OPUS-DSD2: Disentangling Dynamics and Compositional Heterogeneity for Cryo-EM Single Particle Analysis

Zhenwei Luo[1,2,3], Yiqiu Zhang[3], Gang Xu[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: Neural volumes, structural heterogeneity, self-supervised learning, dynamics

Abstract

Cryo-electron microscopy (cryo-EM) datasets preserves a wide range of structural heterogeneities in macromolecules, arising from both large-scale dynamics and compositional changes. The deep learning framework OPUS-DSD has been developed to characterize the structural heterogeneities in cryo-EM data. However, different types of variations often entangle together in its results. To address this, here we propose OPUS-DSD2, an algorithm which incorporates dynamics model into OPUS-DSD to enhance the disentanglement of large-scale dynamics and compositional differences. OPUS-DSD2 retains the convolutional decoder from OPUS-DSD to represent compositional changes and fine-scale dynamics through 3D volumes. Additionally, it introduces a dynamics decoder based on multi-layer perceptron to model large-scale dynamics via deformation fields. OPUS-DSD2 employs separate latent spaces for training these decoders. Comparative tests on real datasets have shown that OPUS-DSD2 effectively separates dynamics and compositional changes into distinct latent spaces, and encodes compositional changes in different regions of macromolecules to different principal components of composition latent space. OPUS-DSD2 integrates with the original OPUS-DSD framework, and is available at https://github.com/alncat/opusDSD.

Introduction

Cryo-electron microscopy (cryo-EM) has emerged as a powerful tool for structural determination of biological macromolecules, providing unprecedented insights into the complex world of cellular machinery[1]. However, cryo-EM still faces many challenges which limit its

resolving power. One of the major challenges in cryo-EM data analysis is the structural heterogeneity inherent in macromolecules. This heterogeneity, which arises from the inherent flexibility of macromolecules[2,3], and the existence of multiple functional states can significantly complicate the interpretation of cryo-EM datasets and hinder the extraction of high-resolution structural information. Therefore, the development of methodologies to resolve structural heterogeneity is of paramount importance in the field of structural biology[1]. The challenge is to discern and segregate different conformations or states present in the sample, which mixes in the collected data. Traditional methods for addressing structural heterogeneity in cryo-EM datasets often involve computationally intensive classification and particle subtraction procedures[4-8], which can be both time-consuming and prone to error. Moreover, these methods often require substantial human intervention and can only produce finite number of states, making them less suitable for high-throughput cryo-EM studies and modelling continuous structural changes.

In recent years, deep learning has shown great promise in addressing complex problems across a wide range of fields due to its great modelling capacity[9]. In the context of cryo-EM, deep learning methods have the potential to provide automated solutions for handling structural heterogeneity. Various types of deep learning approaches which represent 3D structural variations using neural networks have been proposed. First of all, there are end-to-end neural networks which can translate 2D images into 3D volumes, such as cryoDRGN[10] and OPUS-DSD[11]. Though both methods capture structural variations using 3D volumes directly, cryoDRGN uses neural network to represent the Fourier transform of the 3D volume, whereas OPUS-DSD uses neural network to represent the 3D volume using 3D convolutional neural network (3DCNN) directly. Both methods have shown the capacity of resolving discrete and continuous structural heterogeneities. Other types of methods include Multi-CryoGAN[12],

e2gmm[13], and 3DFlex in cryoSPARC[14]. Multi-CryoGAN is a method based on generative adversarial network[15]. E2gmm uses a set of gaussians to model the 3D structure. 3DFlex assumes each particle is generated from a common structure according to its deformation field which can be modelled by multi-layer perceptron (MLP)[16].

The 3D structural variations in cryo-EM datasets can be mainly categorized into two types according to their origins. The first type of variation is the compositional changes, which are caused by the binding or disassociation of subunit in the macromolecular complex. The second type of variation results from the dynamics of complex, where the relative positions and orientations of subunits in the complex undergo dynamical changes. OPUS-DSD[11] describes both kinds of structural variations within a single latent space using the 3D volume representation. It leveraged various priors to encourage the smoothness of latent space and improve the quality of 3D reconstruction. Though it can successfully resolve various structural heterogeneities on real datasets, different types of variations are entangled in its results. This paper introduces the next generation of OPUS-deep structural disentanglement (OPUS-DSD2), which extends the work of OPUS-DSD.

OPUS-DSD2 strives to overcome the drawbacks of 3D volume representation in OPUS-DSD. The 3D volume represents 3D structure through the distribution of voxel values over grid. To represent the dynamics of a 3D structure, a series of volumes with continuous changes of voxel values should be reconstructed. The large-scale movement of a 3D structure often incurs a large number of voxel changes which make the volume representation become ineffective. Hence, it is necessary to find alternative parameterizations which can greatly reduce the number of free parameters to improve the modelling efficiency and accuracy. Fortunately, the dynamics of macromolecule often follows the trajectories of one or a few low-frequency normal modes[17].

A common approach is decomposing dynamics of macromolecules as the rigid-body movements of multiple subunits, which was adopted by the multi-body refinement of RELION for cryo-EM data[18], and the multi-group TLS method in crystallographic refinement[19]. The parameters for modelling the dynamics of macromolecules comes down to the rigid-body movement parameters of subunits. Another parameterization of dynamics is proposed by the 3DFlex in cryoSAPRC[14], which approximates the displacements of voxels using neural network. By leveraging the modelling capacity of MLPs, it also successfully models continuous dynamics of 3D structures in cryo-EM. Besides, Calero *et al.* used Zernike polynomials and spherical harmonics as basis for fitting the dynamics of macromolecules in cryo-EM dataset[20,21].

OPUS-DSD2 represents the compositional changes and dynamics using separate models with their dedicated latent encodings. Inspired by the designs of Spatial Transformer[22] and Neural Volumes[23], we propose to explicitly disentangle structural heterogeneity as composition and dynamics. The composition is modelled using volume, while the dynamics is modelled using spatial transformation defined by deformation field. Spatial transformer can produce the 3D structure with specific variations by applying the spatial transformation on the volume. Hence, it can capture the structural landscape while explicitly separating different heterogeneities. To accommodate the independence of different variations, OPUS-DSD2 encode them into two separate latent spaces. It also leverage different models which can accurately characterize associated variations to resolve underlying heterogeneities of cryo-EM data.

In summary, we present a structural heterogeneity analysis method which harnesses the power of neural-network-based real space representation. It can explicitly disentangle the compositional changes and large-scale dynamics from cryo-EM macromolecular datasets. In the **Results** section, we provide an overview of our approach, and delve into the architecture of

OPUS-DSD2 for learning latent spaces pertaining to composition and dynamics. Our systematic testing on real cryo-EM datasets showcases structural heterogeneity resolving result of much higher quality compared to the previous version.

Results

Network Architecture

OPUS-DSD2 introduces a dynamics decoder which models dynamics of macromolecules into the neural network architecture of OPUS-DSD[11]. It converts a 2D cryo-EM image into a 3D volume as following. The encoder takes a 2D cryo-EM image as input, and estimates the distributions of the composition latent code in a n-dimensional space and the dynamics latent code in an m-dimensional space. The two latent codes are assumed to follow Gaussian distributions, whose means $z \in \mathbb{R}^n$ and standard variations $\sigma \in \mathbb{R}^n$ are estimated by encoder[24]. The two latent codes are sampled from estimated distributions and supplied into corresponding decoders. The composition latent code is supplied into the composition decoder to generate a 3D volume (**Fig.1a**), which is represented by a discrete grid of voxels, $V(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathbb{R}^3$ denotes a grid point in 3D space. Meanwhile, the sampled dynamics latent code is supplied into the dynamics decoder to generate a deformation field (**Fig.1a**), which is represented by 3D vectors on a discrete grid, i.e., $\boldsymbol{g}(\boldsymbol{x}): \mathbb{R}^3 \rightarrow \mathbb{R}^3$. A 3D vector at a grid point of the deformation field records its source grid point where the output voxel value for this point comes from. More specifically, the deformation field is assumed to be generated by the rigid-body dynamics of a set of subunits as it is described in the **Deformation field** subsection in **Methods**. The parameters

for rigid-body movement of each subunit are predicted by the dynamics decoder. Finally, spatial transformer[22] is leveraged to apply the spatial transformation defined by deformation field $\boldsymbol{g}$ to the 3D volume $V(\boldsymbol{x})$, producing a deformed 3D volume. The deformed 3D volume is transformed into a 2D reconstruction with specified pose and CTF parameters following the differentiable cryo-EM image formation model. The neural architecture of OPUS-DSD2 can be trained end-to-end by reconstructing each input image and minimizing the squared reconstruction error.

