CryoFIFA: deep curriculum learning for cryo-EM fine and fast ab-initio reconstruction

Fang Kong[1,2,*], Chenlong Wang[2], Le song[2,3,*], Chuangye Yan[1,4,*]

[1]Beijing Frontier Research Center for Biological Structure, State Key Laboratory of Membrane Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

[2] BioMap Research, California, USA

[3]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE

[4]Lead contact.

[*]Correspondence: yancy2019@tsinghua.edu.cn (C.Y.), songle@biomap.com (L.S.), and kongf21@mails.tsinghua.edu.cn (F.K.)

Abstract

Cryo-electron microscopy (Cryo-EM) single particle analysis (SPA) is currently the primary method for determining structures of macromolecular complexes ranging from tens of kilodaltons (kDa) to several megadaltons (MDa). Traditional cryo-EM three-dimensional (3D) reconstruction and refinement methods are based on maximum likelihood approaches. Recently, deep learning-based methods began to emerge in the field of cryo-EM reconstruction. Here, we propose cryoFIFA, a deep curriculum learning-based approach for cryo-EM fine and fast 3D reconstruction. CryoFIFA enables the ab-initio reconstruction of macromolecules ranging from ~150 kDa to over 2 MDa, and significantly enhances reconstructing speed through supporting multi-GPU processing. It also demonstrates superior performance in the reconstruction of near-atomic resolution map for G-protein coupling receptors (GPCR) complex. CryoFIFA provides a framework for exploring new paradigms in deep learning-driven cryo-EM reconstruction, highlighting the potential for future applications of deep learning in cryo-EM analysis.

Main

Recent advancements in cryo-electron microscopy (cryo-EM) have enabled scientists to resolve macromolecules structures at near-atomic resolution[1-3]. In this technique, macromolecules are vitrified and imaged using a high-voltage electron microscope. The resulting micrographs contain particle images with unknown projection orientations. A crucial step for structure determination is the reconstruction of 3D density maps from the collected two-dimensional particle images, which is particularly challenging because both the particle poses and the density maps are initially unknown. Compounding this difficulty is the fact that these particle projections are often obscured by heavy noises.

Currently, the primary software packages for cryo-EM reconstruction include RELION[4] and cryoSPARC[5], along with others such as EMAN2[6,7]. RELION relies heavily on a pose-search-based framework for 3D classification and structure refinement, whereas cryoSPARC utilizes stochastic gradient descent (SGD) for ab-initio structure determination and employs a branch-and-bound algorithm for rapid refinement.

Deep learning has rapidly advanced in addressing challenges in computer vision and natural language processing, and it shows significant potential for tackling mathematical or physics-based problems as well. Several deep learning-based methods have been developed for the cryo-EM field. For instance, Topaz[8,9] has been used for

denoising and particle picking, DeepEMhancer[10] for map sharpening, Blush regularization[11] for map denoising, and ModelAngelo[12] for automated model building.

However, one of the key challenges in cryo-EM—ab-initio reconstruction, which involves reconstructing 2D particle images into 3D density maps—remains difficult to address using deep learning approaches. Some studies, such as cryoAI[13] and cryoDRGN2[14], have demonstrated the effectiveness of deep learning-based approaches for solving ab-initio reconstruction problems. CryoAI is the first end-to-end framework that takes particle images as input and outputs density maps. However, it does not show extensive results on experimental datasets, leaving the question open as to whether deep learning is sufficiently robust to serve as a reliable method for cryo-EM reconstruction.

CryoDRGN2 implements two tasks: ab-initio reconstruction and structural dynamics analysis. The ab-initio reconstruction task involves traditional pose search, which is computationally intensive. According to its official documentation, cryoDRGN2 requires 20 hours to process 218,000 particles with a box size of 128 pixels on a single V100 GPU. Subsequent work, such as cryoFIRE[15], aims to resolve structure dynamics from ab-initio heterogeneous reconstruction but employs a backbone similar to cryoAI and cryoDRGN2, which may also lack sufficient validation by experimental datasets. It is also important to note that no deep learning-based approaches have addressed the reconstruction problem for small membrane proteins, leaving various superfamilies, including transporters and G-protein coupling receptors (GPCRs), unaddressed.

To further explore the potential of applying deep learning for cryo-EM reconstruction, innovations from both cryo-EM reconstruction and deep learning should be integrated. Here, we propose an advanced, deep curriculum learning-based ab-initio reconstruction method for cryo-EM Fine and Fast reconstruction, named cryoFIFA (Fig.1). CryoFIFA adopts an autoencoder network trained in a curriculum learning-based manner to enhance reconstruction quality. We have evaluated cryoFIFA using various experimental datasets, demonstrating its robust capabilities for ab-initio reconstruction. By using other downstream deep learning-based sharpening tools, we have achieved near-atomic resolution reconstruction of a GPCR complex through a fully deep learning-driven approach. This work can represent an important step toward fully deep learning-based cryo-EM structure determination.

Results

CryoFIFA adopts a curriculum learning-based strategy for ab-initio reconstruction

Although several studies have explored image-encoder and volume-decoder frameworks for ab-initio reconstruction, these approaches have shown only limited results on experimentally collected cryo-EM data. In this work, we proposed a curriculum learning procedure to optimize the network training, complemented by several reconstruction techniques, to enhance the capabilities of deep learning for cryo-EM ab-initio reconstruction (Fig. 1).

In the Encoder, we employed a ResNet network for feature extraction, a two-layer SIREN network for further feature processing, and two parallel SIREN networks for predicting rotation and translation (Fig. 1b, *left panel*; 1c). In the Decoder, we utilized the same FourierNet as described in cryoAI[13], where two separate SIREN networks were implemented for predicting the values of volume density (Fig. 1b, *right panel*).

