A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0

Jasenko Zivanov * [1,2,3], Joaquín Otón * [1,4], Zunlong Ke[1,5], Kun Qu[1,5], Dustin Morado[1,5], Daniel Castaño-Díez[3,6], Andriko von Kügelgen[7], Tanmay A. M. Bharat[1,7], John A. G. Briggs[1,5], and Sjors H. W. Scheres[1]

[1]MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
[2]Laboratory of Biomedical Imaging (LIB), EPFL, Lausanne, Switzerland
[3]BioEM lab, Biozentrum, University of Basel, Switzerland
[4]ALBA Synchrotron, Barcelona, Spain
[5]Max Planck Institute of Biochemistry, Martinsried, Germany
[6]Biofisika Institute (CSIC, UPV/EHU), Bilbao, Spain
[7]Sir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom

February 28, 2022

Abstract

We present a new approach for macromolecular structure determination from multiple particles in electron cryo-tomography (cryo-ET) data sets. Whereas existing subtomogram averaging approaches are based on 3D data models, we propose to optimise a regularised likelihood target that approximates a function of the 2D experimental images. In addition, analogous to Bayesian polishing and CTF refinement in single-particle analysis, we describe approaches that exploit the increased signal-to-noise ratio in the averaged structure to optimise tilt series alignments, beam-induced motions of the particles throughout the tilt series acquisition, defoci of the individual particles, as well as higher-order optical aberrations of the microscope. Implementation of our approaches in the open-source software package RELION aims to facilitate their general use, in particular for those researchers who are already familiar with its single-particle analysis tools. We illustrate for two cases that our approaches allow structure determination from cryo-ET data to resolutions sufficient for *de novo* atomic modelling.

1 Introduction

In recent years, electron cryo-microscopy (cryo-EM) has allowed the 3D imaging of an increasing number of biological macromolecules at resolutions sufficient for *de novo* atomic modelling. This development, originally driven by advances in detector technology, was further facilitated by novel, robust image processing algorithms. In single-particle analysis, images of multiple copies of isolated macromolecular complexes, or particles, that are suspended in random orientations in a thin layer of vitreous water are combined in a 3D reconstruction. Nowadays, many aspects of single-particle analysis workflows can be performed with only minimal human supervision, e.g. the detection, extraction and initial classification of particles in the images [27, 2, 24], 3D reconstruction [27, 14], as well as refinement of the optical parameters [27, 29, 14, 21] and per-particle tracking of electron beam-induced motion [26, 28]. Many of the algorithms that underlie these modern methods are built on solid statistical foundations that require few tuneable parameters. This decreases the need for operator expertise and provides objectivity, as well as robustness, in obtaining optimal structures.

The single-particle approach is, however, limited to investigating isolated protein complexes that are purified to relative homogeneity. To examine these complexes in their biological context, electron cryo-tomography (cryo-ET) may be used instead. In the tomographic approach, the sample is tilted multiple times during image acquisition, yielding a so-called tilt series of images from which a 3D tomogram can be computed. In the same manner as single-particle analysis, repeated occurrences of particles in those tomograms can then be aligned and averaged to obtain higher-resolution reconstructions. This process is referred to as *subtomogram averaging*. Unlike the field of single-particle analysis, many of the tools used for subtomogram averaging still require considerable levels of expertise from the operator, often in order to tune parameters that arise from heuristics in the underlying

*These authors contributed equally

algorithms. This not only provides a barrier for new scientists entering the field, but can also lead to the calculation of suboptimal structures.

Compared to single-particle analysis, subtomogram averaging faces several unique challenges. In addition to estimating the position and orientation of each particle, the algorithm also has to consider the geometry of the tilt series. Typically, this is solved through a set of pre-processing steps that include estimation of contrast transfer function (CTF) parameters and alignment of the tilt series, followed by the reconstruction of, often inconveniently large, tomograms for the entire field of view. Smaller subtomograms surrounding selected particles are then extracted from the tomograms and used in a separate process of subtomogram alignment and averaging. The separation between tomogram reconstruction and subtomogram averaging can lead to an accumulation of errors, as errors in the CTF estimation or tilt series alignments are hard to correct. In addition, because the sample cannot be rotated 180 degrees within the microscope, the subtomograms contain empty regions in Fourier space, the so-called missing wedge, which are difficult to deal with in subtomogram averaging.

A fundamental problem with subtomogram averaging as described above is that it transforms the original 2D image data into 3D subtomograms, which are then used as a substitute for experimental data in the alignment algorithm. RELION-3 introduced the concept of a 3D CTF to describe the transfer of information from the 2D images to the subtomograms, which dealt to some extent with the missing wedge and the loss of information through interpolations in the reconstruction algorithm [4]. A drawback of the 3D CTF approach is that it does not deal correctly with the lower resolution regions of Fourier space, where information from different tilt images overlaps. A statistically more attractive approach would be to formulate the optimisation target function directly as a function of the actual 2D images that are measured in the microscope. This has been proposed in an approach called constrained single-particle cryo-ET [1], where individually boxed particles from the tilt series images are processed as in single-particle analysis, but their relative orientations are kept fixed. To deal with unknowns in the relative orientations of the particles from the tilt series images, as well as their CTFs, the program M recently introduced new optimisation approaches that compare reference projections against the 2D particle images [21]. M still relies on RELION-3 for alignment and classification of 3D subtomograms that are recalculated from the optimised parameters in M. Nevertheless, this iterative approach allows subtomogram averaging to resolutions that approach those observed for single-particle analysis, even for particles in complex cellular environments [21].

Here, we describe a new approach in RELION-4.0 that approaches subtomogram averaging as the optimisation of a regularised likelihood function that depends directly on the 2D images of the tilt series. In order to do so at acceptable computational, as well as implementation costs, we have altered the main refinement program in RELION-4.0 to work with so-called *pseudo subtomograms*: explicitly constructed sets of 3D data arrays that contain sums of CTF pre-multiplied tilt series images, together with auxiliary arrays that contain the corresponding sum of squared CTFs and how often each 3D voxel has been observed. Pseudo subtomograms no longer aim to represent the actual scattering potential of the underlying particles. Instead, they represent a convenient way to implement the 2D approach within the existing RELION code. Evaluation of the pseudo subtomograms by RELION-4.0 approximates the likelihood of observing a hypothetical particle in the images of the entire tilt series, given the current model. Using that likelihood as a metric, operations equivalent to those in single-particle analysis can now be performed on tomographic data, e.g. 3D initial model generation, 3D classification or high-resolution refinement. In addition, we describe new methods for optimising parameters of the tilt series that exploit the increased signal-to-noise ratio in the average structure. Besides optimisation of the tilt series alignment itself, we also describe methods analogous to CTF refinement [27, 29] for refining descriptors of the optical properties (defocus, astigmatism and higher-order aberrations) and a method akin to Bayesian Polishing [28] to model beam-induced particle motion throughout the tilt series. Once all these parameters have been optimised, new pseudo subtomograms can be constructed and the alignment can be repeated. The resulting iterative image processing workflow is similar to existing approaches for single-particle analysis in RELION.

2 Methods

2.1 Particle alignment and averaging

RELION performs maximum-a-posteriori estimation to find the set of model parameters Θ that maximise the probability of observing the experimental images $\mathbb{X}$. Using Bayes' theorem, we define a regularised likelihood optimisation target function as:

$$P(\Theta|\mathbb{X}) = P(\mathbb{X}|\Theta)P(\Theta), \tag{1}$$

where $P(\Theta)$ expresses prior information about the model, i.e. that the reconstructed map has limited power in Fourier space, and $P(\mathbb{X}|\Theta)$ is the likelihood of observing the data given the model. A marginalised likelihood function is used, where one integrates over the unknown alignments ϕ of each individual particle. For simplicity, these integrals are omitted from the notations used in this paper.

