In recent years, new algorithms have successfully transformed 2D cryoEM projections into a heterogeneous conformation of states from a single particle dataset.[1-7] The unique structure and size of ROOL for an RNA-only particle has suggested that a thorough application of these novel heterogeneity algorithms could reveal a conformational landscape.[8]

Section 1: 2D classification

The most traditional method for uncovering heterogeneity among a single particle dataset is through expectation-maximization (EM) or gradient-driven classification algorithms. [9,10]The two distinct "open" and "closed" forms of ROOL were first revealed from cryoSPARC's 2D EM classification jobs run on the initial particle dataset.[9] From the 2D projections, it became visually evident that some particles formed fully connected spheres, while others had separated along the C2 symmetrical axis (**Figure 1**).
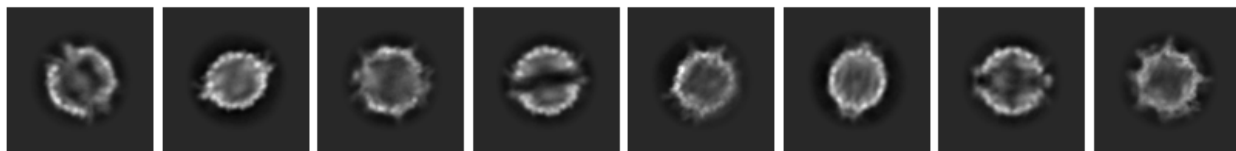
Figure 1: From a cryoSPARC 2D classification job, one can clearly distinguish "open" and "closed" classes.

Section 2: Ab-initio reconstruction

Subsequent *ab-intio reconstruction* in cryoSPARC initialized with two and three classes confirmed this hypothesis. The first reconstruction job with 10A maximum resolution on 14,842 particles (P17 J83) picked from a series of refinement jobs from a traditional cryoEM processing workflow was initialized with two classes and separated the particles into 8,066 (open) and 6,776 (closed) particles. To confirm consistency of results, the same custom parameters were used in three subsequent jobs (J83, J89, and J90) with a small variation in the number of particles classified into either consensus form.

Table 1: Classification of refined particle stacks into consensus "open" and "closed" form from ab-initio reconstruction.

	Open	Closed
J83	8,066	6,776
J84	8,066	6,776
J89	8,080	6,762
J90	8,049	6,793
Average	8065.25 +/- 0.5625	6776.75 +/- 13.16

To assess if any distinct hidden conformations were classified into either the open or closed form, multiple iterations of *ab-initio* reconstruction initialized with three classes was performed (Table 2). Consensus volumes revealed that the algorithm segregated the dataset into open, closed, and noisy classes.

Table 2: Classification results from cryoSPARC's *ab-initio* reconstruction job initialized with three classes at a maximum resolution of 10 Å, initial minibatch size of 300, and final minibatch size of 1500.

	Class 1 (open)	Class 2 (closed)	Class 3 ("junk")
J85	7,605	6,546	691
J86	7,726	6,504	612
J87	7,804	6,479	559
J88	7,733	6,543	566
Average	7717 +/- 79.36	6518 +/- 28.67	607 +/- 58.16
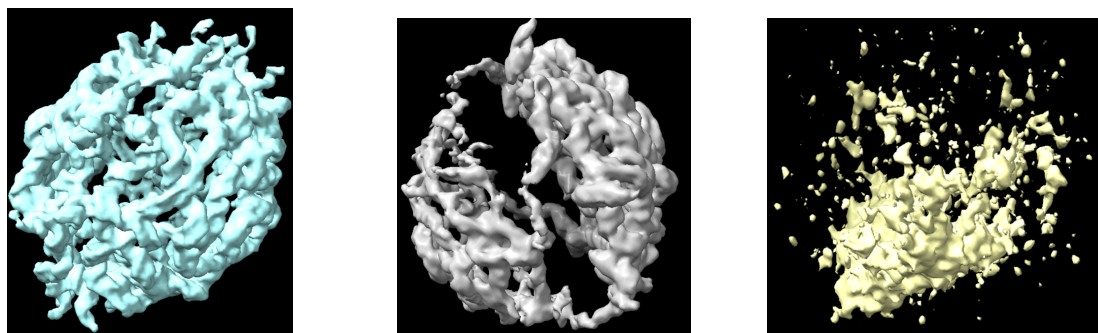
Figure 2: The three consensus volumes from cryoSPARC's *ab-initio* reconstruction with three classes, 10 Å max resolution, 300 minibatch size, 1500 final minibatch size, and the remaining default parameters (see METHODS).

After removing particles used to reconstruct the noisy class, many subsequent NU refinement jobs on both the open and closed classes were performed. Following these refinement jobs, *ab-initio* reconstruction of just the refined open particles into 3 classes revealed that there may be compositional heterogeneity among particles investigated (J291, J292 k=3), and (J293, J294 k=2).

However, the large standard deviation of particles labeled in class 1 or class 2 from the *ab-initio* reconstruction jobs with k = 3 suggests that a number of hidden conformations could be present among the dataset. Additionally, because *ab-initio* reconstruction realigns all particles before performing the EM algorithm to segregate the particles into the number of initialized classes chosen by the user, it is much more computationally burdensome.

Section 3: 3D Classification

Refined and aligned ROOL particle stacks with poses from prior *ab-initio* reconstruction, 3D alignment, and non-uniform alignment jobs are further subjected to a rigid heterogeneity analysis through cryoSPARC's 3D classification job. Similar to the 2D classification job, cryoSPARC's 3D classification job applies a hybrid online and batch expectation maximization algorithm to sort 2D particle projections into 3D classes.[9]

Eleven iterations of cryoSPARC's 3D classification were performed, varied by the number of initial classes chosen by the user (indicated by the variable k), as well as unique filter resolutions. Five jobs initialized with $k=10$, four with $k=2$, and two with $k=3$. Classification jobs with $k=10$ were performed at the following resolutions: 4A, 6A, 8A, and 10A. Outputs from those classification runs consisted of multiple junk classes (4-6 per run), with clearly two resolved classes (open and closed form) as evident from the high number of particles, ~100-1000, jumping from similar classes between iterations 5 and 6. In addition to visual inspection, consensus structures from similar classes were verified by performing a cross correlation analysis.

Table 3: Particle distributions from 3D classifications with k=10. Magenta classes were manually identified as closed, teal as open, and red as "junk".

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
435 (10A)	1237	2	8	3386	6186	6	2419	141	0	4
436 (8A)	4690	3	5712	1191	7	16	428	1330	0	12
437 (6A)	1826	1	2733	6839	0	1306	0	9	457	218
438 (4A)	14	3512	553	2	1964	2332	844	6	3700	461
448 (6A) 2000 batch size	2539	6921	1	1	1748	129	760	561	10	712

Cross correlation values that were used to validate class labels are in tables 4i-v below.

Tables 4: Cross correlation values of consensus class volumes from 3D classification jobs (i → iv. are J435, 436, 437, 438, 448, respectively)

Column1	map_1.mrc	map_2.mrc	map_3.mrc	map_4.mrc	map_5.mrc	map_6.mrc	map_7.mrc	map_8.mrc	map_9.mrc	map_10.mrc
map_1.mrc	1	0.34402594	0.140030563	0.879466176	0.326468021	0.278969109	0.873800099	0.411315858	0.191575468	0.160933852
map_2.mrc	0.34402594	1	0.334418625	0.328677446	0.297627151	0.536050081	0.337662965	0.363026887	0.605488777	0.606279254
map_3.mrc	0.140030563	0.334418625	1	0.126468316	0.123436101	0.163688689	0.137277395	0.158129752	0.284853667	0.272835732
map_4.mrc	0.879466176	0.328677446	0.126468316	1	0.258403897	0.266834885	0.895583272	0.350856155	0.180138037	0.149300084
map_5.mrc	0.326468021	0.297627151	0.123436101	0.258403897	1	0.258388907	0.299656391	0.679206192	0.150660068	0.127850309
map_6.mrc	0.278969109	0.536050081	0.163688689	0.266834885	0.258388907	1	0.278958142	0.300166696	0.328981221	0.294861704
map_7.mrc	0.873800099	0.337662965	0.137277395	0.895583272	0.299656391	0.278958142	1	0.38660872	0.187710986	0.158640236
map_8.mrc	0.411315858	0.363026887	0.158129752	0.350856155	0.679206192	0.300166696	0.38660872	1	0.171236634	0.165924594
map_9.mrc	0.191575468	0.605488777	0.284853667	0.180138037	0.150660068	0.328981221	0.187710986	0.171236634	1	0.778147697
map_10.mrc	0.160933852	0.606279254	0.272835732	0.149300084	0.127850309	0.294861704	0.158640236	0.165924594	0.778147697	1