During training, to encourage the smoothness of latent spaces, the composition and dynamics latent codes are regularized by the structural disentanglement prior[11] from OPUS-DSD and Kullback-Leibler (KL) divergence with respect to (w.r.t) standard gaussian[24,25] as in variational autoencoder (VAE). The composition decoder is the same as in OPUS-DSD[11]. The dynamics decoder is rather simple and consists of a series of MLP and nonlinearities (**Fig.1b**). The dynamics latent code is transformed into a 512-dimensional representation by an MLP followed by LeakyReLU nonlinearity. The 512-dimensional representation is projected to a $N_{\text{body}} \times 6$-dimensional vector by an MLP, where N_{body} denotes the number of rigid bodies defined for the macromolecule. The $N_{\text{body}} \times 6$-dimensional vector is reshaped into N_{body} 6-dimensional vectors, each of which defines the rotation and translation parameters for a subunit incurred by dynamics. The first three dimensions of the 6-dimensional output are used to construct the quaternion to represent the rotation of the subunit, while the last three dimensions of the 6-dimensional output are used to determine the translation of the subunit.

Pre-catalytic spliceosome

OPUS-DSD2 is used to analyze the structural heterogeneity of the pre-catalytic spliceosome (EMPIAR-10180)[26]. We trained a 16-dimensional latent variable model, where 12 dimensions are used for composition latent space and the remaining 4 dimensions are kept for dynamics latent space. The consensus refinement result for model training is downloaded from EMPIAR-10180[27]. The subunits for defining the deformation field are segmented according to the masks deposited in EMPIAR-10180[18]. We first compared the reconstruction qualities of the volumes from the 12-dimensional model of OPUS-DSD2 to the volumes from the model of OPUS-DSD with the same number of dimensions. Distinct conformations are revealed by principal component analysis (PCA) on the composition latent space. At the positive end of the first principal component (PC1) of OPUS-DSD2 and PC1 of OPUS-DSD, the reconstructed template volume corresponds to an 'open' conformation where the SF3b subunit stays on the core body, while the reconstructed volume at the negative ends of PCs corresponds to a 'closed' conformation where SF3b subunit folds onto the back of core body. **Fig.2a** shows the 'open' conformation reconstructed by OPUS-DSD2 and OPUS-DSD. Contoured at the same level, the reconstruction of OPUS-DSD2 shows clearly separated helices in front view in the top panel. In contrast, the reconstruction of OPUS-DSD presents blurred densities at the same regions (**Fig.2a**). The difference become more evident in the enlarged views inside the boxed regions. The reconstruction from OPUS-DSD2 shows more resolved structural folds in the SF3b subunit (**Fig.2a**). **Fig.2b** shows the 'closed' conformation reconstructed by OPUS-DSD2 and OPUS-DSD. Similarly, in the front view, the reconstruction from OPUS-DSD2 shows clearer and more continuous structures at both Core and Helicase subunits (**Fig.2b**). In the rear view, the improved structural features in the reconstruction of OPUS-DSD2 compared to the reconstruction from OPUS-DSD are noticeable (**Fig.2b**). Traversal along different PCs in the

composition latent space of OPUS-DSD2 corresponding different changes of components in structures. We compared the traversals along PC3 and PC5 side by side, and observed emergences of two different structural elements between the Foot and the Helicase (**Supplementary Video 1**).

We also compared the reconstruction qualities of different methods on the KMeans[28] centroids (**Fig.2c** and **Fig.2d**). The latent spaces of the results of both methods are clustered into 10 classes using KMeans algorithms (**Fig.2c**). The reconstructions from OPUS-DSD2 are of higher resolutions than the reconstruction from OPUS-DSD. Similar reconstructions from the results of different methods are selected for close-up comparison (**Fig.2d**). In regions marked by red dashed boxes, the reconstruction from OPUS-DSD2 presents structural features with higher resolution than reconstructions from other methods (**Fig.2d**).

It's worth noting that the qualities of reconstructions differ at KMeans centers and the ends of PCs. This is caused by that KMeans and PCA are different ways to sample the latent space. The centers from KMeans clustering are in dense regions, while the points on principal components are manually set. Therefore, the reconstructions at centers from KMeans clustering are generally of higher resolution, while the reconstructions corresponding to points at two ends of PCs are of lower resolution owing to the sparseness of the latent encodings around them.

Another important functionality of OPUS-DSD2 is reconstructing dynamics directly by traversing along the PCs of dynamics latent space. We performed PCA on the dynamics latent space. Class 2 from KMeans clustering in the composition latent space was chosen as the template conformation to be deformed. Traversing the first PC reveals the concerted folding of SF3b and Helicase towards Core (**Supplementary Video 2**), and traversing the second PC reveals the folding of Helicase towards Core, while the SF3b rotating around its center

(**Supplementary Video 3**). The dynamics reconstructed by OPUS-DSD2 shows continuous deformation of densities near boundaries between subunits, which is in sharp contrast to the dynamics reconstructed by Multi-body refinement in RELION[18], where densities near boundaries between subunits are broken.

Plasmodium falciparum 80S ribosome

OPUS-DSD2 is used to analyze the structural heterogeneity of *Plasmodium falciparum* 80S (Pf80S) ribosome (EMPIAR-10028)[29]. We trained a 16-dimensional latent variable model, where composition latent space is of 12 dimensions and dynamics latent space is of 4 dimensions. We followed the subunit segmentation scheme in the multibody refinement of Pf80S[18], except that the Head and SSU subunits are merged as a single subunit in our paper. Hence, the deformation field in our analysis is comprised of movements of two subunits. We first compared the reconstruction qualities of volumes from the 12-dimensional model of OPUS-DSD2 to volumes from the model of OPUS-DSD with the same number of dimensions. PCA is performed on both latent spaces. Reconstructions at points on the PC3 from OPUS-DSD2 and the PC1 from OPUS-DSD are compared. We denoted the conformations at the positive end of PCs, with magnitude being 1, as 'POS'. At the same contour level, the reconstructions from both methods have similar compositions at this point (**Fig.3a**). However, in both viewing directions, the reconstruction from the result of OPUS-DSD2 shows higher resolution structures as exemplified by the enlarged views inside dashed boxes (**Fig.3a**). For example, the separation between parallel helices is more evident in the reconstruction from OPUS-DSD2 (**Fig.3a**). Next, we denoted the conformations at the center of PCs, with magnitude being 0.1, as 'NEU' (**Fig.3b**).

The reconstruction by OPUS-DSD2 still demonstrates superior qualities (**Fig.3b**). By traversing along different PCs of the latent space from OPUS-DSD2, different patterns of compositional changes were revealed (**Supplementary Video 4**). The PC1 mainly encodes the compositional changes within the Head in SSU, that is, several RNAs go from presence to missing during the twisting of Head (**Supplementary Video 4**). The PC3 mainly encodes the compositional changes of an RNA strand connecting SSU and LSU. The RNA strand slides from SSU to LSU when traversing along the PC3 (**Supplementary Video 4**).

We compared the qualities of reconstructions on centroids from KMeans clustering in latent spaces of both methods. Both latent spaces are clustered into 10 classes using KMeans algorithms. Volumes are generated by supplying the centroids into the decoders of OPUS-DSD2 and OPUS-DSD, and are visualized at the same contour level in **Fig.3c**. The reconstructions by OPUS-DSD2 consistently present more structural details. In the selected reconstructions on KMeans centroids, the reconstruction by OPUS-DSD2 shows more separated structural elements. (**Fig.3d**).

We performed PCA on the dynamics latent space, and visualized the learned multi-body dynamics by traversing along PCs. Class 4 from KMeans clustering was chosen as the starting conformation to be deformed. The relative rotation between the LSU and SSU can be revealed by traversing the second PC of the dynamics latent space (**Supplementary Video 5**).