Our curriculum learning strategy focuses on two key components: (1) gradually increasing the diameters of masks applied in Fourier space (Fig. 1d), and (2) implementing a loss-function-based approach to filter out junk particles in every minibatch (Fig. 1e). The first strategy, similar to the Frequency Marching[16] method used in cryoDRGN2, simplifies training by initially passing only low-frequency signals. By applying frequency masks to the input coordinate plane before feeding them into the Decoder, we can also reduce the memory requirements during early epochs of training. This procedure works particularly well in conjunction with the second strategy. In cryo-EM, low-frequency signals primarily contribute to the overall particle shape and intensity. Particles with lower loss function scores in the early epochs are more likely to have strong low-frequency signals, making it easier for convolution-based feature extractors to derive meaningful features.

Additionally, we employed two specialized reconstruction techniques for challenging cases, such as small membrane proteins. The first technique involved applying a shifted

mask based on the predicted translation to modify the input particles, which was particularly useful when multiple particles were present or the particle was not well-centered. The second technique introduced gradually reduced disturbances to the predicted pose during early training epochs, helping the network achieve better results by escaping the negative influences caused by poorly initialized parameters or inaccurately predicted poses.

For testing the performance of cryoFIFA, we present reconstruction results on both simulated and experimental datasets. The simulated datasets, which consist of two challenging tasks involving small proteins (<100 kD), help validate the capability of cryoFIFA. The experimental datasets, comprising complexes with various molecular weights and symmetries, demonstrates that CryoFIFA exhibits broad reconstruction capability. Finally, we tested CryoFIFA on a GPCR-G protein complex (~150 kD) and achieved near-atomic resolution result.

CryoFIFA ab-initio reconstruction on simulated datasets

To test the ability of cryoFIFA to reconstruct high-resolution details, we initially evaluated our method on two simulated datasets generated from a small soluble protein, streptavidin[17], and a small membrane protein, GABA transporter 1 (GAT1)[18] (Fig. 2). For both proteins, we used cryoSPARC to simulate 100,000 particles with an SNR of 0.1, which closely resembles experimental cryo-EM particle images. The defocus values were randomly generated, ranging from 1. 2 to 3.0 μm.

For streptavidin, which has been widely utilized for cryo-EM reconstruction validation. We used cryo-EM density map (EMD-20907) to generate particles and the corresponding atomic model (PDB:6J6K) for validation. The simulated particles from the published 2.6 Å streptavidin map have a pixel size of 1.072 Å and a box size of 128 pixels (Fig. 2a). For GAT1, we used the cryo-EM density map (EMD-33671) to generate particles and the corresponding atomic model (PDB:7Y7V) for validation. The reported 2.2 Å GAT1 map represents a high quality for such a small membrane protein. The original pixel size of the published map is 1.0825 Å, and the box size is 200 pixels. We downsampled the simulated particles to a box size of 128, resulting in a corresponding pixel size of approximately 1.691 Å (Fig. 2e).

We showed that CryoFIFA successfully reconstructed streptavidin at ~2.9 Å and GAT1 at ~3.4 Å based on the model-to-map FSC (FSC=0.5). The published atomic models also fit well with the reconstructed maps (Fig. 2b-d, 2f-h), allowing for the identification of large residues. This high-resolution reconstruction of such challenging cases on simulated datasets highlights CryoFIFA's potential for cryo-EM ab-initio reconstruction.

CryoFIFA ab-initio reconstruction on experimental datasets

We further evaluated our method on different types of experimental datasets, including the ribosome[19], the U4/U6.U5 triple small nuclear ribonucleoprotein (tri-snRNP) [20], the

recombination activating gene 1 (RAG1)-RAG2 complex[21], the transient receptor potential (TRP) channels TRPV1[22] and TRPM8[23], apoferritin, the mastigoneme[24], the GPCR complex[25], and the B^{act} spliceosome[26]. Most of these data can be downloaded from EMPIAR[27], and some are from other research groups (Extended Data Table 1). All using particles mentioned in this section were cropped into a box size of 128 pixels, resulting in pixel sizes varying from ~2 to 4 Å.

Previously, experimental datasets tested in cryoAI only included large complexes: ribosome (3.2 MDa) and tri-snRNP (2 MDa). In cryoDRGN2, the tested datasets included one more: the RAG complex (300kDa). We first tested cryoFIFA on these three datasets, and the results showed that cryoFIFA performed well in reconstructing these cases (Fig. 3a, *left panel*). It is also important to notice that no membrane proteins were tested in previous deep learning-based reconstruction methods. Therefore, we tested two more challenging cases: TRPV1 (380 kDa) and TRPM8 (513 kDa). The reconstructions were equally good as those from cryoSPARC (Fig. 3a, *right panel*).

We also included apoferritin (480 kDa) with octahedral symmetry and a fibril glycan-protein mastigoneme. Although cryoFIFA currently only supports C1 symmetry, it was able to reconstruct those higher ordered symmetry particles as well (Fig. 3b). We also tested two more experimental proteins complexes, GPCR-G protein complex and the B^{act} spliceosome, using cryoSPARC, cryoAI, cryoDRGN2, and cryoFIFA (Fig. 3c). For the GPCR-G protein complex, we chose β_2-adrenergic receptor-G_s protein complex

(~150 kDa)[25] without the fusion of T4 lysozyme. These results indicate that CryoFIFA succeeded across all types of datasets, closely matching the results of cryoSPARC (Fig. 3a-c).

