

STRUCTURAL BIOLOGY

Correction of preferred orientation-induced distortion in cryo-electron microscopy maps

Dongjie Zhu^{1†}, Weili Cao^{1,2†}, Junxi Li^{1,2}, Chunling Wu^{1,2}, Duanfang Cao^{1,2}, Xinzheng Zhang^{1,2*}

Reconstruction maps of cryo-electron microscopy (cryo-EM) exhibit distortion when the cryo-EM dataset is incomplete, usually caused by unevenly distributed orientations. Prior efforts had been attempted to address this preferred orientation problem using tilt-collection strategy and modifications to grids or to air-water interfaces. However, these approaches often require time-consuming experiments, and the effect was always protein dependent. Here, we developed a procedure containing removing misaligned particles and an iterative reconstruction method based on signal-to-noise ratio of Fourier component to correct this distortion by recovering missing data using a purely computational algorithm. This procedure called signal-to-noise ratio iterative reconstruction method (SIRM) was applied on incomplete datasets of various proteins to fix distortion in cryo-EM maps and to a more isotropic resolution. In addition, SIRM provides a better reference map for further reconstruction refinements, resulting in an improved alignment, which ultimately improves map quality and benefits model building.

INTRODUCTION

Single-particle cryo-electron microscopy (cryo-EM) has been well developed and has generated many high-resolution reconstructions of purified proteins (1–4). The reconstructions require protein projections from various directions to cover the entire reciprocal space (5). However, many proteins embedded in ice exhibit preferred orientations (6–8), which leads to incomplete projection datasets, meaning the dataset has missing components in Fourier space. It was generally recognized that the hydrophobic (9–11) and/or the electrostatic interaction (12, 13) between air-water interfaces (AWIs)/solid-water interface and proteins causes this problem. When preferred orientation occurred, protein adsorbed to AWIs at specific orientations (14, 15). The protein densities would distort, as a result when incompleteness in Fourier space happens (16). Different experimental methods (12, 17–23) have been developed to eliminate or reduce this problem, but they are time-consuming, complex, and do not guarantee success. Reconstruction from an incomplete dataset suffers from anisotropic resolution (17, 24, 25), which causes density distortion in cryo-EM maps that can be measured by three-dimensional (3D) Fourier shell correlation (FSC), directional FSC (17, 26, 27), and sampling compensation factor (SCF) (24, 28).

The 2D electron crystallography also suffers from incompleteness in reciprocal space, often called the “missing cone” or “missing wedge” problem (29, 30). This problem was generally caused by lacking of highly tilted images (31). An algebraic reconstruction method via iterative reconstruction and applying restraints named “projective constraint optimization” (PCO) (32, 33) has been developed to overcome this problem by applying restraints in both real and reciprocal space during reconstruction to correct the map distortion caused by the missing cone problem. Nevertheless, 2D electron crystallography acquires diffraction images with good signal-to-noise ratios (SNRs) in measured amplitudes, with no phase information (34). In cryo-electron tomography (cryo-ET), several approaches—including using maximum likelihood estimation (35), deep-learning (36, 37), and

shrinkwrap (non-negativity) (38), which was adopted into PCO—have been developed to compensate the distortion from missing wedge. In these methods, the orientation of each image in the tilt series was given, and projection missing wedge was provided by a well-defined tilting range. As a result, the regions that are filled with or lacking of values from projections could easily be determined. Nevertheless, in single-particle analysis (SPA), the orientation of each image was obtained from the refinement process, which defined the orientation distribution. Therefore, in an incomplete dataset, the orientation distribution would have no such prior knowledge as tomographic data did and might be suffered from alignment error. Other than that, current deep learning approaches trained their models for recovering below-nanometer resolution information in cryo-ET. Nonetheless, the recovery ability of near-atomic resolution signal in SPA was uncertain for these models due to low micrograph SNR at this resolution (39). As a result, current approaches designed for cryo-ET could not be directly used.

Here, we present a procedure called signal-to-noise ratio iterative reconstruction method (SIRM) to reduce map distortion induced by preferred orientation. This procedure contains two steps, including removing misaligned particles and an iteratively reconstruction method based on the SNR of each Fourier component. Our results showed that the SIRM effectively recovers the density of cryo-EM reconstruction map after removing misaligned particles in incomplete datasets, resulting in more isotropic resolution and improved density. The corrected reconstructions were used as models for the iterative refinement, which furtherly reduced the alignment error and improved the overall resolution.

RESULTS

Incomplete data generates systematic misalignment

Several kinds of algebraic reconstruction methods have been developed to correct the artifact caused by missing wedge problem in tomographic reconstruction. However, different from tomography, in SPA, alignment error might arise during refinement in an incomplete dataset. This type of misalignment could be systematic and relevant to severity of missing orientation, ultimately leading to signal contamination to the filled orientations. Therefore, this contamination could

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

¹National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 100101 Beijing, China. ²University of Chinese Academy of Sciences, 100049 Beijing, China.

*Corresponding author. Email: xzzhang@ibp.ac.cn

†These authors contributed equally to this work.

be amplified, resulting in map distortion during the algebraic reconstruction. To test the alignment accuracy in an incomplete dataset, we used a high-resolution ribosome 50S dataset (EMPIAR-10509) (40) as a reference. Because large particles such as the ribosome should have very little Euler angular error in an orientational-complete dataset, and angular changes during the 3D refinement of an incomplete dataset could be regarded as errors and attributed to incompleteness, we generated seven incomplete datasets by selectively removing images of protein particles that had specific orientations from complete datasets that produced the reference reconstructions. Each incomplete dataset contained a missing cone from 30° to 60° with a step of 5° (also known as MC-30° to MC-60°; see Materials and Methods). This method created cryo-EM data with missing cone. Particles were removed from the dataset based on the “tilt” Euler angle. Taking the MC-40° incomplete dataset as an example, it retained particles with tilt ranging more than 0° to 50° and 130° to 180°. Therefore, the total 80° of tilt from 50° to 130° was missing, equivalence to a 40° missing cone. The degree of incompleteness of this dataset was measured by the point spreading function and the value of E_{od} (41), as shown in fig. S4. After that, we used the reconstruction from the complete dataset and low pass-filtered it to 20 Å as the initial model to refine the ribosome particles in each incomplete dataset using RELION (42, 43). After convergence of refinement, we computed changes in Euler angles against the reference from the complete dataset. The error-of-angle (EoA) curves were generated by computing geodesic angular distances using functions from RELION, which involved calculating the differences in orientation between particles in the incomplete dataset and the reference dataset, shown on Fig. 1A.

At 30° and 35°, the shapes of EoA curves were sharp and their peaks were closed to zero. With the enlargement of missing cone, the EoA curves began to be flatten, and their peaks tended to move away from zero. The EoA distributions of each incomplete dataset were shown at fig. S1. As the growing of missing cone, especially above 40° (fig. S1, D to G), the angular accuracy area marked in light-green, yellow, and red on the upside of the right spherical were worsen notably, indicating that the misalignment was not completely random; otherwise, the color should be almost the same as the 30° and 35° (fig. S1, A and B) showed. When converting all EoAs into vectors and summed them up, the results also suggested the occurrence of nonrandom misalignment (table S1). It was possible that the orientation estimations were affected by map distortion during refinement of the incomplete dataset, which led to systematic misalignment and exacerbated density distortion.

Another phenomenon that is worth noting was that for almost all incomplete datasets, there were small numbers of particles located inside the missing orientations, especially in fig. S1 (A and B). However, the prior knowledge we had suggested that there should have been no particle at all. If the rate of these particles was sufficient, then the final reconstructions could also be damaged.

Two-step strategy removes misaligned particles in incomplete data

From the results above, it can be concluded that misaligned particles have relatively large quantities in an incomplete dataset. To reduce overall misalignment, inspired by the cross-validation method (44, 45), we developed a strategy called “validation.” It removed particles with a change of angle (CoA) larger than a threshold between RELION and cryoSPARC (45, 46) after convergence of their own refinement, as shown in Fig. 1C. The threshold could be determined

by three- to five-fold accuracy angles or trial-and-error. After using validation, the EoA curves were shown on Fig. 1B, and EoA distributions were shown on fig. S2. All suggested that validation effectively removed misaligned particles and improved alignment accuracy, especially in the most intensive regions (fig. S2, E to G), similar to the red ones showed in fig. S1 (E to G).

Even after validation, there were small number of particles that stayed in the missing orientations (fig. S2, A and B), which could be eliminated based on the assumption that correct particles only existed near the concentrated regions. According to this hypothesis, we developed a strategy called “mask picking.” Specifically, a mask could be generated to define the concentrated regions, which excluded particles with orientation outside the mask (Fig. 1D and Materials and Methods). To check the existences of these particles in the real cryo-EM dataset, we drew four varieties of soluble proteins about their angular distributions (fig. S3). All their orientations exhibited major concentrated regions with sparser distributions further away acting as an evenly distributed background. The concentrated regions indicated that the proteins were trapped by the AWIs at specific orientations as described in previous studies (10, 12, 47–50). When the proteins were trapped by AWI, the number of particles with different orientation follows Boltzmann distribution. Therefore, the number of proteins with energy unfavorable orientations should be decreased notably instead of exhibiting an evenly distributed background. Thus, we assumed that the particles in the background are from misalignment and could be removed by the mask picking. Mask picking is an option, which can be turned off if the assumption is not suitable for a protein.

Iterative reconstruction compensates map distortion

After taking effort to remove misaligned particles, the incompleteness of data that caused by preferred orientation contributed most in map distortion, which will be described in the latter part. We thus developed an algebraic reconstruction method (the SIRM) for SPA to correct this distortion.