SARS-CoV-2 spike protein trimer

OPUS-DSD2 is used to analyze the structural heterogeneity of a SARS-CoV-2 spike (S) protein trimer (EMPIAR-10492)[30]. S protein trimer is comprised of two regions, S1 and S2

(Fig.4a). S1 region can be further divided into two types of subunits, N-terminal domain (NTD) and C-terminal domain (CTD) (**Fig.4a**). S1 CTDs serve as the receptor-binding domain, and is found to present two different compositions in this dataset by previous study[30], which have three closed CTDs and two closed CTDs, respectively. Using the consensus refinement deposited in EMPIAR-10492, we trained a 16-dimensional latent variable model, where the composition latent space is of 12 dimensions and the dynamics latent space is of 4 dimensions. S trimer is segmented to 4 subunits as it is shown in **Fig.4a**. KMeans clustering was performed on the composition latent space to classify particles into 16 classes (**Fig.4b**). The reconstructions produced by the decoder of OPUS-DSD2 at cluster centers reveal different structures of S trimer, some of which show three closed CTDs while some have one CTD missing (**Fig.4c**). Two orthogonal movement modes can be found in the dynamics reconstructed by OPUS-DSD2 for S trimer. Traversing the first principal component (PC1) of the dynamics latent space shows two S1 subunits squeezing towards each, which leads to the close of S trimer (**Supplementary Video 6**). Traversing the third principal component of the dynamics latent space reveals three S1 subunits expanding to outer space, which leads to the open of S trimer (**Supplementary Video 6**). The dynamics within S1 NTD and S2 are embedded in the composition latent space. For example, traversing the third principal component of composition latent space corresponds to the open and close of NTD, and the elongation and contraction of the helix bundle in S2. (**Supplementary Video 7**).

Discussion

In this paper, we presented a method, OPUS-DSD2, for resolving structural heterogeneity using a deformable volume representation. The OPUS-DSD2 framework marks a significant step forward in addressing structural dynamics and compositional changes in cryo-EM data. By introducing the deformation field that models multi-body movements to the framework of OPUS-DSD, it can disentangle dynamics and compositional changes using neural networks in self-supervised manner. This modelling approach also drastically improves the qualities of reconstruction, and can even determine the unresolved subunit in consensus model to higher resolutions.

An important objective of structural heterogeneity analysis is to disentangle distinctive modes of structural variations and associate them to independent latent factors, thus enabling an easy understanding about structural changes in complex. To achieve this, OPUS-DSD2 uses dual latent spaces to encode dynamics and compositional changes, respectively. In contrast, OPUS-DSD models both using a single latent space. When traversing along principal components of the latent space of OPUS-DSD, these two kinds of variations are often entangled together. In contrast, experiments on real datasets using OPUS-DSD2 demonstrate that the principal components of its composition latent space mainly encode the compositional changes in different regions of the structure, while principal components of its dynamics latent space model different modes of large-scale dynamics. Hence, by letting each latent space capture a specific kind of variation, and by encouraging each component in the latent space to capture different changes, OPUS-DSD2 successfully disentangles the dynamics from compositional changes, greatly increasing the interpretability of the learned latent spaces. The dual-latent-space approach provides a comprehensive solution for handeling the complex variations commonly observed in cryo-EM datasets.

The deformable volume, which consists of a deformation field and a volume, greatly enhances the capacity of OPUS-DSD2 for modelling multi-faceted structural heterogeneities in cryo-EM data. Firstly, the deformation field with our dynamics model can efficiently represent large-scale variations using a small number of parameters. Secondly, as many macromolecular complexes indeed behave as a conglomeration of independently moving rigid bodies, the deformation model of OPUS-DSD2 based on rigid multi-body dynamics is valid and effectively decomposes the complex's overall dynamics[17-19]. Lastly, in contrast to the multi-body dynamics model proposed in RELION[18], the movements of grid points near the boundaries between subunits are continuously blended in the model of OPUS-DSD2. In the dynamics reconstructed by OPUS-DSD2, densities near the boundaries between subunits deform continuously. Consequently, the rigid-body movement model in OPUS-DSD2 accurately account large-scale dynamics in a physically valid manner. Meanwhile, structural variations at finer scales that are independent of dynamics can be captured by the neural network that models compositional changes. This dual approach to modeling dynamics and compositional changes is synergistic. As the composition of a conformation being more accurately modelled, the determination of movement relative to the starting conformation becomes more precise, leading to an overall increase in the modelling quality. This is evident when comparing the density map from OPUS-DSD2 to the density map from OPUS-DSD, showcasing the superior modelling quality of OPUS-DSD2. OPUS-DSD2 can also leverage the results of structural heterogeneity analysis produced by OPUS-DSD, which uncovers the most distinguishable macromolecular movements in cryo-EM data. OPUS-DSD2 can employ these preliminary modes to define the partition of rigid subunits and their rotation axes in dynamics, iterating to refine these results further.

In conclusion, the introduction of OPUS-DSD2 offers a neural-network-based paradigm in cryo-EM data analysis, and allows for a disentangled modelling of dynamics and compositional changes. However, as with any novel method, it is necessary to verify its performance with a broad range of macromolecular complexes and under different experimental conditions. This, in turn, will provide a rigorous assessment of its generalizability and robustness.

Methods

Notations

$x, \boldsymbol{x}$ and X will be denoted as scalars, vectors and matrices, respectively. $\|\cdot\|$ represents the Euclidean norm of a tensor. $\mathcal{X}$ denotes the Fourier transform of the multidimensional array X.

Deformation field

OPUS-DSD2 represents a 3D conformation by explicitly modelling its composition and deformation. The composition of a 3D conformation can be represented by a volume on a discrete 3D grid, which is referred as template, and the deformation in relative to the template can be represented by a 3D deformation field. Each voxel in the deformation field records its original position in the template volume. The 3D conformation in OPUS-DSD2 is sampled from the template according to the deformation field. Formally, the 3D conformation can be given by the composition of two functions $V(\boldsymbol{x} + \boldsymbol{g}(\boldsymbol{x})): \mathbb{R}^3 \rightarrow \mathbb{R}$, where $V(\cdot)$ is the template which maps

a grid point $\boldsymbol{x}$ to a voxel value, and $\boldsymbol{g}(\boldsymbol{x}): \mathbb{R}^3 \to \mathbb{R}^3$ is the deformation field which record the displacement between the voxel $\boldsymbol{x}$ in the deformed conformation and its source voxel $\boldsymbol{x} + \boldsymbol{g}(\boldsymbol{x})$ in template. Next, we give a detailed definition about the deformation field.

In OPUS-DSD2, the deformation of macromolecule is assumed to be compromised of individual movements of multiple subunits. The 3D deformation field is constructed through several steps. Firstly, a 3D consensus model is segmented into a set of pre-defined subunits, each of which has a specific shape with masses inside and undergoes rigid-body movement. Mathematically, the subunit can be defined as a function on the 3D grid, which is of the form

$$f(\boldsymbol{x}) = \begin{cases} 0, \boldsymbol{x} \in S \\ m(\boldsymbol{x}) > 0, \boldsymbol{x} \notin S \end{cases}, \qquad \text{(Equation 1)}$$

where S refers to the set of grid points belongs to the subunit, and $m(\boldsymbol{x})$ is the mass at point $\boldsymbol{x}$. We can further define some statistics about the subunit. For a subunit with center $\boldsymbol{c}_i \in \mathbb{R}^3$, and suppose its principal axes form a matrix P_i, where each row is one principal axis, its semi-principal diameters can be defined as

$$\boldsymbol{\sigma}_i = \sqrt{\frac{\sum (P_i(\boldsymbol{x} - \boldsymbol{c}_i))^2 f(\boldsymbol{x})}{\sum f(\boldsymbol{x})}}, \qquad \text{(Equation 2)}$$

where the summations are over all grid points in the 3D grid. For a grid point $\boldsymbol{x} = \{x, y, z\} \in \mathbb{R}^3$ in a subunit i, suppose the subunit undergoing a rigid-body displacement where the subunit is rotated by a rotation matrix $R_i \in SO(3)$, and translated by the translation vector $\boldsymbol{t}_i \in \mathbb{R}^3$, let the center of mass (COM) of the subunit be $\boldsymbol{c}_i$, the rigid-body displacement of this point can be expressed as

$$\boldsymbol{x}' - \boldsymbol{x} = (R_i - I)(\boldsymbol{x} - \boldsymbol{c}_i) + \boldsymbol{t}_i. \qquad \text{(Equation 3)}$$

For a grid point in a large complex which is comprised of multiple subunits, each of which undergoes independent movement, its displacement should be the combination of those

movements. We propose to blend the movement from a subunit at a grid point according to its distance w.r.t the center of subunit. Specifically, for a grid point $\boldsymbol{x}$, suppose the principal axes of subunit i form a matrix P_i as preceding, the displacement incurred by the movement of subunit i decays as,

$$w_i(\boldsymbol{x}) = \exp\left(-\frac{1}{T}\left(P_i(\boldsymbol{x}-\boldsymbol{c}_i)\right)^T \Sigma^{-1}\left(P_i(\boldsymbol{x}-\boldsymbol{c}_i)\right)\right), \qquad \text{(Equation 4)}$$

where x_j is the jth component of the grid point $\boldsymbol{x}$, Σ is a diagonal matrix with its diagonal element $\Sigma_{jj} = \sigma_{i,j}$, i.e., the semi-principal diameter of subunit i along the jth axis, and T is a constant for controlling the falloff of the weight function. The weight function should be normalized if there are n subunits in the complex. The normalized weight function can be expressed as,

$$\overline{w}_i(\boldsymbol{x}) = \frac{w_i(\boldsymbol{x})}{\sum_{i=1}^{N} w_i(\boldsymbol{x})}. \qquad \text{(Equation 5)}$$

