Deep learning based mixed-dimensional GMM for characterizing variability in CryoEM

Muyuan Chen[1] and Steven Ludtke[1]

[1] Verna Marrs and McLean Department of Biochemistry and Molecular Biology, Baylor College

of Medicine, Houston, Texas, USA

Abstract

The function of most protein molecules involves structural flexibility and/or dynamic interactions with other molecules. CryoEM provides direct visualization of individual macromolecules in different conformational and compositional states. While many methods are available for classification of discrete states, characterization of continuous conformational changes or large numbers of discrete state without human supervision remains challenging. Here we present a machine learning algorithm to determine a conformational landscape for proteins or complexes using a 3-D Gaussian mixture model mapped onto 2-D particle images in known orientations. Using a deep neural network architecture, this method can automatically resolve the structural heterogeneity within the protein complex and map particles onto a small latent space describing conformational and compositional changes. This system presents a more intuitive and flexible representation than other manifold methods currently in use. We demonstrate this method on several different biomolecular systems to explore compositional and conformational changes at a range of scales.

Introduction

Cryo-electron microscopy (CryoEM) is used to image biological macromolecules in a near-native state and is ostensibly capable of resolving structures to near-atomic resolution. However, most macromolecules possess substantial conformational and/or compositional variability as part of their biological function. A CryoEM micrograph contains a snapshot of many macromolecules, each frozen at a random point on its conformational and/or compositional landscape. This presents the difficulty that the detail visible in any single structure solved using CryoEM data will be limited by the conformational variability among the particles making it up. However, with more complete analysis, this fact can be turned into an advantage, as the individual data intrinsically explore a large portion of the conformational landscape of the system. With appropriate methods, achieving an ensemble of individually more details structures could be achieved.

Many methods have been developed to address the heterogeneity problem in SPR[1–3]. Perhaps the oldest and most commonly used method is multi-model refinement/3-D classification, in which multiple 3-D reference maps are used and each particle is compared to the projections of each map in each iteration[4–7] This can either be done simultaneously with particle orientation determination or making use of fixed orientations from an initial single model refinement. In focused classification, a mask is also used[8]. These methods often work quite well when the system falls into a small number of discrete states, such as the two states associated with ligand binding. However, to work well, the number of discrete states should be small, and the quality of the seed volumes often has an impact on the results.

Another common practice is to perform multi-model refinement, then rely on a human to discard particles representing states judged to be "bad"[9,10]. This process is typically repeated multiple times until a single map with improved resolution is achieved. The assumption is that this method produces one structure at high resolution representing the most populous state in the data, simply ignoring any other information present in the data. This clearly has the disadvantage of imposing human bias on the results and while the resulting map generally has improved resolution, it clearly presents an incomplete picture of the macromolecule being studied.

In addition to 3-D classification, multi-body refinement can be used to resolve local structural variability caused by conformational changes[11]. This technique relies heavily on researchers' prior knowledge about domain segmentation of the protein and requires the domains of interest to be large enough to provide signal required for local orientation assignment.

Finally, we have seen the recent emergence of manifold embedding techniques to address the problem of structural variability[12–16]. These methods are fairly new and varied in their mathematical methods. While they have shown promising results, they also face difficulties in mapping the manifolds to interpretations.

In this manuscript we present a strategy leveraging deep learning technology to map 2-D data directly to a 3-D Gaussian mixture model (GMM). This produces a representation where

conformational and compositional variabilities can be directly and intuitively related back to the data representation.

Methods

One of the difficulties in SPR heterogeneity analysis is the mathematical representation of protein conformations. If we consider the motion of an object from position A to position B along a simple linear path, it should be possible to represent the position on the path with a single value. However, when we represent this motion in images or volumes, the motion becomes a pattern of pixels becoming brighter and dimmer along the path in a complex sequence. Simple image analysis methods, such as principal component analysis (PCA), can readily identify the pixels involved in such a motion, but cannot readily map the highly nonlinear sequence of pixel variations back to the single degree of freedom we know must exist.

The second difficulty lies in the fact that in single particle analysis, variations occur in 3-D space, and yet the individual measurements are 2-D projections, lacking information about the location of the moving object in the 3rd dimension. However, a sufficiently large ensemble of images in different orientations and states contains sufficient information to constrain both the 3-D structure and variability.

Rather than attempting to determine paths in image space, we instead impose a Gaussian model at a specified level of detail, and then identify changes of Gaussian location and intensity which are self-consistent with the ensemble of particles. In this Gaussian mixture model (GMM) each function is defined by five variables: its 3-D coordinates, amplitude, and width. The width parameter is typically fixed, representing the level of detail in the representation. In this approach, the local motion of a domain would be represented by a simple change in location of the Gaussians making it up.

Converting the Gaussian representation into an image representation (a projection) is a trivial process, whereas the inverse process, starting with a single image and producing 3-D Gaussian locations, is underdetermined. The inverse problem is sufficiently constrained only when a large ensemble is considered. To solve this sparse and nonlinear inverse problem, we make use of deep learning methodologies (Fig 1). This design requires only the definition of a loss function describing the agreement between individual images and specific configuration of 3-D Gaussians. We make use of the Fourier ring correlation (FRC) metric[17] in the loss function, which has the additional advantage of being insensitive to microscope contrast transfer function (CTF) artifacts so long as the image is reasonably stigmated with minimal drift, and phase-flipping corrections have been applied to the particle images.

The network design involves 2 components. First, a decoder, which maps a small latent vector, into a set of 5N Gaussian parameters. The latent vector is simply a reduced dimensionality representation of the 3-D configuration of the molecule. In linear analysis, each component in the latent vector can represent one degree of freedom in the macromolecule. However, with the nonlinearity provided by the network, it is possible for local regions in the latent space to represent discrete states without a direct linear mapping.

The second network component is the encoder, which maps 2-D images, via their derivatives, into latent vectors. Thus, a 2-D image can be input to the encoder to produce a latent vector. The latent vector then passes through the decoder to produce 5N Gaussian parameters, which immediately provide a 2-D projection or 3-D volume as desired. This mapping process is constrained by the latent vector representation, and the set of particles mapped into this latent space will form a manifold, conceptually similar to other manifold methods in CryoEM. However, due to the nonlinearities and our enforcement of a GMM with a specified level of detail, it becomes possible to probe systems in very specific ways, which would be difficult using competing methods. For example, parameters of specific Gaussian components can be held constant, such that the GMM considers only variability in specific regions, or looks for correlations between specific regions.