The SIRM contained three steps. The first step was improving alignment accuracy, including validation and mask picking, which were described above. Being optional, if the alignment of a dataset was very accurate, then users could skip this step. The second stage was the setup stage, where we computed the SNR and conducted reliability of each Fourier component. The iterative reconstruction process was run in stage three. During the setup stage, we had to determine which orientation was not presented in a SPA dataset. According to the central section theorem, each Fourier-transformed projection slice fills a central section in the 3D Fourier space of a reconstruction (5). Supposing all slices had roughly the same SNR, the SNR of a Fourier component would depend on the number of slices filled in. As a result, we could use the SNR to determine the missing Fourier component. The SNR of each Fourier component was computed (see Materials and Methods), as depicted in Fig. 2A. A SNR weighting function within a range of zero to one in reciprocal space was generated (see Materials and Methods) based on the SNR of each Fourier component. The Fourier components of the original reconstruction then multiplied by their SNR weights to create a weighted Fourier transform of the original reconstruction (WFTO). In the WFTO, if a Fourier component contained insufficient slices or located at beyond the resolution limit, its SNR weight would be close to zero, leading to a close-to-zero reliability. Conversely, if the component’s SNR exceeded a given value, the corresponding reliability would be set to one, meaning no change to this

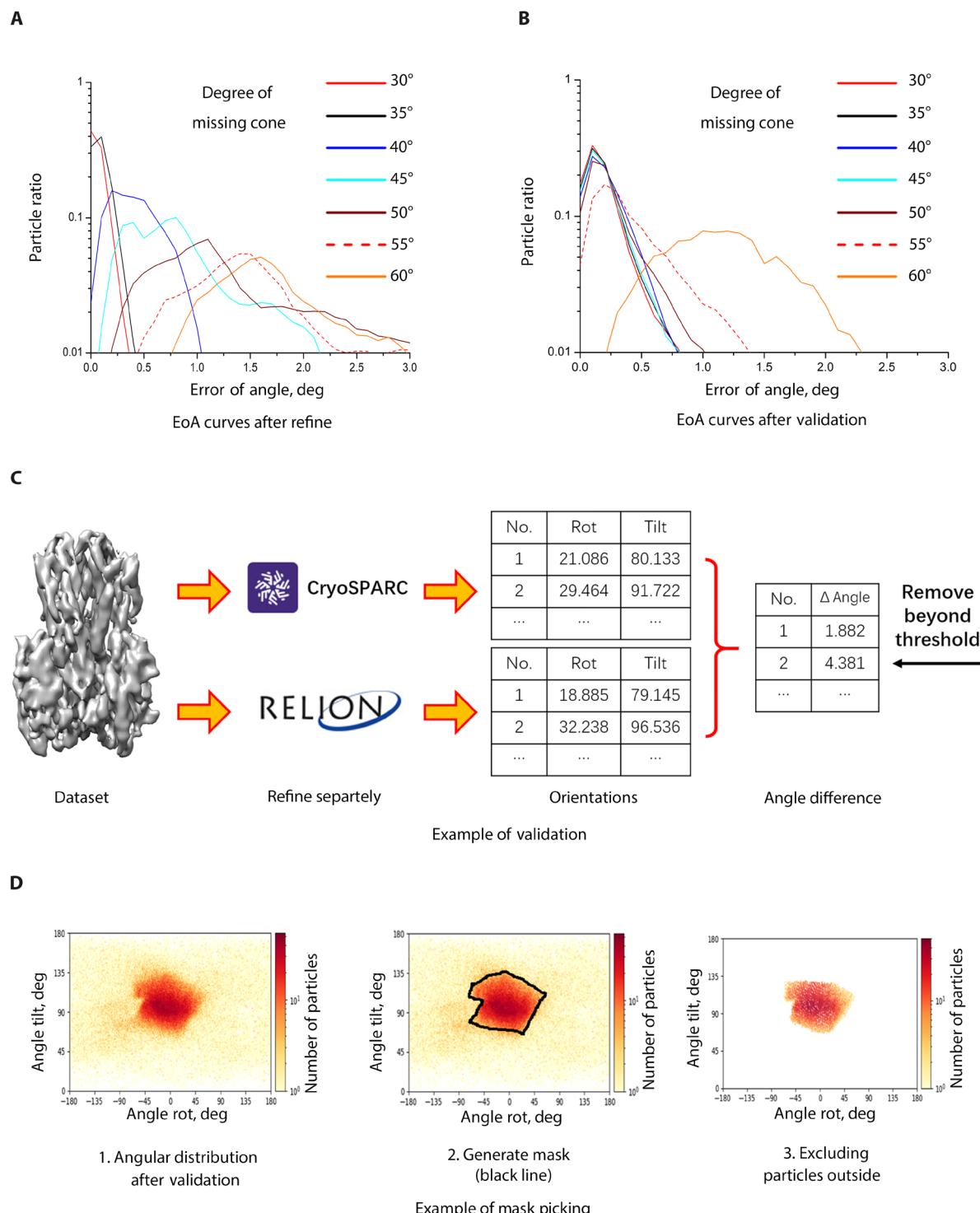
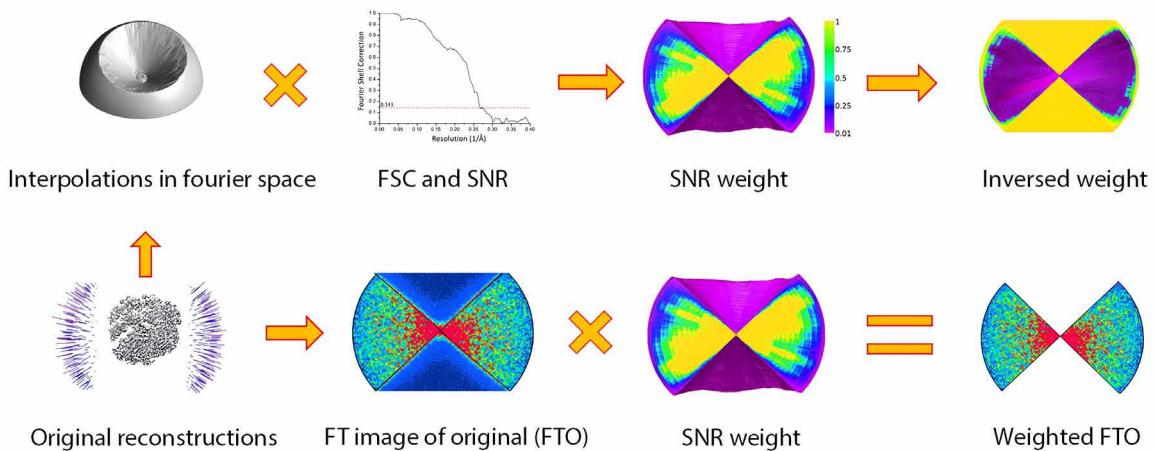


Fig. 1. SIRM removed misaligned particles. (A) EoA curves after conventional RELION refine. (B) EoA curves after validation with threshold of 1°. The EoA curves were computed against the reference (the complete dataset). Both graphs used logarithm scale in the y axis for better representation. (C) Basic example of validation. (D) Basic example of mask picking.

A

Step 1: Improve alignment of particles. Validation, mask picking, etc.

Step 2: Setup

**B**

Step 3: Iteration

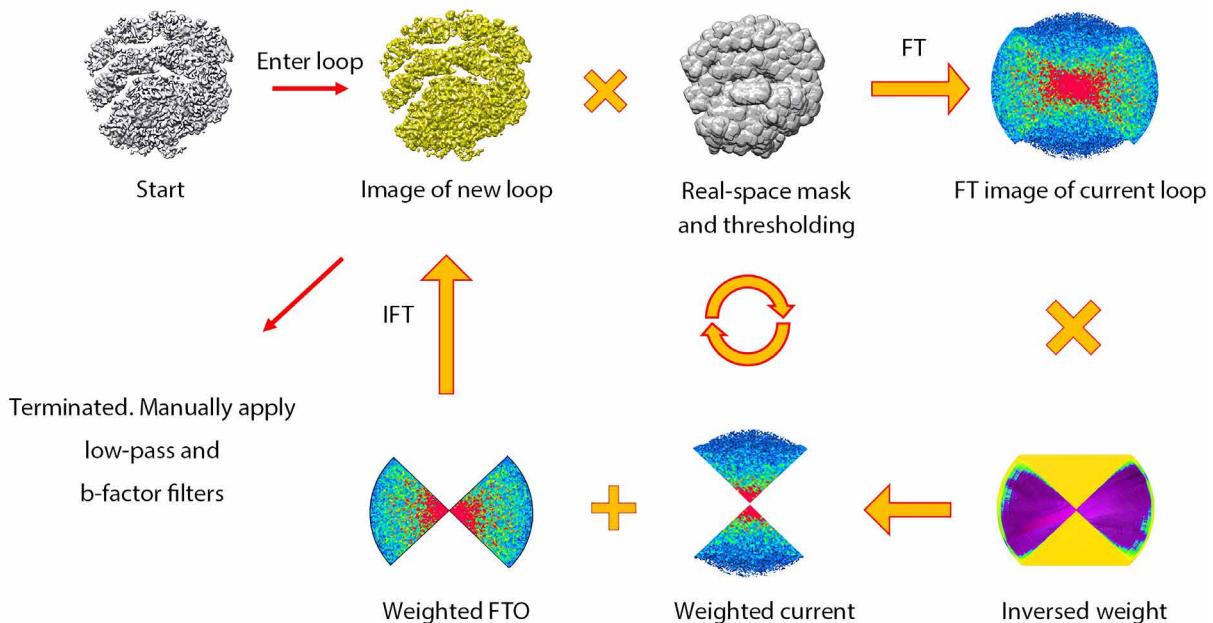


Fig. 2. Work flow of SIRM. (A) Setup procedure. Interpolations in Fourier space and SNR curve were from conventional reconstruction. FTO was multiplied to SNR weight to get the WFTO. Improving alignment procedure had been described on Fig. 1 (C and D). (B) Iteration step. Loops were started with the conventional reconstruction. FT and IFT is short for Fourier transformation and inversed Fourier transformation. Weighted current was computed by multiplying FT image of current loop and inversed weight.

component in the following iterative steps. Apart from the WFTO, an inverse weighting function was computed based on the SNR weighting (see Materials and Methods), which would be used as described below.