To enhance reconstruction speed, cryoFIFA supports multiple GPUs by using Distributed Data Parallel in Pytorch[28]. We compared the running time among cryoSPARC, cryoAI, cryoDRGN2, and cryoFIFA on the activated spliceosome (B^{act} complex) and the GPCR-G protein complex datasets (Extended Data Table 1). For B^{act} complex, which contains only 20,000 particles, cryoFIFA is the fastest, achieving a result where most features can be clearly identified. For the GPCR, which contains approximately 100,000 particles, the running time of cryoSPARC is shorter than that of cryoFIFA, but cryoFIFA shows better ab-initio reconstruction than cryoSPARC and other deep learning-based methods.

Near-atomic resolution reconstruction of the GPCR-G protein complex

Several deep learning-based methods have been developed for cryo-EM map sharpening. We aimed to validate whether the unsharpened output volume from cryoFIFA could serve as input for these methods and whether the quality of the map generated by cryoFIFA could be further improved by sharpening. We performed cryoFIFA reconstruction on the GPCR-G protein complex dataset using the unbinned particles with a pixel size of 1.0825 Å. The output volume was then fed into DeepEMhancer[10] for sharpening, and the results showed that it indeed exhibited high-

resolution quality (Fig. 4a,b). The sharpened map aligns well with the previously published model (PDB: 3SN6) (Fig. 4c). The density shows clear features of α-helix and allows for the identification of large interacting residues (Fig. 4d).

A local refinement module was integrated into the training network for obtaining the high-resolution map (Fig. 3c, 4a). We employed a double-encoder framework for high-resolution reconstruction. The first encoder takes temporally downsampled particles (typically using a box size of 128) as input. The second encoder takes original particles as input and predicts the rotation and translation deviations to further optimize the predicted values from the first encoder, making the second encoder serve as a local refinement module. CryoFIFA was initially trained using only the first encoder during the early epochs. Once the current reconstruction diameter reaches the downsampled box size, the parameters of the first encoder are fixed, and the local refinement module is integrated into the training process. The Decoder parameters are also frozen for several fade-in epochs to avoid disruption from the poor initialization of the local refinement module. By integrating additional modules, cryoFIFA shows how a fully deep learning-based approach worked for high resolution reconstruction.

Discussion

In this paper we have developed a novel deep learning method trained in a curriculum-based manner for cryo-EM ab-initio reconstruction, showcasing its potential across a range of experimental datasets.

For the two small proteins, streptavidin and GAT1, cryoFIFA currently works only on simulated datasets, while for the two experimental datasets, it still cannot produce a reasonable map. An image denoising module may be required to improve the signal-to-noise ratio of small particles. Additionally, using a tight mask instead of a circular mask may also help improve the performance of cryoFIFA in such cases.

It is also worthy trying methods like Blush regularization[11] when updating reconstructions. Blush regularization has already shown advantages in map denoising for improving map resolution, and incorporating such a process may further enhance the performance of deep learning-based methods. This could be also the key step to making cryoFIFA effective for small proteins.

We also noticed that the periphery or flexible regions in large complexes can be well captured at low resolution by cryoFIFA. In the cases of the ribosome, tri-snRNP, mastigoneme, and B^{act} spliceosome, the features of these regions are finer than those produced by traditional methods (Fig. 3). This may aid in the global visualization and interpretation of large cryo-EM maps and advance local refinements on these flexible regions by providing more rigid and condensed local volumes.

The current version of cryoFIFA is focused on ab initio volume generation. Looking ahead, our future direction could expand to two key areas: 3D classification and high-

resolution local refinement. We can modify the current network to enable cryoFIFA to support multiple reference generation for 3D classification. A dynamic mask strategy, similar to that used in cryoSPARC, may improve local refinement in cryoFIFA. Integrating other deep learning-based methods, such as cryoSieve[29] for particle sorting, could also benefit the training process. By advancing in these directions, we hope to enhance the efficiency and effectiveness of cryoFIFA, facilitating the study of deep learning approaches used in the cryo-EM field.

Acknowledgements

This work was funded by the National Natural Science Foundation of China (32341016, 32171204, C.Y), the National Key R&D Program of China (2020YFA0509301, C.Y.), Beijing Frontier Research Center for Biological Structure, State Key Laboratory of Membrane Biology, Tsinghua University Initiative Scientific Research Program (20221080032, 20231080037, C.Y.), and Start-up funds from Tsinghua-Peking Center for Life Sciences and Tsinghua University.

Author Contributions: C.Y., L.S. and F.K conceived the project. F.K. designed most of the experiments. All other authors contributed to experiment design. F.K. wrote almost all the codes and performed all experiments. C.W. helped in some scripts analysing and environmental tests. All authors analysed the data. F.K. and C.Y. wrote the manuscript. C.Y., F.K. and L.S. contributed to the manuscript finalization. All authors read and approved the final manuscript.

Conflict of Interest Statement: The authors declare no competing interests.