	map_1.mrc	map_2.mrc	map_3.mrc	map_4.mrc	map_5.mrc	map_6.mrc	map_7.mrc	map_8.mrc	map_9.mrc	map_10.mrc
map_1.mrc	1	0.193777204	0.270022213	0.869846404	0.142899051	0.312106162	0.828684926	0.302678585	0.109393641	0.27226302
map_2.mrc	0.193777204	1	0.19754447	0.19352442	0.288675398	0.299079627	0.215240747	0.198873341	0.329335928	0.396755129
map_3.mrc	0.270022213	0.19754447	1	0.287196189	0.186144799	0.230498239	0.305548579	0.89022541	0.128658742	0.301024884
map_4.mrc	0.869846404	0.19352442	0.287196189	1	0.146900669	0.318049967	0.820044398	0.321862221	0.115785003	0.275637895
map_5.mrc	0.142899051	0.288675398	0.186144799	0.146900669	1	0.18201527	0.159372643	0.191068217	0.134943902	0.297184944
map_6.mrc	0.312106162	0.299079627	0.230498239	0.318049967	0.18201527	1	0.330089569	0.243808702	0.188376367	0.332553029
map_7.mrc	0.828684926	0.215240747	0.305548579	0.820044398	0.159372643	0.330089569	1	0.340843946	0.130877987	0.291916519
map_8.mrc	0.302678585	0.198873341	0.89022541	0.321862221	0.191068217	0.243808702	0.340843946	1	0.138061285	0.304806769
map_9.mrc	0.109393641	0.329335928	0.128658742	0.115785003	0.134943902	0.188376367	0.130877987	0.138061285	1	0.277118504
map_10.mrc	0.27226302	0.396755129	0.301024884	0.275637895	0.297184944	0.332553029	0.291916519	0.304806769	0.277118504	1

	map_1.mrc	map_2.mrc	map_3.mrc	map_4.mrc	map_5.mrc	map_6.mrc	map_7.mrc	map_8.mrc	map_9.mrc	map_10.mrc
map_1.mrc	1	0.139734	0.867911	0.281613	0.142774	0.854246	0.140166	0.129806	0.821641	0.392639
map_2.mrc	0.139734	1	0.137907	0.125349	0.964811	0.145288	0.996601	0.294217	0.160002	0.137724
map_3.mrc	0.867911	0.137907	1	0.271086	0.141261	0.858068	0.138619	0.124469	0.82083	0.38323
map_4.mrc	0.281613	0.125349	0.271086	1	0.128056	0.284065	0.127621	0.124525	0.304971	0.738453
map_5.mrc	0.142774	0.964811	0.141261	0.128056	1	0.148664	0.976189	0.259877	0.163246	0.144781
map_6.mrc	0.854246	0.145288	0.858068	0.284065	0.148664	1	0.145858	0.128369	0.821148	0.396735
map_7.mrc	0.140166	0.996601	0.138619	0.127621	0.976189	0.145858	1	0.285313	0.160879	0.14125
map_8.mrc	0.129806	0.294217	0.124469	0.124525	0.259877	0.128369	0.285313	1	0.134346	0.14686
map_9.mrc	0.821641	0.160002	0.82083	0.304971	0.163246	0.821148	0.160879	0.134346	1	0.416202
map_10.mrc	0.392639	0.137724	0.38323	0.738453	0.144781	0.396735	0.14125	0.14686	0.416202	1

Column1	map_1.mrc	map_2.mrc	map_3.mrc	map_4.mrc	map_5.mrc	map_6.mrc	map_7.mrc	map_8.mrc	map_9.mrc	map_10.mrc
map_1.mrc	1	0.296861053	0.345463067	0.284953833	0.299072385	0.318690211	0.30928424	0.475168258	0.311926872	0.332240731
map_2.mrc	0.296861053	1	0.310222596	0.151704609	0.873483837	0.287821293	0.851681113	0.24230966	0.264226407	0.307964146
map_3.mrc	0.345463067	0.310222596	1	0.150804549	0.315051973	0.852578342	0.32834962	0.319134891	0.857171953	0.815525413
map_4.mrc	0.284953833	0.151704609	0.150804549	1	0.149863511	0.14011851	0.156232655	0.265249133	0.13827692	0.128935963
map_5.mrc	0.299072385	0.873483837	0.315051973	0.149863511	1	0.29408434	0.844781578	0.244286776	0.271443754	0.315384299
map_6.mrc	0.318690211	0.287821293	0.852578342	0.14011851	0.29408434	1	0.306227565	0.296915889	0.899595201	0.828670502
map_7.mrc	0.30928424	0.851681113	0.32834962	0.156232655	0.844781578	0.306227565	1	0.249144003	0.280504346	0.325474799
map_8.mrc	0.475168258	0.24230966	0.319134891	0.265249133	0.244286776	0.296915889	0.249144003	1	0.298926413	0.326240957
map_9.mrc	0.311926872	0.264226407	0.857171953	0.13827692	0.271443754	0.899595201	0.280504346	0.298926413	1	0.845689893
map_10.mrc	0.332240731	0.307964146	0.815525413	0.128935963	0.315384299	0.828670502	0.325474799	0.326240957	0.845689893	1

	map_1.mrc	map_2.mrc	map_3.mrc	map_4.mrc	map_5.mrc	map_6.mrc	map_7.mrc	map_8.mrc	map_9.mrc	map_10.mrc
map_1.mrc	1	0.272906	0.206296	0.16167	0.869418	0.350745	0.845784	0.837812	0.234476	0.840857
map_2.mrc	0.272906	1	0.248057	0.200535	0.276648	0.686826	0.290619	0.295623	0.269243	0.293011
map_3.mrc	0.206296	0.248057	1	0.263744	0.213109	0.318306	0.219719	0.218293	0.43831	0.217833
map_4.mrc	0.16167	0.200535	0.263744	1	0.169484	0.253235	0.171961	0.172191	0.337424	0.178058
map_5.mrc	0.869418	0.276648	0.213109	0.169484	1	0.356302	0.846348	0.840842	0.24009	0.843709
map_6.mrc	0.350745	0.686826	0.318306	0.253235	0.356302	1	0.367107	0.375217	0.321131	0.373661
map_7.mrc	0.845784	0.290619	0.219719	0.171961	0.846348	0.367107	1	0.832944	0.24578	0.833388
map_8.mrc	0.837812	0.295623	0.218293	0.172191	0.840842	0.375217	0.832944	1	0.255016	0.832347
map_9.mrc	0.234476	0.269243	0.43831	0.337424	0.24009	0.321131	0.24578	0.255016	1	0.248899
map_10.mrc	0.840857	0.293011	0.217833	0.178058	0.843709	0.373661	0.833388	0.832347	0.248899	1

Class labels were assigned from visual inspection of volumes. Closed class cells are highlighted yellow, while open class cells are highlighted magenta. Tables i, ii, iii, and v., jobs processed at 10, 8, 6, and 10A respectively, recovered two separate open classes, each with a wide distribution of particles assigned to each class. Table v recovered 4 open classes. Tables i - v recovered 3, 3, 5, 3, 5, classes of closed particles, respectively. For J435, J437, and J448, the classes qualitatively labeled as open showed relatively "low" cross-correlation values, 0.68, 0.73, and 0.67 respectively, suggesting that they could be representing distinct intermediate conformations of the open form of ROOL. However, the inconsistent distribution of particles to each open class between replicates of classification casts doubt on this hypothesis. The relatively low cross correlation score for the two open classes is likely due to the variability of density among the solvent exposed helices, which unlike the closed conformation, are not tightly

packed with other motifs in the final tertiary structure. The separate classes qualitatively labeled "closed" for each trial have cross-correlation values all above 0.83 for every job processed. However, the consistent recovery of at least three separate closed classes for all five jobs indicates that within the particles classified as "closed," there could remain a sub-landscape of heterogeneous structures present.

