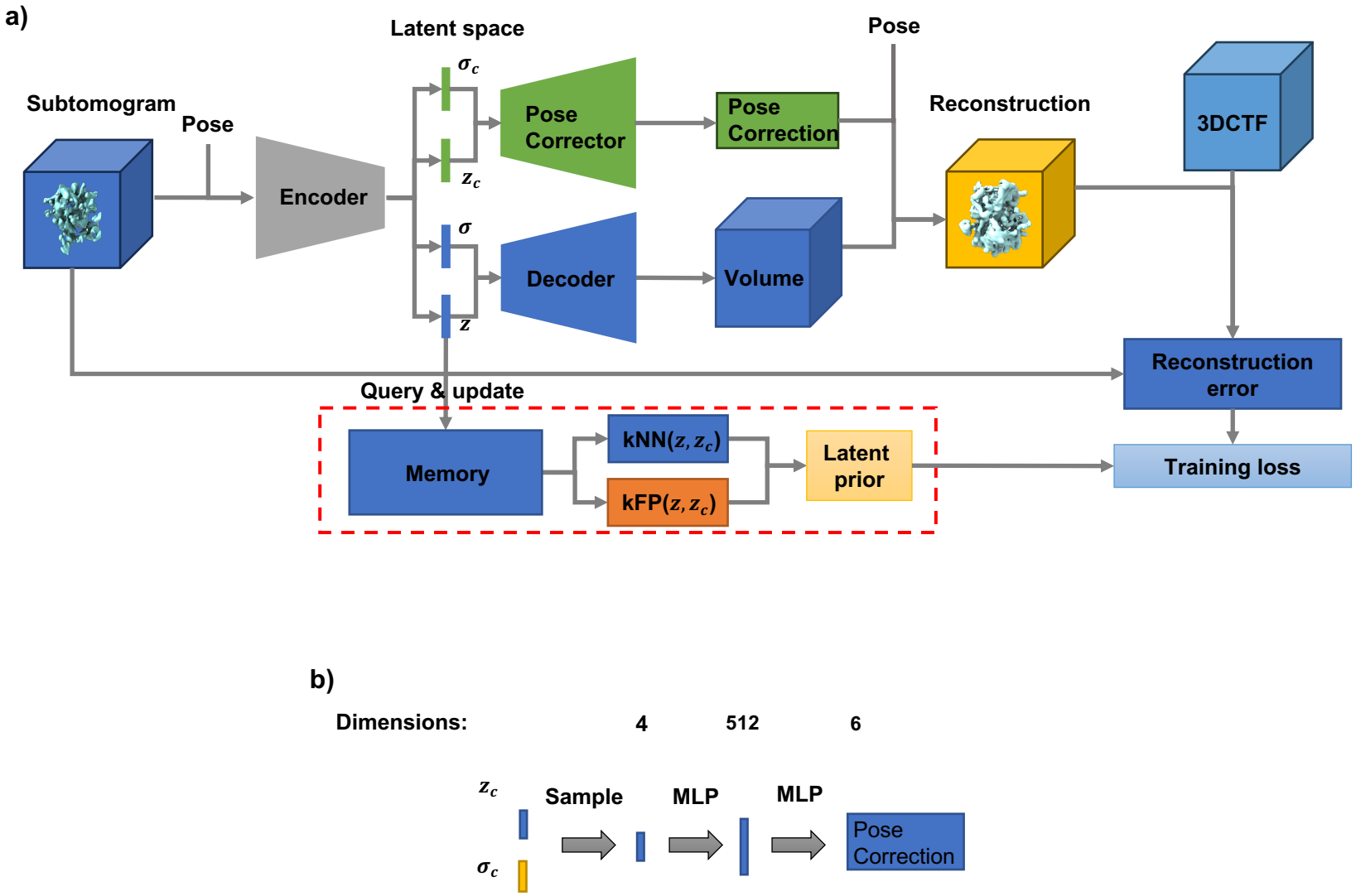

a)
Latent space
Pose
Subtomogram
Pose
Encoder
σ_c
z_c
σ
z
Pose Corrector
Pose Correction
Decoder
Volume
Reconstruction
3DCTF
Reconstruction error
Query & update
Memory
kNN(z, z_c)
kFP(z, z_c)
Latent prior
Training loss
b)
Dimensions:
4
512
6
z_c
σ_c
Sample
MLP
MLP
Pose Correction

