Check for updates

Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells

Dimitry Tegunov [1 ✉], Liang Xue [2,3], Christian Dienemann[1], Patrick Cramer [1 ✉] and Julia Mahamid [2 ✉]

Cryo-electron microscopy (cryo-EM) enables macromolecular structure determination in vitro and inside cells. In addition to aligning individual particles, accurate registration of sample motion and three-dimensional deformation during exposures are crucial for achieving high-resolution reconstructions. Here we describe M, a software tool that establishes a reference-based, multi-particle refinement framework for cryo-EM data and couples a comprehensive spatial deformation model to in silico correction of electron-optical aberrations. M provides a unified optimization framework for both frame-series and tomographic tilt-series data. We show that tilt-series data can provide the same resolution as frame-series data on a purified protein specimen, indicating that the alignment step no longer limits the resolution obtainable from tomographic data. In combination with Warp and RELION, M resolves to residue level a 70S ribosome bound to an antibiotic inside intact bacterial cells. Our work provides a computational tool that facilitates structural biology in cells.

Cryo-EM[1] is a widely used method for macromolecular structure determination[2,3]. Two types of data are commonly analyzed to obtain high-resolution maps. First, samples are prepared at concentrations where individual particles can be distinguished in two-dimensional (2D) projections captured in a transmission electron microscope (TEM), and fractionated exposures at constant stage orientation (frame series) are typically acquired. Such data are then subjected to single-particle analysis (SPA). Second, samples containing multiple particles stacked along the projection axis, or samples that capture portions of crowded cellular environments, favor a tomographic approach to distinguish the particles in three dimensions. Here, the microscope stage is tilted to different angles between subexposures (tilt series). Each subexposure also comprises a frame series (tilt movie). Analysis of recurring structures in this data type has been implemented as subtomogram averaging (STA)[4–6].

In SPA, many noisy projections of similar particles observed under different orientations are iteratively aligned, classified and averaged to reconstruct three-dimensional (3D) maps of the macromolecules' Coulomb potential[7]. SPA refinement algorithms assume that each observation shows a single particle in isolation, and can thus be treated independently from other particles[8]. The same assumption is made in the closely-related STA workflow[9–11], where the reference of a single particle is aligned to each subtomogram and surrounding particles are treated as noise.

As samples are irradiated with electrons, beam-induced motion (BIM) leads to changes in particle positions and orientations[12]. If left uncorrected, these changes decrease the apparent image quality and limit the map resolution. Exposure fractionation into multiple frames captures the particles along their trajectories, allowing for accurate motion registration and the reversal of the detrimental effects of BIM[13,14]. Unfortunately, the granularity of the motion

model is limited by the low signal per particle. Although each particle's trajectory is unique, correlations exist on a local scale and can be used to regularize the motion model[13,15]. It is thus beneficial to exploit these correlations and treat the contents of a micrograph or tomogram as a multi-particle system embedded in the same physical space rather than isolated particles.

At the data preprocessing stage, the motion model can be fitted based on raw data using reference-free approaches[13,14,16–20]. Frame series are aligned in two dimensions, whereas tilt series are aligned and used to reconstruct tomograms. Extracted particles are fed into SPA or STA pipelines to obtain 3D references. Reference-based alignment can then improve the model accuracy by aligning the raw data to high-resolution reference projections. Such algorithms exist for both frame and tilt-series data[6,15,21,22], and improve the accuracy by enforcing local smoothness between particle trajectories on different spatio-temporal scales. However, most implementations remain different for frame and tilt-series data, and are limited to one reference species even in highly heterogeneous datasets. They are further decoupled from other parts of the refinement process, including rotational alignment and contrast transfer function (CTF) fitting, leading to a fragmented workflow and decreased convergence speed, limiting the final map resolution.

Here we present M, a software tool that integrates reference-based refinement of particle motion trajectories with other parts of the structure determination pipeline. We formulate our approach explicitly in a multi-particle framework, which simultaneously optimizes particle poses and hyperparameters describing physically plausible sample deformation within the entire field of view. This allows us to unify the processing of frame and tilt series, define a set of intuitive regularization constraints such as spatial and temporal resolution and include any number of particle species at different resolutions. Coupled with a robust approach to CTF correction

[1]Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. [2]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [3]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. ✉e-mail: dteguno@mpibpc.mpg.de; patrick.cramer@mpibpc.mpg.de; julia.mahamid@embl.de
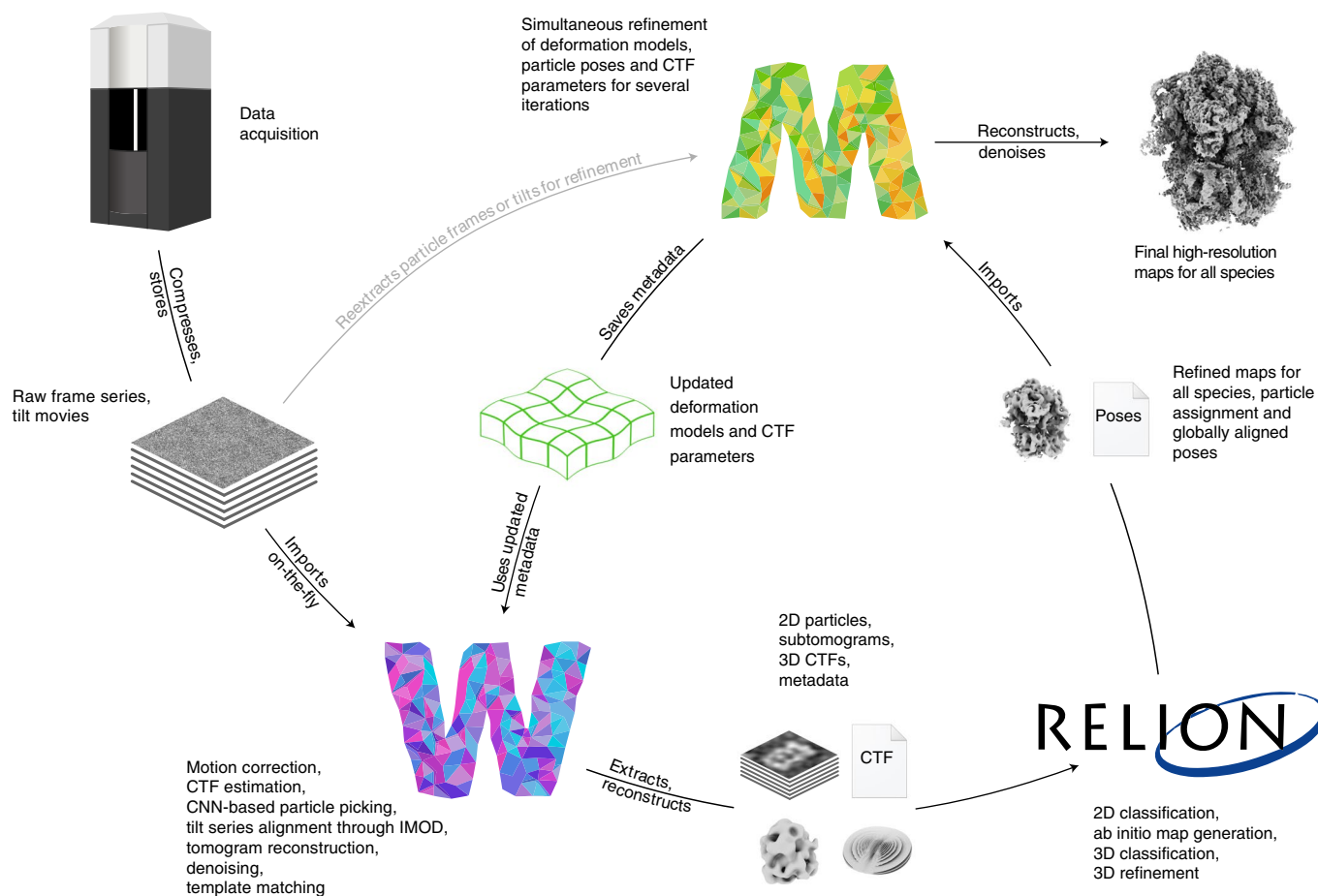
Fig. 1 | The Warp–RELION–M pipeline for frame and tilt-series cryo-EM data refinement. Electron microscopy data are preprocessed on-the-fly in Warp, which then exports particles as images or subtomograms. For tilt series, 3D CTF volumes containing the missing wedge and tilt-dependent weighting information are generated for each particle. Particles are imported in RELION where they can be subjected to a multitude of processing strategies, resulting in 3D reference maps, global particle pose alignments and class assignments. The particle population encompassing all classes is then imported in M, where reference-based frame or tilt image alignments are performed simultaneously with further refinement of particle poses and CTF parameters. Finally, M produces high-resolution reconstructions that can be used for model building. The improved alignments can be used in Warp to reexport particles for further, more accurate classification in RELION.

and with neural network-based map denoising, M achieves higher resolution on several datasets compared to other methods[6,21,23,24]. We demonstrate how various features of M contribute to these improvements, and achieve the same high resolution for frame and tilt-series data given similar numbers of particles. We also use M to visualize a 70S ribosome bound to an antibiotic in its native cellular context at residue-level resolution from tilt-series data.

Results
Overall design. M forms the last part of a cryo-EM data preprocessing and map refinement pipeline, preceded by Warp[19] and RELION[25] or compatible tools (Fig. 1). Warp performs initial reference-free motion correction and CTF estimation on frame series or tilt movies. For tilt series, Warp, starting with v.1.1.0, calls routines from IMOD[26] to perform tilt-series alignment, estimates per-tilt CTF using the tilt angles as constraints and reconstructs tomographic volumes. Warp then picks particles using a convolutional neural network (CNN) or template matching, and exports them as dose-weighted images or volumes depending on the data type. Particle poses and classes are then determined in RELION[27]. All classes are imported into M to perform a more accurate, reference-based, multi-particle frame or tilt-series refinement and obtain the final high-resolution maps. Optionally, improved

alignments can be applied to reexport particles for further classification in RELION, or to pick additional particles in tomograms.

M provides a graphical user interface that allows users to create, import, export and manage data. Projects are organized as 'populations', which contain 'data sources' and 'species'. A data source is a set of frames or tilt series that stem ideally from the same sample grid and acquisition session. A species is a distinct type of macromolecule, or its compositional and conformational substate. The refinement evolution is tracked as a directed graph, parts of which can be stored in different locations while remaining uniquely connected through cryptographic hashes.

Multi-particle system modeling. M considers the entire field of view as a physically connected multi-particle system (Fig. 2a). The particles can belong to different species, which can be of varying size, symmetry and resolution. The particles are subject to the same global transformations including stage translation and rotation, and locally correlated transformations caused by BIM. M performs a reference-based registration of these transformations (Fig. 2b), and reverses them when back-projecting individual particle images to obtain more accurate reconstructions.

In frame series, all transformations occur in the same image reference frame. Their combined effects are parametrized as a
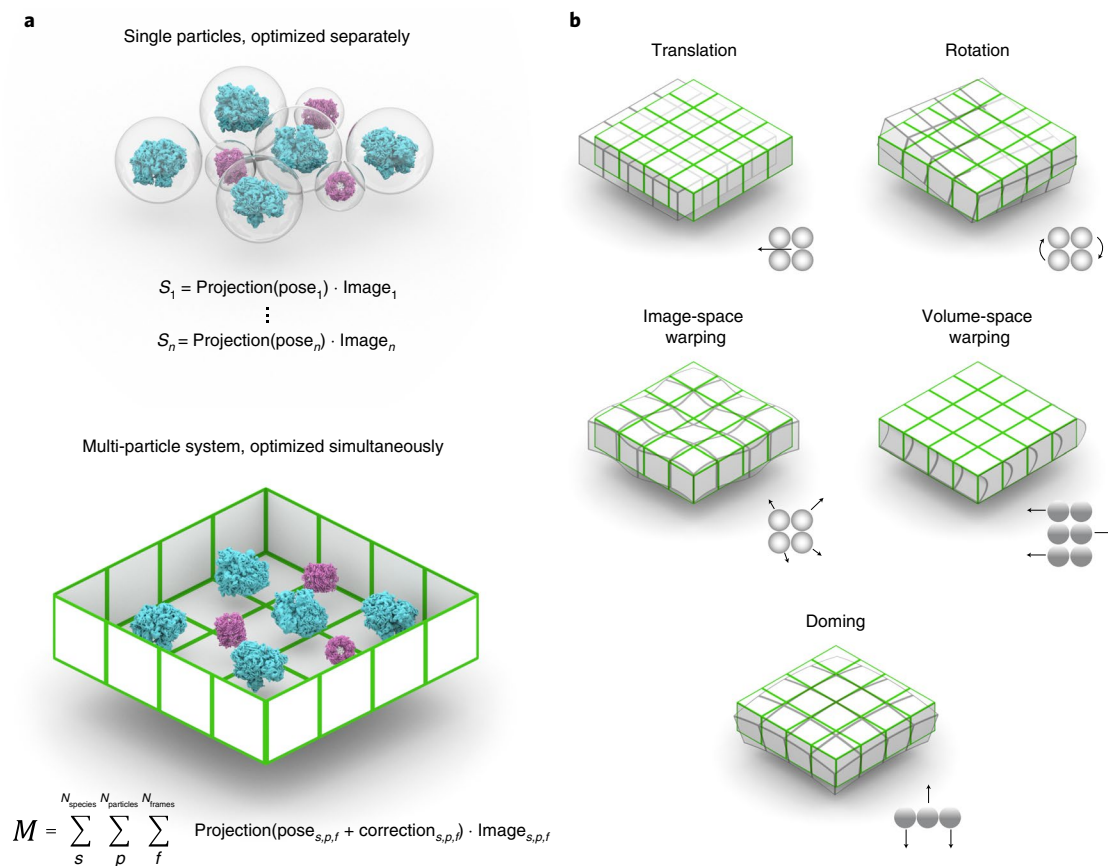
Fig. 2 | Multi-particle system modeling and optimization. a, Previous algorithms treat particles as isolated entities and optimize their poses using separate cost functions (top). In M's multi-particle refinement framework, all particles within the field of view are treated as parts of the same physical volume. Their poses and hyperparameters describing the beam-induced deformation of the volume are optimized simultaneously using a single cost function (bottom). **b**, The multi-particle system deformation model incorporates several modes: global movement and rotation to account for inaccuracies in stage movement between frames and stage rotation between tilts; image-space warping to model local nonlinear deformation in the 2D reference frame of a frame or tilt image; volume-space warping to model the movement of overlapping particles perpendicular to the projection axis (tilt series only) and doming to account for the hypothesized bending of a thin sample along the projection axis (frame series only).

pyramid of 3D cubic spline grids (Extended Data Fig. 1), to combine fast, global stage motion with slow, local BIM. This model is similar to Warp's reference-free alignment, but fits more parameters due to the increased signal of high-resolution references. In addition to image-space warping, M can fit doming-like motion[12] (Fig. 2b) implemented as parameter grids for defocus and orientation offsets.

