Supplementary Material
CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images

Axel Levy[1,2,*], Frédéric Poitevin[1,*], Julien Martel[2,*], Youssef Nashed[3], Ariana Peck[1], Nina Miolane[4], Daniel Ratner[3], Mike Dunne[1], and Gordon Wetzstein[2]

[1]LCLS, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
[2]Stanford University, Department of Electrical Engineering, Stanford, CA, USA
[3]ML Initiative, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
[4]University of California Santa Barbara, Department of Electrical and Computer Engineering, Santa Barbara, CA, USA

A Experimental Details

CryoAI and all the baselines are run on a single Tesla V100 GPU with 8 CPUs. Fourier Shell Correlations between two voxel grids (aligned with an exhaustive search over $SO(3) \times \mathbb{R}^3$) are measured using the software EMAN v2.91 [14].

A.1 Generation of Ground Truth Volume from Atomic Model

The python library gemmi [15] was used to assemble a complete atomic model of the 80S Ribosome [16] from its small (PDB:3J7A) and large (PDB:3J79) subunits deposited in different files in the Protein Data Bank (PDB) [3]. The atomic models of the pre-catalytic Spliceosome (PDB:5NRL) [7] and the SARS-CoV-2 spike ectodomain structure (PDB:6VYB) [2] were used without pre-processing after being downloaded from the PDB.

ChimeraX [6] was used to simulate a noise-less electrostatic potential volume discretized on a cubic grid from each atomic model. First, an empty cubic grid of desired spacing and extent is centered on the center of mass of the atomic model. Second, a map of same grid spacing is simulated from the atomic model and resampled on the first map, before being saved to a MRC file [1].

A.2 CryoAI

Each minibatch contains 32 images when $L = 128$ and 128 images when $L = 64$. We use the Adam optimizer [4], with a learning rate of 10^{-4}.

A.3 CryoSPARC

CryoSPARC v3.2 [8] was used. We followed the typical workflow: *Import* particle stacks, perform *Ab Initio* reconstruction before *Homogeneous Refinement*. Default parameter values were used except when noted in the text.

A.4 CryoDRGN2 [18]

All results from cryoDRGN2 are directly reported from the main paper or the supplementary materials, because source code for cryoDRGN2 is currently not available.

A.5 CryoPoseNet [5]

Volumes are represented with a voxel grid in real space and the image formation model is simulated in real space in the decoder. The encoder outputs rotations R_i in the 6-dimensional space $\mathcal{S}^2 \times \mathcal{S}^2$. Each minibatch contains 32 images when $L = 128$ and 128 images when $L = 64$. The Adam optimizer [4] is used, with a learning rate of 10^{-3}.

B Encoder Architecture

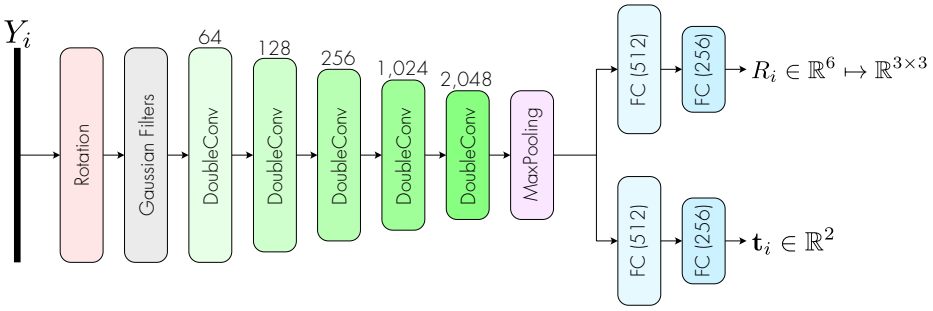

Fig. S1. Architecture of the encoder. The "rotation" layer duplicates each image and rotates one of the copies. "Gaussian filters" are low-pass filters. "FC" stands for "fully connected".

The encoder takes as input an image Y_i and outputs a rotation matrix R_i and a translation vector $\mathbf{t}_i$. The rotation is represented in the 6-dimensional space $\mathcal{S}^2 \times \mathcal{S}^2$ [19] and converted in a matrix using the PyTorch3D library [9]. The architecture of the encoder is summarized in Fig. S1. Each image is first duplicated by a "rotation" layer, which applies an in-plane rotation of π to one of the duplicates. Those duplicates are then concatenated batch-wise. Each image is filtered by a set of 5 Gaussian filters of size 11 with cutoff frequencies distributed geometrically between 0.1 and 10 pix^{-1}. Empirically, we found this multi-scale representation to lead to more robust convergence of our framework in some cases. This set of filtered images is then fed into five consecutive "DoubleConv" layers generating respectively 64, 128, 256, 1024 and 2048 channels. A "DoubleConv" layer contains two convolutional layers with kernels of size 3 and a max-pooling layer dividing the height and the width of each image by

2. The last "DoubleConv" layer is followed by another max-pooling layer after which the dimension of each image is finally divided by $2^6 = 64$. The feature vector has, at this point, 8182 dimensions when the input image has a size of 128^2 pixels. This feature vector is finally fed into two separated fully connected neural networks with ReLU activation functions and two hidden layers of sizes 512 and 256.