Fig. 1

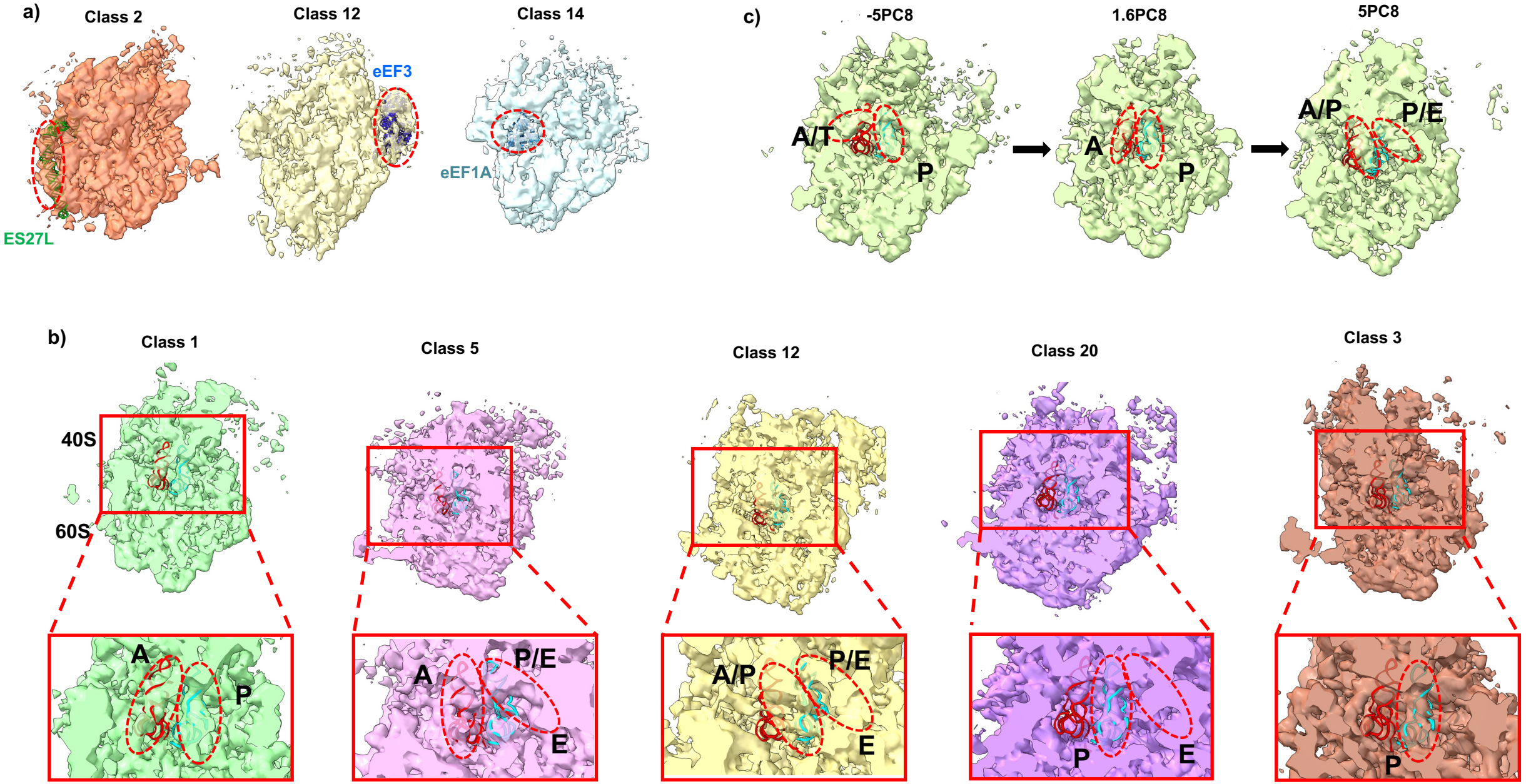

a)
Class 2
Class 12
Class 14
eEF3
eEF1A
ES27L
c)
-5PC8
1.6PC8
5PC8
A/T
P
A
P
A/P
P/E
b)
Class 1
Class 5
Class 12
Class 20
Class 3
40S
60S
A
P
A
P/E
E
A/P
P/E
E
P
E
P

Fig. 2

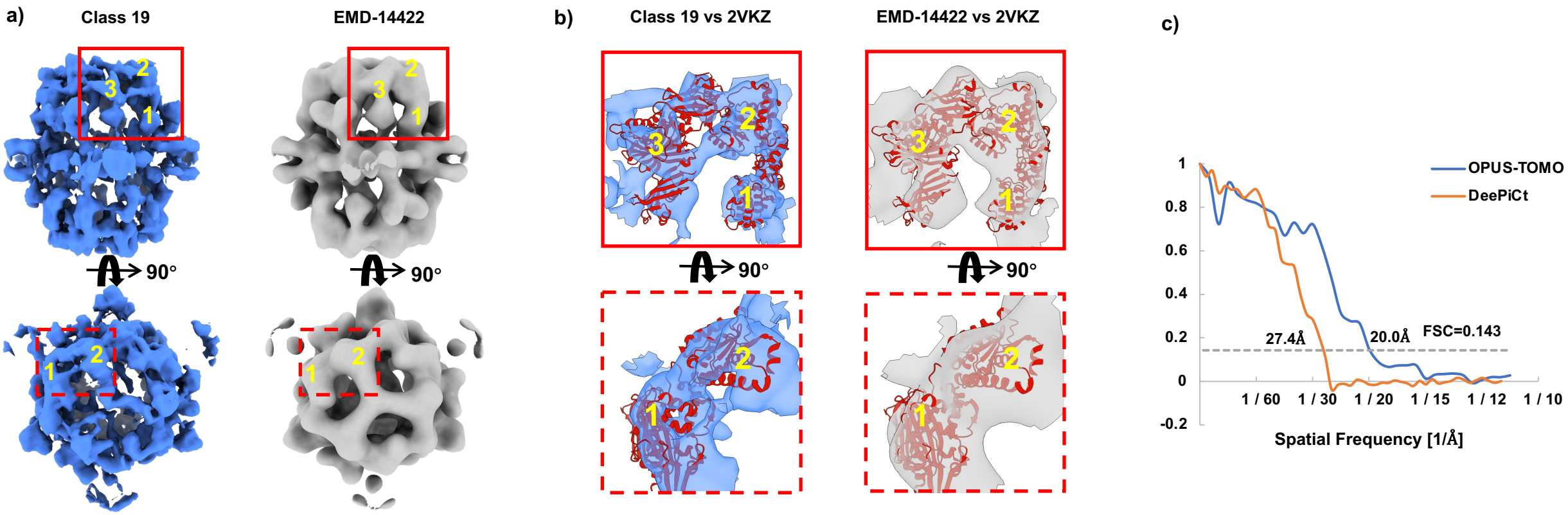

a)
Class 19
EMD-14422
90°
90°
b)
Class 19 vs 2VKZ
EMD-14422 vs 2VKZ
90°
90°
c)
OPUS-TOMO
DeePiCt
1
0.8
0.6
0.4
0.2
0
-0.2
27.4Å
20.0Å
FSC=0.143
1 / 60
1 / 30
1 / 20
1 / 15
1 / 12
1 / 10
Spatial Frequency [1/Å]