For tilt series, M distinguishes image-space and volume-space effects. Additionally, a coarse model can be fitted for every tilt movie to account for the substantial deformation captured in each exposure. Volume-space transformations are resolved in 3D as a function of the accumulated exposure. Because M does not average particle frames or tilts in intermediate steps, per-particle translation and rotation trajectories can be fitted. The temporal resolution of the trajectories can be set for each species depending on the particle's size and thus the signal available per particle.

We show the benefit of considering the particles of multiple species in refinement using frame series of apoferritin (AF-f) (Methods). We artificially split the apoferritin population in two species comprising 5 or 95% of the particles (Extended Data Fig. 2a and Supplementary Table 1), and assumed no structural similarity between the two species during refinement. Refining the 5% species alone produced a 3.2 Å map, while adding the 95% species to the multi-particle system improved the map calculated from the 5% species to 2.8 Å (Extended Data Fig. 2b).

Correction of electron-optical aberrations. In addition to a geometric deformation model, M fits CTF parameters and higher-order aberrations including beam tilt. For frame series, defocus is optimized per particle, similar to cisTEM[23] and recent RELION versions[24]. For tilt series, defocus is optimized per tilt, similar to the capability offered in emClarity[6]. For both data types, astigmatism, anisotropic pixel size and higher-order aberrations are fitted per series.

CTF correction at high defocus can introduce artifacts if the chosen particle box size is too small to retain high-resolution Thon rings, leading to their aliasing (Extended Data Fig. 3a). M automates the selection of a sufficiently large box size at which the data are premultiplied by an aliasing-free CTF. The images are then cropped in real space. To match the underlying CTF of these images, correctly band-limited CTF[2] images are constructed in a similar way (Extended Data Fig. 3a) to be used for refinement and reconstruction.

We show the benefit of this approach by reconstructing a map from a high-defocus tilt series of HIV1 virus-like particles (EMPIAR-10164, Supplementary Table 1). Using a box size twice the particle diameter, the resolution is limited to 3.9 Å as the average sign error of the aliased CTF increases (Extended Data Fig. 3b). Premultiplying the data and CTF at sufficient size and then cropping them improved the resolution to 3.2 Å using the same reconstruction box size. Only premultiplying the data but using an aliased CTF[2] for the Wiener-like reconstruction filter did not decrease the
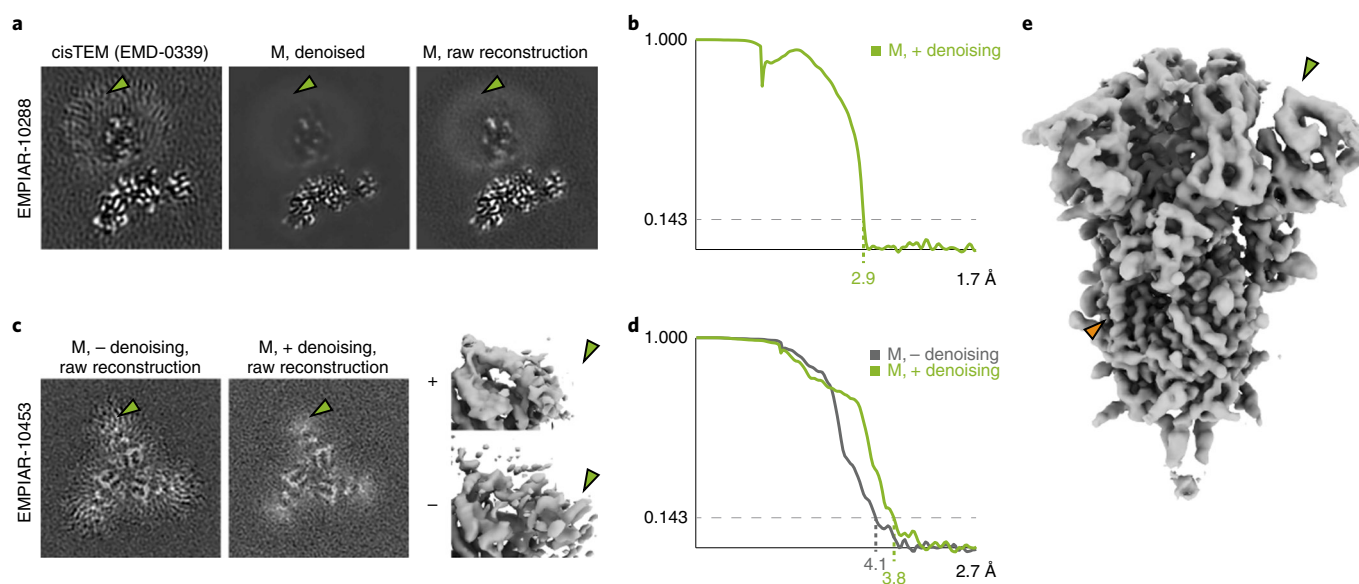
Fig. 3 | Effects of deep learning-based denoising of reconstructions during refinement. a, 2D *xy* slices through 3D reconstructions of the cannabinoid receptor 1-G membrane protein[35]. The original refinement in cisTEM (left) introduced artifacts in the highly disordered lipid region (green arrow). The denoised map (middle) and the raw reconstruction before denoising (right) used in the last refinement iteration in M using 149,308 particles (around 15% fewer than in the original study) are devoid of the artifacts because the denoising filtered and downweighed the low-resolution region. **b**, FSC between the half-maps independently refined in M, showing a global resolution of 2.9 Å. A value of 3.0 Å was reported in the original study, with no FSC curve included with the deposited map. **c**, 2D *xy* slices and isosurface renderings of the S1 domain in SARS-CoV-2 spike protein[36] reconstructions. Refinement in M without denoising introduced visible artifacts (left, bottom right) in the region (green arrows), which had lower resolution than the rest of the protein. Using denoising, the artifacts were avoided (center, top right). **d**, FSC between the half-maps refined in M with and without denoising, showing an improvement in global resolution from 4.1 to 3.8 Å when using denoising. **e**, Isosurface rendering of the entire denoised SARS-CoV-2 reconstruction with a global resolution of 3.8 Å. Through the denoising process, the more disordered S1 domain (green arrow) was filtered to lower resolution compared to other parts where side chains are visible (orange arrow).

nominal resolution in this case. However, for algorithms that would use aliased models during refinement and classification, we expect these effects to be noticeable. This approach improved the estimated per-tilt-series weighting factors for high-defocus data to the level of low-defocus data for the entire dataset (Extended Data Fig. 3c).

Optimization procedure. M optimizes all hyperparameters describing geometric deformation, electron-optical aberrations and particle pose trajectories, simultaneously by applying a gradient-descent optimization using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm[28]. The target function is the sum of normalized cross-correlations (NCCs) between all extracted particle images contained in a field of view, and reference projections at angles and shifts defined by the particles' poses and deformation hyperparameters (Fig. 2a).

At the end of an optimization iteration, similar to the Fourier ring correlation approach[29,30], M calculates the per-Fourier component NCC between reference projections and image data. This is used to optimize exposure- and tilt-dependent data weighting, and reconstruct half-maps using the updated model, correcting for Ewald sphere curvature[31]. Because the NCC is resolved in 2D, anisotropic weights can be fitted to make better use of the first frames, which are often affected by strong, unidirectional motion (Extended Data Fig. 4).

Map denoising and local resolution. Instead of using a traditional Fourier shell correlation (FSC) approach for local resolution estimation[32], M trains a CNN-based denoiser using a species' half-maps to filter them to local resolution for the next refinement iteration (Methods). The denoiser applies the noise2noise[33] training regime to independently refined[34] half-maps obtained at the end of each

iteration in M by back-projecting extracted images from the original frames or tilts. Because each half-map is denoised independently, no common artifacts are introduced and amplified over subsequent refinement iterations.

M's denoising was assessed on the cannabinoid receptor 1-G[35] dataset (EMPIAR-10288, Supplementary Table 1). The original 3.0 Å map (EMD-0339) showed overfitting artifacts in the lipid bilayer (Fig. 3a). Processing with Warp, RELION and M led to only slightly improved resolution of 2.9 Å (Fig. 3b), while removing the overfitting artifacts in M's final reconstruction (Fig. 3a).

Denoising was also tested on tilt series of SARS-CoV-2 virions (EMPIAR-10453, Supplementary Table 1). The S1 domain of the spike protein is conformationally heterogeneous and has notably lower resolution than the stable parts. Processing with Warp, RELION and M led to a 3.8 Å map (Fig. 3c–e), improving on the originally obtained[36] 4.9 Å. Repeating the refinement in M without denoising decreased the global resolution to 4.1 Å and generated visible overfitting artifacts in the S1 domain (Fig. 3c,d). This is in line with improvements recently demonstrated using different approaches to local filtering[37,38].

Contribution of different model parameters to map resolution. Apoferritin frame and tilt-series data collected from the same grid square under identical conditions (datasets AF-f and AF-t) (Methods), were used to estimate the contribution of different groups of hyperparameters to map resolution (Fig. 4 and Supplementary Table 1). For frame series, particles extracted following reference-free alignment in Warp and refined in RELION (without polishing and CTF refinement) provided a baseline resolution of 2.75 Å, which was improved by accumulating the following sets of optimizable parameters in M: the reference-based
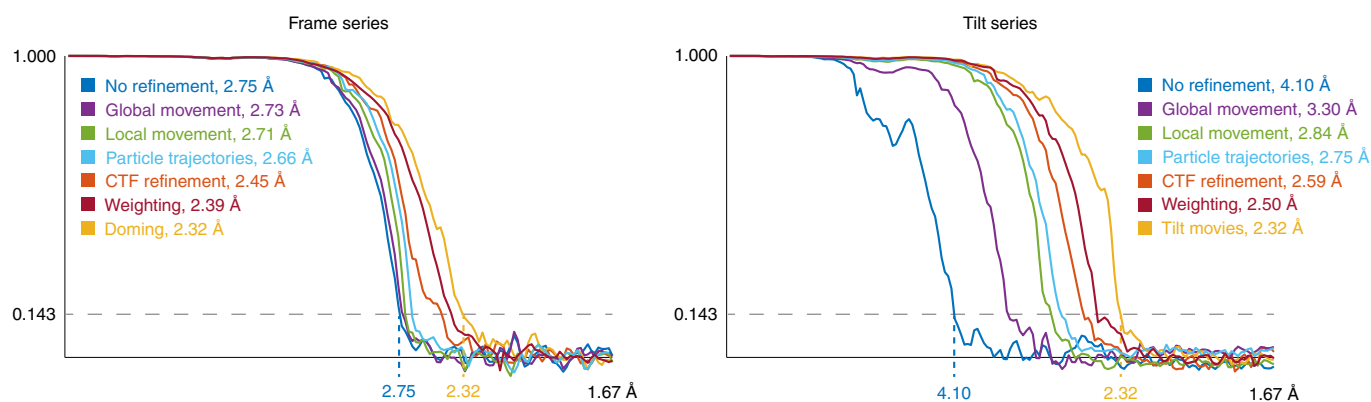
Fig. 4 | Contributions of individual multi-particle system model components to map resolution. FSC between half-maps for frame- and tilt-series apoferritin data obtained through extending the set of optimizable parameter groups. Starting with the 'No refinement' baseline, in top-down order in the legend, a new group of parameters was added, while keeping the previously added groups, and refinement was performed from scratch. The resolution for each step is given in the legend.
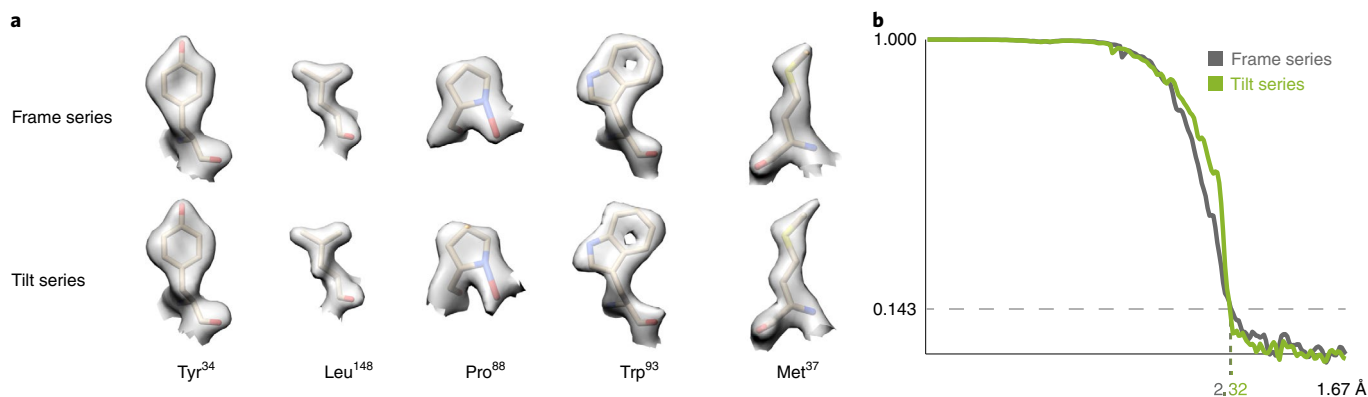

Fig. 5 | M achieves similar resolution for frame- and tilt-series data of an apoferritin sample. a, Representative side-chain densities observed in the frame- and tilt-series maps. **b,** Comparison between the global FSC curves for each map.

global motion alignment provided 2.73 Å; relaxing this constraint to allow local motion alignment provided 2.71 Å; resolving individual particle pose trajectories provided 2.66 Å; fitting per-particle defocus and per-frame-series astigmatism and beam tilt provided 2.45 Å; data-driven anisotropic weight estimation provided 2.39 Å and resolving doming-like motion provided 2.32 Å.

For tilt series, reference-free tilt movie alignment in Warp, patch tracking-based tilt-series alignment in IMOD, and refinement in RELION provided a baseline resolution of 4.1 Å, which was then improved by accumulating the following optimizations in M: reference-based global tilt image alignment provided 3.3 Å; relaxing this constraint to allow local image-space warping provided 2.84 Å; resolving individual particle poses provided 2.75 Å; fitting per-tilt defocus and astigmatism, and per-tilt-series beam tilt provided 2.59 Å; data-driven anisotropic weight estimation provided 2.50 Å and reference-based tilt movie alignment provided 2.32 Å. Volume-space warping was not tested because the particles were arranged in a single 2D layer.