Figure Legends

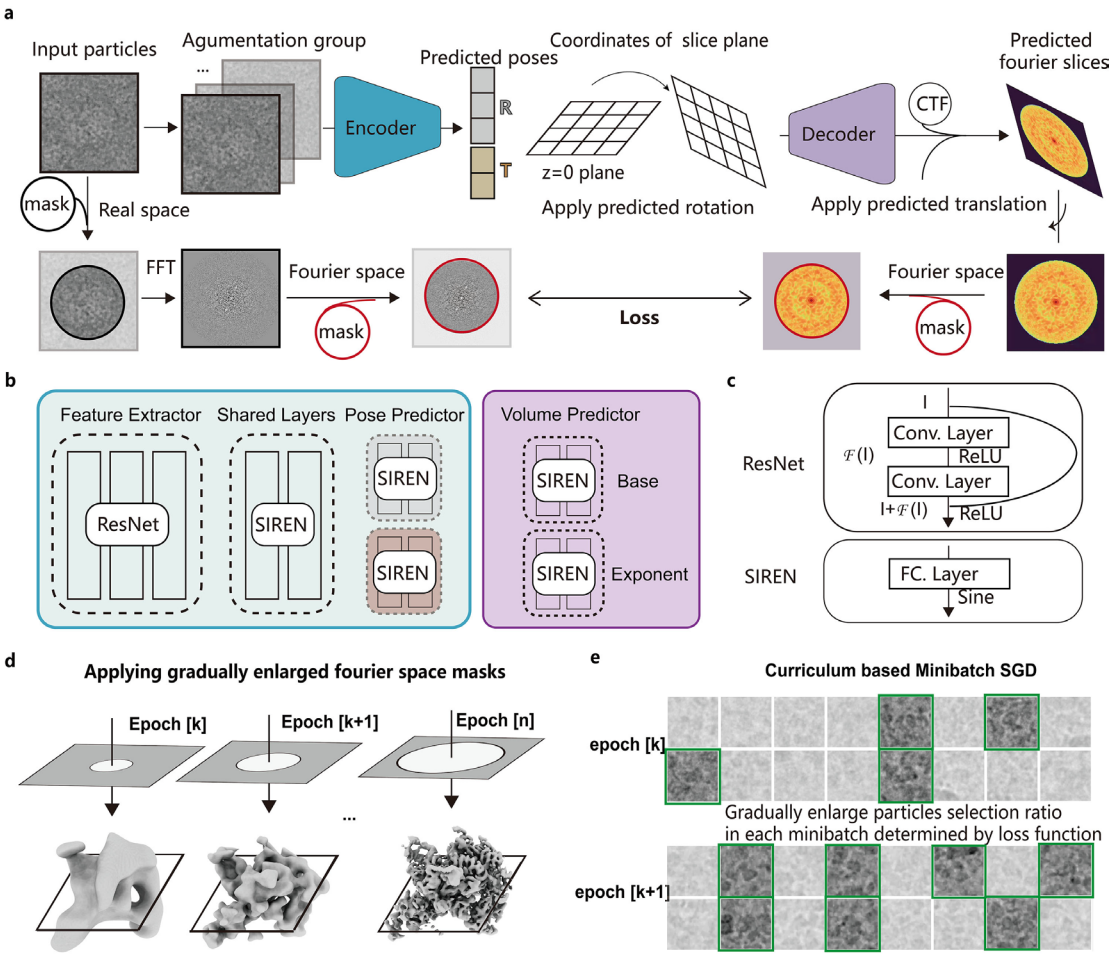

Fig. 1 | CryoFIFA workflow. **a**, Overall framework of CryoFIFA. The total framework is an Encoder-Decoder network. Particles are first preprocessed and augmented before fed into Encoder. Predicted rotation is applied on a series of coordinates representing the center slice of z=0 plane. Rotated coordinates within current frequency mask are fed into Decoder to predict the values of center slice in Fourier space. **b**, **c**, Detailed structure of cryoFIFA network. **d**, A series of gradually enlarged masks applied in Fourier space. **e**, Curriculum based minibatch SGD to filter out junk particles by loss function.

