

Overcoming the preferred-orientation problem in cryo-EM with self-supervised deep learning

Received: 18 April 2024

Yun-Tao Liu^{1,2,4}, Hongcheng Fan^{1,2,4}, Jason J. Hu^{1,2,3} & Z. Hong Zhou^{1,2}  

Accepted: 10 October 2024

Published online: 18 November 2024

 Check for updates

While advances in single-particle cryo-EM have enabled the structural determination of macromolecular complexes at atomic resolution, particle orientation bias (the ‘preferred’ orientation problem) remains a complication for most specimens. Existing solutions have relied on biochemical and physical strategies applied to the specimen and are often complex and challenging. Here, we develop splIsoNet, an end-to-end self-supervised deep learning-based software to address map anisotropy and particle misalignment caused by the preferred-orientation problem. Using preferred-orientation views to recover molecular information in under-sampled views, splIsoNet improves both angular isotropy and particle alignment accuracy during 3D reconstruction. We demonstrate splIsoNet’s ability to generate near-isotropic reconstructions from representative biological systems with limited views, including ribosomes, β-galactosidases and a previously intractable hemagglutinin trimer dataset. splIsoNet can also be generalized to improve map isotropy and particle alignment of preferentially oriented molecules in subtomogram averaging. Therefore, without additional specimen-preparation procedures, splIsoNet provides a general computational solution to the preferred-orientation problem.

Recent advances in cryogenic electron microscopy (cryo-EM) hardware^{1–4} and image-processing algorithms^{5–8} have enabled atomic-structure determination of well-behaved macromolecular complexes in their near-native states^{3,4}, transforming cryo-EM into a mainstream technique of structural biology. Ideally, biological complexes are evenly distributed with random orientations in a thin layer of vitreous ice on a cryo-EM grid⁹. However, this is rarely the case; instead, cryo-EM specimens typically suffer from the preferred-orientation problem, which is characterized by uneven or biased distribution of particle orientations¹⁰. Preferred orientation generally arises from the interaction of macromolecules with the air–water interface (AWI) or the support film–water interface during specimen preparation^{11–13} (Fig. 1a). Because electrostatic and hydrophobicity properties vary

across the surface of the protein, certain protein surfaces are preferred to be exposed on the AWI^{14–16}. This preferred binding to the AWI leads to orientation bias¹⁵.

The preferred-orientation problem can severely compromise the quality and accuracy of three-dimensional (3D) reconstructions, especially for low-symmetry or asymmetric macromolecules, resulting in artifacts in the reconstruction, such as skewed secondary structures, broken peptides or nucleotide chains, and distorted side chains or nucleobase densities¹⁷. In some severe cases, the preferred-orientation problem leads to a reconstructed map with a distorted or incorrect shape¹⁰, which deteriorates the alignment of particles during the 3D refinement process, resulting in artifacts and even failure to obtain a reconstruction. These artifacts misrepresent the true structure of

¹Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA, USA. ²California NanoSystems Institute, University of California, Los Angeles, CA, USA. ³Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA.

⁴These authors contributed equally: Yun-Tao Liu, Hongcheng Fan.  e-mail: Hong.Zhou@UCLA.edu

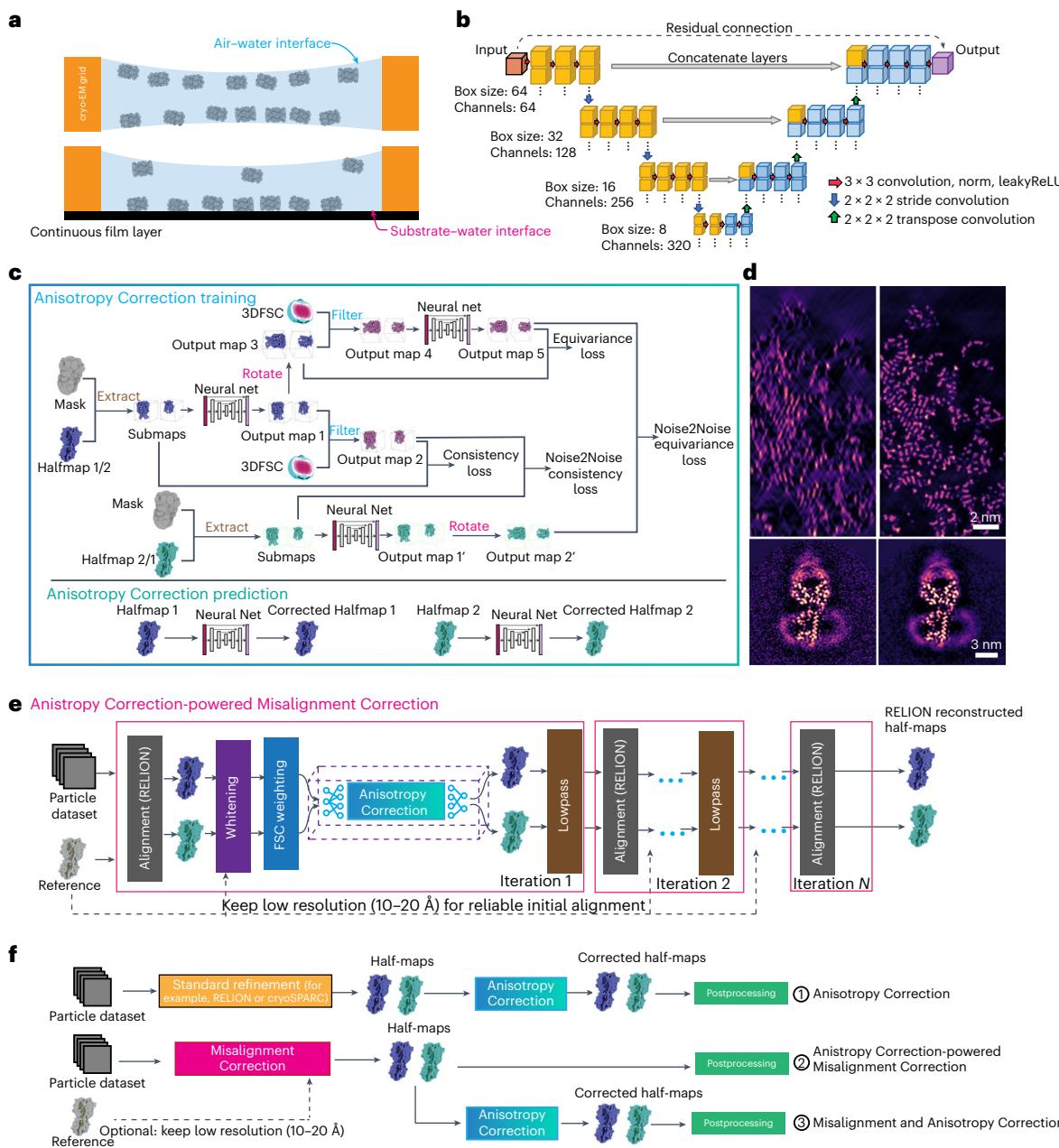


Fig. 1 | Workflow of splIsoNet. **a**, Illustrations of two common scenarios in which particles exhibit preferred orientations on cryo-EM grids. **b**, Architecture of the U-net neural network used in splIsoNet. **c**, A diagram showing the Anisotropy Correction algorithm, highlighting the four loss functions. **d**, 2D slices of maps before and after splIsoNet Anisotropy Correction. Top, a subregion of simulated

ribosome. Bottom, denoising effect of splIsoNet with EMD-20806 (ref. 57). **e**, A diagram showing Anisotropy Correction-powered Misalignment Correction. **f**, Three workflows for using splIsoNet to tackle the preferred-orientation problem.

the macromolecules and make atomic modeling difficult. Therefore, overcoming the prevailing challenge of preferred orientation is crucial for high-throughput and accurate cryo-EM structure determination. Addressing this issue will broaden cryo-EM's applicability, enabling it to be used to study difficult biological specimens, enhance our understanding of molecular processes, and accelerate structure-guided drug design.

Past efforts to address the preferred-orientation problem have resorted to biochemical or physical approaches. Biochemical solutions include using detergents to occupy the air–water interface^{12,14,18} or chemically modifying the protein specimen¹⁹, but these techniques are specimen-specific, typically require exhaustive screening, and can be detrimental to specimen quality. Physical approaches

to alleviating the preferred-orientation problem abound, including coating cryo-EM grids with support films^{11,20–24}, using time-resolved vitrification devices^{25–29}, collecting data from thicker ice regions³⁰, and tilting the grids during image acquisition^{10,31}. However, these methods may be labor-intensive and expensive, and often have unintended consequences, such as higher background noise in images and a different type of preferred orientation induced by a support film–water interface. Therefore, the preferred-orientation problem lacks a general, straightforward solution, owing to the specificity and complexity of existing methods.

Here, we introduce a self-supervised deep learning method called single-particle IsoNet (splIsoNet) to restore angular isotropy in specimens that exhibit preferred orientations. splIsoNet addresses both

an isotropic reconstruction and particle misalignment caused by the preferred-orientation problem. splsoNet consists of two modules, (1) map Anisotropy Correction and (2) Anisotropy Correction-powered particle Misalignment Correction. When applied to single-particle cryo-EM, splsoNet generated better reconstructions for various biological macromolecules—such as a 3.5-Å-resolution reconstruction from a non-tilted hemagglutinin (HA) dataset, which was previously intractable—by improving alignment accuracy and angular isotropy. When applied to subtomogram averaging of uneven orientation distribution, it also improves the quality of the reconstructed map. The splsoNet computational pipeline is specimen-independent and offers a sought-after solution to the preferred-orientation problem in single-particle cryo-EM and single-particle cryogenic electron tomography (cryo-ET) (that is, subtomogram averaging).

Results

Strategy of splsoNet to address two ramifications of the preferred-orientation problem

Even in anisotropic reconstructions, maps generated by cryo-EM contain rich and recurring information from structural features, such as secondary structures, amino acids, and atoms, in well-sampled views. The organization of these features is constrained by hydrogen bonds, van der Waals forces, and electrostatic and other interactions to minimize free energy, in a manner consistent with biophysical laws. With splsoNet, we demonstrate that it is possible to use modern deep learning methods to recover under-sampled information by merging information from well-sampled views of similar features present in the reconstruction, thus remedying the preferred-orientation problem.

The preferred-orientation problem has two major ramifications that hinder single-particle analysis. The first ramification is anisotropic reconstruction. According to the central slice theorem, the Fourier transform of a two-dimensional (2D) projection (that is, a particle image) corresponds to a 2D slice in 3D Fourier transform of the cryo-EM map at its corresponding orientation³². A uniformly distributed particle orientation corresponds to a well-sampled 3D Fourier space, whereas an uneven particle orientation distribution corresponds to uneven sampling in Fourier space, causing a distorted 3D reconstruction in real space. The second ramification is particle misalignment. Standard cryo-EM workflows require iterative particle orientation determination or alignment, where the maps reconstructed from one iteration are used as references for the next iteration^{5,7}. Thus, distorted maps reconstructed from the particles suffering from the preferred-orientation problem lead to errors in particle alignment in the following iterations. The particle-misalignment ramification has also been demonstrated in previous studies^{33,34}. In some special cases, such as the EMPIAR-10096 dataset, this ramification is so deleterious that one study has argued that it is “the major contributor to preferred orientation caused artifacts”³⁴.

In splsoNet, we implemented two modules—(1) Anisotropy Correction (Fig. 1c,d) and (2) Anisotropy Correction-powered Misalignment Correction (Fig. 1e)—to remedy these ramifications of the preferred-orientation problem. These modules can be used alone or in tandem—first Misalignment Correction and then Anisotropy Correction to overcome the preferred-orientation problem, as demonstrated below (Fig. 1f).

The Anisotropy Correction module

The Anisotropy Correction module requires two half-maps, a 3D Fourier shell correlation (3DFSC) volume¹⁰, and solvent mask as input. For user convenience, we implement a fast 3DFSC algorithm in splsoNet. Regions in a 3DFSC volume with low intensity values indicate that these regions in the Fourier space of a map are under-sampled, whereas an isotropic map will generate an isotropic 3DFSC volume. Hence, we use 3DFSC as a proxy of directional resolution and map isotropy. Applying an anisotropic 3DFSC volume as a filter on a map

in Fourier space approximates a cryo-EM map with uneven orientation sampling.

The Anisotropy Correction module trains a deep neural network of U-net architecture³⁵ (Fig. 1b). The inputs into this module are the two half-maps from a standard cryo-EM reconstruction or subtomogram averaging pipeline (Fig. 1c). Each half-map is randomly divided into smaller submaps as training data for splsoNet. The loss function of the network training is defined as the weighted sum of four losses: consistency loss³⁶, equivariance loss³⁶, noise2noise³⁷ consistency loss, and noise2noise equivariance loss. By minimizing the loss function, the neural network simultaneously accomplishes denoising and information recovery (Fig. 1d and Extended Data Fig. 1). The conceptual design of consistency and equivariance losses is based on the end-to-end, self-supervised framework of equivariant imaging proposed previously³⁶. In splsoNet, consistency loss allows information to be populated into the Fourier regions with missing data without preventing overfitting. Incorporation of equivariance loss into our design ensures reliability of the populated information without introducing overfitting artifacts (Extended Data Figs. 1 and 2).

Minimizing consistency loss ensures that the 3DFSC-filtered, network-predicted map approximates the original map; therefore, the recovered information will reside only in the original under-sampled area in the Fourier space. To calculate the consistency loss, each extracted submap is passed through the neural network to generate the initial output map (referred to as output map 1). Output map 1 is then filtered by the 3DFSC filter, resulting in a 3DFSC-filtered and network-predicted map (output map 2). Consistency loss is defined as the difference between the output map 2 and the input submap.

By minimizing the equivariance loss, the network learns how to recover information leveraging the fact that characteristic features, such as amino acids and nucleic acids, should be similar regardless of their positions and orientations. To calculate this loss, output map 1, mentioned above, is rotated computationally to produce output map 3. Then, output map 4 is obtained by applying the 3DFSC filter to output map 3. Output map 4 is processed through the neural network to generate output map 5. Minimization of equivariance loss ensures that output map 5 matches output map 3.

Recent cryo-EM and cryo-ET packages, such as Cryo-CARE³⁸, Topaz-denoise³⁹, WARP/M^{8,40}, and RELION5 (ref. 41), have implemented denoising convolutional neural networks in the absence of explicit access to ground-truth images with the noise2noise framework³⁷. Noise2noise relies on pairs of noisy images to extract information about their shared signal. In the splsoNet implementation, the output maps generated from two independent determined 3D half-maps are treated as a pair for noise2noise network training (Fig. 1c). Specifically, the noise2noise consistency loss calculates the difference between output map 2 and the submap extracted from the other half-map. To calculate the noise2noise equivariance loss, the submap from the other half-map is passed through the neural network to produce output map 1', which is then rotated to produce output map 2'. The noise2noise equivariance loss is calculated to minimize the difference between output map 2' and output map 5. This denoising algorithm acts as a local filter, improving map quality and suppressing artifacts (Fig. 1d).