To further support or cast doubt upon the hypothesis that there remains particle projections of intermediate conformations between the open and closed forms, four 3D classification jobs initialized with k=2 at 10, 8, 6, and 4A filter resolutions were processed. Each job resulted in an identical distribution of particles : 6,332 in a closed form, and 7,057 in an open form. The consistency among the results even with varying the filter resolution suggests that there are not any distinct intermediate conformations of ROOL in energetically accessible states that lie within this dataset. This conclusion was further validated by two 3D classification jobs initialized with three classes. *The particles were separated by the same distribution as the 2-class jobs, with the third class produced being assigned no particles at the final iteration.*

Section 4: 3DVar

A new algorithm embedded in cryoSPARC's software package is 3DVA (3D Variability Analysis) which uses probabilistic PCA (PPCA) to generate a consensus volume V_0 and a user chosen number (k) of eigenvolumes V_k trained on aligned particle stacks.[1] Iterations of expectation-maximization (EM) are performed to assign weights of each variable eigenvolume V_k to each 2D projection image. These weights are fundamentally latent coordinates of each particle projection that can be used to reproduce a unique 3D reconstruction when sampled from the latent space. An advantage of 3DVA to other neural network heterogeneous algorithms is that the variable components are trained on the entire dataset before latent coordinates are assigned, which prevents overfitting of variable component weights to individual particles.

Nine iterations of 3DVA were performed on the ROOL dataset, three were processed on all particles, open and closed, three closed, and three open. Variables changed within each iteration of 3DVA were the number of modes (V_k) and particle masking.

Table 5: 3DVA jobs with parameter descriptions. Note that S, C, and I refer to "simple," "cluster," and "intermediate" display modes.

	Consensus Mask	Particles	Filter Res	Modes	Display jobs
J569	J493	All	6	3	581(S) 582(C) 583(I)
J577	J493	Closed (439)	6	3	584(S) 585(C) 586 (I)
J578	J493	Open (439)	6	3	587(S) 588(C) 589(I)
J561	J439	All	6	3	564(S) 565(C) 566(I)
J562	J439	Closed (439)	6	3	571(S) 573(C) 574(I)
J563	J439	Open (439)	6	3	572(S) 575(C) 576(I)

| J579	J439	All	6	6	594(S) 596(C) 597(I)
J592	J439	closed	6	6	598(S) 599(C) 600(I)
J593	J439	open	6	6	601(S) 602(C) 603(I)

Modes from iterations trained on all particles confirmed conclusions from the rigid classification jobs: there were no eigenvolumes that were recovered to recreate structures that were clearly in between both open and closed classes (Supplementary information).
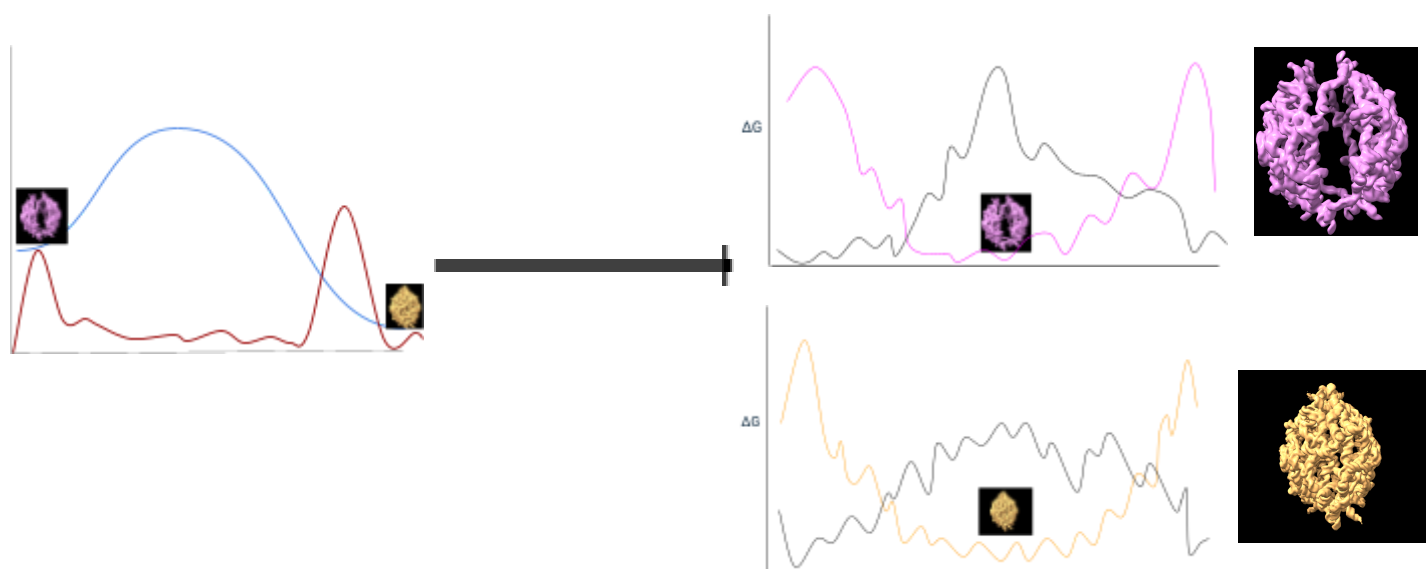
Figure 3: Probing the conformational landscapes of ROOL particles from open and closed classes separately with heterogeneity algorithms. Figure on the left represents a simplified energy landscape of ROOL (blue) superimposed with an estimated relative population of states (red) to illustrate the hypothesis that open and closed conformations represent favorable thermodynamic states separated by a kinetic barrier. Figures on the right represent individual simplified energetic landscapes of the open (top, pink) and closed (bottom, yellow) states with black lines representing an estimated relative population of particles in corresponding states. Our analysis changed from trying to recover states within the poorly populated red region between both open and closed states (left), to the lowly populated peripheral regions of either primary structure (right).

J564 simple volume series → traversing latent coordinate space across each generated eigenvolume (V_k) .
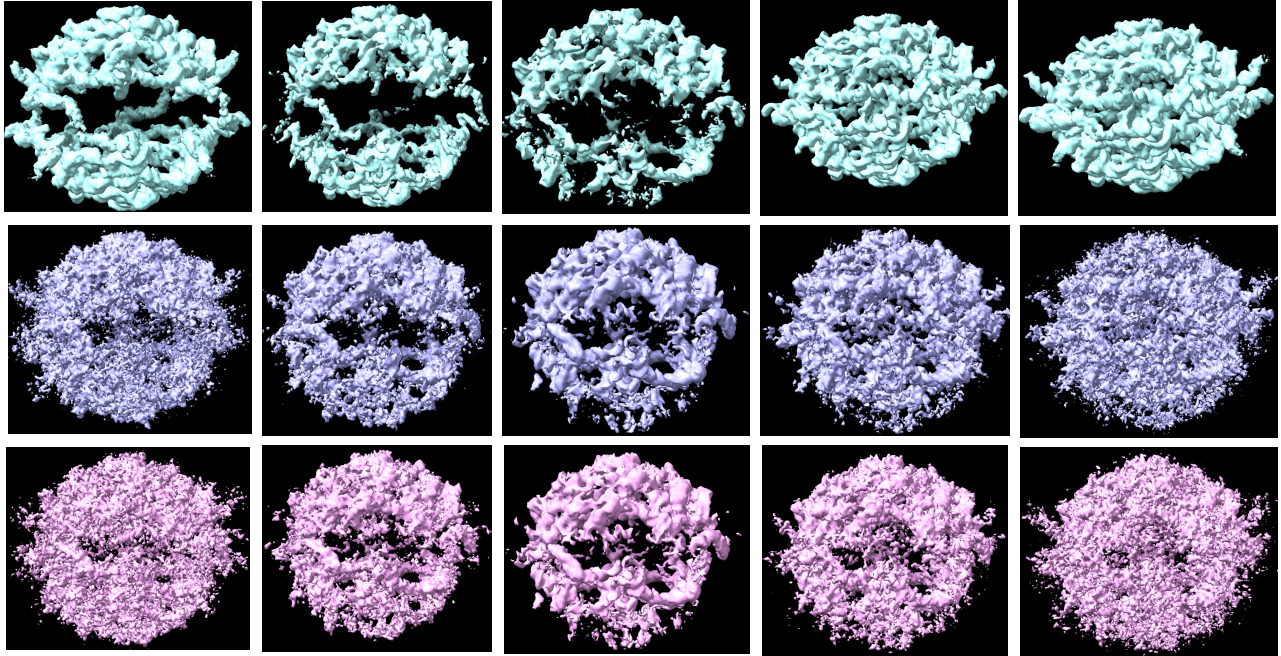
Figure 4: Frames 0, 5, 10, 15, and 19 (*note: a bit uneven because wasn't paying attention to 0-indexing*) from traversing the principal component eigenvolume V_k