In a complex with n subunits, the displacement of a grid point is the linear combination of movements of different subunits, which can be written as,

$$\boldsymbol{x}' - \boldsymbol{x} = \sum_{i=0}^{n} \overline{w}_i\left((R_i - I)(\boldsymbol{x}-\boldsymbol{c}_i) + \boldsymbol{t}_i + \boldsymbol{x} - (\boldsymbol{c}_i + \boldsymbol{t}_i)\right)\left[(R_i - I)(\boldsymbol{x}-\boldsymbol{c}_i) + \boldsymbol{t}_i\right], \qquad \text{(Equation 6)}$$

where $\overline{w}_i$ controls the scale of the movement of subunit i. We hence defined the deformation field $\boldsymbol{g}$, with which the deformed grid can be represented as $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{g}(\boldsymbol{x})$. In OPUS-DSD2, the rigid-body transformation parameters, R_i and $\boldsymbol{t}_i$, are predicted by the neural network. They should be restrained to rule out unrealistic deformations. We assume that the translational modes of subunits are generated by rotating COMs of subunits in relative to the COM of a common

fixed subunit, i.e., $\boldsymbol{t}_i = (R_{i,t} - I)(\boldsymbol{c}_i - \boldsymbol{c}_f)$, where the COM $\boldsymbol{c}_i$ is rotated by $R_{i,t}$ in relative to COM $\boldsymbol{c}_j$, and the translation of subunit can be obtained as the difference between the rotated COM and the original COM. The translation defined in this way retains the distance between COMs of subunits. The determination of the translation of each subunit comes down to estimating two rotation matrices. We can further limit the range of motion by restraining the rotation matrix to be an identity matrix, which inevitably hurts the expression power of the deformation model. To maximize the capacity of the deformation model under the identity restraint, the rigid-body movement should be defined in an appropriate reference frame, i.e., the z axis should align with the rotation axis. The rotation matrix which generates the translation of COM of subunit can be decomposed as $R_{i,t} = A_i^T R'_{i,t} A_i$, where the rotation matrix A_i aligns the rotation axis of the COM of subunit to the z axis, and $R'_{i,t}$ is the rotation matrix predicted by neural network. The rotation matrix of the subunit can be decomposed similarly as $R_i = A_i^T R'_i A_i$, where R'_i is the rotation matrix predicted by neural network. Our method adopts quaternion to represent rotation. The neural network in decoder outputs the last three dimensions of the quaternion, while the first dimension of quaternion is fixed to be a positive value, 8, which can regularize it to be an identity rotation.

The images collected by cryo-EM are 2D projections of a 3D molecular structure which undergoes a series of transformations from a template 3D volume. For a molecular structure V', let the template volume be V, suppose V' is rotated from V by rotation matrix R_t and deformed by the deformation field $\boldsymbol{g}$, the voxel in the 3D molecular structure has the following relation w.r.t the voxel in the template volume,

$$V'(\boldsymbol{x}) = V\left(R_t^T \boldsymbol{x} + \boldsymbol{g}(R_t^T \boldsymbol{x})\right). \qquad \text{(Equation 7)}$$

The 3D molecular structure can then be sampled from the template volume using spatial transformer. The 2D projection of the molecular structure can be obtained by summation along the Z axis, namely,

$$X(\{x, y\}) = \sum_{z=1}^{N} V'(\{x, y, z\}). \qquad \text{(Equation 8)}$$

In real cryo-EM experiments, the images are not in perfect focus, the objective lens presents aberration and there are significant amount of noises when recording images. Hence, the 2D projections are should be corrupted by the contrast transfer function which describes the effects of underfocus and the aberration of objective lens. In Fourier domain, let the Fourier transform of 2D projection be $\mathcal{X}$, the Fourier transform of the experimental image $\mathcal{X}'$ is of the form[31]

$$\mathcal{X}'_{h,k} = \mathrm{CTF}_{h,k}\mathcal{X}_{h,k} + \epsilon_{h,k}, \qquad \text{(Equation 9)}$$

where $\mathrm{CTF}_{h,k}$ is the contrast transfer function at frequency vector $[h, k]$, and $\epsilon_{h,k}$ is the noise introduced during experiment. We hence define an image formation model to generate cryo-EM image from a template volume according the specified deformation. It is fully differentiable and enables us to learn the template volume and deformation field simultaneously.

Training Objective

The main objective of OPUS-DSD2 is learning the template volumes and deformation fields using only 2D images. In OPUS-DSD2, two latent spaces are designated to encode the template volume and deformation field, respectively. The neural networks generate these objects conditioned on the corresponding latent codes. Let the neural network that represents the template volume be V, and the neural network that represents the deformation field be $\boldsymbol{g}$, for a particle, the image X reconstructed by these two neural networks is of the form,

$$X(\{x,y\}|\boldsymbol{z},\boldsymbol{z}^d,R_t) = \sum_{z=1}^{N} V(R_t\boldsymbol{x} + \boldsymbol{g}(R_t\boldsymbol{x}|\boldsymbol{z}^d))|\boldsymbol{z}), \qquad \text{(Equation 10)}$$

where $\boldsymbol{x} = \{x, y, z\}$ represents the coordinate of a voxel, $\boldsymbol{z}$ is the latent code of the particle for the network V, and $\boldsymbol{z}^d$ is the latent code of the particle for the network $\boldsymbol{g}$. The latent spaces and neural networks can be trained by minimizing the difference between the image reconstructed by the image formation model and the experimental image, that is,

$$\min_{V,\boldsymbol{g}} \left\| \mathrm{CTF}_i \mathcal{X}_i(\boldsymbol{z}_i, \boldsymbol{z}_i^d, R_t) - \mathcal{X}_i' \right\|^2 \qquad \text{(Equation 11)}$$

where $\boldsymbol{z}_i$ is the latent code for the template volume of image i, and $\boldsymbol{z}_i^d$ is the latent code for the deformation field of image i. However, the latent codes $\boldsymbol{z}_i$ and $\boldsymbol{z}_i^d$ are unknown. In practice, to efficiently train the content generation neural networks and guarantee the smoothness of latent space, variational autoencoder (VAE) architecture is adopted. The distributions of latent codes $\boldsymbol{z}$ and $\boldsymbol{z}^d$ are inferred by an encoder network f, and their distributions are restrained to be close to the standard gaussian distribution. In OPUS-DSD, we proposed a structural disentanglement prior to enforce the smoothness of latent space of template volumes[11]. Similarly, OPUS-DSD2 leverages a combination of the training objective of β-VAE[25] and the structural disentanglement prior to learn a smooth latent space for representing template volume. More specifically, let the dimension of 3D volume be N^3, the training objective of OPUS-DSD2 is of the form,

$$\begin{aligned} \min_{f,V,\boldsymbol{g}} E_{f(\boldsymbol{z}_i, \boldsymbol{z}^d|X_i')} - \log \sum_{\theta' \in NN(\theta)} \exp -\frac{1}{2} \left\| \mathcal{X}_i' - \mathrm{CTF}_i \mathcal{X}_i(\boldsymbol{z}_i, \boldsymbol{z}_i^d, R_t) \right\|^2 \qquad \text{(Equation 12)} \\ + \lambda_{KL} D_{KL}(f(\boldsymbol{z}_i, \boldsymbol{z}_i^d|X_i')|\big|N(0,I)\big) + \lambda J\left(f\left(\boldsymbol{z}_i, \boldsymbol{z}_i^d\middle|X_i'\right)\right) \\ + \frac{\lambda_{tv}}{N} \sum_{\boldsymbol{x}} \left\| \frac{\partial V(\boldsymbol{x}|\boldsymbol{z}_i)}{\partial \boldsymbol{x}} \right\|_2 + \frac{\lambda_{l_1}}{N} \sum_{\boldsymbol{x}} \|V(\boldsymbol{x}|\boldsymbol{z}_i)\|_1, \end{aligned}$$

where the first term is expectation of error between the experimental image X_i' and the reconstruction over the distribution of latent codes, $N(0, I)$ is the standard gaussian distribution, D_{KL} is the KL divergence between the distributions of latent codes given by the encoder network f and the standard gaussian distributions, λ_{KL} is the restraint strength for the KL divergence, λ is the restraint strength of structural disentanglement prior for the latent codes, λ_{tv} is the restraint strength for the total variation of the template volume, and λ_{l_1} represents the restraint strength of the L_1 norm of the template volume. It is also worth noting that the expectation term in (Equation 12) is intractable, and is approximated by the reparameterization trick during training [24].