C Implicit Representation in Fourier Domain

C.1 Electrostatic Potentials in Fourier Space

The electrostatic potential $V(\mathbf{r})$ derives from the distribution of charges following the Poisson-Boltzmann equation [10]. We know that the smoothness of the function V at low-to-medium resolution will translate into a rapid decrease of its Fourier coefficients $\hat{V}$. In particular, if we assume that V is a smooth function of $\mathbf{r}$, in the sense that it is an α-Lipschitz function, Sampson and Tuy [12] demonstrated that the norm of its Fourier transform $|\hat{V}(\mathbf{k})|$ decreases as least in $1/k^\beta$ for all $\beta < \alpha$. In Fig. S2, we analyze the Fourier transform of the electrostatic potential of an adenylate kinase molecule ($L = 128$). We observe that $|\hat{V}(\mathbf{k})|$ indeed decreases rapidly with $|\mathbf{k}|$, which implies that $\hat{V}$ varies on several orders of magnitude.

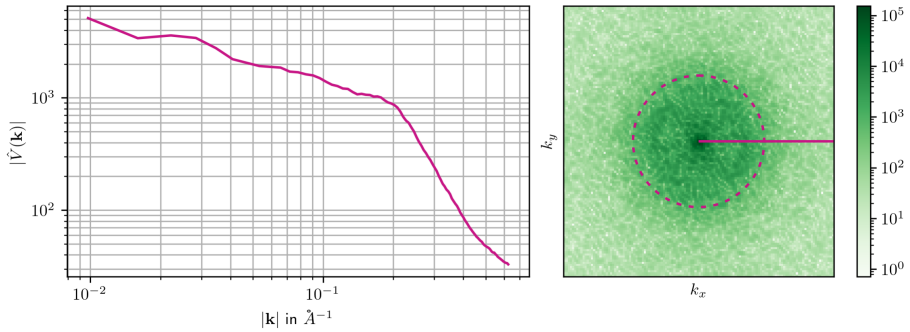

Fig. S2. (Left) $\hat{V}$ is the Fourier transform of the electron scattering potential of adenylate kinase (128^3 voxels of size 3.2 Å). The radial mean of $|\hat{V}|$ decreases following a power law of $|\mathbf{k}|$ and varies over 2 orders of magnitude. (Right) Slice $(k_x, k_y, 0)$ of $|\hat{V}|$, varying over 5 orders of magnitude.

C.2 Performance of FourierNet on 2D images

We show here that the architecture proposed in Fig. 2 is relevant for representing the Fourier transform of "natural" 2D signals by comparing it with a

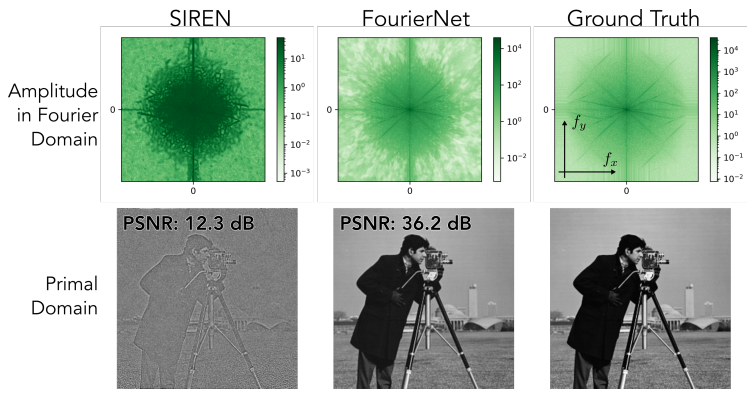

Fig. S3. Approximations of the Fourier transform of a "natural" image with a SIREN [13] and a FourierNet, and their inverse Fourier transforms. The SIREN poorly fits signals that vary over several orders of magnitude.

simple SIREN. Using gradient-based optimization we optimize the weights of a SIREN [13] and of a FourierNet to approximate the Fourier transform of a real-world 2D image. For the comparison to be relevant, we use the same number of optimizable parameters (300k) in the SIREN and in the FourierNet. The results of the experiment are shown in Fig. S3. SIRENs are built in a way such that the weights follow a normalized distribution and they can only efficiently represent normalized functions. The ground truth Fourier transform varies on more than 6 orders of magnitude, leading to a poor reconstruction when taking the inverse Fourier transform of the approximated function with a SIREN. On the opposite, the FourierNet can precisely fit the given Fourier transform, leading to a quantitatively better reconstruction in primal (real) domain. Artifacts, however, can still be observed at high frequency.