We conclude that accurately registering image-space deformation is essential for obtaining high-resolution maps from frame and tilt-series data, whereas modeling other effects leads to smaller improvements that may only become important in the sub-5-Å resolution range. Initial reference-free alignment is less accurate for tilt series than for frame series. However, it allows obtaining initial reference maps and particle poses that can be further

refined in M. Given similar amounts of particles, M achieved the same resolution with very similar map features (Fig. 5) from either frame or tilt-series data. Thus, collecting data as tilt series does not incur a resolution penalty. However, because tilt series are slower to acquire[39] and commonly used for crowded, thick samples, we expect maps derived from tilt series to remain at lower resolution on average.

Comparison with RELION on atomic-resolution frame-series data. M's frame-series performance was assessed on apoferritin data previously processed with RELION 3.1 (EMPIAR-10248, Supplementary Table 1)[40]. The data acquired on a JEOL microscope with a cold-field-emission gun achieved an atomic resolution of 1.54 Å. At this resolution, we were able to assess the effect of Ewald sphere correction with the single side-band algorithm[31] (Extended Data Fig. 5). Applying it to the reconstruction alone, as done in RELION v.3.0, improved the resolution from 1.44 to 1.41 Å. Considering the sphere curvature during refinement, improved the resolution to 1.34 Å. Coupled with the demonstrated benefits of multi-species refinement and map denoising, this makes M a useful addition to the frame-series SPA pipeline.

The data's high resolution also enabled analysis of the sample's doming behavior, showing that the defocus of the entire field of view changed by over −25 Å during the first 7.5 e⁻ Å⁻² of exposure (Extended Data Fig. 6a), corresponding to the sample moving away
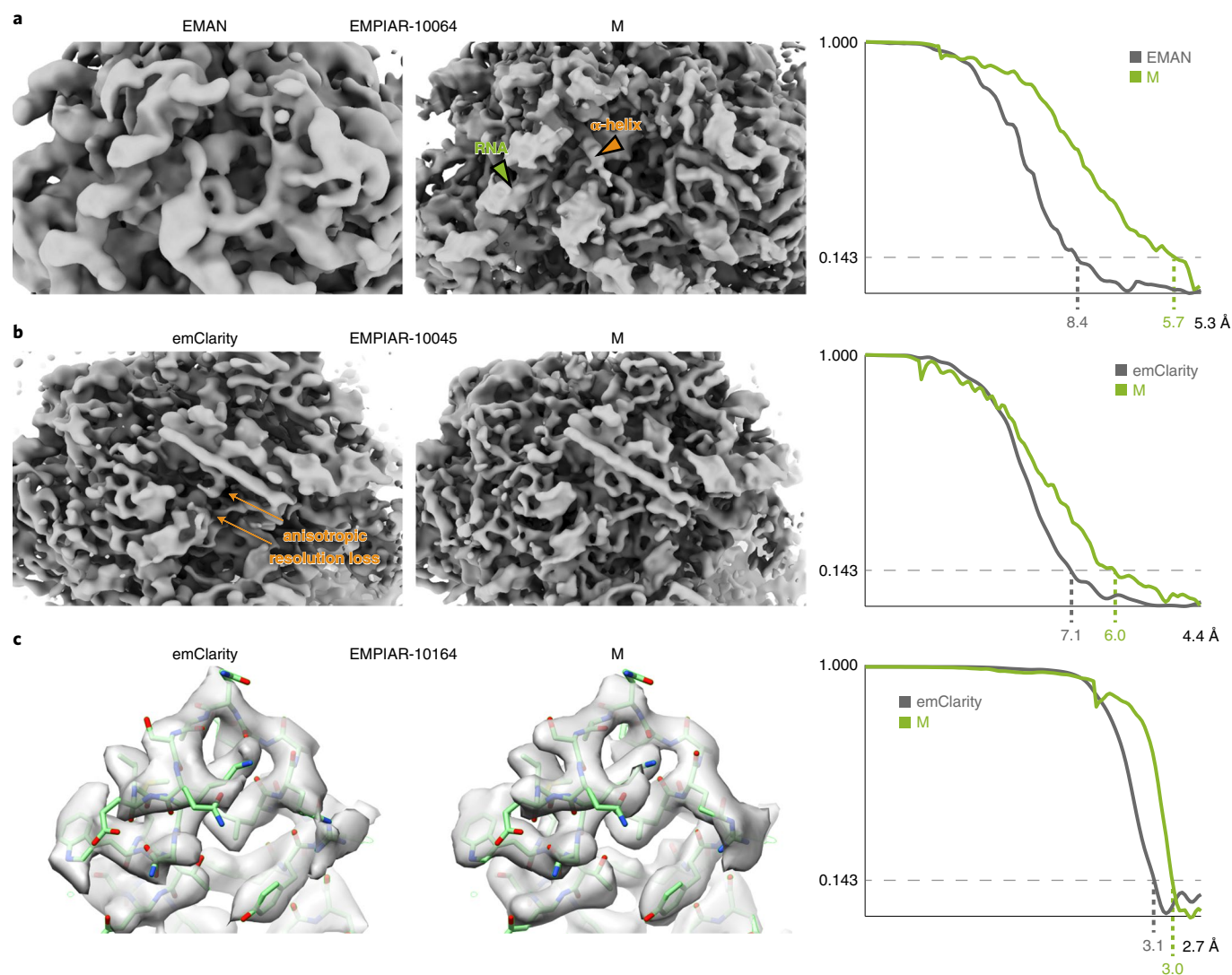
Fig. 6 | Comparison of maps obtained from published tilt series using M or other software. a, 80S ribosome data from EMPIAR-10064 were used to benchmark tilt-series processing in EMAN (EMD-0529). M achieved higher resolution, accompanied by visibly better resolved features such as RNA (green arrow) and α-helices (orange arrow). **b**, 80S ribosome data from EMPIAR-10045 were used to benchmark emClarity. The originally published map (EMD-8799, not shown) exhibited strong resolution anisotropy. A recently updated map[42] still suffered from resolution anisotropy ('smearing' direction indicated by orange arrows). M achieved higher and more isotropic resolution, aiding the map's interpretability. **c**, HIV-1 capsid-SP1 data from EMPIAR-10164 were used to benchmark emClarity (EMD-8986). M achieved slightly higher resolution with roughly 30% of the particle number used by emClarity. Doubling the number of particles did not increase the resolution. PDB 5L93 was rigid-body fitted into the maps for visualization.

from the electron source. A more localized, steadily increasing bending of the center relatively to the periphery followed, reaching a difference of $-16\,\text{Å}$ after $37.5\,\text{e}^-\,\text{Å}^{-2}$ (Extended Data Fig. 6b,c).

Comparison with other tools for tilt-series data refinement. M's performance on tilt series was compared with the EMAN2 (ref. [21]) and emClarity[6] packages on data used in the respective publications (Fig. 6 and Supplementary Table 1). EMAN2 reached a resolution of 8.4 Å on an in vitro 80S ribosome sample (EMPIAR-10064), while emClarity reached 8.6 Å for the same data[6], improving on a previous 13 Å result[41]. M improved the resolution to 5.7 Å and produced a map with secondary structure elements and helical RNA groves (Fig. 6a). We attribute some of this improvement to M's application of constraints between individual particle tilt images, which is absent in EMAN2.

emClarity reached a resolution of 7.8 Å on purified 80S ribosomes[6] (EMPIAR-10045), and was later improved to 7.1 Å[42] to

7.1 Å, surpassing the original 12.9 Å result[43]. M reached 6.0 Å, accompanied by improved resolution isotropy and map features (Fig. 6b). We attribute the improved isotropy to M's denoising-based filtering, whereas emClarity uses an FSC-based approach that may have to be tuned more conservatively to achieve the desired robustness.

emClarity also reached 3.1 Å on isolated HIV-1 capsid-SP1 assemblies (EMPIAR-10164), improving on previous[44] 3.9 and 3.4 Å results[45]. M reached 3.0 Å, accompanied by local improvements in map quality (Fig. 6c). We attribute the slight improvement to M's more accurate deformation model and simultaneous optimization of all parameters, in contrast to emClarity's separate steps for full image alignment and particle alignment.

M enables the visualization of an antibiotic bound to 70S ribosomes at 3.5 Å in cells. M's performance on tilt-series data of intact cells was assessed using data of chloramphenicol (Cm)-treated *Mycoplasma pneumoniae*[46] (Extended Data Fig. 7a). M refined
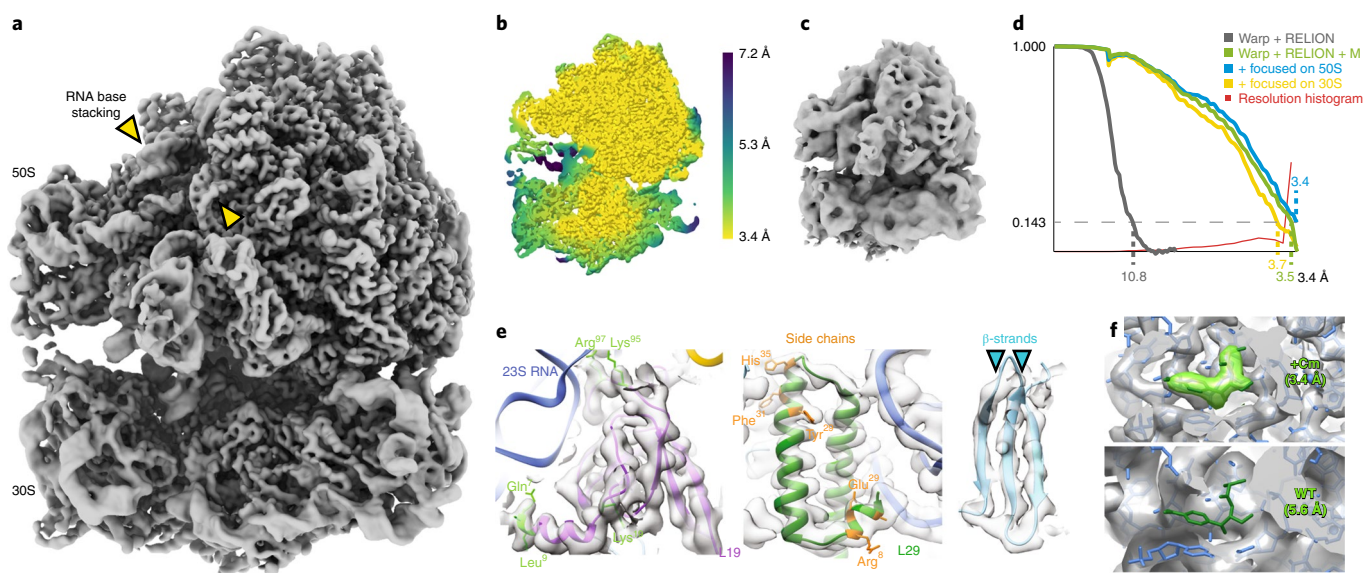
Fig. 7 | M. pneumoniae 70S ribosome-antibiotic map at 3.5 Å refined with the new Warp–RELION–M pipeline from tilt-series dataset of intact cells.
a, Isosurface representation of the 3.5 Å resolution map. **b**, Isosurface of the map colored by local resolution. Despite stalling of the ribosome that is induced by antibiotic binding, residual ratcheting leads to higher resolution in the large 50S subunit, which dominates the alignment, and lower resolution in the small 30S subunit. **c**, Isosurface of a 10.8 Å map derived from the same dataset using only Warp and RELION. **d**, FSC curves showing the resolution improvement achieved through global and focused refinement in M. The overlaid local resolution histogram shows that a large portion of the map is resolved close to the data's Nyquist limit of 3.4 Å. **e**, High-resolution features, such as large amino-acid side chains (in green and orange) and well-separated β-strands (cyan arrows), are resolved at a level expected for this resolution range. **f**, Atomic model of a Cm-bound 70S ribosome (PDB 4v7t) fitted into the 3.4 Å 50S map (top) shows correspondence of map density (light green) to the Cm molecule (dark green). Fitting the same model into a 5.6 Å 70S ribosome map of untreated *M. pneumoniae* cells (EMD-10683, bottom) does not show any density for Cm, providing a negative control.

the 70S ribosome to 3.5 Å (Fig. 7a,d) based on 17,890 particles from 65 tomograms, and a *B* factor[47] of 86 Å² (Extended Data Fig. 7b and Supplementary Table 1). The large 50S ribosomal sub-unit dominated the alignment and had a higher average resolution (Fig. 7b,d), with much of its core reaching the 3.4 Å Nyquist limit. Independent refinement of the 30S and 50S subunits improved the resolution to 3.7 and 3.4 Å, respectively (Fig. 7d). In contrast, processing these data with Warp and RELION alone led to a 10 Å map (Fig. 7c). M's result constitutes a dramatic increase in struc-tural detail (Fig. 7e); features typical for this resolution range, such as amino-acid side-chain stubs and individually resolved β-strands (Fig. 7e), are observed in the map. A rigid-body fit of an *Escherichia coli* 70S ribosome–Cm structure (PDB 4v7t) con-firmed the presence of the Cm molecule at its expected binding site (Fig. 7f), marking the first direct visualization of a drug bound to its target inside a cell. The density was absent in a reconstruction from untreated *M. pneumoniae*[46] (Fig. 7f). Therefore, tilt-series data of an intact, roughly 160 nm thick (Extended Data Fig. 7c), cellular specimen can lead to residue-level resolution structures of macro-molecules in their native biological context.

Discussion
Our results demonstrate that treating cryo-EM frame and tilt series as multi-particle systems rather than sets of isolated particles, and integrating their reference-based refinement with particle alignment and CTF refinement improves map resolution. The new framework removes previous technical limitations of tilt-series data processing, allowing the achievement of resolution on par with state-of-the-art frame-series results, provided similar amounts and quality of data. Because correlation between a 3D reference and a subtomogram was shown to be equivalent to the weighted sum of correlations between reference projections and the image series from which the subtomogram was reconstructed[48], we were able to

formulate the refinement process in M identically for frame and tilt series. Processing of image series rather than reconstructed sub-tomograms can lower the computational complexity[48] and enable flexible refinement of tilt angles and offsets.

Although M's refinement is constrained based on multi-particle assumptions, its forward model and reconstruction algorithms, as well as RELION's 3D classification, assume isolated particles. While this is rarely an issue for in vitro data, refinement of crowded cellular data stands to benefit from extending these algorithms as well. Future work on reconstruction and classification algorithms within M's flexible optimization framework may address this short-coming by modeling the multi-particle system explicitly to achieve higher resolution.