Training

Our networks are trained end-to-end using only 2D images. The starting point of training is a set of 2D images after consensus 3D refinement in RELION[6] or cryoSPARC[32]. The pose parameters of images are determined w.r.t to a consensus reference model. Before training, a specific fraction of images is randomly assigned to the training set which is used to train the neural networks, while the remaining images are assigned to the validation set which solely serves to evaluate the reconstruction quality of neural network. A customized batching process is designed to approximately compute the structural disentanglement prior. The images are classified into 48 different projection classes according to the first two Euler angles of their rotation w.r.t the consensus model using HEALPix[33]. During training, a projection class is randomly chosen, and a batch of images within the projection class is sampled for training. This batching process guarantees the images in a batch come from the same projection class, and

approximates the intra-class comparison term of latent codes in the structural disentanglement prior. A sampled batch of images undergo a series of data augmentation operations which was defined in OPUS-DSD[11]. The augmented images are fed into the encoder network to compute the means and standard deviations of their composition and dynamics latent codes. We sample latent codes according to computed distributions and supply them into the composition and dynamics decoders, which generate template volume and deformation field, respectively. The final volume for the input image is reconstructed by deforming the template volume according to the deformation field. The deformed volumes are projected into 2D reconstructions following the image formation model which is described before. The reconstructions together with input images without augmentation form the reconstruction loss function. We also added the priors for latent codes, and the total variation[34] and L_1 norms[35] for the template volumes with specified weights to the training objective according to (Equation 12). The total loss function is optimized by the Adam optimizer[36].

Hyperparameters' settings

It is of critical importance to setting suitable hyperparameters for training OPUS-DSD2 since they balance image fitting and latent space regularization. Suitable hyperparameters can guarantee the quality of heterogeneity discovery. Since reconstruction losses are computed over a batch of images with high level of noises, while the priors are computed over the denoised signals, the reconstruction losses are much noisier than the priors. To balance the variances of reconstruction losses and the priors while avoiding posterior collapse[37], the scale of hyperparameters should be proportional to the SNRs of cryo-EM datasets. Moreover, the SNRs

of cryo-EM datasets are varying due to the masses of macromolecules and experimental setup. The hyperparameters can be conveniently set on an absolute scale considering the SNR of dataset. Using a batch of N images, the SNR can be estimated as,

$$\mathrm{SNR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left\|\tilde{X}_i\right\|^2}{\left\|X_i - \tilde{X}_i\right\|^2}, \qquad \text{(Equation 13)}$$

where $\tilde{X}_i$ is the image reconstructed by OPUS-DSD2, and X_i is the experimental image. $\left\|X_i - \tilde{X}_i\right\|^2$ estimates the norm of noises in the experimental image, while $\left\|\tilde{X}_i\right\|^2$ estimates the norm of signals. The restraint strength for KL divergence λ_{KL} is then set in relative to the estimated SNR, and of the form $\lambda \times \mathrm{SNR}$.

We fixed the restraint strength for the total variation of 3D volume $\lambda_{tv} = 0.5$ and the restraint strength for LASSO $\lambda_{l_1} = 0.15$. The learning rate was set to 10^{-4} that was decayed by 0.95 at each epoch for all experiments. The batch size was set to 64 during the training. All trainings were performed on 4 Nvidia V100 GPUs. The number of samples for interclass comparison was set to 12800. The numbers of kNN for interclass or intraclass comparison were set to 128 or 4, respectively, while the number of kFP for interclass comparison was set to 512. The latent codes of particles were updated with a momentum of 0.7.

Acknowledgements

The research was partially supported by National Key Research and Development Program of China (No.2021YFF1200400), Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01) and ZJLab. The authors specially thank James Krieger for his contribution to the code improvement of OPUS-DSD2.

Author Contributions

Z.L., Q.W., and J.M. conceived the work. Z.L. designed the algorithm and implemented the software. Z.L. performed the experiments. All authors wrote the final manuscript.

Figure Legends

Figure 1. Architecture of OPUS-DSD2 a). **Overall architecture.** Pose refers to the projection direction of input with respect to the consensus model. CTF refers to its contrast transfer function. z is the mean vector of composition latent code. σ is the standard deviation vector of composition latent code. z_d is the mean vector of dynamics latent code. σ_d is the standard deviation vector of dynamics latent code. $\mathrm{kNN}(z, z_d)$ refers to the K nearest neighbor of the concatenation of z and z_d. $\mathrm{kFP}(z, z_d)$ refers to the K farthest points of the concatenation of z and z_d. For better visualization, the link between the dynamics latent code and the memory is omitted. For simplicity, the standard gaussian prior for latent distribution as well as the smoothness and sparseness priors for the 3D volume are omitted in this chart. **b)**. **Architecture of the dynamics decoder of OPUS-DSD2.** This diagram shows the dynamics decoder that translates the latent encoding z_d into a deformation field. MLP denotes the multi-layer perceptron. RGD denotes the rigid-body dynamics model that transforms the predicted rigid-body movement parameters into the 3D deformation field. All MLPs except last one are with LeakyReLU nonlinearities with negative slope 0.2.

Figure 2. Heterogeneity analysis on the pre-catalytic spliceosome (EMPIAR-10180). a). Open conformations of spliceosome generated along the PC1 of latent space from OPUS-DSD2 and the PC1 of latent space from OPUS-DSD. The open conformations are generated at PCs with amplitude -1.1. **b)**. Closed conformations of spliceosome generated along the PC1 of the latent space from OPUS-DSD2 and the PC1 of latent space from OPUS-DSD. The closed conformations are generated at PCs with amplitude

-1.1. The regions with large differences in the qualities of densities are marked and highlighted by dashed boxes. All maps are contoured at the same level. **c)**. UMAP[38] visualizations of the 12-dimensional latent spaces of all particles encoded by OPUS-DSD2 and OPUS-DSD. Solid black dot represents the cluster center for labelled class. The corresponding reconstructions at the cluster centers are shown as well. **d)**. Closed conformations of spliceosome generated at the centroids of KMeans clusters in latent spaces learned by OPUS-DSD2, OPUS-DSD and CryoDRGN.

Figure 3. Heterogeneity analysis on *Plasmodium falciparum* 80S ribosome (EMPIAR-10028). a). POS conformations of Pf80S ribosome generated along the PC3 of latent space learned by OPUS-DSD2 and the PC1 of latent space learned by OPUS-DSD. The POS conformations are generated at PCs with amplitude 1. **b)**. NEU conformations of Pf80S ribosome generated along the PC3 of latent space learned by OPUS-DSD2 and the PC1 of latent space learned by OPUS-DSD. The NEU conformations are generated at PCs with amplitude 0.1. The regions with large differences in the qualities of densities are marked and highlighted by dashed boxes. All maps are contoured at the same level. **c)**. Conformations of Pf80S ribosome generated at the centroids of KMeans clusters in latent spaces learned by OPUS-DSD2, OPUS-DSD and CryoDRGN. **d)**. UMAP visualizations of the 12-dimensional latent spaces of all particles encoded by OPUS-DSD2 and OPUS-DSD. Solid black dot represents the cluster center for labelled class. The corresponding reconstructions at the cluster centers are shown as well.

Figure 4. Heterogeneity analysis on SARS-CoV-2 spike protein trimer (EMPIAR-10492). a). The segmentation of SARS-Cov-2 spike protein trimer. **b)**. UMAP visualization of the 12-dimensional composition latent space learned by OPUS-DSD2. Solid black dot represents the cluster center for labelled class. **c)**. Structures of S trimer generated by the decoder of OPUS-DSD2 at the centroids of KMeans clusters in the composition latent space. Red dashed circles mark the missing S1 CTD in the corresponding structures.

Supplementary Video Legends

Supplementary Video 1. Traversal of the third and fourth principal components of the composition latent space of the pre-catalytic spliceosome corresponds distinct compositional changes between Helicase and Foot.

Supplementary Video 2. Traversal of the first principal component of the dynamics latent space of the pre-catalytic spliceosome reveals the concerted folding of SF3b and Helicase onto the Core.

Supplementary Video 3. Traversal of the second principal component of the dynamics latent space of the pre-catalytic spliceosome reveals the folding of Helicase onto the Core and horizontal swing of the SF3b.