Fig. 3

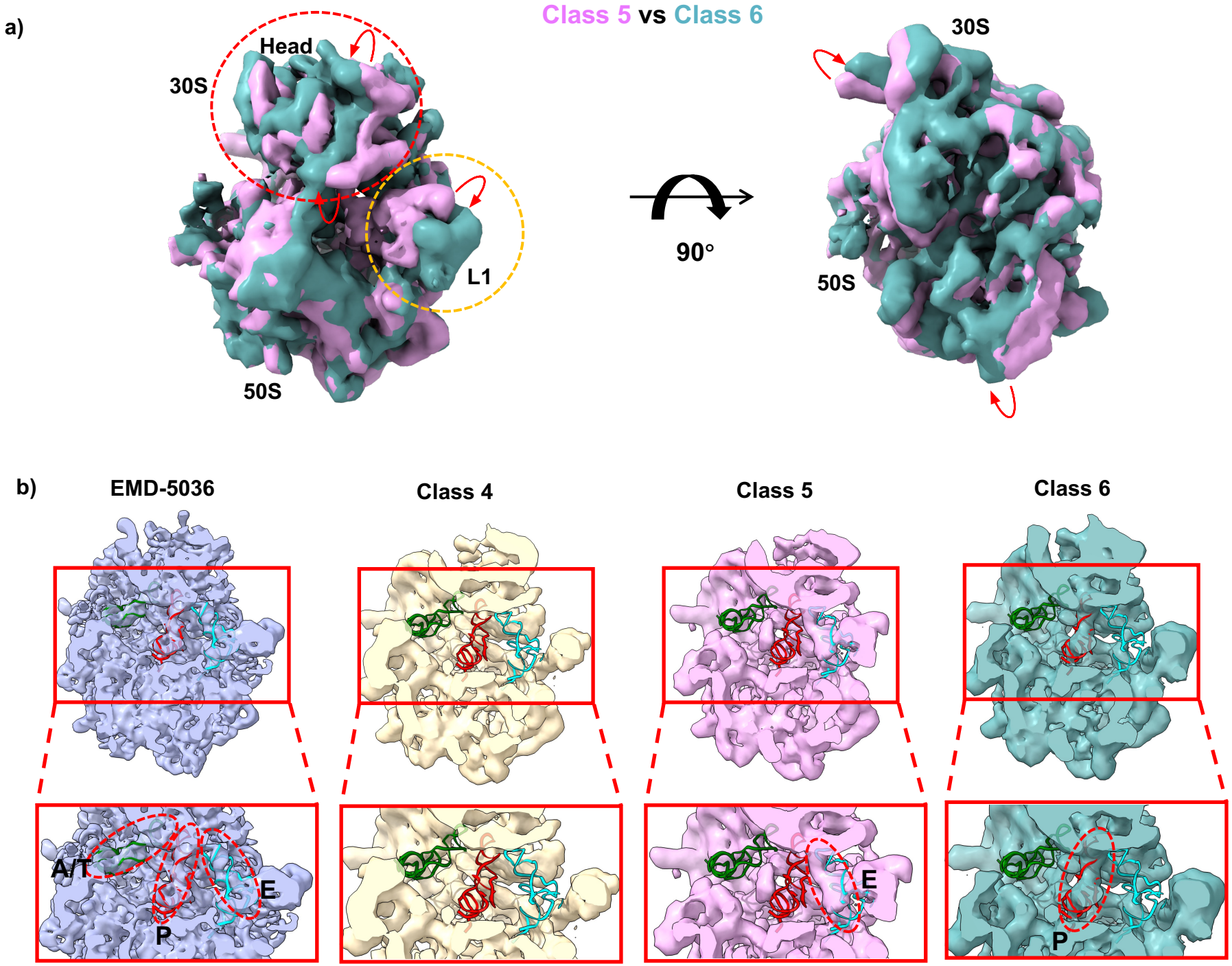
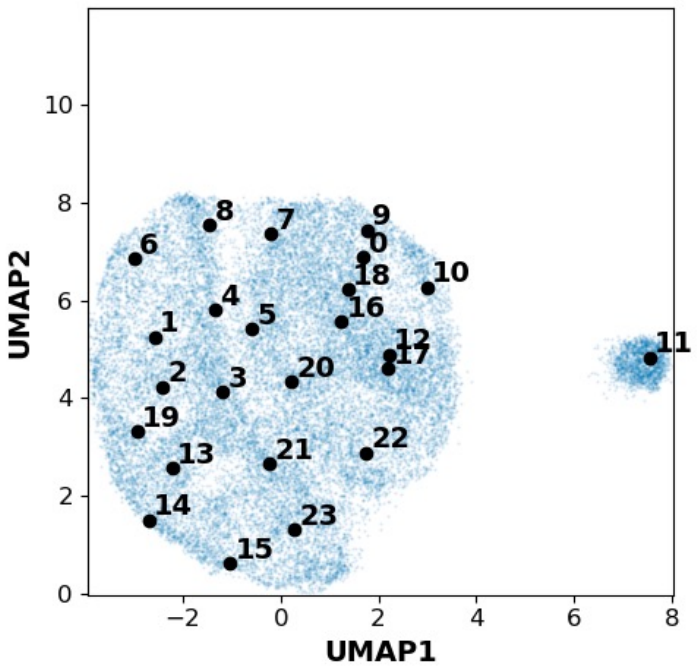
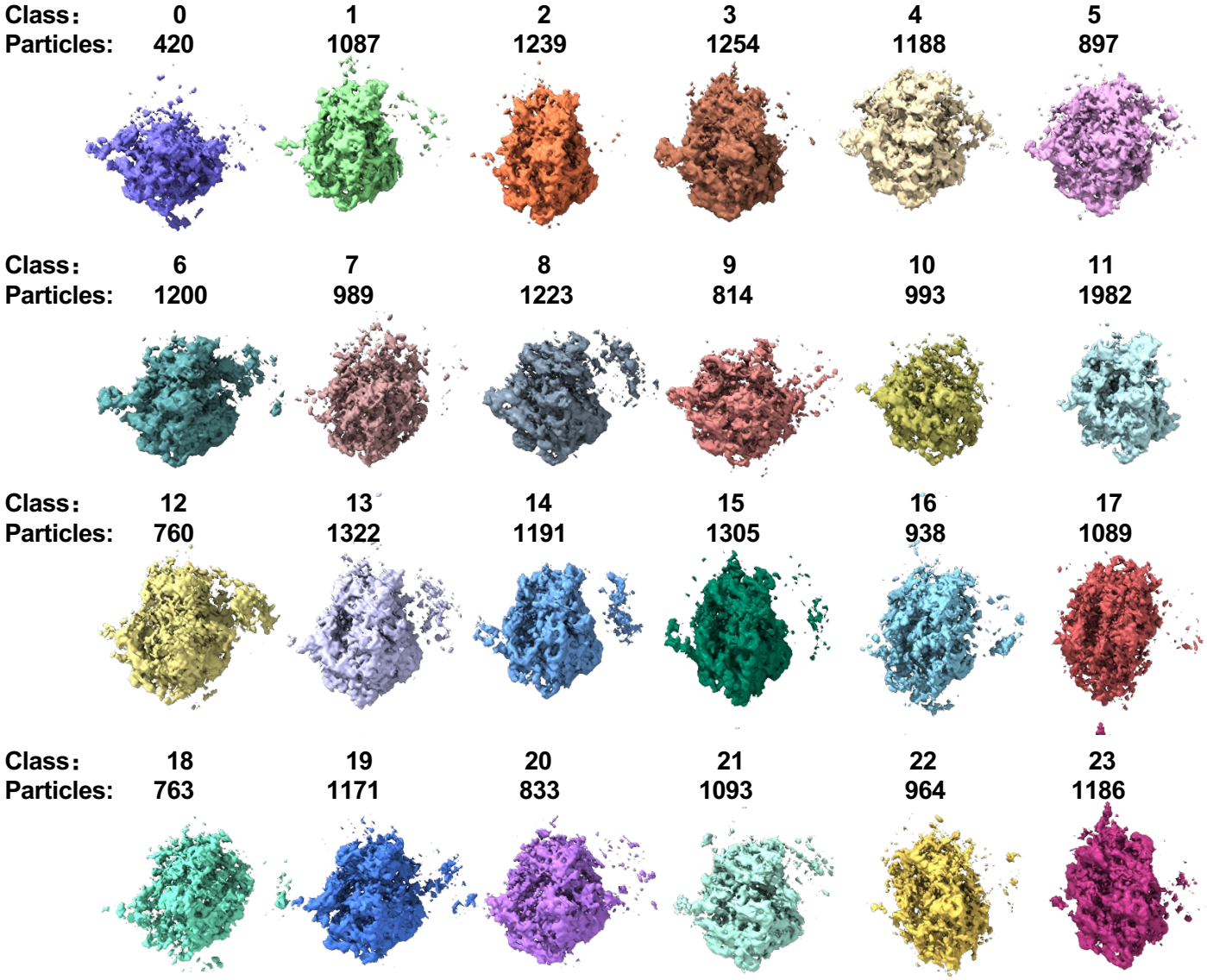