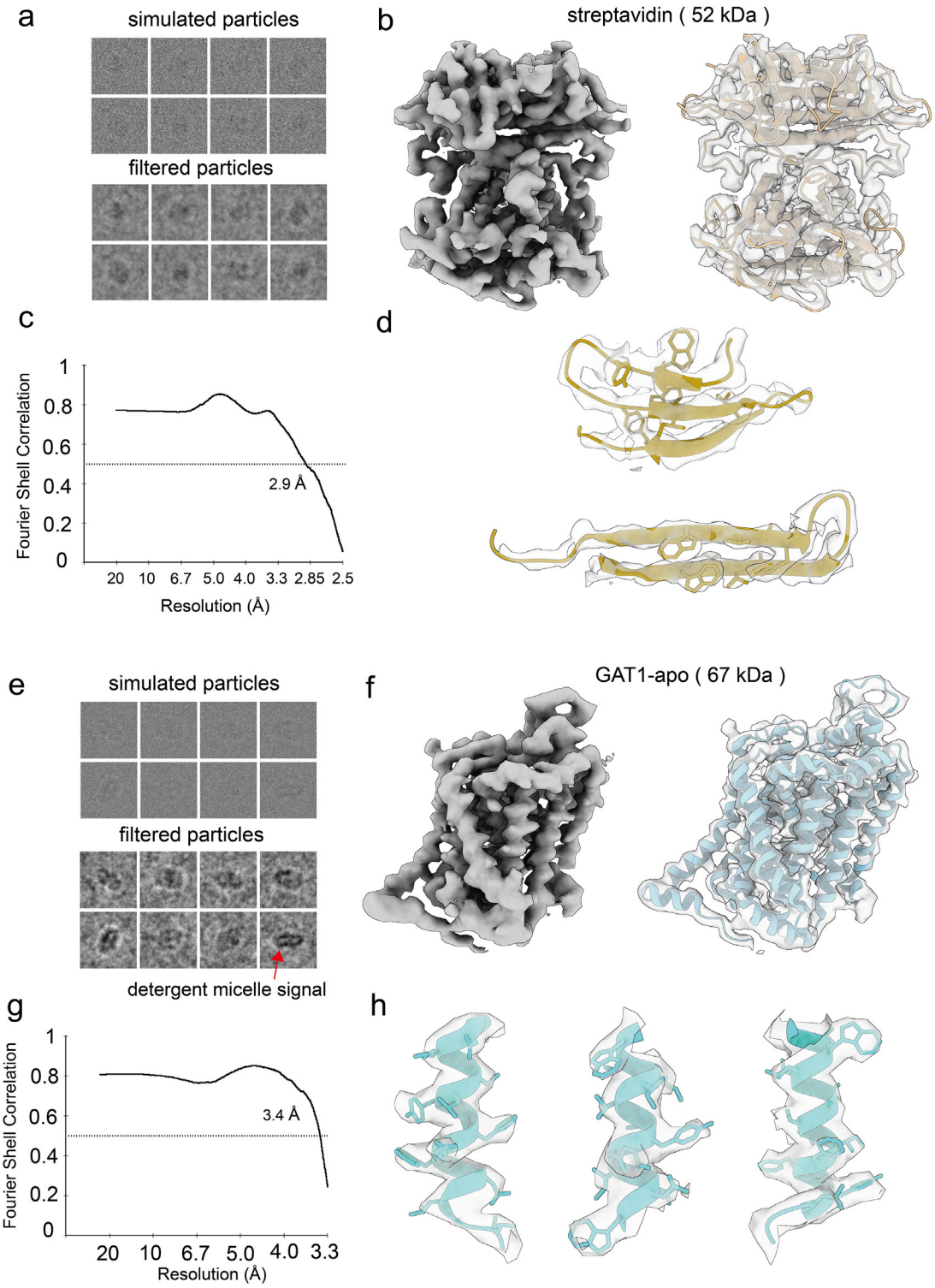

Fig. 2 | Reconstruction results on simulated datasets. **a**, Simulated particles and low-pass filtered particles in streptavidin dataset. **b**, Reconstructed map of streptavidin and model comparison. **c**, Model-to-map FSC plot for streptavidin dataset. **d**, Representative densities showing residue features. **e**, Simulated particles

and low-pass filtered particles in GAT1 dataset. **f**, Reconstructed map of GAT1 and model comparison. **g**, Model-to-map FSC plot for GAT1 dataset. **h**, Representative densities showing residue features.

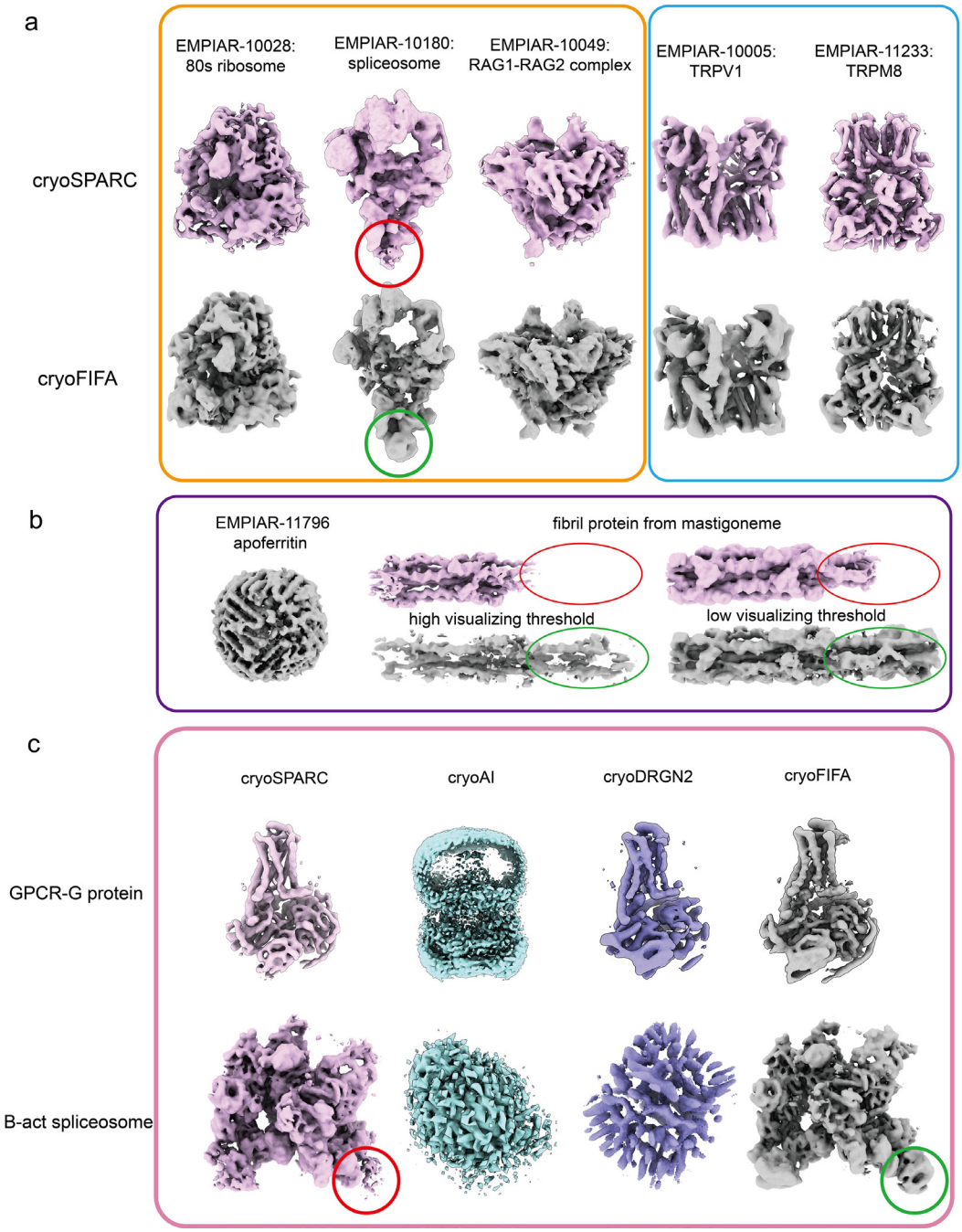

Fig. 3 | CryoFIFA reconstruction on various types of experimental datasets. **a**, Comparison of the reconstructions between cryoSPARC and cryoFIFA. Experimental datasets, including the ribosome, the tri-snRNP, the RAG1-RAG2 complex, the TRPV1, and the TRPM8, were tested. **b**, Reconstruction results on particles from proteins with special symmetry. Apoferritin and mastigoneme were tested. **c**, Comparison of the reconstructions among cryoSPARC, cryoAI, cryoDRGN2, and cryoFIFA. The GPCR-

G protein complex and the B^{act} spliceosome were tested. Low-resolution regions of periphery and flexible domains in the reconstructions are indicated by green or red circles.