In the iterative reconstruction step, we initially applied an adjustable non-negative restraint (51, 52) to the reconstruction map of the current iteration to set any negative pixel values to zero (The original 3D reconstruction map was used in the first iteration). Similar steps have been used for 2dx/Focus (32, 33) software, LoTToR (52), and NUDIM (53) to correct missing wedge for 2D electron crystallography or cryo-ET data. A real-space mask could also have been applied as part of the first restraint to remove the noise. Next, the map was Fourier-transformed, and the second restraint was enforced in Fourier space. The Fourier components were multiplied by the inversed weighting function. The resulting components were added to the WFTO from the setup step. After applying this second restraint, the resulting map in Fourier space was then inverse Fourier-transformed to generate a new reconstruction map for the next iteration (Fig. 2B). The loop was terminated either by reaching a maximum iteration number or by convergence of the residual (53).

SIRM corrects map distortion without alignment error

To assess SIRM performance, we generated incomplete cryo-EM datasets of 2.2-Å apo-ferritin (54) using C1 symmetry, the 40°-tilted influenza hemagglutinin (HA) trimer (EMPIAR-10097) (17), and the ribosome 50S dataset described above, with MC-35°, MC-40°, and MC-45°. The orientation of each particle was provided by the refinement results from the complete datasets using SPA. To avoid changes in resolution, we randomly selected the same number of particles for all tests if comparison of resolution or map quality occurred, and there was no further description. After conventional reconstruction of particles via RELION software from complete and incomplete datasets, SIRM was used to correct density distortions in the cryo-EM maps from incomplete datasets. The results are compared below.

Visual inspection of the reconstructions from the MC-40° and MC-45° datasets revealed severe distortion, as comparison shown in Figs. 3 (A, E, and F), 4 (A, E, and F), and 5 (A, E, and F). Distorted side chains were elongated perpendicular to the missing directions, generating erroneous connections to other densities. It even formed a square-like artifact in the apo-ferritin incomplete datasets. This made it difficult to trace the main chain in the map. All the incomplete datasets also exhibited some level of discontinuities or main-chain fractures and visible distortions (Figs. 3, D to F; 4, D to F; and 5, D to F). In contrast, densities from corrected reconstructions exhibited distinct side-chain densities similar to those of reference reconstructions from complete datasets (Figs. 3, G to I, 4, G to I, and 5, G to I), and the apo-ferritin results did not show the square-like artifact anymore. The SIRM-processed HA-trimer dataset at MC-45° exhibited a slightly lower map quality than that at 40° (Fig. 4I). However, it still exhibited notable improvement in removing distortion when compared to the uncorrected map. This suggested that the ability to remove distortion decreased with increased data incompleteness. As for the ribosome datasets, the uncorrected maps all exhibited uncorrected linking along the vertical axis, which was disappeared in the corrected maps.

We performed quantitative analysis by calculating 3D-FSCs and directional FSCs (17) between uncorrected/corrected reconstruction and reference in the HA-trimer and the ribosome 50S incomplete datasets. They revealed strong anisotropic resolution as shown in figs. S5 and S6, which was greatly diminished in the 3D-FSC of the corrected reconstruction, even in the MC-45° dataset. We also

examined the correlation between the corrective power of SIRM and the map resolution in the MC-40° HA-trimer dataset. The results indicated that as the resolution increased, the recovery capability of SIRM improved (fig. S13).

SIRM corrects map distortion of ribosome with alignment error

In real scenarios, cryo-EM data almost certainly contain misalignment, sometimes leading to extraordinary high SNR in should be empty Fourier component. This high SNR would prevent from accurate determination of missing orientation and interrupt our correction. To test how SIRM worked in actual datasets, we made tests in ribosome dataset mentioned above with MC-45°. We removed all alignment information and used a heavily low-passed (80 Å) initial model for 3D classification, followed by autorefinement to emulate conventional SPA procedure. After convergence of refinement by cryoSPARC, we used RELION to perform a run of focused refinement. Validation and mask picking were tested, where the threshold of CoA in validation was set to 1°. The remaining orientations of particles were shown in fig. S8F.

By computing the EoA curve against the reference, the average and corresponding SD of the EoA from conventional refinement were 2.12° and 10.61° (Fig. 6A black line), respectively. The very large SD was occurred because many particles had an EoA greater than 2° and did not appear on the graph. Comparing to ribosome results above (fig. S1D and table S1), we noticed that after the extra step of de novo 3D classification, the alignment error was amplified. When we applied the validation procedure, the average and SD were reduced to 0.32° and 0.30° (Fig. 6A red line), respectively. The validation process removed approximately 10,800 particles of near 38,800, and mask picking removed almost none in this case. Distributions of EoA (fig. S7, A to C) suggested the effectiveness of validation.

We also tested whether correcting distortion in the reference of each iteration of refinement by RELION provided a better alignment by integrated SIRM into a modified version of RELION (see Materials and Methods). When we enabled SIRM during refinement, the averaged EoA and corresponding SD decreased to 0.89° and 7.53°, respectively, before using validation. The corresponding distribution of EoA also showed a more evenly distribution angular accuracy across the filled region (fig. S7D). After validation and mask picking, the average EoA was furtherly decreased to 0.19°, and SD also notably decreased to 0.21° (Fig. 6A blue line). EoA distribution suggested that validation and mask picking removed most of the highly erroneous particles (fig. S7E). This result recommended that implementing SIRM during iterative refinement was able to improve alignment accuracy when incompleteness existed. Remaining particles were re-reconstructed with SIRM. The improved densities map exhibited more details, especially in connectivity and separation between two complemented RNA chains (Fig. 6, C and D; fig. S9; and movies S1 and S2). The 3D-FSC and directional FSC graphs of SIRM reconstructions showed substantial improvements in directional resolution for missing regions comparing to conventional reconstructions (Fig. 6B, and fig. S8, A, B, and E).

SIRM corrects map distortion on smaller proteins

Smaller proteins had lower SNR and less angular accuracy, which may limit the correcting power of SIRM. To find out, we tested SIRM on several smaller proteins. We used the same HA-trimer dataset

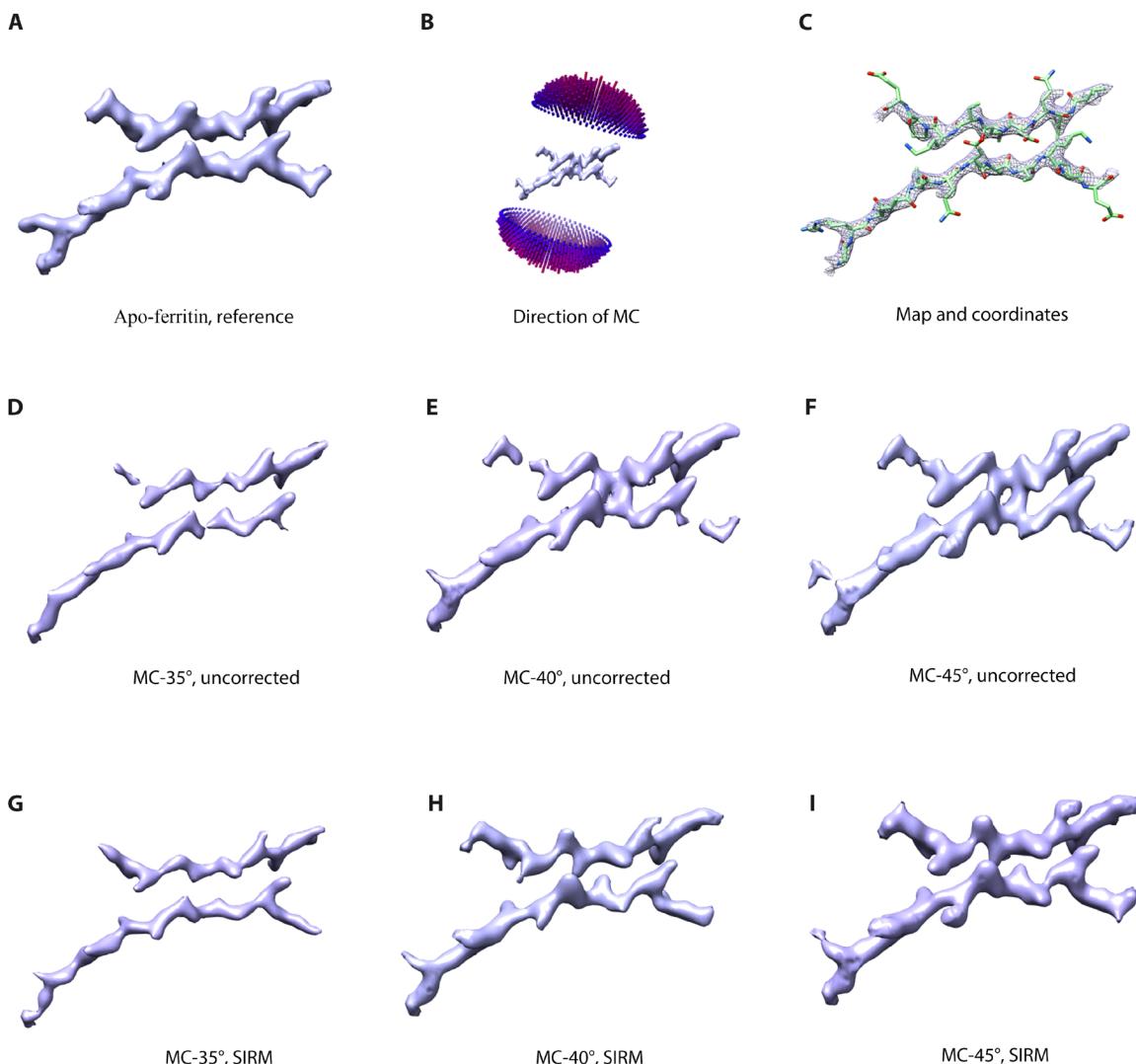


Fig. 3. Density maps of apo-ferritin incomplete datasets without alignment error. (A) Reference density, obtained from the complete C1 reconstruction. (B) Orientation distribution of the 45° missing cone was shown. (C) Sequences followed Protein Data Bank (PDB) 1MFR (64) and residues from 77 to 85 on chain L and 74 to 85 on chain V. The reference map was used here. (D to I) Densities from MC-35°, MC-40°, and MC-45° with accurate alignments. Three uncorrected maps were reconstructed by conventional relion_reconstruct program. Maps marked with SIRM were corrected by SIRM. SCF value: MC-35°, 0.665; MC-40°, 0.484; and MC-45°, 0.459.

mentioned above, with MC-40°. Alignment information in this dataset was all removed as discussed for the ribosome dataset above. SIRM were enabled during autorefinement. The validation procedure removed 43,940 of 93,569, whose CoA larger than 1.5° between cryo-SPARC3 and RELION, followed by mask picking. Comparing to results from cryoSPARC, maps reconstructed by SIRM had better connectivity, less distorted side chains, and nearly no misconnected main chains (fig. S10, B and F). The 3D-FSC and directional FSC graphs showed better isotropic results in the prospective of SD and directional FSC in the z axis (fig. S11, A to D). The curve of EoA against the reference also suggested improved overall alignment (fig. S11E).