This network structure is conceptually similar to the concept of an autoencoder[18], in which the network is trained directly from raw data, with no need for ground truth. The goal in an autoencoder is for the network representation to match the corresponding input image as closely as possible. In our case, the input data is 2-D, and our network output is the full 3-D GMM. While this 3-D model can recreate 2-D projections for training, the GMM output is far richer than any individual 2-D image. To achieve this result, a slightly different network training strategy is required.

We begin by training the decoder to provide an initial neutral 3-D context for the 2-D data. In the decoder training process, a fixed vector of zero is provided as input in the latent space, representative of the 3-D map's neutral conformation. The network is then trained to produce 5N Gaussian parameters best matching the neutral structure. To avoid becoming trapped in a local minimum, we begin with a downsampled version of the map, then progressively increase the sampling as the training process converges. The decoder is trained using an ADAM optimizer[19] with the FRC between the GMM and the provided map as a loss function. When complete, the decoder will produce an accurate representation of the neutral map with an input latent vector of zero, limited only by the number of Gaussians permitted by the model.

With the decoder trained, we move to include the encoder. The goal of this procedure is for the 3-D variability of the specimen to be represented by the latent space vector. The training protocol is seeded with the particles and orientation parameters from a standard single particle refinement. The assigned orientations for the particles can be imported from a standard EMAN2 or Relion refinement[20,21]. Instead of using raw particle images directly we use a perturbation-like model. For each particle, we compute the gradient of the FRC between the particle image and GMM with respect to the Gaussian parameters of the neutral model. These gradients become the input to the encoder for that particle. The gradient vectors are computed in the coordinate system of the GMM, so they are intrinsically invariant to translation and rotation of the raw particles. The loss function is, again, the FRC between the particle and Gaussian projection. For training, the encoder is initialized with random values producing a latent vector close to, but not exactly, zero.

The particle data and Gaussian parameters clearly will not agree perfectly, due to both noise present in the 2-D particle images as well as the conformational and compositional variability in

the specimen. As noise will be completely random within each particle, whereas the conformational and compositional variability will follow patterns represented in many particles, the latent space should preferentially train for variabilities actually present in the data. We do not require the orientations to be truly optimal at this point, as when one part of the structure is moving with respect to another, the concept of a single correct orientation does not exist. Once the complete network has been trained to represent the variabilities in the data with the given orientations, another training cycle can be run where the particle orientations are refined against the dynamic Gaussian model (Fig S1). This process can be iterated, though in practice, it is unlikely to take more than one or two iterations before the orientations and variability parameters agree to the best extent possible, given the limitations of the latent vector representation.

With a traditional PCA representation of variability in image space[22,23], the dimensionality of even a simple local motion within a structure will be very high since the motion involves many pixels undergoing nonlinear variations in intensity. As discussed above, with the GMM representation each independent motion should require, at most, a single variable in the latent space. Thus, our default of a 4-D latent vector can represent 4 infinite independent variabilities. However, given the nonlinearity of the system and the fact that molecular variation tends to be highly constrained, it is readily possible for a single variable to possess multiple features across its domain. Thus, once all of the particles have been mapped into the latent space via the encoder, it is still necessary to perform additional analysis on the particle distribution within the latent vector space. Any dimensional reduction algorithm could be used for this purpose. PCA applied to either the latent or GMM spaces is one straightforward approach for visualization and segmentation. Even with the nonlinearity of the network, we still have the constraint that similar configurations will be close to each other in the latent space and less similar configurations should be further apart. That is, we still expect continuous variabilities in structure to appear as manifolds in the latent space. Any latent vector can be easily visualized either immediately as its GMM representation or by reconstruction of the particles falling in a local region in the latent space.

Results

To validate the performance of our method, we made use of three data sets which are publicly available through EMPIAR. Each of these test data sets exhibits different sorts of variability, with the majority of the observed differences being well known and well understood. We also observe some additional motions not reported in the original study, but generally consistent with our understanding of the underlying system.

Ribosome assembly

For this test, we used a L17-Depleted 50S Ribosomal Intermediate dataset (EMPIAR-10076)[24], to demonstrate the method's ability to identify discrete variability, such as partial complex formation/ligand binding. We began with a structure determined using normal single particle methods in EMAN2 to 3.3 Å using the entire dataset after removing obvious ice contamination (124,900 particles). This structure was low-pass filtered to 8 Å, then used to generate a GMM with 3082 Gaussians. The specific number of Gaussians was empirically determined, based on the targeted level of detail. Any Gaussians falling outside a specified mask are excluded from

the final model. Since most of the structural variabilities within this dataset are the presence/absence of individual ribosomal components, we initially permitted only the Gaussian amplitudes to vary. After training, we took the 4-D latent space vectors (Fig S3) and used UMAP[25] to reduce the space to 2-D in order to further explore the structural variability of the system. From the embedded space, particles were clearly separated into visible clusters in this analysis. The particles in each of these clusters were then used to produce a 3-D reconstruction representing the cluster. The observed structural differences recapitulate known states[24] of ribosome assembly as shown in Fig 2.

While the points form clear clusters in the 2D conformation space, such classification only represents large scale structural differences, and more subtle compositional changes can also be observed from particles within the same class. For example, we manually selected three points along a line in one of the clusters with the central protuberance domain and reconstructed an averaged structure from the 2000 particles closest to each of the points in the embedded space. The resulting structures show the introduction of h68-70 and h76-78 of the 23s rRNA[24]. Interestingly, selecting 3 points along a nearly parallel line in a different cluster, one without the central protuberance domain, we observe the introduction of the same rRNA helices in the structures along the selected line (Fig 3b).

Finally, we examined conformational changes within the system. One of the factors that limits the resolvability of the averaged ribosome is the smearing effect of the dynamic central protuberance domain of the ribosome. To study this, we continued the network training with the Gaussian positions now permitted to vary in this domain, including only particles where the central protuberance domain is present. Note that the network model always includes the full set of 5 parameters for each Gaussian, but any of these components can be held constant. This additional analysis identified a clear tilting motion of ~8 degrees of this domain (Fig 3d).

Spliceosome

To test the performance of our method on large scale conformational changes we made use of pre-catalytic spliceosome data (EMPIAR-10180)[26]. We began with the particle orientation assignments and averaged structure determined using EMAN2 to 4.6 Å using the full dataset (327,490 particles). As resolution in CryoEM is a measure of self-consistency rather than visible detail, it is possible to achieve relatively high measured resolutions even in the presence of significant motion blurring, even when the structure clearly lacks detail. The density map was lowpass filtered to 13 Å and represented by 2048 Gaussian functions. We used PCA to reduce the neural network latent space to 2D after training for visualization of the subspace with the most significant variation. Compared to nonlinear dimension reduction methods, PCA conveniently preserves the inverse transform, so the eigenvectors can be mapped back to the Gaussian parameter space and the corresponding motion trajectories can be easily visualized. The first eigenvector from PCA shows a correlated motion of the helicase domain and the SF3b subunits, similar to the motion trajectories reported in previous studies.