a)
Class 5 vs Class 6
Head
30S
L1
50S
90°
30S
50S
b)
EMD-5036
Class 4
Class 5
Class 6
A/T
E
P
E
P

Fig. 4

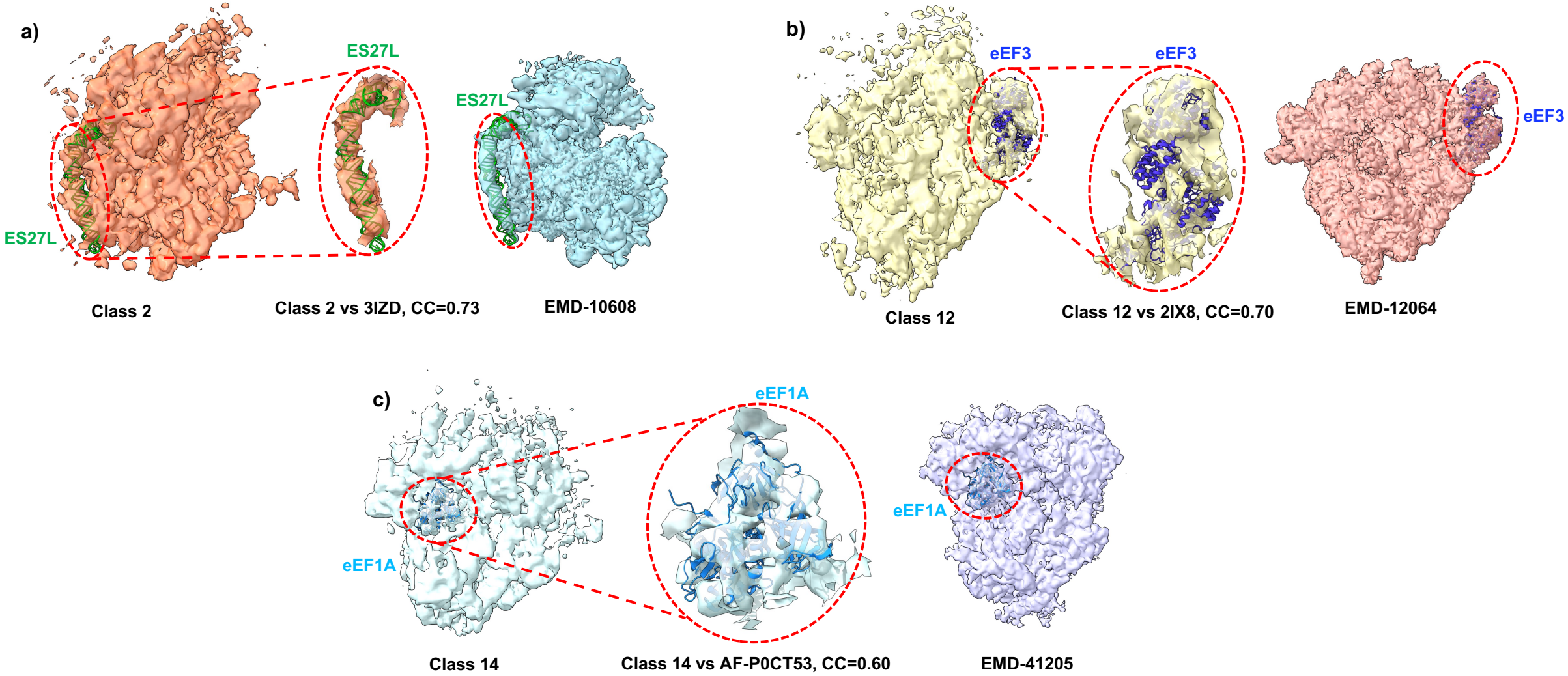
a)



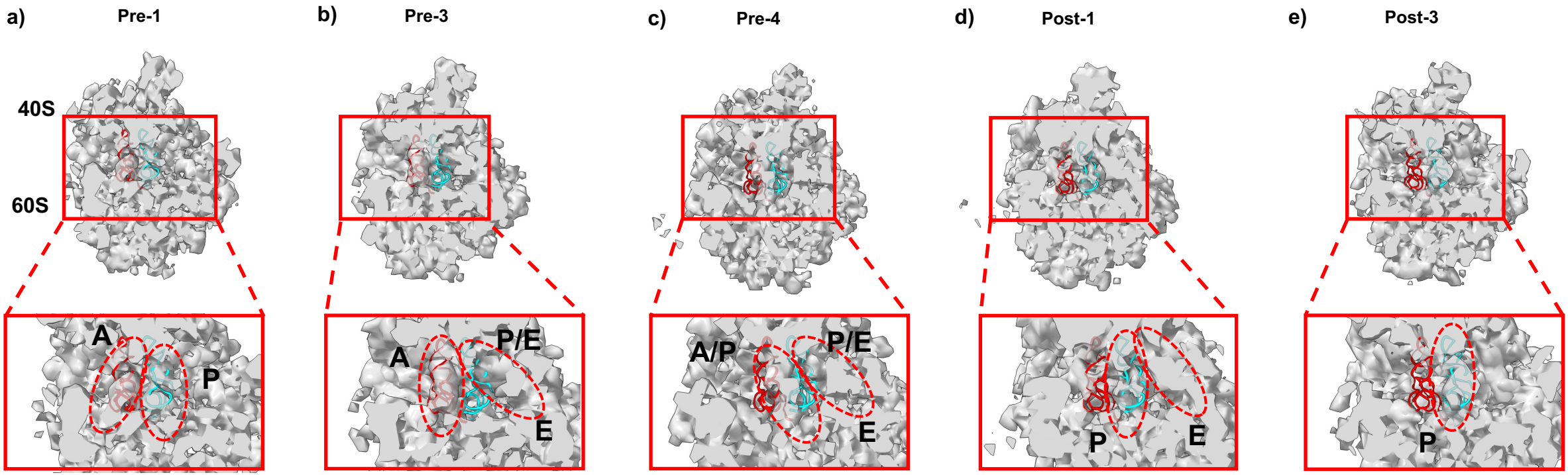
b)