We also used a photonsystem-1 (PS1) dataset (55) to generate a MC-45° dataset with no prior orientational or translational information. By using validation to remove CoA greater than 1.5° that removed 17,160 of 53,592 and mask picking, alignment was improved (fig. S11F). The connection of densities, 3D-FSC, and directional

FSC after SIRM reconstruction also improved compared to results from cryoSPARC3 (figs. S10J and S11, G to J).

In a real case with no known reference, we applied the entire SIRM procedure, especially validation and mask picking to a human SGLT1-MAP17 complex dataset (56, 57). Reconstructed via cryo-SPARC, this dataset exhibited severe preferred orientation, mainly heavily stretched densities in the slightly tilted horizontal direction (Fig. 7A), which agreed with the previous reconstruction EMD-33962 (57). After refinement with SIRM-integrated RELION, validation process dropped particles with CoAs larger than 1.5°, which removed 180,011 of 280,215. It was followed by mask picking that removed 25% of all remaining particles (25,807) with orientations outside a specific area (Fig. 7, C and D, and fig. S12B). To verify the effect of removing “misaligned” particles by validation and mask picking, we performed a conventional reconstruction after removing misaligned particles, and the distortion was reduced

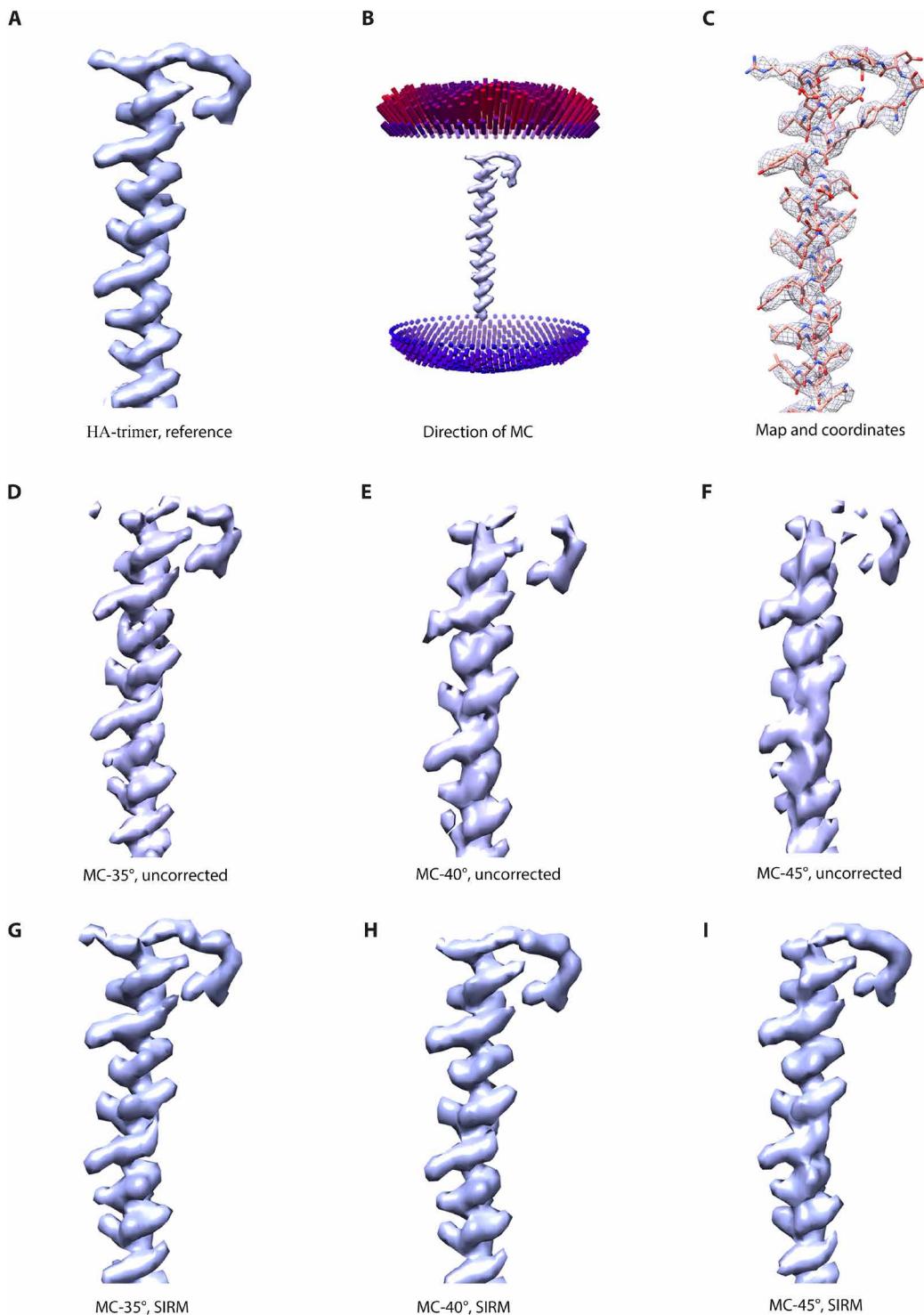


Fig. 4. Density maps of HA-trimer incomplete datasets without alignment error. (A) Reference density, obtained from the complete C3 reconstruction of EMPIAR-10097. (B) Orientation distribution of the 45° missing cone was shown. (C) Sequences followed PDB 4FNK (65) and residues from 58 to 108 on chain D. The reference map was used here. (D to I) Densities from MC-35°, MC-40°, and MC-45° with accurate alignments. Three uncorrected maps with were reconstructed by conventional relion_reconstruct program. Maps marked with SIRM were corrected by SIRM. SCF value: MC-35°, 0.409; MC-40°, 0.394; and MC-45°, 0.416.

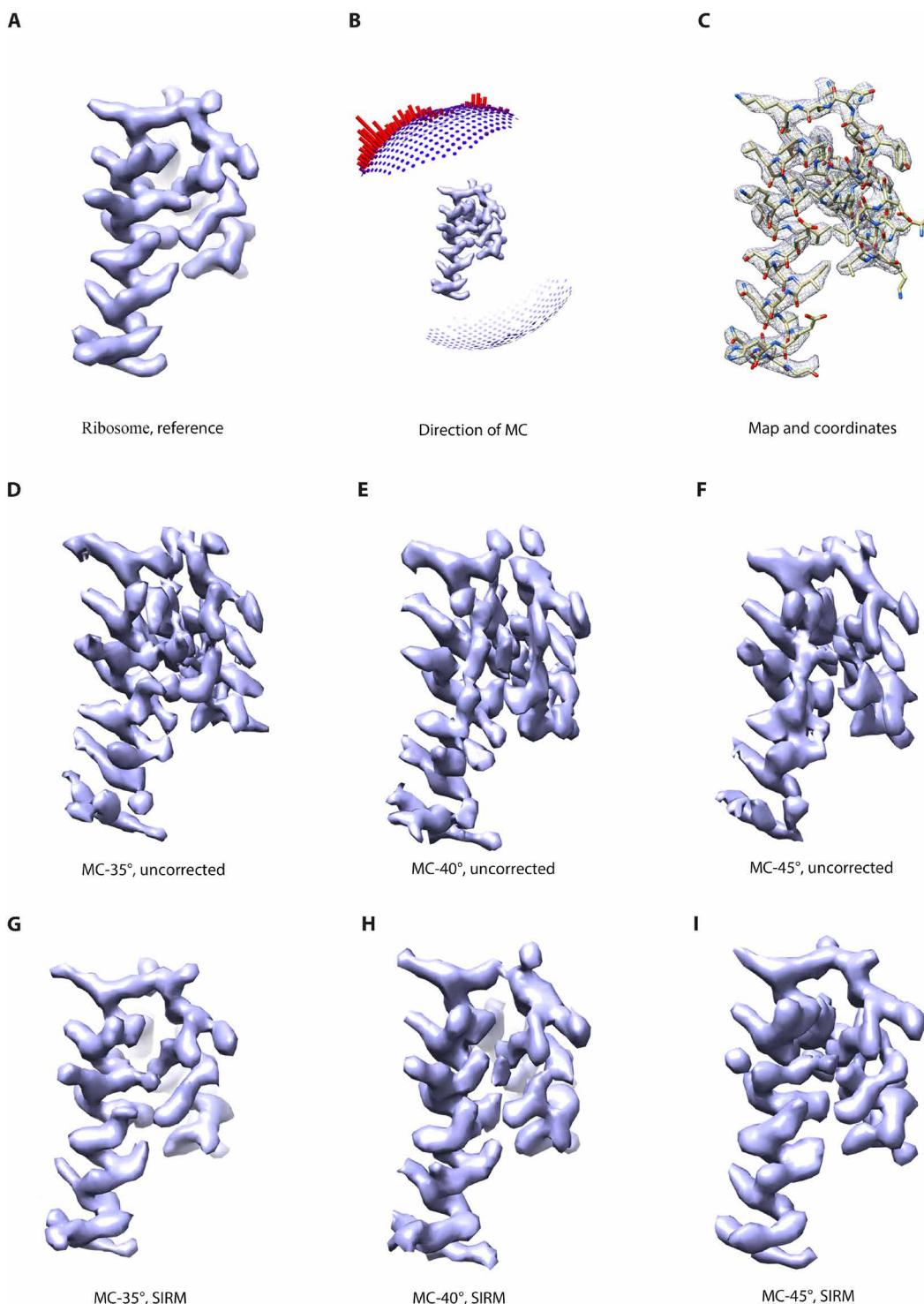


Fig. 5. Density maps of the ribosome 50S incomplete datasets without alignment error. (A) Reference density, obtained from the complete C1 reconstruction of EMPIAR-10509. (B) Orientation distribution of the 45° missing cone was shown. (C) Sequences followed PDB 7k00 and residues from 20 to 63 on chain r. The reference map was used here. (D to I) Densities from MC-35°, MC-40°, and MC-45° with accurate alignments. Three uncorrected maps with were reconstructed by conventional relion_reconstruct program. Maps marked with SIRM were corrected by SIRM. SCF value: MC-35°, 0.939; MC-40°, 0.929; and MC-45°, 0.916.