As seen Figures 4 and **SI-1**, traversing the latent space of principal variable component V_1, or increasing the corresponding weight of V_k for each particle image, represents subtracting particle density from one half of ROOL's open conformation until only a noisy representation of that half is present, and then adding particle density to full form ROOL's closed conformation. This further supports the 2D and 3D classification results that counter the proposed hypothesis of particles representative of intermediate states being present within the dataset. Principal variable components V_2, and V_3 appear to traverse noise present within the particle images.
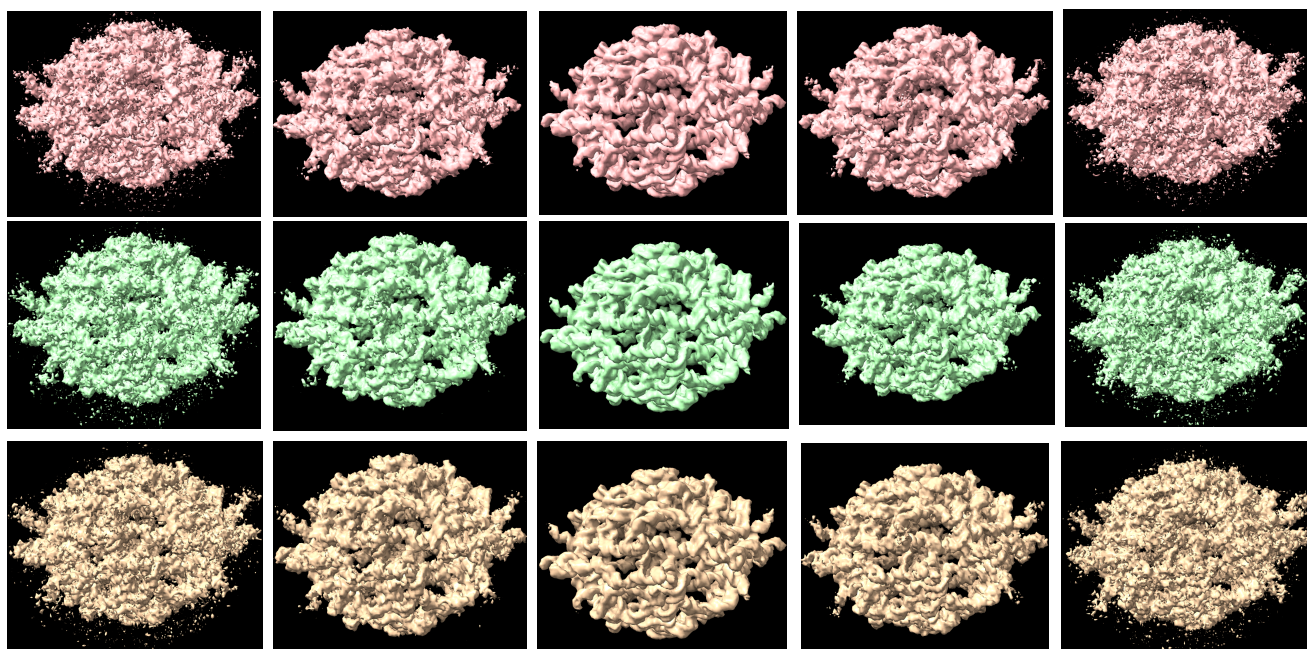
Figure 5: Frames 0, 4, 9, 14, and 19 from traversing each principal variable component eigenvolume V_1, V_2, V_3 from top to bottom for closed particles extracted from classification J439.

As seen in Figure 5, the principal variable eigenvolumes seem to just be adding or subtracting noise within the dataset, suggesting that there is not a wider conformational landscape of closed ROOL particles. This result can be used to help validate a final model map fitted to a high resolution closed consensus structure derived from this dataset.
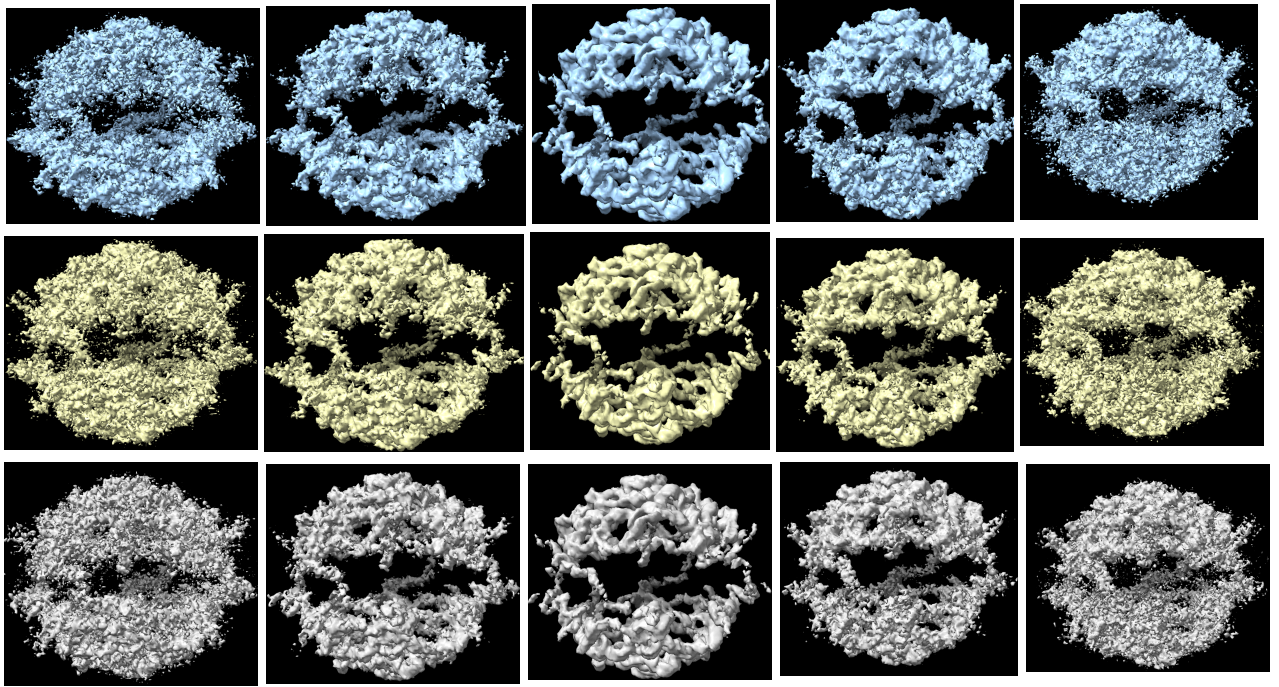
Figure 6: Frames 0, 4, 9, 14, and 19 from traversing each principal variable component eigenvolume V_1, V_2, V_3 from top to bottom for open particles extracted from classification J439.

Similar to the closed particle dataset, volumes extracted from far ends of latent coordinate space along each principal eigenvolume axis have increased presence of noise compared to structures sampled from the center of the latent coordinate axis. While it may appear that volumes extracted from the far ends of the latent space are representative of particles in the intermediate region between open and closed conformations, the particle diameter remains the same throughout the trajectory , confirming that they are just representative of noise from open form particles.

Closed NU Refined with mask
Looking at extracted volume series across all three principal variable eigenvolumes again reveals that each PC represents different modes of noise within both the J584 volume series (closed) and J587 volume series (open).