C.3 Architecture Details

Our neural representation uses two SIRENs with 256 hidden features, mapping $\mathbb{R}^3$ to $\mathbb{R}^2$. The SIREN preceding the exp (see Fig. 2) has 2 hidden layers while the other one contains 3 hidden layers. The total number of parameters is only 330,240, which can be compared to the number of parameters used in a voxel-based representation (slightly faster) of resolution 128: $128^3 = 2{,}097{,}152$ (Table S1 also indicates the runtime on the *kinase ideal* dataset).

Table S1. Comparison between *coordinate-based* and *voxel-based* representations.

Representation	Runtime	# Parameters ($L = 128$)
FourierNet	0:09h	**330,240**
Voxel-grid	**0:06h**	2,097,152

D Avoiding Spurious Planar Symmetries

D.1 Handedness Ambiguity in cryo-EM

In cryo-EM, the interaction of the electron beam with the electrostatic potential V in the orientation R_i corresponds to an orthographic transparent projection described by

$$Q_i = Q(R_i) : (x, y) \mapsto \int_z V(R_i[x, y, z]^T)dz. \tag{1}$$

For a set of projections $\{Q_i\}$, if the associated orientations are not given, there exists an instrinsic ambiguity on the handedness of the volume V [11]. That is to say, one cannot distinguish a set of projections obtained with V from any other set of projections obtained with a "mirrored" version of V.

More specifically, let us fix an orthonormal basis $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ on $\mathbb{R}^3$. We consider a volume $V : \mathbb{R}^3 \to \mathbb{R}$ and define

$$\tilde{V}(x, y, z) = V(F[x, y, z]^T), \tag{2}$$

where

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{3}$$

$\tilde{V}$ is the mirrored version of V with respect to the (x, y) plane. We also define the orientations R_i using Euler angles in the "ZYZ" proper Euler convention. That is, each rotation matrix is parameterized by α_i, β_i, γ_i and

$$R_i = R_{\alpha_i,\beta_i,\gamma_i} = \begin{bmatrix} c_1c_3 + s_1s_2s_3 & c_3s_1s_2 - c_1s_3 & c_2s_1 \\ c_2s_3 & c_2c_3 & -s_2 \\ c_1s_2s_3 - c_3s_1 & c_1c_3s_2 + s_1s_3 & c_1c_2 \end{bmatrix}, \tag{4}$$

where c and s represent cosines and sines ($e.g$ $c_1 = \cos\alpha_i$). Using this definition, we can show that

$$FR_i = \tilde{R}_i F \quad \text{where } \tilde{R}_i = R_{\alpha_i+\pi,\beta_i,\gamma_i+\pi}. \tag{5}$$

Therefore,

$$\begin{aligned} \tilde{Q}_i = Q(\tilde{R}_i) : (x, y) \mapsto & \int_z \tilde{V}(\tilde{R}_i[x, y, z]^T)dz \\ &= \int_z \tilde{V}(\tilde{R}_i[x, y, -z]^T)dz \\ &= \int_z \tilde{V}(\tilde{R}_i F[x, y, z]^T)dz \\ &= \int_z \tilde{V}(FR_i[x, y, z]^T)dz \\ &= \int_z V(R_i[x, y, z]^T)dz \\ &= Q_i(x, y) \end{aligned} \tag{6}$$

The second equality is a change of variable $z \to -z$. In conclusion, the volume V associated with the set of orientations R_i will give the same set of projections as the volume $\tilde{V}$ with the set of orientations $\tilde{R}_i$. For a symmetrical volume ($V = \tilde{V}$), R_i and $\tilde{R}_i$ will give the same projections (Fig. S4, right).