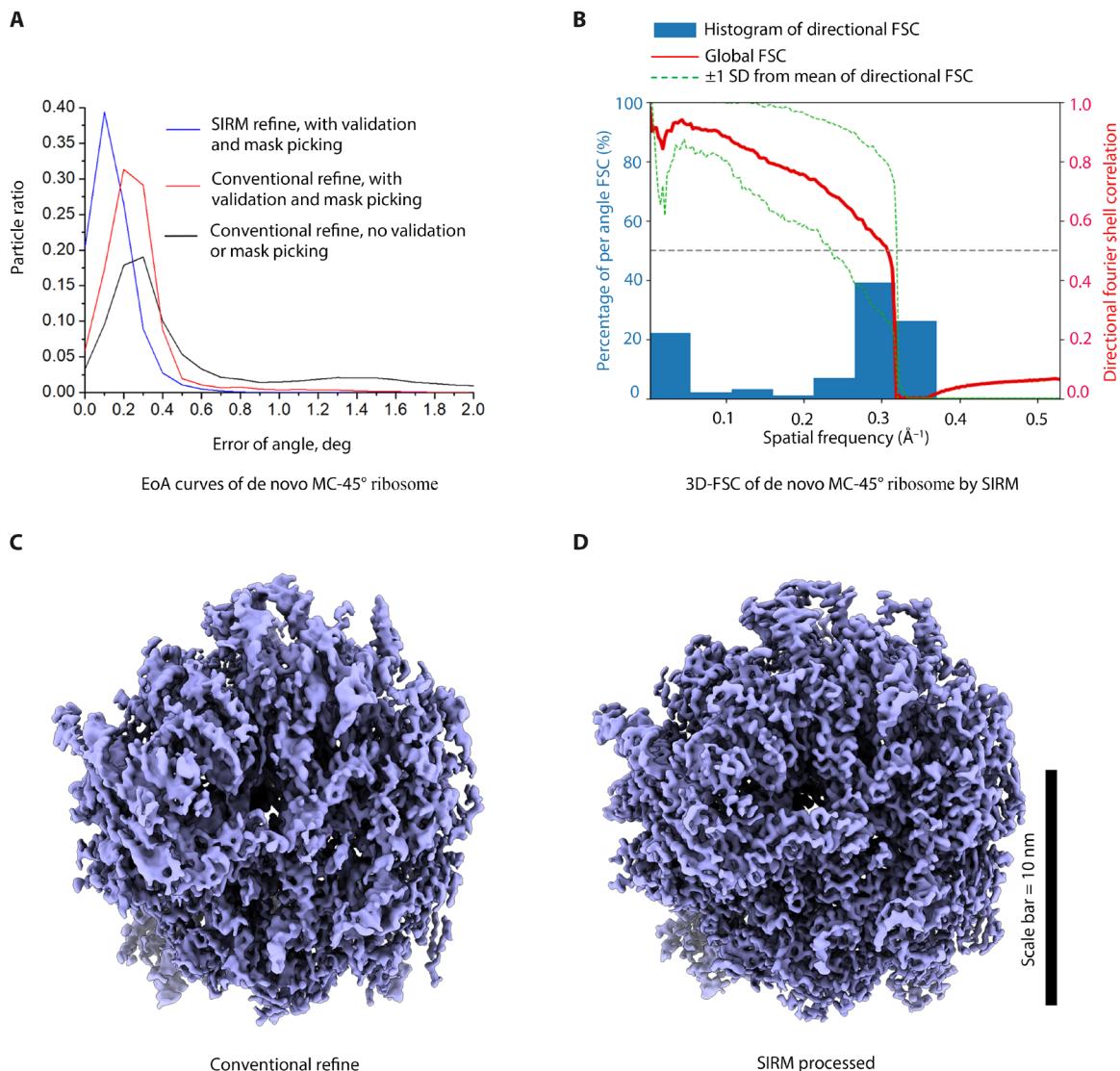


Fig. 6. Information of the incomplete ribosome 50S dataset. (A) Interval of EoA was chosen to 0.1°. The angle itself was obtained by calculating the Euler angular distance between alignment of refined particles to that of reference (from the complete dataset). “SIRM refine” meant we enabled SIRM during autorefine of RELION. Initial alignment was generated by cryoSPARC. (B) 3D-FSC was generated by comparing reconstruction of SIRM to the reference. The resolution cutoff was set to 0.5. (C) Density map of conventional refine. Reconstruction was from the final result of cryoSPARC3. (D) Density map after SIRM processions. Scale bar, 10 nm in length. SCF value: Conventional, 0.943; after removing particles, 0.906.

(fig. S12A). We were thus able to mitigate the stretched densities furtherly using SIRM reconstruction also with improved isotropic resolution (Fig. 7B; fig. S12, D and F; and movies S3 and S4).

DISCUSSION

The correction power of SIRM is related to resolution of reconstruction. Because SIRM belongs to algebraic reconstruction, it solves unknown variables in real space. A higher resolution map provides more precise boundaries of the protein density, which establishes a finer constraint relative to a low-resolution map. It can be inferred that the number of unknown variables belonged to protein densities would decrease with a finer map. Thus, the SIRM’s ability to mitigate map distortion relies on detailed boundary information from a

high-resolution map, provided by a real-space restraint; such a restraint is also related to compressive sensing.

The PCO method has been developed to recover the distortion in 2D crystal datasets. Comparing with 2D crystal data, single-particle data have low SNR and alignment errors, which are major differences. The 2D electron crystallography and cryo-ET record images with certain tilt angles, generally in level of less than 0.3°. In contrast, the alignment error in SPA was affected by the mass of particles, going less accurate with fewer mass. In a HA dataset, the reported accuracy angles were about 1.6°, much greater than tomography. After removing most of the misaligned particles in the SGLT1-MAP17 dataset, even conventional reconstruction was able to exhibit a slighter distorted map. Therefore, to correct the distorted density in an incomplete single-particle dataset, we developed

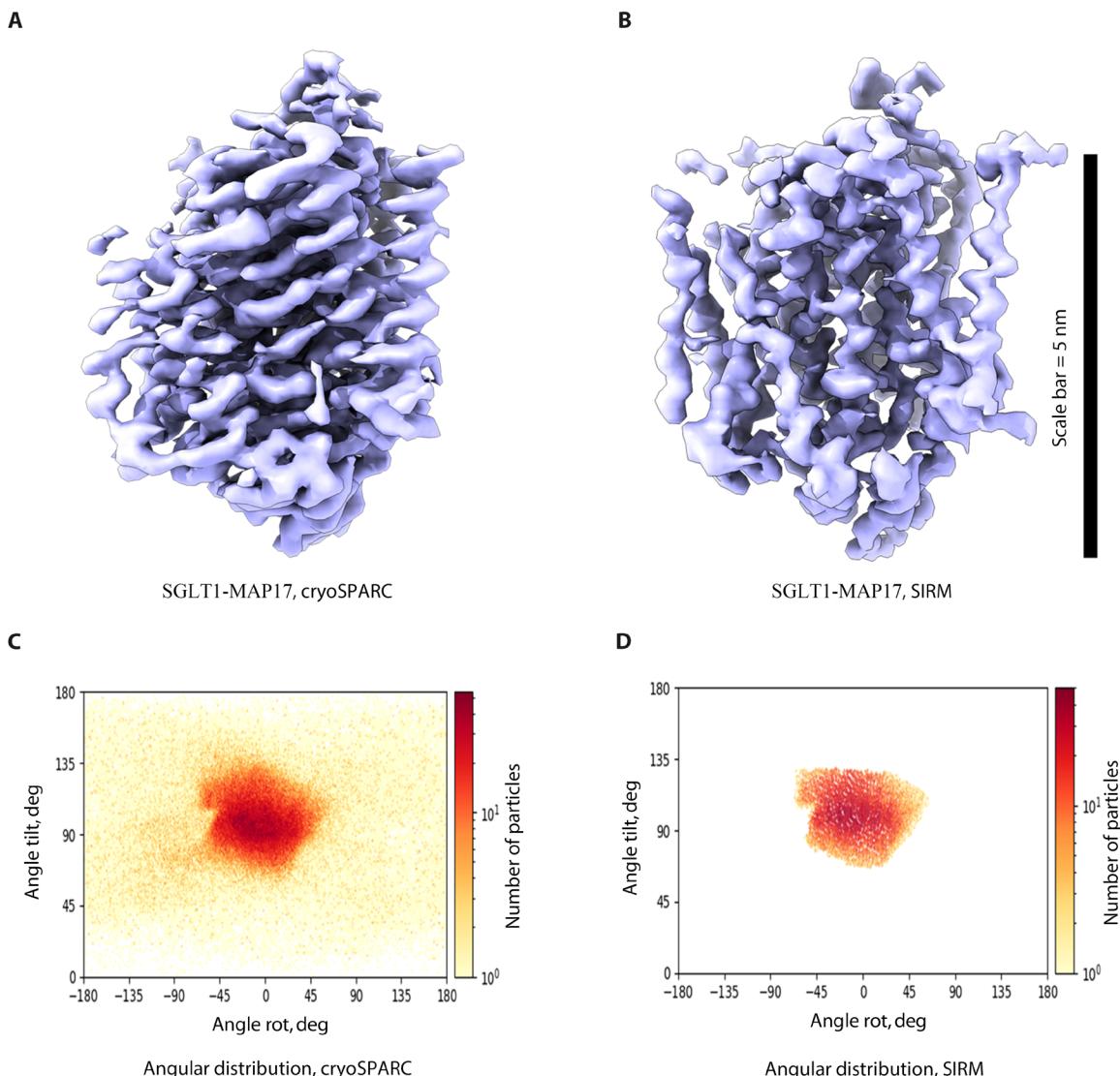


Fig. 7. Use SIRM to a human SGLT1-MAP17 complex dataset. (A and B) Densities of the human SGLT1-MAP17 complex, no sequence was used, and all maps were applied the same b-factor for sharpening and the same low-passed filter. Scale bar, 5 nm in length. (C and D) Angular distribution of the protein. “angdist” (66) software was used to produce these graphs. SCF value: Conventional, 0.904; after removing particles, 0.863.