The data model assumes independent Gaussian noise on the Fourier components of the cryo-EM images of individual particles p. We therefore write the negative log-likelihood of observing a particle in a hypothetical alignment ϕ as a sum over a grid of 2D Fourier pixels $\mathbf{j} \in \mathbb{R}^2$:

$$-\log\big(P(X|\phi)\big) \propto \sum_{\mathbf{j}} \frac{|X_{\mathbf{j}} - \mathrm{CTF}(\mathbf{j})V_{\mathbf{j}}^{(p)}|^2}{\sigma_j^2}, \tag{2}$$

where X is the Fourier transform of the experimental image, CTF the contrast-transfer function, $V_{\mathbf{j}}^{(p)}$ denotes the 2D slice out of the 3D Fourier transform of the known map V into the view of the particle, and σ_j^2 is the noise variance of the frequency band of $\mathbf{j}$. $V_{\mathbf{j}}^{(p)}$ is given by

$$V_{\mathbf{j}}^{(p)} = \exp(i\mathbf{t}_p \cdot \mathbf{j}) V(A_p \mathbf{j}) \tag{3}$$

for a 2D vector $\mathbf{t}_p$ and a 2×3 matrix A_p that respectively encapsulate the particle's position and orientation, and the evaluation of $V(A_p\mathbf{j})$ is achieved through linear interpolation.

In tomography, our aim is to approximate that same likelihood on tilt-series data. The direct equivalent is a sum over the pixels of the relevant regions of all images f from the tilt series:

$$-\log\big(P(X|\phi)\big) \propto \sum_{f,\mathbf{j}} \frac{|X_{f\mathbf{j}} - \mathrm{CTF}_f(\mathbf{j}) V_{f\mathbf{j}}^{(p)}|^2}{\sigma_j^2}. \tag{4}$$

We model the shifts and rotations as compositions of per-particle and per-image components:

$$\mathbf{t}_{pf} = A_f^\intercal \mathbf{T}_{pf} + \mathbf{t}_f \tag{5}$$

$$A_{pf} = R_p A_f, \tag{6}$$

where we keep the per-particle rotation component, R_p, identical for all images in the tilt series, and only vary A_f, the rotational alignment of the tilt series images. In turn, the tilt-series alignment A_f is shared among all particles in a given tilt image. The per-particle part of the translation is modelled as a 3D vector, $\mathbf{T}_{pf} \in \mathbb{R}^3$ that can vary over different tilt images f. This contrasts with single-particle analysis, where beam-induced motion of the particle can be corrected for as a preprocessing step [12, 17, 26, 28], so that each particle is associated with a single 2D translation in a motion-corrected image.

For our pseudo-subtomogram approach, we now approximate the sum over 2D pixels $\mathbf{j}$ and tilt images f in Eq. 4 by a sum over 3D voxels $\mathbf{k}$ in the pseudo-subtomogram:

$$-\log\big(P(X|\phi)\big) \propto \sum_{\mathbf{k}} \frac{|D_{\mathbf{k}} - W_{\mathbf{k}} V(R_p \mathbf{k})|^2}{M_{\mathbf{k}} \sigma_{\mathbf{k}}^2}. \tag{7}$$

Here, the data term D, its weight W, and a multiplicity volume M are 3D arrays in the Fourier domain. Together, they constitute a pseudo subtomogram. They are constructed as follows.

$$D_{\mathbf{k}} = \sum_{f,\mathbf{j}} l(A_{pf}\mathbf{j} - \mathbf{k}) \mathrm{CTF}_f(\mathbf{j}) X_{f\mathbf{j}} \tag{8}$$

$$W_{\mathbf{k}} = \sum_{f,\mathbf{j}} l(A_{pf}\mathbf{j} - \mathbf{k}) |\mathrm{CTF}_f(\mathbf{j})|^2 \tag{9}$$

$$M_{\mathbf{k}} = \sum_{f,\mathbf{j}} l(A_{pf}\mathbf{j} - \mathbf{k}), \tag{10}$$

where $l(\cdot)$ represents linear interpolation with forward mapping, i.e. each 2D Fourier pixel $\mathbf{j}$ is projected into 3D Fourier space, updating the 8 closest voxels. This allows the 2D images to be more finely sampled in Fourier space than the 3D volumes, which is equivalent to larger 2D images than 3D volumes in real space. This is desirable because the high spatial-frequency information is spread out by the effect of the CTF in the 2D images. After the 2D image has been modulated by its CTF again in Eq. 8, half of that signal is relocalised, so it fits into a smaller real-space region, which is equivalent to a less finely sampled Fourier volume. Note that only half the signal can be recovered this way – the other half is shifted twice as far and could only be retrieved by using a very large box, which is impractical even in single-particle analysis.

Ignoring the difference of pre-multiplying the images with their CTFs, Eq (7) aims to be equivalent of Eq (4). It is possible to replace individual Fourier components $X_{f\mathbf{j}}$ by sums over multiple components $D_{\mathbf{k}}$ because of the law of total variance: $|\langle\mathbf{X}\rangle - \mathbf{C}|^2 = \left\langle|\mathbf{X} - \mathbf{C}|^2\right\rangle - \mathrm{Var}(\mathbf{X})$, where $\mathbf{C}$ is a constant and $\mathrm{Var}(\cdot)$ represents the variance. In this case, $\mathrm{Var}(X_{f\mathbf{j}})$ remains constant for any alignment R_p, so it can be neglected. The variance $\sigma_{\mathbf{k}}^2$ is equivalent to σ_j^2, the power of the noise in the individual Fourier components in the 2D images. The additional term $M_{\mathbf{k}}$ in the denominator of Eq (7) arises from the observation that the variance in the sum of $M_{\mathbf{k}}$ independent random variables, each with variance $\sigma_{\mathbf{k}}^2$, is equal to $M_{\mathbf{k}}\sigma_{\mathbf{k}}^2$.

Optimisation of Eq (1) by expectation-maximization [6], while using Eq (7) to construct the likelihood function, leads to the following update formulae for V and σ^2:

$$V_{\mathbf{k}} = \frac{\sum_p D(R_p^\intercal \mathbf{k})/\sigma_k^2}{\sum_p W(R_p^\intercal \mathbf{k})/\sigma_k^2 + 1/\tau_k^2} \tag{11}$$

$$\sigma_k^2 = \frac{1}{N_k} \sum_{\mathbf{k} \in S_k} \frac{1}{M_{\mathbf{k}}} |D_{\mathbf{k}} - W_k V(R_p \mathbf{k})|^2, \tag{12}$$

where the term τ_k^2 arises from the prior $P(\Theta)$ and expresses the expected, frequency-dependent power of the signal; τ_k^2 and σ_k^2 are calculated by averaging over $\tau_{\mathbf{k}}^2$ and $\sigma_{\mathbf{k}}^2$ resp. in hollow spheres of radius k and thickness 1 described by S_k; N_k is the total number of voxels in S_k with $M_{\mathbf{k}} > 0$; and the divisions by τ_k^2 and σ_k^2 in Eq (11) are evaluated element-wise.

2.2 Orientational priors

One advantage of the approach to align pseudo subtomograms is that the coordinate system of the pseudo subtomograms themselves can be chosen arbitrarily. By default, the pseudo subtomograms are created in the same orientation as the tomogram, but the user can chose to align them in a more meaningful way. This is useful, as many proteins are organised in 2D arrays inside the tomograms, for example inside membranes or as part of capsid-like structures. Often, the individual protein molecules inside these arrays exhibit limited rotational freedom with respect to the surface normal of the array, although they may be able to rotate freely around that normal. By constructing the

pseudo subtomograms with their Z axis parallel to the normal of the 2D array, using a rotational prior on the tilt angle will limit the amount of rocking of the particles inside the array. This not only accelerates the refinement, as fewer orientations need to be evaluated, it also makes it possible to solve more challenging structures because fewer solutions are allowed.