We tested the limit of information recovery for Anisotropy Correction with simulated data. A ribosome density map was created from the atomic model of a translating 70S ribosome (Protein Data Bank (PDB): 8BOX)⁴² and used to simulate distorted maps suffering from various levels of the preferred-orientation problem. We applied a missing-cone filter that zeros values in the Fourier space up to certain angles to create three maps with missing cones, mimicking different severities of map anisotropy (Fig. 2a–i). Applying larger (45°, 60°, and 75°) angles of the missing-cone filter results in more severe density elongation (Fig. 2b,e,h and Extended Data Fig. 3a–c) and more distorted rRNA (Fig. 2c,f,i). Applying Anisotropy Correction improved the isotropy of all the three distorted maps, especially for the 45° and 60° missing-cone

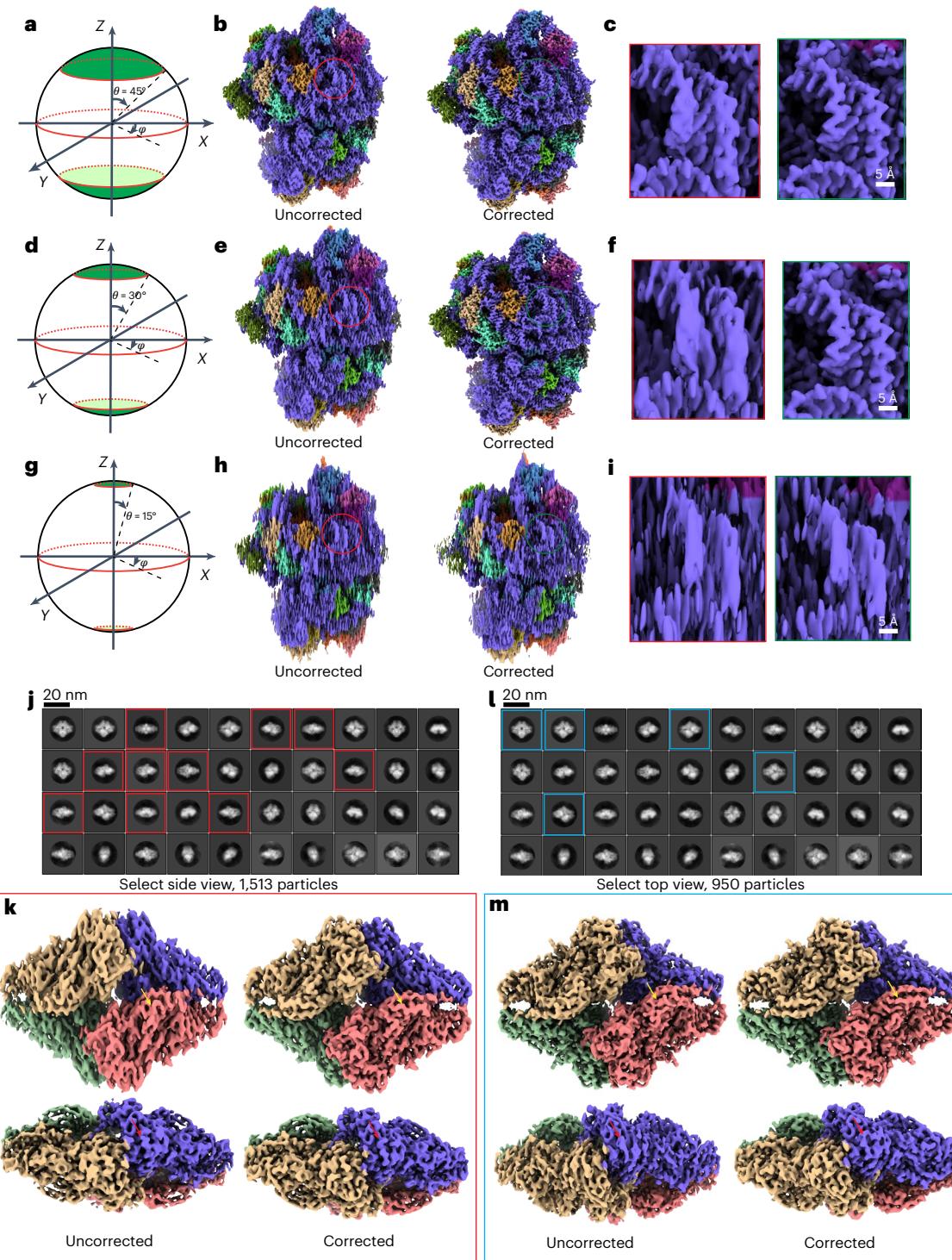


Fig. 2 | Evaluation of information recovery by spIsoNet. **a–i**, Anisotropy correction applied to synthetic data generated from the ribosome structure (PDB: 8BOX)⁴². Three missing-cone datasets were made for comparison. The red panel show zoom-in the density regions of the uncorrected map, and the green

panel show the density regions of corrected map after Anisotropy Correction. **j–m**, Anisotropy Correction was applied to cryo-EM β -galactosidase density maps reconstructed from different views in 2D classifications in **j** and **l**.

filtered maps, where the density elongation is eliminated and both bases and phosphates of nucleotides are well-resolved (Fig. 2b,c,e,f). For the severe situation with -75° to 75° missing cones, spIsoNet only slightly reduced elongation distortion of the RNA backbone (Fig. 2i and Extended Data Fig. 3a). We conclude that spIsoNet's Anisotropy Correction module can faithfully and effectively recover missing information for simulated data suffering from -60° to 60° missing cones.

The Anisotropy Correction-powered Misalignment Correction module

The Anisotropy Correction-powered Misalignment Correction module is an integrated workflow comprising map filtering, Anisotropy Correction, and RELION auto-refine (Fig. 1e), with Anisotropic Correction at its kernel. The results of this module include improved particle-orientation parameters (saved in STAR file) and two reconstructed half-maps from

RELION. After each iteration of 3D refinement, post-processing-like filters including whitening⁴³ and FSC weighting⁴⁴ are applied to the produced half-maps. Then, splIsoNet's Anisotropy Correction module is applied to these filtered half-maps. The resulting corrected half-maps are then low-pass filtered to their corresponding resolution on the basis of the gold-standard FSC criterion of 0.143. The two filtered half-maps are used as reference for the next iteration of orientation estimation. This module has an optional argument of ‘–keep_lowres’ to preserve the reliable low-resolution (10–20 Å) information from the initial reference map to enhance the accuracy of the initial alignment. This option is desirable for specimens that suffer from severe preferred orientation in which a correct initial map could not be obtained. Although low-resolution information from the reference is used for alignment, the final reconstruction is obtained solely from original particles.

Notably, in this module, the neural network used in splIsoNet exploits prior knowledge (consistency and equivariance losses) to regularize cryo-EM reconstruction. The practice of using deep learning to regularize the structure determination process is akin to the denoising algorithms implemented in M⁸ or RELION5 Blush regularization⁴¹. But unlike these methods, we not only have a denoiser neural network, but also incorporate the isotropic prior information to further improve the resolution and density quality of the reconstructed maps (Extended Data Figs. 4–7).

Applications in real-world preferred-orientation problems

Application of the Anisotropy Correction module to a tetrameric protein complex—the β-galactosidase. We tested splIsoNet on an RELION tutorial dataset containing β-galactosidase. By selecting side-view particles and top-view particles from 2D class averages (Fig. 2j,l), we curated two subsets of particles with preferred orientation and performed standard RELION 3D reconstructions (Fig. 2k,m). Because accurate alignment of the particles was already available, we applied only Anisotropy Correction to obtain corrected maps (Fig. 2k,m). After splIsoNet Anisotropy Correction, the maps had reduced density elongation, continuous β-strands, and clearer α-helical pitch. Hence, our tests on the β-galactosidase dataset demonstrate that the Anisotropy Correction module alone robustly reduces artifacts caused by top-view dominant or side-view dominant orientations.

Application to a protein dataset with moderate preferred-orientation problem—hemagglutinin trimer with 40° tilt. The cryo-EM map (EMDB-8731) of an HA trimer from a dataset (EMPIAR-10097) acquired at a 40° stage tilt in a previous study¹⁰ shows relatively poor quality, as illustrated by unclear side chain densities like rod-like α-helices (Fig. 3a,b and Extended Data Figs. 4 and 5). To evaluate the performance of splIsoNet, we first performed Anisotropic Correction on the half-maps and discovered that the corrected map exhibits better quality with higher local resolution and less noise (Extended Data Figs. 4a and 5b). The side chain densities, which cannot be observed in the original map, are discernable in the anisotropy-corrected map (Fig. 3a,b). We then applied the Anisotropy Correction-powered Misalignment Correction module on this particle dataset and obtained a cryo-EM map at 4.1-Å resolution (Fig. 3a) with well-resolved α-helical pitch, clear side chain densities (Fig. 3b), and a near-spherical (0.991) 3DFSC (Fig. 3c,e). The misalignment-corrected map shows improved map-to-model FSC (Extended Data Fig. 4b), as well as larger isotropic regions of Fourier shell occupancy⁴⁵ (FSO; Fig. 3d,f) compared with the original map. Therefore, for this dataset, both Anisotropic Correction and Misalignment Correction improve the quality of cryo-EM reconstructions.

Application to a protein dataset with severe preferred-orientation problem—hemagglutinin trimer without tilt. The untilted dataset (EMPIAR-10096) of the HA trimer has severe preferred orientation (Fig. 4a), leading to a severely distorted cryo-EM map¹⁰ when standard

cryo-EM structure-determination pipelines, including cryoSPARC and RELION, are used (Extended Data Fig. 6 and 7). For example, the RELION-reconstructed cryo-EM map is much shorter than the true structure and does not have an interpretable secondary structure (Fig. 4b). Because splIsoNet's Anisotropy Correction relies on reasonably accurate information within the cryo-EM map, Anisotropy Correction alone does not work well with such a severely distorted map. Instead, we used the splIsoNet's Anisotropy Correction-powered Misalignment Correction module to process the particle dataset, using the HA trimer map reconstructed from the tilt dataset as the reference (Fig. 3a, right). During every iteration of refinement, we kept low-resolution information (up to 10 Å) from the reference for reliable orientation alignment. Misalignment Correction generated a map with the correct shape (Fig. 4b) and improved isotropy (Fig. 4c–f). Consistent with the resolutions determined by the half-map FSC (3.5 Å) and model-to-map FSC (3.6 Å) (Fig. 4h), the map shows discernable amino acid side chains and glycans. These high-resolution features are sufficient to support atomic model building (Fig. 4g). Notably, particle-orientation distributions changed considerably (Fig. 4i,j), indicating that Misalignment Correction identified and corrected many incorrectly assigned orientations in the standard RELION processing pipeline. The particle-misalignment-corrected map can be then treated by Anisotropy Correction, further leading to a denoised structure with more continuous densities (Fig. 4b and Extended Data Figs. 6 and 7).

Interestingly, incorrect orientation assignment is often hard to evaluate and could mislead orientation-distribution-based metrics, such as cryo-EF score. This score quantifies the uniformity of point spread function for a particle-orientation distribution, reflected by data coverage in 3D Fourier space⁴⁶. Typically, a cryo-EF score below 0.5 indicates that particles with this orientation distribution are not able to yield a high-resolution ($>1/(4 \text{ \AA})$) structure. However, although particles processed with standard RELION refinement failed to produce correct structures, their corresponding cryo-EF score is 0.61, surpassing the 0.5 threshold. This overestimation is attributed to the presence of misaligned particles (Fig. 4i). After Misalignment Correction, the particle-orientation distribution has a reduced cryo-EF score of 0.33, falling below the cryo-EF threshold of 0.5 and revealing a truly severe preferred-orientation problem (Fig. 4j). Even with the low cryo-EF score, we still obtained a high-resolution structure with splIsoNet, rendering the cryo-EF threshold of 0.5 no longer applicable. Thus, splIsoNet broadens the utility of cryo-EM by effectively tackling this severe preferred-orientation problem (cryo-EF < 0.5), previously perceived as impossible to overcome in achieving high resolution.

Application to a dataset of an asymmetric, protein–nucleic acid complex. Using the *Acinetobacter baumannii* 70S ribosome dataset (EMPIAR-10406)⁴⁷, we curated a preferential-oriented dataset (Extended Data Fig. 8a) by choosing particles within a specified Euler angles range (rot angles (-20° – 140°) and tilt angles (100° – 160°)). Standard reconstruction using these particles yields a map exhibiting elongated and fragmented densities (Fig. 5a–d and Extended Data Fig. 8b). With Anisotropic Correction, the corrected map showed substantial improvement in map quality, represented by continuous map density, higher local resolution, and less noise (Fig. 5a–d and Extended Data Fig. 8c and 9a,b). We then performed both Misalignment and Anisotropic Correction on this dataset. We observed that this combined protocol improved the reconstruction, with further enhanced visibility of side chains (Fig. 5a–d). We observed that utilizing subtomogram averaging maps from either 70S or 80S ribosomes (Extended Data Fig. 9c,d) as references, while maintaining an initial resolution of 15 Å for alignment (Extended Data Fig. 9e,f), consistently yielded comparable high-quality maps without model bias (Fig. 5a–d and Extended Data Fig. 9g). Thus, splIsoNet improves particle alignment and mitigates anisotropy for asymmetric and nucleic-acid-containing molecules.