SNR weighting functions used in the Fourier space restraint and a protocol to remove misaligned particles, both of which are critical in SIRM.

For an incomplete dataset, a null Fourier component might be filled a few times with misaligned particles. The SNR weight determined via particle numbers might be erroneous for these Fourier components, occasionally misleading our program to recognize those components as valid, which ruins the capability to correct distortion. Mask picking helps to remove these misaligned particles. In our tests, mask picking helped improve the reconstruction most of the time. Especially in the SGLT1-MAP17 case, because of the large quantities of evenly distributed particles, which filled Fourier component incorrectly, if we skipped mask picking, then the iterative reconstruction would become useless. This was also a side prove that what mask picking removed were erroneous particles. We also noticed that another approach (cryoPROS) that improves the accuracy of alignment for preferred orientation proteins

has been proposed by Zhang *et al.* (58), which involves generating pseudo particles with missing orientations and using them during refinement.

Recently, membrane proteins were solved on a flat, native membrane layer directly purified from cells (59). This technique has been successfully applied to the nuclear pore complex by tilting the grid up to 55° (60). Large proteins usually retain reasonably good contrast and sufficient signal for alignment after tilting. However, the image quality significantly deteriorates for small proteins, which affects alignment in the refinement when the tilting angle exceeds 40°, making the datasets incomplete. In reality, however, even if the sample had very substantial preferred orientation (e.g., top-view only), due to the bend of the supporting film, the degree of missing core should be around 70° rather than near 90°. When the sample stage is tilted to a lower angle (e.g., tilt 20° to 30°), the degree of missing cone could be furtherly reduced to ~45°. At this point, SIRM can be used to recover density distortion.

MATERIALS AND METHODS

SNR of grid points in Fourier space

The projection slice theorem indicates that in cases where the particle dataset has preferred orientation, the interpolation number of a Fourier component at some specific directions will be significantly less than that of other directions. Supposing each projection has roughly the same SNR at the same fraction of frequency, we can infer that the SNR of grid points is proportional to the interpolated numbers (24). Therefore, SNR can be used to represent the reliability of data. Let us assume that the increment of SNR is g after the grid point having been interpolated once in a frequency fraction. Consequently, the SNR of a single grid point “ i ” denoted as s_i will be proportional to the corresponding number of interpolated times N_i , which is

$$s_i = N_i \cdot g \quad (1)$$

The average SNR of that fraction can be computed by this equation (61)

$$\text{SNR} = \frac{2\text{FSC}}{1 - \text{FSC}} \quad (2)$$

Then, we have the following equation

$$\text{SNR} = \frac{\sum_i s_i}{N} = \frac{\sum_i N_i \cdot g}{N} \quad (3)$$

where N is the total number of grid points in a certain band of frequency. g can be expressed as

$$g = \frac{N \cdot \text{SNR}}{\sum_i N_i} \quad (4)$$

The SNR of a grid point s_i is

$$s_i = \frac{N_i \cdot N \cdot \text{SNR}}{\sum_i N_i} \quad (5)$$

SNR weight of a component

Now, we can determine whether one grid point is reliable by using a threshold p . We define its SNR weight as the following equation

$$\text{SNRW}_i = \begin{cases} 1, & s_i \geq p \text{ and } s_i \text{ inside resolution} \\ s_i/p, & s_i < p \text{ and } s_i \text{ inside resolution} \\ 0, & s_i \text{ outside resolution} \end{cases} \quad (6)$$

The inversed SNR weight is defined as

$$\text{ISNRW}_i = \begin{cases} 1 - \text{SNRW}_i, & \text{SNRW}_i > 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Principle of SIRM

In the SIRM approach, we assume that the orientation information of all particles is accurate, and missing orientations are the only cause of map distortion. Similar to the PCO method, we assume that the gray value of protein must be greater than a certain number, in most of cases zero. On the basis of these assumptions, we use the following algorithm to correct distortion.

Setup 1: Back-project all particles in Fourier space.

In setup 1 process, all particles are back-projected according to their Euler angles and contrast transfer function (CTF) parameters. This Fourier-transformed map is denoted as F_0 , and the i -th grid point in F_0 is represented as $f_{0,i}$, with its SNR denoted as s_i . Applying inverse Fourier transformation to F_0 yields map in real space, which

is recorded as R_0 . This step is the same as the conventional reconstruction process.

Setup 2: Generate WFTO.

Since we have already known its SNR weight by Eq. 6, we multiply $f_{0,i}$ to SNRW_i to get $f'_{0,i}$. The Fourier map F'_0 is composed of $f'_{0,i}$

$$f'_{0,i} = f_{0,i} \cdot \text{SNRW}_i \quad (8)$$

Because $f'_{0,i}$ will be used during iteration, we denote $f'_{0,i}$ as $f_{\text{WFTO},i}$.

The iteration: Application of the real-space restraint

In j -th iteration, the i -th grid point in R_j is marked as $r_{j,i}$. We assume that $r_{j,i}$ is non-negative. If the real-space threshold was not zero, then $r_{j,i}$ would be subtracted by this threshold first. Then, we have the following constraints

$$r'_{j,i} = \begin{cases} r_{j,i}, & r_{j,i} \geq 0 \text{ and } r_{j,i} \text{ in mask} \\ 0, & r_{j,i} < 0 \text{ or } r_{j,i} \text{ out of mask} \end{cases} \quad (9)$$

We record the density composed of $r'_{j,i}$ as R'_j . It should be noted that we use an apostrophe to indicate the application of restraint for either R_j or F_j .

The iteration: Application of the reciprocal space restraint

In j -th iteration, the i -th grid point in F_j is marked as $f_{j,i}$. After performing Fourier transformation on R'_j , we obtain the reciprocal space map F_j , where $j \geq 1$. We then apply the reciprocal space restraint to obtain F'_j . The $f'_{j,i}$ is mixed with $f_{\text{WFTO},i}$ as follows

$$f'_{j,i} = f_{\text{WFTO},i} + f_{j,i} \cdot \text{ISNRW}_i \quad (10)$$

We then obtain R_{j+1} by inverse Fourier transformation to F'_j . This iteration process should be repeated for more than 30 rounds. Once the loop is complete, we obtain a corrected real-space map R_N , where N is the number of iterations. B-factor and low-pass filter are applied to R_N afterward.

SIRM in modified RELION

The SIRM version of RELION incorporates two major changes. The first one is to the “relion_reconstruct” program. We have added a “SIRM correction” dialog under the “reconstruction” section of RELION. When set to “yes,” users need to provide an additional star file, containing the FSC curve of reconstruction. Users can generate this star-file by using “convert_star_2_SIRM_weight.py” script available on our repository. The threshold of SNR weight can be adjusted or determined by checking the “debug_fsc_snr_weight.mrc” file, which will be generated during the running of reconstruction. More details can be found on the user’s manual. The second change involves the integration of SIRM into “3D autorefine” section. A similar option is presented on “optimization” dialog. During autorefine, the program automatically generates required SNR weight using the fsc_halves_class vector of RELION.

In addition, we provide an option called “Do MWTCF” in the same dialog mentioned above. When enabled, the modified RELION calculates the average interpolation number of each frequency fraction. Components whose interpolation numbers are larger than “MWTCF value” multiplying by average will be used for FSC calculation.

The validation procedure

To validate a particle, we compare its Euler angular distance between cryoSPARC and RELION. Users can use scripts provided on

our repository to complete this process. It needs be noticed that there might be small rotation between reconstruction from cryo-SPARC and RELION. This rotation should be corrected before validation, and this procedure is describe on our repository. Determining the threshold of angular distance is critical. We recommend consulting the value of “accuracy angles” from 3D autorefine of RELION. Typically, we multiply this value by three to five for large protein datasets and one to two for small proteins.

Mask picking particles

The mask picking process involves using a mask on sphere to pick particles about their orientations. This mask could be produced by our scripts. In our experiments, we first used the bild file of a star file after validation. The bild file was generated by “--write_bild” option in “relion_star_handler” program of our modified RELION. We then converted this bild file into a low-passed mrc file, which could be examined by visualization software such as UCSF Chimera (62). Last, an angular list file was produced from the mrc file to filter particles.

Generate incomplete dataset

Particles with tilt angles between angle 1 and angle 2 were removed from the complete dataset using “relion_star_handler --select rlnAngleTilt”. The resulting angle range was equal to the desired angle of incompleteness, with angle 1 and angle 2 being symmetrical to 90°. The alignment information, including orientation and translation, could be kept or removed depending on the usage. If removed, then the alignment would be generated through conventional SPA, including 3D classification and 3D refinement. Another procedure to produce incomplete dataset was keeping particles that formed a missing cone. This could be done by using a script, which calculated the angle between each particle and a certain pair of Euler angle rot/tilt, and then excluding whose angles were larger than a given threshold.