Both NU Refined with mask
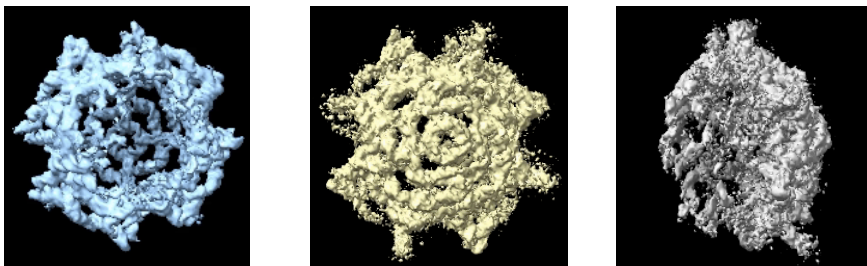
Figure 7: Components 0,1, and 2 from 3-mode 3D variability analysis at a 6A filter resolution for open and closed particles from NU refinement job with top half masked (J581)

There were no unique results from 3DVA jobs initialized to solve 6 eigenvolumes (k=6), further supporting the hypothesis that ROOL particles in the closed and open form conformations are kinetically trapped, with a narrow conformational landscape surrounding these trapped states.

Section 5: cryoDRGN

Sets of open, closed, and both particles were used to train cryodrgn, a variational autoencoder (VAE) based model which maps every particle image in a training set to its own latent coordinate through the encoder, and every latent coordinate to a reconstructed 3D volume through the decoder.

The first training run included the entire particle stack refined with a top half mask from J493. Convergence of the encoder was verified by elbow-like stabilization of the dot product, magnitude, and cosine distance of latent coordinates (Figure 8)
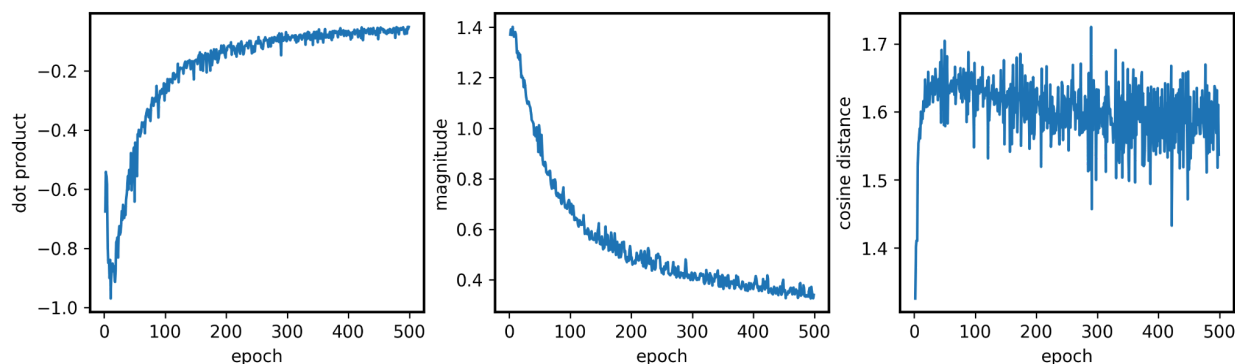
Figure 8: Dot product, magnitude, and cosine distance of latent coordinates used to verify encoder convergence.
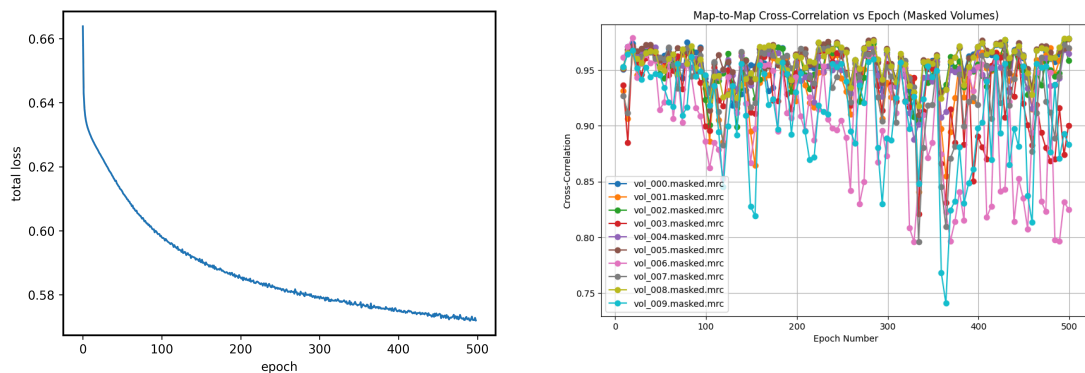
Figure 9: Total loss of VAE (left) used to demonstrate convergence of entire VAE, and cross-correlation of the decoded 3D volumes from 20 GMM k-means clustering centroids from the UMAP of latent coordinates used to verify convergence of the decoder.
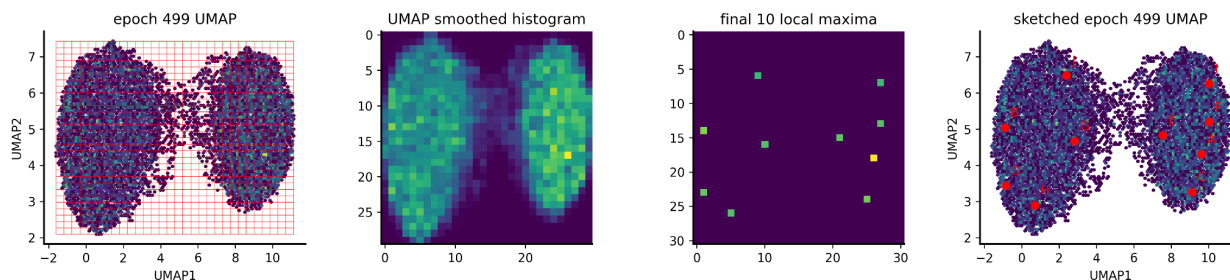

Figure 10: Final UMAP of latent space after 500 epochs of training shows two clear clusters. The left cluster represents open particles, and the right cluster represents closed particles.
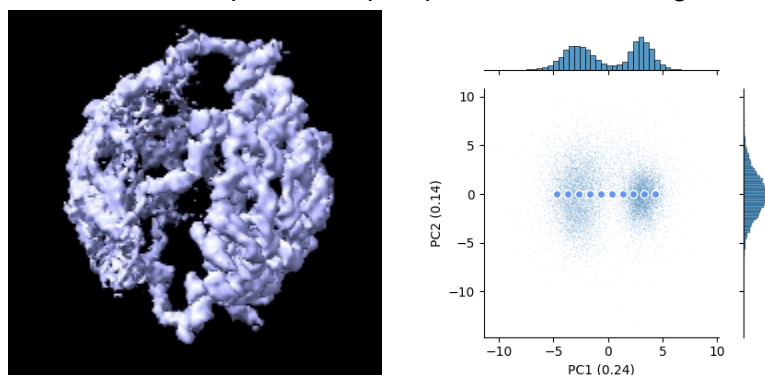