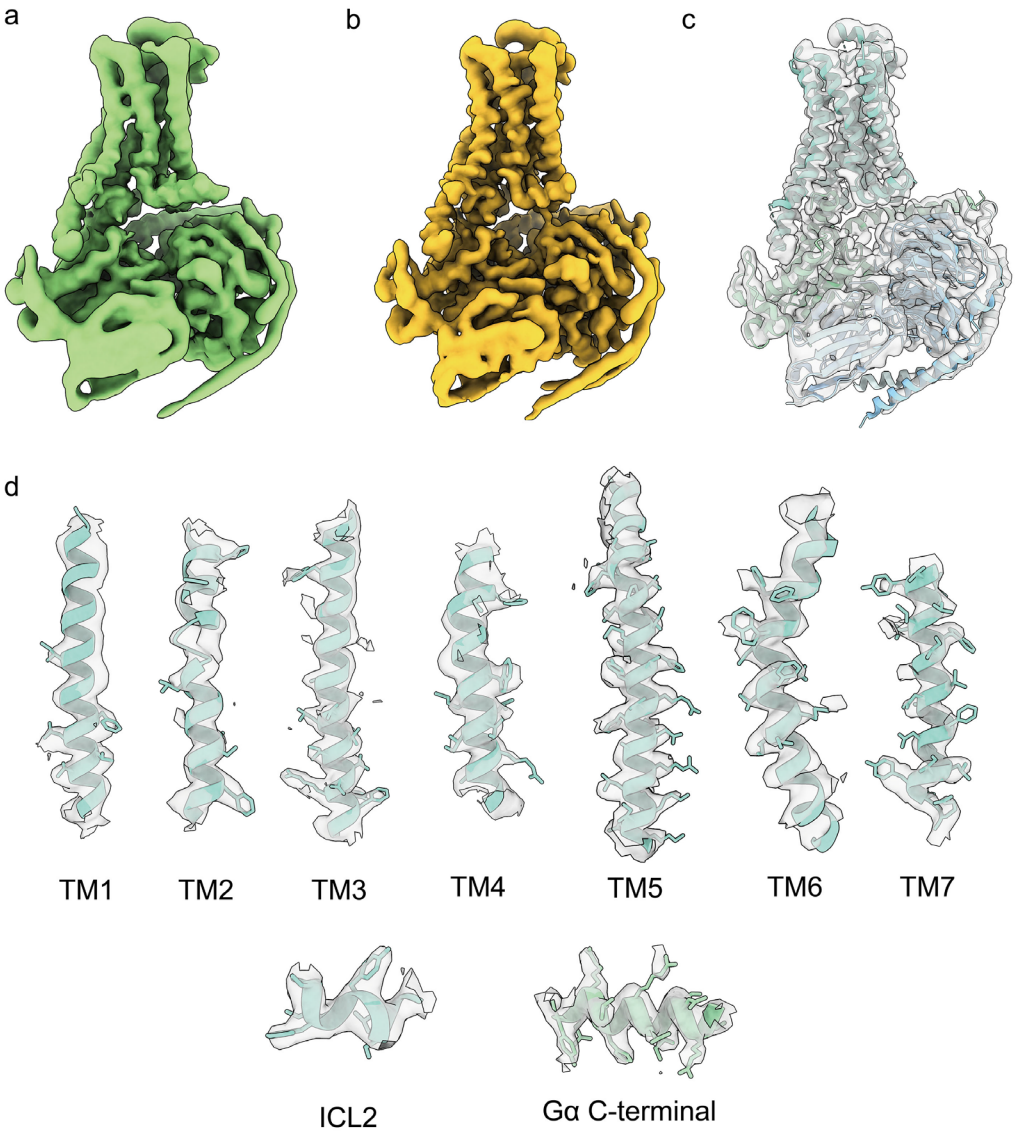

Fig. 4 | Near-atomic resolution reconstruction results on experimental β_2AR-G protein complex dataset. **a**, CryoFIFA output (unsharpened map). **b**, Sharpened map using DeepEMhancer. **c**, Model comparison of the sharpened map against deposited atomic coordinates. **d**, Local densities of sharpened map. Here shows local densities of transmembrane helices (TM) 1-TM7 and intracellular loop (ICL) 2 for the receptor.

Methods

CryoFIFA framework

CryoFIFA could successfully perform ab initio reconstruction of cryo-EM density maps by training a self-supervised neural network in a curriculum learning-based manner. The main architecture adopted is an image-Encoder-Decoder-volume network.

The input of Encoder is cryo-EM particles extracted from micrographs. Encoder network is a ResNet[30] for feature extraction followed by SIREN network for pose information prediction. The rotation is represented in the 6-dimensional space[31], as it has been shown to lead to the best results for the rotation prediction in many related works, such as cryoAI[13].