While the eigenvectors from PCA exhibited several overall modes of motion of the complex, to better interpret the mechanism of the system, it is more interesting to look at the eigen-motion trajectories localized in individual domains. The use of PCA does not change the fact that the

latent space has a non-linear relationship to the motions of the system. Thanks to the characteristics of Gaussian models, we can focus on specific regions in real space. Rather than decomposing the point cloud in the neural network latent space with PCA, we search for origin-crossing vectors in the latent space where the motion of Gaussian coordinates along the line is maximized in the domain of interest but minimized in rest parts of the protein (Fig 4). Since the points from this dataset form a relatively isotropic distribution in the latent space, the vectors found using this method can represent motion as dominant as the eigenvectors from PCA, while keeping the Gaussian functions that are involved in the motion as localized as possible. Furthermore, since the motion trajectories are localized at different domains, the two eigen-motion vectors are also orthogonal to each other.

With the two independent eigen-motion vectors localized in the helicase and SF3b domains, we can further investigate the coordination of the two domains by looking at motion trajectories produced by the linear combinations of the two vectors. Adding the two vectors results in a motion mode that the two domains are moving toward the same direction, similar to the first eigen-motion extracted by PCA from the system. In the alternative combination, the two domains can be seen move apart from each other, a motion mode never reported for the dataset (Fig S4).

SARS-COV-2

Our third, somewhat timely, test system is the spike structure of SARS-Cov-2 (EMPIAR-10492). While the opening of the Receptor Binding Domains (RBD) was not observed in the deposited particles due to the sucrose cushion used in sample preparation[27], the RBDs in the final structure still have weaker density and lower resolution compared to the rest of the protein. 3-D classification was performed on the dataset in the original publication, but only an asymmetrical structure with weak RBD density was reported, and it is unclear what conformational changes take place to weaken the RBD density.

To investigate this question, we performed heterogeneity analysis on the combined particle set of the RBD closed and weaker density state (55,159 particles). To demonstrate that our method is directly compatible with other software, rather than solving the structure again in EMAN2 we make use of the averaged structure and particle orientations from the deposited Relion refinement results. 2188 Gaussian functions were used to model the averaged structure at ~7 Å. To break the C3 symmetry, we treat every particle as three copies at the three symmetrical orientations, and only Gaussian functions at one of the asymmetrical units are allowed to move from the neutral structure, so that every particle is mapped to three points in the conformation space, corresponding to the three asymmetric units.

After training, we perform PCA on the points in conformation space and the eigenvectors show the motion of secondary structure elements at the RBD. Along the first eigenvector, the alpha helix at residue 335-344 can be seen tilting toward the RBD of its adjacent subunit by ~11 degrees. Interestingly, in the averaged structures along the same trajectory, the same helix in one of the neighboring subunits is undergoing the same motion, but in the opposite direction (Fig 5). Since the adjacent subunit was not targeted in the heterogeneity analysis, the presence of correlated motion suggests that the conformational changes of the RBDs in the two subunits

7

are coordinated. Meanwhile, the same domain in the other subunit remains unchanged. On the other hand, the second eigenvector from PCA emphasizes the motion of the alpha helix at residue 364-371, as well as the beta-sheet strain at residue 354-359. Some coordination of motion in the adjacent subunit can also be observed but it is less clear.

It is also worth noting that in the density maps reconstructed from particles at determined conformation states, the RBD at the subunit we focus on has stronger density than that of the other two subunits, suggesting the conformational changes we get are indeed contributing to the weakening of density at the RBD (Fig S2).

Discussion

The major difference between the proposed method and the vast majority of previous CryoEM structure determination methods is that we represent the structure of the molecule as a set of Gaussian functions during refinement. This is analogous to the idea of directly refining the atomistic structure against the raw data, but in a reduced representation based on the scale of the expected variabilities. This representation provides a number of advantages. First, it greatly reduces the number of parameters that needed to represent the molecule at any specified level of detail, limited by the sampling of the image data[28]. For example, in the case of spliceosome, to represent the structure at 13Å using a voxel-based representation, the density map can be downsampled to a cube with a box size of 84, so a total of 592704 floating point parameters is required to render the volume. Using our representation, only 10240 variables are needed for 2048 Gaussian functions used in the model, and the average FSC between the Gaussian model and the density map is above 0.95 for the spatial frequencies under consideration, indicating that it is a very good representation of the density map. Moreover, as the size of protein or the target resolution increases, the parameters needed for the Gaussian representation grows at a much slower rate than that for density maps since atoms in proteins generally occupy 3-D space sparsely.

Second, at low resolution, Gaussian functions are a natural way to model the electron density function of molecules[29–31]. If these methods were extended to atomic resolution representation, it may be necessary to include the atomic form factors in the model, but at intermediate resolutions these subtleties are effectively undetectable. Typical protein structural variability, such as ligand binding and domain motion, can be easily represented as linear trajectories in the Gaussian parameter space. Whereas under voxel-based representations, a high-dimensional nonlinear model is required to depict the motion of a domain along a linear trajectory, especially when the length of the path is longer than the size of the domain of interest. As a result, the complexity of the model required to describe the structural variability of the protein is greatly reduced, easing the effort to train the encoder-decoder neural networks.

Third, due to the mathematical characteristics of Gaussian functions in both real and Fourier space, our representation avoids the artifacts produced by common image processing operations. For example, masking can be performed by selecting a subset of Gaussian functions, without producing significant edge effects in either real or Fourier space. Since the projection operation is performed by transforming the coordinates of the center of Gaussian functions, no interpolation artifacts are introduced by rotation or non-integer translation. This

also makes it easier to apply constraints in both spaces when studying the structural variability of proteins, such as focusing on specific domains in real space or limiting to a range of Fourier frequencies.

Finally, the use of Gaussian models makes the output from the neural networks directly and intuitively interpretable, unlike the typical abstract spaces produced by other "manifold methods"[14,15]. Each point in the conformational space is mapped to a set of Gaussian parameters, which corresponds to a 3-D structure of the molecule at the conformation. This means that for any given point, a representation can be generated either by reconstructing the particle image data in the vicinity of the point, or by directly converting the Gaussians into a density representation. For any two selected points in the confirmation space, it is easy to visualize the differences between those points by plotting the trajectory of coordinate motion or amplitude changes. This can be especially useful in identifying transient conformational changes when there are insufficient particles in the conformation of interest to provide a 3-D reconstruction at sufficient resolution. The Gaussian representation remains equivalently resolved at any point.