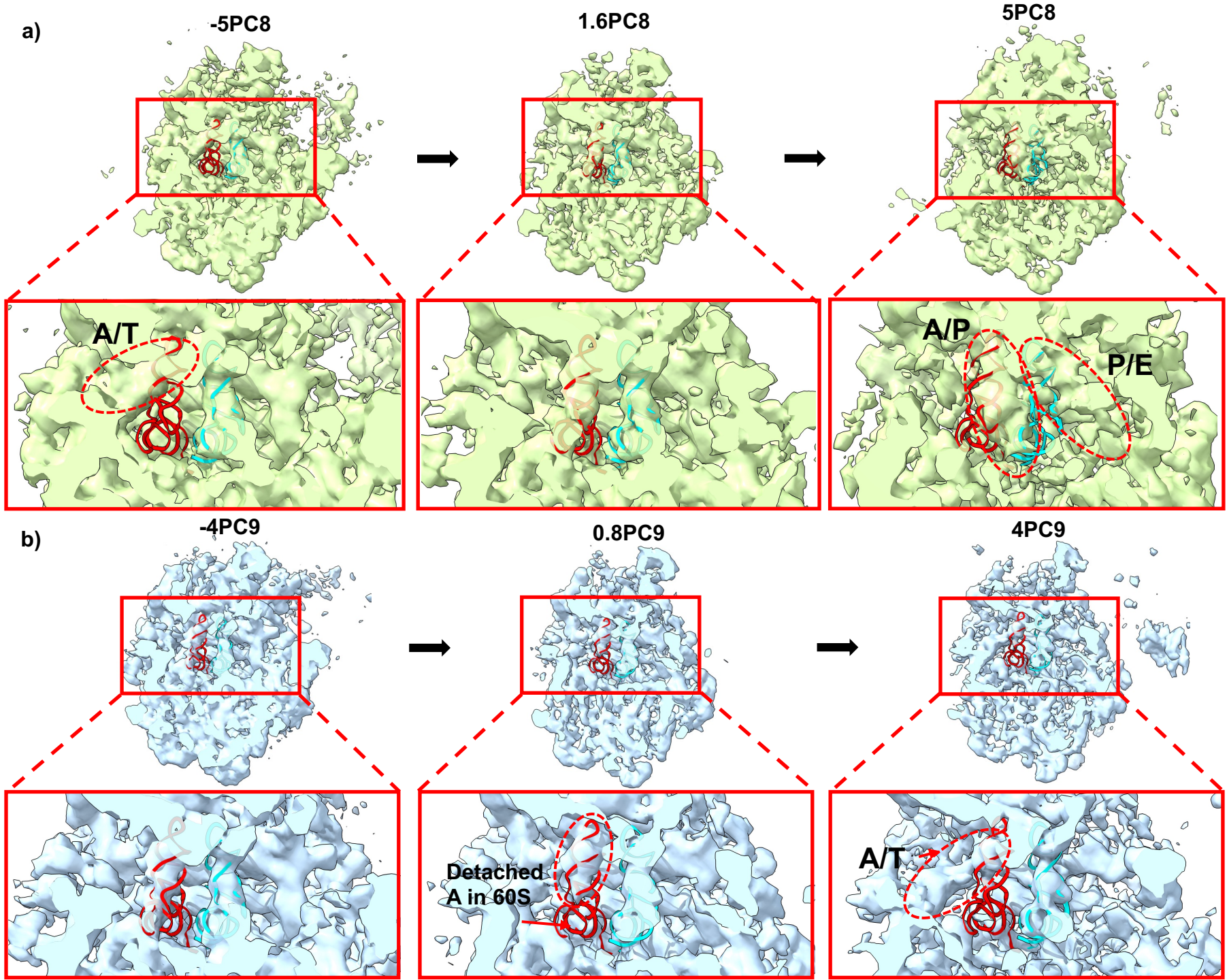
Extended Data Fig. 1



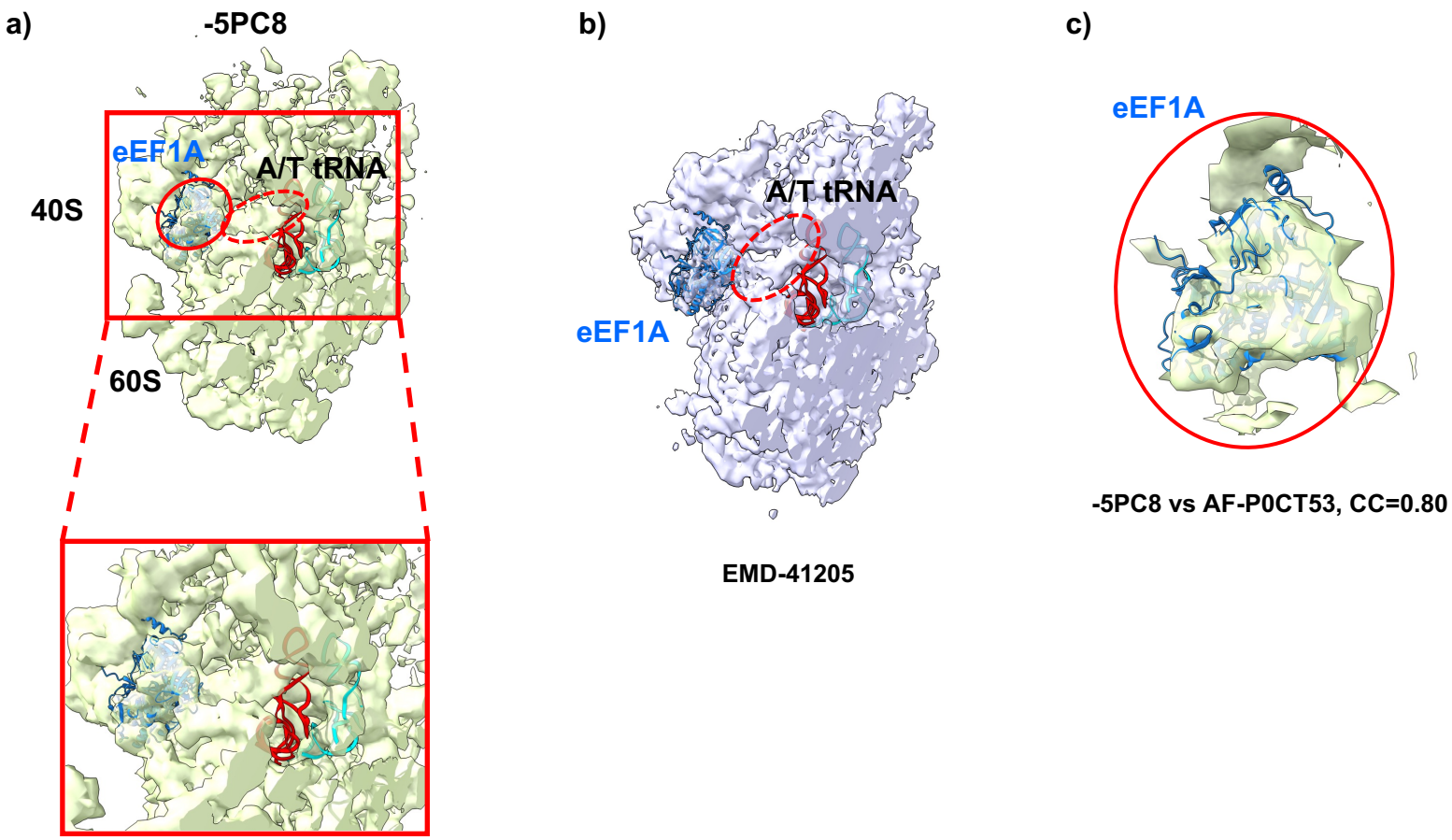
Extended Data Fig.2



Extended Data Fig.3