D.2 Spurious Planar Symmetries and Symmetric Loss

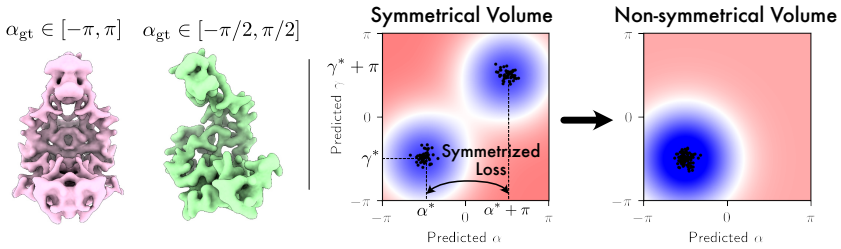

Fig. S4. (Left) Reconstructions obtained from the noisy adenylate kinase dataset ($L = 128$) obtained with cryoAI and the L2 loss, depending on the range of simulated in-plane angles. The model gets stuck in a spuriously symmetrical state when poses are generated on all $SO(3)$ but does not when in-plane rotations are restricted to $[-\pi/2, \pi/2]$. (Right) Heatmap of the L2 loss (per image) and intuition on the role of the symmetric loss. With a symmetrical volume ($V = \tilde{V}$), the energy landscape is periodic and shows an energy barrier between two minima. Each black point qualitatively represents the predicted Euler angles for one image. Without symmetric loss, the model is stuck when the predicted angles are evenly distributed between the two minima. The symmetric loss helps the model to overcome the barrier by creating a "tunnel" in the energy landscape.

When using our framework with a simple L2 loss

$$\mathcal{L} = \sum_{i \in \mathcal{B}} \|\hat{Y}_i - \hat{X}_i\|^2, \tag{7}$$

we observed that the model got stuck in local minima where the volume was showing spurious symmetry planes (see Fig. 5). In order to verify that this behavior was linked to the ambiguity described in the previous section, we generated a simulated dataset with a $\alpha_i \in [-\pi/2, \pi/2]$ instead of $[-\pi, \pi]$. Doing so, the encoder only has to predict α_i in $[-\pi/2, \pi/2]$ (up to a rotation that does not depend on i) to get a correct reconstruction of the volume and will therefore be less likely to mis-classify the orientation $(\alpha_i, \beta_i, \gamma_i)$ as $(\alpha_i + \pi, \beta_i, \gamma_i + \pi)$. As shown in Fig. S4, our framework is able to accurately reconstruct V with a L2 loss in that case.

We devised the symmetric loss (Eq. (10)) using the observation that the model was able to quickly converge when supervised on images Y_i for which $\alpha_i \in [-\pi/2, \pi/2]$. As described in Fig. S5, each image is firstly duplicated and one

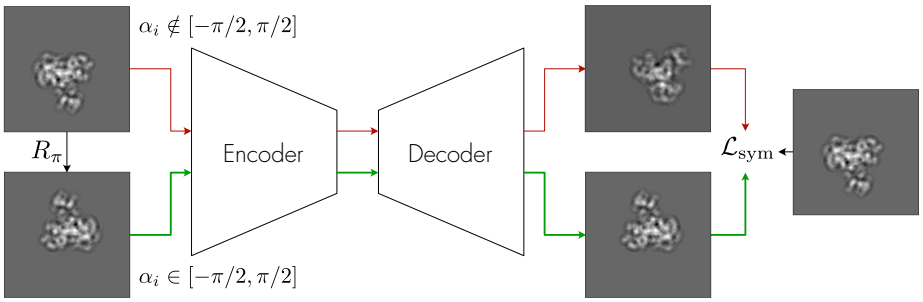

Fig. S5. For each input image, cryoAI produces two copies and rotates one them by π rad. Two poses are predicted by the decoder and two images are reconstructed by the decoder. The symmetric loss only penalizes (in Fourier space) the lowest distance between the reconstructed images and the input image. During the backward pass, gradients are only backpropagated through the pass that gave the most accurate reconstruction (the green path here). Images are shown in real space for clarity.

of the duplicates is rotated by π (rotation layer in Fig. S1). We know that one of the two duplicates has an associated in-plane angle α_i belonging to $[-\pi/2, \pi/2]$ (for some Euler basis). Each duplicate then goes through the whole pipeline and the two predicted images are compared with a L2 loss. The symmetric loss finds the minimum between the two L2 distances. This operation has the effect of disconnecting the loss from the worse predicted image in the computational graph. Therefore, at the backward pass, the gradients will only flow through one the of two paths, actually enabling the model to be supervised on images Y_i for which $\alpha_i \in [-\pi/2, \pi/2]$, without any loss of information. Keeping track of which path was selected by the loss, we know which latent vector (rotation and translation) we should consider at the output of the encoder, when estimating the poses.

Fig. S4 (right) gives an intuition on the role of the symmetric loss. For a symmetrical volume ($V = \tilde{V}$), the L2 loss is periodic for each image ($\mathcal{L}(\alpha, \gamma) = \mathcal{L}(\alpha + \pi, \gamma + \pi)$, see D.1). Therefore, there exists two identical global minima in the energy landscape. The model is stuck when predicted angles are distributed evenly between the two minima. The symmetric loss creates a "tunnel" in the energy landscape, helping the model to overcome the energy barrier and distribute the predicted angles in a way that is consistent with a non-symmetrical volume.