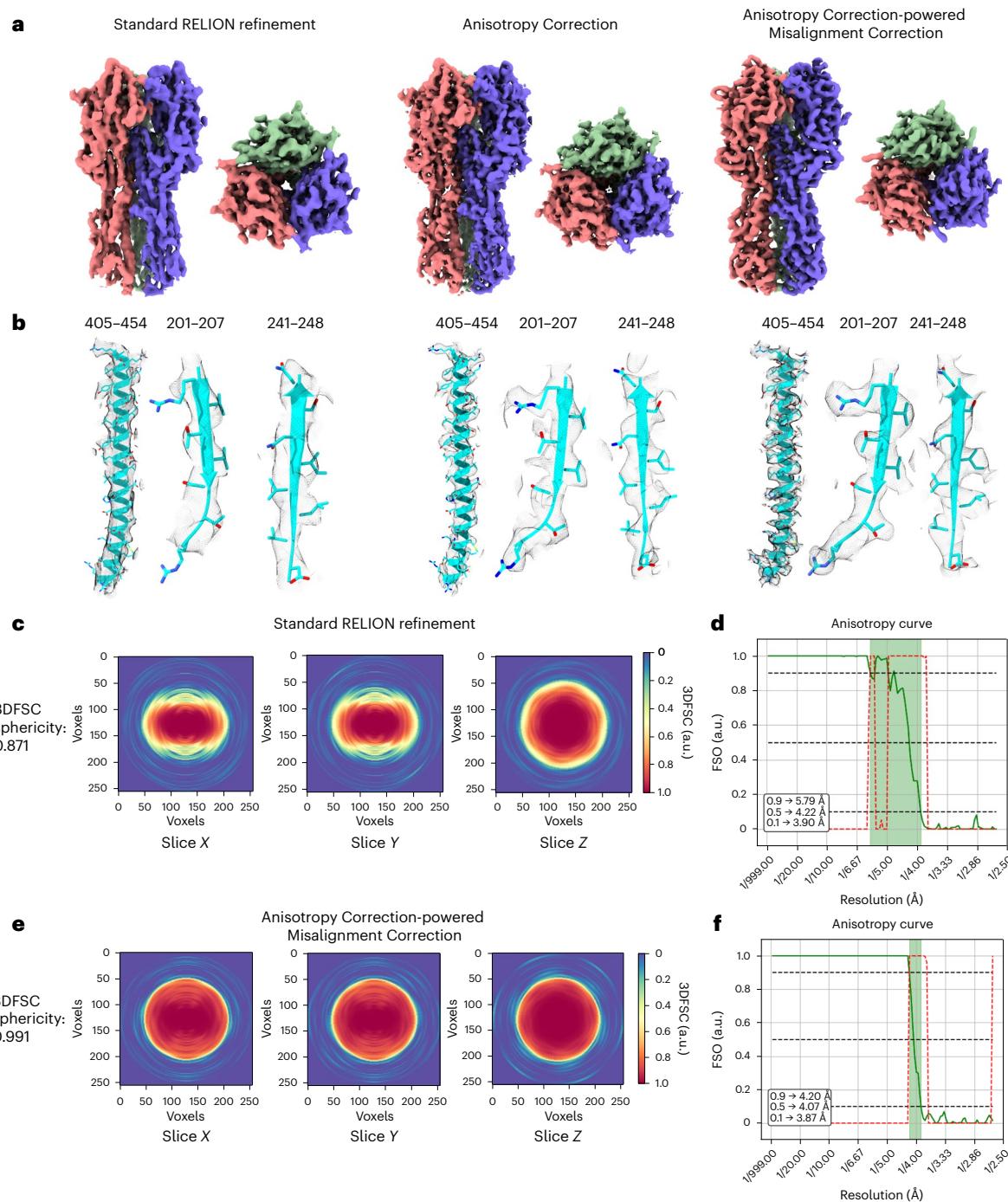


Fig. 3 | Application of splIsoNet to the tilted cryo-EM influenza HA trimer datasets. **a**, cryo-EM map of the HA trimer reconstructed with different methods. From left to right: standard RELION Refinement, Anisotropy Correction, and Anisotropy correction-powered Misalignment Correction. **b**, Representative density regions of the HA trimer structure with the HA atomic model (cyan) fitted. The residue numbers are indicated. From left to right: standard Relion Refinement, Anisotropy Correction, and Anisotropy correction-powered

Misalignment Correction. **c**, Central slices along the *X*, *Y* and *Z* directions of the 3DFSC for the standard RELION Refinement. **d**, FSO (green line) and *P* value for the Bingham test (red dashed line) computed from the standard RELION Refinement result. a.u., arbitrary units. **e**, Central slices along the *X*, *Y* and *Z* directions of the 3DFSC for the splIsoNet misalignment-corrected result. **f**, FSO (green line) and *P* value for the Bingham test (red dashed line) computed from the splIsoNet misalignment corrected result. a.u., arbitrary units.

splIsoNet improves isotropic resolution in subtomogram averaging

Cryo-ET is a method of choice for obtaining *in situ* biological structures by combining views in a tilt series of the same specimen area⁴⁸. Bio-macromolecules are often arranged in specific orientations in cells, rather than being randomly oriented⁴⁹, giving rise to preferred-orientation problems *in situ* and distorted structures even

after subtomogram averaging. We explored the performance of splIsoNet in subtomogram averaging, using the HIV-1 immature capsid dataset (EMPIAR-10164)⁵⁰. Using a subset of five tilt-series, the standard workflow in RELION4 (ref. 51) yields a 3.7-Å-resolution structure without performing CTF refinement and frame alignment (Fig. 6a and Extended Data Fig. 10a). With Misalignment Correction, we obtained an isotropic 3.6-Å resolution structure (Fig. 6e and Extended Data Fig. 10a).

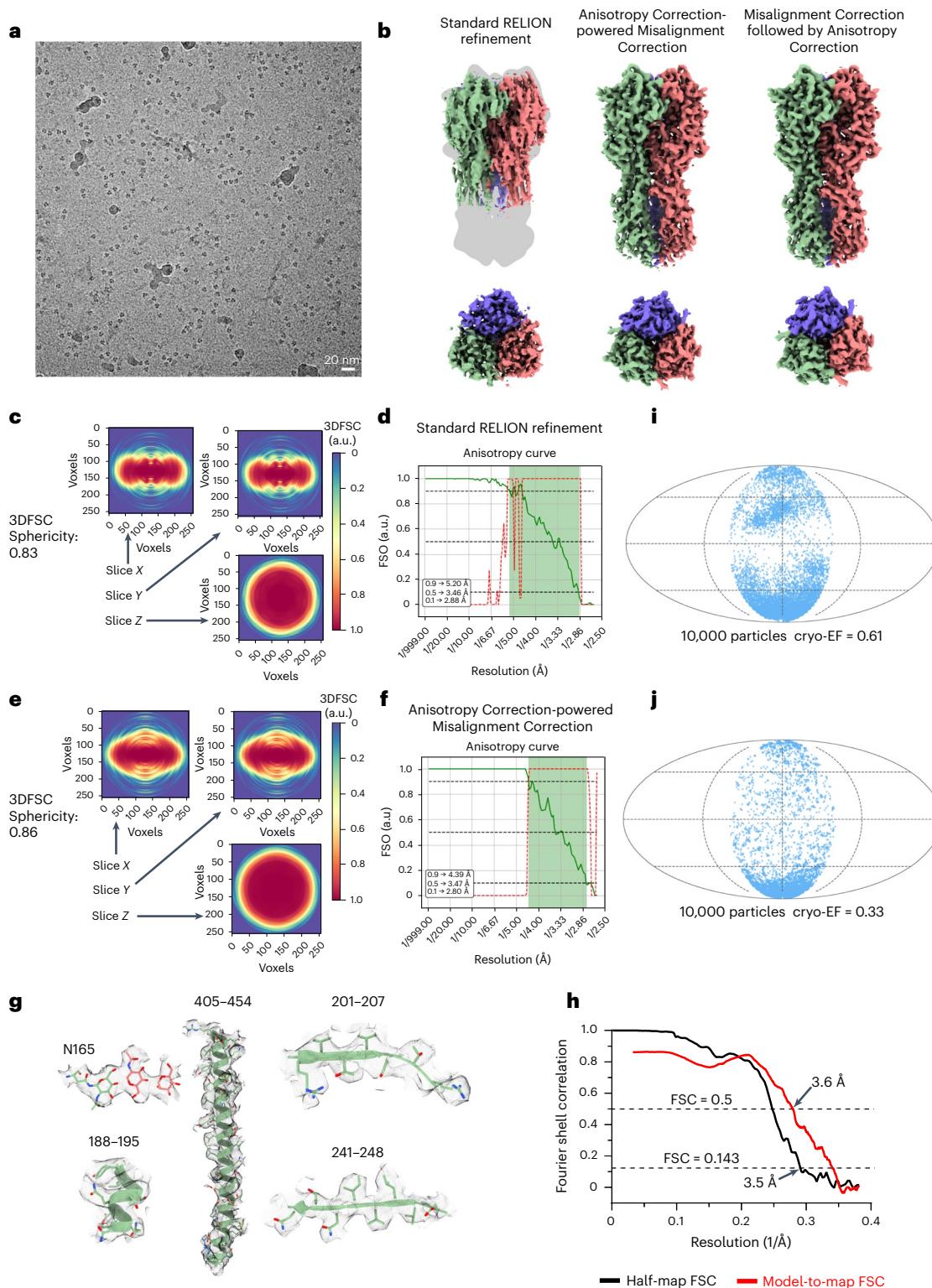


Fig. 4 | Application of splsoNet to the untilted cryo-EM influenza hemagglutinin trimer datasets. **a**, Representative cryo-EM micrograph of the untilted cryo-EM influenza HA trimer dataset. Similar results are observed in 447 micrographs from [EMPIAR-10096](#). **b**, cryo-EM map of the HA trimer reconstructed from different methods. From left to right: standard RELION Refinement, splsoNet Anisotropy Correction-powered Misalignment Correction, and both Misalignment Correction and Anisotropy Correction. **c**, Central slices along the X, Y, and Z directions of the 3DFSC for the standard RELION Refinement. **d**, FSO (green line) and Pvalue for the Bingham test (red dashed line) computed from the standard RELION refinement result. **e**, Central slices along the X, Y, and Z directions of the 3DFSC for the splsoNet misalignment corrected result. **f**, FSO

(green line) and Pvalue for the Bingham test (red dashed line) computed from the splsoNet Anisotropy Correction-powered Misalignment Correction result. **g**, Representative density of selected amino acid residues and glycans (denoted with the residue numbers) from the misalignment corrected HA trimer cryo-EM map docked with the relevant atomic models. **h**, FSC curves of the misalignment-corrected HA trimer map. Black, half-map FSC curve with criterion of 0.143; red, FSC curve calculated between the cryo-EM map and the refined structure model with criterion of 0.5. **i,j**, Orientation distribution results and corresponding cryo-EF scores from a standard RELION refinement (**i**) and from splsoNet Anisotropy Correction-powered Misalignment Correction module (**j**). a.u., arbitrary units.

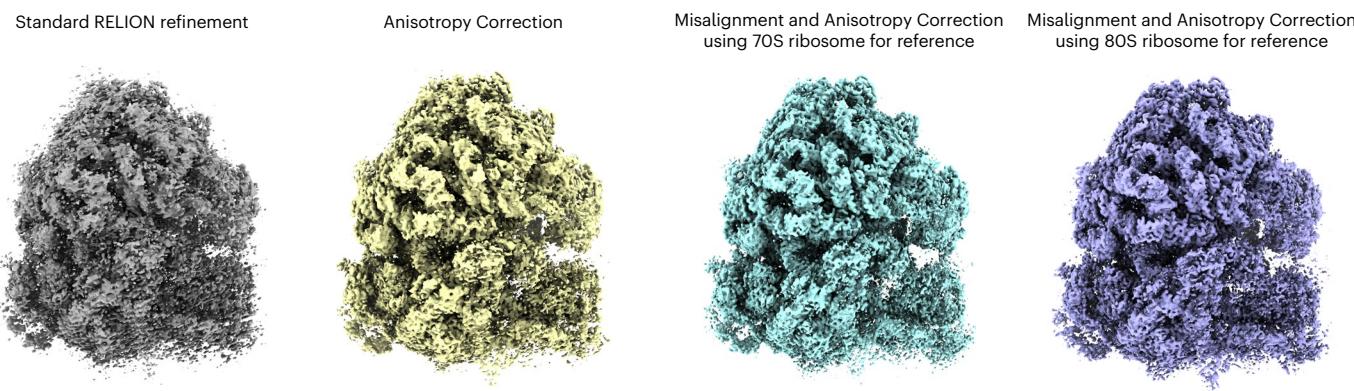
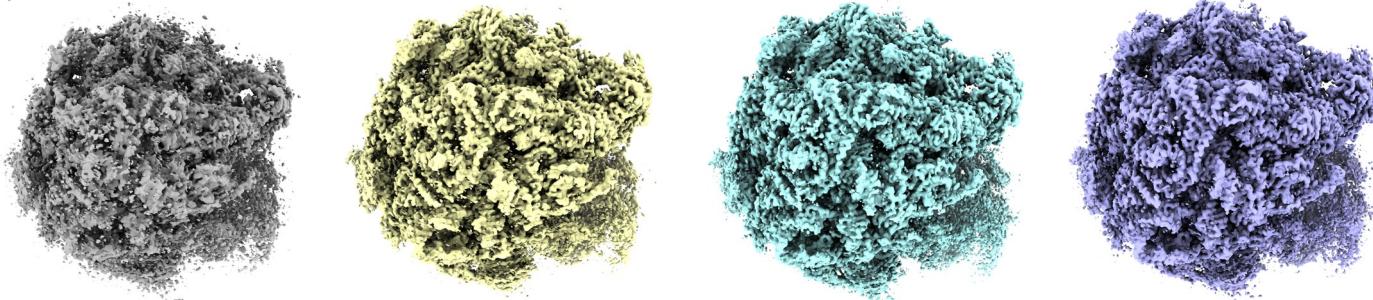
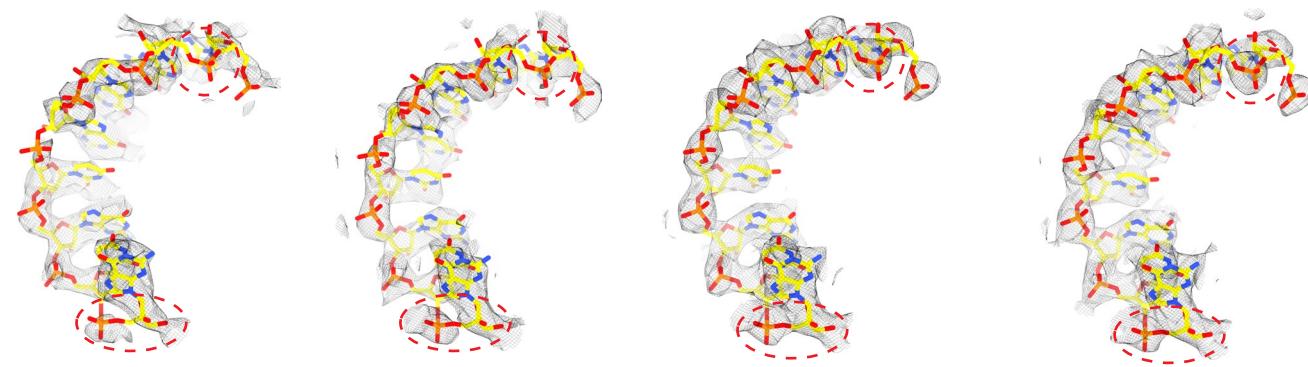
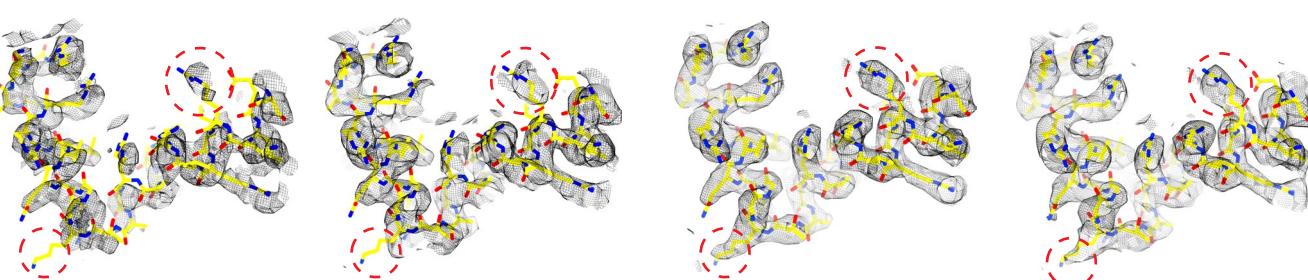
a**b****c****d**

Fig. 5 | Application of splIsoNet to the ribosome dataset. **a,b**, Ribosome maps reconstructed from different reconstruction methods. From left to right: standard RELION Refinement, splIsoNet Anisotropy Correction, Misalignment Correction followed by Anisotropy correction using 70S ribosome as reference,

and Misalignment followed by Anisotropy Correction using 80S ribosome map as reference. **c,d**, Representative density regions with the fitted atomic model (yellow).