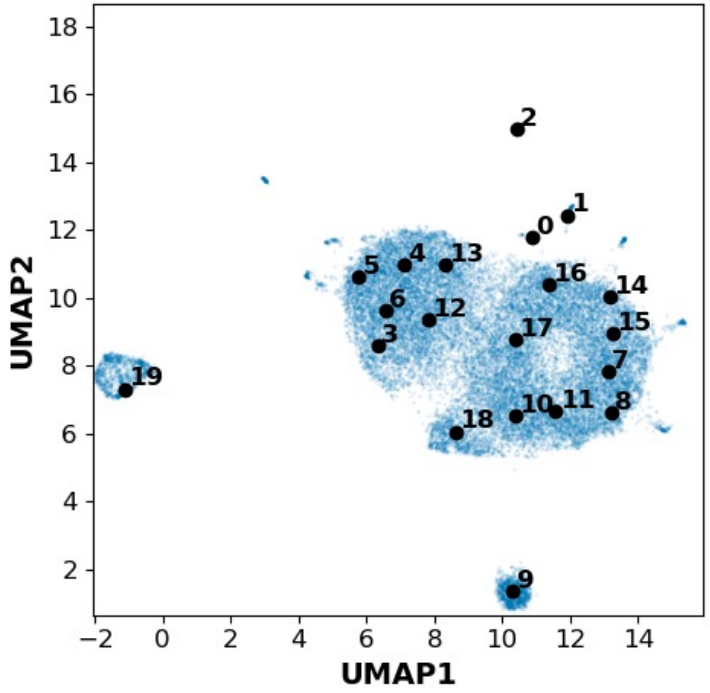


Extended Data Fig.4

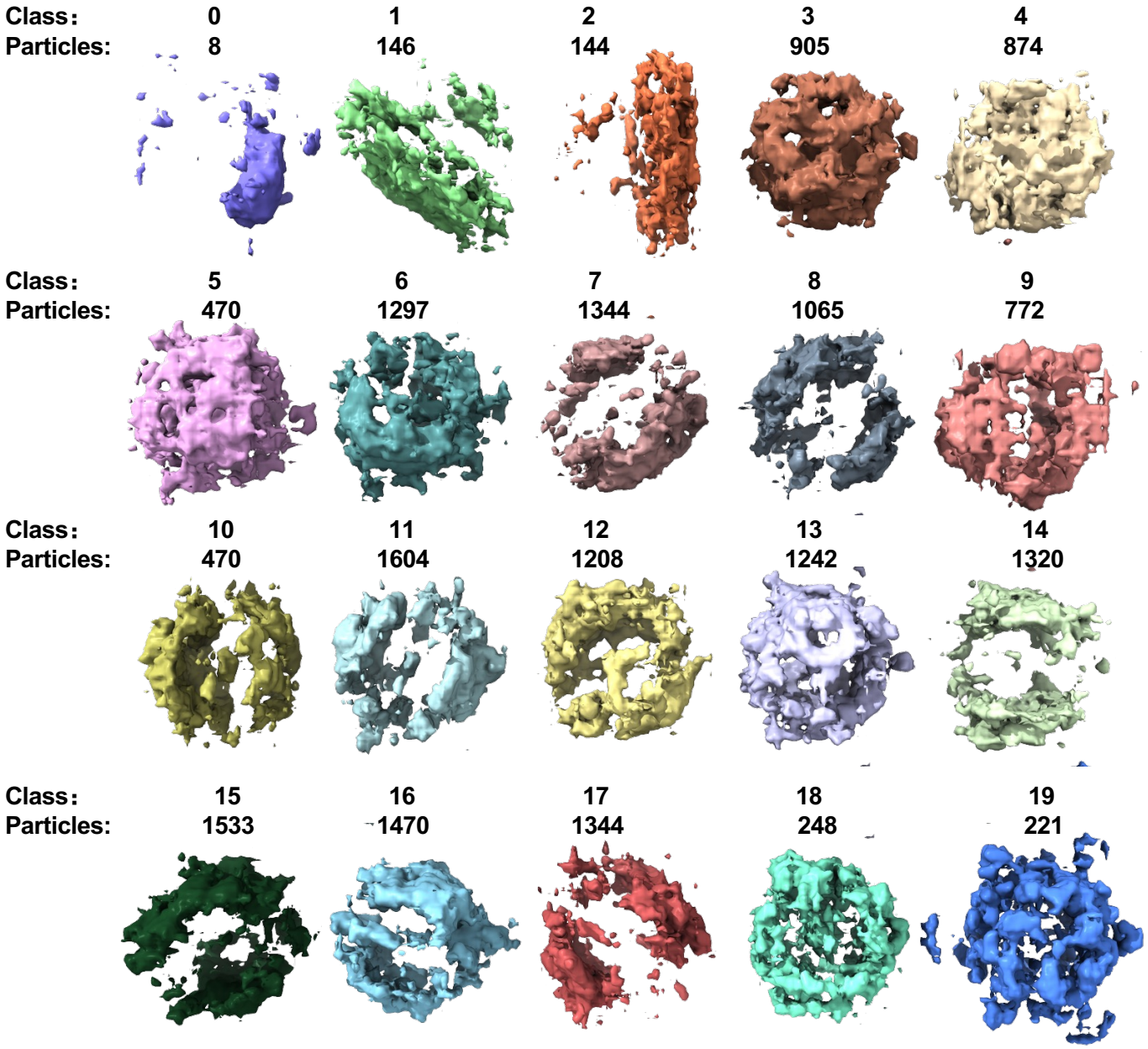


Extended Data Fig.5

a)

