Supplementary Tables

Dataset	***# Particles***	***Image size*** *(pixels)*	***Pixel size*** *(A/pix)*	**|z|**	***Architecture***	***Epochs***	***Total training time*** *(hr; 1 GPU)*	***Figure***
Simulated Uniform	50,000	128	6	1	E: 256x3; D: 512x5	100	6.7	3
Simulated Cooperative	50,000	128	6	1	E: 256x3; D: 512x5	100	6.7	3
Simulated Noncontiguous	50,000	128	6	1	E: 256x3; D: 512x5	100	6.7	3
Simulated Compositional	50,000	128	3	1	E: 256x3; D: 512x5	100	6.7	3
RAG complex (EMPIAR-10049)	108,544	192	1.23	10	1024x3	25	10.4	4
Pf80S ribosome (EMPIAR-10028)	105,247	128	3.76875	10	256x3	50	6.2	N/A
	93,852	256	1.884375	10	1024x3	25	17.1	4
Assembling LSU (EMPIAR-10076)	131,899	128	3.275	1	256x3	50	6.3	5
	131,899	128	3.275	10	256x3	50	6.3	5
	87,328	256	1.6375	10	1024x3	50	31.2	5
	87,328	256	1.6375	10	1024x3	50	31.2	S8
	87,328	256	1.6375	8	1024x3	50	19.7*	S8
Pre-catalytic spliceosome (EMPIAR-10180)	327,490	128	4.2475	10	256x3	50	17.0	6
	139,722	256	2.12375	10	1024x3	50	52.9	6

Supplementary Table 1. Summary of dataset statistics, training hyperparameters, and runtimes for cryoDRGN heterogeneous reconstruction experiments.

The neural network architecture is denoted as $d \times l$, where d indicates the number of nodes per layer and l is the number of hidden layers. The architecture corresponds to both the encoder (E) and decoder (D) MLPs unless otherwise specified.

Total training times were recorded from training on a single Nvidia Tesla V100 32GB memory GPU card on either an Intel Xeon Gold 6130 CPU (2.10GHz, 791GB of RAM) or an IBM Power9 node with 1.2 TB of RAM. The reported training times may be overestimated since the presence of any concurrently running programs was not controlled for. (*) The training time of the third replicate of EMPIAR-10076 is substantially faster as using $|z| = 8$ better satisfies tensor shape constraints for Nvidia Tensor Core hardware acceleration.

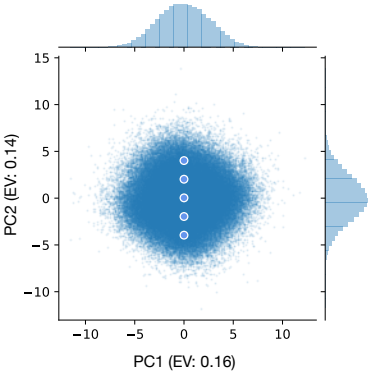
Supplemental Video 1: CryoDRGN reconstruction of the RAG1-RAG2 complex. Structures (left) and their corresponding location in the latent space (right). The series of structures shown are generated at the *k*-means cluster centers of the latent encodings with *k*=20, followed by a trajectory generated with cryoDRGN's graph traversal algorithm.

Supplemental Video 2: CryoDRGN reconstruction of the *Pf*80S ribosome. Structures from two distinct views (top) and their corresponding location in the latent space (bottom). The series of structures shown are generated at the *k*-means cluster centers of the latent encodings with *k*=20, followed by a trajectory generated with cryoDRGN's graph traversal algorithm.

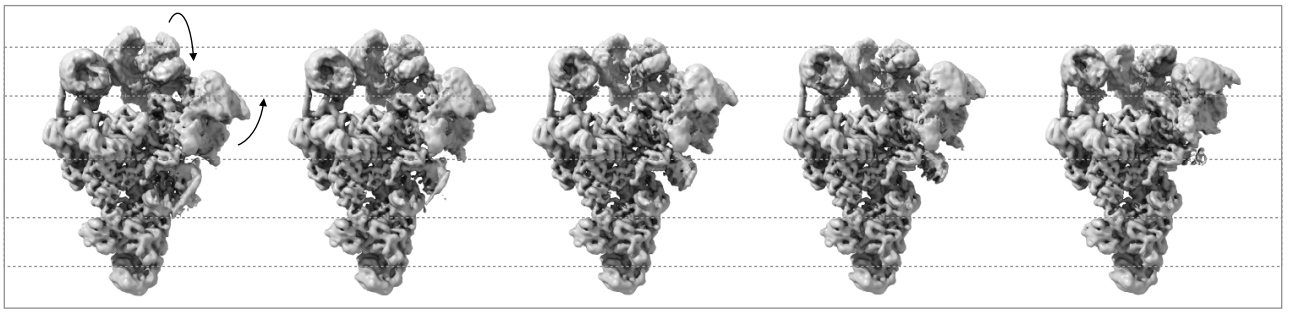
Supplemental Video 3: CryoDRGN reconstruction of the bacterial large ribosomal subunit undergoing assembly. Structures (left) and their corresponding location in latent space (right), which is colored by minor assembly state as in **Figure 5H**. Trajectories are generated with cryoDRGN's graph traversal algorithm through the three parallel assembly pathways assigned in Davis et al[12].

Supplemental Video 4: CryoDRGN reconstruction of the pre-catalytic spliceosome. Front and back view of structures (left) and their corresponding locations in latent space (right). Trajectories show density maps generated from PC1 and PC2 of the latent encodings and cryoDRGN's graph traversal algorithm.

A



B



PC2 = 4.02 PC2 = 2.02 PC2 = 0.03 PC2 = -1.96 PC2 = -3.96

Supplementary Figure 1. Additional structures of the pre-catalytic spliceosome reconstructed by cryoDRGN.

A) PCA projection of the 10-D latent encodings from cryoDRGN as in **Figure 6c** with 5 points along PC2.

B) Structures generated by traversing along PC2 of the latent space representation at points shown in (**A**).