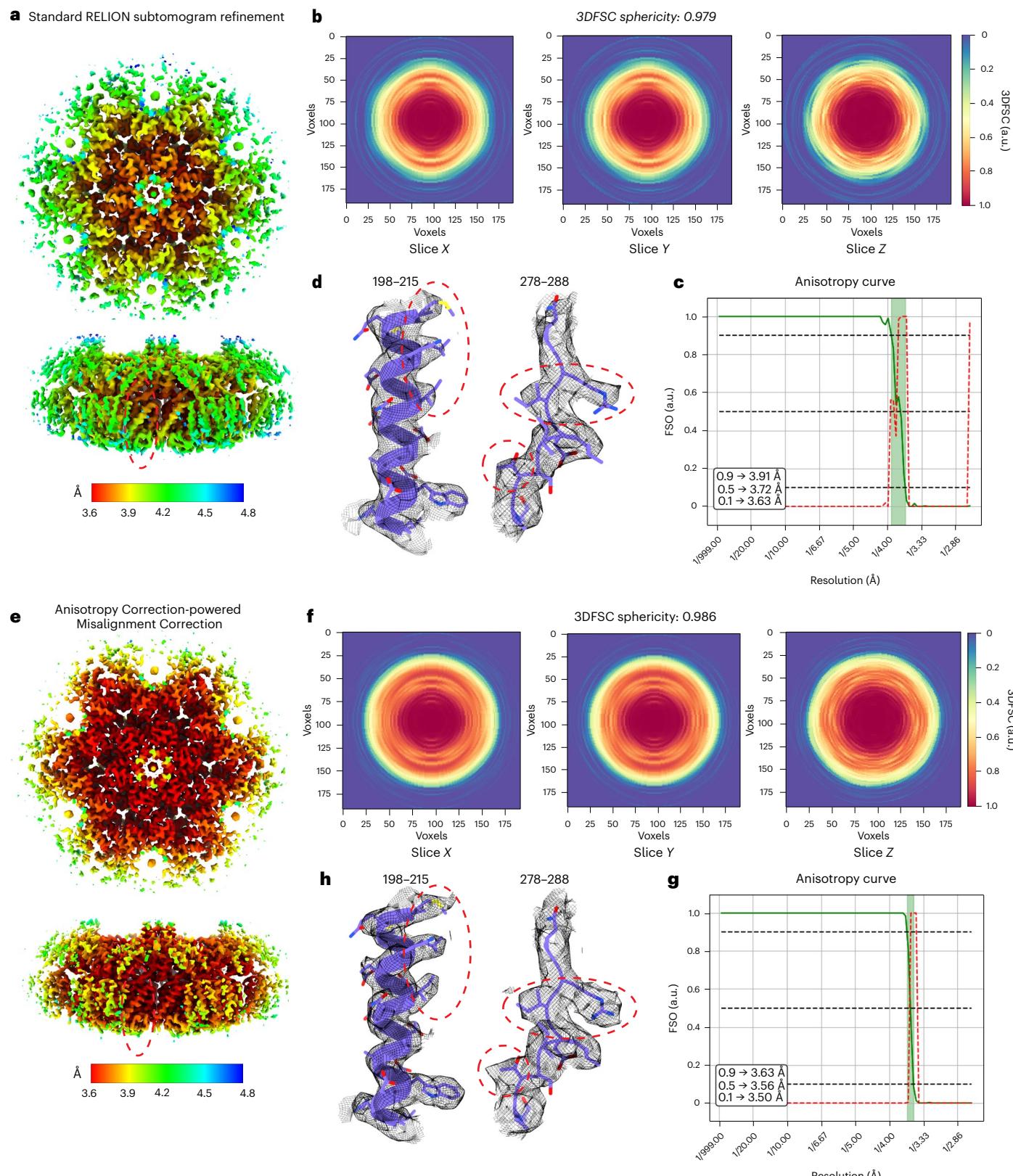


Fig. 6 | Application of splIsoNet to subtomogram averaging. **a**, Local-resolution map of the HIV-1 immature capsid reconstructed from standard RELION subtomogram-averaging. **b**, Central slices along the X, Y and Z directions of the 3DFSC for the standard RELION subtomogram averaging. **c**, FSO (green line) and P value for the BT (red dashed line) computed from the standard RELION4 subtomogram averaging pipeline. **d**, Representative density of amino acid residues (denoted with the residue numbers) docked with the atomic models.

e, Local-resolution map of the HIV-1 immature capsid reconstructed from splIsoNet Anisotropy Correction-powered Misalignment Correction. **f**, Central slices along the X, Y, and Z directions of the 3DFSC for the splIsoNet Anisotropy Correction-powered Misalignment Correction. **g**, FSO (green line) and P value for the Bingham test (red dashed line), computed from the splIsoNet result. **h**, Representative density of selected amino acid residues (denoted with the residue numbers) fitted with the atomic models. a.u., arbitrary units.

Despite only marginal improvement in the overall resolution, the corrected map displayed improvements across various metrics, including half-map FSC, model-to-map FSC, and local resolution (Fig. 6a,e and Extended Data Fig. 10). The structure revealed better-resolved side chain densities (Fig. 6d,h) and exhibited increased 3DFSC sphericity and reduced resolution anisotropy regions within the FSO curve (Fig. 6b,c,f,g). These improvements can be attributed not only to the information retrieval, but also to the denoising effect, which should improve the accuracy of particle-orientation estimation. Thus, splIsoNet can be used for *in situ* structural biology to improve map quality in subtomogram averaging.

Discussion

In contrast to the numerous previous chemical and physical methods to alleviate the seemingly ubiquitous problem of preferred orientation in cryo-EM, splIsoNet now provides a computational method to address challenges in 3D reconstruction caused by the preferred-orientation problem. Although the two modules in splIsoNet can be executed independently, using the Misalignment Correction module followed by the Anisotropy Correction module typically yields optimal results. For particles with severe preferential orientation, the Misalignment Correction module has an option of preserving low-resolution information from the initial reference map to ensure alignment accuracy. The reference map can be obtained from various methods, such as reconstruction from small-particle datasets with stage tilt or subtomogram averaging from small tomography datasets. Regardless of which splIsoNet workflows users choose to solve their specific preferred-orientation problem, the pure computational nature of the workflow should reduce the time from imaging to structure interpretation, allowing investigators to focus on scientific discovery rather than specimen or grid manipulation, which is required for other methods.

Of note, splIsoNet differs from IsoNet⁵² in both applications and algorithms. Self-supervised information recovery was also implemented in the original IsoNet package, which was designed to overcome the missing-wedge problem in electron tomography⁵². In this approach, the network is trained with rotated subtomograms as ground truth and rotated, missing-wedge-imposed subtomograms as input. As detailed above, owing to the complexity of preferred-orientation problem in cryo-EM, here we used a four-loss end-to-end implementation with 3DFSC as a reliability measure in splIsoNet, differing from the iterative one-loss design of IsoNet⁵².

Previous deep learning methods, such as deepEMhancer⁵³ and EMReady⁵⁴, have attempted to enhance the interpretability of cryo-EM maps by training a network using maps simulated from the PDB as ground truth and cryo-EM maps in the EMDB as input. Although these approaches have shown promising results, there are concerns regarding potential biases introduced by relying on knowledge from other structures, which could complicate the interpretation of neural-network-generated cryo-EM maps. By contrast, splIsoNet exploits the rotation equivariance present in the biological structure with self-supervised learning. In essence, a cryo-EM map contains molecular information distributed across a 3D volume, encompassing recurring information, ranging from large structural features, such as α -helices and β -sheets, to small structural features in individual amino acids. By accounting for both consistency and equivariance loss in the network training process, the network learns how to recover information, leveraging the fact that characteristic features should be similar regardless of their positions and orientations, and ensures that the recovered molecular information will go to regions in which information is lacking. The ability of splIsoNet to capture complex patterns and relationships solely from user-provided cryo-EM maps allows it to achieve faithful representations of reality with incomplete experimental data.

In single-particle cryo-EM, the accuracy of reconstruction hinges on incorporation of reliable prior knowledge into the cryo-EM

structure-determination pipeline. From smoothness priors in RELION⁵, to the independent half-map noise assumptions^{41,55}, to low-dimensional manifold of conformation change⁵⁶, to the isotropic prior in splIsoNet, reliable prior knowledge is incorporated into the reconstruction process. Such integration allows for 3D reconstructions from fewer particles and fewer orientation views, enhancing the accuracy, efficiency, and investigative power of cryo-EM and cryo-ET in biological inquiries.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02505-1>.

References

- Ruskin, A. I., Yu, Z. & Grigorieff, N. Quantitative characterization of electron detectors for transmission electron microscopy. *J. Struct. Biol.* **184**, 385–393 (2013).
- Obr, M. et al. Exploring high-resolution cryo-ET and subtomogram averaging capabilities of contemporary DEDs. *J. Struct. Biol.* **214**, 107852 (2022).
- Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
- Nakane, T. et al. Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat. Methods* **18**, 186–193 (2021).
- Drulyte, I. et al. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr. D Struct. Biol.* **74**, 560–571 (2018).
- Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
- Fan, H. & Sun, F. Developing graphene grids for cryoelectron microscopy. *Front Mol. Biosci.* **9**, 937253 (2022).
- Kang, J. S., Zhou, X., Liu, Y. T., Wang, K. & Zhou, Z. H. Theoretical framework and experimental solution for the air-water interface adsorption problem in cryoEM. *Biophys. Rep.* **9**, 215–229 (2023).
- Noble, A. J. et al. Routine single particle CryoEM sample and grid characterization by tomography. *eLife* **7**, e34257 (2018).
- Li, B., Zhu, D., Shi, H. & Zhang, X. Effect of charge on protein preferred orientation at the air-water interface in cryo-electron microscopy. *J. Struct. Biol.* **213**, 107783 (2021).
- Liu, N. & Wang, H. W. Better Cryo-EM specimen preparation: how to deal with the air–water interface? *J. Mol. Biol.* **435**, 167926 (2023).
- Taylor, K. A. & Glaeser, R. M. Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. *J. Struct. Biol.* **163**, 214–223 (2008).
- Glaeser, R. M., Nogales, E. & Chiu, W. *Single-particle Cryo-EM of Biological Macromolecules* (IOP publishing, 2021).

18. Chen, J., Noble, A. J., Kang, J. Y. & Darst, S. A. Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J. Struct. Biol.* **X**, 100005 (2019).
19. Zhang, Z., Shigematsu, H., Shimizu, T. & Ohto, U. Improving particle quality in cryo-EM analysis using a PEGylation method. *Structure* **29**, 1192–1199 (2021).
20. Grassucci, R. A., Taylor, D. J. & Frank, J. Preparation of macromolecular complexes for cryo-electron microscopy. *Nat. Protoc.* **2**, 3239–3246 (2007).
21. Han, B. G. et al. Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. *J. Struct. Biol.* **195**, 238–244 (2016).
22. Fan, H. et al. A cryo-electron microscopy support film formed by 2D crystals of hydrophobin HFBI. *Nat. Commun.* **12**, 7257 (2021).
23. Naydenova, K., Peet, M. J. & Russo, C. J. Multifunctional graphene supports for electron cryomicroscopy. *Proc. Natl Acad. Sci. USA* **116**, 11718–11724 (2019).
24. Lu, Y. et al. Functionalized graphene grids with various charges for single-particle cryo-EM. *Nat. Commun.* **13**, 6718 (2022).
25. Noble, A. J. et al. Reducing effects of particle adsorption to the air-water interface in cryo-EM. *Nat. Methods* **15**, 793–795 (2018).
26. Ravelli, R. B. G. et al. Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nat. Commun.* **11**, 2563 (2020).
27. Rima, L. et al. cryoWriter: a blotting free cryo-EM preparation system with a climate jet and cover-slip injector. *Faraday Discuss.* **240**, 55–66 (2022).
28. Torino, S., Dhurandhar, M., Stroobants, A., Claessens, R. & Efremov, R. G. Time-resolved cryo-EM using a combination of droplet microfluidics with on-demand jetting. *Nat. Methods* **20**, 1400–1408 (2023).
29. Klebl, D. P., Kay, R. W., Sobott, F., Kapur, N. & Muench, S. P. Towards sub-millisecond cryo-EM grid preparation. *Faraday Discuss.* **240**, 33–43 (2022).
30. Huntington, B. et al. Thicker ice improves the integrity and angular distribution of CDC48A hexamers on cryo-EM grids. *Front. Mol. Biosci.* **9**, 890390 (2022).
31. Aiyer, S. et al. Overcoming resolution attenuation during tilted cryo-EM data collection. *Nat. Commun.* **15**, 389 (2024).
32. Frank, J. *Three-dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State* (Oxford University Press, 2006).
33. Sorzano, C. O. S. et al. Algorithmic robustness to preferred orientations in single particle analysis by CryoEM. *J. Struct. Biol.* **213**, 107695 (2021).
34. Hui, Z. et al. Addressing preferred orientation in single-particle cryo-EM through AI-generated auxiliary particles. Preprint at bioRxiv <https://doi.org/10.1101/2023.09.26.559492> (2023).
35. Ronneberger, O., Fischer, P. & Brox, T. U-net convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-assisted Intervention—miccai 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18 234–241 (Springer, 2015).
36. Chen, D., Tachella, J. & Davies, M. E. Equivariant imaging: learning beyond the range space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 4379–4388 (IEEE, 2021).
37. Lehtinen, J. et al. Noise2Noise: learning image restoration without clean data. *Proc. 35th Int. Conf. Mach. Learn.* **80**, 2965–2974 (2018).
38. Buchholz, T.-O., Jordan, M., Pigno, G. & Jug, F. Cryo-CARE: Content-aware image restoration for cryo-transmission electron microscopy data. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 502–506 (IEEE, 2019).
39. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat. Commun.* **11**, 5208 (2020).
40. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
41. Kimanius, D. et al. Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nat. Methods* **21**, 1216–1221 (2024).
42. Fromm, S. A. et al. The translating bacterial ribosome at 1.55 Å resolution generated by cryo-EM imaging services. *Nat. Commun.* **14**, 1095 (2023).
43. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
44. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
45. Vilas, J. L. & Tagare, H. D. New measures of anisotropy of cryo-EM maps. *Nat. Methods* **20**, 1021–1024 (2023).
46. Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun.* **8**, 629 (2017).
47. Nicholson, D., Edwards, T. A., O'Neill, A. J. & Ranson, N. A. Structure of the 70S ribosome from the human pathogen *Acinetobacter baumannii* in complex with clinically relevant antibiotics. *Structure* **28**, 1087–1100 (2020).
48. Wan, W. & Briggs, J. A. Cryo-electron tomography and subtomogram averaging. *Methods Enzymol.* **579**, 329–367 (2016).
49. Liu, Y. T. et al. Mesophasic organization of GABA(A) receptors in hippocampal inhibitory synapses. *Nat. Neurosci.* **23**, 1589–1596 (2020).
50. Schur, F. K. et al. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506–508 (2016).
51. Zivanov, J. et al. A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife* **11**, e83724 (2022).
52. Liu, Y. T. et al. Isotropic reconstruction for electron tomography with deep learning. *Nat. Commun.* **13**, 6482 (2022).
53. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
54. He, J., Li, T. & Huang, S. Y. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nat. Commun.* **14**, 3217 (2023).
55. Ramlau, K., Palmer, C. M., Nakane, T. & Aylett, C. H. S. Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER. *J. Struct. Biol.* **211**, 107545 (2020).
56. Frank, J. & Ourmazd, A. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **100**, 61–67 (2016).
57. Kim, J. et al. Structure and drug resistance of the *Plasmodium falciparum* transporter PfCRT. *Nature* **576**, 315–320 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

Methods

Software implementation

We implemented splIsoNet in Python using Linux as the native operating system. The package can be downloaded from Github (<https://github.com/IsoNet-cryoET/splIsoNet>). A detailed document is provided, accompanied by the splIsoNet package.