Some limitations remain in the current implementation of the method. First, since this method requires the orientation of each particle as input, it only works in situations where a portion of the molecule is rigid enough that a reasonable neutral 3-D structure exists, and reasonable particle orientations can be determined. While this is a safe assumption for most single particle cases, it is also possible that the protein complex of interest is so heterogeneous that the alignment in the initial refinement fails entirely, and particle-projection mismatch is caused by misalignment instead of conformational differences. While a potential solution to this problem is incorporation of orientation determination into the initial network training process, it is difficult to train the model to convergence when both orientation and conformation of the particles are simultaneously considered.

Second, the protocol normally begins with an averaged structure of all particles, assuming this represents the "neutral structure", which is then perturbed by the network. This assumption is not always true. When a domain motion is large enough, regions in the averaged map may be sufficiently spread in space that no Gaussian function is identified in that region when we initialize the neutral model based on the averaged structure. As a result, the model excludes the motion of that region even after full training. Under the current workflow, this can be corrected by selecting an initial "neutral structure" with better density in this region. If dealing with a system with compositional variability, such as multiple ligands which may or may not be present, it is critical that the training volume be one with some density present for all ligands. For systems with conformational variability, the most populous state would be a natural choice. This potential problem can also be reduced by building the neutral Gaussian model directly from aligned particles instead of the averaged structure, although this will incur a time penalty and may lead to a less robust neutral model.

Finally, GPU memory currently limits the largest size of molecule and highest resolution our method can currently handle. Currently, running on an RTX 2080 GPU with 12GB memory, we can use as many as 3200 Gaussian functions with particles sampled at 128 x 128 pixels, and a

batch size of 8 during training. This can represent the structure of the 50S ribosome at ~8 Å resolution, or smaller proteins at higher resolution. So, for most proteins, the method is currently limited to solving motion of domains or alpha helices, but not heterogeneity of individual sidechains.
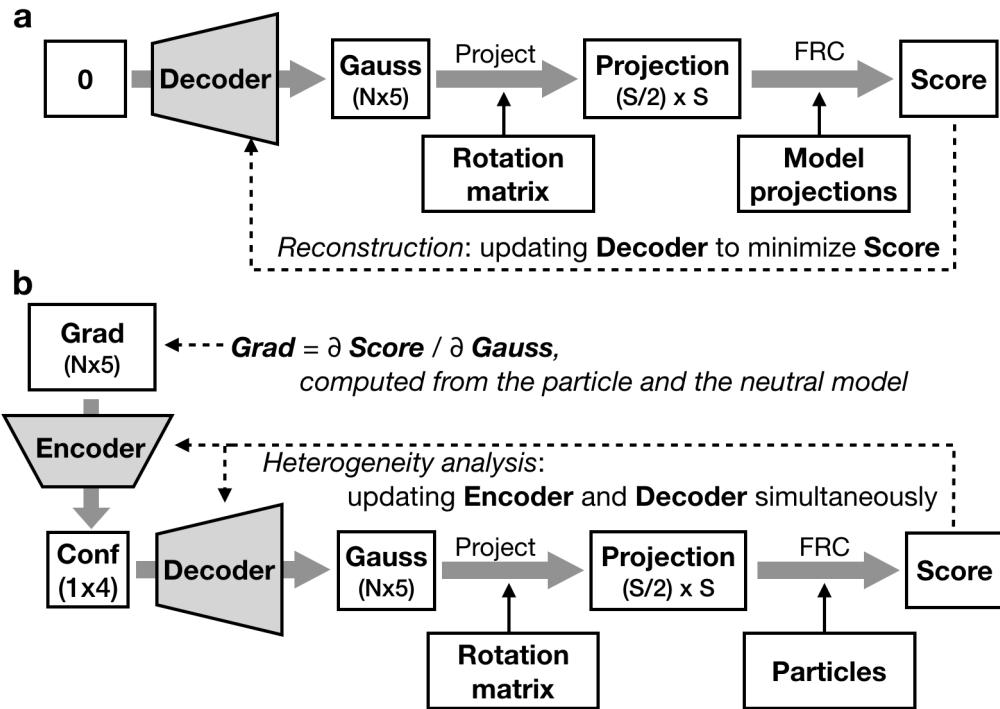
The primary cause for the large memory consumption is the requirement that each Gaussian be represented as a full image, despite being quite localized. This is part of the underlying TensorFlow infrastructure and cannot be remedied at present without using low level programming which may cease to function in future TensorFlow versions. Alternatively, there are some simple ways to moderately reduce memory consumption, but at the potential cost of decreasing the performance of the method. For example, reducing the precision from float32 to float16 reduces memory cost 2-fold, but faces the risk that there may not be enough precision to cover the large dynamic range present in Fourier space. Reducing the batch size during training would also decrease memory usage, but this also weakens the statistical power of the optimizer which may decrease the robustness of the convergence. We expect that continuing evolution of GPU hardware as well as TensorFlow itself will remedy this problem in the near future without requiring such compromises to the method.