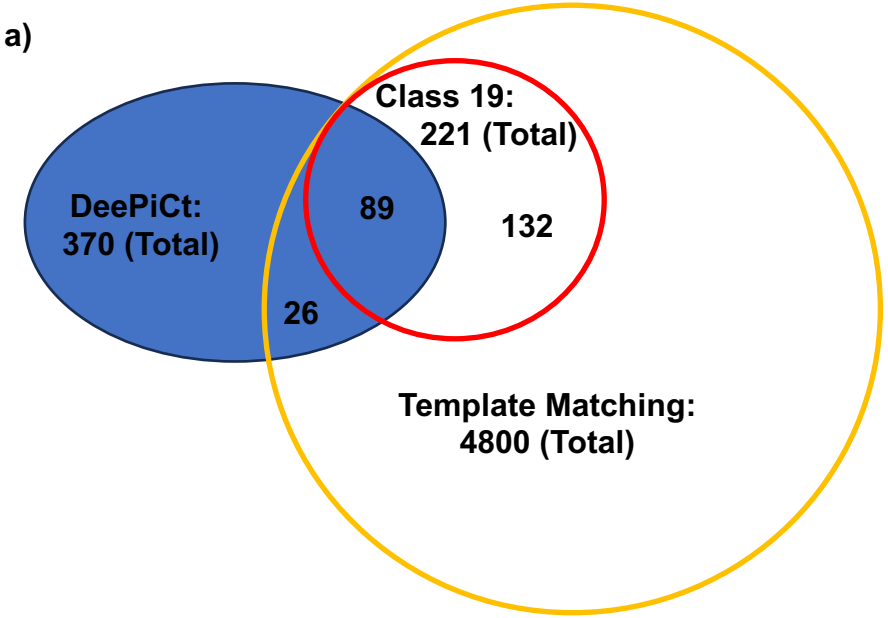


b)

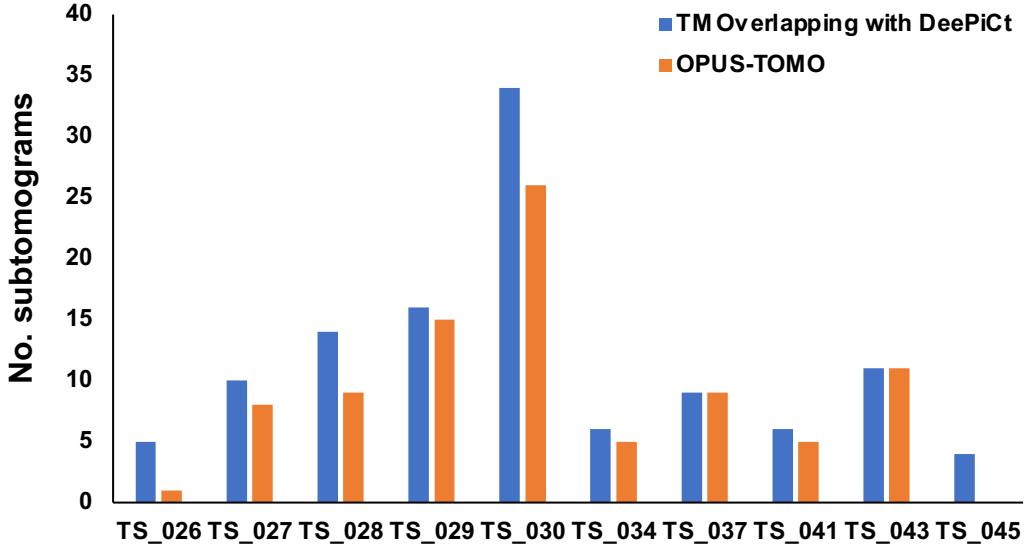


Extended Data Fig. 6

a)

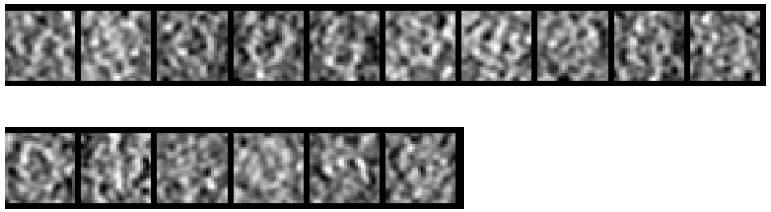


b)

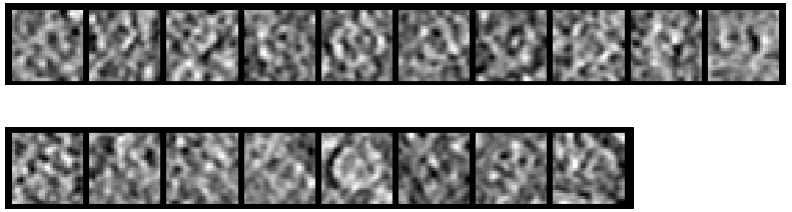


c)

Tilt Series: TS_045



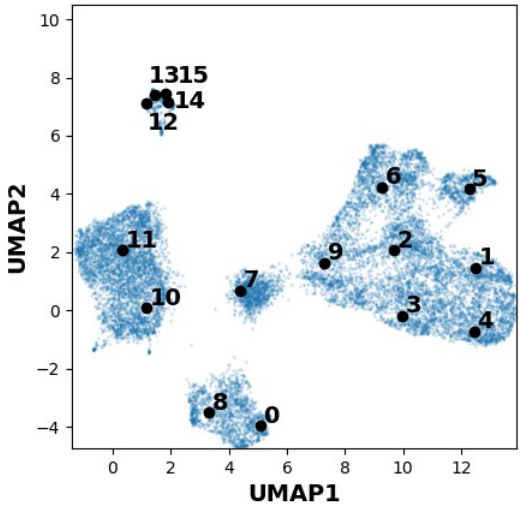
16 subtomograms selected by OPUS-TOMO (Class 19)