The input of Decoder starts from a Fourier space plane coordinates. The plane coordinates are rotated by applying the rotation parameters predicted by the Encoder. The Decoder is trained to learn the mapping between the given coordinates and their corresponding Fourier space values. A shifted mask is applied to both the input particles and the reconstructed shifted projections for loss calculation. A minibatch-based SGD optimizer named Adam is used for parameters update.

Once trained, the Encoder acts as a pose predictor, capable of inferring pose information for all particles in the training dataset. The Decoder reconstructs the Fourier space of the cryo-EM volume with continuous resolution, allowing density values to be fetched

at any coordinates without interpolation errors. This approach is known as implicit neural representation, which has been widely used in scene reconstruction[32] within the computer vision field. Our method extends the capabilities of previously proposed amortized inference for cryo-EM reconstruction.

The loss function is computed between the Fourier transform of the masked original images and the central slice in Fourier space of the predicted volume. CryoFIFA aims to minimize the loss function in terms of:

$$arg\min_{\theta} \ (M_F * (CTF * TE_\theta(X) * D_\theta\big(C * RE_\theta(X)\big) - FT(M_R * X) \ )^2)$$

M_F is the mask series applied in Fourier space and M_R is the mask applied to input particle in real space. Contrast Transfer Function (CTF) are known parameters that can be estimated using existing mathematic tools like CTFFIND4[33], Gctf[34] or patch CTF in cryoSPARC[5]. $TE_\theta(X)$ and $RE_\theta(X)$ provides translation and rotation prediction for each input particles X. C represents the coordinates of a center slice lied on the z =0 plane, D_θ denotes Fourier decoder that decode an input coordinate into a value represented in Fourier space.

Curriculum learning strategies enable fine and fast reconstruction

Curriculum learning is a technique from machine learning where a model is first asked to learn from easy data before transitioning to more difficult data. This is a model optimization strategy to tackle the non-convex optimization nature inherent for deep learning model learning, helping the model learn a better local minimum and thereby

achieve higher accuracy. Inspired by human learning, where children start with simpler materials and gradually progress to more complex subjects, curriculum learning aims to improve the training process in a similar manner[35].

CryoFIFA was trained in resolution gradient manner, achieved by employing a series of circular masks in Fourier space. The incrementally expanding radius of mask effectively implements low-pass filtering across various frequency bands of the images. Consequently, the decoder's input is restricted to coordinate points within the current mask. Additionally, a particle filtering process was implemented based on the loss function score ranking to select good particles in every minibatch. We refer to this strategy as Curriculum Learning-based Minibatch SGD.

For local refinement, we used a two-stage Encoder strategy. The input particles for the first-stage Encoder are processed by a downsampling module for ab-initio reconstruction, while the second Encoder uses full-sized particles for local refinement.

Data preprocessing

For better prediction of pose information during the training, we designed a two-step data preprocessing approach: low-pass filtering and rotation augmentation. We use Butterworth filtering for low-pass filtering the input particles. The output particles groups will have 2 channels, the original one and the low-pass filtered copy. We provide two strategies for rotation augmentation: degree-fixed rotation and random rotation.

The degree-fixed rotation strategy involves 8 x 45-degree or 4 x 90-degree rotations, each doubled by a flip operation. This strategy works well on most datasets, such as the ribosome and GPCR. The random rotation strategy involves completely random rotation augmentation (usually ~8x), also doubled by a flip operation. This strategy works well on specialized datasets, such as apoferritin.

Training cryoFIFA

CryoFIFA was trained on an 8 x Tesla V100 -32G GPU machine with 512G RAM. Dataset details and training parameters can be found in Extended Data Table 2.

References:

1 Kühlbrandt, W. The Resolution Revolution. *Science* **343**, 1443-1444 (2014). https://doi.org/doi:10.1126/science.1251652

2 Cheng, Y., Grigorieff, N., Penczek, Pawel A. & Walz, T. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* **161**, 438-449 (2015). https://doi.org/10.1016/j.cell.2015.03.050

3 Merk, A. *et al.* Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165**, 1698-1707 (2016). https://doi.org/10.1016/j.cell.2016.05.040

4 Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180**, 519-530 (2012). https://doi.org/https://doi.org/10.1016/j.jsb.2012.09.006

5 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). https://doi.org/10.1038/nmeth.4169

6 Tang, G. *et al.* EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology* **157**, 38-46 (2007). https://doi.org/https://doi.org/10.1016/j.jsb.2006.05.009

7 Bell, J. M., Chen, M., Baldwin, P. R. & Ludtke, S. J. High resolution single particle refinement in EMAN2.1. *Methods* **100**, 25-34 (2016). https://doi.org/https://doi.org/10.1016/j.ymeth.2016.02.018

8 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods* **16**, 1153-1160 (2019). https://doi.org/10.1038/s41592-019-0575-8

9 Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nature Communications* **11**, 5208 (2020). https://doi.org/10.1038/s41467-020-18952-1

10 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology* **4**, 874 (2021). https://doi.org/10.1038/s42003-021-02399-1

11 Kimanius, D. *et al.* Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nature Methods* **21**, 1216-1221 (2024). https://doi.org/10.1038/s41592-024-02304-8

12 Jamali, K. *et al.* Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450-457 (2024). https://doi.org/10.1038/s41586-024-07215-4

13 Levy, A. *et al.* in *Computer Vision – ECCV 2022.* (eds Shai Avidan *et al.*) 540-557 (Springer Nature Switzerland).

14 Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. in *2021 IEEE/CVF International Conference on Computer Vision (ICCV).* 4046-4055.

15 Levy, A., Wetzstein, G., Martel, J., Poitevin, F. & Zhong, E. D. Amortized Inference for Heterogeneous Reconstruction in Cryo-EM. *Adv Neural Inf Process Syst* **35**, 13038-13049 (2022).