Despite these minor limitations, this method represents an easy to use mechanism for exploring macromolecular variability in CryoEM with results which can be easily and intuitively interpreted. The user can define the resolution of interest, easily approaching features at any level of detail, within hardware limits. The next obvious development for this method would be to operate on CryoET data, to permit similar studies in the context of the cellular environment, but technically this adaptation is not entirely straightforward to achieve due to the high noise levels in individual tilts and the increase in the amount of coordinated image data this would entail. All of the GMM operations are available through the program e2gmm_refine.py, and a graphical interface for interactive examination of results and exploring changes in parameters is provided by e2gmm.py. All of the necessary software is provided as part of EMAN2.39. A tutorial with sample data is available at http://eman2.org.
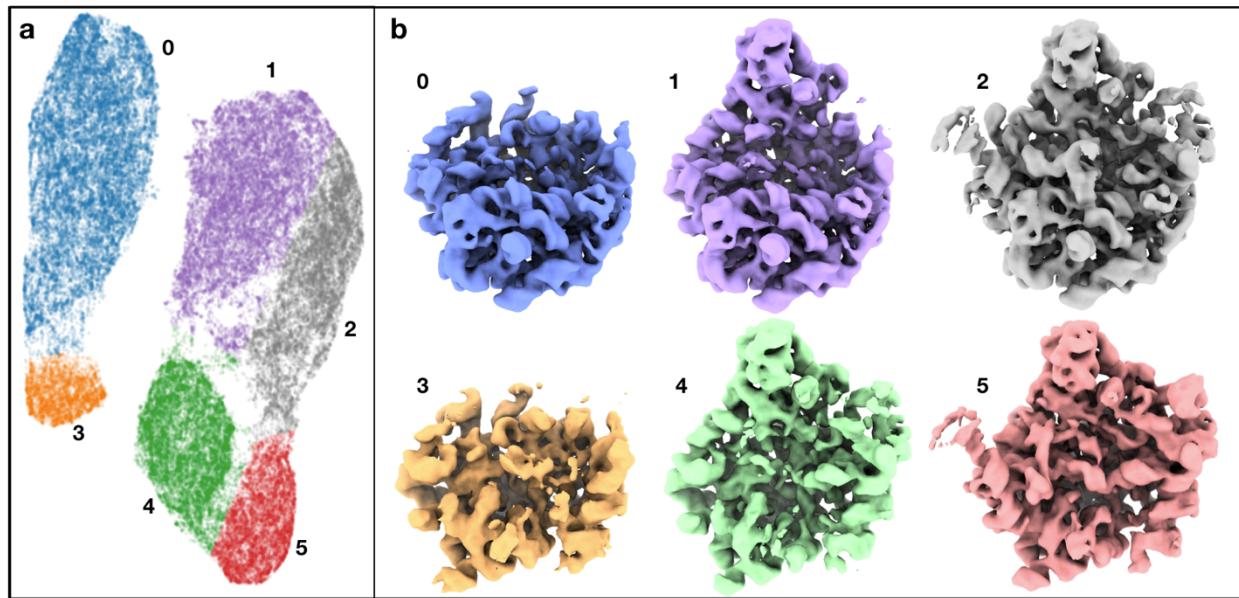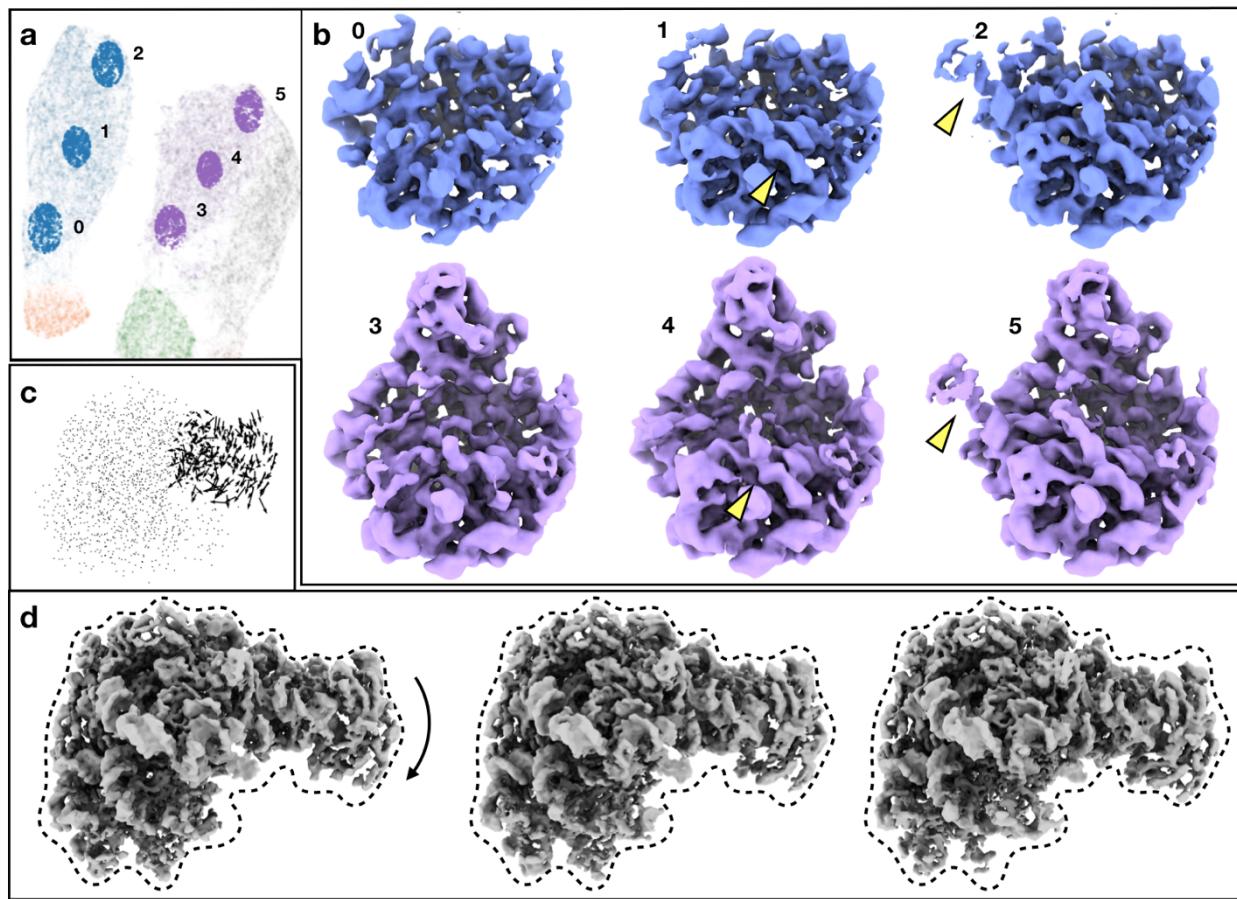
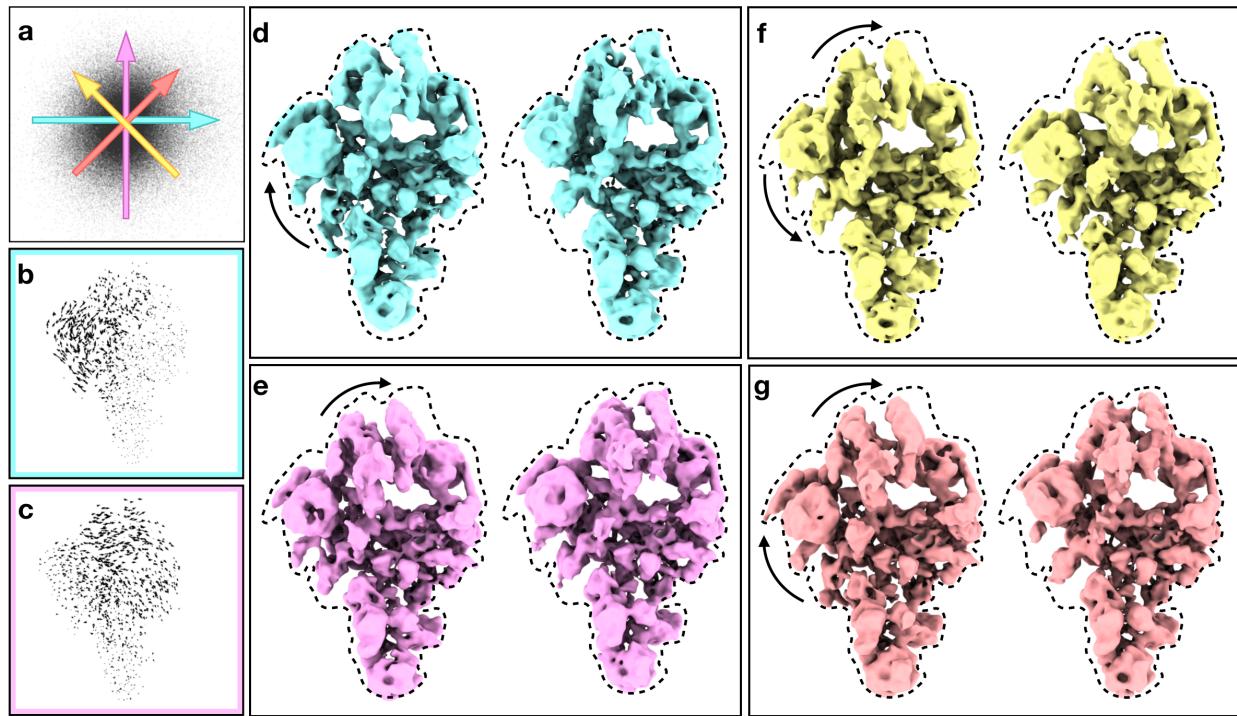## Acknowledgement