2.3 Tilt series refinement

Averaging over multiple particles leads to an increased signal-to-noise ratio in the estimated density map V. We also implemented procedures that exploit V for subsequent re-estimation of parameters that describe the tilt series. These procedures do not require pseudo subtomograms, but are performed by comparing projections of the density maps directly with the (Fourier) pixels of 2D boxes that are extracted from the tilt series images, with a sufficient size to hold the CTF-delocalised signal. The various tilt series parameters are then estimated by minimising the negative log-likelihood as defined in Eq. 4, i.e. the sum over noise-weighted square differences between the prediction and the observation.

The tilt-series properties that can be refined fall into two broad categories: optical and geometrical. The optical refinement concerns the different parameters of the CTF, while the geometrical refinement aims to optimise the alignment of the tilt series, as well as the beam-induced motion of the individual particles. Both sets of algorithms are closely related to the corresponding single-particle algorithms in RELION: optical-aberration refinement [27, 29] and Bayesian Polishing [28], respectively. In spite of the similarity between the algorithms, the models that are optimised differ significantly from single-particle analysis. Details of the implementation of the optical and geometrical refinement algorithms are given in the Appendix.

CTF refinement for tomographic data in RELION-4.0 includes optimisation of scale factors that model frequency-dependent radiation damage, defocus, astigmatism, and higher-order symmetrical and antisymmetrical aberrations. Although individual particles within a field of view are at distinct defoci in the tilt series images, their relative defoci are known from the geometry of the tilt series and the known 3D positions of the particles in the tomogram. Therefore, one can efficiently perform defocus estimation in a single pass, considering all particles in a tilt series image simultaneously. In order to do so, we modified procedures that were developed for higher-order aberration estimation in single-particle analysis [29], where the information from all particles in each tilt series image is condensed into two images that are used to estimate a common phase shift (see Appendix).

Similar procedures can also be used to model higher-order symmetrical and antisymmetrical aberrations in the tomographic data. Analogously to our single-particle approach, they are modeled using Zernike polynomials and estimated in the same way. Because the higher-order aberrations are often only a limiting factor at relatively high spatial frequencies, a large number of particles are needed to estimate them reliably. Optimally, higher-order aberrations would thus be estimated globally, over the entire data set, and only for cases that yield high-resolution averages. Only aberrations that change too quickly during collection make it necessary to split the tomograms into optics groups, and estimate the aberrations per optics group. Typically, the third-order antisymmetrical aberrations are the most important ones, i.e. trefoil and axial coma, which can both be caused by a tilted beam. The resolution gains that these optimisations will yield depend on the microscope (mis)alignment. Provided alignment has been performed reasonably well, higher-order aberration correction will probably be most useful for reconstructions that extend beyond 3Å resolution.

The geometric alignment includes both the (rigid) rotational and translational re-alignment of the tilt series images, as well as the modelling of beam-induced motion of individual particles throughout the tilt series. For the latter, we neglect rotations of the particles, and only model beam-induced translations. By doing so, we can precompute the likelihood of each particle being in each position around its original one, and then look for an alignment that simultaneously maximises the sum of those likelihoods over all tilt series images and all particles, as well as a prior that ensures spatially coherent motion. This allows us to evaluate the likelihood of a hypothetical particle-position by looking up a single interpolated value in an image. In this formulation, the problem becomes equivalent to the Bayesian Polishing approach that we originally developed for single-particle analysis, except for the inclusion of a third spatial dimension for the motion.

3 Results

3.1 HIV immature capsid

We tested the workflow above on the cryo-ET dataset that was used to determine the structure of the immature capsid lattice and spacer peptide 1 (CA-SP1) regions of the Gag polyprotein from human immunodeficiency virus 1 (HIV-1) [18] (EMPIAR-10164). We used the same subset of 5 tomograms that were also used to assess the NovaCTF [22] and Warp [20] programs. Introducing 3D CTF correction, and using the alignment parameters from the original analysis by Schur et al, NovaCTF reported a resolution of 3.9 Å [22]. The Warp program introduced local and global motion correction in the tilt series images, as well as optimisation of CTF parameters. The combination of Warp and subtomogram alignment and averaging in RELION-3 led to a resolution of 3.8 Å.

We used tilt series projections after movie frame alignment from the original analysis [18], without any other preprocessing step, along with the tilt series alignment data, performed with IMOD package [11], and CTF parameters estimation using CTFFIND4

[15]. We used 13,320 particles from the 5 tomograms subset, reconstructed an initial reference map using the original published particle alignment and filtered it to 5 Å. A first alignment in 3D auto-refine, followed by averaging of the initial pseudo subtomograms, led to a resolution of 3.6 Å. This average was then used for a full cycle of pseudo-subtomogram improvement and realignment. We first applied CTF refinement to optimise the defoci of all particles. This improved the resolution only marginally. Subsequent optimisation of the tilt series geometry, including modelling local particle motion, improved the resolution to 3.5 Å. And finally, re-alignment of newly generated pseudo-subtomograms led to a resolution of 3.4 Å. A second cycle of these three steps provided 3.3 Å, while a third cycle converged to 3.2 Å (Figure 1a). Geometrical refinement was performed estimating local particle motion. Considering deformations did not show additional improvement. In the first cycle, where improvements in both CTFs and geometry are most obvious, the order of applying those optimisations did not alter the final result for this data set. These data and results are also distributed as part of the subtomogram tutorial in RELION-4.0, as described on https://relion.readthedocs.io/en/release-4.0/.

Analysis of the complete dataset generated a structure at 3.0 Å resolution (not shown), which is the same resolution obtained using the M and RELION-3 workflow [21], and is likely limited by flexibility and asymmetry in the CA hexamer.

3.2 *Caulobacter crescentus* S-layer

We also applied our approach to thin cellular appendages of *C. crescentus* bacteria known as stalks, which have previously been imaged using cryo-ET [3]. The cell body and cell stalks of *C. crescentus* cells are covered by a nearly hexagonal, paracrystalline array known as the surface layer (S-layer) [19]. The structure of the S-layer was solved using a combination of X-ray crystallography, cryo-EM single particle analysis and subtomogram averaging, revealing how the S-layer is attached to bacterial cells by an abundant glycolipid called lipopolysaccharide (LPS) [3, 23]. Previously, cryo-ET of the S-layer, using 110 tilt series collected with a dose-symmetric scheme, yielded 51,866 hexamers of the S-layer. Using subtomogram averaging methodology described previously, which is based on a constrained cross-correlation approach implemented in the AV3 Matlab suite [8], specifically optimised for the analysis of macromolecules arranged in a lattice [25], a 7.4 Å reconstruction of the S-layer was obtained, in which alpha-helices were resolved [3]. This reconstruction was improved by application of NovaCTF [22], leading to a 4.8 Å reconstruction, in which large amino acid residue side chains were resolved [23]. Moreover, density for an LPS molecule was observed near the putative LPS-binding residues of the S-layer, in agreement with a cryo-EM single particle structure of an *in vitro* reconstituted complex [23].

We used the tilt series after movie frame alignment from the initial analysis [3], without additional pre-processing, along with the tilt series alignments performed within IMOD [11], CTF parameters from CTFFIND4 [15], and the Euler angle assignments and subtomogram co-ordinates from the original analysis. These parameters were imported into RELION-4.0, followed by multiple cycles of pseudo-subtomogram generation and refinement, analogous to the immature HIV dataset described above, leading to a 5.6 Å reconstruction of the S-layer hexamer (Figure 2a). Next, we defined a mask around the central pore of the S-layer, corresponding to the inner domain bound to LPS, to perform focused refinements. Another cycle of pseudo-subtomogram reconstruction, CTF refinement and refinement within the new mask improved the resolution to 4.4 Å. Accounting for per-particle motions with additional cycles of pseudo-subtomogram improvements and refinements increased the resolution of the central pore to 4.0 Å, and the inner domain of the S-layer to 3.7 Å. Further 3D classification without alignments identified a subset of 42,990 subtomograms that gave a 3.5 Å resolution reconstruction of the inner S-layer.