16 Barnett, A., Greengard, L., Pataki, A. & Spivak, M. Rapid Solution of the Cryo-EM Reconstruction Problem by Frequency Marching. *SIAM Journal on Imaging Sciences* **10**, 1170-1195 (2017). https://doi.org/10.1137/16M1097171

17 Han, Y. *et al.* High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. *Proceedings of the National Academy of Sciences* **117**, 1009-1014 (2020). https://doi.org/10.1073/pnas.1919114117

18 Zhu, A. *et al.* Molecular basis for substrate recognition and transport of human GABA transporter GAT1. *Nature Structural & Molecular Biology* **30**, 1012-1022 (2023). https://doi.org/10.1038/s41594-023-00983-z

19 Wong, W. *et al.* Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, e03080 (2014). https://doi.org/10.7554/eLife.03080

20 Plaschka, C., Lin, P.-C. & Nagai, K. Structure of a pre-catalytic spliceosome. *Nature* **546**, 617-621 (2017). https://doi.org/10.1038/nature22799

21 Ru, H. *et al.* Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell* **163**, 1138-1152 (2015). https://doi.org/https://doi.org/10.1016/j.cell.2015.10.055

22 Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-112 (2013). https://doi.org/10.1038/nature12822

23 Diver, M. M., Cheng, Y. & Julius, D. Structural insights into TRPM8 inhibition and desensitization. *Science* **365**, 1434-1440 (2019). https://doi.org/10.1126/science.aax6672

24 Huang, J. *et al.* Structure-guided discovery of protein and glycan components in native mastigonemes. *Cell* **187**, 1733-1744.e1712 (2024).

https://doi.org/10.1016/j.cell.2024.02.037

25 Rasmussen, S. G. F. *et al.* Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549-555 (2011). https://doi.org/10.1038/nature10361

26 Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149-155 (2017). https://doi.org/10.1126/science.aak9979

27 Iudin, A. *et al.* EMPIAR: the Electron Microscopy Public Image Archive. *Nucleic Acids Research* **51**, D1503-D1511 (2023). https://doi.org/10.1093/nar/gkac1062

28 Paszke, A. *et al.* in *Advances in Neural Information Processing Systems 32* (eds H. Wallach *et al.*) 8024–8035 (Curran Associates, Inc., 2019).

29 Zhu, J. *et al.* A minority of final stacks yields superior amplitude in single-particle cryo-EM. *Nature Communications* **14**, 7822 (2023). https://doi.org/10.1038/s41467-023-43555-x

30 He, K., Zhang, X., Ren, S. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 770-778.

31 Zhou, Y., Barnes, C., Lu, J., Yang, J. & Li, H. in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 5738-5746.

32 Mildenhall, B. *et al.* in *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I* 405–421 (Springer-Verlag, Glasgow, United Kingdom, 2020).

33 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015). https://doi.org/10.1016/j.jsb.2015.08.008

34 Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* **193**, 1-12 (2016). https://doi.org/https://doi.org/10.1016/j.jsb.2015.11.003

35 Bengio, Y., Louradour, J., Collobert, R. & Weston, J. in *Proceedings of the 26th Annual International Conference on Machine Learning* 41–48 (Association for Computing Machinery, Montreal, Quebec, Canada, 2009).

Extended Data Table 1 | Running time comparison of cryoSPARC, cryoAI, cryoDRGN2 and cryoFIFA.

	B^{act} spliceosome (20 k particles, box size: 128 pixel)	GPCR-G protein complex (100 k particles, box size: 128 pixel)
cryoSPARC	~ 21 minutes	**~ 24 minutes**
cryoAI	> 8 hours	> 8 hours
cryoDRGN2	~ 1.5 hours	~ 8 hours
cryoFIFA	**~ 20 minutes**	~ 1.5 hours

The B^{act} spliceosome and GPCR-G protein complex datasets were compared. CryoFIFA was run with an 8x V100 GPU setup, whereas other software was run with a single GPU, as they only support single GPU configurations.

Extended Data Table 2. Datasets details and training parameters of cryoFIFA.

	Streptavidin (simulated)	**GAT1 (simulated)**	**Ribosome**	**Tri-snRNP**	**RAG1-RAG2**	**TRPV1**	**TRPM8**	**Apo-ferritin**	**Mastigonemes**	**GPCR-G protein**	**B^{act} spliceosome**
Data source	EMD-20907	EMD-33671	EMPIAR-10028	EMPIAR-10180	EMPIAR-10049	EMPIAR-10005	EMPIAR-11233	EMPIAR-11796	Nieng Yan's Group	Xiangyu Liu's Group	Yigong Shi's Group
Particle number	100,000	100,000	105,277	124,340	108,544	35,645	42040	102,000	168,000	100,000	19,310
Box size	128	128	128	128	128	128	128	128	128	128 / 228	128
Pixel size	1.072	1.691	3.768	4.247	1.845	2.431	2.6475	2.075	4.33	1.928 / 1.0825	3.3125
Steps & Epochs (Fourier masks)	32 x 3	32 x 3	32 x 3	32 x 2	32 x 5	32 x 3	32 x 6	32 x 3	30 x 5	32 x 3 / 57 x2	32 x 3
Start resolution	20	100	100	40	100	30	120	30	60	30 / 30	60
End resolution	2.4	3.4	8	9	4	5	6	4.4	8.7	4 / 3	7
Batch size (per GPU)	32	64	32	64	64	16	32	32	128	32	32
Augmentation	8 x 45°	8 x 45°	8 x 45°	8 x 45°	8 x 45°	4 x 90°	8 x 45°	8 x RD	8 x 45°	8 x 45°	8 x 45°

RD stands for random