M's ability to resolve ribosomes in cells at a resolution previously considered exclusive to samples of isolated particles demonstrates that structures can be, in an ideal case, visualized directly inside cells at a resolution suitable to arrive to atomic models. The ribo-some is an outlier in terms of size and abundance in cells. Whereas smaller complexes may be refined to similar resolutions in prin-ciple (Fig. 6c), the number of such instances may be limited by the scarcity and heterogeneity of many complexes and the difficulty of localizing them in crowded cellular environments. Concentrating proteins in cells is likely to perturb the organism. The only way to overcome this is to collect more data; although sample preparation and data acquisition are becoming more streamlined, collecting enough particles of a rare protein complex to reach high resolu-tion may prove impractical for an individual research team. To help overcome this limitation, M offers data pooling and distrib-uted processing mechanisms to allow the community to share data and explore their potential. We show that including more particle species in a multi-particle refinement improves the resolution of all species involved. Thus, everyone stands to benefit from having more proteins identified and refined in shared data.

In conclusion, M can be combined with the established programs Warp and RELION into a powerful pipeline for cryo-EM data processing, that includes a comprehensive and transferable tilt-series workflow. This workflow avoids conversion of file formats and conventions between different software packages, enabling nonexpert users to achieve state-of-the-art results. It has the potential to achieve residue-level resolution maps of particles inside cells and to capture macromolecular machines in action within their native environment. Together with complementary approaches, it further establishes the foundation for the emerging field of high-resolution structural biology in cells.

Online content

Any methods, additional references, Nature Research 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/s41592-020-01054-7.

References

1. Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A. & Homo, J. C. Electron microscopy of frozen water and aqueous solutions. *J. Microsc.* **128**, 219–237 (1982).
2. Danev, R., Yanagisawa, H. & Kikkawa, M. Cryo-electron microscopy methodology: current aspects and future directions. *Trends Biochem. Sci.* **44**, 837–848 (2019).
3. Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **294**, 5181–5197 (2019).
4. Bharat, T. A. & Scheres, S. H. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protocols* **11**, 2054–2065 (2016).
5. Castaño-Díez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–51 (2012).
6. Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat Methods* **15**, 955–961 (2018).
7. Frank, J., Goldfarb, W., Eisenberg, D. & Baker, T. S. Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy* **3**, 283–290 (1978).
8. Frank, J. Single-particle reconstruction of biological molecules—story in a sample (Nobel Lecture). *Angew. Chem. Int. Edn* **57**, 10826–10841 (2018).
9. Knauer, V., Hegerl, R. & Hoppe, W. Three-dimensional reconstruction and averaging of 30S ribosomal subunits of *Escherichia coli* from electron micrographs. *J. Mol. Biol.* **163**, 409–430 (1983).
10. Oettl, H., Hegerl, R. & Hoppe, W. Three-dimensional reconstruction and averaging of 50S ribosomal subunits of *Escherichia coli* from electron micrographs. *J. Mol. Biol.* **163**, 431–450 (1983).
11. Leigh, K. E. et al. Subtomogram averaging from cryo-electron tomograms. *Methods Cell Biol.* **152**, 217–259 (2019).
12. Brilot, A. F. et al. Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **177**, 630–637 (2012).
13. 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).
14. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
15. Bartesaghi, A., Lecumberry, F., Sapiro, G. & Subramaniam, S. Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure* **20**, 2003–2013 (2012).
16. Mastronarde, D.N. in *Electron Tomography* (ed. Frank, J.) 163–185 (Springer, 2006).
17. Lawrence, A., Bouwer, J., Perkins, G. & Ellisman, M. Transform-based backprojection for volume reconstruction of large format electron microscope tilt-series. *J. Struct. Biol.* **154**, 144–67 (2006).
18. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
19. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152 (2019).
20. Fernandez, J.-J., Li, S. & Agard, D. A. Consideration of sample motion in cryo-tomography based on alignment residual interpolation. *J. Struct. Biol.* **205**, 1–6 (2019).
21. Chen, M. et al. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat. Methods* **16**, 1161–1168 (2019).
22. Zhang, L. & Ren, G. IPET and FETR: experimental approach for studying molecular structure dynamics by cryo-electron tomography of a single-molecule structure. *PLoS ONE* **7**, e30249 (2012).
23. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
24. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
25. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
26. 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).
27. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).
28. Nocedal, J. Updating quasi-Newton matrices with limited storage. *Math. Comput.* **35**, 773–773 (1980).
29. Saxton, W. O. & Baumeister, W. The correlation averaging of a regularly arranged bacterial cell envelope protein. *J. Microsc.* **127**, 127–138 (1982).
30. van Heel M., Keegstra W., Schutter W. & van Bruggen E. J. F. Arthropod hemocyanin structures studied by image analysis. In *Structure and Function of Invertebrate Respiratory Proteins, EMBO Workshop 1982* (ed. Wood E.J.) *Life Chem. Rep.* 69–73 (1982).
31. Russo, C. & Henderson, R. Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy* **187**, 26–33 (2018).
32. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
33. Lehtinen, J. et al. Noise2Noise: learning image restoration without clean data. In *Proc. 35th Intl Conf. Machine Learning, PMLR* **80**, 2965–2974 (2018).
34. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
35. Krishna Kumar, K. et al. Structure of a signaling cannabinoid receptor 1-G protein complex. *Cell* **176**, 448–458.e412 (2019).
36. Turoňová, B. et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203–208 (2020).
37. Ramlaul, 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).
38. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
39. Eisenstein, F., Danev, R. & Pilhofer, M. Improved applicability and robustness of fast cryo-electron tomography data acquisition. *J. Struct. Biol.* **208**, 107–114 (2019).
40. Kato, T. et al. CryoTEM with a cold field emission gun that moves structural biology into a new stage. *Microsc. Microanal.* **25**, 998–999 (2019).
41. Khoshouei, M., Pfeffer, S., Baumeister, W., Forster, F. & Danev, R. Subtomogram analysis using the Volta phase plate. *J. Struct. Biol.* **197**, 94–101 (2017).
42. Himes, B.A. change log, '2018-Nov-14' entry *emClarity Wiki* https://github.com/bHimes/emClarity/wiki (2020).
43. Bharat, T. A., Russo, C. J., Lowe, J., Passmore, L. A. & Scheres, S. H. Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. *Structure* **23**, 1743–1753 (2015).
44. Schur, F. K. et al. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506–508 (2016).
45. Turonova, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4A. *J. Struct. Biol.* **199**, 187–195 (2017).
46. O'Reilly, F. J. et al. In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554–557 (2020).
47. 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).
48. Sánchez, R.M., Mester, R. & Kudryashev, M. in *Image Analysis* (eds Felsberg, M. et al.) 415–426 (Springer International Publishing, 2019).

Methods

Data management. M requires data sources initialized based on a Warp project folder. Beside a list of frame/tilt-series items, it stores the deformation model to be refined. M saves the refined deformation model for each item in the same XML metadata files previously created by Warp. Due to a shared code base, Warp can use the updated model when calculating new frame-series averages or tomographic reconstructions. Multiple data sources of either type can be combined in a single population to facilitate the sharing and pooling of valuable datasets capturing complex cellular environments that can contribute to far more than one project, but do not contain enough data for any single project on their own. To account for minor pixel size miscalibrations between different microscopes, the pixel size can be refined alongside other parameters in M.

A species is initialized from the refinement results of RELION or other compatible software, taking the unfiltered half-maps, a mask and the particle coordinates and poses (that is, translations and rotations) as a starting point. The state of a species after each refinement iteration comprises the reconstructed half-maps, the weights of the trained denoising model, various filtered and sharpened maps, a denoised map and a list of particle coordinates and poses with multiple temporal sampling points if desired. Various map metrics, including global, local and anisotropic resolution, are calculated. The particles reference their data source items by their data hash to avoid naming conflicts between different data sources.

To enable multiple users to collaborate and pool their results, M tracks precisely the chain of refinements and other operations on data. After each refinement iteration, a 'commit' is generated to save the new state. Similar to version-control systems such as Git[49], the commit's hash is based on the exact state of the system committed. The hash of each data source item is calculated from the raw data, the refined deformation and imaging models, and the hashes of all species used for their refinement. The hash of each species is calculated based on the half-maps, the weights of the denoising model, the particle coordinates and poses and the hashes of all data source items contributing information. The hashes can be used to verify a graph representing all steps that led to a particular state of a data source or species. Similar to the 'pull request' mechanism in Git, species can be added to a population taking into account potential physical collisions with existing particles. This enables the maintenance of a centralized population repository from which multiple users can obtain prealigned data sources, identify new particle species or reclassify existing particles into more states and contribute the results back to the repository.

Deformation model. For frame-series data, deformation of the multi-particle system is modeled in the xy plane only, with a pyramid (Extended Data Fig. 1) of cubic spline grids[19] $G_{F,j}(\delta,i)$ (where j is the index within the pyramid, δ is the spatial interpolation coordinate and i is the temporal interpolation coordinate) going from high temporal/low spatial to low temporal/high spatial resolution. This accounts for the fast-changing, global stage movement, and the slowly-developing, local BIM. Furthermore, translation and rotation of individual particles as a function of exposure can be modeled with two to three control points depending on the particle size and overall exposure.

The model for tilt-series data is more complex, owing to the higher potential for perturbations in the system between individual tilt exposures. As the mechanical rotation of the microscope stage and the estimated orientation of the tilt axis are imperfect, the assumed stage orientation can be randomly off in every tilt. M thus refines an independent set of stage rotation angle corrections ω_i for every tilt i. These corrections only affect the particle orientations to avoid redundancy, as the induced changes in the projected particle positions can be fully modeled by a deformation grid that must already be used for other purposes.

Similarly, stage translation varies randomly between individual tilts. BIM patterns can be very different across adjacent tilt images as additional exposures are taken for focusing and tracking in between. Particle positions can further deviate due to other imaging artifacts, such as wrongly calibrated magnification anisotropy[50]. M uses an 'image warp' grid of cubic splines G_{TI} with a spatial resolution of 3–5 control points in x and y and per-tilt temporal resolution to model these geometric displacements in image space collectively. Furthermore, in vitro and in situ sample types for which tilt series are commonly used contain multiple overlapping layers of particles. Some deformations of densely filled volumes, such as shearing or bending in the z dimension when viewed at a high tilt angle, cannot be modeled accurately by xy translations in image space. M uses an additional 'volume warp' grid G_{TV}, implemented as a four-dimensional grid of control points with quadrilinear interpolation between them that is anchored in volume space rather than image space. Hence it rotates with the sample and can model slow, continuous deformation that affects the particles' projected positions in image space. As with frame-series data, per-particle translation and rotation as a function of exposure is also modeled for tilt series.

Finally, a single tilt image exposure is usually fractionated in multiple frames, making it a tilt movie. At 1–3 e⁻ Å⁻², the exposure in a single tilt movie is usually short, but still requires additional modeling to compensate motion. M parametrizes the xy translation as a combination of a grid with no spatial and per-frame temporal resolution, and a grid with a spatial resolution of 3×3 and a temporal resolution of three. Stage and particle orientations are assumed to remain constant

throughout a tilt movie, as the biggest beam-induced changes have been shown to occur in the very beginning of each of the short exposures[20]. Overall, the number of parameters for tilt series is larger than for frame series, requiring a higher particle density to achieve equivalent accuracy.

Imaging model. The ability to model imaging conditions such as defocus, astigmatism, magnification or higher-order aberrations is equally important for obtaining high-resolution reconstructions. Frame and tilt series offer different advantages for refining some of these parameters.

For particles in frame-series data, the z coordinate and thus the relative offset from the global defocus of the micrograph is unknown. Although local defocus estimation based on amplitude spectrum fitting has been shown to increase resolution[19], reference-based refinement of per-particle defocus can lead to a further increase in resolution[24]. M refines per-particle defocus and a per-series astigmatism for frame series, assuming constant values throughout the series.

Tilt series provide accurate z coordinates for all particles. However, the initial amplitude spectra-based global defocus estimates for each tilt have lower accuracy due to very short exposures, and cannot be assumed to remain constant throughout the series due to stage movement and refocusing. Furthermore, these estimates can be biased by contrast-rich objects that are not the particles of interest, such as a carbon film below or above the particles, or the platinum coating layer for FIB-thinned samples[51]. The astigmatism can also change between tilts due to fluctuating electron optics. M refines per-tilt defocus and astigmatism for tilt series, and calculates per-particle tilt CTFs based on these values and the z coordinate of a particle's position transformed according to the fitted stage orientation. Particles in tilt series can potentially have more accurate defocus values because the number of parameters that can be fitted scales with the number of tilts or particles for tilt or frame series, respectively. In many cases, the number of tilts will be substantially lower than the number of particles.

In both frame and tilt series, M also models per-series anisotropic magnification and higher-order optical aberrations. Refinement of a global set of Zernike polynomials representing the aberrations based on a 2D phase residual image calculated from all particles in a dataset has been shown to improve the resolution for slightly misaligned microscopes[52]. Within individual tilt series, beam tilt can vary as it is applied to compensate stage misalignments during tracking. Unfortunately, the signal in individual tilts is insufficient for accurate beam tilt estimation, and such an option is not implemented in M.

Optimization procedure. M seeks to maximize the following target function M, which is essentially a weighted, NCC between all particle images and the corresponding reference projections:

$$M = \frac{\sum_s \sum_p \sum_i A_{s,p,i} \cdot B_{s,p,i}}{\sqrt{\sum_s \sum_p \sum_i |A_{s,p,i}|^2 \times \sum_s \sum_p \sum_i |B_{s,p,i}|^2}},$$

$$A_{s,p,i} = W_i \times P(s, \Theta_{p,i}, \tau),$$

$$B_{s,p,i} = T \times \mathrm{FT}\big(\mathrm{FT}^{-1}\big(W_i \times \mathrm{CTF}(i, \Lambda_{p,i}) \times \mathrm{AS}_i^{-1} \times I(i, \Lambda_{p,i})\big) \times D(d_s)\big),$$

where s is a particle species, p is a particle of that species and i is the index of a frame or tilt in a series; $\cdot$ denotes the dot product between two complex vectors, where the complex numbers are treated as pairs of scalars; $|\ldots|$ denotes the L_2 norm; W is the anisotropic exposure- and tilt angle-dependent weighting of frame or tilt i; P is a projection operator in Fourier space sampling a central slice of the volume of species s at orientation Θ, taking into account the anisotropic scaling τ, bent to account for the Ewald sphere curvature determined by the species' diameter; $\times$ denotes scalar multiplication; T is the complex-valued beam tilt compensation; FT denotes the discrete Fourier transform; CTF is the real-valued CTF taking into account the defocus at position Λ and the astigmatism in frame or tilt i; AS is the real-valued, rotational average over the amplitude spectra of all particle images of all species extracted from tilt i or the average of all aligned frames, used for spectrum whitening, scaled and cropped to the respective species size and resolution; I is the Fourier transform of a particle image extracted from frame or tilt i at position δ, cropped to the respective species resolution, and D is a soft circular mask with particle diameter d.