Image processing for datasets without alignment

We generated incomplete datasets that includes HA-trimer with MC-40°, ribosome 50S MC-45°, and PS1 MC-45°, and an initial model of the ribosome and PS1 was generated using “3D initial model” in RELION. The HA-trimer model was from our reconstruction to EMPIAR-10097, which resulted in a 3.0-Å resolution map. All models were low pass-filtered to 80 Å and used for “3D classification” in RELION, with C3 symmetry used for the HA-trimer dataset and C1 for the rests. After the reported resolution reached around 8 Å, homogeneous refinement in cryoSPARC3 was performed for each dataset using the models from the last iteration in 3D classification. Default settings were used, except for symmetry and disabling per-particle defocus refinement because all defoci of particles should be accurate enough. After convergence, “csparc2star.py” in pyem (63) package was used to convert outputs of cryoSPARC to star files. Further 3D autorefine in SIRM-RELION was performed, with models from cryoSPARC being low-passed to 10 Å. The angular step size was set to 1.8° for both initial angular sampling and local searches from autosampling. The “Do SIRM correction” and Do MWTCF were enabled. “Threshold of the SIRM SNR weight” was set to 1.0 for ribosome and 5.0 for other datasets. MWTCF value was set to 0.5. Reported resolution after post-process was 3.45 Å for HA-trimer, 3.16 Å for ribosome, and 3.63 Å for PS1. FSC weight file required by SIRM reconstruction was generated after post-process. Validation and mask picking were then performed.

For ribosome dataset, validation threshold was set to 1.0°. For PS1 and HA-trimer dataset, thresholds were set to 1.5°. Masks that required by mask picking were generated. The orientational lists for picking had step of 1.875°, and the picking threshold had tolerance of 1.0°, slightly above half the step. SIRM reconstructions were carried out after validation and mask picking. “--maxres” was set to 2.8 for ribosome and PS1 and 3.0 for HA to reduce computational cost. Additional real-space mask in SIRM correction was used for each dataset. Iteration number was set to 60 for HA-trimer, 80 for PS1, and 500 for ribosome. The resulting reconstructions were low pass-filtered to 3.6 Å for ribosome, PS1, and HA-trimer.

Supplementary Materials

This PDF file includes:

Figs. S1 to S13

Table S1

Legends for movies S1 to S4

Legends for data files S1 to S4

References

Other Supplementary Material for this manuscript includes the following:

Movies S1 to S4

Data files S1 to S4

REFERENCES AND NOTES

- Y. Cheng, Single-particle Cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015).
- D. Lyumkis, Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **294**, 5181–5197 (2019).
- T. Nakane, A. Kotecha, A. Sente, G. McMullan, S. Masiulis, P. M. G. E. Brown, I. T. Grigoras, L. Malinauskaitė, T. Malinauskas, J. Miehling, T. Uchański, L. Yu, D. Karia, E. V. Pechnikova, E. de Jong, J. Keizer, M. Bischoff, J. McCormack, P. Tiemeijer, S. W. Hardwick, D. Y. Chirgadze, G. Murshudov, A. R. Aricescu, S. H. W. Scheres, Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
- K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
- J. Frank, *Three Dimensional Electron Microscopy of Macromolecular Assemblies* (Oxford Univ. Press, 2006).
- R. M. Glaeser, How good can cryo-EM become? *Nat. Methods* **13**, 28–32 (2016).
- R. M. Glaeser, B.-G. Han, Opinion: Hazards faced by macromolecules when confined to thin aqueous films. *Biophys. Rep.* **3**, 1–7 (2017).
- Y. Liu, X. Meng, Z. Liu, Deformed grids for single-particle cryo-electron microscopy of specimens exhibiting a preferred orientation. *J. Struct. Biol.* **182**, 255–258 (2013).
- K. A. Taylor, R. M. Glaeser, Electron microscopy of frozen hydrated biological specimens. *J. Ultrastruct. Res.* **55**, 448–456 (1976).
- R. M. Glaeser, B. G. Han, R. Csencsits, A. Kililea, A. Pulk, J. H. D. Cate, Factors that influence the formation and stability of thin, Cryo-EM specimens. *Biophys. J.* **110**, 749–755 (2016).
- R. M. Glaeser, Proteins, interfaces, and cryo-EM grids. *Curr. Opin. Colloid Interface Sci.* **34**, 1–8 (2018).
- B. Li, D. Zhu, H. Shi, X. Zhang, Effect of charge on protein preferred orientation at the air-water interface in cryo-electron microscopy. *J. Struct. Biol.* **213**, 107783 (2021).
- R. Bromberg, K. Cai, Y. Guo, D. Plymire, T. Emde, M. Puzio, D. Borek, Z. Otwinowski, The His-tag as a decoy modulating preferred orientation in cryoEM. *Front. Mol. Biosci.* **9**, 912072 (2022).
- A. J. Noble, V. P. Danedy, H. Wei, J. Brasch, J. Chase, P. Acharya, Y. Z. Tan, Z. Zhang, L. Y. Kim, G. Scapin, M. Rapp, E. T. Eng, W. J. Rice, A. Cheng, C. J. Negro, L. Shapiro, P. D. Kwong, D. Jeruzalmi, A. des Georges, C. S. Potter, B. Carragher, Routine single particle CryoEM sample and grid characterization by tomography. *eLife* **7**, e34257 (2018).
- E. D’Imprima, D. Floris, M. Joppe, R. Sánchez, M. Grininger, W. Kühlbrandt, Protein denaturation at the air-water interface and how to prevent it. *eLife* **8**, e42747 (2019).
- N. Boisset, P. A. Penczek, J. C. Taveau, V. You, F. De Haas, J. Lamy, Overabundant single-particle electron microscope views induce a three-dimensional reconstruction artifact. *Ultramicroscopy* **74**, 201–207 (1998).
- Y. Z. Tan, P. R. Baldwin, J. H. Davis, J. R. Williamson, C. S. Potter, B. Carragher, D. Lyumkis, Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
- R. B. G. Ravelli, F. J. T. Nijhuis, R. J. M. Henderikx, G. Weissenberger, S. Thewessem, A. Gijssbers, B. W. A. M. Beulen, C. López-Iglesias, P. J. Peters, Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nat. Commun.* **11**, 2563 (2020).