Overview of Anisotropy Correction module

The input of splIsoNet's Anisotropy Correction are two unfiltered half-maps and a soft solvent mask. It first calculates 3DFSC volume using the two half-maps to represent the anisotropic resolution for the reconstruction. Then, a 'reconstruct' step is performed to train a network to recover the missing information quantified based on 3DFSC¹⁰. This 'reconstruct' step, performed on the half-maps, generates a neural network and the anisotropy-corrected half-maps (Fig. 1c and Extended Data Figs. 1 and 2). The subsequent post-processing of the half-maps for sharpening and gold-standard FSC determination can be performed in other cryo-EM software, such as RELION⁵.

Fast 3DFSC implementation

3DFSC¹⁰ is an algorithm that quantitatively assesses directional resolution anisotropy of density maps. The algorithm compiles a set of one-dimensional FSC curves computed over distinct angular directions into an 3D array, which is visualized as a 3D density volume.

For user convenience and to make our program standalone, we implemented the 3DFSC algorithm in the splIsoNet natively. We used a cKDTree algorithm⁵⁸ to increase the speed of processing. It typically runs for 30 seconds for maps with 400³ voxels on 16 CPU cores.

In the following sessions, a 3D volume (x) with the 3DFSC filter applied is denoted as Ax , which means applying this filter in Fourier space, to mimic the preferred-orientation-induced artifacts on map x :

$$Ax := iFFT(3DFSC \cdot FFT(x))$$

Where FFT and iFFT are 3D Fourier transform and 3D inverse Fourier transform respectively.

The neural network architecture in splIsoNet

The neural network used in splIsoNet is based on U-net³⁵ (Fig. 1b), which is well recognized in biomedical image restoration and segmentation. The main building blocks of our U-net implementation are 3D convolution layers with non-linear activation functions called leaky rectified linear units, which are applied per voxel. Those convolution layers have kernel sizes of 3 × 3 × 3. Three 3D convolution layers are stacked together to form a convolution block in our network, which can extract features in cryo-EM maps. By stacking the convolutional blocks, the U-net is built on the basis of encoder-decoder architecture (Fig. 1b). The encoder path is a set of convolution blocks and strided convolution layers that compress 3D volumes. Strided convolution layers reduce the spatial size of the input of this layer by 2 × 2 × 2, allowing the network to learn more abstract information. One encoder block is composed of a convolution block followed by a strided convolution layer. Three encoder blocks constitute the entire encoding path. After the encoder path, the 3D volumes are processed with a convolution block and enter the decoder path of the network. The decoder path is symmetrical to the encoder but uses transpose convolution layers, opposite to strided convolution layers, to enlarge the dimension of features. Although the down-sampling of the 3D volumes captures the essence of the features, high-resolution information is lost by stride convolution operations. In U-net, skip-connections that concatenate the feature layers of the same dimension in two paths are implemented to preserve high-resolution information.

In the following sessions, the network is represented as $f_\theta()$, which captures the neural network as f with parameter θ , which

represents network weights and biases that are updated throughout the training process.

Network training for Anisotropy Correction

Firstly, voxel values of input half-maps are normalized to a s.d. of 1 and mean of 0. By default, 1,000 randomly distributed 3D coordinates are randomly generated within the regions defined by a user-provided mask. Then, cubic volumes (default 64³ voxels), centered at the generated coordinates, are saved as submaps. The network is trained to minimize a loss function. The loss function is minimized by employing Adam optimizer with an initial learning rate of 0.0004. The neural network training is performed on one or multiple GPUs and typically consists of 30 epochs.

The loss function consists of four terms: two for information recovery and two for noise2noise based denoising.

The first term is the data consistency loss:

$$Lc = \|Af_\theta(x^0) - x^0\|_1$$

In this formula, x^0 represents the input submaps in one half-map, A denotes the 3DFSC filter described above, and f_θ represents the neural network parameterized by θ . This term in this expression enforces data consistency, ensuring that the application of the 3DFSC filter to the predicted map results in the raw map.

The second term is the rotation equivariance loss:

$$Le = \|f_\theta(ATf_\theta(x^0)) - Tf_\theta(x^0)\|_1$$

This loss takes the rotated map and tends to recover information from the removed information in the rotated maps. T stands for random rotation operator. To calculate the rotation equivariance loss, we create a rotated and corrected submap (R map, that is $Tf_\theta(x)$) by applying the network to an input submap and randomly rotating the resulting map. After applying the 3DFSC filter to the R map, we pass it through the network again for recovery. The rotation equivariance terms minimize the difference between the recovered map and the original R map.

The third and the fourth terms are similar data consistency loss and rotation equivariance loss, but comparing between the submaps from half-map1 and half-map2 for the noise2noise based denoising:

$$Lc_{n2n} = \|Af_\theta(x^0) - x^1\|_1$$

$$Le_{n2n} = \|f_\theta(ATf_\theta(x^0)) - Tf_\theta(x^1)\|_1$$

Where in these terms x^0 and x^1 are submaps from different half-maps, but at the same location. Noise will be reduced using these two terms which benefits splIsoNet information recovery.

The final loss is represented as:

$$L = Lc + \alpha Le + \beta(Lc_{n2n} + \alpha Le_{n2n})$$

Where α and β are hyperparameters. The α is balancing between data consistency and rotation equivariance, and beta defines how much level of denoising should be considered in training. By default, α and β values equal to 1 and 0.5, respectively.

After the last epoch of training, The splIsoNet program splits the half-maps into smaller 3D chunks and applies the network to them separately. Then, output 3D chunks are stitched to produce the final output. To avoid the line artifact between adjacent chunks caused by the loss of information on the edges of subtomograms, we implemented a seamless reconstruction method called the overlap-tile strategy³⁵, which predicts the overlapping chunks to avoid the edge effect. The output-corrected two half-maps, as well as the trained neural network, are saved for further post-processing.

Anisotropy Correction-powered Misalignment Correction module

In the Anisotropy Correction-powered Misalignment Correction module, we integrate splsoNet in RELION's 3D refinement pipeline. Taking the advantage of ‘–external_reconstruct’ function in `relion_refine`, the `relion_wrapper.py` in splsoNet will be executed after each iteration of RELION refinement when RELION refinement reaches a finer angular sampling, that is ‘healpix_order’ becomes 3.

After each RELION iteration, the splsoNet is executed in four steps: spectral whitening, FSC weighting, anisotropy correction, and low-pass filtering. Spectral whitening equalized the Fourier power of half-maps for all the spectral frequency from 10 Å up to the resolution determined by $FSC = 0.143$ criterion, similar to map whitening implemented in *cisTEM*⁴³. Those maps are then processed by FSC weighting⁴⁴. Because the FSC weighting is performed only on the half-maps, the $\sqrt{\frac{2FSC}{1+FSC}}$ weighting proposed in Rosenthal and Henderson is not performed⁴⁴, instead the half-maps are directly filtered with the FSC. After the FSC weighting, map Anisotropy Correction is performed for those half-maps. Those maps are then low-pass filtered to a resolution of $FSC = 0.143$ and serve as references for the subsequent iterations. Users can choose not to perform either spectral whitening, FSC weighting, or low-pass filtering, if necessary.

In some cases, owing to the severe preferred-orientation problems, RELION fails to perform reliable image alignment and to generate relatively correct initial maps. As a result, we provided an option called ‘–keep_lowres’. This option allows the low-resolution information of the reference maps to be preserved during subsequent RELION refinement iterations. Given a relatively uniformly sampled reference at about 10–20 Å resolution, RELION will be less likely to perform erroneous image alignment. To avoid differences in the intensity between the reference map and the map reconstructed from the particle dataset that could cause errors in alignment, we designed a function called spectral matching that enforces the spectral power distribution of the reference maps the same as the map reconstructed using the particle datasets in each RELION iteration. This ‘–keep_lowres’ is performed before the whitening step, when specified.

Benchmark with simulated ribosome map

To benchmark splsoNet’s performance with an asymmetric complex, we downloaded the structure of the a translating 70S ribosome (PDB: 8BOX)⁴² and generated a density map with a pixel size of 1 Å and a box size of $400 \times 400 \times 400$ from the atomic model. We then applied a missing-cone filter that removed orientational information in Fourier space up to certain degrees to create three maps with missing cones (Fig. 2a,d,g), mimicking different degrees of map anisotropy conditions in practice. Because we have only one map for each simulation, we cannot calculate 3DFSC. Instead, we used missing-cone volumes as a quantification of the missing information in Fourier space. The splsoNet α parameter is set to 1. Also, because there is only one map, the β parameter is not applicable. A total of 30 epochs are performed to correct each simulated map.

Benchmark with β-galactosidase datasets

We also tested splsoNet with the β-galactosidase dataset, which was used as the RELION single-particle tutorial dataset. To validate whether splsoNet can improve the map quality, which is affected by varying views of preferred orientation, we created two datasets with preferred orientation by selecting side-view particles (1,513 particles) and top-view particles (950 particles) from the 2D class averages (Fig. 2j–m).

We conducted standard RELION 3D reconstruction to generate two anisotropic maps, as depicted in (Figs. 2k,m). Subsequently, we applied map Anisotropy Correction to obtain the corrected maps, as shown in (Figs. 2k,m). The Anisotropy Correction was executed with parameters set to $\alpha = 1$ and $\beta = 0.5$, limiting information recovery resolution up to

3.5 Å and running for 30 epochs. Both side-view dominant and top-view dominant maps corrected by Anisotropy Correction show improved map quality and reduced density smearing and elongation on the missing-information direction.

Benchmark with the HA-trimer tilted datasets

We first downloaded the 4.2-Å-resolution HA trimer cryo-EM map (EMDB-8731)¹⁰ and performed Anisotropy Correction directly. The Anisotropy Correction parameters are $\alpha = 1$ and $\beta = 0.5$, with limit information recovery resolution to 3.5 Å. After Anisotropy Correction, post-processing is performed using RELION with default parameters for structure comparison.

The EMPIAR-10097 HA-trimer dataset was collected using a grid-tilting strategy¹⁰. We imported 130,000 particles into RELION4 and performed Misalignment Correction, and the healpix-order became 3. We set the splsoNet parameters as follows: $\alpha = 1$, $\beta = 0.5$, and epochs = 5. We retrained the splsoNet model with every RELION4 refinement iteration using the RELION-generated half-maps. We finally obtained the 4.1-Å-resolution map as measured by gold-standard FSC. The same particles were processed by homogeneous refinement in cryoSPARC with default settings or RELION5 Blush regularization to generate respective reconstructed maps.

Benchmark with HA-trimer non-tilted dataset

The EMPIAR-10096 cryo-EM HA-trimer dataset was collected without grid tilting¹⁰. We imported 130,000 particles into RELION4 and performed the Misalignment Correction using ‘–external_reconstruction’ and ‘–keep_lowres’ command when the healpix-order became larger than 3. The ‘–keep_lowres’ command allows us to incorporate the low-resolution initial map into each round of RELION4 reconstruction to enhance the initial 3D alignment accuracy. The initial reference is the map reconstructed from the tilt dataset (Fig. 3a, right panel) and filtered to 10-Å resolution. The splsoNet Misalignment Correction parameters were set as follows: $\alpha = 1$, $\beta = 0.5$, and epochs = 5. We retrained the splsoNet model with every RELION4 refinement iteration. After RELION refinement, we suspect that there are some bad particles. Thus, we performed 3D classification without alignment for four classes. We selected two of the resulting classes with a total of 85,358 particles for another round of the Misalignment Correction with the same parameter as before. We finally obtained a 3.45-Å-resolution map. We further performed the Anisotropy Correction to improve the cryo-EM map. The parameters of Anisotropy Correction are $\alpha = 1$, $\beta = 0.5$, and epochs = 30 and limit recovery resolution to 3.5 Å. The particles were also processed by homogeneous refinement in cryoSPARC with default settings or RELION5 Blush regularization to generate respective reconstructed maps.

Benchmark with asymmetric ribosome dataset

This dataset (EMPIAR-10406) contains the 70S ribosome from the human pathogen *A. baumannii* in complex with amikacin⁴⁷. We manually created a preferred-orientation datasets by selecting the particles with specific ranges of Euler angles (Extended Data Fig. 8a): rot angles (-20° – 140°) and tilt angles (100° – 160°). We followed the standard single particle RELION protocol to perform the 3D auto-refinement using 23,177 particles. We then executed the Misalignment Correction using splsoNet with ‘–external_reconstruction’ and ‘–keep_lowres’ commands when the healpix-order became 3. The initial resolution for the reference is set to 15 Å, and the information lower than this resolution is maintained through the intermediate RELION refinement process. The splsoNet parameters were set as follows: $\alpha = 1$, $\beta = 0.5$, and epochs = 5. Neural networks in splsoNet are retrained after every RELION refinement iteration to adapt to the updated half-maps. To test the effect of incorporating the low-resolution information from the initial reference map, we chose two different subtomogram-averaging 3D cryo-EM maps for comparison: one is the 70S ribosome structure from

Escherichia coli (EMDB-16139)⁵⁹ and another is the 80S ribosome structure from the *Saccharomyces cerevisiae*.

To generate a reference map from the 80S ribosome, we reprocessed the EMPIAR-10045 dataset⁶⁰ as follows. The *S. cerevisiae* 80S ribosome tilt-series were aligned using IMOD⁶¹, and the defocus values for each tilt were estimated using GCTF⁶². We picked 3,288 ribosomal particles using TomoTwin software⁶³ and performed the subtomogram averaging using RELION4 (ref. 51). We followed the standard RELION4 subtomogram refinement procedure and achieved a 7.23-Å resolution for the yeast 80S ribosome. After Misalignment Correction, the maps were then subjected to Anisotropy Correction with default parameters: $\alpha = 1$, $\beta = 0.5$, epochs = 30, and limit_res = 3.5 Å.