Supplementary Video 4. Traversal of the first and the third principal components of the composition latent space of Pf80S ribosome reveal different compositional changes.

Supplementary Video 5. Traversal of the second principal component of the dynamics latent space of the Pf80S ribosome reveals the relative rotation between the SSU and LSU.

Supplementary Video 6. Traversal of the first and the third principal components of the dynamics latent space of S trimer reveal two orthogonal modes of movements.

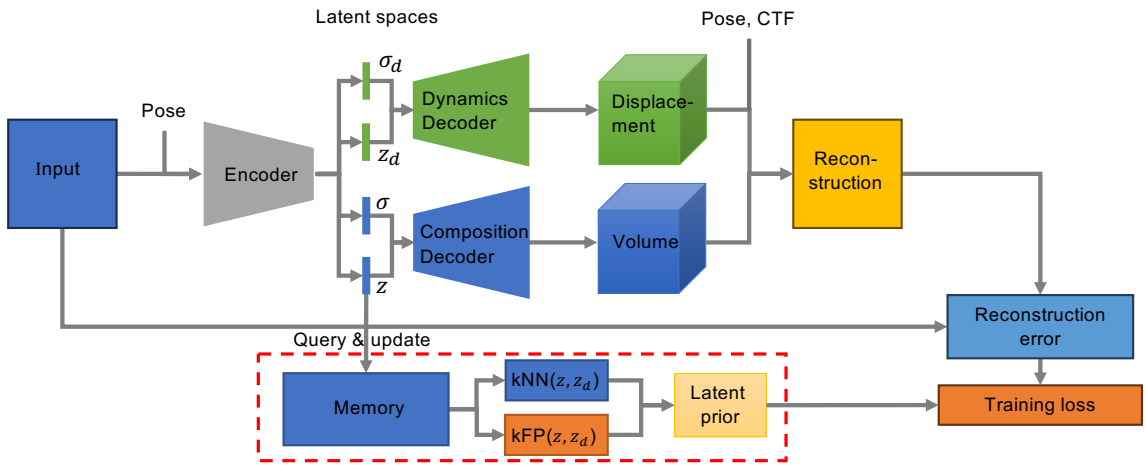
Supplementary Video 7. Traversal of the third principal component of the composition latent space of S trimer reveals the open and close of the S1 NTD, and the elongation and contraction of S2.

References

1 Glaeser, R. M. How good can cryo-EM become? *Nat. Methods* **13**, 28 (2015).
2 McCammon, J. A. & Harvey, S. C. *Dynamics of proteins and nucleic acids*. (Cambridge University Press, 1988).
3 Ma, J. New advances in normal mode analysis of supermolecular complexes and applications to structural refinement. *Current Protein and Peptide Science* **5**, 119-123 (2004).
4 Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377-388 (2013). https://doi.org:10.1016/j.jsb.2013.07.005
5 Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406-418 (2012).
6 Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519-530 (2012).
7 Bai, X., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife* **4** (2015). https://doi.org:10.7554/eLife.11182
8 Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods in enzymology* **579**, 125-157 (2016).
9 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436-444 (2015).
10 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
11 Luo, Z., Ni, F., Wang, Q. & Ma, J. OPUS-DSD: deep structural disentanglement for cryo-EM single-particle analysis. *Nat. Methods*, 1729-1738 (2023).
12 Gupta, H., Phan, T. H., Yoo, J. & Unser, M. Multi-CryoGAN: reconstruction of continuous conformations in cryo-EM using Generative Adversarial Networks. *Computer Vision – ECCV 2020 Workshops*, 429-444 (ECCV, 2020).
13 Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**, 930-936 (2021). https://doi.org:10.1038/s41592-021-01220-5
14 Punjani, A. & Fleet, D. J. 3D Flexible Refinement: structure and motion of flexible proteins from cryo-EM. *Microscopy and Microanalysis* **28**, 1218 (2022).
15 Goodfellow, I. *et al.* Generative adversarial nets. *Advances in neural information processing systems* **27** (2014).
16 Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review* **65**, 386 (1958).
17 Ma, J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **13**, 373-380 (2005).
18 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7** (2018). https://doi.org:10.7554/eLife.36861
19 Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D: Biological Crystallography* **62**, 439-450 (2006).
20 Calero, D. H. *et al.* Continuous heterogeneity analysis of CryoEM images through Zernike polynomials and spherical harmonics. *Microscopy and Microanalysis* **27**, 1680-1682 (2021).
21 Herreros, D. *et al.* Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials. *Nature Communications* **14**, 154 (2023).
22 Jaderberg, M., Simonyan, K., Zisserman, A. & Kavukcuoglu, K. Spatial transformer networks. *Advances in neural information processing systems* **28** (NeurIPS, 2015).

23 Lombardi, S. *et al.* Neural volumes. *ACM Trans. Graph.* **38**, 1-14 (2019). https://doi.org:10.1145/3306346.3323020
24 Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. *International Conference for Learning Representations* (ICLR, 2014).
25 Higgins, I. *et al.* β-VAE: Learning basic visual concepts with a constrained variational framework. *International Conference on Learning Representations* (ICLR, 2017).
26 Haselbach, D. *et al.* Structure and conformational dynamics of the human spliceosomal Bact complex. *Cell* **172**, 454-464. e411 (2018).
27 Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617-621 (2017).
28 Lloyd, S. P. Least squares quantization in pcm. *IEEE Trans. Inf. Theory* **28**, 129-137 (1982). https://doi.org:10.1109/TIT.1982.1056489
29 Wong, W. *et al.* Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, e03080 (2014).
30 Ke, Z. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498-502 (2020).
31 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216-221 (2015). https://doi.org:10.1016/j.jsb.2015.08.008
32 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290 (2017).
33 Gorski, K. M. *et al.* HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal* **622**, 759 (2005).
34 Rudin, L. I., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* **60**, 259-268 (1992).
35 Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288 (1996).
36 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *International Conference for Learning Representations* (ICLR, 2015).
37 Lucas, J., Tucker, G., Grosse, R. & Norouzi, M. Understanding posterior collapse in generative latent variable models. (2019, ICLR).
38 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://arxiv.org/abs/1802.03426. (2018).

a)


Latent spaces
Pose, CTF
Pose
Input
Encoder
σ_d
z_d
σ
z
Dynamics Decoder
Composition Decoder
Displace-ment
Volume
Recon-struction
Reconstruction error
Training loss
Query & update
Memory
kNN(z, z_d)
kFP(z, z_d)
Latent prior

b)