Similar target functions in previous literature[23,25] used $P \cdot \mathrm{CTF}$ to model the contents of I. However, in M's implementation I is premultiplied by CTF to avoid CTF aliasing despite using small particle windows. This change does not affect the numerator part of M due to the associativity of complex number multiplication; its impact on the denominator part of M does not affect the achieved resolution in any way. It also avoids the additional memory footprint of storing precalculated CTFs, or the computational overhead of calculating them on-the-fly.

M can consider the Ewald sphere curvature during refinement if this is made necessary by a large species and/or high resolution[53]. In this case, two copies of $\mathrm{CTF} \cdot I$ are prepared using the single side-band algorithm[31]: $\mathrm{CTF}_P \cdot I$ and $\mathrm{CTF}_Q \cdot I$. To calculate the cost function, one is correlated with a bent central slice P, and the other with a central slice bent in the opposite direction. The resulting cost

functions M_P and M_Q are then added. As with previous implementations[24], the absolute handedness for the correction must be provided by the user.

For frame series, the position and orientation of particle p in frame i are calculated as:

$$\Lambda_{p,i} = \lambda_p(i) + \sum_j G_{OF,j}(\lambda_p(i), i) + \sum_j G_{F,j}(\lambda_p(i), i) + Z_p,$$

$$\Theta_{p,i} = \theta_p(i),$$

where λ is the value of the refined particle position trajectory interpolated at the accumulated exposure of frame i; G_{OF} is a deformation grid pyramid produced by Warp's original reference-free alignment that is not altered in M refinement; G_F is a deformation grid pyramid that is refined in M; Z is the refined defocus value of particle p that is added as the z coordinate to its position and θ is the value of the refined particle orientation trajectory interpolated at the accumulated exposure of frame i.

For tilt series, the position and orientation of particle p in tilt i are calculated as:

$$\Lambda_{p,i} = R(\Omega_i) \times (\lambda_p(i) + G_{TV}(\lambda_p(i), i) - C_V) + C_i + G_{TI}(\lambda_p(i), i) + Z_i,$$

$$\Theta_{p,i} = R^{-1}(R_{xyz}(\omega_i) \times R(\Omega_i) \times R(\theta_p(i))),$$

where R and R_{xyz} construct a rotation matrix based on a set of Euler and xyz angles, respectively, and R^{-1} calculates a set of Euler angles based on a rotation matrix; C_V is the center of the volume in which the multi-particle system is anchored and C_i is the center of the full tilt image; Z_i is the refined defocus value of tilt i that is added to the z coordinate of the transformed particle position; Ω is the stage orientation determined in the initial, reference-free tilt-series alignment that is not altered in M refinement and $\times$ denotes matrix multiplication here.

For frames in tilt movie i, the position of particle p in frame k is calculated as:

$$\Lambda_{p,k} = \Lambda_{p,i} + \sum_j G_{OF,i,j}(\Lambda_{p,i}, k) + \sum_j G_{TF,i,j}(\Lambda_{p,i}, k),$$

where G_{OF} is the deformation grid pyramid produced by Warp's original reference-free alignment of the tilt movie that is not altered in M refinement and G_{TF} is a deformation grid pyramid for the tilt movie that is refined in M.

Due to the very large number of parameters, M uses L-BFGS[28] to perform almost all of the optimization. Only the initial defocus search is done exhaustively over a limited range to avoid getting trapped in a local optimum because of the quickly oscillating nature of the CTF. Every L-BFGS search iteration requires the calculation of a partial derivative of the target function with respect to each optimizable parameter. Reevaluating M twice per parameter to compute the gradient with the central differences numerical scheme would be very computationally expensive. Like Warp, M takes a computational shortcut for most of the parameters.

Before optimization starts, M calculates the partial derivatives of the x and y components of all $\Lambda_{p,i}$ with respect to all warping grid parameters and all control points of a particle's position trajectory that affect them. Similarly, the partial derivatives of the individual Euler angle components of all $\Theta_{p,i}$ with respect to all stage angle correction parameters and all control points of a particle's orientation trajectory are calculated. As each parameter influences only a small fraction of particle frames or tilts, most of the derivatives are zero. They are excluded from the precalculated lists to avoid unnecessary computation. Then, during optimization, once per search iteration, the partial derivative of $(A \cdot B)/\sqrt{|A|^2|B|^2}$ for each particle frame or tilt is calculated with respect to x, y and the Euler angles. This amounts to evaluating M ten times. A useful approximation for the derivative for each parameter η can then be calculated as follows:

$$\frac{\partial M}{\partial \eta} = \frac{\sum_s \sum_p \sum_i \sum_\alpha \frac{\partial \left(A_{s,p,i} \cdot B_{s,p,i}/\sqrt{|A_{s,p,i}|^2|B_{s,p,i}|^2}\right)}{\partial \alpha} \times K_{s,p,i} \times |A_{s,p,i}| \times |B_{s,p,i}|}{\sum_s \sum_p \sum_i \sum_\alpha |A_{s,p,i}| \times |B_{s,p,i}|},$$

$$K_{s,p,i} = \frac{\partial(\Lambda_{p,i} \partial \Theta_{p,i})_\alpha}{\partial \eta},$$

where $\alpha \in (x, y, \phi, \vartheta, \psi)$, that is, one of the translation axes or Euler angles; $||$ denotes the concatenation of two tuples; $(\dots)_\alpha$ denotes the selection of component α from a tuple.

The deformation parameters make up the bulk of all parameters. Parameters such as absolute magnification and beam tilt do not benefit from the same shortcut and their derivatives must be calculated independently with the central differences scheme. The CTF-related parameters are few, but the calculation of their derivatives is especially expensive because it requires the particles to be reextracted at an aliasing-free size, premultiplied by the altered CTF and cropped to refinement size, all involving expensive Fourier transform steps. M calculates the values of M by adding up the results from small batches of particles. This allows the cost of the

first Fourier transform at aliasing-free size to be amortized over all optimizable CTF parameters, as its result is reused for all subsequent calculations. The gradients for all per-particle or per-tilt defocus and astigmatism parameters can all be calculated in the same pass as each of them affects only one particle or tilt.

If defocus is to be optimized, an iterative grid search can be executed before the L-BFGS optimization starts. The search runs for five iterations. For the first iteration, a range of ± 300 nm around the current values is sampled in 10-nm steps. For each subsequent iteration, the search step is halved and a range of plus or minus the new search step around the two best values for each particle or tilt from the previous iteration is sampled.

Memory footprint considerations. Traditional SPA refinement treats every particle as an isolated entity, thus requiring no more than one particle to be held in memory at any given time if parallelization is not considered. A multi-particle approach, however, needs to rapidly evaluate the state of the entire multi-particle system during refinement. The particle frame /tilt series need to be stored in memory because reextracting and reprocessing them for every evaluation would be too inefficient. While an in vitro sample usually contains a single layer of proteins with up to 1,000–2,000 particles in a field of view, a densely packed in situ volume has the potential to contribute tens of thousands of particles to refinement if enough species can be identified. The image size is selected to be twice the particle diameter to account for signal delocalization and interpolation artifacts, leading to substantial overlap even in the single-layer case. At high refinement resolution, the memory requirements of all extracted particle frame/tilt series in a system can vastly exceed those of the original data, rising to tens or even hundreds of gigabytes.

Although M uses GPUs for acceleration wherever possible, currently available consumer-level cards offer up to 12 GB, which would be insufficient in many cases. Therefore, the extracted particle frame/tilt series are held in 'pinned' (that is, page-locked) CPU memory where they can be transparently accessed by the GPU. Despite the low bandwidth of CPU–GPU memory transfers, the GPU does not experience a notable performance penalty when correlating them to reference projections. This is because the particle data accesses are sequential and highly coalesced, whereas the creation of reference projections on-the-fly accesses the GPU memory randomly, creating significant overhead. As faster CPU–GPU interfaces are being developed, the penalty should become more negligible in the future.

Still, memory requirements can become too high even for CPU memory. To reduce the footprint, M exploits the varying information content of frames/tilts over the course of a series. As sample damage from radiation is accounted for by applying a Gaussian (B factor) weighting function in Fourier space[14,24], the contribution of higher-frequency components becomes negligible at high exposure. M crops extracted particle images in Fourier space to a resolution that corresponds to the weighting function value falling below 0.25, resulting in considerable space savings once high resolution is reached. Assuming an increase in the weighting B factor of 4 Å^2 per 1 e$^-$ Å^{-2} of accumulated exposure, the maximum useful frequency at exposure d is $f_{max} = \sqrt{\ln(4)/d}$, and the image size m scales with a factor of $\min(1, f_{max}/f_{refine})$. Thus, the upper bound for memory consumption in case of low refinement resolution and/or low overall exposure is $O(m^2d)$, while the lower bound is $\Omega(m^2\ln(d))$ in case of high refinement resolution and/or high overall exposure.

Avoiding CTF aliasing. Cryo-EM data of thin biological specimens are usually acquired at defocus to achieve phase contrast. In the absence of a phase plate device, and often in the case of in situ tomography, defocus values can exceed 4 μm to enable better visual interpretation of the raw data. Higher defocus results in stronger delocalization of the signal in real space, as reflected by faster oscillations of the CTF in Fourier space. As the CTF oscillates between -1 and 1, combining signals with different defoci would result in an average value of zero at higher spatial frequencies. Thus, a phase shift of π must be applied to frequency components modulated by negative CTF values before averaging. Furthermore, it is desirable to compute the reconstruction as a weighted average, using the CTF for the weighting. Multiplying the Fourier transform of a particle image by the corresponding real-valued CTF achieves both goals.

Current SPA packages advise the user to select the particle box size as 1.5–2 the particle diameter to account for Fourier-space interpolation artifacts, not considering the image defocus. When an image is cropped around a particle, the Fourier-space modulation pattern becomes band-limited to the new window size. If CTF oscillations are too fast to be resolved, the band-limited values for the amplitudes of the corresponding frequency components will converge to zero. Even worse, the analytical 2D CTF model used in refinement and reconstruction is not band-limited, and contains solely aliasing artifacts past the Fourier-space Nyquist frequency instead of converging to zero. This can put a hard limit on the achievable resolution for small particles and those acquired at high defocus that is independent of the actual data quality.

This problem can be mitigated by selecting a box size large enough to avoid CTF aliasing[54] at the highest defocus value in a dataset. However, the required size m can exceed 1,000 px at high resolution or defocus, slowing refinement algorithms whose complexity and memory footprint are $O(m^2)$ and $O(m^3)$, respectively. This

increase can be entirely avoided by premultiplying particle images by the CTF at an aliasing-free size, and cropping them to a smaller size for refinement or reconstruction. As the modulation pattern is CTF^2 after premultiplication, the band-limited oscillations will converge to 0.5 instead of 0. The 2D CTF model used in refinement and reconstruction must be similarly band-limited to match the data. As M operates on all particles of an entire frame/tilt series at a time and extracts the particle images on-the-fly, such considerations are made automatically for the currently needed resolution.

The minimum box size needed for CTF correction at a given resolution is dictated by the maximum oscillation rate of the CTF within the available spatial frequency range. This is not necessarily the oscillation rate at the highest spatial frequency as φ is not a monotonic function; a combination of low underfocus and high C_s will cause the oscillations to slow and accelerate again at higher spatial frequencies. The oscillation rate can be calculated as the first derivative of φ. In practice, it is easier to evaluate $d\varphi/dk$ numerically within the relevant range of spatial frequencies to find its maximum absolute value. To fully resolve the oscillation, one period must be rasterized onto at least 2 pixels, that is, the window size must be chosen such that $\max(d\varphi/dk) = 2\pi/2px$. While this guarantees a fully resolved CTF in one dimension, a CTF rasterized on a Cartesian 2D grid has an anisotropic sampling rate. At its lowest, that is, along the diagonals, it requires $\sqrt{2}$ the sampling rate of the one-dimensional case.

Before particle extraction, the size padding factor at which the images will be premultiplied by the CTF has to be determined, taking into consideration the maximum defocus value expected in a frame/tilt series and the expected maximum resolution. During refinement, the latter is set to the refinement resolution. For the final reconstruction, it is set to 1.25× the current global resolution. Particles are extracted using the calculated minimum box size (or twice the particle diameter in case that value is larger), and premultiplied by the CTF in Fourier space. Then the inverse Fourier transform is applied, the particles are cropped to the refinement or reconstruction size in real space, and transformed back to Fourier space for refinement. The band-limited CTF^2 model is prepared by simulating the function at the same aliasing-free size in Fourier space, cropping its inverse Fourier transform in real space, and taking the real components of the result's Fourier transform.

Data-driven weighting. To account for radiation damage as a function of accumulated exposure, or increasing sample thickness as a function of the stage orientation, several heuristics and empirical approaches have been proposed[14,24,43]. By default, M adopts the heuristic introduced[43] in RELION 1.4. The B factor is increased by 4 $Å^2$ per 1 e^- $Å^{-2}$ of exposure, and each tilt is weighted as $\cos \vartheta$. Once high resolution is reached, the weights can be estimated empirically using a reference correlation-based approach similar to the one introduced[24] in RELION v.3.0.

In a departure from RELION's scheme, the normalized correlation is calculated between particle images and reference projections at the end of a refinement iteration are not combined across the entire dataset. It is kept as a 2D image to enable the fitting of anisotropic weights rather than averaging rotationally. The correlation data can then be recombined in different ways to calculate different kinds of weight. Furthermore, because M supports the refinement of multiple species with different resolution, the per-species correlation vectors for each frame or tilt need to be combined. This is done by weighting each one by the FSC calculated between the half-maps of the respective species. This produces a set of vectors $NC_{d,i,k}$, where d is the series, i is the frame or tilt and, optionally, k is the tilt movie frame.