Benchmark with HIV VLP tomography dataset

This dataset (EMPIAR-10164) contains the immature HIV-1 dMACANC virus-like particles (VLPs)⁵⁰, which are also in the RELION4 subtomogram averaging tutorial dataset⁵¹. We followed the standard RELION4 refinement procedure according to the tutorial and achieved a 3.7-Å-resolution map. We then used the same RELION4 parameters and executed the Misalignment Correction with ‘–external_reconstruction’ when the healpix-order became 4. The splIsoNet parameters were set as follows: $\alpha = 1$, $\beta = 0.5$, epochs = 20. We retrained the splIsoNet model with every RELION refinement iteration. We finally obtained a 3.6-Å-resolution map. We used the unsharpened maps for comparison between the standard RELION refinement and Misalignment Correction.

3D visualization

IMOD⁶¹ was used to visualize 2D slices of cryo-EM density maps. UCSF ChimeraX⁶⁴ was used to visualize the reconstructed cryo-EM density maps in three dimensions. The 3DFSC and FSO were calculated within the Scipion software⁶⁵. The local-resolution maps were calculated by ResMap⁶⁶, and local-resolution estimation was done in RELION⁵.

Statistics and reproducibility

No statistical methods were used to predetermine sample size. We confirm that all results in this study can be reproduced using publicly available datasets and our standard protocol with the splIsoNet software.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The tutorial dataset for Anisotropy Correction and Misalignment Correction can be downloaded from Zenodo: <https://zenodo.org/records/12640059> (ref. 67). All data that support this study are available from the corresponding authors upon request. The cryo-EM density maps of the HA trimer from EMPIAR-10096 and EMPIAR-10097 have been deposited in the EMDB under the accession numbers EMD-45551 and EMD-45555, respectively.

Code availability

The splIsoNet is available at the GitHub under MIT license (<https://github.com/IsoNet-cryoET/splIsoNet>), and can be downloaded from Zenodo: <https://zenodo.org/records/12640059> (ref. 67).

References

58. Maneewongvatana, S. & Mount, D. M. in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 105–123 (American Mathematical Society, 1999).
59. Khavnekar, S. et al. Multishot tomography for high-resolution *in situ* subtomogram averaging. *J. Struct. Biol.* **215**, 107911 (2023).
60. Bharat, T. A. & Scheres, S. H. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protoc.* **11**, 2054–2065 (2016).
61. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
62. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
63. Rice, G. et al. TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining. *Nat. Methods* **20**, 871–880 (2023).
64. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
65. de la Rosa-Trevin, J. M. et al. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).
66. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
67. Liu, Y. T. splIsoNet version 1.0 with tutorial data. Zenodo <https://doi.org/10.5281/zenodo.12640059> (2024).

Acknowledgements

This project is supported by a grant from the US National Institutes of Health (R01GM071940 to Z.H.Z.). We thank P. Zhou for helpful discussions. We acknowledge using computational resources from the Electron Imaging Center for NanoSystems (EICN) of the California NanoSystems Institute (CNSI) at UCLA.

Author contributions

Y.-T.L. conceptualized the method, wrote code, processed data, made illustrations and documentation, and wrote the paper; H.F. came up with the misalignment process, processed data, made documentation and illustrations, and wrote the paper; J.J.H. assisted in data processing, figure generation and paper writing; Z.H.Z. oversaw the project, interpreted the results and wrote the paper. All the authors edited and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

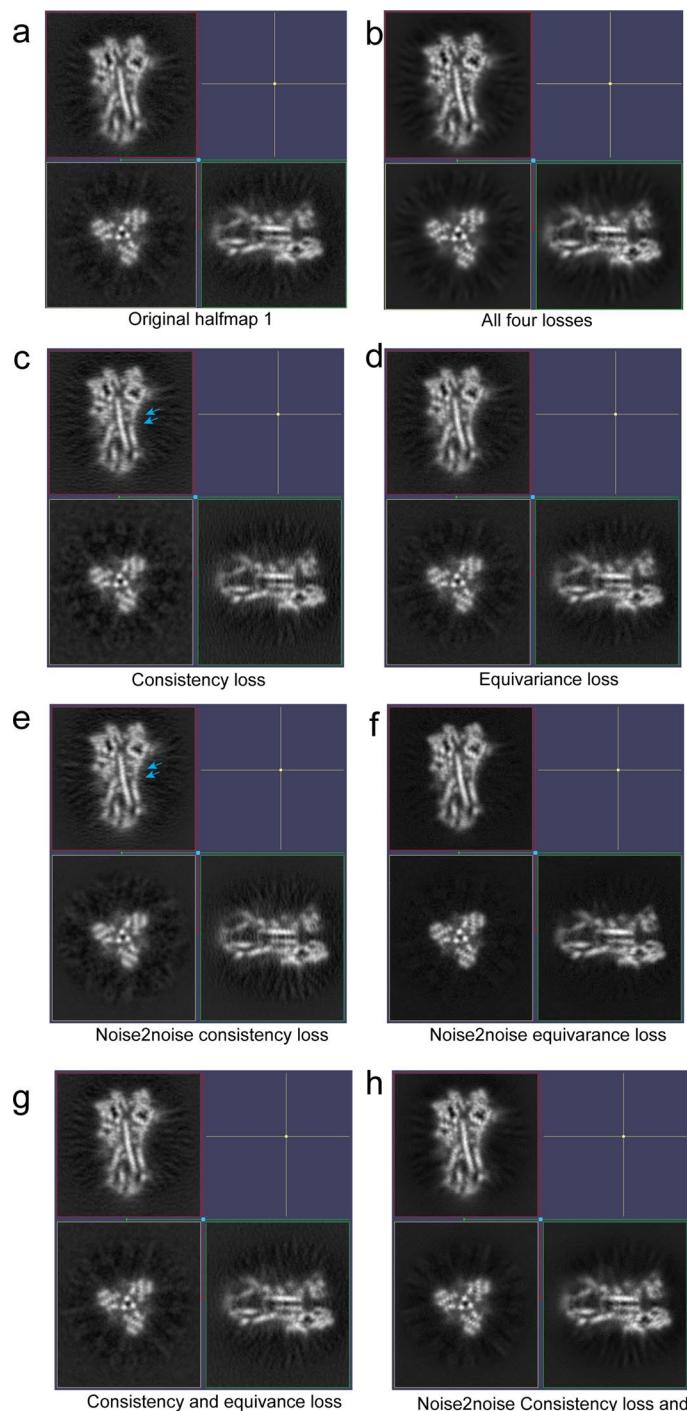
Extended data is available for this paper at <https://doi.org/10.1038/s41592-024-02505-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02505-1>.

Correspondence and requests for materials should be addressed to Z. Hong Zhou.

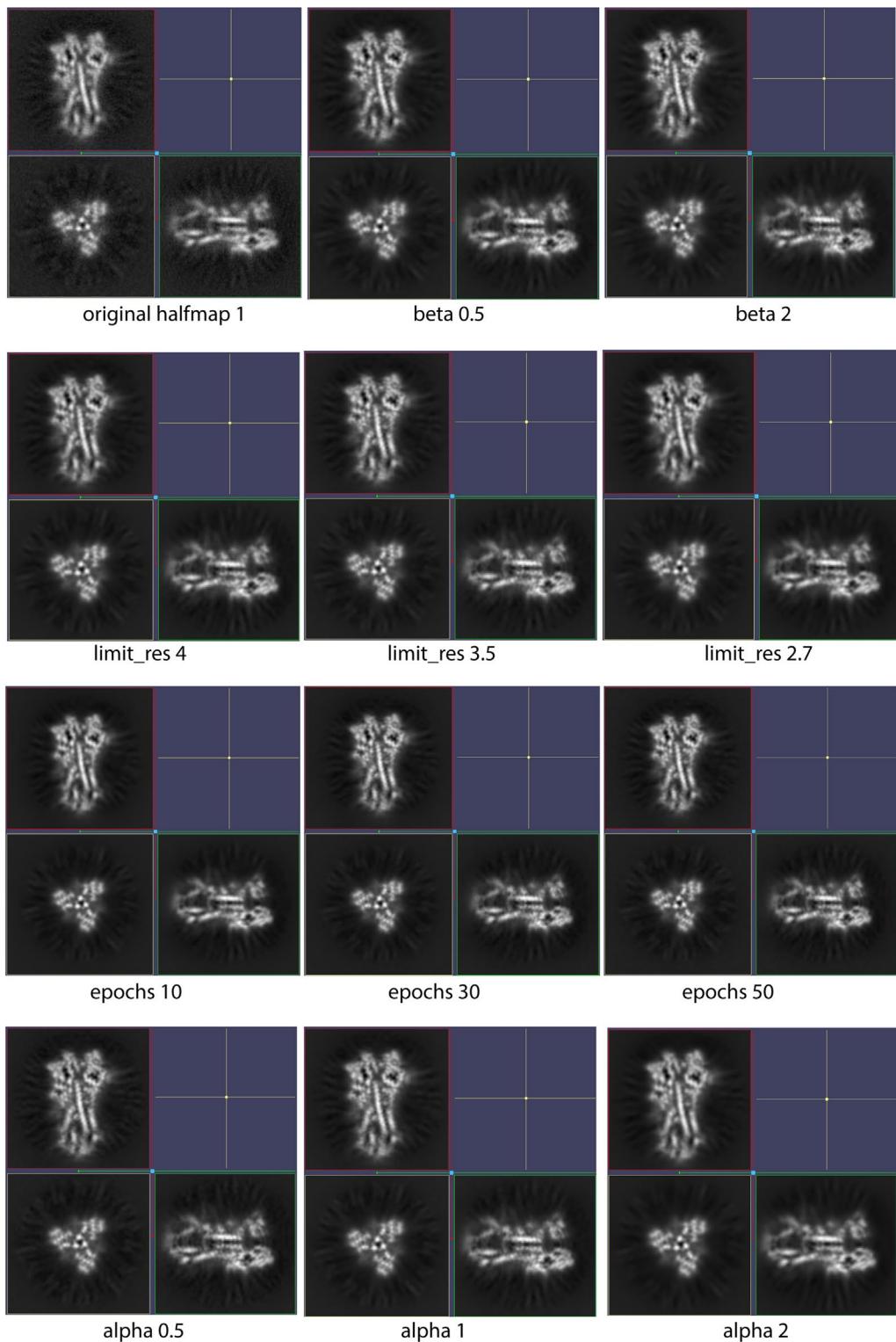
Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Arunima Singh, in collaboration with the *Nature Methods* team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

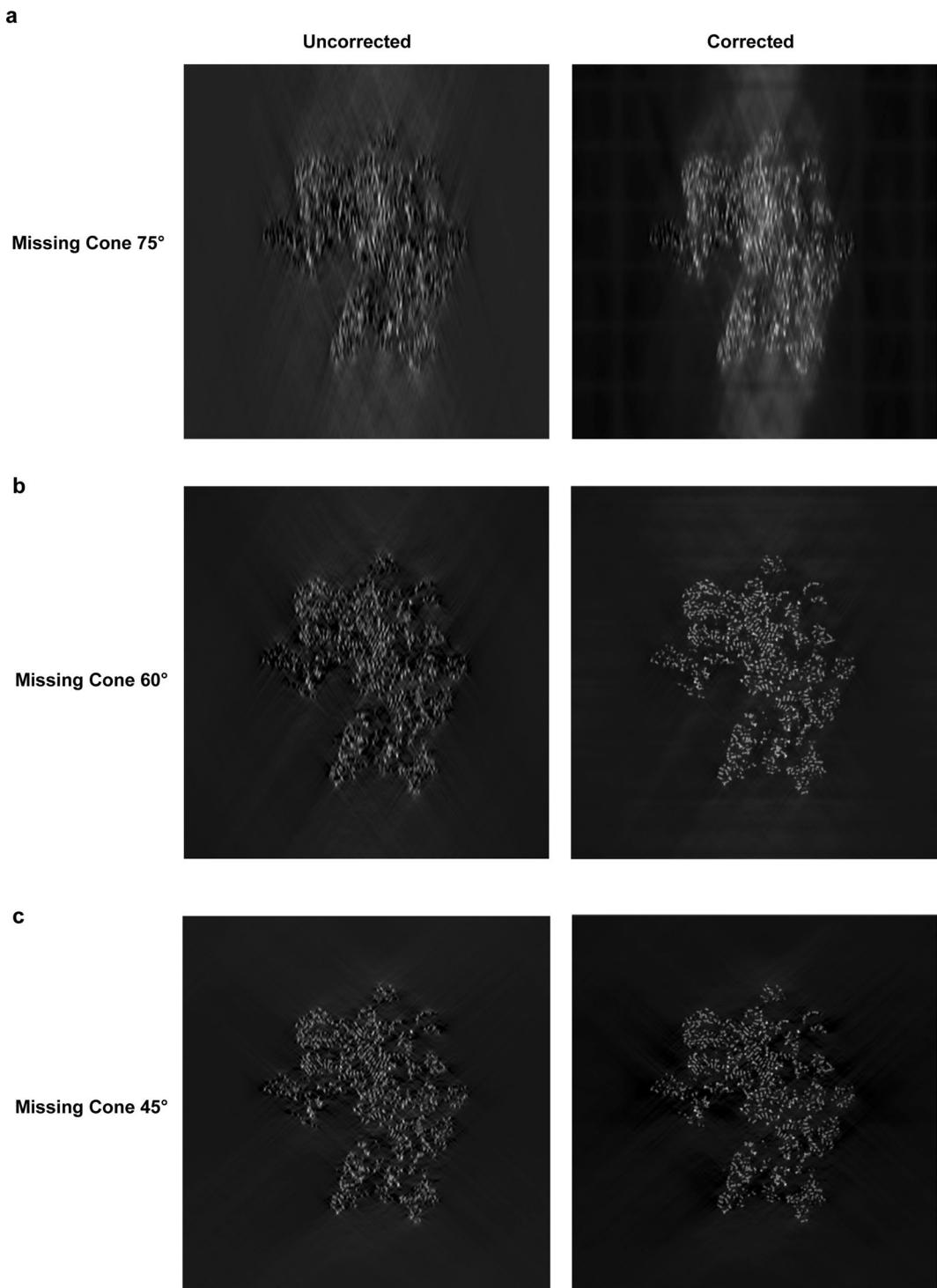

Extended Data Fig. 1 | Effects of different loss functions in splIsoNet.