D.3 Symmetric Loss on PoseVAE

In [18], Zhong *et. al.* proposed to use a variational auto-encoder to predict the pose and named their technique "PoseVAE". Visual reconstruction show that their reconstructed volume gets stuck in symmetric states. We reproduced their method and added the symmetric loss thereby enabling PoseVAE to work on the *hand pointer* dataset (Fig. S6).

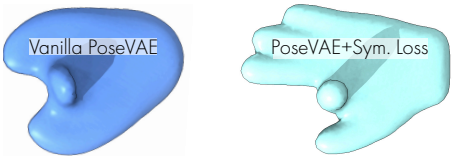

Fig. S6. Comparison between L2 loss (vanilla) and the symmetric loss with the PoseVAE method on the *hand pointer* dataset, both from [18].

E Datasets

Table S2 summarizes the parameters of the simulated and experimental datasets we used. Fig. S7 (resp. Fig. S8) shows in more details the statistics of the poses, the defoci and the shifts in the dataset for the simulated spliceosome (resp. experimental 80S).

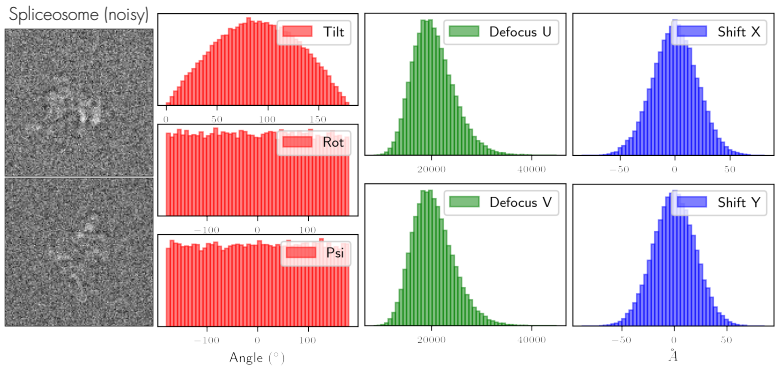

Fig. S7. Statistics and samples from the simulated noisy spliceosome dataset.

F Additional Results

F.1 Full Evaluation of Poses

We compared in Fig. 3 the runtime of cryoAI and cryoSPARC to reach a resolution of 10 Å on the reconstructed volume with the simulated noisy 80S dataset. In that experiment, cryoAI may reach convergence before processing all images in the dataset. For the sake of completeness, we run a second experiment, whose results are shown in Fig. S9. Once cryoAI has reached the threshold resolution of 10 Å, we use our model in evaluation mode (the computational graph is not maintained anymore, which decreases the memory cost of a forward pass and

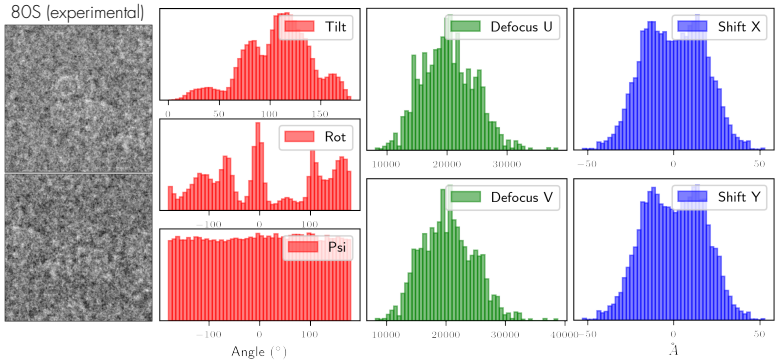

Fig. S8. Statistics and samples from the experimental 80S dataset.

Table S2. Parameters for our datasets. L is the image size, N is the number of images in the dataset. We give samples of the datasets in Fig. S7 and S8 to show the level of noise.

	Dataset	L	N	Å/pix.	Shift?	SNR (dB)
Simulated	80S noisy	128	10k-9M	3.77	N	0
	Spike ideal	128	50,000	3.00	N	∞
	Spike noisy	128	50,000	3.00	N	-10
	Spliceosome ideal	128	50,000	4.25	Y	∞
	Spliceosome noisy	128	50,000	4.25	Y	-10
	Kinase ideal	64	10,000	3.20	N	∞
	Kinase noisy	128	10,000	3.20	N	-5
Experimental	80S EMPIAR-10028	256	105,247	1.89	Y	NA

enables us to increase the batch size) to predict the poses associated with all images in the dataset. We report the time required to evaluate the whole dataset with cryoAI and with cryoSPARC. With cryoAI, we show the time required for CPU-to-GPU and GPU-to-CPU data transfers since this time could potentially be compressed with smart data handling. When a solid-state drive (SSD) is available, cryoSPARC can significantly decrease the time of *ab initio* reconstruction with particle caching. We can deduce from Fig. S8, that the runtime per image is 1.1 ms (2:35h / 9M) for cryoAI vs. 4.7 ms (11:36h / 9M) for cryoSPARC ab initio without SSD, which goes to show the computational benefits of using an encoder to estimate poses.