The procedure then iteratively calculates $\overline{NC}$ as:

$$\overline{NC} = \frac{\sum_d \sum_i \sum_k NC_{d,i,k} \times G(\mathbf{B}_d + \mathbf{B}_i + \mathbf{B}_k) \times W_d \times W_i \times W_k \times \overline{CTF}_{d,i}}{\sum_d \sum_i \sum_k G(\mathbf{B}_d + \mathbf{B}_i + \mathbf{B}_k) \times W_d \times W_i \times W_k \times \overline{CTF}_{d,i}}$$

and optimizes the weighting parameters to minimize the following cost function:

$$C = \sum_d \sum_i \sum_k \left| NC_{d,i,k} - \overline{NC} \times G(\mathbf{B}_d + \mathbf{B}_i + \mathbf{B}_k) \times W_d \times W_i \times W_k \right|,$$

where × denotes scalar multiplication; G is an anisotropic 2D Gaussian B factor weighting function; $\mathbf{B}$ is a vector describing the B factor along the x and y axes and their rotation; W is a scalar weight and $\overline{CTF}$ is the weighted average of all particle CTFs in one frame or tilt. The B factors in each group are constrained such that the highest value in a group is set to zero.

In this default formulation, the weighting scheme allows to assign separate weights not only to individual frames/tilts, but also to weight the contribution of an entire series. For data with high particle density, this scheme can be extended to assign different weights to frames/tilts of each individual series. Anisotropic B factors improve the weighting of frames with substantial intra-frame motion (Extended Data Fig. 4). Combined with per-series, per-frame weighting, such granularity allows to rescue more information from the first few frames of an exposure if parts of them are less affected by BIM.

Map reconstruction. Previous refinement packages took two different approaches to map reconstruction from frame- and tilt-series data. For frame series, weighted averages were prepared either directly from the initial, reference-free alignments or were based on a 'polishing' procedure[24]. These 2D averages were then weighted based on a 2D CTF model and a spectral signal-to-noise ratio term[25], and back-projected to obtain the reconstruction. For tilt series, the algorithms operated on intermediate per-particle 3D reconstructions (subtomograms) with fixed translational and rotational offsets between individual tilt images. These 3D subtomograms were then weighted based on a 3D CTF model[43] and a spectral signal-to-noise ratio term, and back-projected to obtain the reconstruction.

M seeks to unify the handling of both types of data and uses the original, noninterpolated 2D data at every step, including reconstruction. For tilt series, this approach avoids any artifacts from intermediate interpolation and reconstruction steps. For frame series, the requirement for identical orientation of all particle frames no longer exists as they are not averaged in 2D, enabling the modeling of particle orientation as a function of exposure. Only for individual tilt movie frames a shortcut is taken to save memory and computation, and they are preaveraged in 2D using the approach described for Warp[19] after a separate multi-particle refinement of the respective tilt movie.

Thus, for the reconstruction, individual particle frames or tilts are weighted by an exposure-dependent function to account for radiation damage, and an aliasing-free 2D CTF model (Data-driven weighting) that incorporates the exact defocus and astigmatism values for that position and frame/tilt. The weighted data are then back-projected through Fourier-space summation, accounting for Ewald sphere curvature. The reconstruction is finalized by dividing the summed data component by the summed weights component[25].

Map denoising. Reconstructions of biological specimens derived from cryo-EM data rarely have homogeneous resolution throughout all parts of the macromolecule. Using a map filtered to its global resolution for particle alignment can have detrimental effects. Poorly resolved regions, such as floppy protein domains or the lipid bilayer around transmembrane domains, will make the alignment worse by adding noise to reference projections below the refinement resolution. In the case of fully independent half-maps[34], the noise patterns that the particles will be aligned against are independent, and amplifying them over several iterations only has the potential of making the resolution worse. In the case of refinement with merged half-maps[23] where overfitting is avoided by limiting the refinement resolution, the poorly resolved regions may be well below that limit, leading to a common, overfitted noise pattern in both half-maps.

Past attempts at filtering maps based on local resolution estimates for refinement[55,56] applied FSC-based approaches[32] to estimate the local resolution and performed the filtering in the Fourier domain. As only one set of estimates can be made based on one pair of half-maps, any spurious patterns in the estimated values will be introduced into both half-maps when the filtering is performed. The locality and accuracy of the estimates depends on the window size[32]. A smaller window increases locality at the expense of accuracy. Once introduced, the noise pattern can become amplified over multiple iterations, leading to overestimated local resolution and phantom features that can be misinterpreted. More advanced regularization schemes have been proposed[37,38] since to deal with this problem.

M implements a new approach to map filtering that uses neural network-based denoising. The recently proposed noise2noise training principle[33] allows the training of differentiable denoiser models without a noise-free ground truth, using only two independently noisy observations. It has been successfully applied to micrograph[19] and tomogram[19,57] denoising. The implementation in M uses gold-standard[34] half-map reconstructions, which represent another obvious case of two independently noisy observations of the same signal, and are interchangeably used as input and target in training. The reconstructions are obtained at the end of each refinement iteration in M by back-projecting extracted images from the original frames or tilts, using the particle half-sets carried over from RELION at the beginning of the workflow. We find that a denoiser trained on one pair of half-maps not only matches closely the result of conventional global resolution filtering when applied to maps with homogeneous resolution, but also provides locally smooth, artifact-free local resolution filtering. As such models can train on and denoise sets of micrographs or tomograms with different defocus values and thus different noise models, they can also recognize and adapt to different noise levels within the same reconstruction. In another important departure from FSC-based methods, the denoising step is applied to the half-maps independently and the denoiser sees only one of them at a time. Thus, even if some spurious pattern is introduced as part of the denoising, it is independent between the half-maps.

The neural network architecture, implemented in TensorFlow v.1.10, is identical to the one used for tomogram denoising in Warp. A separate denoising model is maintained for every species and trained only on the respective pair of half-maps. The model is initialized with random values and trained for 800 iterations upon the creation of a new species. It is later retrained for another 800 iterations after every refinement. Spectrum whitening is applied to the maps before training to restore high-frequency amplitudes[23], similar to B-factor-based sharpening[47]. During training, 64^3 pixel volumes are extracted from both maps at the same random position and orientation, and presented to the network as input and output in mini-batches of three. The random orientations make sure the network learns the noise model rather than merely learning the average map.

The learning rate for the Adam optimizer is exponentially decreased from 10^{-3} to 10^{-5} throughout the training. For the denoising of each half-map, the map is partitioned in 64^3 px windows overlapping by 24 px, denoised, and the results from each window are inserted into the output volume. Regardless of regions with above-average resolution being potentially present, the refinement resolution is set conservatively to the global map resolution. In addition to the two half-maps for refinement, a denoised average map is also prepared by applying the same denoising model to the average of the spectrum-whitened half-maps.

Assessment of map denoising. Frame-series data were downloaded for the EMPIAR-10288 entry (Fig. 3a,b). Frame alignment and local CTF estimation were performed in Warp with a spatial resolution of 5×5. Then, 1,033,994 particles were picked with a retrained BoxNet model in Warp and exported at 1.5 Å px^{-1}. The 2D classification, 3D classification and refinement were performed in RELION using EMD-0339 as the initial reference. Next, 149,328 particles corresponding to the best 3D class were imported in M. The particle poses were given a temporal resolution of 2, the deformation grid resolution was set to 2×2, and refinement of all parameters was performed for five iterations (Supplementary Table 1). Data-driven weight estimation was performed to assign unique weights to every frame index.

Prealigned tilt movies were downloaded for the EMPIAR-10453 entry (Fig. 3c,d). Gold fiducials were picked with BoxNet in Warp, and fiducial-based tilt-series alignment was performed in IMOD. Tilt-series CTF estimation and reconstruction of full tomograms at 12 Å px^{-1} was performed in Warp. A binary classifier based on a 3D CNN (in development, not part of Warp and M) was trained using five manually segmented tomograms to segment the SARS-CoV-2 virions. Another 3D CNN-based binary classifier was trained on manually picked spike protein positions in seven tomograms. Automatically picked spike protein positions were cross-referenced with the segmented virions to remove particles further away than 200 Å, obtaining 38,742 particles. Subtomograms were reconstructed at 5 Å px^{-1} for refinement in RELION. After ab initio map generation, 3D refinement was performed, reaching the 10 Å Nyquist limit. The results were imported in M, where a $1 \times 1 \times 41$ image warping grid and particle poses were optimized for two iterations. Subtomograms were reconstructed at 5 Å px^{-1} using the improved alignments, and subjected to classification into four classes in RELION. 22,998 particles from two classes showing the spike trimer were imported in M, where a $3 \times 2 \times 41$ image warping grid, CTF and particle poses were optimized for four iterations with C3 symmetry (Supplementary Table 1). For the comparison, the refinement procedure was modified to omit the denoising step. Refinement was then restarted at 10 Å and performed for five iterations using the same settings.

Acquisition of apoferritin benchmark data. To compare the resolution achievable with frame and tilt-series data and assess individual algorithms implemented in M, we acquired two datasets of human heavy-chain apoferritin: AF-f (frame series) and AF-t (tilt series). To make sure that any observed differences came from data type and processing strategies rather than local variance in sample quality, neighboring holes within the same grid square were used for both datasets.

GST-tagged apoferritin was overexpressed in *E. coli*, captured on gluthathione-sepharose beads after cell lysis, cleaved off the resin by tobacco etch virus (TEV) protease and purified to homogeneity by size exclusion chromatography in 50 mM Tris-HCl pH 7.5, 100 mM NaCl and 0.5 mM TCEP.

Then, 3 μl of apoferritin at 3.8 mg ml^{-1} were applied to freshly glow discharged R 1.2/1.3 holey carbon grids (Quantifoil) at 4 °C and 100% relative humidity followed by plunge-freezing in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). The sample concentration resulted in a dense, single-layered hole coverage. Data were collected on a Titan Krios TEM (Thermo Fisher Scientific) operated at 300 kV and a magnification resulting in a calibrated pixel size of 0.834 Å. The energy filter (Gatan) was operated in zero loss mode with a slit width of 20 eV. The K3 direct electron detector (Gatan) was operated in counting mode with a freshly acquired reference for gain correction. The exposure rate was adjusted to 20 e$^-$ px^{-1} s^{-1}. SerialEM[58] was used for frame and tilt-series acquisition.

Positions for both datasets were selected to be distributed evenly over the same grid area to maximize the similarity in ice thickness and particle density. For AF-f, 150 frame series were collected with a total series exposure of 32 e$^-$ Å^{-2}, fractionated in 40 frames. For AF-t, 135 tilt series ranging from -40 to $+40°$ were collected in a grouped dose-symmetric scheme[59] with a group size of two and in 2° steps. Each tilt was exposed to 2.7 e$^-$ Å^{-2}, fractionated in three frames.

Comparison between frame and tilt-series performance. Using dataset AF-f, frame-series alignment and local CTF estimation were performed in Warp with a spatial resolution of 8×5, owing to the rectangular format of the K3 chip. Next, 22,122 particles were picked with a retrained BoxNet model in Warp and exported at full resolution in 512 px boxes. Global 3D refinement with octahedral symmetry was performed in RELION v.3.0. The results were imported in M. The particle poses were given a temporal resolution of 3, the deformation grid resolution was set to 6×4, and refinement of all parameters was performed for five iterations (Supplementary Table 1). Data-driven weight estimation was performed to assign unique weights to every series and frame index.

Using dataset AF-t, tilt movie frame alignment was performed in Warp using a model without spatial resolution. Initial tilt-series alignment was performed in IMOD using patch tracking on 6× binned images with default settings. Tilt-series CTF estimation was performed in Warp. Then 18,991 particles were picked using Warp's 3D template matching in full tomograms reconstructed at 10 Å px^{-1}. Subtomograms and 3D CTF volumes were exported at 2 Å px^{-1} using 140 px boxes. Global 3D refinement with octahedral symmetry was performed in RELION v.3.0. The results were imported in M. The particle poses were given a temporal resolution of three, the image warp grid resolution was set to $6 \times 4 \times 41$, and refinement of all parameters was performed for five iterations, including tilt movie frame alignment in the last two iterations (Supplementary Table 1). Data-driven weight estimation was performed to assign unique weights to every series and tilt index.

Assessment of multi-species refinement. Particles from each frame series of the AF-f dataset were split in 5 and 95% subpopulations, resulting in species with 3,710 and 70,497 particles, respectively. Frame alignments and particle poses previously obtained from Warp and RELION were reused. In the first scenario, the 5% species was refined alone. In the second scenario, the 5% species was corefined with the 95% species. Both species were assumed to be structurally independent and did not contribute particles to each other's reconstructions. For both tested scenarios, a 6×4 starting grid for the deformation was used, the resolution of all species was set to 4.0 Å and only one refinement iteration was performed in M to avoid possible benefits from the higher resolution the 95% species would reach after the first iteration.

Comparison with RELION on atomic-resolution frame-series data. Frame-series data were downloaded for the EMPIAR-10248 entry and preprocessed in Warp. Then, 109,437 particles were exported at 0.6 Å px^{-1} using 466 px boxes and refined in RELION. The resulting particle poses and half-maps were imported in M and refined for five iterations starting with a resolution of 3.0 Å in the first iteration. A starting grid of 4×4 was used for the deformation model, and the number of frames was truncated to 25. All CTF-related parameters were refined, including doming, per-series beam tilt and a 3×3 grid model for local astigmatism (Supplementary Table 1). For the last two iterations, anisotropic per-series and per-frame *B* factor weights were estimated. The final iteration was completed in around 24 h, using four GeForce 2080 Ti GPUs. The original mask deposited with EMD-9865 was used to estimate the final resolution.

To analyze the doming behavior, fitted doming model parameters were averaged across the dataset. Because doming was fitted after per-particle defocus, which was dominated by frames 3–4 due to weighting, the values were normalized by subtracting those of frame 1 from all. As a larger, planar inclination spanning the field of view was observed in the fits in addition to the more local bending of the center relative to the periphery, a plane was fitted into each frame's values and subtracted from them before quantifying the doming.

Comparison with other tools for tilt-series data refinement. Tilt-series data were downloaded for the EMPIAR-10064 entry. Initial tilt-series alignment was performed in IMOD using manually picked gold fiducials on 4× binned images with default settings. Tilt-series CTF estimation was performed in Warp. Next, 3,566 particles were picked using Warp's 3D template matching in full tomograms reconstructed at 10 Å px^{-1}. Subtomograms and 3D CTF volumes were exported at 5.0 Å px^{-1}. Global 3D refinement reached a resolution of 13 Å. The results were imported in M. The particle poses were given a temporal resolution of three, the image warp and volume warp grid resolutions were set to $8 \times 8 \times 41$ and $4 \times 4 \times 2 \times 20$, respectively, and refinement of all parameters was performed for five iterations (Supplementary Table 1). Data-driven anisotropic weight estimation was performed to assign unique weights to every series and tilt index.

The processing of EMPIAR-10045 tilt series was performed in exactly the same way as descried in the previous paragraph for EMPIAR-10064, using 3,058 particles (Supplementary Table 1).