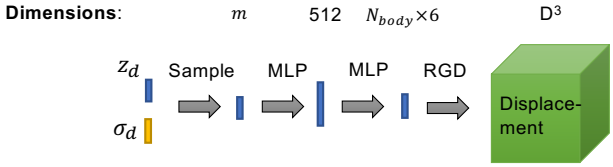

Dimensions:
m
512
$N_{body}\times 6$
D³
z_d
σ_d
Sample
MLP
MLP
RGD
Displace-ment

Fig.1

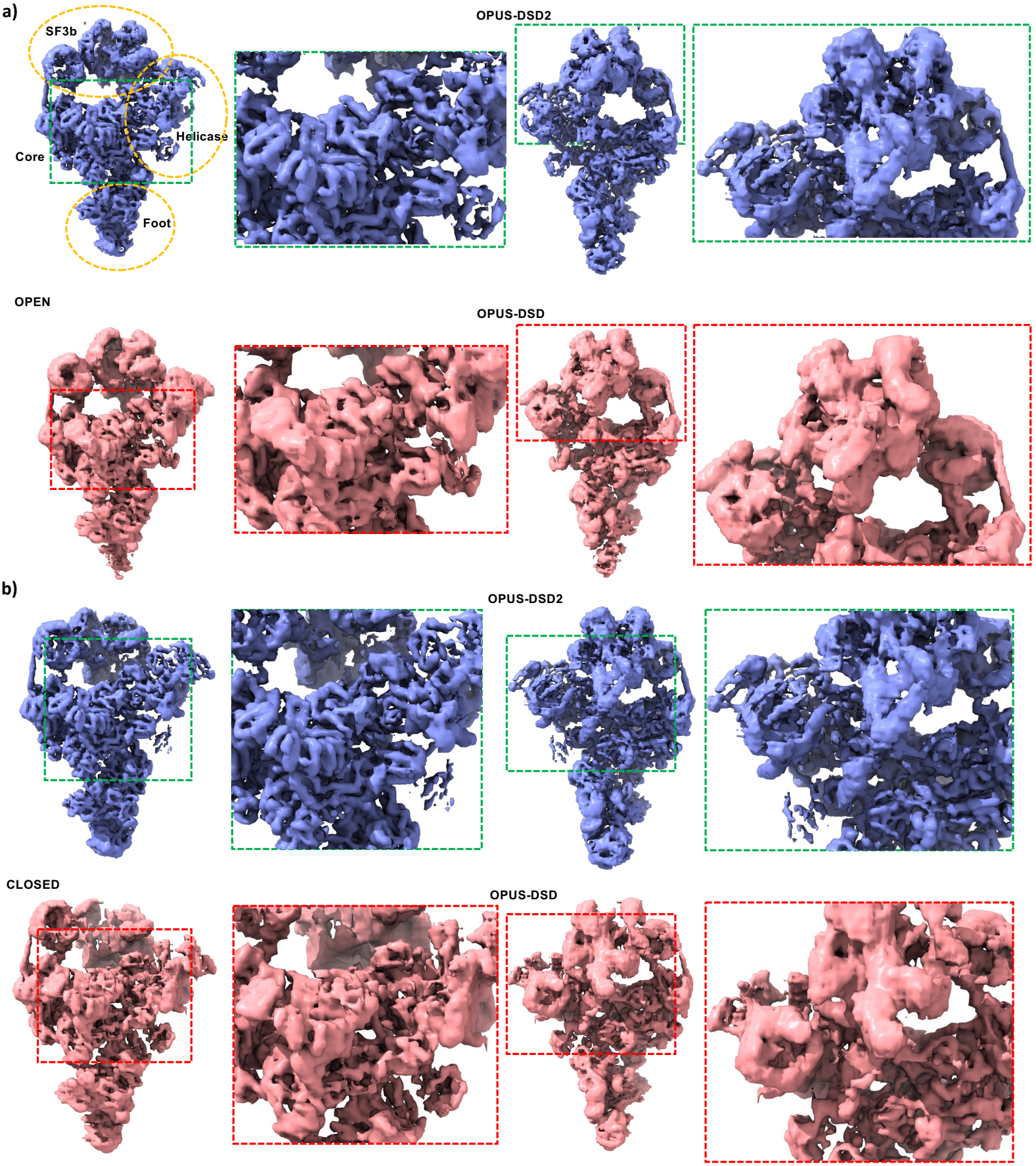

a)
SF3b
Helicase
Core
Foot
OPUS-DSD2
OPEN
OPUS-DSD
b)
OPUS-DSD2
CLOSED
OPUS-DSD

Fig.2

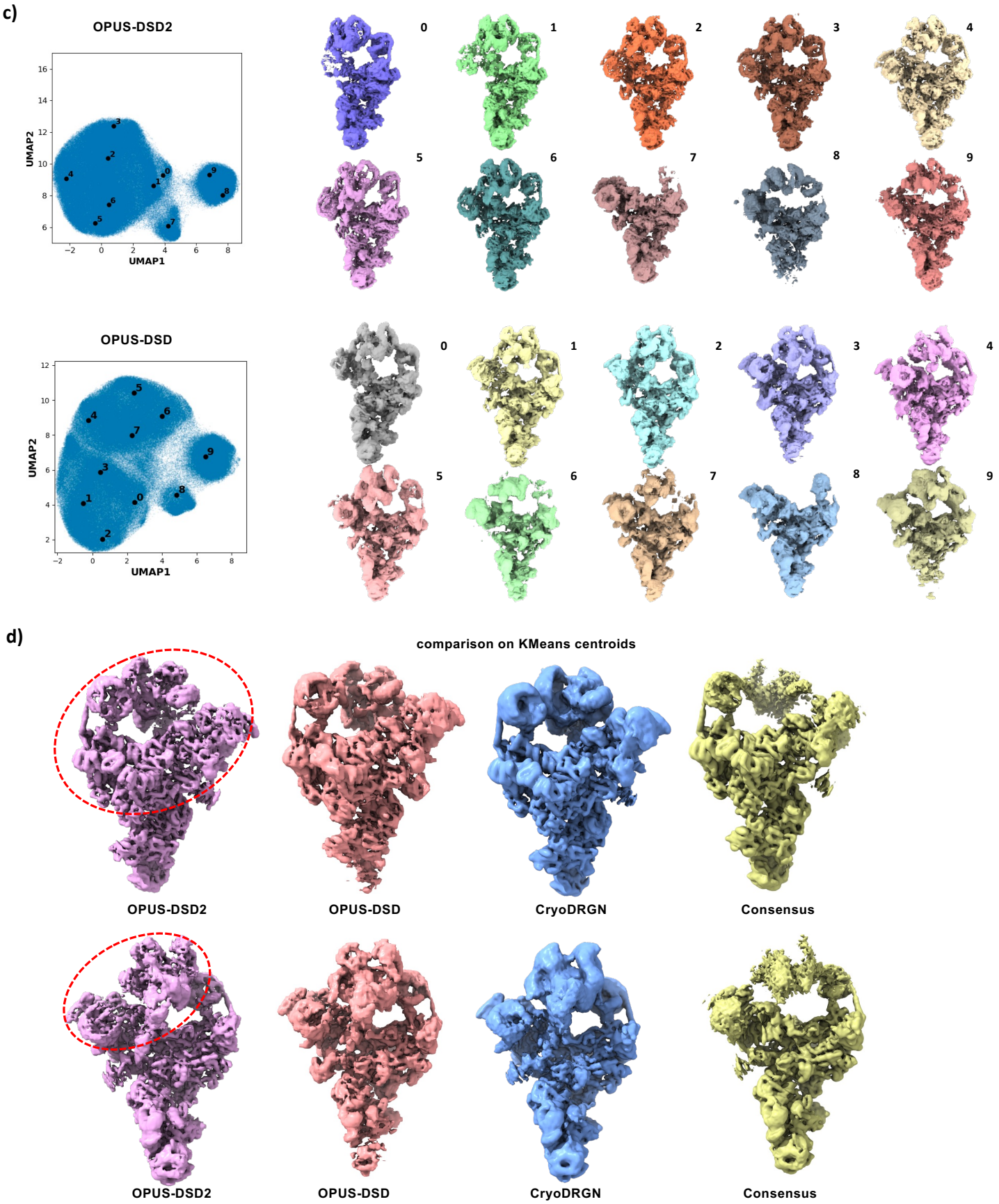

c)
OPUS-DSD2
UMAP2
UMAP1
0
1
2
3
4
5
6
7
8
9
OPUS-DSD
UMAP2
UMAP1
d)
comparison on KMeans centroids
OPUS-DSD2
OPUS-DSD
CryoDRGN
Consensus
OPUS-DSD2
OPUS-DSD
CryoDRGN
Consensus