F.2 Experimental 80S

Additional Results. Table S3 and Fig. S10 show quantitative and qualitative results obtained with the experimental dataset of the 80S. In the absence of a ground truth volume, we perform a volume reconstruction with cryoAI using published poses.

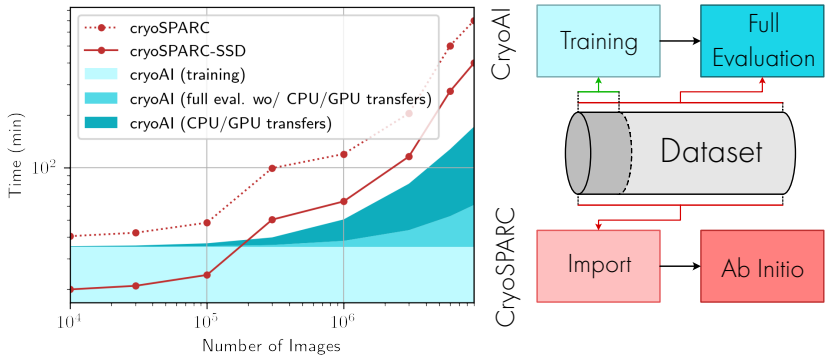

Fig. S9. (Left) Time required to estimate all the poses in the simulated noisy 80S dataset, with cryoAI and cryoSPARC. CryoAI switches to evaluation mode once a resolution of 10 Å is achieved on the volume. We indicate the time spent transferring data between the CPUs and the GPU during evaluation with cryoAI. CryoSPARC-SSD speeds up computations by caching particles on local SSDs. (Right) Comparison of workflows. CryoAI converges before seeing the whole dataset and can process rapidly the remaining images in evaluation mode. CryoSPARC imports (and formats) the dataset before processing it.

Table S3. Accuracy of pose and volume estimation for experimental 80S data. Resolution (Res.) is reported using the FSC = 0.143 criterion, in Å (↓). Rotation (Rot.) error is the median square Frobenius norm between predicted and published poses matrices R_i (↓). Translation (Trans.) error is the mean square L2-norm, in Å (↓). "cryoAI + cryoSPARC" refers to cryoAI (*ab initio*) + cryoSPARC (refinement).

80S (exp.)	cryoSPARC	cryoDRGN2	cryoAI	cryoAI + cryoSPARC
Res. Å	**7.54**	**7.54**	7.91	**7.54**
Rot.	**0.0001**	0.0008	0.004	**0.0001**
Trans.	**0.0008**	0.002	0.005	**0.0008**

Input resolution. We observed that cryoAI did not properly converge when fed with input images of size $L = 128$. Increasing the input size to 256 provides more information to the encoder for pose estimation but also implies making 256^2 queries per image to the FourierNet. For the computation to fit on a 40 Gb GPU, we need to decrease the batch size to 8, which makes the gradient-descent too stochastic and prevents the model from converging. Our solution was to use an image size of 256 on the encoder size and keep an image size of 128 on the decoder size. Once convergence is reached, voxel grids of sizes 256^3 or 128^3 can be queried in the FourierNet. However, in order to stay consistent with the image size used during training, we chose to output volumes of sizes 128^3. For the comparison with cryoSPARC to be fair, we use images of sizes 256 and downsample the volumes reconstructed by cryoSPARC from 256^3 to 128^3. CryoDRGN2 only reports the results obtained with $L = 128$.