**Figure 11:** PCA 1 traversal video (left) and plot of volumes used to form frames for video (right)

PCA 1 traversal further supports the conclusion from 3D classification and 3DVA results that there are no intermediate states present within the ROOL dataset used for this reconstruction.
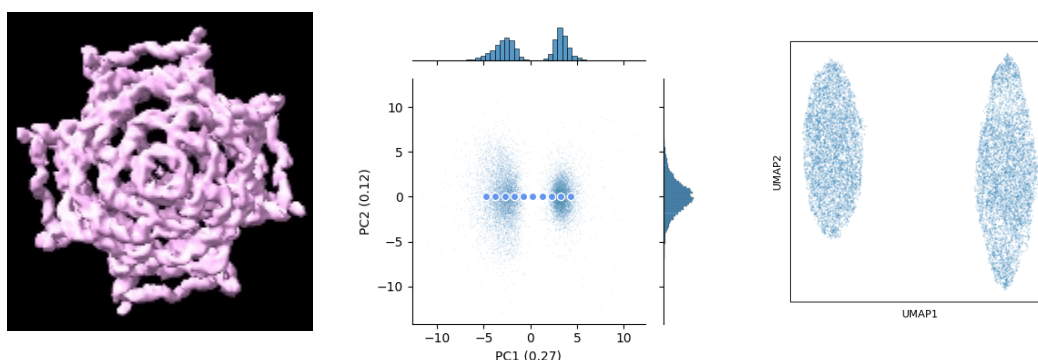
Figure 12: PC1 traversal video (far left), coordinates sampled for PC1 traversal video (middle) (notice coordinates sampled from white space), and UMAP graph of latent space showing two distinct and separate clusters. 06_cryodrgn128
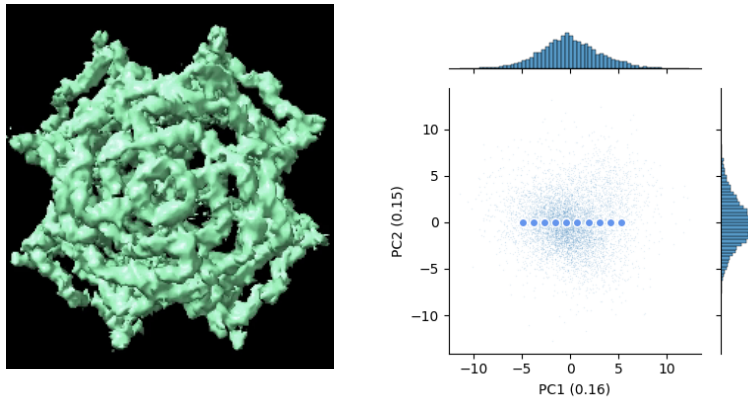
Figure 13: PC1 traversal video from open particles only (no mask) (08_cryodrgn128) with plot of coordinates sampled to generate corresponding volumes (right).
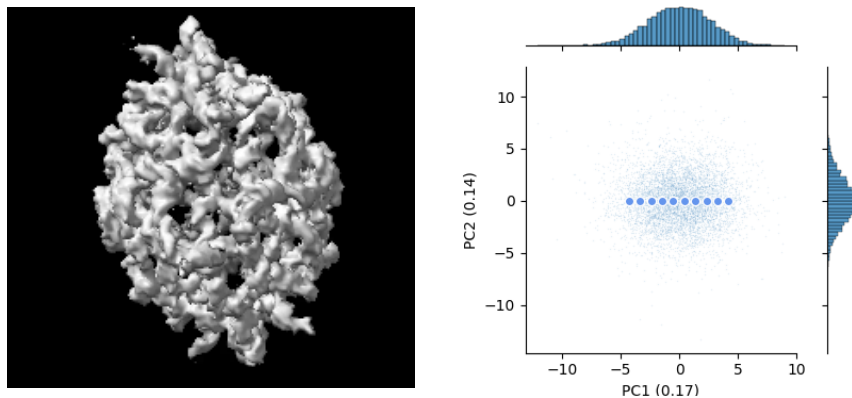
Figure 14: PC1 traversal video from closed particles only (no mask) (11_cryodrgn128) with plot of coordinates sampled to generate corresponding volumes (right)

Table 6 : Summary of cryoDRGN training iterations. Results from highlighted rows are used to draw conclusions above. More detailed results from each job displayed in the table can be found as folders in the virtual desktop 147 /data/gtully/ROOL/cryodrgn_training/

	Down_sampled particles	Z dimensions	Image size	num_epochs	Mask?
01_cryodrgn128	J493	8	128	500	Yes
01_cryodrgn_256	J493	8	256	25	Yes

| 02_cryodrgn256	J493	4	256	25	Yes
03_cryodrgn256 (ERROR in name)	J493	4	128	100	Yes
04_cryodrgn256	J439_closed	8	256	25	no
05_cryodrgn	J439_closed	4	256	500	no
06_cryodrgn	J439_both	8	128	49	no
07_cryodrgn	J439_both	4	128	50	no
08_cryodrgn	J439_open	8	128	500	no
09_cryodrgn	J439_open	4	128	50	no
10_cryodrgn128	J439_closed	8	128	200	no
11_cryodrgn128	J439_closed	8	128	500	NO
12_cryodrgn256	J439_closed	4	256	500	no
13_cryodrgn256	J439_open	4	256	500	no

Section 6: 3DFlex

We trained 3DFlex, a deep learning based ML model embedded in cryoSPARC on open, closed, and both particles separated from a previous 3D classification job with k=2 (J439).[3] The initial aim of cryoSPARC's 3Dflex algorithm is to use the recovery of heterogeneity among the particle dataset to correct for flexible regions "blurred-out" by the standard *ab-initio* reconstruction algorithm. Of the model systems trained on 3DFlex by the algorithm developers, the largest range of motion captured was 20Å by the tri-snRNP spliceosome particles.[3] In comparison, ROOL subunits vary in position by about 40Å between the open and closed form, so we are uncertain if the algorithm design can reliably capture these large motions. As summarized in **Figure 15**, the 3D flex model takes in 2D particle projections, a consensus 3D structure, and a tetrahedral mesh of the consensus structure. The model then learns a deformation flow field for each particle image to produce a convected 3D map for each particle image. Particle image latent coordinates, the deformation flow field, and the canonical 3D map are jointly trained on an MSE loss function between original 2D images and projections of the convected 3D map using the experimentally determined poses, CTF, and noise parameters previously determined for each image.
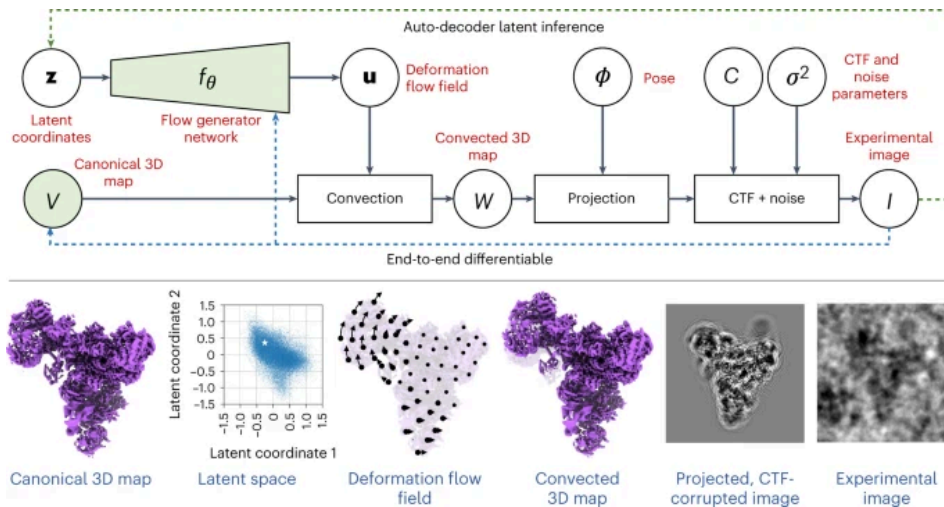
Figure 15: 3DFlex model.[3]

We first followed the 3D flex training pipeline (3D flex data prep, mesh prep, training, generation, and reconstruction) with particles isolated in closed and open form, using standard parameters for mesh creation and flex training. We found that the heterogeneous conformation of states generated from these initial training runs recovered little range of motion or conformation of states beyond the consensus structures for both the open and closed classes. However, we postulate that because there must exist a pathway between the two consensus classified structures, there is likely, within a population of thousands of particles, some particles that should be found along the pathway between open and closed. As a result, we relaxed the rigidity constraints within the algorithm for training.

For open form particles only, we decreased the default rigidity prior on regions of lowest density from 0.5 to 0.3, and increased the tetrahedron grid span number from 20 to 40 (both in the mesh prep job). We then decreased the rigidity penalty in training from 2 to 0.1. A slightly more evident range of motion became present when viewing volume series generated along two orthogonal latent dimensions.
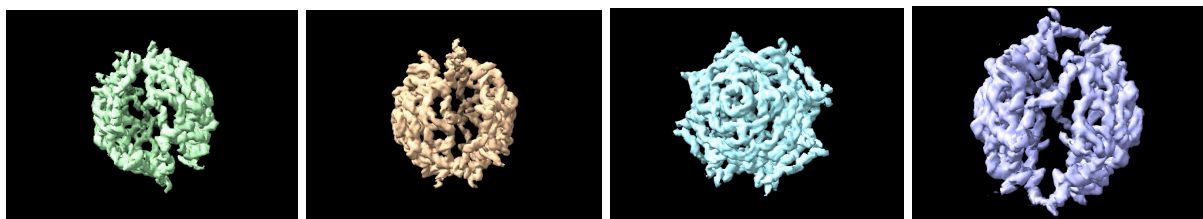
Series 1 and 2 from default dimensions (extracted from Flex Generator job 462) , compared with 1 and 2 from relaxed parameters (extracted from Flex Generator job 482) .