The 3.5 Å map is in excellent agreement with the single particle structure of the *in vitro* reconstituted complex, including the LPS binding site [23]. Furthermore, divalent metal ions, known to be tightly bound to the inner S-layer [10], are resolved (Figure 2b). Surprisingly, at lower isosurface contour levels, we also observed a second LPS binding site (Figure 2c-d). The size and shape of this density agrees with the structure of the LPS O-antigen, illustrating how improved subtomogram averaging in RELION-4.0 can help uncover new biology.

4 Discussion

We formulate the problem of averaging over multiple identical particles in tomographic tilt series in an empirical Bayesian framework that is based on a statistical model for the two-dimensional experimental data. The same statistical framework has proven effective in reducing the number of tunable parameters and in obtaining high-quality reconstructions from single-particle data [7]. The two-dimensional data model describes the experimental images better than alternative approaches that use 3D reconstructed subtomograms as an intermediate. One example of a problem with the intermediate 3D data model is the need for missing wedge correction, which arises from the observation that the experimental images were acquired, incompletely, in three dimensions. Artifacts related to missing wedge correction may affect the classification of distinct structural states, by separating particles according to the orientation of their missing wedge. By using a 2D data model, missing wedge correction is no longer required. Instead, the problem approaches that of single-particle analysis, where projections from different orientations and of different structural states are sorted out simultaneously. Provided the 3D Fourier transform of the distinct classes

is fully sampled through the orientation distribution of the raw particles, likelihood optimisation techniques have been highly successful in tackling this problem in single-particle analysis [16, 7]. Therefore, the methods presented here may be particularly advantageous over existing subtomogram averaging methods when multiple distinct structural states exist in the set of particles.

In practice, the implementation in RELION-4.0 does not use stacks of 2D projection images as input for the refinement program that performs alignment and classification. Instead, the concept of pseudo-subtomograms is introduced, where the tilt series images are Fourier transformed, pre-multiplied with their CTF, and inserted as a slice into a 3D Fourier volume according to the best current estimates for the tilt series geometry. In order to approach the likelihood of observing the individual 2D tilt series images, the likelihood calculation from the 3D pseudo subtomogram requires separate storage of the accumulated squares of the CTFs, as well as a multiplicity term that keeps track how often each pixel in Fourier space is sampled. In the limit of infinitely fine sampling of the Fourier transform, the resulting likelihood function only differs from the exact one by a constant offset. The use of pseudo subtomograms allowed re-using existing code for subtomogram averaging in RELION, while input stacks of 2D images would have required significant software development efforts. Nevertheless, in the future one might still choose to implement a true 2D version of the code, as the number of Fourier pixels to examine, and hence the computational cost, will decrease under the currently common tomography setups. Specifically, a 2D implementation is more efficient when the number of tilt images is small in comparison to the particle size measured in pixels. For example, the current implementation could be used to process pairs of tilted images, but a 2D implementation would be more efficient. However, if one were to collect tilt series in a continuous manner [5], then the current implementation is preferable.

Besides the alignment and classification of individual particles, the methods described here also deal with re-estimation of parameters that describe the optical and geometrical features of the tilt series. As soon as an initial average structure has been obtained, its increased signal-to-noise ratio can be exploited to determine these parameters more accurately than what is possible from the raw tilt series images alone. The implementations in RELION-4.0 again follow those previously implemented for single-particle analysis, where CTF refinement is used for re-estimation of the tilt series images CTFs, and algorithms akin to Bayesian polishing are used to re-estimate the tilt-series alignment, as well as the movement of individual particles throughout the tilt series acquisition process. As better tilt series parameters will allow better pseudo-subtomograms, particle alignment and classification are iterated with the optimisation of the tilt series parameters.

Similar tilt series and CTF optimisation approaches have been implemented in the program M. Compared to M, RELION-4.0 uses computationally more efficient algorithms that do not require the computational power of a GPU. However, both in tomography and in SPA, RELION-4.0 only models beam-induced translations of the particles, whereas M also models beam-induced rotations. Since SPA routinely reaches 2 Å resolutions without modelling beam-induced rotations, we assumed that the effect of rotations of individual particles throughout the tilt series is not large enough to warrant their correction at typical tomography resolutions. In cases where the data do allow for unusually high resolutions and a GPU is available, M could still be used in a postprocessing step, following alignment and classification of the individual particles in RELION. It is likely that adaptation of M, in order to function with the pseudo-subtomograms proposed here, would lead to increased synergy between the two programs. In the meantime, external tools to convert from M parameters to RELION-4.0 are already available (https://github.com/joton/reliontomotools).

Besides the reduction in tunable parameters that is characteristic of the Bayesian approach, its uptake by researchers that are new to the field is further facilitated through the implementation of a graphical user-interface. This interface is already widely used for single-particle analysis, and has been extended for the processing of tomographic data in RELION-4.0. Apart from the calculations that will be familiar to users of single-particle analysis, e.g. 3D classification, 3D initial model generation, 3D auto-refinement, the new interface also provides convenient access to the tomography-specific versions for CTF refinement and Bayesian polishing, as well as pre-processing operations to calculate the pseudo-subtomograms and to import data and metadata from pre-processing operations in IMOD [11]. To further facilitate the uptake of this new software by the community, we have provided an online tutorial that uses the publicly available HIV immature capsid data set to describe and illustrate all steps necessary to obtain the results described in Figure 1.

In summary, we introduce new methods for subtomogram averaging to resolutions that are sufficient for *de novo* atomic modelling, and increase the accessibility of this emerging technique. We envision that our methods will allow more researchers to calculate better structures from tomographic data, which will aid the next revolution in structural biology, where macromolecular complexes are imaged, not in isolation, but in their biologically relevant environment.

Acknowledgements

We are grateful to the MRC-LMB EM facility for help with data acquisition; and to Jake Grimmett and Toby Darling for help with high-performance computing. This work was funded by the UK Medical Research Council (MC_UP_A025_1013 to S.H.W.S.; and MC_UP_1201/16 to J.A.G.B.), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-

CoG-2014, grant 648432, MEMBRANEFUSION to J.A.G.B.); the Swiss National Science Foundation (grant 205321_179041/1 to D.C.-D.) and the Max Planck Society (to J.A.G.B.). T.A.M.B. is a recipient of a Sir Henry Dale Fellowship, jointly funded by the Wellcome Trust and the Royal Society (202231/Z/16/Z). J.Z. was partially funded by the European Union's Horizon 2020 research and innovation program (ERC-ADG-2015, grant 692726, GlobalBioIm to Michael Unser). T.A.M.B. thanks the Vallee Research Foundation, the Leverhulme Trust and the Lister Institute of Preventative Medicine for support.

Appendix

CTF Refinement

CTF refinement in RELION-4.0 optimises the following parameters: scale, defocus, astigmatism and higher-order (even and odd) optical aberrations. Since, save for the difference in defocus, the same CTF needs to be valid for an entire micrograph of particles, similar optimisations can be applied as in our single-particle algorithms. All the relevant information is first consolidated into a minimal form using linear transformations, and the final, typically non-linear, optimisation is then performed on that minimal form.