Figures

Fig 1. Model diagram. **(a)** Training workflow during reconstruction. **(b)** Training workflow during heterogeneity analysis.
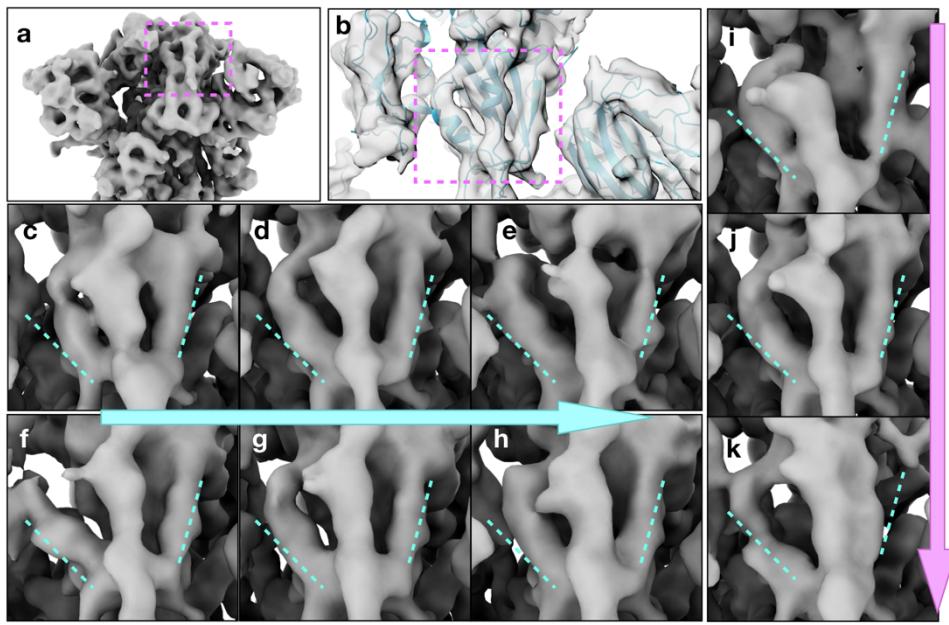
Fig 2. Classification of assembling ribosomes. **(a)** 2D embedding of particles from the 4-D latent space, colored by labels from clustering. **(b)** Averaged 3-D structures produced using the 2-D particles in each class, filtered to 8A. See also supplementary movie 1.

Fig 3. Exploration of subtle structural variability in the ribosome dataset. **(a)** Location of particles sampled from the 2D embedding of the conformation space. **(b)** Averaged structures reconstructed from the sampled particles. Yellow arrows point to the major differences between the structures. **(c)** Motion trajectories of Gaussian coordinates in the central protuberance domain from the first eigenvector of conformational heterogeneity analysis. **(d)** Averaged structures of the particles at points along the motion trajectory. The dotted envelope is fixed to better visualize the changes in each map. See also supplementary movie 2.

Fig 4. Structural variability analysis of spliceosome. **(a)** Distribution of particles in the 2D space formed by the selected base vectors. Colored arrows correspond to different motion trajectories shown in (d-g). **(b-c)** Motion trajectories of Gaussian coordinates from the two base vectors. **(d-e)** Averaged structures reconstructed from particles along the two base vectors, showing the conformational change corresponding to **(b-c)** in 3-D. **(f-g)** Averaged structures reconstructed from particles along two different combinations of the base vectors, corresponding to the colored arrows in **(a)**. Dotted envelopes are fixed to better visually judge variations between maps. See also supplementary movie 3.

Fig 5. Structural variability analysis of the spike protein of SARS-COV-2. **(a)** Average structure of the spike protein, showing the RBD of the subunit that the analysis focuses on. **(b)** Density map of the target RBD overlaid with the molecular model (PDB: 6zwv). **(c-e)** Conformation change of the target RBD along the first eigenvector. **(f-h)** Conformation change of the one of the RBDs that are not targeted along the first eigenvector. **(i-k)** Conformation change of the target RBD along the second eigenvector. See also supplementary movie 4.

1. Ludtke, S. J. Single-Particle Refinement and Variability Analysis in EMAN2.1. *Methods Enzymol.* **579**, 159–189 (2016).

2. Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol.* **579**, 125–57 (2016).

3. Jonić, S. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Curr. Opin. Struct. Biol.* **43**, 114–121 (2017).

4. Gabashvili, I. S., Agrawal, R. K., Grassucci, R. & Frank, J. Structure and structural variations of the Escherichia coli 30 S ribosomal subunit as revealed by three-dimensional cryo-electron microscopy. *J. Mol. Biol.* **286**, 1285–91 (1999).