Fig.2

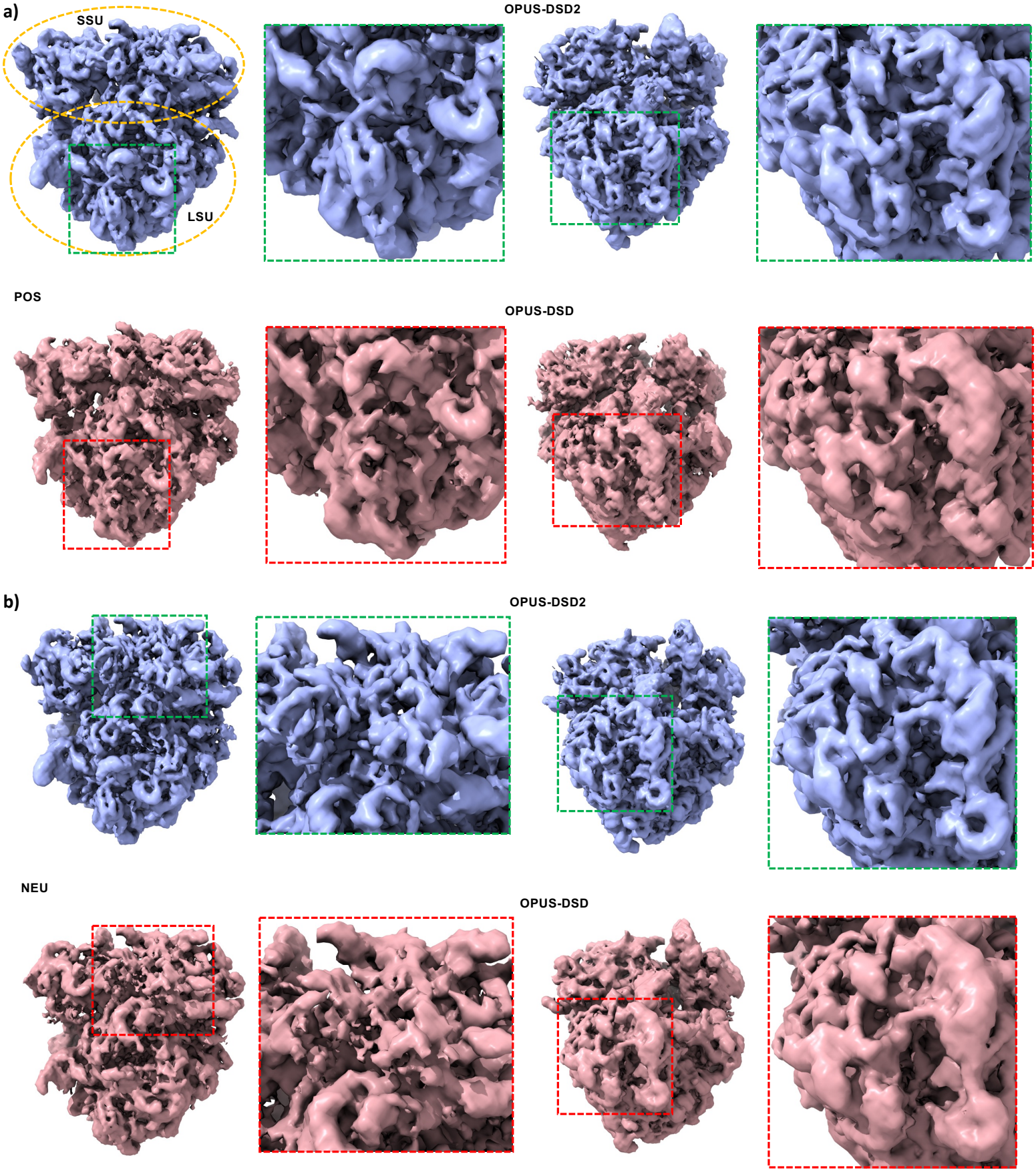


Fig.3

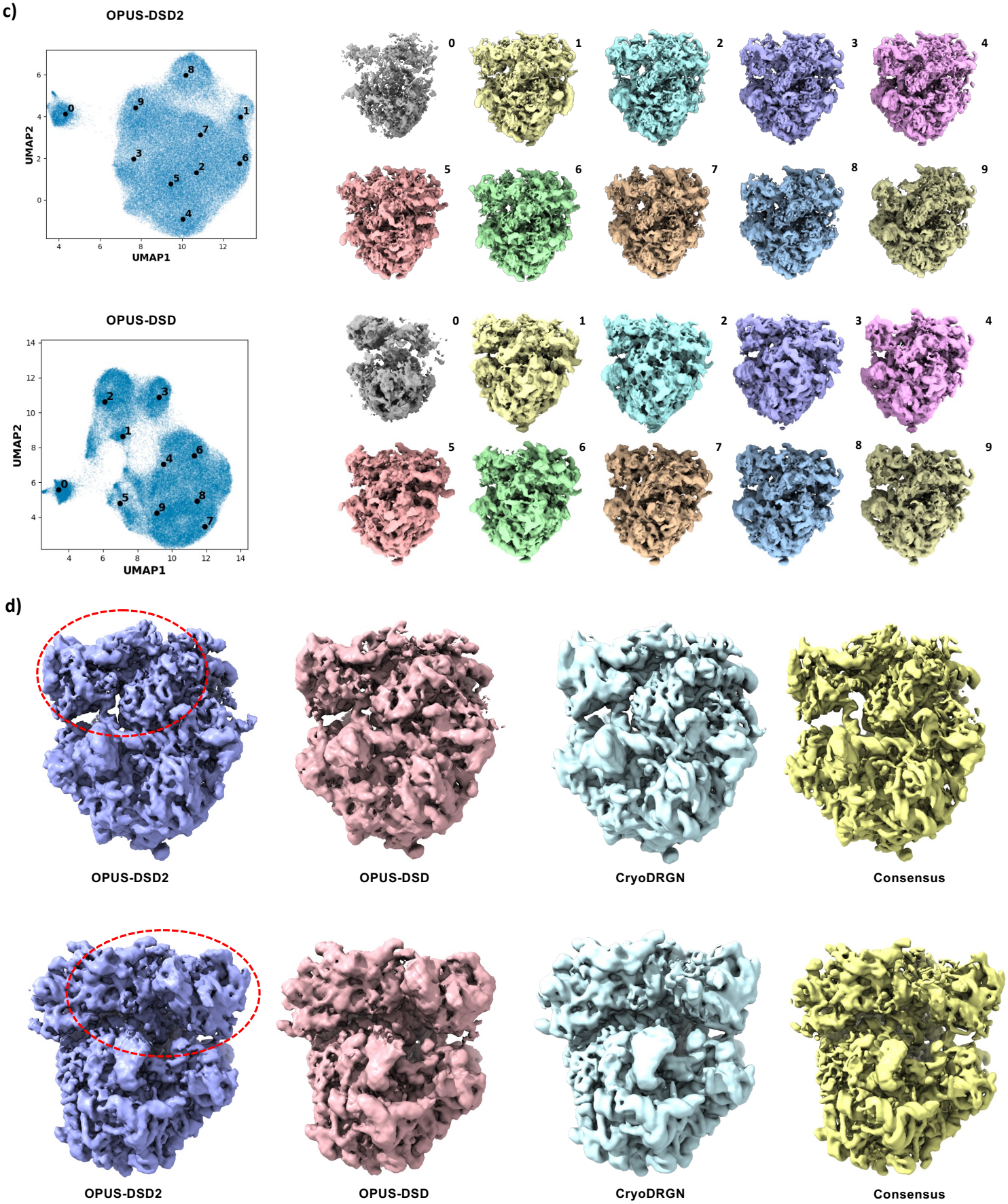

c)
OPUS-DSD2
UMAP1
UMAP2
OPUS-DSD
UMAP1
UMAP2
d)
OPUS-DSD2
OPUS-DSD
CryoDRGN
Consensus
OPUS-DSD2
OPUS-DSD
CryoDRGN
Consensus

Fig.3

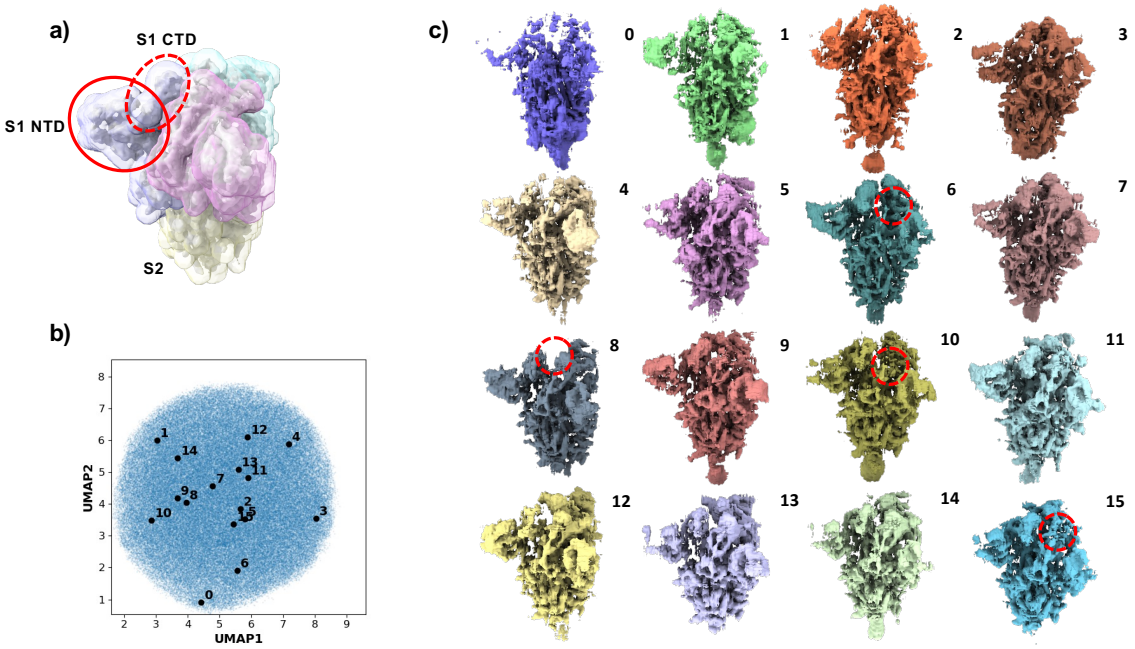


Fig.4