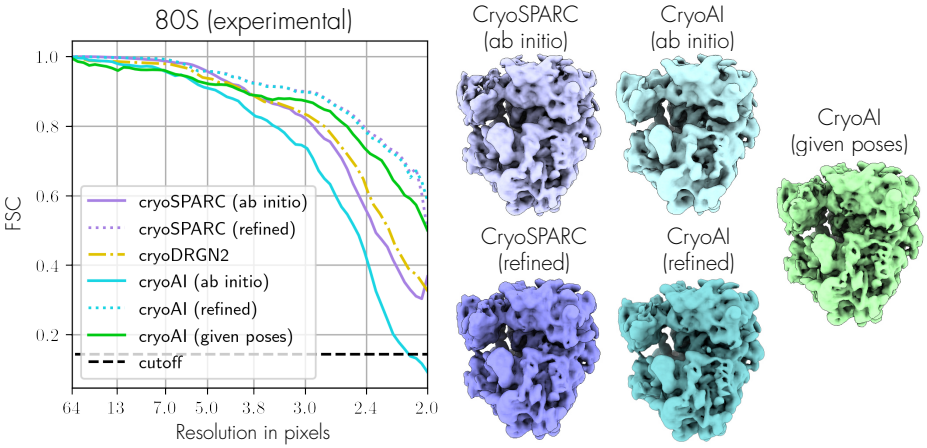

Fig. S10. (Left) FSC reconstruction-to-reconstruction on the experimental dataset of 80S. (Right) Reconstructed volumes visualized with UCSF ChimeraX [6].

F.3 Experimental Precatalytic Spliceosome

We downloaded images of the precatalytic spliceosome from EMPIAR-10180 [7], and downsampled to $D = 128$ (4.25 Å/pix). We performed homogeneous reconstruction with cryoAI on the filtered set of 139,722 images available at [17]. Particle images are shifted by their published poses, since the particles in this dataset are significantly out of center [18]. Results are shown in Fig. S11. We report the half-to-half FSC. We note the presence of a blurry zone in the reconstructed volume, which correlates with the zone where the molecule can fold.

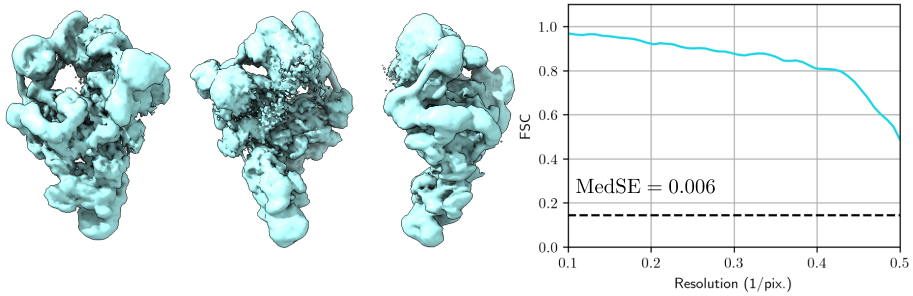

Fig. S11. Qualitative reconstruction on the EMPIAR-10180 dataset [17].

F.4 Simulated Datasets

We show in Fig. S12-S13 quantitative and qualitative results obtained with our simulated datasets with cryoAI and cryoSPARC. We quantitatively study the impact on the noise level on a small (10k images) synthetic dataset of the 80S ribosome ($L = 128$) in Table S4. We can see that the convergence time increases with the noise while the final resolution decreases slightly.

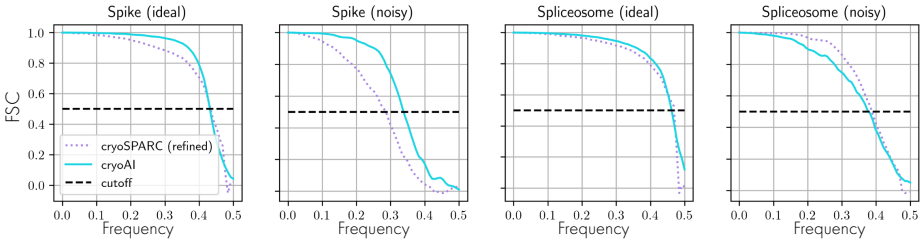

Fig. S12. Fourier Shell Correlations reconstruction-to-ground-truth on simulated datasets, with a cutoff at FSC = 0.5.

Table S4. Impact of the noise on the runtime and the resolution with the synthetic 80S dataset.

SNR	0dB	−5dB	−10dB	−15dB
Runtime	0:20h	0:29h	0:45h	2:07h
Res. (pix)	2.15	2.44	2.79	3.21

Molecular graphics and analyses performed with UCSF ChimeraX [6], developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