As a result, we decided to train 3D flex on open and closed particles together, with both default rigidity parameters as well as relaxed parameters.
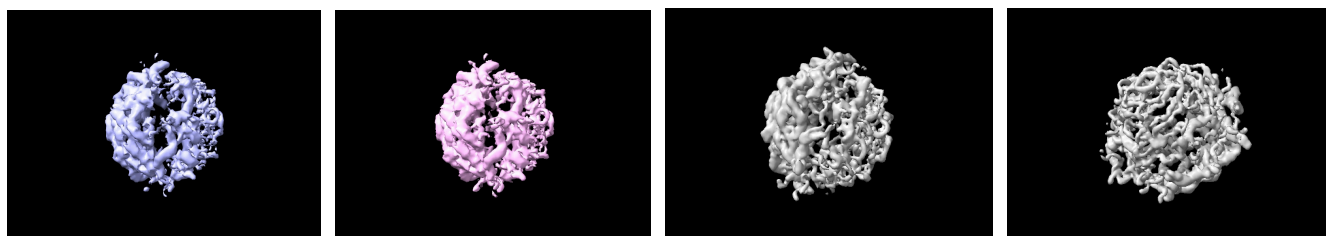
Above first two series along default trajectory along latent coordinates 1 and 2, with the default parameters and relaxed parameters.

Increasing the model size from 2 to 4 latent dimensions, as well as relaxing the rigidity parameter even more (down to 0.1), allowed for recovery of an almost complete open-closed range of motion. Additionally, flex reconstruction helped improve the resolution on the periphery of the molecules (see figures below).

Table 7: Summary of 3D Flex training jobs performed on ROOL dataset from J434 NU refinement, and labeled open or closed from J439 classification job.

	Particle label	LD	Rigidity (prior)	Rigidity(lambda)	tetrahedron	Initial low res flex Generator job	Flex reconstruction job	High res flex generate job	Job Time (closest hr)
J451	C	2	Min = 0.5	2	20	J452	J453	J619	3
J455	C	2	Min = 0.5	2	20	J463	J604	J618	3
J458	C	2	Min =0.5	1	20	J476	J605	J617	3
J468	C	8	Min =0.5	2	20	J483	J606	J616	10
J461	O	2	Min =0.5	2	20	J462	J615	J631	4
J467	O	4	Min =0.5	2	20	J478	J614	J630	7
J480	O	2	Min =0.3	0.1	40	J482	J613	J629	10
J485	O	4	Min =0.1	0.1	40	J503	J610	J612	17
J466	B	2	Min =0.5	2	20	J469, J470, J473	J471	J609	3

| J475 | B | 2 | Min =0.3 | 2 | 40 | J477 (mid-training) J481 | J611 | 628 | 10 |
| J487 | B | 4 | Min =0.1 | 0.1 | 40 | J504 | J505 | J506 | 17 |

Following the typical workflow of 3D flex, improvement of resolution is evaluated from FSC scores from the cryoSPARC's flex reconstruction job.

Closed Particles, No Mask
From training J468 on closed particles only (parameters can be seen in **Table 7**), 3DFlex appears to slightly improve final resolution of consensus structure as can be seen by side by side FSC plots of rigid and flex reconstructed volumes (**Figure 15**) . Additionally, increasing the latent dimensions of the model increases resolution improvement from the flex model, as J468 was trained with 8LD had GSFSC Resolution of 5.01A, while J455 trained with 4LD had a GSFSC Resolution of 5.08A, compared to the GSFSC rigid resolution of 5.12A.
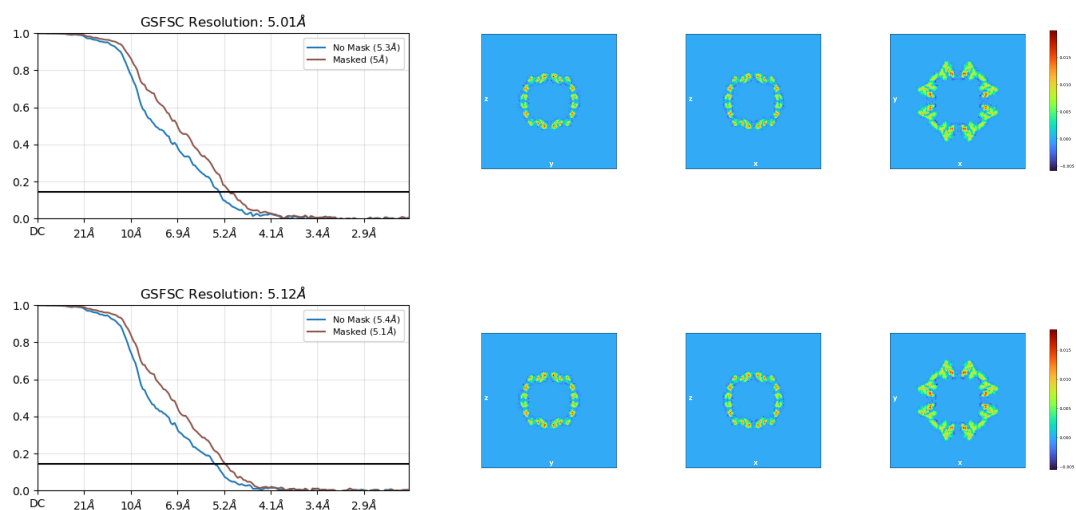
Figure 15: Top- Flex generated volume FSC plot (left) and flex reconstruction FSC-filtered full map (right). Bottom- rigid reconstruction FSC (right) and rigid-reconstruction FSC-filtered full map..
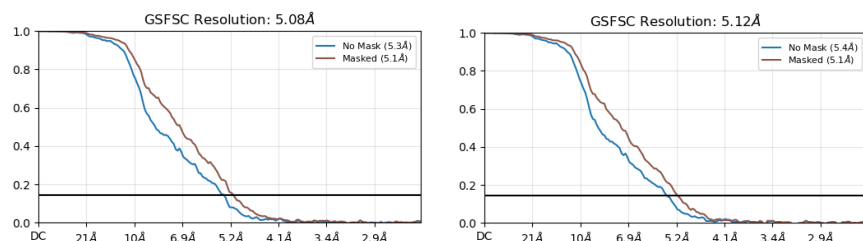
Figure 16: Flex-generated FSC plot from closed-particle training J455 (parameters found in Table 7) (left) and corresponding rigid-generated FSC plot (right)

Open Particles, No Mask

As seen in **Figure 17** below, the resolution of reconstructed volumes from 3DFlex on open particles improves from flex-generated volumes compared to rigid generated volumes. This improvement also increases as the number of model latent dimensions increases, from a 0.91A to a 1.11A improvement from a 2LD to 4LD model.
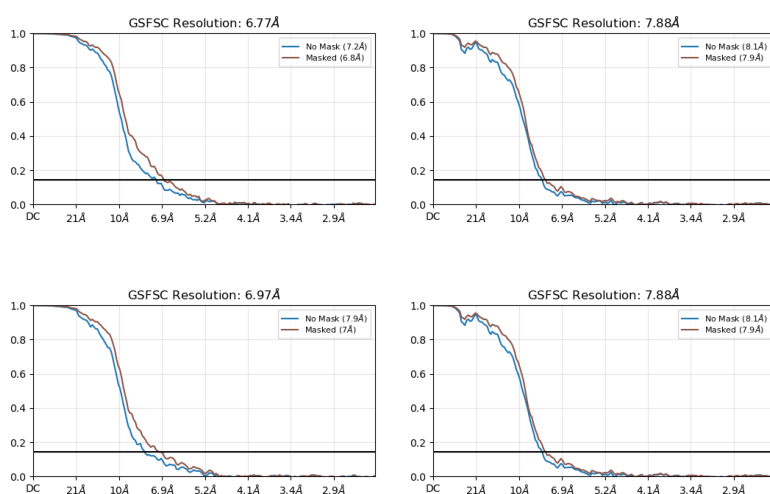

Figure 17: FSC flex-generated (left) and rigid (right) plots from J614(full parameters can be seen in **Table 7**), with 4LD (top) and same results from J615 (bottom). Both jobs are trained on open particles only.