Orthogonal slices of the cryoEM half-map obtained with splIsoNet *Anisotropy Correction* trained with different parameters. (a) the original cryoEM half-map. The parameters used are as follows, All four losses: $\alpha = 1, \beta = 0.5$ (b); consistency loss: $\alpha = 0, \beta = 0$ (c); equivariance loss: $\alpha = 1000, \beta = 0$ (d); noise2noise

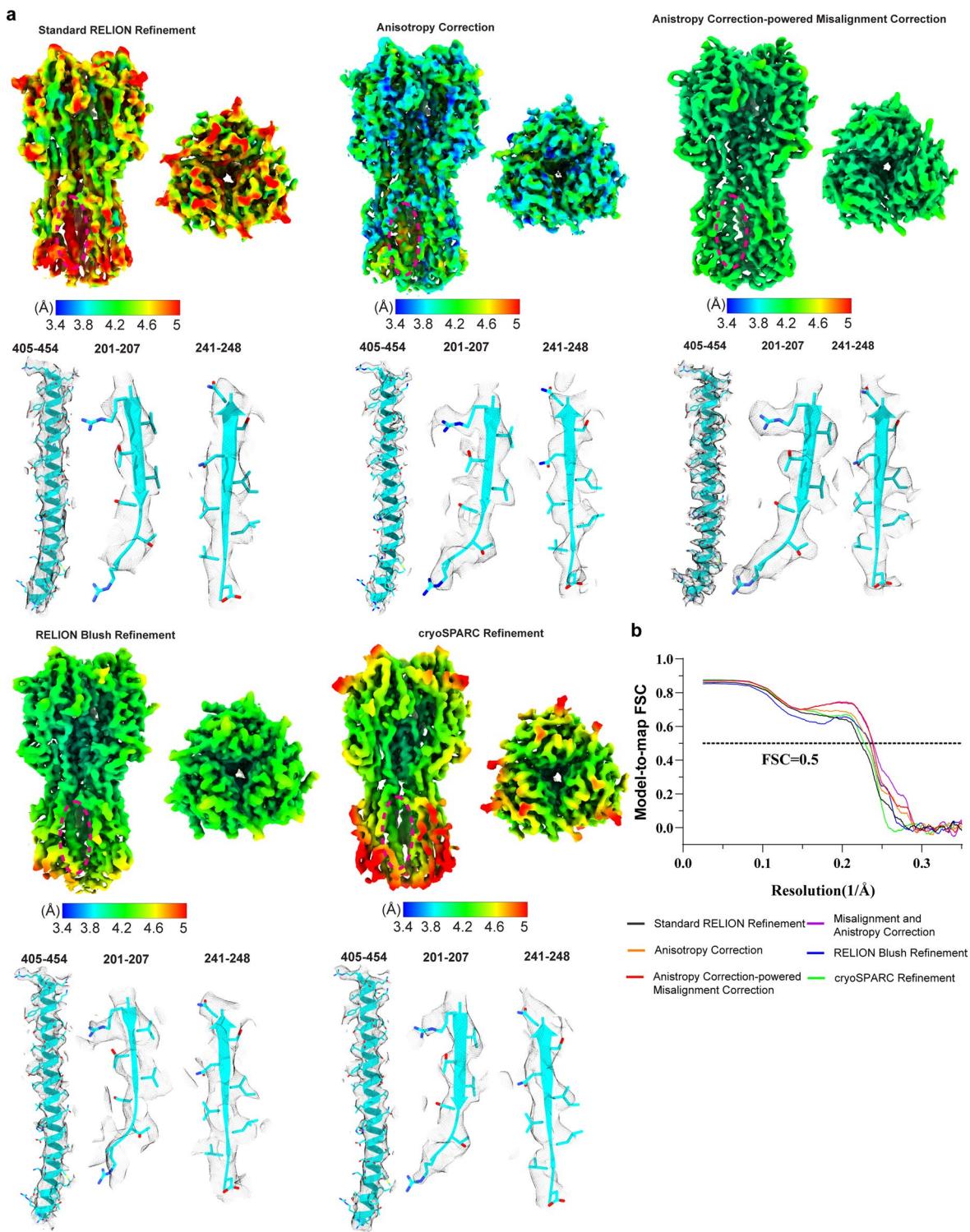
consistency loss: $\alpha = 0, \beta = 1000$ (e); noise2noise equivariance loss: $\alpha = 1000, \beta = 1000$ (f); consistency and equivariance losses: $\alpha = 1, \beta = 0$ (g); noise2noise consistency and equivariance losses: $\alpha = 1, \beta = 1000$ (h). Blue arrows show small artifacts. The default parameters in *Anisotropy Correction* are used if not specified.



Extended Data Fig. 2 | Sensitivity of splsoNet Anisotropy Correction with different α , β , epochs, and limit resolution values. Orthogonal slices of cryoEM half-map corrected using splsoNet Anisotropy Correction with different parameters. The default parameters are $\alpha = 1$, $\beta = 0.5$, epochs=30, limit_res=3.5. If the parameters are not indicated in the figure the default parameters are used.

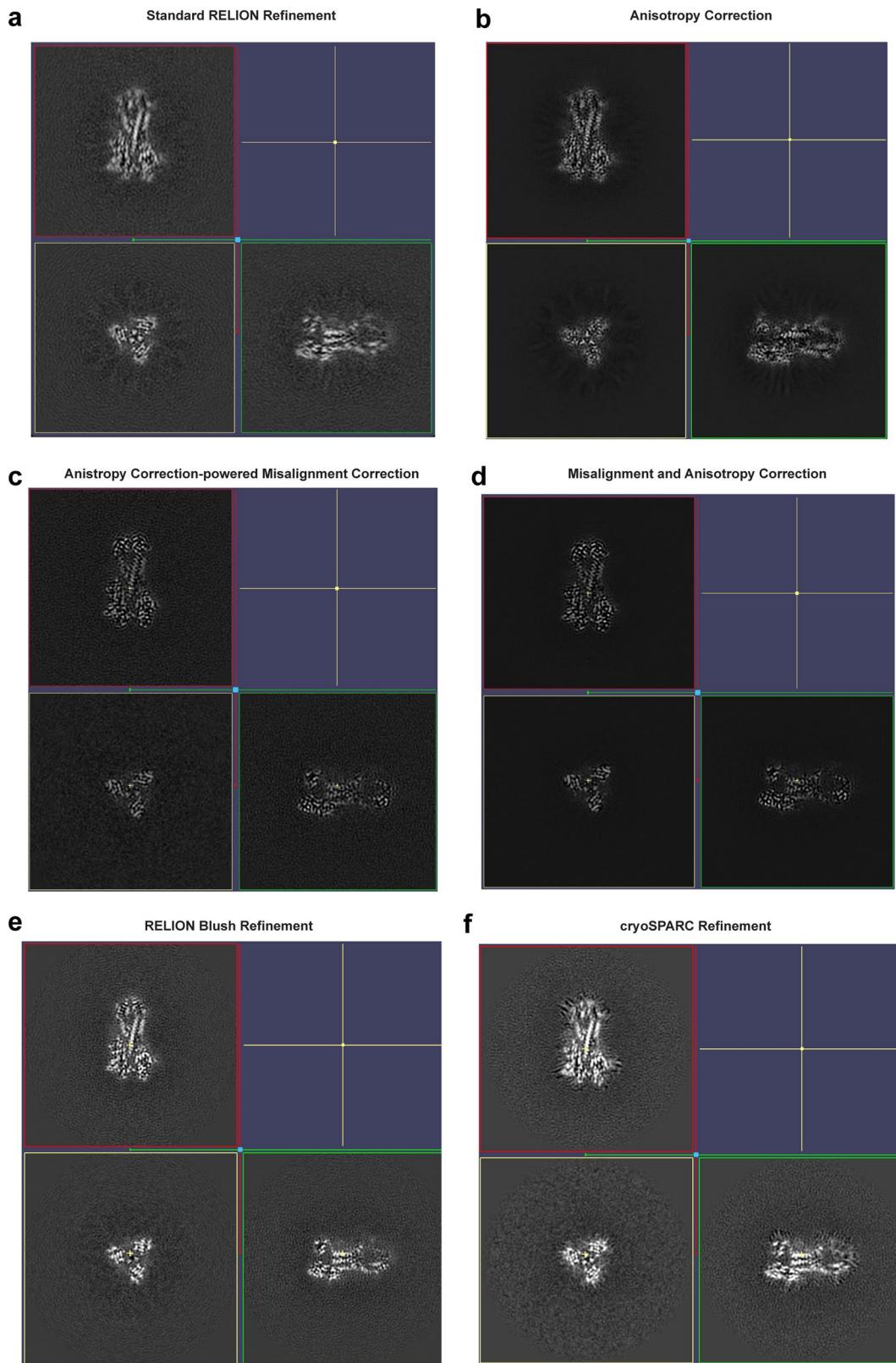


Extended Data Fig. 3 | Effect of the spIsoNet Anisotropy Correction on the simulated ribosome maps. a-c, Left panels show the sections of XZ planes for the uncorrected maps simulated with different missing cones, while the right panels show the sections on XZ planes for the maps with *Anisotropy Correction*.



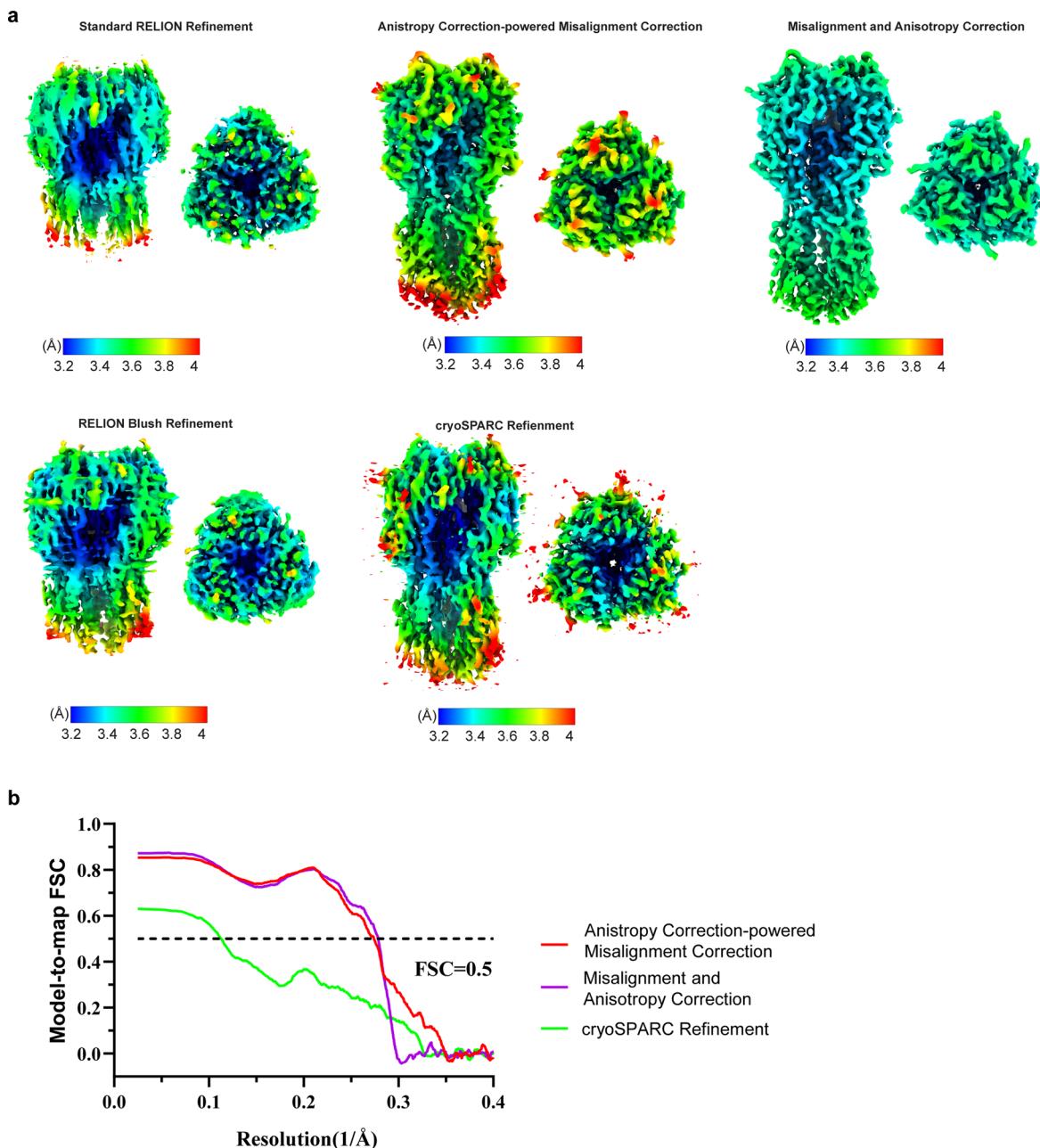
Extended Data Fig. 4 | splsoNet Anisotropy Correction and Misalignment Correction on the HA trimer dataset collected with stage tilt. **a**, Local resolution map and atomic model fitting of HA trimer reconstructed by different methods: standard RELION Refinement, splsoNet Anisotropy Correction, and

splsoNet Anisotropy Correction-powered Misalignment Correction, RELION Brush refinement; cryoSPARC homogeneous refinement. **b**, Model-to-map FSC curves of maps in **a**.