We formulate the CTF for frame f of particle p as follows:

$$\mathrm{CTF}_{pf}(\mathbf{j}) = -\alpha_f \tau_f(\mathbf{j}) \sin(\gamma_{pf}(\mathbf{j})) \exp(i\rho_f(\mathbf{j})), \quad (13)$$

where α_f describes the overall scaling factor, $\tau_f(\mathbf{j})$ the empirical radiation-damage model as defined by Grant and Grigorieff [9], $\gamma_{pf\mathbf{j}}$ the symmetrical phase delay component and $\rho_{f\mathbf{j}}$ the antisymmetrical one. Note that only γ varies between particles. This is because it contains the quadratic defocus term that depends on the position of the particle. The phase delays are parametrised the same way as in single-particle analysis in RELION-3 – as a combination of explicitly named low-order terms and higher-order Zernike polynomials:

$$\gamma_{pf}(\mathbf{j}) = \pi\lambda\mathbf{j}^\intercal D_{pf}\mathbf{j} + \frac{\pi}{2}C_s\lambda^3|\mathbf{j}|^4 - \chi_f + \sum Z_n^m(\mathbf{j}), \quad (14)$$

$$D_{pf} = \begin{bmatrix} \delta z_p + a_1 & a_2 \\ a_2 & \delta z_p - a_1 \end{bmatrix}. \quad (15)$$

As before, the astigmatic-defocus matrix D_{pf} is decomposed into a defocus term δz_p and two linear astigmatism terms, a_1 and a_2, while C_s describes the spherical aberration of the microscope, χ_f a constant phase offset (owing to amplitude contrast and a phase plate, if one is used), λ is the wavelength of the electron and Z_n^m are the higher-order even Zernike terms. One key difference to single-particle analysis is that the defocus term δz_p is no longer a free parameter for each particle, but it instead depends on the already known 3D position of the particle. Therefore, in tomography, the defocus term is only estimated once per tilt-image, and all the particles contribute to that estimate.

The scaling factor α_f is estimated by computing the following two sums for each micrograph and dividing them (the † symbol indicates complex conjugation):

$$\alpha_f = \frac{G_f}{H_f} \quad (16)$$

$$G_f = \sum_{p,\mathbf{j}} \frac{1}{\sigma^2_{|\mathbf{j}|}} \mathrm{Re}(X_{pf\mathbf{j}}(\mathrm{CTF}'_{pf}(\mathbf{j})V^{(p)}_{pf\mathbf{j}})^\dagger) \quad (17)$$

$$H_f = \sum_{p,\mathbf{j}} \frac{1}{\sigma^2_{|\mathbf{j}|}} |\mathrm{CTF}'_{pf}(\mathbf{j})V^{(p)}_{pf\mathbf{j}}|^2. \quad (18)$$

Note that the CTF' used in Eq. 18 is missing its scale factor:

$$\mathrm{CTF}'_{pf}(\mathbf{j}) = -\tau_f(\mathbf{j}) \sin(\gamma_{pf}(\mathbf{j})) \exp(i\rho_f(\mathbf{j})). \quad (19)$$

Alternatively, we also allow the user to fit the parameters of Lambert's extinction model to the data instead, assuming perfectly flat samples of constant thickness. In that case, the CTF scale in image f of tomogram t is expressed as a function of the beam luminance α_0, sample normal $\mathbf{n}_t$ and optical sample thickness κ_t:

$$\alpha_{tf}(\alpha_0, \kappa_t, \mathbf{n}_t) = \alpha_0 \kappa_t^{\frac{1}{|\mathbf{n}_t \cdot \mathbf{z}_f|}} \quad (20)$$

If this option is used, then the CTF scales of all the tilt series in the data set are estimated together. The beam luminance α_0 is modeled globally, while the sample thickness and normal are allowed to differ among tomograms, but not between the images of a tilt series. The vector $\mathbf{z}_f$ points in viewing direction of tilt image f. Note that this model does not allow for separating the geometrical sample-thickness from its extinction factor, so we can only estimate the product of the two. Also, the ice normal is required to be perpendicular to the estimated tilt axis of the tilt series, since its component pointing in the direction of the axis is indistinguishable from an increase in ice thickness or opacity. This global optimisation is performed using the sums G_{tf} and H_{tf} computed in Eqs. 17 and 18, where the subscript t indicates tilt series t. This is done by finding a global value of α_0 and values of κ_t and $\mathbf{n}_t \in \mathbb{S}^2$ for all tomograms that produce α_{tf} which minimise the following quantity and thus maximise the overall likelihood in Eq. 4:

$$\sum_{t,f} (H_{tf}\alpha_{tf}(\alpha_0, \kappa_t, \mathbf{n}_t) - G_{tf})^2. \quad (21)$$

To perform defocus estimation efficiently, we apply the optimisations we originally developed for the estimation of higher-order aberrations in single-particle analysis [29]. It allows us to determine a collective offset to γ for a large set of particles that all have different values of γ. Specifically, it allows the change to the log-likelihood arising from changing the value of γ at any Fourier pixel to be expressed as a pair of 2D images, independently of the number of particles. Therefore, each pixel of each particle only needs to be considered once. After that, the log-likelihood can be evaluated by iterating over the pixels of a single image.

In single-particle analysis, this approach is used to estimate the higher-order aberrations that are shared among all the particles in a data set. In tomography, we also use this approach to condense the information from all the particles in a tilt image (all of which exhibit slightly different defoci), into two such images, and to then determine the optimal change to γ very efficiently using a nonlinear algorithm.

The two condensed images R and $\widehat{\mathbf{t}}$ that we compute are the same as the ones in single-particle analysis, except for the inclusion of the noise power σ^2. The definitions are repeated here for the sake of completeness. Note that each pixel $\mathbf{j}$ of R contains a real symmetrical 2×2 matrix and each pixel of $\widehat{\mathbf{t}}$ a $\mathbb{C}^2$ vector:

$$R_{f\mathbf{j}} = \sum_p \frac{1}{\sigma_j^2} |\widetilde{V}_{f\mathbf{j}}^{(p)}|^2 \mathbf{d}_{pf\mathbf{j}} \mathbf{d}_{pf\mathbf{j}}^\intercal \tag{22}$$

$$\widehat{\mathbf{t}}_{f\mathbf{j}} = -R_{f\mathbf{j}}^{-1} \sum_p \frac{1}{\sigma_j^2} \mathrm{Re}(X_{pf\mathbf{j}}^\dagger \widetilde{V}_{f\mathbf{j}}^{(p)}) \mathbf{d}_{pf\mathbf{j}}, \tag{23}$$

where $\mathbf{d}_{pf\mathbf{j}} \in \mathbb{R}^2$ describes the point on the unit circle corresponding to the initial phase angle $\gamma^{(0)}$, which is given by the initial CTF parameters:

$$\mathbf{d}_{pf\mathbf{j}} = \begin{bmatrix} \cos(\gamma_{pf}^{(0)}(\mathbf{j})) \\ \sin(\gamma_{pf}^{(0)}(\mathbf{j})) \end{bmatrix}. \tag{24}$$

The predicted 2D images $\widetilde{V}_f^{(p)}$ contain the effect of the initial CTFs, except for the symmetrical aberration:

$$\widetilde{V}_{f\mathbf{j}}^{(p)} = -\alpha_f \tau_f(\mathbf{j}) \exp(i\rho_f(\mathbf{j})) V_{f\mathbf{j}}^{(p)}. \tag{25}$$