Open and Closed, No Mask

As seen in **Figure 18** below, the resolution of reconstructed volumes from 3DFlex on open and closed particles improves from flex-generated volumes compared to rigid generated volumes. This improvement also increases as the number of model latent dimensions increases, from a 0improvement from a 2LD to 4LD model.
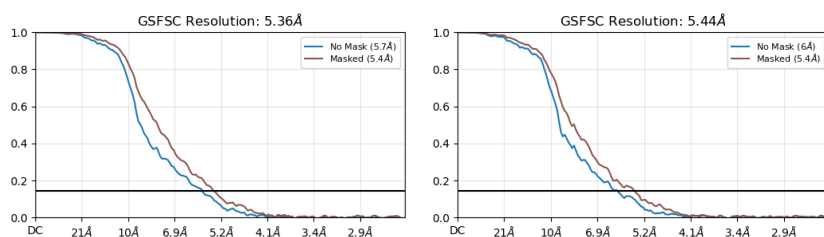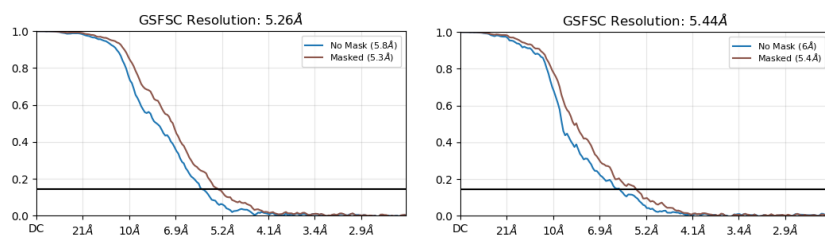
Figure 18: Top: FSC plots from J471 with 2LD on open and closed particles (flex generated on left, rigid generated on right) and the same thing from J505 with 4LD on the bottom.

Masking particles

To prevent potential bias in the 3DFlex training jobs on open and closed particles together (decoder is just capturing density being translated from open to close), NU refinement jobs with a mask over one half of the ROOL particles was used for further training. The purpose of this procedure is so that the particles are not aligned on the center of mass, but aligned from the reference point of the top half.

Procedure for constructing mask: I aligned both the Class 0 Class 1 volumes from J439, put a spherical eraser over the top half where all regions are aligned, erased all that is not in the sphere, and created a mask.

Volume of this partial particle : Volume J491.mrc
Mask: J492_mask.mrc

Table 8: Summary of 3DFlex jobs performed with ROOL particles refined from J493 NU(open and closed), J495 NU(closed), J496 NU (open), and J497 HR (open and closed), J498 HR (closed), J499 HR (open). Original labels of open and closed assigned from J439 3D classification with k=2.

	NU or Homo	Open/closed	LD	Rigidity (prior)	Rigidity penalty	tetrahedron	Job time	Flex generator job	Flex reconstruct job	Flex generator job
J536	NU	both	2	0.1	0.1	20	3	J537	J620	J632
J552	NU	both	4	0.1	0.1	40	15	J558	J559	J621
J525	NU	closed	2	0.5	2	20	3	J539	J607	J622

J527	NU	open	2	0.5	2	20	3	J541	J623	J633
J530	Homo (J497)	both	2	0.5	2	20	3	J542	J624	J634
J546	Homo	both	4	0.1	0.1	40	15	J557 and J560	J625	J635
J549	Homo	closed	2	0.5	2	20	3	J553	J608	J626
J550	Homo	open	2	0.5	2	20	3	J554	J627	J636

Conclusion

We conclude that the open and closed conformations extracted from this ROOL cryoEM SPA are kinetically trapped states. However, we acknowledge that these results do not mean a negligible number of intermediate states of ROOL are populated in biological environments, but our dataset of particle projections from the cryo-EM microscope did not capture sufficient population of these states, and thus we can confidently fit atomic models to our high resolution model maps of open and closed particles from this dataset.

Supplementary Figures
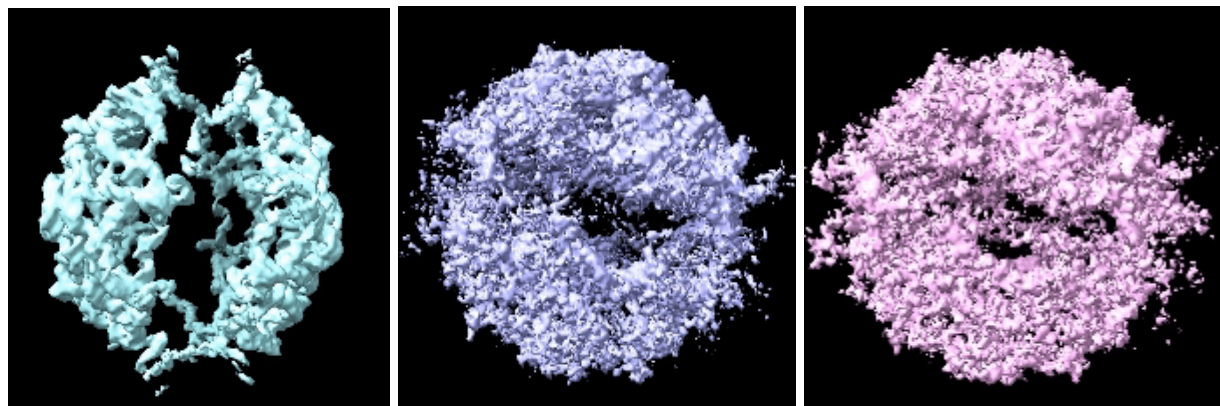
SI I: Traversing the latent coordinates (or weights) of three variability component eigenvolumes (V_k) generated from 3DVA of open and closed particles.

Bibliography

(1)   Punjani, A.; Fleet, D. J. 3D Variability Analysis: Resolving Continuous Flexibility and Discrete Heterogeneity from Single Particle Cryo-EM. J. Struct. Biol. 2021, 213 (2), 107702. https://doi.org/10.1016/j.jsb.2021.107702.

(2)   Zhong, E. D.; Bepler, T.; Berger, B.; Davis, J. H. CryoDRGN: Reconstruction of Heterogeneous Cryo-EM Structures Using Neural Networks. Nat. Methods 2021, 18 (2), 176–185. https://doi.org/10.1038/s41592-020-01049-4.

(3)   Punjani, A.; Fleet, D. J. 3DFlex: Determining Structure and Motion of Flexible Proteins from Cryo-EM. Nat. Methods 2023, 20 (6), 860–870. https://doi.org/10.1038/s41592-023-01853-8.

(4)   Schwab, J.; Kimanius, D.; Burt, A.; Dendooven, T.; Scheres, S. H. W. DynaMight: Estimating Molecular Motions with Improved Reconstruction from Cryo-EM Images. bioRxiv, 2023, 2023.10.18.562877. https://doi.org/10.1101/2023.10.18.562877.

(5)   Frank, J.; Ourmazd, A. Continuous Changes in Structure Mapped by Manifold Embedding of Single-Particle Data in Cryo-EM. Methods 2016, 100, 61–67. https://doi.org/10.1016/j.ymeth.2016.02.007.

(6)   Chen, M.; Ludtke, S. J. Deep Learning-Based Mixed-Dimensional Gaussian Mixture Model for Characterizing Variability in Cryo-EM. Nat. Methods 2021, 18 (8), 930–936. https://doi.org/10.1038/s41592-021-01220-5.

(7)   Herreros, D.; Lederman, R. R.; Krieger, J. M.; Jiménez-Moreno, A.; Martínez, M.; Myška, D.; Strelak, D.; Filipovic, J.; Sorzano, C. O. S.; Carazo, J. M. Estimating Conformational Landscapes from Cryo-EM Particles by 3D Zernike Polynomials. Nat. Commun. 2023, 14 (1), 1–10. https://doi.org/10.1038/s41467-023-35791-y.

(8)   Zhang, W.-B.; Cheng, S. Z. D. Giant Is Different: Size Effects and the Nature of Macromolecules. Giant (Oxf.) 2020, 1 (100011), 100011. https://doi.org/10.1016/j.giant.2020.100011.

(9)   Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A. cryoSPARC: Algorithms for Rapid Unsupervised Cryo-EM Structure Determination. Nat. Methods 2017, 14 (3), 290–296. https://doi.org/10.1038/nmeth.4169.

(10)   Kimanius, D.; Dong, L.; Sharov, G.; Nakane, T.; Scheres, S. H. W. New Tools for Automated Cryo-EM Single-Particle Analysis in RELION-4.0. Biochem. J. 2021, 478 (24), 4169–4185. https://doi.org/10.1042/bcj20210708.