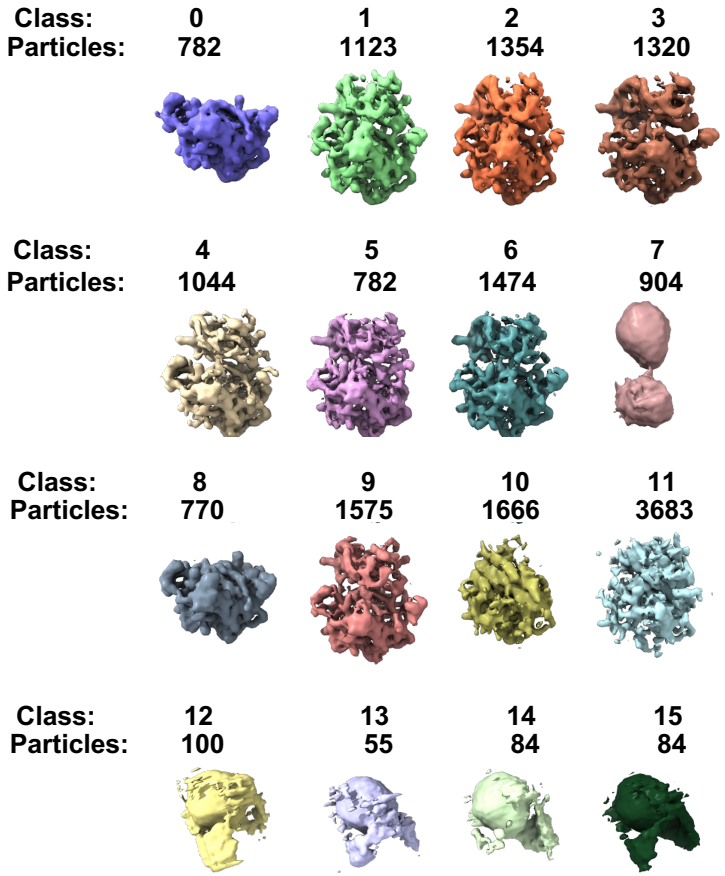
18 subtomograms selected by DeePiCt

Extended Data Fig.7

a)



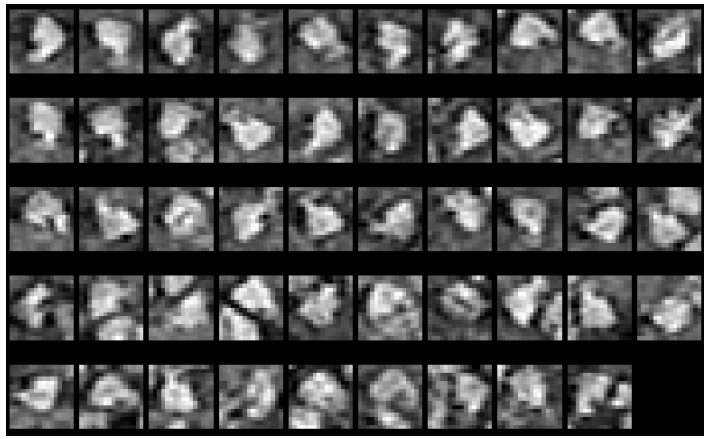
b)

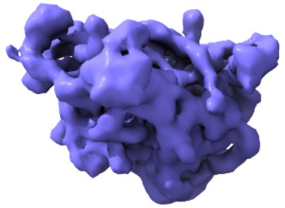


Extended Data Fig.8

a)

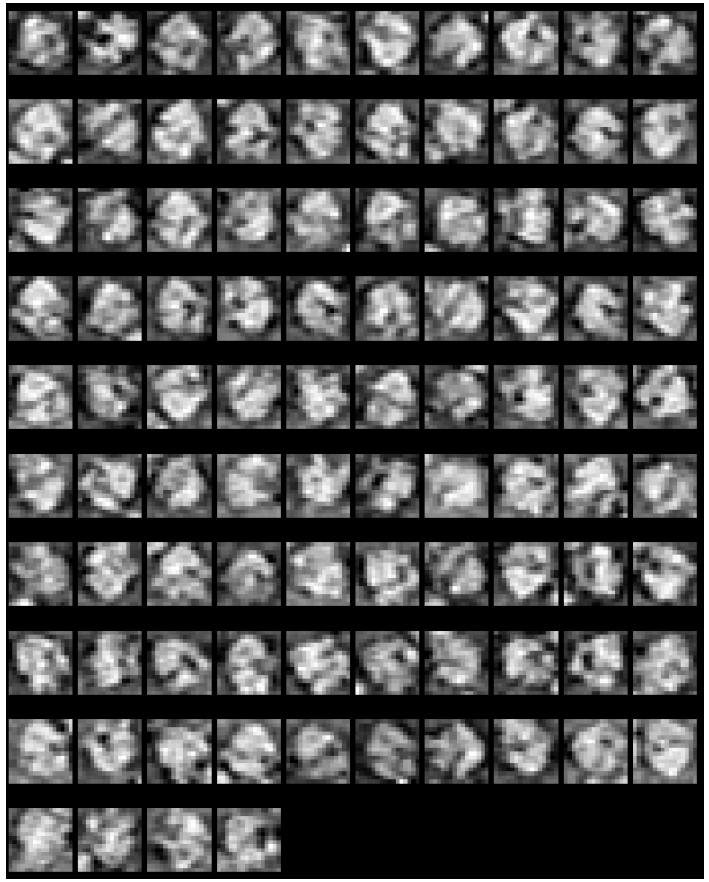
Class 0

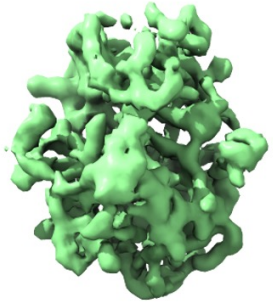




b)

Class 1





Extended Data Fig.9

Tilt series	Template matching	DeePiCt	Template matching overlapping with DeePiCt	Class 19	Class 19 overlapping with DeePiCt
TS_026	400	16	5	3	1
TS_027	400	56	10	10	8
TS_028	600	76	14	46	9
TS_029	400	46	16	26	15
TS_030	400	64	34	34	26
TS_034	400	37	6	19	5
TS_037	600	22	9	17	9
TS_041	400	18	6	30	5
TS_043	600	17	11	20	11
TS_045	600	18	4	16	0
Sum	4800	370	115	221	89

Extended Data Table 1