The vector-valued condensed image $\widehat{\mathbf{t}}_f$ describes the most likely phase shift γ for each pixel $\mathbf{j}$, expressed as a point on a circle, while the matrix-valued one, R_f, describes the anisotropic weight of that information. With these two condensed images computed for a given tilt image f, the change to the likelihood defined in Eq. 4 resulting from the change to the phase delay γ at any pixel can be expressed as a quadratic form. Therefore, we look for a change δD to the astigmatic-defocus matrices D_{pf} which produces phase delays that minimise that quadratic form:

$$C_f = \sum_{\mathbf{j}} \mathbf{e}_{\mathbf{j}}^\intercal(\delta D) R_{\mathbf{j}} \mathbf{e}_{\mathbf{j}}(\delta D), \tag{26}$$

where the per-pixel error $\mathbf{e}_{\mathbf{j}}(\delta D)$ is given by the deviation from the optimal phase shift $\widehat{\mathbf{t}}_{\mathbf{j}}$:

$$\mathbf{e}_{\mathbf{j}}(\delta D) = \begin{bmatrix} \cos(\delta\gamma) - \mathrm{Re}(\widehat{\mathbf{t}}_{\mathbf{j}}) \\ \sin(\delta\gamma) - \mathrm{Im}(\widehat{\mathbf{t}}_{\mathbf{j}}) \end{bmatrix} \tag{27}$$

$$\delta\gamma = \mathbf{j}^\intercal \delta D \mathbf{j} \tag{28}$$

As an alternative to fitting D independently for each tilt image, our program also allows the user to apply an L^2 regulariser to the δD_f of the different tilt images in the same series. In that case, the sum in Eq. 26 runs over all the pixels $\mathbf{j}$ of all the frames f. This helps to stabilise the CTFs of the higher tilts, but it risks impairing the estimates of the CTFs of the more important lower tilts. Formally, this is done by minimising the following cost:

$$C_{\mathrm{glob}} = \sum_{f,\mathbf{j}} \mathbf{e}_{f\mathbf{j}}^\intercal(D_f) R_{f\mathbf{j}} \mathbf{e}_{f\mathbf{j}}(D_f) + \lambda \sum_f |D_f - \widehat{D}|^2, \tag{29}$$

Since the early frames carry more information than the later ones, their values in R are typically significantly greater. Therefore, using this formulation, they automatically assume a greater weight in the estimation. The optimal weight for the regulariser itself, λ, depends on the specific hardware setup and cannot be measured from the data, however.

Geometric Refinement

Analogously to Bayesian Polishing, the log likelihood of a particle being observed at a position $\mathbf{s}$ is given by twice its cross correlation with the predicted image:

$$\log(P(X|\mathbf{s})) = 2\mathrm{CC}(\mathbf{s}) \tag{30}$$

$$\mathrm{CC} = \mathrm{IFT}(wXV^{(p)\dagger}) \tag{31}$$

$$w_{\mathbf{j}} = \frac{1}{\sigma_j^2}. \tag{32}$$

To keep the problem differentiable, the cross correlation CC is always accessed using cubic interpolation. After the inverse Fourier transformation, each such cross-correlation table is cropped to a smaller size to make storing many such tables feasible, and the memory throughput efficient. The size of the tables can be controlled by the user, and should be set to the maximal positional error expected in the data set.

The geometrical model that is optimised this way projects points $\mathbf{s} \in \mathbb{R}^3$ in the tomogram to 2D positions $\mathbf{p}_f \in \mathbb{R}^2$ in each tilt image:

$$\mathbf{p}_f = W(\mathbf{l}_f) \tag{33}$$

$$\mathbf{l}_f = J_f \begin{bmatrix} \mathbf{s} \\ 1 \end{bmatrix}. \tag{34}$$

The initial linear projection $\mathbf{l}_f$ is obtained by multiplying $\mathbf{s}$ with a 2×4 matrix J_f, and then optionally shifted by the non-linear image distortion W. The cost C_{align} that is being minimised consists of the sum over all (negative) cross-correlation values of all particles in all images plus all regularisers for all regularised parameters:

$$C_{\mathrm{align}} = -\sum_{p,f} \mathrm{CC}(\mathbf{p}_f - \mathbf{p}_f^{(0)}) + \mathfrak{R} \tag{35}$$

Although our framework supports arbitrary projection matrices J_f, our optimisation algorithm only looks for orthogonal rotations to the initial projection matrix. This is achieved by parametrising that rotation using Tait-Bryan angles, not Euler angles. The disadvantage of Euler angles is that they are gimbal locked in the initial configuration where all three angles are zero, i.e. the first and third angle refer to the same axis. The rigid alignment of the tilt image is never regularised, because we do not assume to have any prior information on it.

The distortion field W can take different forms. We have implemented models that express the distortion using the Fourier basis, a cubic spline basis and an affine linear one. The intended purpose of these deformations is to model distortions of the image that arise at the image forming lenses at the bottom of the optical system. An imperfect calibration of these lenses is likely to go unnoticed as long as the microscope is only used for single-particle analysis, because the same particle is never observed at starkly different positions during the collection of a single-particle movie. In tomography, a given particle may appear at any image position in any tilt image, so arbitrary deformations to the 2D image become relevant. We expect these deformations to be stable over time, so the intended purpose of the deformation field W is to model only one such deformation per tilt series. Optionally, we also allow the user to instead model a different deformation for each tilt image, but we have not encountered any data sets where this has produced an improvement. The deformation fields are optionally regularised by penalising the squared coefficients of the respective model. This limits the extent of deformation, and it forces the system to explain changes in position through particle motion, rather than image deformations.

The quantity that we do expect to change during the collection of the tilt series is instead the 3D position of the particle, $\mathbf{s}_f$. Analogously to Bayesian Polishing, we model this change as motion over time. The position in image f is given by the sum over its per-frame velocities $\mathbf{v}_f \in \mathbb{R}^3$ up to that point. Note that the velocity vector $\mathbf{v}_f$ refers to the motion between images f and $f+1$:

$$\mathbf{s}_f = \mathbf{s}_0 + \sum_{f'=0}^{f'<f} \mathbf{v}_f. \tag{36}$$

It is important to note that the tilt images are implicitly assumed to be ordered chronologically. In practice, this is usually not given, so the images are reordered by the program based on the cumulative radiation dose of each image.

As in Bayesian Polishing, the motion vectors themselves are expressed in a collective coordinate sytem for all the particles. This allows the spatial smoothness of a hypothetical motion field to be evaluated and used as a prior. For a more detailed derivation, we refer to the paper on Bayesian Polishing for single-particle analysis [28]. The formal details will be given in the following for the sake of completeness and to highlight differences to the original formulation.

The collective coordinate system for particle motion is obtained through a low-rank approximation of a Gaussian Process. This is done by constructing and then diagonalising the $P \times P$ covariance matrix S for a set of initial particle positions (where P is the number of particles). The entries of S contain the value of the following square-exponential covariance function for each pair of particles p and q:

$$S_{p,q} = \sigma_V^2 \exp(-|\mathbf{s}_p - \mathbf{s}_q|^2/\sigma_D). \tag{37}$$

Optionally, the user can instead also use the original formulation without the square inside the exponential:

$$S'_{p,q} = \sigma_V^2 \exp(-|\mathbf{s}_p - \mathbf{s}_q|/\sigma_D). \tag{38}$$

The former option forces particles in immediate proximity to move more similarly, but it allows for a greater discrepancy at greater distances. Both the single-particle and the tomography implementations allow the user to choose either function, but the default has changed from the latter to the former in tomography. This choice was motivated by both empirical observations and the fact that the square-exponential kernel produces fewer meaningful deformation components, which speeds up the optimisation for tomograms with very large numbers of particles.

We perform a singular-value decomposition of the covariance matrix S,

$$S = U \Lambda W^\intercal, \tag{39}$$

which allows us to construct the coordinate system as follows:

$$\mathbf{b}_i = \sqrt{\lambda_i}\mathbf{w}_i, \tag{40}$$

where $\lambda_i \in \mathbb{R}$ is the i^{th} singular value and $\mathbf{w}_i \in \mathbb{R}^P$ the corresponding singular vector. Basis vectors with very small λ_i are discarded here to speed up the optimisation. Let P' represent the number of remaining basis vectors. In this coordinate system, the set of all particle velocities in a tilt image $V_f \in (\mathbb{R}^3)^P$ can be expressed as $V_f = BQ_f$, where Q_f is a $P' \times 3$ coefficient matrix that encodes the velocity in three spatial dimensions. Note that the same basis B is shared between all three dimensions and all frames. In this coordinate system, the negative log-likelihood of a configuration of particle velocities is given by the Frobenius norm of the coefficient vector, $||Q_f||$, i.e. the sum of the squares of its entries. Therefore, the motion regulariser takes a very simple form:

$$\mathfrak{R}_{\text{motion}} = \sum_f ||Q_f||. \tag{41}$$

The acceleration regulariser that would penalise changes in velocity from one frame to the next in single-particle analysis has been omitted from tomography. This is because, unlike a single-particle movie, the tilt images are not collected in one continuous exposure. Since they are being exposed individually, there is no reason to assume that the particle velocities will be continuous between them. Two further differences to the original Bayesian Polishing are hidden in the notation: the covariance is now based on the 3D distances between the particles, and the coefficient matrix Q contains three columns instead of two.

As in the original Bayesian Polishing approach, the complete alignment of the tilt series is performed by finding parameters that minimise C_{align} from Eq. 35 using L-BFGS [13]. The set of parameters that are being optimised always includes the three Tait-Bryan angles for each tilt image and the set of initial particle positions. The latter are essential, because all the information we have about their 3D positions is derived

from the tilt images themselves, so changing the alignment requires the particles to be able to shift to more likely positions. Estimating the image deformations and particle motion is optional. If they are being estimated, then a set of deformation coefficients is fitted either to each tilt image or to each tilt series, while a set of motion coefficients is fitted to each image transition.

In addition to this local, L-BFGS based refinement, we also offer two methods to align only the 2D shifts of all tilt images globally. This means that instead of trying to obtain the optimal alignment through small changes to the initial one, we instead look for the best possible image shift overall, keeping all other parameters constant. This is helpful when individual tilt images are so badly aligned that a local optimisation cannot converge to the globally optimal position. Note that the initially assumed angles are rarely as incorrect as the image shifts, since the angles can be controlled more effectively through the experimental setup.

There are two variants to this method. If the sample contains few particles per tomogram, then the best results are obtained by predicting an entire micrograph and computing its cross correlation with the original one. The maximum value in that large cross-correlation image then indicates the optimal image shift. This approach can in theory deal with arbitrarily large misalignments. If the sample is very dense, however, then this whole-micrograph approach can fail. In that case, better results are obtained by instead adding up the small, per-particle cross correlation images defined in Eq. 31, and finding the maximum in that sum. This latter approach can only correct for misalignments smaller than half the box size of the particle, and it often produces inferior results on samples with few particles per tomogram.

References

[1] Alberto Bartesaghi, Federico Lecumberry, Guillermo Sapiro, and Sriram Subramaniam. Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure*, 20(12):2003–2013, 2012.

[2] T Bepler, A Morin, M Rapp, J Brasch, L Shapiro, A.J. Noble, and B Berger. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11):1548–7105, 2019.

[3] Tanmay A M Bharat, Danguole Kureisaite-Ciziene, Gail G Hardy, Ellen W Yu, Jessica M Devant, Wim J H Hagen, Yves V Brun, John A G Briggs, and Jan Löwe. Structure of the hexagonal surface layer on caulobacter crescentus cells. *Nat Microbiol*, 2:17059, Apr 2017.

[4] Tanmay AM Bharat and Sjors HW Scheres. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in relion. *Nature protocols*, 11(11):2054–2065, 2016.

[5] Georges Chreifi, Songye Chen, Lauren Ann Metskas, Mohammed Kaplan, and Grant J Jensen. Rapid tilt-series acquisition for electron cryotomography. *Journal of structural biology*, 205(2):163–169, 2019.

[6] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[7] Rafael Fernandez-Leiro and Sjors HW Scheres. Unravelling biological macromolecules with cryo-electron microscopy. *Nature*, 537(7620):339–346, 2016.

[8] Friedrich Förster and Reiner Hegerl. Structure determination in situ by averaging of tomograms. *Methods Cell Biol.*, 79:741–767, 2007.

[9] Timothy Grant and Nikolaus Grigorieff. Measuring the optimal exposure for single particle cryo-EM using a 2.6 å reconstruction of rotavirus vp6. *Elife*, 4, 2015.

[10] Matthew Herdman, Andriko von Kügelgen, Danguole Kureisaite-Ciziene, Ramona Duman, Kamel El Omari, Elspeth F Garman, Andreas Kjaer, Dimitrios Kolokouris, Jan Löwe, Armin Wagner, Phillip J Stansfeld, and Tanmay A M Bharat. High-resolution mapping of metal ions reveals principles of surface layer assembly in caulobacter crescentus cells. *Structure*, Nov 2021.

[11] James R Kremer, David N Mastronarde, and J Richard McIntosh. Computer visualization of three-dimensional image data using imod. *Journal of structural biology*, 116(1):71–76, 1996.

[12] Xueming Li, Paul Mooney, Shawn Zheng, Christopher R Booth, Michael B Braunfeld, Sander Gubbens, David A Agard, and Yifan Cheng. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods*, 10(6):584, 2013.

[13] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

[14] A Punjani, JL Rubinstein, DJ Fleet, and MA Brubaker. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods*, 14(3):290–296, 2017.

[15] Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, nov 2015.

[16] SHW Scheres, H Gao, M Valle, GT Herman, PBB Eggermont, J Frank, and JM Carazo. Disentangling conformational states of macromolecules in

3d-em through likelihood optimization. *Nature Methods*, 4(1):270–29, 2007.

[17] Sjors HW Scheres. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife*, 3, 2014.

[18] Florian KM Schur, Martin Obr, Wim JH Hagen, William Wan, Arjen J Jakobi, Joanna M Kirkpatrick, Carsten Sachse, Hans-Georg Kräusslich, and John AG Briggs. An atomic model of hiv-1 capsid-sp1 reveals structures regulating assembly and maturation. *Science*, 353(6298):506–508, 2016.

[19] J Smit, H Engelhardt, S Volker, S H Smith, and W Baumeister. The s-layer of caulobacter crescentus: three-dimensional image reconstruction and structure analysis by electron microscopy. *J Bacteriol*, 174(20):6527–38, Oct 1992.

[20] Dimitry Tegunov and Patrick Cramer. Real-time cryo-electron microscopy data preprocessing with warp. *Nature methods*, 16(11):1146–1152, 2019.

[21] Dimitry Tegunov, Liang Xue, Christian Dienemann, Patrick Cramer, and Julia Mahamid. Multi-particle cryo-em refinement with m visualizes ribosome-antibiotic complex at 3.5 å in cells. *Nature Methods*, 18(2):186–193, 2021.

[22] Beata Turoňová, Florian KM Schur, William Wan, and John AG Briggs. Efficient 3d-ctf correction for cryo-electron tomography using novactf improves subtomogram averaging resolution to 3.4 å. *Journal of structural biology*, 199(3):187–195, 2017.

[23] Andriko von Kügelgen, Haiping Tang, Gail G Hardy, Danguole Kureisaite-Ciziene, Yves V Brun, Phillip J Stansfeld, Carol V Robinson, and Tanmay A. M. Bharat. In situ structure of an intact lipopolysaccharide-bound bacterial surface layer. *Cell*, 180(2):348–358.e15, 01 2020.

[24] T Wagner, F Merino, M Stabrin, T Moriya, C Antoni, A Apelbaum, P Hagel, O Sitsel, T Raisch, D Prumbaum, D Quentin, D Roderer, S Tacke, B Siebolds, E Schubert, TR Shaikh, P Lill, C Gatsogiannis, and S Raunser. Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Commun Biol*, 2(1):2399–3642, 2019.

[25] William Wan, Larissa Kolesnikova, Mairi Clarke, Alexander Koehler, Takeshi Noda, Stephan Becker, and John A G Briggs. Structure and assembly of the ebola virus nucleocapsid. *Nature*, 551(7680):394–397, 11 2017.

[26] Shawn Q Zheng, Eugene Palovcak, Jean-Paul Armache, Kliment A Verba, Yifan Cheng, and David A Agard. Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods*, 14(4):331, 2017.

[27] Jasenko Zivanov, Takanori Nakane, Björn O Forsberg, Dari Kimanius, Wim JH Hagen, Erik Lindahl, and Sjors HW Scheres. New tools for automated high-resolution cryo-em structure determination in relion-3. *elife*, 7:e42166, 2018.

[28] Jasenko Zivanov, Takanori Nakane, and Sjors HW Scheres. A bayesian approach to beam-induced motion correction in cryo-em single-particle analysis. *IUCrJ*, 6(1):5–17, 2019.

[29] Jasenko Zivanov, Takanori Nakane, and Sjors HW Scheres. Estimation of high-order aberrations and anisotropic magnification from cryo-em data sets in relion-3.1. *IUCrJ*, 7(2):253–267, 2020.

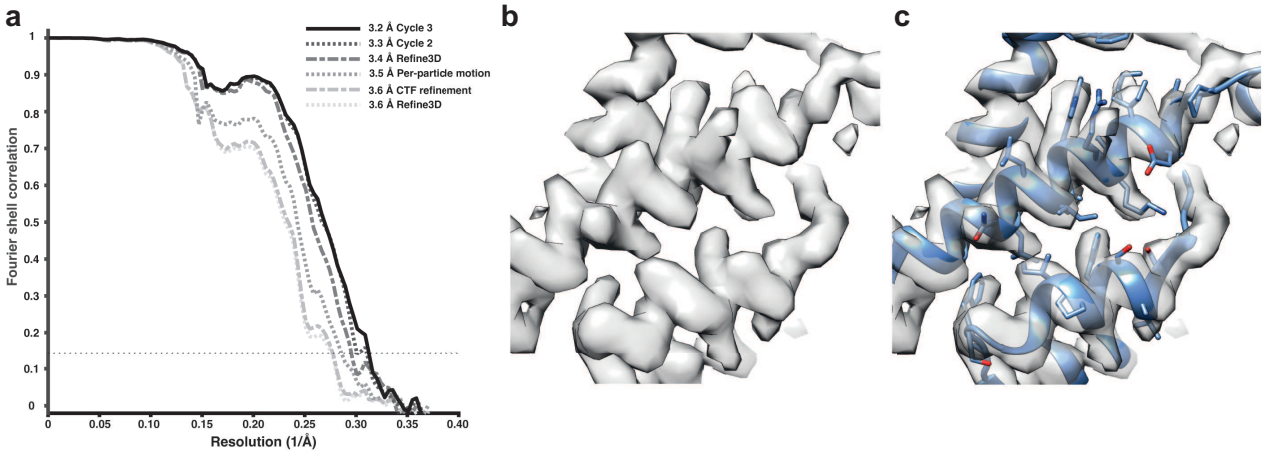

Figure 1: Subtomogram averaging of the HIV immature capsid. *(a) FSC for resolution estimation of iteratively improved reconstructions using the new RELION-4.0 workflow.* ***(b)*** *Representative region of reconstructed density in the final map.* ***(c)*** *The same density as in (b), together with the atomic model.*

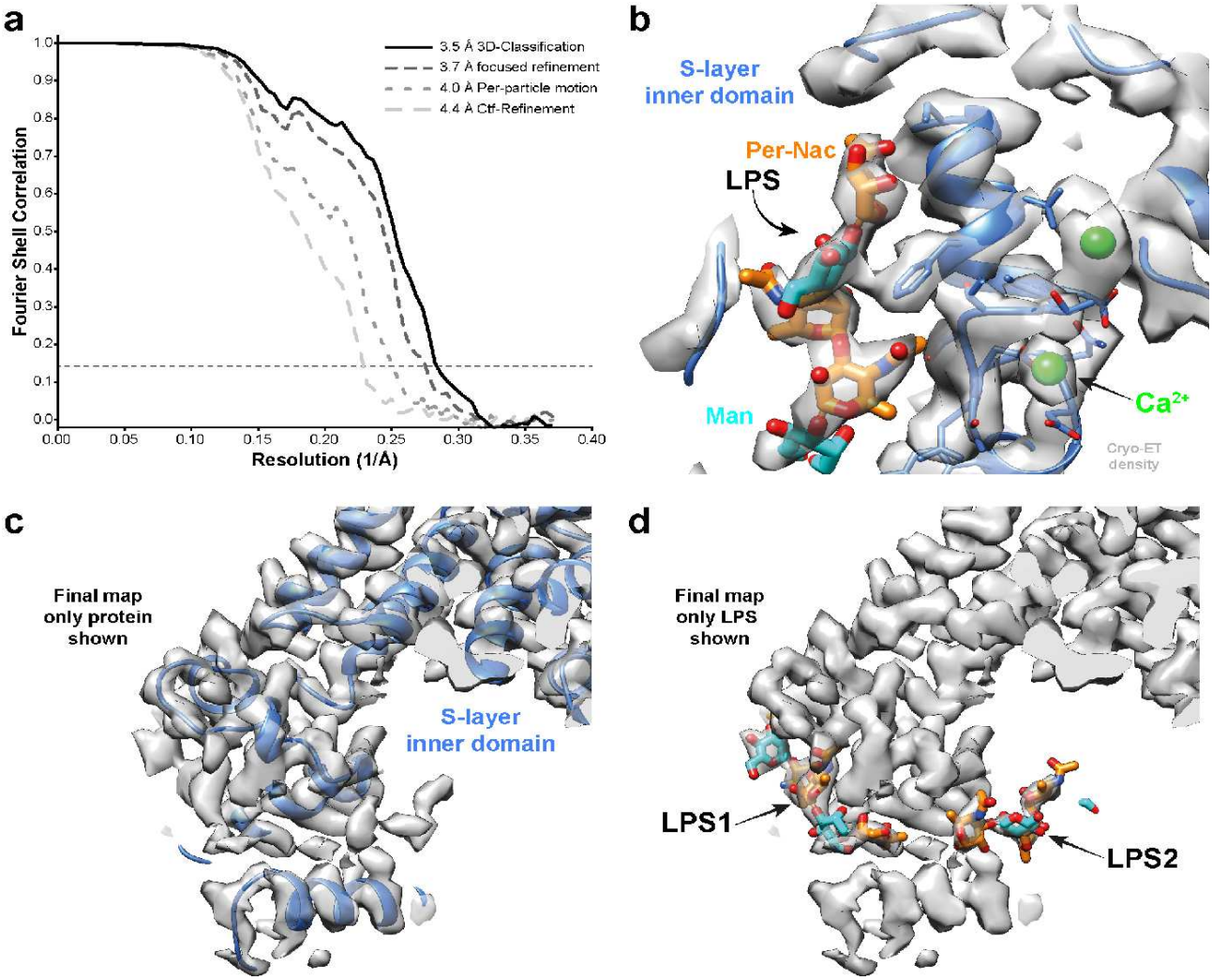

Figure 2: Subtomogram averaging of the* C. crescentus *S-layer from cell stalks. *(a) FSC for resolution estimation of iteratively improved reconstructions using the new RELION-4.0 workflow, tested on the S-layer inner domain.* ***(b)*** *Densities for the previously identified LPS (cyan and orange) and Ca^{2+} ions (green) in prior cryo-EM single particle analyses are resolved.* ***(c,d)*** *The final map shows two densities for bound LPS O-antigen chains. Panel (c) shows only the S-layer protein as blue ribbon and (d) shows LPS O-antigen as orange and cyan sugars corresponding to the N-acetyl-perosamine and mannose moeities respectively.*