5. Scheres, S. H. W. *et al.* Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348**, 139–149 (2005).

6. Chen, D.-H., Song, J.-L., Chuang, D. T., Chiu, W. & Ludtke, S. J. An expanded conformation of single-ring GroEL-GroES complex encapsulates an 86 kDa substrate. *Structure* **14**, 1711–22 (2006).

7. Scheres, S. H. W. *et al.* Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat. Methods* **4**, 27–29 (2007).

8. Penczek, P. A., Frank, J. & Spahn, C. M. T. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol.* **154**, 184–194 (2006).

9. Lu, P. *et al.* Three-dimensional structure of human γ-secretase. *Nature* **512**, 166–170 (2014).

10. Dong, Y. *et al.* Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49–55 (2019).

11. 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).

12. Fu, T. M. *et al.* Cryo-EM Structure of Caspase-8 Tandem DED Filament Reveals Assembly and Regulation Mechanisms of the Death-Inducing Signaling Complex. *Mol. Cell* **64**, 236–250 (2016).

13. Lederman, R. R. & Singer, A. Continuously heterogeneous hyper-objects in cryo-EM and 3-D movies of many temporal dimensions. (2017).

14. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: Reconstruction of heterogeneous structures from cryo-electron micrographs using neural networks. *bioRxiv* 2020.03.27.003871 (2020). doi:10.1101/2020.03.27.003871

15. Punjani, A. & Fleet, D. J. 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *bioRxiv* 2020.04.08.032466 (2020). doi:10.1101/2020.04.08.032466

16. Dashti, A. *et al.* Retrieving functional pathways of biomolecules from single-particle

snapshots. *Nat. Commun.* **11**, 4734 (2020).

17. Van Heel, M. Similarity measures between images. *Ultramicroscopy* **21**, 95–100 (1987).

18. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. *Proc. 25th Int. Conf. Mach. Learn. - ICML '08* 1096–1103 (2008). doi:10.1145/1390156.1390294

19. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014).

20. Bell, J. M., Chen, M., Baldwin, P. R. & Ludtke, S. J. High resolution single particle refinement in EMAN2.1. *Methods* **100**, 25–34 (2016).

21. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).

22. van Heel, M. & Frank, J. Use of multivariate statistics in analyzing the images of biological macromolecules. *Ultramicroscopy* **6**, 187–194 (1981).

23. Penczek, P. A., Kimmel, M. & Spahn, C. M. T. Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *Structure* **19**, 1582–1590 (2011).

24. Davis, J. H. *et al.* Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell* **167**, 1610-1622.e15 (2016).

25. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).

26. Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621 (2017).

27. Ke, Z. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498–502 (2020).

28. Kawabata, T. Gaussian-input Gaussian mixture model for representing density maps and atomic models. *J. Struct. Biol.* **203**, 1–16 (2018).

29. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).

30. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384–1403 (2019).

31. Bonomi, M. *et al.* Bayesian Weighing of Electron Cryo-Microscopy Data for Integrative Structural Modeling. *Structure* **27**, 175-188.e6 (2019).

Methods

Neural network structure and training parameters

The structure of neural networks and the parameters during the two phases of training process are user adjustable. By default, the encoder and decoder both have four fully-connected hidden layers, each with 512 units. A dropout layer with a rate of 0.3, as well as a batch normalization layer is included before the final output layer of both networks. The ReLu function is used for activation in each layer, except for the output layer of the decoder, which uses a sigmoid activation function, so the Gaussian parameters are constrained to [-1, 1]. During the training process, the default learning rate is 0.0001, with an L2 regularization of 0.001. A small random variable is also added to the latent space vector before it is inputted to the decoder, as a way to enforce the continuity of particle distribution in the latent space. An additional regularization factor is applied to the standard deviation of the width of Gaussian functions to prevent large variations.

Time requirements

For EMPIAR-10076, starting from a completed single particle refinement, the first round of heterogeneity analysis, which focuses on only the amplitude changes of Gaussians takes ~3 hours on a GeForce RTX 2080 TI GPU, including the reconstruction of Gaussian model and the dimension reduction process. ~10 CPU hours (< 1 hour on a 12-core workstation) was required to embed the encoder latent space to 2D, clustering and reconstruction of 3-D density maps. The second round of heterogeneity analysis focusing on the conformational change in the central protuberance domain also required ~3 hours on the GPU and <1 hour on the 12 core workstation.

For the EMPIAR-10180 dataset, the heterogeneity analysis required ~3 GPU-h and ~30 CPU-h for the reconstruction of the density maps along the four reported motion trajectories.

The provided Relion alignment was used for the EMPIAR-10492 dataset. Heterogeneity analysis required ~2 GPU-h plus ~10 CPU-h.

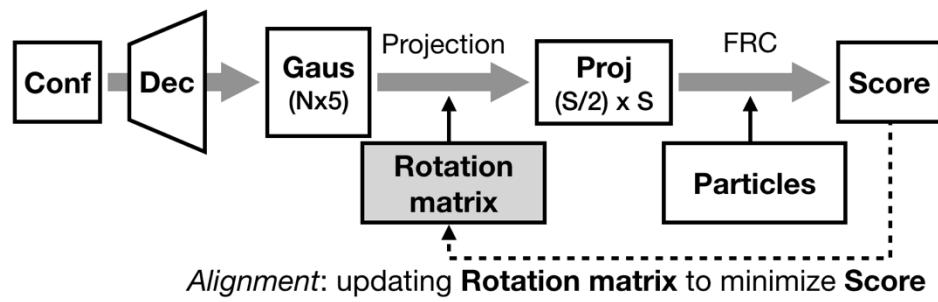*Alignment*: updating **Rotation matrix** to minimize **Score**