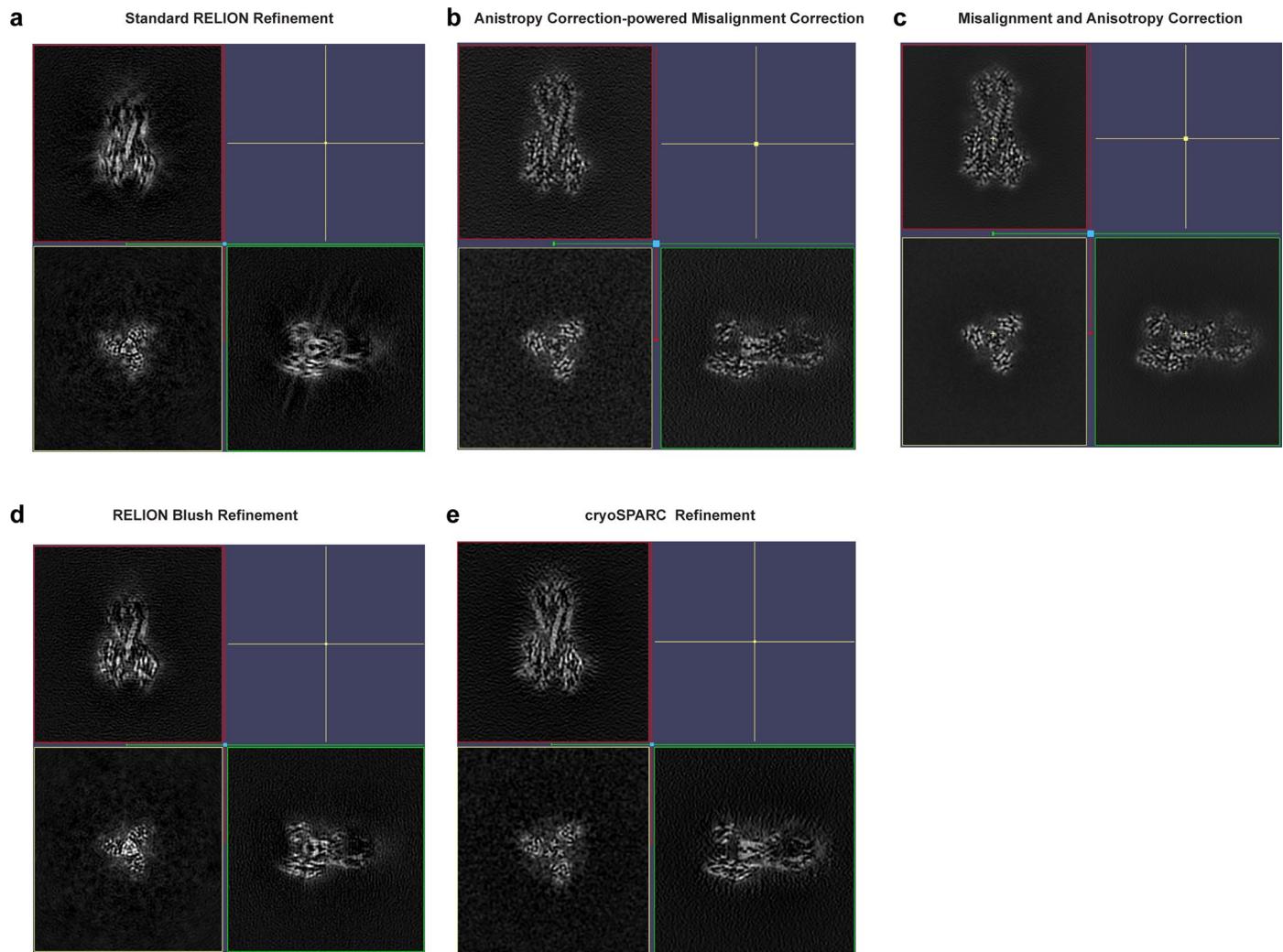


Extended Data Fig. 5 | Orthogonal map slices showing the effect of *spiloNet* on the tilted cryoEM influenza hemagglutinin (HA) trimer dataset.
a-f, Sections of orthogonal planes for the maps reconstructed by different refinement methods: standard RELION Refinement (a), Anisotropy Correction

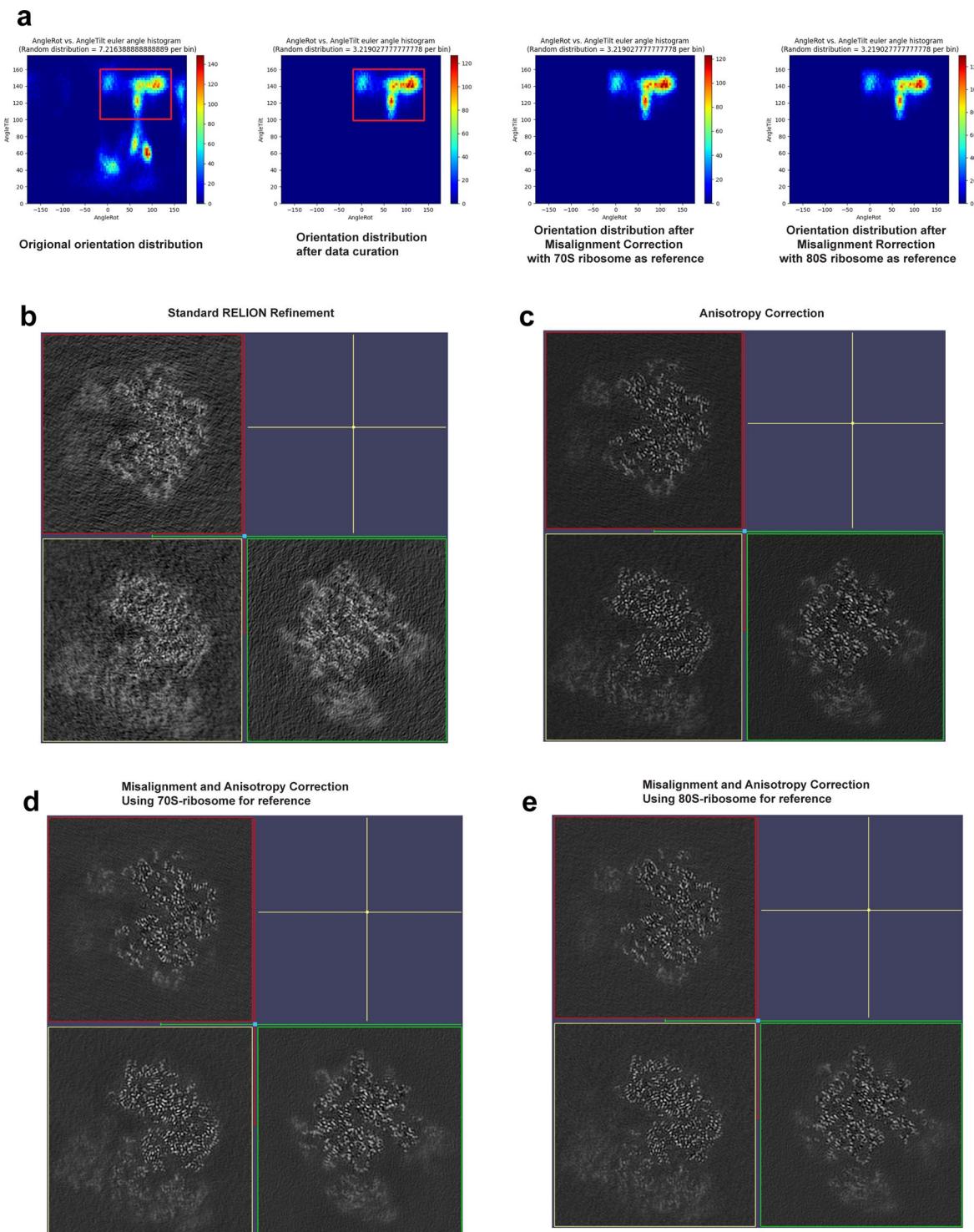
(b), Anisotropy correction-powered Misalignment Correction (c), the combination of Misalignment and Anisotropy correction (d), RELION Brush refinement (e), cryoSPARC homogeneous refinement (f).



Extended Data Fig. 6 | Effect of spIsoNet on the untilted cryoEM influenza hemagglutinin (HA) trimer dataset. **a**, Local resolution map of HA trimer reconstructed by different methods. **b**, model-to-map FSC for the maps reconstructed with *Misalignment Correction*, the combination of *Misalignment and Anisotropy correction*, and cryoSPARC homogeneous refinement.

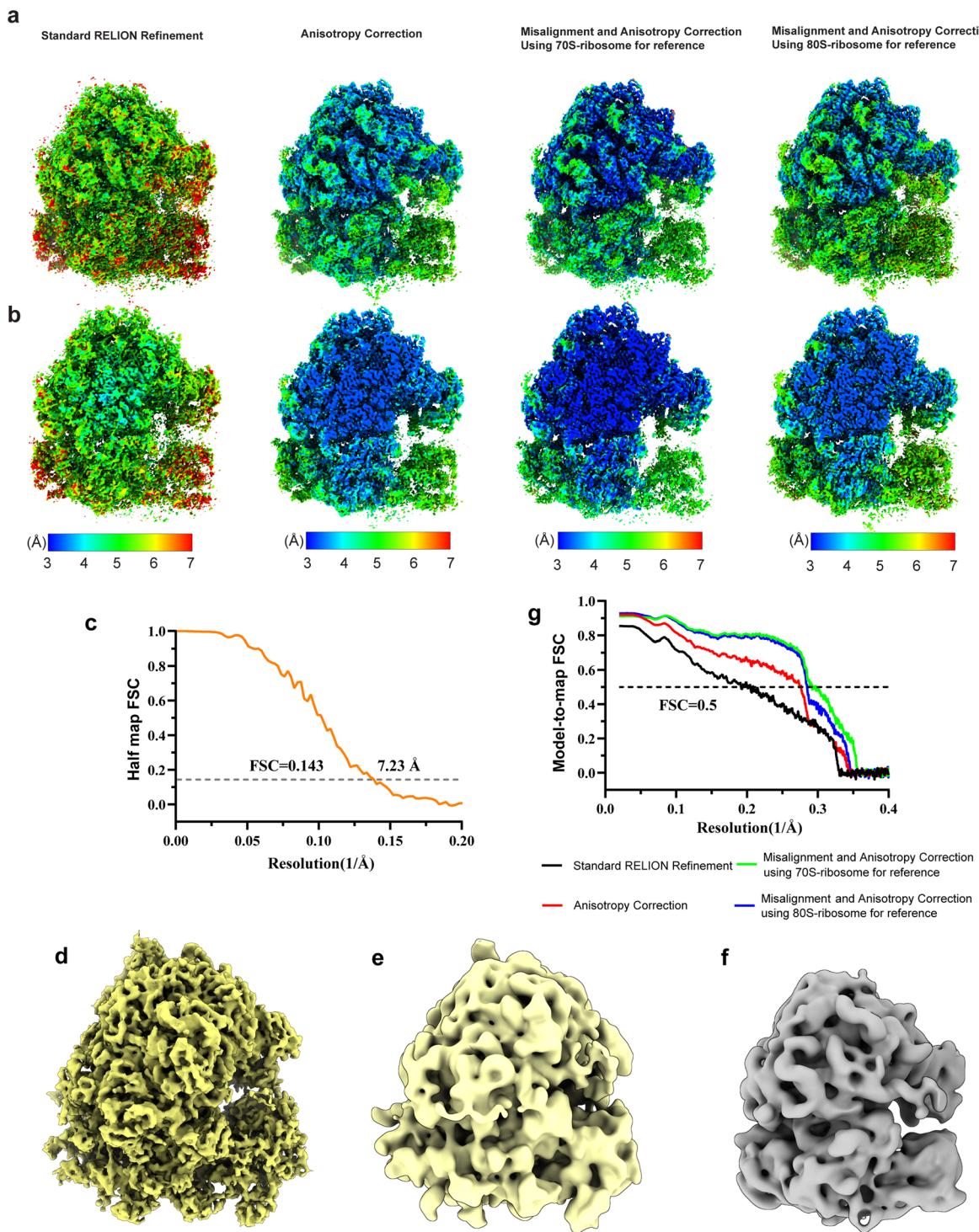


Extended Data Fig. 7 | Orthogonal slices showing the effect of spIsoNet on the untilted cryoEM influenza hemagglutinin (HA) trimer dataset. a–e, Sections of orthogonal planes for the maps reconstructed by different refinement methods: standard RELION Refinement (a), Anisotropy correction-powered Misalignment Correction (b), the combination of Misalignment and Anisotropy correction (c), RELION Blush refinement (d), cryoSPARC homogeneous refinement (e).



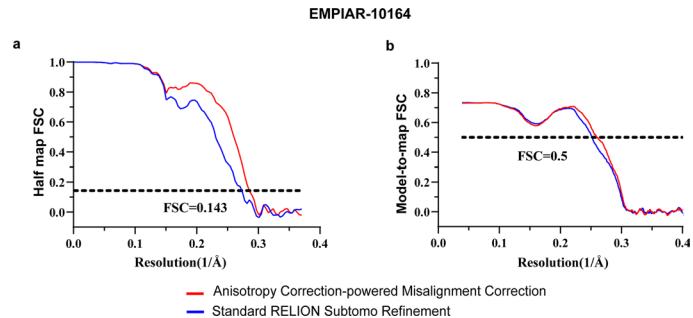
Extended Data Fig. 8 | Effect of Anisotropy and Misalignment Correction on the experimental ribosome datasets. **a**, Orientation distribution (rot and tilt Euler angles) for original particle dataset of ribosome, curated particle dataset, curated particle dataset after *Anisotropic Correction-powered Misalignment Correction* with different initial reference. **b-e**, Orthogonal sections of ribosome

reconstructed by different methods: standard RELION Refinement (**b**), spIsoNet *Anisotropy Correction* (**c**), spIsoNet *Misalignment Correction* and *Anisotropic Correction* using 70S ribosome as initial reference for particle alignment (**d**) and *Misalignment Correction* and *Anisotropic Correction* using 80S ribosome map as initial reference for particle alignment (**e**).



Extended Data Fig. 9 | spIsoNet Anisotropy and Misalignment Correction on the experimental ribosome datasets. **a–b**, Local resolution map of ribosome reconstructed by different methods: standard RELION refinement, spIsoNet Anisotropy Correction, spIsoNet Misalignment Correction and Anisotropy correction using 70S ribosome as initial reference for particle alignment, and Misalignment Correction and Anisotropy correction using 80S ribosome map as

initial reference for particle alignment. **c–d**, Subtomogram average map of the 80S yeast ribosomes using EMPIAR 10045 with its FSC curves. **e–f**, The map of 80S (**e**) and 70S (**f**) low passed to 15 Å and used as initial reference for particle alignment. **g**, model-to-map FSC for maps reconstructed by different methods in (a).



Extended Data Fig. 10 | Fourier Shell Correlation (FSC) plots for HIV-1 immature capsid datasets. **a**, Gold-standard FSC curves of the cryoEM HIV-1 immature capsid map with criterion of 0.143. Blue, RELION standard refinement; Red, splIsoNet *Anisotropy Correction-powered Misalignment Correction*. **b**, Model-to-map Fourier shell

correlation curves calculated between the cryoEM map and the refined atomic model with criterion of 0.5. Blue, RELION standard refinement; Red, splIsoNet *Anisotropy Correction-powered Misalignment Correction*.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No software was used for data collection

Data analysis RELION 4.0, RELION 3.1, IMOD 4.9.9, GCTF 1.18, ChimeraX 1.6.1, ResMap 1.1.5, TomoTwin 0.8.0, Scipion 3.0, splsoNet 1.0 (github link: <https://github.com/soNet-cryoET/splsoNet>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The tutorial dataset for Anisotropy and Misalignment Correction can be downloaded from Zenodo: <https://zenodo.org/records/12640059>. All data that support this study are also available from the corresponding authors upon request. The cryo-EM density maps of HA-trimer from EMPIAR-10096 and EMPIAR-10097 have been deposited in the EMDB under the accession numbers EMD-45551 and EMD-45555, respectively.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	<input type="text" value="N/A"/>
Reporting on race, ethnicity, or other socially relevant groupings	<input type="text" value="N/A"/>
Population characteristics	<input type="text" value="N/A"/>
Recruitment	<input type="text" value="N/A"/>
Ethics oversight	<input type="text" value="N/A"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. The sample size information is already available in publicly database EMDB and EMPIAR.
Data exclusions	No data exclusions
Replication	No replication was performed. The result can be reproduced following the protocol of our software.
Randomization	No randomization was performed. Randomization is not relevant to this study because the software processes input data to generate structural information based on predefined algorithms. The results depend on the data provided rather than experimental conditions or group comparisons that would typically require randomization.
Blinding	Blinding was not performed in this study because the software operates in an automated and objective manner, where the outcomes are generated solely based on the input data and computational algorithms. Since there is no subjective interpretation of results or experimental groups that could introduce bias, blinding is not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.