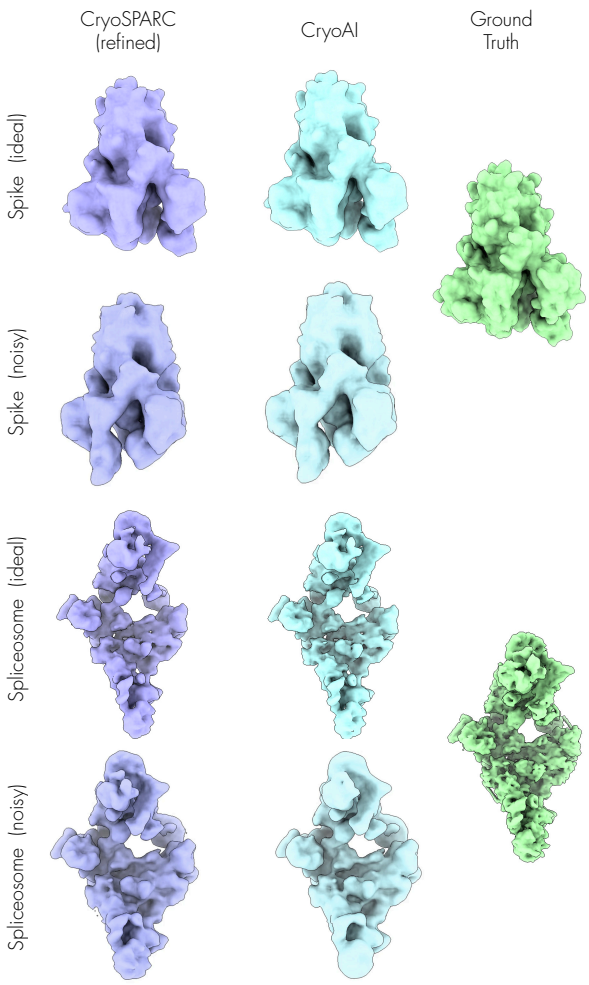

Fig. S13. Qualitative reconstructions on simulated datasets.

References

1. A, C., R, H., Mastronarde D, e.a.: Mrc2014: Extensions to the mrc format header for electron cryo-microscopy and tomography. J Struct Biol. **192**, 146–150 (2015)
2. AC, W., YJ, P., MA, T., A, W., AT, M., D, V.: Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. Cell (2020)
3. Berman, H., Westbrook, J., Feng, Z., Gilliland, G., T.N. Bhat, H.W., Shindyalov, I., Bourne, P.: The protein data bank. Nucleic Acids Research **28**, 235–242 (2000)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
5. Nashed, Y.S.G., Poitevin, F., Gupta, H., Woollard, G., Kagan, M., Yoon, C.H., Ratner, D.: CryoPoseNet: End-to-End Simultaneous Learning of Single-Particle

Orientation and 3D Map Reconstruction From Cryo-Electron Microscopy Data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 4066–4076 (2021)
6. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E.: Ucsf chimerax: Structure visualization for researchers, educators, and developers. Protein Science **30**, 70–82 (2021)
7. Plaschka, C., Lin, P., Nagai, K.: Structure of a pre-catalytic spliceosome. Nature (2017)
8. Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A.: cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods **14**, 290–296 (2017)
9. Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W.Y., Johnson, J., Gkioxari, G.: Accelerating 3D Deep Learning with PyTorch3D. arXiv: 2007.08501 (2020)
10. Ren, P., Chun, J., Thomas, D.G., Schnieders, M.J., Marucho, M., Zhang, J., Baker, N.A.: Biomolecular electrostatics and solvation: a computational perspective. Quarterly Reviews of Biophysics **45**(4), 427–491 (2012). https://doi.org/10.1017/S003358351200011X
11. Rosenthal, P.B., Henderson, R.: Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. Journal of molecular biology **333**, 721–745 (2003)
12. Sampson, G., Tuy, H.: Fourier transforms and their Lipschitz classes. Pacific Journal of Mathematics **75**, 519–537 (1978)
13. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in Neural Information Processing Systems **33**, 7462–7473 (2020)
14. Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., Ludtke, S.J.: EMAN2: an extensible image processing suite for electron microscopy. Journal of Structural Biology **157**, 38–46 (2007)
15. Wojdyr, M.: Gemmi - library for structural biology [software]. `https://github.com/project-gemmi/gemmi` (2021)
16. Wong, W., Bai, X.c., Brown, A., Fernandez, I.S., Hanssen, E., Condron, M., Tan, Y.H., Baum, J., Scheres, S.H.: Cryo-em structure of the *Plasmodium falciparum* 80s ribosome bound to the anti-protozoan drug emetine. eLife **3**, e03080 (jun 2014). https://doi.org/10.7554/eLife.03080, `https://doi.org/10.7554/eLife.03080`
17. Zhong, E.D., Bepler, T., Berger, B., Davis, J.H.: Data for "CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks" (Jan 2021). https://doi.org/10.5281/zenodo.4355284, `https://doi.org/10.5281/zenodo.4355284`
18. Zhong, E.D., Lerer, A., Davis, J.H., Berger, B.: Cryodrgn2: Ab initio neural reconstruction of 3d protein structures from real cryo-em images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4066–4075 (2021)
19. Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the Continuity of Rotation Representations in Neural Networks. arXiv: 1812.07035 (2020)