Fig S1. Workflow for local orientation refinement based on the Gaussian model and conformation of each particle. Here the conformation of particles and the decoder is obtained from the heterogeneity analysis training step.
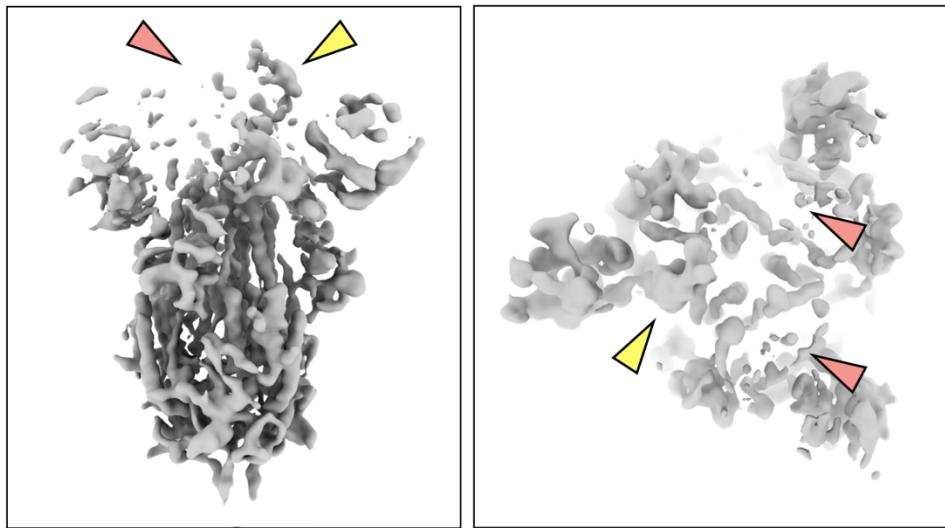
Fig S2. Structure of SARS-COV2 spike at one point in the continuous motion, visualized at high isosurface threshold. Note that the RBD of the subunit the heterogeneity analysis focuses on (yellow arrow) is still solid while the other two RBDs (red arrows) already vanish. This suggests the continuous motion is contributing to the weakening of density at the RBD.
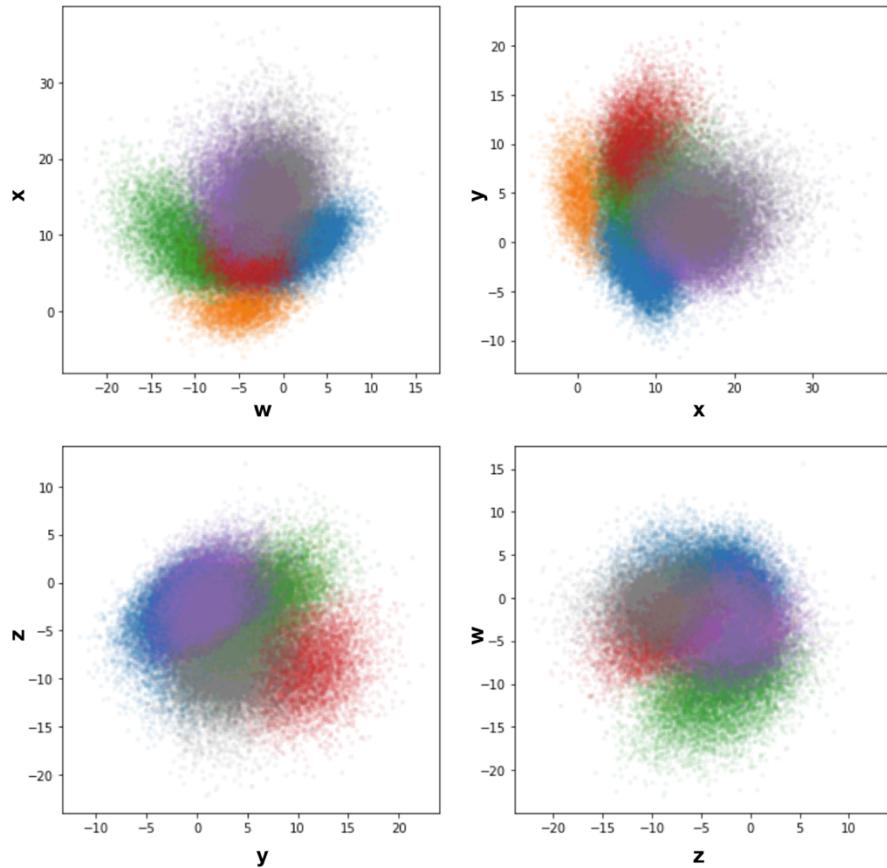
Fig S3. 50S ribosome particle distribution in the 4D encoder latent space, colored by the classification results shown in Fig 2.
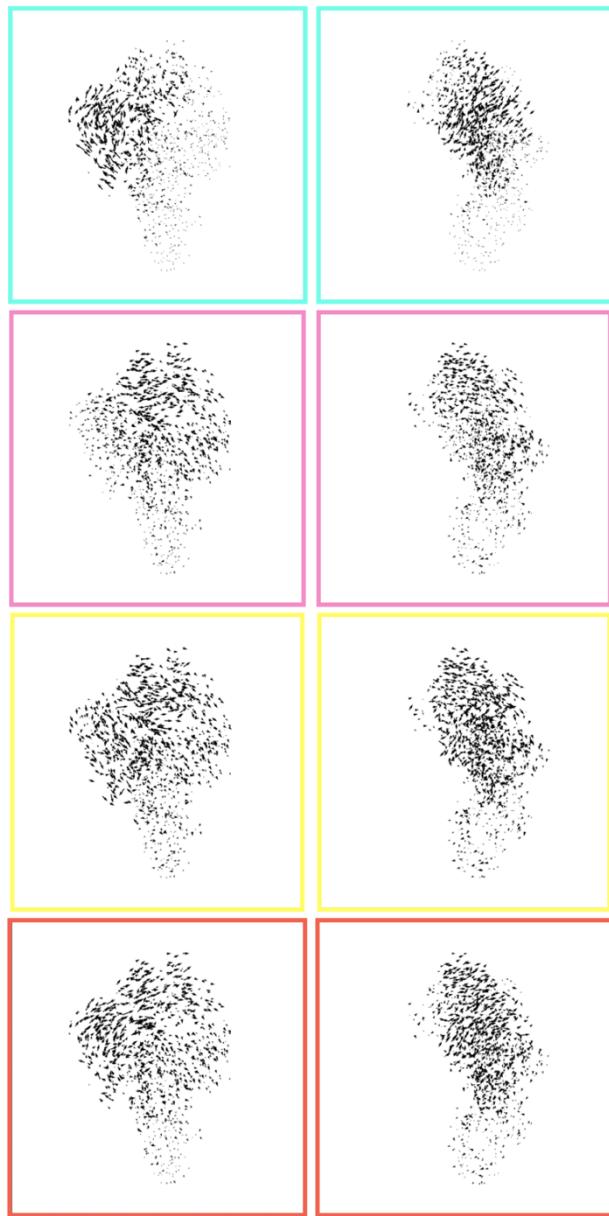
Fig S4. Front and side views of the motion trajectory vectors from the four identified motion modes of the spliceosome dataset shown in in Fig4 d-g.