Tilt-series movie data were downloaded for the EMPIAR-10164 entry. Tilt movie frame alignment was performed in Warp using a model without spatial resolution. Initial tilt-series alignment was performed in IMOD using gold fiducials automatically picked in Warp, on 6x binned images with default settings. Tilt-series CTF estimation was performed in Warp. A total of 130,658 particles were picked using Warp's 3D template matching with a template derived from EMD-3782 in full tomograms reconstructed at 10 Å px^{-1}. Subtomograms and 3D CTF volumes were exported at 5 Å px^{-1} using 56 px boxes. Global 3D refinement with C6 symmetry was performed in RELION v.3.0 and reached the 10 Å Nyquist limit. The results were imported in M. The particle poses were given a temporal resolution of three, the image warp and volume warp grid resolutions were set to $8 \times 8 \times 41$ and $3 \times 3 \times 3 \times 20$, and refinement of all parameters was performed for five iterations, including tilt movie frame alignment in the last two iterations (Supplementary Table 1). Data-driven anisotropic weight estimation was performed to assign unique weights to every series, tilt index and tilt frame index.

Acquisition and refinement of *M. pneumoniae* in situ tilt-series data. Data previously used in another study[46] were reanalyzed with the release version of M. As described there, *M. pneumoniae* strain M129 (ATCC 29342) cells were grown

on 200 mesh gold grids coated with a holey carbon support (R 2/1, Quantifoil). Cells were cultivated at 37 °C in modified Hayflick medium: 14.7 g l⁻¹ of Difco PPLO (Becton Dickinson), 20% (v/v) Gibco horse serum (New Zealand origin, Life Technologies), 100 mM HEPES-Na (pH 7.4), 1% (w/w) glucose, 0.002% (w/w) phenol red and 1,000 U ml⁻¹ of freshly dissolved penicillin G. Cm (Sigma-Aldrich) was added 15 min before vitrification, at a final concentration of 0.5 mg ml⁻¹. Grids were quickly washed with PBS buffer containing 10 nm protein A-conjugated gold beads (Aurion), blotted from the back side for 2 s and plunged into mixed liquid ethane/propane at liquid N_2 temperature with a manual plunger (Max Planck Institute of Biochemistry). The cryo-EM grids were stored in a sealed box in liquid N_2 before usage.

Tilt-series data were collected on a Titan Krios TEM operated at 300 kV (Thermo Fisher Scientific) equipped with a field-emission gun, a Gatan K2 Summit direct detector and a Quantum post-column energy filter (Gatan). Images were recorded in exposure fractionation, counting mode using SerialEM v.3.7.2. Tilt series were acquired with a dose-symmetric scheme using dedicated scripts[59] with the following settings: TEM in nano-probe mode, magnification 81,000 with a calibrated pixel size of 1.7 Å, energy filter in zero loss mode, defocus range of 1.5 to 3.5 µm, tilt range −60° to 60° with 3° tilt increment and constant exposure per tilt and a total exposure of 120 e⁻ Å⁻². In total, 65 tilt series were collected from Cm-treated cells.

Raw tilt movies were processed in Warp. De novo tilt-series alignment was performed in IMOD using gold fiducials picked automatically with Warp's BoxNet, and the results were imported in Warp, where the tilt-series CTFs were estimated. Using full tomograms reconstructed at 10 Å px⁻¹, two tomograms were denoised using Warp's Noise2Map tool to pick the ribosome particles manually. Using these coordinates, subtomograms were exported from Warp to RELION to obtain an initial reference. This reference was used to perform template matching in Warp at 10 Å px⁻¹. In addition, a binary classifier based on a 3D CNN was trained on the two manually picked tomograms to remove false positives (membranes, carbon hole edges and so on) from the template matching results. In total, 24,202 particles were obtained this way. Subtomograms for all particles were exported from Warp to RELION and aligned against the previously refined low-resolution reference. No classification was performed. The results were imported in M. There, global movement and rotation, a 5×5×41 image-space warping grid, a 8×8×2×10 volume-space warping grid, as well as particle pose trajectories with three temporal sampling points were refined over five iterations (Supplementary Table 1). Starting with iteration 3, CTF parameters were also refined. At the beginning of iteration 4, reference-based tilt movie alignment was performed, resulting in a 3.7 Å map. Using the improved alignments, subtomograms were reconstructed at 3 Å px⁻¹. Classification into five classes was performed in RELION. Then 17,890 particles from the two best classes were imported in M and refined for another iteration using the same settings to obtain a 3.5 Å map. The final iteration was completed in around 6 h, using four GeForce 2080 Ti GPUs. Afterward, focused refinements were performed in M with masks limited to the 30S and 50S subunits, optimizing only image warping and particle poses.

To calculate the Rosenthal–Henderson[47] plot, deformation, weighting and CTF parameters from the last iteration of 70S refinement were kept. The number of particles was reduced by excluding entire tilt series from the dataset, thus keeping the average particle density per series constant. Resolution was reset to 10 Å at the beginning of each subset's refinement, and only the particle pose trajectories were optimized for three iterations.

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

Data availability
Maps were deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers 11603 (AF-t), 11611 (AF-f), 11652 (EMPIAR-10248), 11653 (EMPIAR-10045), 11654 (EMPIAR-10064), 11655 (EMPIAR-10164), 11656 (EMPIAR-10288), 11651 (EMPIAR-10453), 11650, 11998, 11999 (Cm-treated *M. pneumoniae* 70S ribosome, 30S and 50S subunits, respectively). Raw data for AF-t and AF-f were deposited in the EMPIAR database under accession number 10491, and raw data of Cm-treated *M. pneumoniae* under accession number 10499. Data from previous studies reanalyzed here were obtained from EMPIAR (10045, 10064, 10164, 10248, 10288, 10453). Maps from previous studies used for comparisons were obtained from the EMDB (0339, 0529, 3782, 8986, 9865, 10683). Atomic models from previous studies used for comparisons were obtained from PDB (4V7T, 5L93).

Code availability
M and Warp binaries, source code, and user guide are available as Supplementary Software. Updated versions can be found at https://github.com/cramerlab/warp and https://warpem.com.

References
49. Git—free and open source distributed version control system. https://git-scm.com (2020).
50. Grant, T. & Grigorieff, N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.* **192**, 204–208 (2015).
51. Mahamid, J. et al. Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016).
52. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCr. J.* **7**, 253–267 (2020).
53. DeRosier, D. Correction of high-resolution data for curvature of the Ewald sphere. *Ultramicroscopy* **81**, 83–98 (2000).
54. Penczek, P. A. et al. CTER—rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140**, 9–19 (2014).
55. Ludtke, S. J. Single particle refinement and variability analysis in EMAN2.1. *Methods Enzymol.* **579**, 159–189 (2016).
56. Schilbach, S. et al. Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* **551**, 204–209 (2017).
57. Buchholz, T.-O., Jordan, M., Pigino, G. & Jug, F. Cryo-CARE: content-aware image restoration for cryo-transmission electron microscopy data. In *Proc. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 502–506 (2019).
58. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
59. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).

Author contributions
D.T. designed M's architecture and algorithms, and carried out all implementation and application. L.X. and J.M. provided tilt-series data of Cm-treated *M. pneumoniae*, and assisted in testing *M* and interpretation of the maps. D.T. and L.X. solved the Cm-bound ribosome structure. C.D. collected apoferritin frame and tilt series, and analyzed the frame-series data using RELION. P.C. provided scientific environment, funding and additional interpretations and implications. D.T., P.C. and J.M. wrote the manuscript with input from all authors.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41592-020-01054-7.

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-020-01054-7.

Correspondence and requests for materials should be addressed to D.T., P.C. or J.M.

Peer review information Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
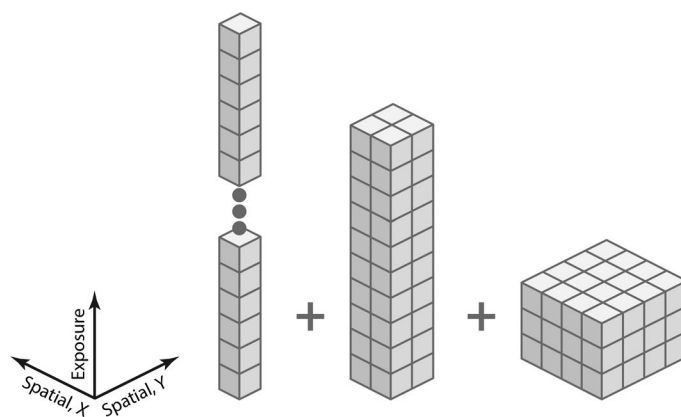
Extended Data Fig. 1 | Example of a parameter grid pyramid that models in-plane motion in a frame-series. A pyramid results from a combination of several grids to model the in-plane motion occurring in a frame-series with 40 frames as a function of position and dose. Each cubical cell represents a sampling point. The top grid has full temporal (per-frame exposure) and no spatial resolution to model fast, global motion (left, 1x1x40, shown truncated). For subsequent grids, temporal resolution is reduced by a factor of 4 and spatial resolution is doubled to model slower, local motion (center, 2x2x10; right 4x4x3). The spatial resolution of the first grid can be set higher if there is enough particle signal to fit.
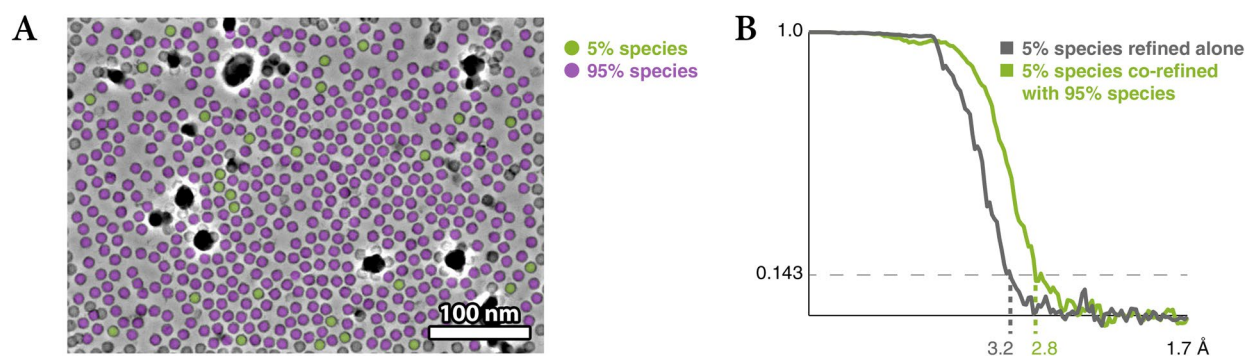
Extended Data Fig. 2 | Benefits of considering more particles per micrograph through multi-species refinement. Apoferritin frame-series were refined using a small 5% sub-population of the particles alone, and together with another 95% sub-population that improved the accuracy of the multi-particle system hyperparameters, but did not contribute particles to the 5% half-maps. **a**, Exemplary micrograph (n = 150 micrographs collected from the same sample) showing the distribution of the 2 sub-populations within a frame-series. **b**, FSC curves between the half-maps of the 5% population in both scenarios, showing the benefit of multi-species refinement.
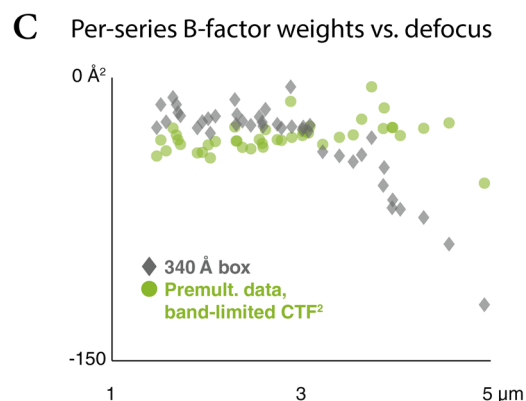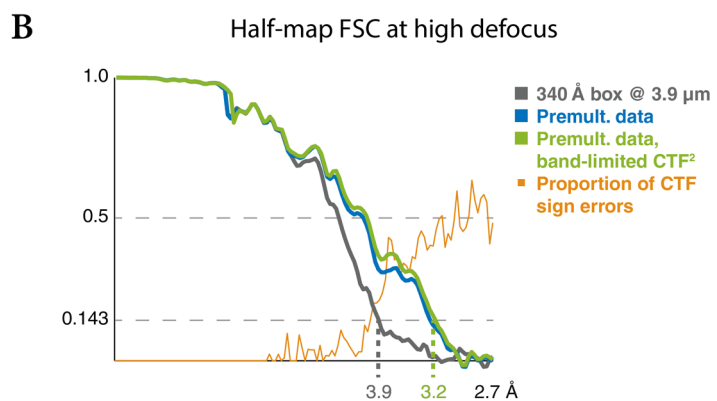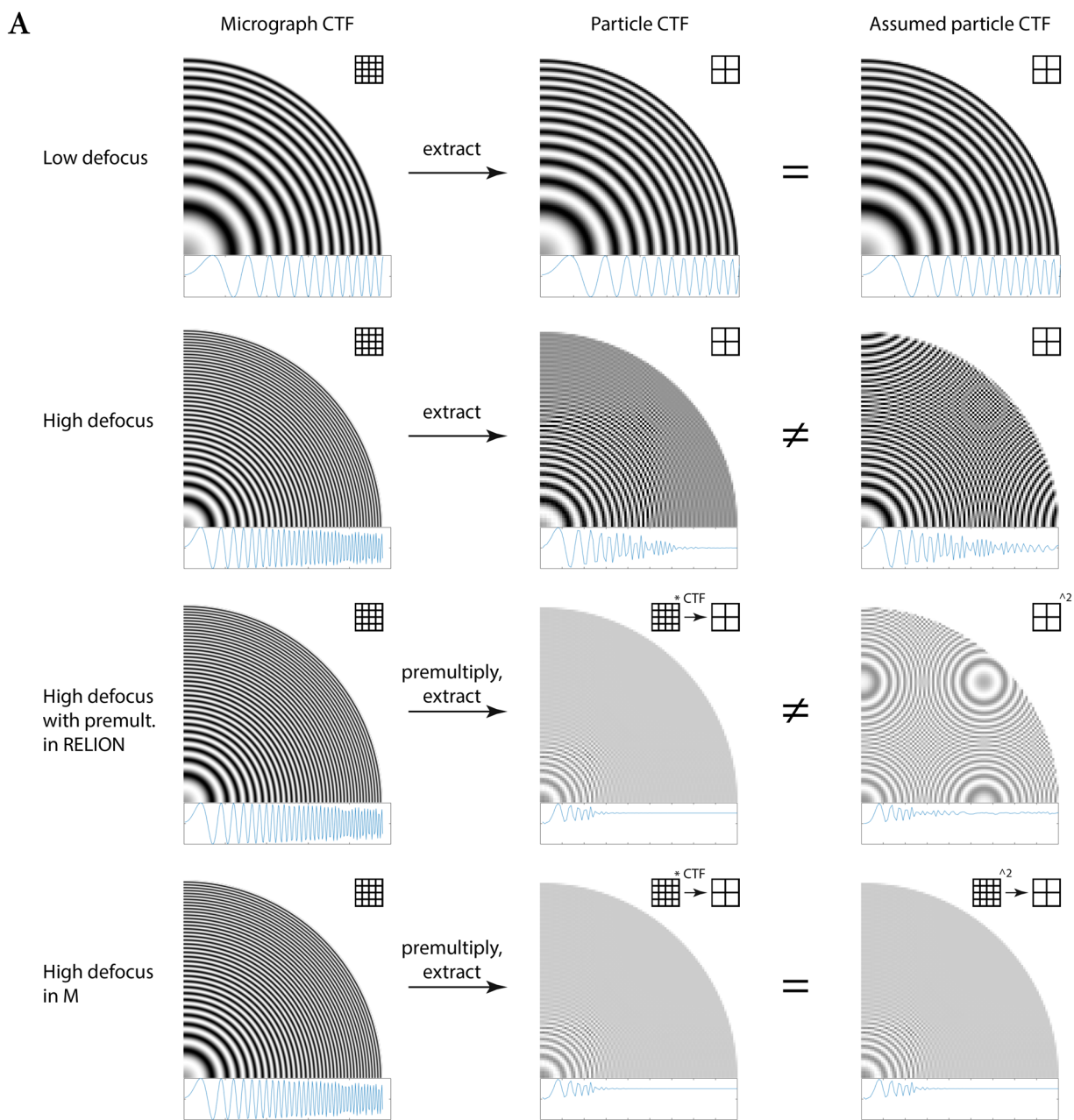
**Extended Data Fig. 3 | See next page for caption.**