19. K. Naydenova, P. Jia, C. J. Russo, Cryo-EM with sub-1 Å specimen movement. *Science* **370**, 223–226 (2020).
20. C. J. Russo, L. A. Passmore, Progress towards an optimal specimen support for electron cryomicroscopy. *Curr. Opin. Struct. Biol.* **37**, 81–89 (2016).
21. L. Zheng, N. Liu, X. Gao, W. Zhu, K. Liu, C. Wu, R. Yan, J. Zhang, X. Gao, Y. Yao, B. Deng, J. Xu, Y. Lu, Z. Liu, M. Li, X. Wei, H. W. Wang, H. Peng, Uniform thin ice on ultraflat graphene for high-resolution cryo-EM. *Nat. Methods* **20**, 123–130 (2023).
22. F. Wang, Y. Liu, Z. Yu, S. Li, S. Feng, Y. Cheng, D. A. Agard, General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24269–24273 (2020).
23. H. Fan, F. Sun, Developing graphene grids for cryoelectron microscopy. *Front. Mol. Biosci.* **9**, 937253 (2022).
24. P. R. Baldwin, D. Lyumkis, Non-uniformity of projection distributions attenuates resolution in Cryo-EM. *Prog. Biophys. Mol. Biol.* **150**, 160–183 (2020).
25. C. O. S. Sorzano, D. Semchonok, S. C. Lin, Y. C. Lo, J. L. Vilas, A. Jiménez-Moreno, M. Gragera, S. Vacca, D. Maluenda, M. Martínez, E. Ramírez-Aportela, R. Melero, A. Cuervo, J. J. Conesa, P. Conesa, P. Losana, L. del Caño, J. J. de la Morena, Y. C. Fonseca, R. Sánchez-García, D. Strelak, E. Fernández-Giménez, F. de Isidro, D. Herreros, P. L. Kastritis, R. Marabini, B. D. Bruce, J. M. Carazo, Algorithmic robustness to preferred orientations in single particle analysis by CryoEM. *J. Struct. Biol.* **213**, 107695 (2021).
26. J. L. Vilas, H. D. Tagare, J. Vargas, J. M. Carazo, C. O. S. Sorzano, Measuring local-directional resolution and local anisotropy in cryo-EM maps. *Nat. Commun.* **11**, 55 (2020).
27. M. Beckers, C. Sachse, Permutation testing of Fourier shell correlation for resolution estimation of cryo-EM maps. *J. Struct. Biol.* **212**, 107579 (2020).
28. P. R. Baldwin, D. Lyumkis, Tools for visualizing and analyzing Fourier space sampling in Cryo-EM. *Prog. Biophys. Mol. Biol.* **160**, 53–65 (2021).
29. T. S. Baker, R. Henderson, Electron cryomicroscopy in *International Tables for Crystallography* (Wiley, 2006), vol. F, chap. 19.6, pp. 451–463.
30. E. V. Orlova, H. R. Saibil, Structural analysis of macromolecular assemblies by electron microscopy. *Chem. Rev.* **111**, 7710–7748 (2011).
31. M. Barth, R. K. Bryan, R. Hegerl, Approximation of missing-cone data in 3D electron microscopy. *Ultramicroscopy* **31**, 365–378 (1989).
32. B. R. Gipson, D. J. Masiel, N. D. Browning, J. Spence, K. Mitsuoka, H. Stahlberg, Automatic recovery of missing amplitudes and phases in tilt-limited electron crystallography of two-dimensional crystals. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* **84**, 011916 (2011).
33. N. Bhatia, R. D. Righetto, R. McLeod, D. Caujolle-Bert, D. Castano-Diez, K. N. Goldie, H. Stahlberg, Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol.* **198**, 124–133 (2017).
34. D. B. Williams, C. B. Carter, *The Transmission Electron Microscope* (Springer, 2009); <https://link.springer.com/book/10.1007/978-0-387-76501-3>.
35. R. Yan, S. V. Venkatakrishnan, J. Liu, C. A. Bouman, W. Jiang, MBIR: A cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. *J. Struct. Biol.* **206**, 183–192 (2019).
36. Y. T. Liu, H. Zhang, H. Wang, C. L. Tao, G. Q. Bi, Z. H. Zhou, Isotropic reconstruction for electron tomography with deep learning. *Nat. Commun.* **13**, 6482 (2022).
37. G. Ding, Y. Liu, R. Zhang, H. L. Xin, A joint deep learning model to recover information and reduce artifacts in missing-wedge sinograms for electron tomography and beyond. *Sci. Rep.* **9**, 12803 (2019).
38. S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, J. C. H. Spence, X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B* **68**, 140101 (2003).
39. Y. Cheng, Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876–880 (2018).
40. Z. L. Watson, F. R. Ward, R. Méheust, O. Ad, A. Schepartz, J. F. Banfield, J. H. D. Cate, Structure of the bacterial ribosome at 2 Å resolution. *eLife* **9**, e60482 (2020).
41. K. Naydenova, C. J. Russo, Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun.* **8**, 629 (2017).
42. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
43. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, S. H. W. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
44. B. Falkner, G. F. Schröder, Cross-validation in cryo-EM-based structural modeling. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8930–8935 (2013).
45. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: Adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
46. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
47. E. D’Imprima, W. Kühlbrandt, Current limitations to high-resolution structure determination by single-particle cryoEM. *Q. Rev. Biophys.* **54**, e4 (2021).
48. A. J. Noble, H. Wei, V. P. Danedy, Z. Zhang, Y. Z. Tan, C. S. Potter, B. Carragher, Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* **15**, 793–795 (2018).
49. D. P. Klebl, M. S. C. Gravett, D. Kontziampasis, D. J. Wright, R. S. Bon, D. C. F. Monteiro, M. Trebbin, F. Sobott, H. D. White, M. C. Darrow, R. F. Thompson, S. P. Muench, Need for speed: Examining protein behavior during CryoEM grid preparation at different timescales. *Structure* **28**, 1238–1248.e4 (2020).
50. B. G. Han, R. M. Glaeser, Simple assay for adsorption of proteins to the air–water interface. *J. Struct. Biol.* **213**, 107798 (2021).
51. Y. Jiang, R. Hovden, P. Ercius, D. Wang, Y. Yu, H. D. Abruna, D. A. Muller, V. Elser, Filling the missing wedge in tomography: A constraint-based reconstruction method for 3D TEM/STEM imaging. *Microsc. Microanal.* **19**, 768–769 (2013).
52. X. Zhai, D. Lei, M. Zhang, J. Liu, H. Wu, Y. Yu, L. Zhang, G. Ren, LoTToR: An algorithm for missing-wedge correction of the low-tilt tomographic 3D reconstruction of a single-molecule structure. *Sci. Rep.* **10**, 10489 (2020).
53. Z. Geng, Z. She, Q. Zhou, Z. Dong, F. Zhan, H. Zhang, J. Xu, Z. Gao, Y. Dong, NUDIM: A non-uniform fast Fourier transform based dual-space constraint iterative reconstruction method in biological electron tomography. *J. Struct. Biol.* **213**, 107770 (2021).
54. C. Wu, H. Shi, D. Zhu, K. Fan, X. Zhang, Low-cooling-rate freezing in biomolecular cryo-electron microscopy for recovery of initial frames. *QRB Discov.* **2**, e11 (2021).
55. X. Su, J. Ma, X. Pan, X. Zhao, W. Chang, Z. Liu, X. Zhang, M. Li, Antenna arrangement and energy transfer pathways of a green algal photosystem-I-LHCl supercomplex. *Nat. Plants* **5**, 273–281 (2019).
56. Y. Niu, W. Cui, R. Liu, S. Wang, H. Ke, X. Lei, L. Chen, Structural mechanism of SGLT1 inhibitors. *Nat. Commun.* **13**, 6440 (2022).
57. W. Cui, Y. Niu, Z. Sun, R. Liu, L. Chen, Structures of human SGLT in the occluded state reveal conformational changes during sugar transport. *Nat. Commun.* **14**, 2920 (2023).
58. H. Zhang, D. Zheng, Q. Wu, Z. Shi, M. Hu, Addressing preferred orientation in single-particle cryo-EM through AI-generated auxiliary particles. *bioRxiv* 2023.09.26.559492 (2023).
59. B. Krishnarjuna, A. Ramamoorthy, Detergent-free isolation of membrane proteins and strategies to study them in a near-native membrane environment. *Biomolecules* **12**, 1076 (2022).
60. X. Zhu, G. Huang, C. Zeng, X. Zhan, K. Liang, Q. Xu, Y. Zhao, P. Wang, Q. Wang, Q. Zhou, Q. Tao, M. Liu, J. Lei, C. Yan, Y. Shi, Structure of the cytoplasmic ring of the Xenopus laevis nuclear pore complex. *Science* **376**, eabl8280 (2022).
61. N. Grigorieff, Resolution measurement in structures derived from single particles. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56** (Pt 10), 1270–1277 (2000).
62. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
63. D. Asarnow, E. Palovcak, Y. Cheng, asarnow/pyem: UCSF pyem v0.5. (Zenodo, 2019); <https://doi.org/10.5281/zenodo.3576630>.
64. Y. Ha, D. Shi, G. W. Small, E. C. Theil, N. M. Allewell, Crystal structure of bullfrog M ferritin at 2.8 Å resolution: Analysis of subunit interactions and the binuclear metal center. *J. Biol. Inorg. Chem.* **4**, 243–256 (1999).
65. D. C. Ekiert, A. K. Kashyap, J. Steel, A. Rubrum, G. Bhabha, R. Khayat, J. H. Lee, M. A. Dillon, R. E. O’Neil, A. M. Faynboym, M. Horowitz, L. Horowitz, A. B. Ward, P. Palese, R. Webby, R. A. Lerner, R. R. Bhatt, I. A. Wilson, Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* **489**, 526–532 (2012).
66. G. Gaulier, angdist: Plot the 2D histogram of Euler angles covered by a set of cryo-EM particles (Zenodo, 2020); <https://doi.org/10.5281/zenodo.4395763>.
67. J. Wang, J. Ma, Z. Cheng, X. Meng, L. You, M. Wang, X. Zhang, Y. Wang, A CRISPR evolutionary arms race: Structural insights into viral anti-CRISPR/Cas responses. *Cell Res.* **26**, 1165–1168 (2016).
68. K. H. Chan, V. Petrychenko, C. Mueller, C. Maracci, W. Holtkamp, D. N. Wilson, N. Fischer, M. V. Rodnina, Mechanism of ribosome rescue by alternative ribosome-rescue factor B. *Nat. Commun.* **11**, 4106 (2020).

Acknowledgments: We thank L. Chen at Molecular Medicine, Peking University, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Beijing, China for providing the human SGLT1-MAP17 complex dataset. We thank J. Ma at Shenzhen Bay Laboratory, Shenzhen, China and C. C. Huang at the Institute of Biophysics, Chinese Academy Sciences, Beijing, China for providing anticas-3 and UL9 + DNA datasets. We thank L. F. Kong at the Institute of Biophysics, Chinese Academy Sciences, Beijing, China for cryo-EM data management. **Funding:** The project was funded by the National Natural Science Foundation of China (31930069, 32150010, and 32325027 to X.Z. and 32371278 to D.C.), the National Key R&D Program of China (2021YFA1301501 to W.C. and 2017YFA0504700 to X.Z.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB37040101 to X.Z.), the Key Research Program of Frontier Sciences at the Chinese Academy of Sciences (ZDBS-LY-SM003 to X.Z.), the Basic Research Program Based on Major Scientific Infrastructures, CAS (JZHKYPT-2021-05 to X.Z.), and the Key Laboratory of Biomacromolecules, Chinese Academy of Sciences (ZGD-2023-05 to X.Z.). **Author contributions:** Conceptualization: X.Z. and D.Z. Methodology: X.Z., D.Z., and W.C.

Investigation: D.Z., W.C., J.L., and C.W. Resources: X.Z., W.C., and C.W. Software: D.Z. and W.C. Formal analysis: D.Z. and W.C. Validation: X.Z., W.C., D.Z., and C.W. Visualization: W.C., D.Z., D.C., and C.W. Supervision: X.Z. Writing—original draft: D.Z., W.C., C.W., and D.C. Writing—review and editing: D.Z., W.C., and C.W. Data curation: D.Z. and W.C. Funding acquisition: X.Z. Project administration: X.Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The EM density maps have been deposited in the Electron Microscopy Data Bank under the code EMD-38080 (D_1300042682, SIRM reconstruction of the MC-45° de novo processed ribosome 50S dataset), EMD-38081 (D_1300042687, conventional reconstruction of the MC-45° de novo processed ribosome 50S dataset), EMD-39299 (D_1300045682, conventional reconstruction of the human SGLT1-MAP1 complex dataset), EMD-38082 (D_1300042689, SIRM reconstruction of the human SGLT1-MAP17 complex dataset), EMD-38083 (D_1300042692, SIRM reconstruction of the MC-40° de novo processed HA-trimer dataset), EMD-38084 (D_1300042696, conventional reconstruction of the MC-40° de novo processed HA-trimer dataset), EMD-38085 (D_1300042711, SIRM reconstruction of the MC-45° de novo processed PS1 dataset), EMD-38086 (D_1300042704,

conventional reconstruction of the MC-45° de novo processed PS1 dataset). The corresponding EMDB files, validation reports, and a readme file were also included in the Supplementary Materials and other supporting files. Source codes of modified version of RELION (SIRM-RELION), user's manual, and other scripts used are available at https://github.com/homurachan/SIRM_RELION. A permanent archive of the software repository is available at DOI: 10.5281/zenodo.10642618. A permanent archive of all the graphical lines in standard machine-readable format is available at DOI: 10.5061/dryad.73n5tb354. Supplemental data files S1 to S4 are available on Zenodo at DOI: 10.5281/zenodo.14900160. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 17 November 2023

Accepted 6 June 2024

Published 26 July 2024

10.1126/sciadv.adn0092

Correction (6 March 2025): Due to the large file size of data S1 to S4, these files have been moved to a Zenodo repository and are available at DOI: 10.5281/zenodo.14900160. The data and materials availability statement has been updated to include this information online and in the PDF.