Extended Data Fig. 3 | CTF correction at low and high defocus. High-resolution information is delocalized at high defocus. Choosing an insufficiently large particle box size results in loss of that information. In Fourier space, this results in CTF oscillations becoming too fast to be resolved at the sampling rate provided by the small box, averaging to 0. M chooses the box size automatically for each frame- or tilt-series' defocus, pre-multiplies the data and simulated CTF by the CTF to eliminate the oscillations and localize the signal, and then crops the data to the desired map size. This avoids the pitfall of losing map resolution due to an inappropriately chosen box size. **a**, Visualization of the delocalization and aliasing effects in Fourier space as 2D and rotationally averaged 1D CTFs; grids depict sampling rate. At low defocus (row 1), all signal is localized within the box and no aliasing is seen in the simulated CTF used for the image formation model during refinement. At high defocus (row 2), high-resolution signal is delocalized outside the small particle box. Once the particle is extracted, the fast CTF oscillations are averaged to 0 and high-resolution information is lost. At the same time, the simulated CTF is filled with aliasing artifacts because it is not low-pass filtered in the same way. If the particle data are pre-multiplied by the CTF at a box size large enough to contain all signal and resolve all CTF oscillations (row 3), as can be done optionally in RELION, all particle signal is contained in the box after cropping it to a smaller size, and the CTF averages to 0.5. However, the simulated CTF^2 does not match this and contains aliasing artifacts. M applies the pre-multiplication to both particle data and simulated CTF in a larger box before cropping (row 4) to avoid the mismatch. **b**, FSC between the half-maps reconstructed from HIV1 virus-like particles of a single high-defocus (3.9 μm) tilt-series in an insufficiently large box. Using data extracted without pre-multiplication, as is currently common, limits the resolution to 3.9 Å (grey). Pre-multiplying both particle data and CTF in a larger box, as automated in M, improves the result to 3.2 Å (green). Pre-multiplying only particle data is only slightly worse here (blue), but would likely lead to noticeably worse results in RELION as the aliased CTF^2 would be used in the image formation during refinement. The FSC curves diverge as the proportion of CTF sign errors (orange) increases. **c**, Relation between tilt-series defocus and associated contribution of high-resolution information to the reconstruction. For the larger dataset, not pre-multiplying the data results in a strong correlation, where high-defocus data is down-weighted to contribute less (grey). The correlation disappears when pre-multiplication is applied, so more tilt-series contribute high-resolution information (green).

Frame 1 Frame 3

2D correlation
reference vs. data

Fitted anisotropic
weighting model
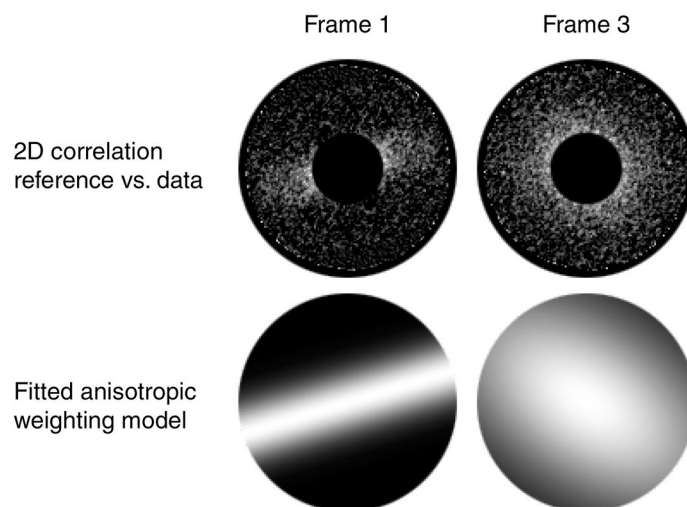
Extended Data Fig. 4 | Examples of anisotropic B-factor weighting. Normalized 2D cross-correlation between reference projections and data, averaged over all particles in a single frame is shown for the 1st and 3rd frame of the same exposure. Values in the low-frequency region are excluded to reduce the value range. The fitted B-factor is highly anisotropic for the 1st frame because of intra-frame motion: 0 Å^2 and -62 Å^2 along X and Y, respectively. For the 3rd frame, the fit is much more isotropic due to lack of intra-frame motion, but some high-resolution information is lost to radiation damage: -8 Å^2 and -10 Å^2 along X and Y, respectively.
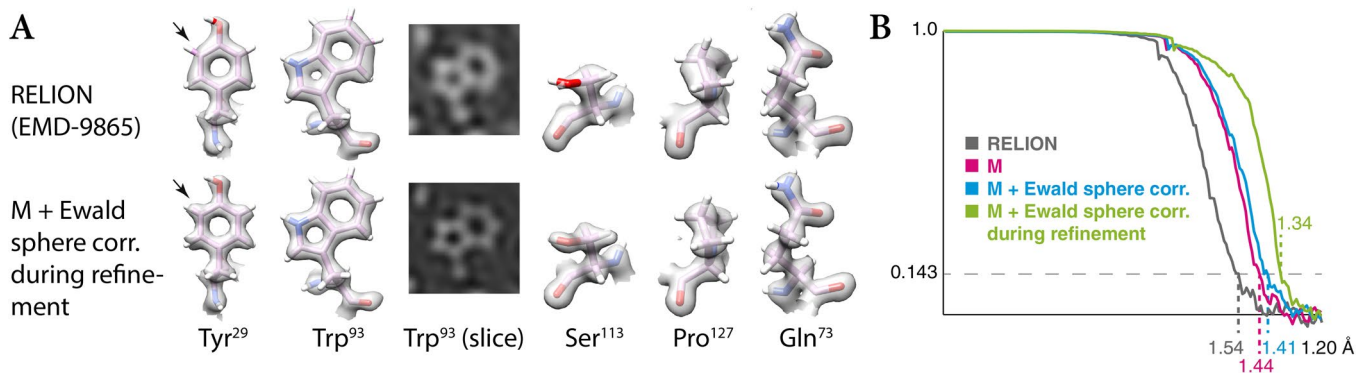
Extended Data Fig. 5 | Comparison with RELION on atomic-resolution frame-series data. Atomic-resolution data of apoferritin previously refined with RELION 3.1 to 1.54 Å (EMD-9865) were processed with *M* to achieve a resolution of 1.34 Å. **a**, Examples of side-chain densities produced by RELION (top) and M (bottom), showing cases of improved atomic features such as one of the hydrogens in Tyr[29] (black arrow). **b**, FSC between the half-maps produced by RELION (grey) and *M* (green), showing a general improvement in resolution through M.
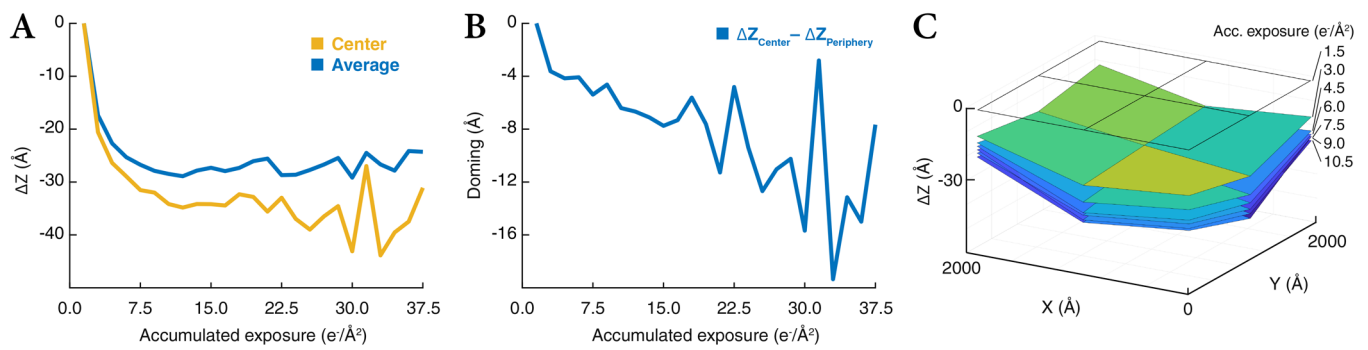
Extended Data Fig. 6 | Quantification of the doming effect in frame-series data. Doming models describing per-frame, spatially resolved (3×3 points) defocus offsets fitted during the refinement of atomic-resolution data of apoferritin (EMPIAR-10248) were averaged across the dataset, showing significant changes in the CTF during exposure. **a**, Defocus change plotted against the accumulated exposure show a fast change in both the central point and the average of the entire field of view's 3×3 points at the beginning of the exposure. After the first 7.5 e⁻/Å² of exposure, the average change stabilizes, while the central point continues to decrease in defocus. **b**, When corrected for global inclination, the difference between the central and peripheral defocus change indicates a steady increase in doming within the field of view as a function of accumulated exposure. **c**, Surface rendering of the spatially resolved defocus change for the first 7 frames shows an inclination of the entire field of view, as well as a more localized dent in the center. The observed change in the CTF can also be caused by electrostatic lensing effects due to sample charging, and further experiments are necessary to investigate the exact nature of doming.
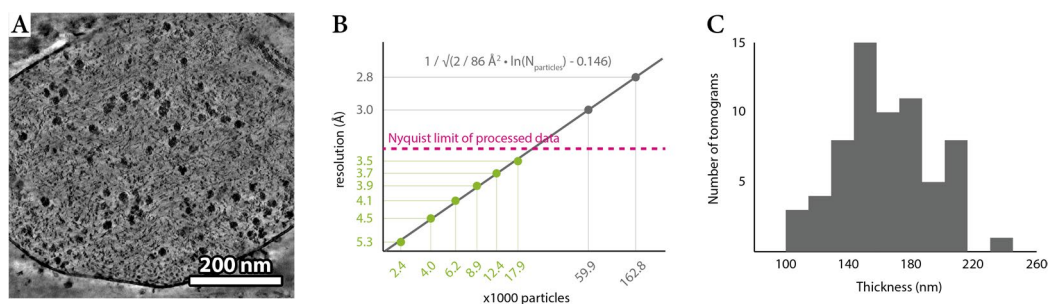
Extended Data Fig. 7 | Overview of the *M. pneumoniae* dataset. a, 2D XY slice through an exemplary denoised tomogram (n = 65 tilt-series collected from the same sample. Each tilt-series captures a single cell). **b**, Resolution plotted against the number of particles shows that 5 Å can be obtained with less than 3000 large, asymmetric particles in cells. Extrapolation beyond the Nyquist limit of the data (magenta line) is speculative, but indicates that 3 Å could be surpassed with less than 100,000 particles, given data with higher magnification. **c**, Histogram of manually measured cell thickness values from 65 tomograms.

Corresponding author(s): Dimitry Tegunov, Patrick Cramer, Julia Mahamid

Reporting Summary

Nature Research 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 Research policies, see <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The <u>exact sample size</u> (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	An indication of 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 statistics including <u>central tendency</u> (e.g. means) or other basic estimates (e.g. regression coefficient) AND <u>variation</u> (e.g. standard deviation) or associated <u>estimates of uncertainty</u> (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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on <u>statistics for biologists</u> may be useful.

Software and code

Policy information about <u>availability of computer code</u>

Data collection	Apoferritin and M. pneumoniae data were acquired using SerialEM (10.1016/j.jsb.2005.07.007)
Data analysis	M and Warp 1.0.9, described in this study; RELION 2.1, doi:10.7554/eLife.18722; RELION 3.0, doi:10.7554/eLife.42166; TensorFlow 1.5 (https://tensorflow.org)

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research <u>guidelines for submitting code & software</u> for further information.

Data

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coulomb potential maps were deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers 11603 (AF-t), 11611 (AF-f), 11652 (EMPIAR-10248), 11653 (EMPIAR-10045), 11654 (EMPIAR-10064), 11655 (EMPIAR-10164), 11656 (EMPIAR-10288), 11651 (EMPIAR-10453), 11650 (in situ 70S

ribosome). Raw data for AF-t and AF-f were deposited in the EMPIAR database under accession number 10491, and the raw in situ data of Cm-treated M. pneumoniae cells were deposited under accession number 10499.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	The number and identity of the particles that went into each refined map were determined through 3D classification, as described under Methods, Benchmarking.
Data exclusions	No manual data exclusion was performed.
Replication	No replication was performed as only one data set was available for every experiment and all analysis algorithms were deterministic.
Randomization	During 3D map refinements, data were randomly split in 2 groups following the "gold standard" protocol (doi:10.1038/nmeth.2115) to enable resolution estimation through cross-validation (Fourier shell correlation). Randomization was not applicable to other parts of the study.
Blinding	No blinding was performed as the exact identities of the specimen imaged need to be known for the analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging