Low-dose cryo-electron ptychography of proteins at sub-nanometer resolution

Berk Küçükoğlu[1], Inayathulla Mohammed[1], Ricardo C. Guerrero-Ferreira[1,2], Stephanie M. Ribet[3], Georgios Varnavides[3,4], Max Leo Leidl[5,6], Kelvin Lau[7], Sergey Nazarov[8], Alexander Myasnikov[8], Carsten Sachse[5,9,10], Knut Müller-Caspary[6], Colin Ophus[3], Henning Stahlberg[1,*]

1. Laboratory of Biological Electron Microscopy, Institute of Physics, SB, EPFL, and Dep. of Fund. Microbiol., Faculty of Biology and Medicine, UNIL, Rt. de la Sorge, 1015 Lausanne, Switzerland.
2. Current address: Robert P. Apkarian Integrated Electron Microscopy Core, Emory University School of Medicine, 1521 Dickey Drive NE, Atlanta, Georgia 30322, USA
3. National Center for Electron Microscopy (NCEM), Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
4. Miller Institute for Basic Research in Science, University of California, Berkeley, CA 94720, USA
5. Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons (ER-C-3): Structural Biology, Jülich, Germany.
6. Department of Chemistry and Center for NanoScience, Ludwig-Maximilians-Universität München, Butenandstr. 11, 81377 München, Germany
7. Protein Production and Structure Core Facility (PTPSP), School of Life Sciences, EPFL, Rte Cantonale, 1015, Lausanne, Switzerland
8. Dubochet Center for Imaging Lausanne, EPFL and UNIL, EPFL VPA DCI-Lausanne, 1015 Lausanne, Switzerland
9. Institute for Biological Information Processing (IBI-6): Cellular Structural Biology, Jülich, Germany
10. Department of Biology, Heinrich Heine University, Düsseldorf, Germany.

* Correspondence should be sent to: henning.stahlberg@epfl.ch

Abstract

Cryo-transmission electron microscopy (cryo-EM) of frozen hydrated specimens is an efficient method for the structural analysis of purified biological molecules. However, for particles smaller than 50kDa, single particle cryo-EM often fails due to the limited imaging contrast of individual micrographs. For dose-resilient samples often studied in the physical sciences, electron ptychography – a coherent diffractive imaging technique using 4D scanning transmission electron microscopy (4D-STEM) – has recently demonstrated resolution down to tens of picometers for thin specimens imaged at room temperature.

Here we applied ptychographic data analysis to frozen hydrated single protein particles, reaching sub-nanometer resolution 3D reconstructions. We employed low-dose cryo-EM with an aberration-corrected, convergent electron beam to collect 4D-STEM data for our reconstructions. The high speed of the electron detector allowed us to record large datasets of electron diffraction patterns with substantial overlaps between the interaction volumes of adjacent scan position, from which the scattering potentials of the samples were iteratively reconstructed. The reconstructed micrographs show strong contrast enabling the reconstruction of the structure of apoferritin protein at up to 5.8 Å resolution. Ptychography is a promising tool to study the structure of smaller frozen hydrated protein particles, and, once combined with electron tomography tilt series, bears the potential to provide unique insight into the ultrastructure of vitrified biological tissue.

Main

Cryo-transmission electron microscopy (cryo-EM) has revolutionized life sciences and pharmaceutical research in academia and industry. Proteins as single particles in thin vitrified, aqueous layers can be structurally analyzed at near-atomic resolution by transmission electron microscopy (TEM) in a matter of hours, if the proteins are available in sufficiently large quantities and homogeneity, and if the protein particles have a molecular weight larger than 50 kDa [1]. If these conditions are not met, cryo-EM often fails. Cryo-EM analysis of proteins suffers from the low signal-to-noise ratio (SNR) in the images because proteins are weak-phase objects and highly fragile under the electron beam. Typically, the electron fluence must be limited to below 20 $e^{-}/Å^{2}$, if a protein is to be imaged at high resolution. Data can be recorded with slightly higher doses if fractionation of the electron dose is used, and recorded frames are resolution weighted with dose-dependent frequency filters before averaging, to obtain higher-contrast images [2]. Cryo-EM furthermore suffers from particle movements under the electron beam, which can be corrected partly during image processing through motion correction in dose-fractionated "movies" [3–5]. Finally, conventional cryo-EM suffers from the oscillating contrast transfer function (CTF), which is dampened towards higher resolution from the limited spatial and temporal coherence of the beam, the detector modulation transfer function (MTF), and non-corrected specimen movements, among others. Phase plates, such as Volta or laser phase plates [6,7] partly improve the CTF for low-resolution components, but so far have not yet shown improvements in final resolution.

An alternative data acquisition scheme is scanning transmission electron microscopy (STEM), which uses a focused electron probe that is scanned across the sample, while electron detectors record the number of electrons scattered to a certain angle covered by the detector pixel(s) as a function of the probe position [8]. Bright-field (BF) STEM images provide only weak phase contrast [9], and dark-field (DF) STEM yields high mass-thickness contrast, yet at low dose efficiency, so that STEM was until recently of limited use in the life sciences. STEM Z-contrast imaging, employing annular detectors covering high scattering angles, is based on Rutherford scattering, where the detector signal is approximately proportional to the square of the atomic number and linear to the number of atoms within the interaction volume. This linearity of the high-angle annular DF (HAADF) STEM signal has been exploited for mass measurements of protein particles that had been freeze-dried on ultra-thin carbon film supports [10,11]. Early attempts of high-resolution low-dose aberration-corrected HAADF STEM imaging showed only amplitude contrast and was found unsuitable for life sciences imaging [12]. However, cryo-STEM tomography, combining BF and DF STEM with sample tilt series to compute a 3D reconstruction of the vitrified specimen [13], has shown promising results for thicker biological specimens up to 1 µm diameter. More recently, integrated differential phase-

contrast (iDPC) STEM was applied to cryo-EM specimens, reaching 3.5 Å resolution for protein 3D reconstructions from data recorded with a quadrant-detector [8,14].

Beyond monolithic or segmented detectors, the use of pixelated detectors in STEM allows to record a 2D image of the diffracted probe for each 2D probe position, enabling advanced analysis in this so-called 4D-STEM method [15]. For example, electron dose-efficient recovery of the specimen's scattering phase can be achieved by ptychographic processing of the recorded diffraction patterns [16]. For materials sciences specimens, this method has recently established resolution records in the picometer range [17,18]. Besides the inverse multi-slice schemes used in that work, which consider multiple scattering events during reconstruction but require comparably high electron doses, several other algorithms exist that employ the single-scattering projection assumption. The latter can be divided into direct methods, such as single-sideband ptychography [19,20], Wigner distribution deconvolution [21,22], or parallax reconstructions from tilt-corrected BF STEM data [23–25], to provide phase information with real-time capabilities, or iterative methods, such as the computationally more demanding proximal gradient algorithms (*e.g.*, difference map [26] and ePIE [27]).

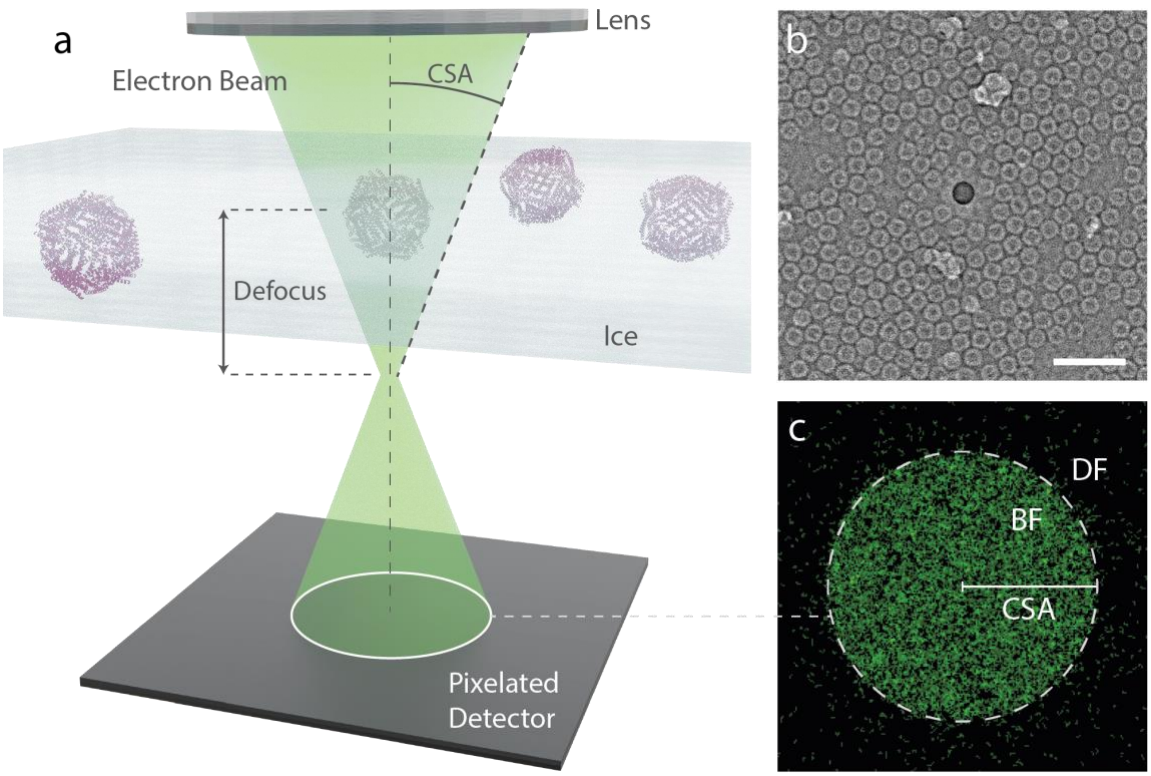
4D-STEM with an aberration-free electron beam that is focused onto the specimen typically scans the sample with a narrow Airy disk profile, limited to only a short vertical extension, depending on the convergence semi-angle (CSA) used to form the probe [8]. 4D-STEM can also be performed with a strongly defocused electron probe (**Figure 1**), so that the convergent electron beam illuminates the specimen with a disk large enough to expose an entire typical single-particle protein in the vitrified cryo-EM specimen as one single scattering experiment. The probe can then be scanned across the specimen surface at a step size in the nm range, so that diffraction patterns are recorded from probe positions that share significant overlap in the (x,y) sample plane. This overlap enables phasing of the recorded data for structure reconstruction, and it eliminates the twin-image problem or mirrored reconstructions.

4D-STEM has recently gained significant momentum from the availability of ultrafast pixelated electron detectors. Extending 4D-STEM to study beam-sensitive biological cryo-EM specimens, however, faces several experimental and computational hurdles. For low-dose experiments, scan strategies must be optimized for most efficient handling of specimen motion and specimen charging. Data processing workflows have to be established that can deal with the very low signal-to-noise ratio in the low-dose datasets and with the enormous quantity of recorded data, reaching terabytes in size for one sample. Nevertheless, significant progress in life sciences 4D-STEM has recently been demonstrated by the application of 4D-STEM ptychography to frozen-hydrated viruses [16,28].

Here, we performed low-dose 4D-STEM studies with a defocused beam and ptychographic

data processing of frozen-hydrated single particle protein samples (**Fig. 1**) and show that structures of proteins at sub-nm resolution can be achieved from less particles than usually imaged in cryo-EM. We implement cryo-electron ptychography with an aberration corrected cryo-EM instrument and fast hybrid pixel detector, and we discuss the current state of automated data collection and data processing.

Fig. 1: Experimental setup of 4D-STEM.



a, A vitrified life sciences sample is illuminated with a convergent electron beam. Electrons are recorded using a pixelated detector. The convergence semi angle (CSA) is in the range of a few mrad. The beam is focused to a plane above or below sample with a defocus in the micrometer range. **b**, An example ptychographic reconstruction of apoferritin. Scale bar: 50 nm. **c**, The sum of three 4D-STEM diffraction patterns. A typical diffraction pattern shows most of the electrons in the bright field (BF) disk of the radius corresponding to the CSA of 6.1 mrad, and some electrons are scattered to higher angles to the dark field (DF) .

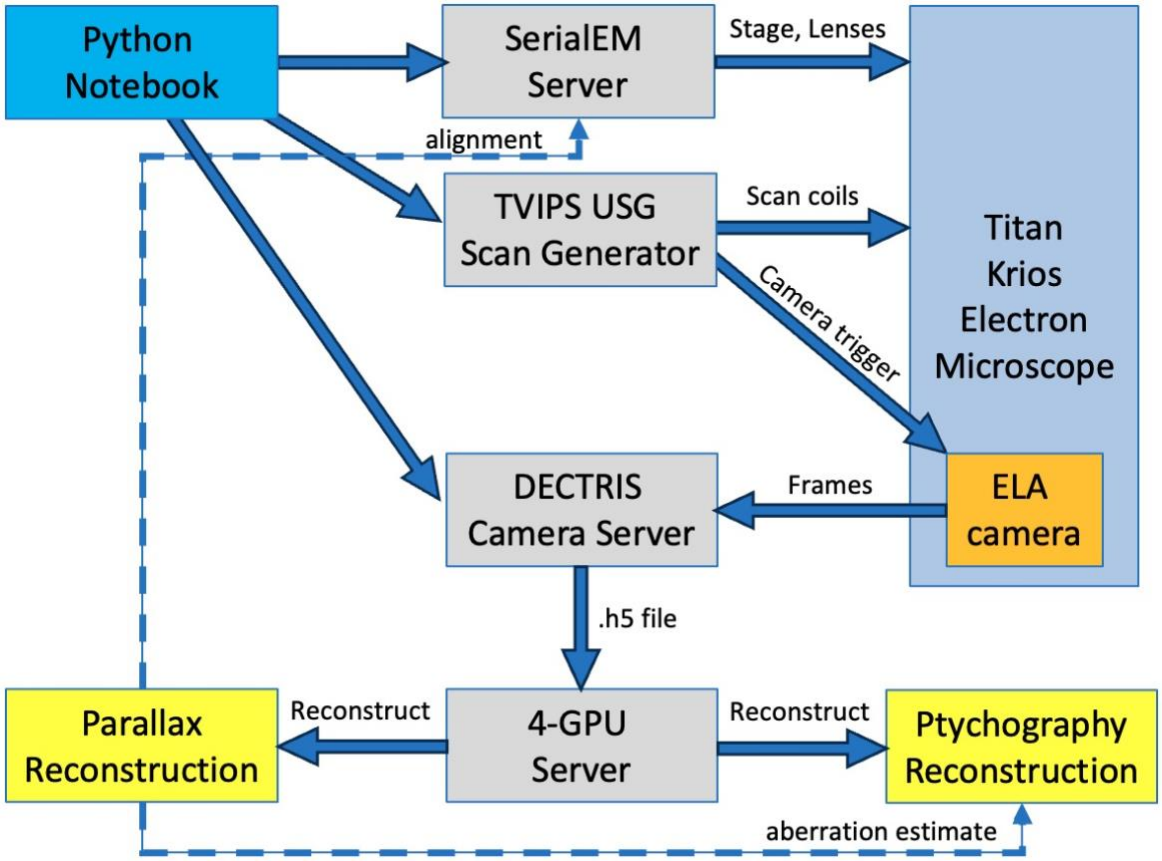
Ptychography of cryo-EM samples

Ptychographic data were recorded with a probe-corrected electron microscope with a CSA of 4 to 6 mrad and a defocus of ~2 µm, resulting in a beam diameter on the sample of ~16 nm for the 4 mrad CSA. Data were collected with a beam dwell time of 250 µs, and a STEM step size of 2 nm, corresponding to 84% beam overlap for two adjacent circular probes. Using a fluence of ~0.5 $e^-/Å^2$ per probe position, an average total electron fluence of ~35 $e^-/Å^2$ was applied to the sample during one scan of 128x128 positions. The low electron dose and parameters used in our experiments meant that on average less than 20% of the detector pixels received an electron for one probe position, making the recorded diffraction patterns very noisy.

A dataset of gold nanoparticles on a carbon film was collected before beginning experiments to calibrate rotation angles and CSA, and a diffraction pattern with the specimen removed from the beam path was recorded to serve as estimation for the probe-forming aperture in the iterative reconstruction algorithm. Data collection was then performed with a setup as shown in **Fig. 2**. The software SerialEM [29] was used to control the Titan Krios instrument, take real-space TEM images of the grid, and compile a table of X,Y coordinates, indicating the specimen locations that have ice holes with suitable ice thickness and a high particle density. After this, the instrument was switched to STEM mode with a convergent beam, and a custom Python script was used for automated 4D-STEM data collection. Data collection was performed at approximately 144 specimen locations per hour, each comprising a 128x128 position 4D-STEM scan, and resulted in approximately 310 GB of raw data per hour before compression.

Data collection was monitored in real-time by computing a parallax (or tilt-corrected BF-STEM) reconstruction with the py4DSTEM software [30]. This enabled fast, yet reliable, phase reconstructions to guide acquisition, and provided the probe estimate necessary for the more computationally-intensive ptychographic reconstructions. Higher-resolution ptychography reconstructions of the scattering potentials were calculated offline with py4DSTEM on a 4-GPU computer, obtaining up to 70 ptychographic reconstructions per hour.

Fig. 2: The ptychography data collection workflow.



A python client controls the electron microscope via SerialEM and coordinates the TVIPS USG scan generator, the Dectris ELA camera recording, and the multi-GPU processing of recorded diffraction patterns. Dotted line: The parallax reconstruction is used to provide on-the-fly feedback about the data collection and serve as aberration estimate.

Structural analysis of proteins

The cryo-electron ptychography pipeline was applied to apoferritin (AF) protein, the phi92 bacteriophage, and the tobacco mosaic virus (TMV). To test the quality of the AF protein and its suitability for high-resolution data collection, apoferritin was first imaged by conventional TEM (CTEM), before being studied using the 4D-STEM pipeline.

Conventional Cryo-EM imaging of Apoferritin

This spherical apoferritin protein particle of octahedral 24-fold symmetry has a diameter of approx. 12 nm and a particle weight of 485 kDA. It shows a highly robust structural integrity when prepared for cryo-EM imaging, making it the ideal test specimen for high-resolution assays [31,32]. Conventional TEM (CTEM), using dose-fractionated cryo-EM on a Titan Krios equipped with a 300kV cold-FEG electron source and a Falcon4i direct electron detector was

used to record 9,846 dose-fractionated real-space images, from which 1'899'133 AF particles were automatically selected and analyzed with the software cryoSPARC [33], resulting in a 3D map from 411'705 particles at a resolution of 1.09Å (**Extended Data Figs. 2 and 3**). Data showed a ResLog B-factor of 28 $Å^2$, indicating that the AF sample and the obtained cryo-EM grid preparations were of outstanding quality and not limiting resolution for a structural analysis (**Supplementary Table 1).**

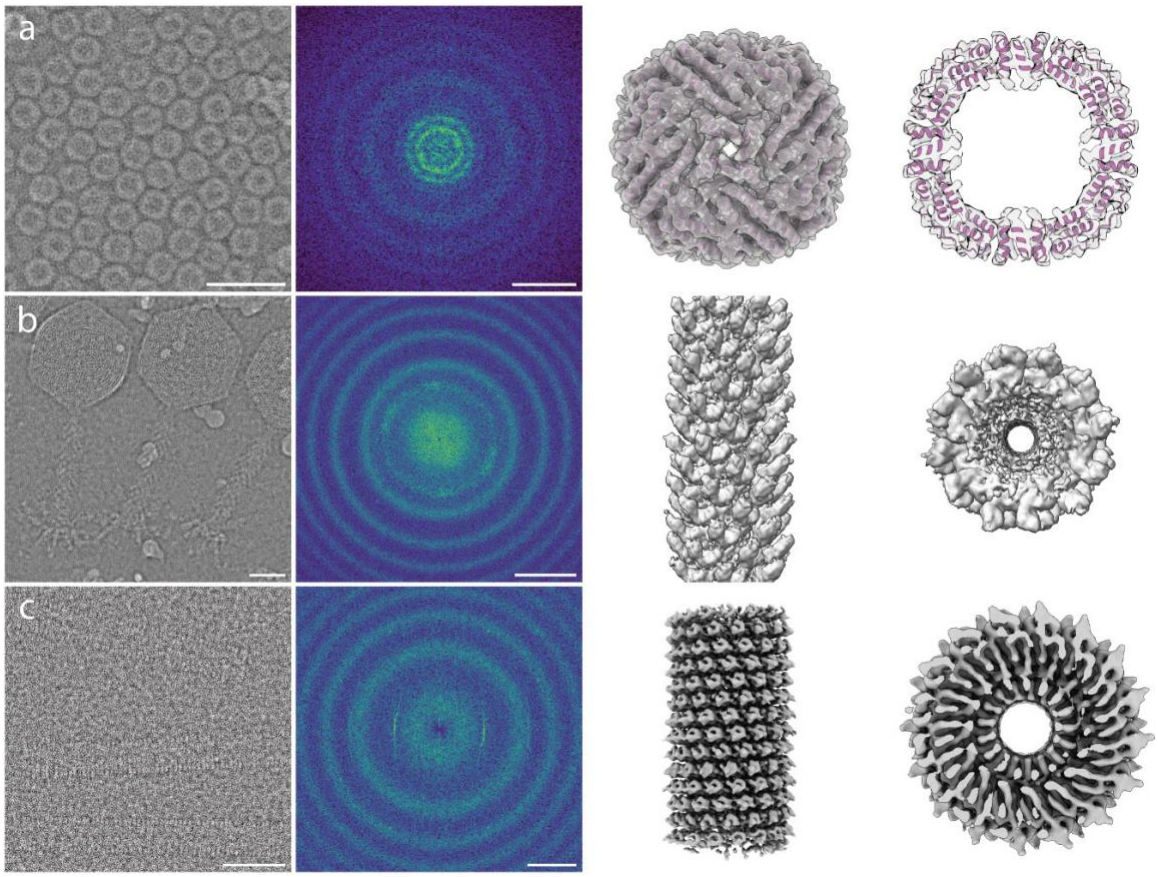
Cryo-electron ptychography of Apoferritin, Phi92 bacteriophage, and TMV

4D-STEM was then used on the same AF specimen, but with a different, Cs probe corrected Titan Krios, recording data with a CSA of 4 mrad. Online parallax reconstructions computed during data collection provided instant feedback on the quality of the data collection. Offline computed higher-resolution reconstructions were done by ptychography computation of the sample scattering potentials, which are obtained when assuming an amplitude function of unity and computing the resulting image only based on the reconstructed phase information [23]. Obtained 2D potential images were subjected to automated particle picking, resulting in 13’655 particles. Particle classification, alignment and 3D reconstruction using cryoSPARC [33] resulted in a 3D map of the protein from 11’552 particles at a resolution of 5.8 Å (**Fig. 3a, Extended Data Figs. 4 and 7**, and **Supplementary Table 1**).

4D-STEM was also used to study the Phi92 bacteriophage, using a CSA of 5.1 mrad. Data processing, using helical reconstruction for the bacteriophage's sheath was conducted using cryoSPARC [33], while the capsid reconstruction was performed using RELION [34,35], resulting in a reconstruction of the Phi92 sheath from 1’600 particles at 8.4 Å resolution, and of the Phi92 capsid from 182 particles at 12.1 Å resolution (**Fig. 3b, Extended Data Figs. 5 and 7,** and **Supplementary Table 1**).

4D STEM was also applied to study the tobacco mosaic virus (TMV), using a CSA of 6.1 mrad. Processing of the ptychography potential maps with RELION [34] resulted in a reconstruction from 2’120 particles at 6.4 Å resolution (**Fig. 3c, Extended Data Figs. 6 and 7,** and **Supplementary Table 1**).

Fig. 3: Cryo-electron ptychography reconstructions of proteins.



a, Apoferritin. **b**, Bacteriophage Phi92 tail protein. **c**, Tobacco mosaic virus (TMV). From left to right: Ptychographic image reconstruction (Scale bars are 30 nm); The same image Fourier transformed, showing circular amplitude oscillations (Scale bars are 0.5 nm^{-1}); The 3D reconstruction of the protein; A cross-section of the 3D reconstruction for **a** and a bottom view of the 3D reconstructions for **b** and **c**.

Discussion

Semi-automated 4D-STEM was implemented as a cryo-electron ptychography pipeline and applied to several biological test specimens. For reconstructing micrographs from recorded datasets, the single-slice ptychography reconstruction implemented in the py4DSTEM software was used. This algorithm refines the estimate for the *Probe* and the *Object* from the measured intensities in the recorded diffraction patterns as amplitude constraints in Fourier space, and by exploiting the knowledge about the position overlap in real space. The *Probe* and *Object* are iteratively refined, using a stochastic gradient descent algorithm. In the initiation of the projection algorithm, an empirically determined hyper-parameter tuning is also integrated, which improves convergence [23]. For details, see the **Online Methods**.

For CTEM imaging, the contrast transfer function (CTF) of the instrument is well characterized and its correction provides access to the highest resolution, including correction of higher-order aberrations [35]. Leidl et al. (2023) recently demonstrated that for thicker biological specimens, multiple electron scattering needs to be taken into account if structural data significantly beyond 1.2 Å resolution were to be achieved [36].

For 4D-STEM ptychography, the CTF of the iterative, gradient-based, regularized, ptychography in the presence of noise as employed here, is not established. The CTF is dependent on the 4D-STEM reconstruction algorithm as-well as acquisition parameters used. Computed Fourier transforms of the 2D ptychography potential maps obtained in this work show strong, ring-like amplitude oscillations for the background noise (**Fig. 3**). Our protein reconstructions corroborate the assumption that the employed 4D STEM data collection scheme and ptychography reconstruction algorithm transfers the specimen information to the reconstructed potential maps via a purely positive transfer function without zero crossings, but with varying amplitude, characterized by a contrast that is low for lowest spatial frequencies terms, rising to a broad maximum at medium resolution, and falling off with amplitude oscillations towards higher resolution [16,28,37]. Here, we ignored the effect of any CTF during image processing. Further work in analyzing, understanding and correcting the contrast in cryo-electron ptychography will be required.

Electron ptychography has established resolution records for materials sciences specimens, while the algorithms for signal reconstruction and our understanding of the mechanism of contrast formation are still under active development. For life sciences samples, cryo-electron ptychography promises to also give access to the highest resolution and contrast. However, significant technical challenges remain that should be addressed to further improve the resolution of the method. These include the implementation of dose fractionation during data collection, which would further reduce the signal-to-noise ratio of the collected diffraction

pattern; reconstructed sub-frames should then be resolution-weighted and averaged, as commonly done in real-space cryo-EM [2,3]. Further required optimizations concern the level of scan precision (noise in the scan coordinates), data collection speed (camera speed and data transfer speed), and post-processing algorithms under the very low dose constraints for life sciences specimens.

The frozen hydrated preparations of the proteins apoferritin, phi92 phage, and TMV studied here by 4D-STEM and ptychography resulted in 3D reconstructions at sub-nm resolutions, obtained from very few imaged particles. The ResLog B-factor analysis of the recorded data (**Online Methods** and **Supplementary Table 1**) suggests that a larger 4D-STEM dataset of the apoferritin protein might be able to reach 3.1 Å resolution. Nevertheless, this resolution would still be worse than what is achievable by conventional cryo-EM. However, here we applied 4D-STEM with a CSA of only 4 mrad to increase specimen height variation tolerance and data collection efficiency; a higher CSA would increase resolution. We also applied 4D-STEM without correcting for the resolution-limiting effects of specimen beam damage or beam-induced particle motion, and we did not correct for the CTF of the ptychography method, which would be required to fulfill the assumptions of equal noise contributions across all resolution ranges for the maximum likelihood processing underlying the cryoSPARC and RELION software packages. Dose fractionation and dose-dependent frequency weighting, combined with correction for particle movement and a CTF correction that normalizes the noise profile, would all increase the resolution performance of the 4D-STEM ptychography method.

The strong contrast of cryo-electron ptychography evident in the reconstructed images, the low numbers of particles needed to obtain sub-nm reconstructions, and the trustworthiness of the projections obtained from the purely positive CTF suggests that 4D-STEM combined with cryo-electron ptychography tomography bears great potential for 3D reconstructions of biological tissue samples. For cellular samples, where the uniqueness of the sample does not allow the averaging of signal from multiple particles, and where thicker slices of the sample are often of interest, the cryo-electron ptychography is especially promising to gain insight into the structural arrangement of smaller structures in cells and tissue.

Online content

Methods, additional references, Nature Portfolio reporting summaries, source 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/….

References

1. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for microscopy of unstained biological molecules. (1992).

2. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. *Elife* **4**, e06980 (2015).

3. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* **177**, 630–7 (2012).

4. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

5. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

6. Danev, R. & Baumeister, W. Cryo-EM single particle analysis with the Volta phase plate. *Elife* **5**, (2016).

7. Schwartz, O. *et al.* Laser phase plate for transmission electron microscopy. *Nat. Methods* **16**, 1016–1020 (2019).

8. Lazić, I. *et al.* Single-particle cryo-EM structures from iDPC-STEM at near-atomic resolution. *Nat. Methods* **19**, 1126–1136 (2022).

9. Evans, J. E. *et al.* Low-dose aberration corrected cryo-electron microscopy of organic specimens. *Ultramicroscopy* **108**, 1636–44 (2008).

10. Müller, S. A. & Engel, A. Mass Measurement in the Scanning Transmission Electron Microscope: A Powerful Tool for Studying Membrane Proteins. *J. Struct. Biol.* **121**, 219–30 (1998).

11. Engel, A. Biological application of scanning probe microscopes. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 79–108 (1991).

12. Buban, J. P., Ramasse, Q., Gipson, B., Browning, N. D. & Stahlberg, H. High-resolution low-dose scanning transmission electron microscopy. *J. Electron Microsc. (Tokyo)* **59**, 103–12 (2010).

13. Wolf, S. G. & Elbaum, M. CryoSTEM tomography in biology. *Methods Cell Biol.* **152**, 197–215 (2019).

14. Rose, H. H. Future Trends in Aberration-Corrected Electron Microscopy. *Philos. Trans. Math. Phys. Eng. Sci.* **367**, 3809–3823 (2009).

15. Ophus, C. Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond. *Microsc. Microanal.* **25**, 563–582 (2019).

16. Zhou, L. *et al.* Low-dose phase retrieval of biological specimens using cryo-electron ptychography. *Nat Commun* **11**, 2773 (2020).

17. Diederichs, B., Herdegen, Z., Strauch, A., Filbir, F. & Müller-Caspary, K. Exact inversion of partially coherent dynamical electron scattering for picometric structure retrieval. *Nat. Commun.* **15**, 101 (2024).

18. Chen, Z. *et al.* Electron ptychography achieves atomic-resolution limits set by lattice vibrations. *Science* **372**, 826–831 (2021).

19. Rodenburg, J. M., McCallum, B. C. & Nellist, P. D. Experimental tests on double-resolution coherent imaging via STEM. *Ultramicroscopy* **48**, 304–314 (1993).

20. Strauch, A. *et al.* Live Processing of Momentum-Resolved STEM Data for First Moment Imaging and Ptychography. *Microsc. Microanal.* **27**, 1078–1092 (2021).

21. Rodenburg, J. M. & Bates, R. H. T. The theory of super-resolution electron microscopy via Wigner-distribution deconvolution. *Philos. Trans. R. Soc. Lond. Ser. Phys. Eng. Sci.* **339**, 521–553 (1997).

22. Bangun, A., Baumeister, P., Clausen, A., Weber, D. & Dunin-Borkowski, R. Wigner Distribution Deconvolution Adaptation for Live Ptychography Reconstruction. *Microsc. Microanal.* **29**, (2023).

23. Varnavides, G. *et al.* Iterative Phase Retrieval Algorithms for Scanning Transmission Electron Microscopy. Preprint at https://doi.org/10.48550/arXiv.2309.05250 (2023).

24. Spoth, K. A., Nguyen, K. X., Muller, D. A. & Kourkoutis, L. F. Dose-Efficient Cryo-STEM Imaging of Whole Cells Using the Electron Microscope Pixel Array Detector. *Microsc. Microanal.* **23**, 804–805 (2017).

25. Yu, Y., Spoth, K., Muller, D. & Kourkoutis, L. Dose-efficient tcBF-STEM imaging with real-space information beyond the scan sampling limit. *Microsc. Microanal.* **27**, 758–760 (2021).

26. Elser, V. Phase retrieval by iterated projections. *JOSA A* **20**, 40–55 (2003).

27. Maiden, A. M. & Rodenburg, J. M. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy* **109**, 1256–1262 (2009).

28. O'Leary, C. M. *et al.* Contrast transfer and noise considerations in focused-probe electron ptychography. *Ultramicroscopy* **221**, 113189 (2021).

29. Mastronarde, D. N. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. *Microsc. Microanal.* **9**, 1182–1183 (2003).

30. Savitzky, B. H. *et al.* py4DSTEM: A Software Package for Four-Dimensional Scanning Transmission Electron Microscopy Data Analysis. *Microsc. Microanal.* **27**, 712–743 (2021).

31. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).

32. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).

33. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

34. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–30 (2012).

35. 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. *IUCrJ* **7**, 253–267 (2020).

36. Leidl, M. L., Sachse, C. & Müller-Caspary, K. Dynamical scattering in ice-embedded proteins in conventional and scanning transmission electron microscopy. *IUCrJ* **10**, 475–486 (2023).

37. Yang, H., Pennycook, T. J. & Nellist, P. D. Efficient phase contrast imaging in STEM using a pixelated detector. Part II: optimisation of imaging conditions. *Ultramicroscopy* **151**, 232–239 (2015).

38. Danev, R., Yanagisawa, H. & Kikkawa, M. Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. *Trends Biochem. Sci.* **44**, 837–848 (2019).

39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

40. Yamashita, K., Palmer, C. M., Burnley, T. & Murshudov, G. N. Cryo-EM single-particle structure refinement and map calculation using Servalcat. *Acta Crystallogr. Sect. Struct. Biol.* **77**, 1282–1291 (2021).

41. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

42. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–45 (2003).

Online Content

Methods

Protein expression and purification

For apoferritin, the plasmid encoding the mouse heavy chain apoferritin was graciously donated by Masahide Kikkawa [38]. The plasmid was transformed into *E. coli* BL21 (DE3) cells. A saturated overnight culture was inoculated into 500 mL LB media containing Kanamycin, and grown at 37 ºC and induced with 1 mM IPTG for 3 hours after reaching an OD600 of 0.6. The pellet was harvested at 5000 x g and resuspended in 20 mL Buffer A (300 mM NaCl, 20 mM HEPES 7.5, 10 mM EDTA), lysed by sonication and then centrifuged for 30 minutes at 20000 x g at 4 ºC. The supernatant was collected and heated at 70 ºC for 10 minutes in a 50 mL tube on a heatblock until milky. The supernatant was mixed by inversion and heated for 5 more minutes. The mixture was centrifuged for 30 minutes at 20000 x g at 4 ºC. The supernatant (~20mL) was precipitated with 6.65 g of ammonium sulfate, stirring at 4 ºC for 10 minutes. This corresponds to ~50-55% ammonium sulfate saturation at 20 ºC. The mixture was centrifuged for 30 minutes at 20000 x g at 4 ºC. The pellet was then resuspended in 2 mL of Buffer A and dialyzed with a 12-14 kDa membrane in 2 L of Buffer A but containing only 2 mM EDTA. The sample was further purified by size exclusion chromatography and loaded directly onto a Superose 6 10/300 GL column with a mobile phase of 150 mM NaCl, 10 mM HEPES 7.5. Peak fractions that corresponded to 24-mer-sized apoferritin were pooled and reinjected for a second round of size exclusion chromatography. Peak fractions were concentrated to 8 mg/mL, using a 100 kDa concentrator and frozen at -80 ºC.

The bacteriophage Phi92 sample, initially obtained by Petr Leiman (Texas University Medical Branch, TX, USA), was a kind gift from Davide Demurtas (EPFL, Lausanne, Switzerland).

The tobacco mosaic virus (TMV) sample, initially obtained by Ruben Diaz-Avalos (La Jolla Institute for Immunology, CA, USA), was a kind gift from Philippe Ringler (University of Basel, Switzerland).

Real-space cryo-EM data collection and processing

3 µL aliquots of the apoferritin at 80 mg/ml were applied onto Ultrafoil gold 1.2/1.3 grids, and plunge frozen in liquid ethane with a Leica GP2 plunger (4 °C, 100% rel. humidity, 30 s waiting time, 3 s blotting time).

For conventional (real-space) cryo-EM, data were collected in automated manner using EPU v.3.0 on a cold-FEG fringe-free Thermo Fisher Scientific (TFS) Titan Krios G4 TEM, operating at 300 kV with aberration-free image shift (AFIS), and recording data with a Falcon IV (TFS) electron counting direct detection camera in electron counting mode. The data were collected at a nominal magnification of 250 kx (pixel size is 0.3 Å at the specimen level) and a total dose of 40 ($e^-/Å^2$) for each exposure. The data from the 1.5 s exposures were stored as electron

event recording (EER) files. The data from 3 × 3 neighboring ice holes in the carbon films were collected using beam and/or image shifting, while compensating for the additional coma aberration (AFIS). The data were collected with the nominal defocus range of −0.3 to −0.9 µm. With an average throughput of 922 images per hour, a total number of 9'846 movies were collected within one 12-h session.

In total, 9'846 movies were imported into cryoSPARC v4.4.1 (eer_upsamp_factor=2, eer_num_fractions=80). Upsampled EER movies were drift-corrected and dose-weighted with Patch Motion Correction (res_max_align=4, output_fcrop_factor=1/2), and the contrast transfer function (CTF) was estimated with Patch CTF Estimation (amp_contrast=0.07, res_max_align=3, df_search_min=1000, df_search_max=20000).

After rejection of bad images based on CTF Fit resolution (2 - 5 Å) and Relative Ice Thickness (1 - 1.2) parameters, 8'609 images were used for further processing. Initial particle picking was performed on a subset of 50 images with Blob Picker, using ring blobs of 110-120 Å diameters, NCC score of 0.3, Power Score of 261–712, and a minimum relative separation distance (diameter) of 0.5. Picked particles were extracted with a box size of 720 pixels, Fourier cropped to 120 pixels (binning 6x6 times) and subjected to the streaming reference-free 2D classification.

Two best 2D class averages were selected as templates for the Template Picker, using a particle diameter of 126 Å and an angular sampling of 10 degrees. In total, 1'899'133 particles were picked. Picked particles were inspected with Inspect Particle Picks, and the best 1'104'665 particles were extracted with a box size of 720 pixels, Fourier cropped to 120 pixels, and subjected to the 3 rounds of reference-free 2D classification and Heterogeneous Refinement. A small subset of the 5'005 particles was used for the ab-initio structure building with 3 classes.

The best 544'845 particles were re-extracted with a box size of 800 pixels and further refined with Homogeneous Refinement to the resolution of 1.21 Å (B-factor = 23 $Å^2$). Custom parameters of Homogeneous Refinement included the application of octahedral (O) symmetry, per-particle defocus optimization using a Defocus Search Range of 4000 Å, per-group CTF parameter optimization with fitting the beam tilt, beam trefoil, spherical aberration, anisotropic magnification, and correction for the curvature of the Ewald sphere.

Refined particles were split into 9 optical groups with Exposure Group Utilities using Action “split”. Grouped particles were further refined with Homogeneous Refinement to the resolution of 1.15 Å (B-factor = 19 $Å^2$). Custom parameters of Homogeneous Refinement include application of octahedral (O) symmetry, an initial low-pass filter of 15 Å, a dynamic mask threshold of 0.3, a dynamic mask near value of 4 and far value of 8, defocus refinement, global

CTF refinement, and correction for the curvature of the Ewald sphere.

Refined particles were used as an input to Local Motion correction using a maximum alignment resolution of 2 Å and a B-factor of 20 $Å^2$ during alignment. Overlapping particles were removed with a minimum separation distance of 120 Å. The remaining 411'705 particles were subjected to Reference Based Motion Correction with default settings and further refined with Homogeneous Refinement (defocus refinement, global CTF refinement, and correction for the curvature of the Ewald sphere) to the resolution of 1.09 Å (B-factor = 18 $Å^2$).

An atomic model for apoferritin was re-built manually from the previously deposited model (PDB ID 8J5A) with COOT v0.9.6 [39], iterated with rounds of refinement with Servalcat [40] and Molprobity v4.5.2 [41]. After each round of refinement, side-chains were manually inspected and adjusted in COOT with Molprobity.

4D-STEM data collection

Ptychographic data were recorded with a probe-corrected TFS Titan Krios, equipped with a cold FEG electron source, TVIPS Universal Scan Generator (USG), Falcon 4i camera (TFS), and a Dectris ELA hybrid pixel detector. The microscope was operated in STEM mode with a CSA of 4 to 6 mrad and a defocus of ~2 µm, resulting for the 4 mrad CSA in a beam diameter on the sample of ~16 nm. Data were collected with a beam dwell time of 250 µs, and a STEM step size of 2 nm, corresponding to 84% beam overlap for two adjacent circular probes. Using a fluence of ~0.5 $e^-/Å^2$ per probe position, an average total electron fluence of ~35 $e^-/Å^2$ was applied to the sample during one scan. To avoid the physical gaps between the segments of the ELA detector, data were recorded on only one of the detector segments of 256x256 pixels. The STEM camera length was set so that the edge of that 256x256 px square was at 6.58 mrad. The low dose regime and parameters used in our experiments resulted in less than 20% of the detector pixels, on average, recording an electron at all. Dose fractionation, not employed here, could have been implemented by repeatedly scanning the same specimen location several times with an even lower dose, which would have further reduced the dose per pattern.

To improve convergence of the ptychographic reconstruction, especially under low-dose conditions, we optimized the alignment of the STEM probe before data acquisition, using the *Probe Corrector S-CORR* software by TFS on a standard gold cross-grating sample. Optimization of the aberrations up to twice the CSA that was later used for data collection, here to a semi-angle of 8 mrad, took approximately 10 minutes. We then collected a calibration dataset before beginning experiments to aid the optimization in the py4DSTEM software. Finally, a vacuum image of the probe was recorded in diffraction space to serve as estimate for the probe-forming aperture in the iterative reconstruction algorithm.

For data collection, the software SerialEM [29] was used to control the Titan Krios instrument and record a grid atlas at low magnification in image mode. Suitable grid squares on that atlas were manually selected, and for each of these, SerialEM was used still in conventional TEM mode with parallel illumination, to record grid square maps. These maps were then used to semi-automatically compile a table of X,Y coordinates, indicating the specimen locations where ice holes had suitable ice thickness and a high particle density. After this, the instrument was switched to STEM mode with a convergent beam, and a home-built Python script was used for automated 4D-STEM data collection, performing the following steps:

1. Controlling the microscope via SerialEM commands to position the instrument specimen stage.
2. Preparing the ELA detector for capturing frames.
3. Activating the TVIPS scan generator for acquisition.
4. Compressing and saving the recorded diffraction patterns for further analysis.

Data collection was performed at approx. 144 specimen locations per hour and resulted in approx. 310 GB of raw data per hour before compression. Data collection could be further accelerated by recording larger ptychography scans at each stage position, and by further accelerating the camera recording and data storage speeds, which all would have increased the amount of data recorded.

Data analysis

Data collection was monitored in real-time by computing a parallax reconstruction [23] with the py4DSTEM software [30]. In addition, a higher resolution ptychography reconstruction was calculated offline with py4DSTEM, with the following steps:

1. **Data loading and flat-field correction:** Electron diffraction patterns from one scan in h5 format were loaded, flat-field normalized, and hot/dead pixel corrected by replacing values of identified non-functional pixels with median values computed from their neighbors. Scan metadata such as scan step size were stored separately.
2. **Organizing 4D-STEM data:** The dataset was then reorganized into a four-dimensional array based on the scanning coordinates. Binning of diffraction patterns was not performed.
3. **Extraction of key parameters:** The py4DSTEM package was then used to compute the average diffraction pattern, identify the brightfield disk, and calculate the reciprocal space pixel size.
4. **Parallax reconstruction:** A parallax reconstruction was computed to give a first reconstruction image and to estimate probe defocus.

5. **Running ptychographic reconstruction:** Starting from the parallax reconstruction, a ptychographic reconstruction was initiated and iterated until convergence, typically reached in 30 iterations, as monitored by an L2 norm of the difference between modeled and measured diffraction intensities as error metric for single-slice ptychography. The probe function in the ptychography reconstruction was initialized based on provided aberrations (C1 in our case) for the phase, and the known amplitudes from the vacuum measurements of the probe, which were recorded for each data collection session. The probe's Fourier amplitude was replaced after each iteration with the amplitude from the vacuum probe measurements. The phase of the probe was also regularized after each update step by fitting a low-order surface expansion (up to radial and angular order 3) and removing residuals [23].
6. **Image storage:** Finally, the reconstructed micrograph was cropped if necessary and saved for subsequent analysis.

The workload was distributed across four GPUs, each processing a single dataset. This parallel approach allowed the processing of 70 datasets per hour with a 4-GPU computer.

4D-STEM ptychography structure determination of proteins

For apoferritin, ninety cryo-electron ptychography datasets (i.e., 4D-STEM scans) were acquired using SerialEM. A grid atlas was initially acquired using the Falcon4i camera, from which suitable grid squares were manually identified to compile grid square maps and identify suitable acquisition locations. These locations had an ice thickness between 50 and 100 nm. At these locations, automated ptychography scans with a CSA of 4 mrad were started, using an average total electron fluence of ~35 $e^-/Å^2$ per scan.

Ptychography reconstructions of the sample potential maps, which are obtained when assuming an amplitude function of unity and computing the resulting image only based on the reconstructed phase information, were obtained at a pixel size of 1.5 Å/px. These potential maps were imported as MRC image files into cryoSPARC [33]. Micrographs were subjected to automated particle picking, resulting in 38'425 particles (**Extended Data Fig. 4**). 2D classification identified 10 classes. The class averages were used as templates to refine the automated particle picking, resulting in 13'665 particles. After another 2D classification and selection of the best classes, a total of 11'552 particles were used for ab-initio 3D structure determination. No further correction for any remaining contrast transfer function was applied.

The resulting 3D map was analyzed by gsFSC with the 0.143 cutoff criterion, showing a final resolution of 5.8 Å of the map (**Fig. 3a,** and **Extended Data Figs. 4 and 7**, and **Supplementary Table 1**).

For the Phi92 bacteriophage, eighty-three 4D-STEM datasets were acquired using the same

methodology as before, but at a CSA of 5.1 mrad. The micrographs were recorded at a total electron dose of approximately 49 $e^-/Å^2$. The pixel size of the final micrographs was 1.44 Å/px.

The 2D ptychographic micrographs obtained were imported as MRC image files into both cryoSPARC [33] and RELION [34,35] for subsequent analysis. The helical reconstruction process of the bacteriophage's sheath was conducted using cryoSPARC, while the capsid reconstruction was performed using RELION.

Initial manual particle picking of the bacteriophage sheath in cryoSPARC yielded 2'420 particles. Following 2D classification and selection of the best classes, 1'600 particles were utilized for ab-initio 3D reconstruction. The helical parameters employed in this process included a helical rise of 35.30 Å and a twist of 26.3 degrees. The final processing stage involved non-uniform refinement (NU-Refine) with a Z fraction set at 50%, resulting in an 8.4 Å resolution map of the Phi92 sheath.

Concurrently, manual particle picking of the capsid in RELION resulted in the identification of 356 particles. After a similar 2D classification process, 182 particles were selected for the final 3D reconstruction. This approach yielded a 12.1 Å resolution map of the Phi92 capsid (**Fig. 3b**, **Extended Data Figs. 5 and 7,** and **Supplementary Table 1**).

For the tobacco mosaic virus (TMV), forty-three 4D-STEM datasets were acquired, following the same protocol as before, but with a CSA of 6.1 mrad. The datasets were acquired at a total dose of approximately 32 $e^-/Å^2$ per micrograph, and with a pixel size of 1.15 Å/px.

These 2D ptychographic micrographs were imported as MRC image files into the RELION software. The initial phase of manual particle picking of TMV in RELION resulted in 3'400 particles. Subsequent 2D classification and the selective process for the best classes reduced the number to 2'120 particles, which were then used for the ab-initio 3D reconstruction. Using helical parameters for the processing of a helical rise of 1.38 Å and a twist of 22.036 degrees resulted in a 6.4 Å resolution map of the TMV (**Fig. 3c, Extended Data Fig. 6 and 7,** and **Supplementary Table 1**).

ResLog B-factor estimation

The information content of the recorded data from conventional TEM and 4D-STEM ptychography imaging was evaluated by calculation of the $FSC_{0.143}$, the Guinier-plot amplitude falloff B-factor, and the Reconstruction "ResLog" B-factor [31,42]. With the latter, the number “#” of particles required to reach a resolution of "d" Angstrom, can be calculated as:

$$\# = \left(\frac{1}{N_{\text{asym}}}\right) \cdot \left[\left(\frac{\langle S \rangle}{\langle N \rangle}\right)^2 \cdot \frac{30\,\pi}{N_e \cdot \sigma_e \cdot d}\right] \cdot e^{\frac{B}{2 \cdot d^2}}$$

with N_{asym} being the number of asymmetric units in the particles, $\frac{\langle S \rangle}{\langle N \rangle}$ being the signal-to-noise ratio of the electrons in the recorded data, $\sigma_e = 0.004\ \text{Å}^2$ being the elastic cross-section of carbon, N_e being the used electron dose, d being the real-space resolution in Å, and B being the ResLog B-factor in Å^2.

This can be simplified to:

$$\# = \mathrm{const} \cdot e^{\frac{B}{2 \cdot d^2}}.$$

The ResLog B-factor in Å^2 can be calculated, based on the relation:

$$\ln(\#) = B \cdot \frac{1}{2 \cdot d^2} + \mathrm{offset}, \quad \text{with} \quad \mathrm{offset} = \ln(\mathrm{const}).$$

When computing the resolution d of a reconstruction with smaller subsets of # number of particles and plotting $\ln(\#)$ as a function of $\frac{1}{2 \cdot d^2}$, the ResLog B-factor becomes apparent as the slope of a line fitted through the experimental data points.

The predicted theoretically achievable resolution when expanding the dataset to more particles could then be obtained with:

$$d = \sqrt{\frac{B}{2 \cdot (\ln(\#) - \mathrm{offset})}}\ .$$

The computed $FSC_{0.143}$, Guinier-plot B-factors, ResLog B-factors, and the predicted resolution for 1'000'000 particles are given in **Supplementary Table 1**.

Data availability

For the apoferritin analysis by real-space cryo-transmission electron microscopy, the raw cryo-EM images were deposited to the Electron Microscopy Public Image Archive (EMPIAR) with accession code EMPIAR-XXX. The cryo-EM map was deposited to the Electron Microscopy Data Bank (EMDB) with accession code EMDB-19436. The atomic model was deposited to the Protein Data Bank (PDB) with accession code PDB-8RQB.

For the apoferritin analysis by cryo-electron ptychography, raw data are available EMPIAR XXX and the reconstructed volume at EMD-19425.

For the Phi92 analysis by cryo-electron ptychography, raw data are available at EMPIAR XXX. The reconstructed sheath volume is at EMDB-19430 and the reconstructed capsid volume at EMD-19423.

For the TMV analysis by cryo-electron ptychography, raw data are available EMPIAR XXX

and the reconstructed volume at EMD-19413.

Code availability

The py4DSTEM package is available at https://github.com/py4dstem/py4DSTEM. The data processing Jupyter Notebook utilized here is available at https://github.com/LBEM-CH/4D-BioSTEM.

Acknowledgements

We acknowledge fruitful discussions with Duncan Alexander, Cécile Hébert, and Marco Cantoni, EPFL Lausanne.

Author Contributions

B.K. designed the data acquisition and processing workflow, and integrated the components of the experimental setup. B.K., I.M., and R.C.G. collected ptychography data. I.M. performed the single particle analysis of the ptychographically acquired data. S.R., G.V., and C.O., are the developers of the py4DSTEM software and provided essential support in the processing of ptychographic data. K.L. purified and prepared the Apoferritin sample. S.N. and A.M. collected TEM images and ran Single Particle Analysis (SPA) on the Apoferritin sample. M.L.L., C.S. and K. M.-C. provided expertise and support on diffractive imaging. H.S. designed the experiments, provided resources, and oversaw data analysis and interpretation. B.K. and H.S. wrote the manuscript with input from all authors.

Funding

This work was supported by the Swiss National Science Foundation, grant 200021_200628 and by the European Union (ERC 4D-BioSTEM, No. 101118656). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

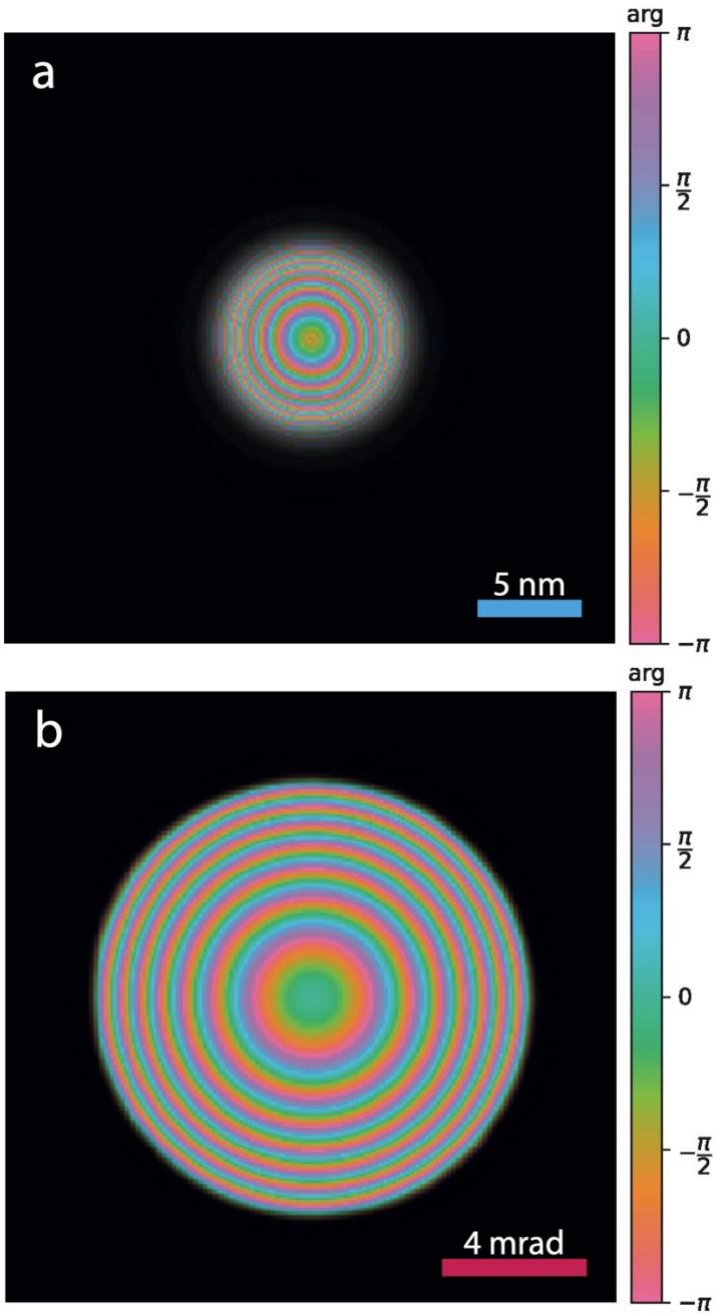
Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. SMR and CO acknowledge support from the U.S. Department of Energy Early Career Research Program. GV acknowledges support from the Miller Institute for Basic Research in Science,

Competing Interest

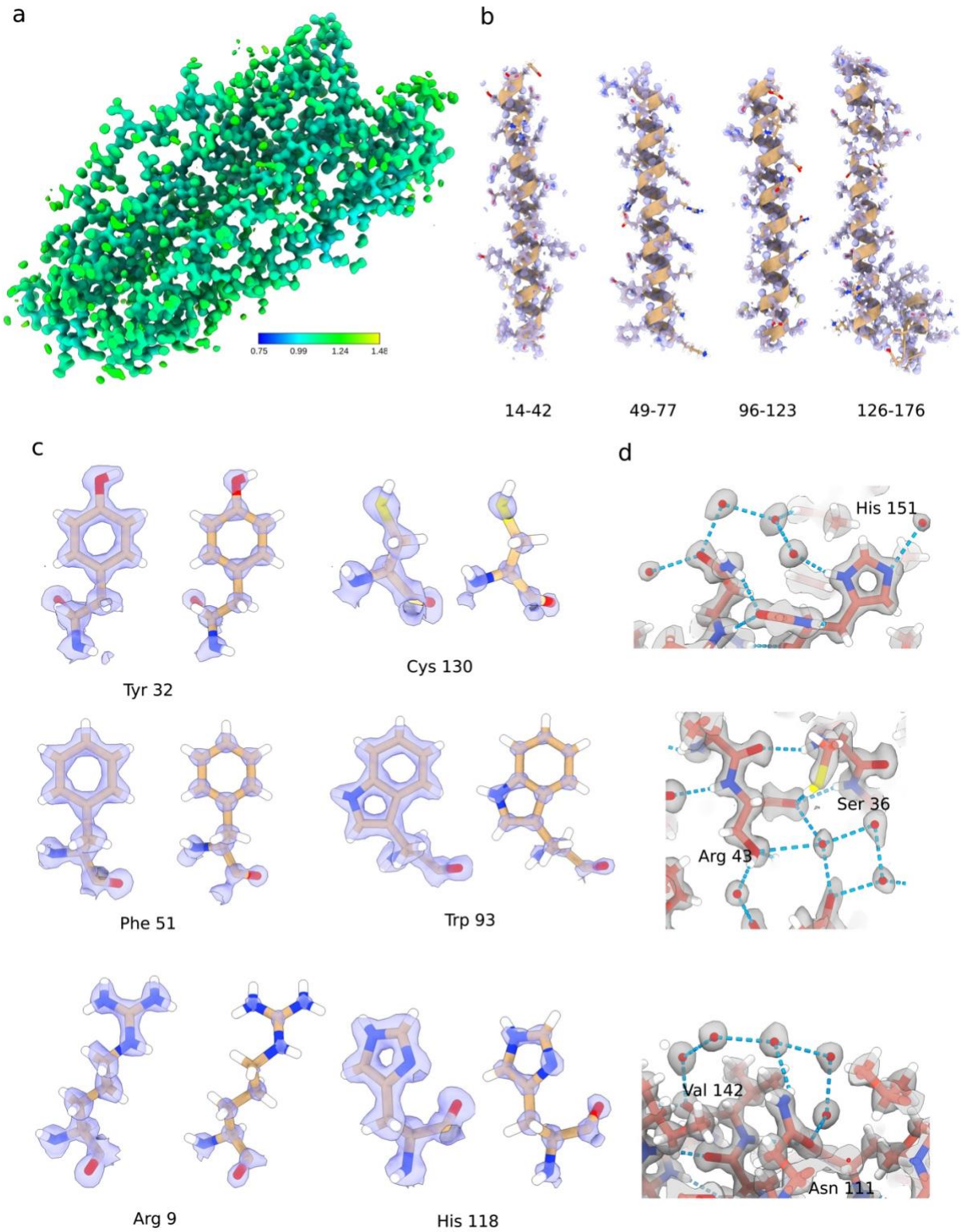
The authors declare no competing interests.

Extended Data

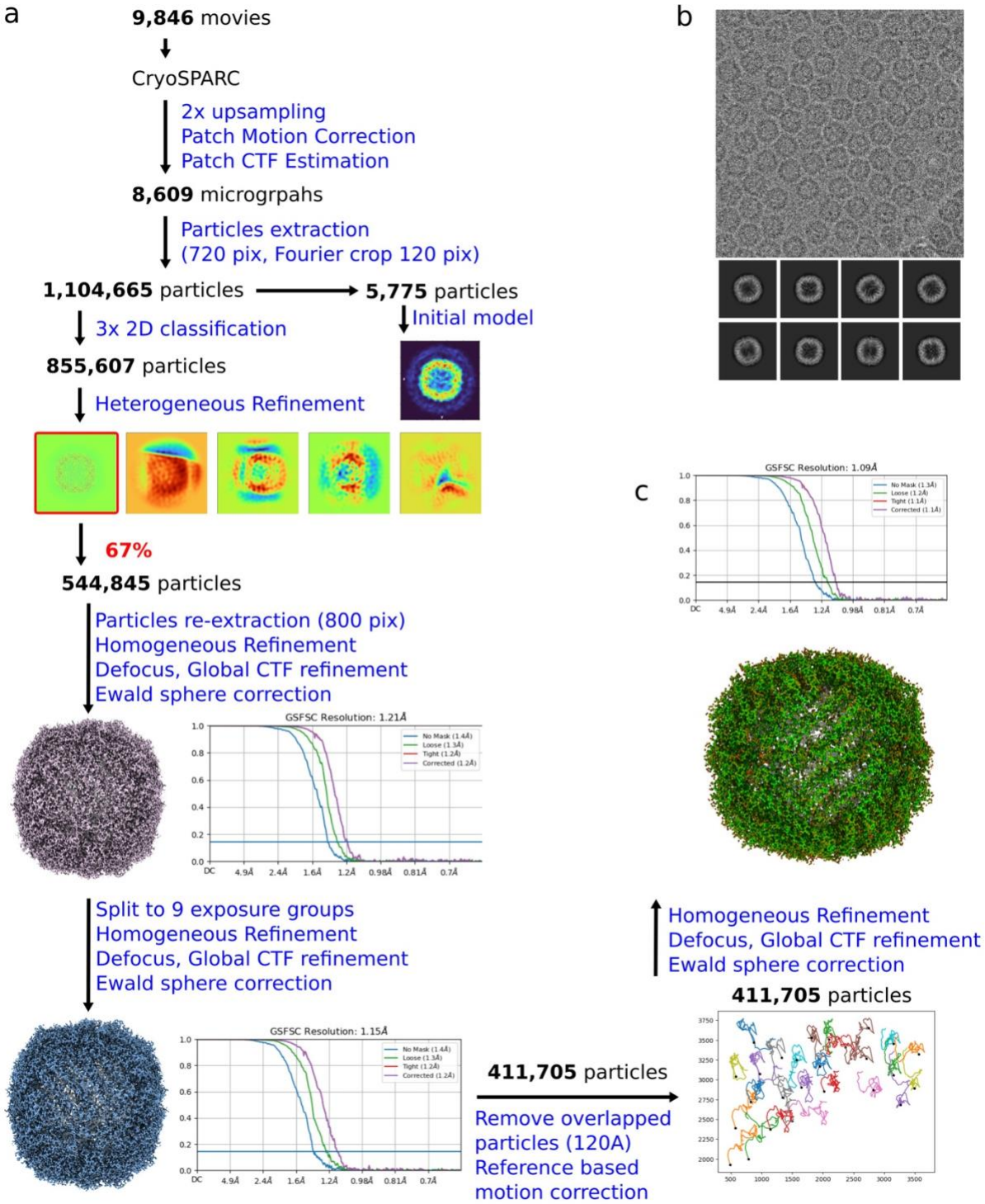
Extended Data Fig. 1: The 4D-STEM electron probe from a micrograph recorded at 6.1 mrad CSA and -758 nm defocus. a, Amplitude and phase of the reconstructed real-space electron probe on the sample. **b**, Amplitude and phase of the reconstructed Fourier probe on the detector.



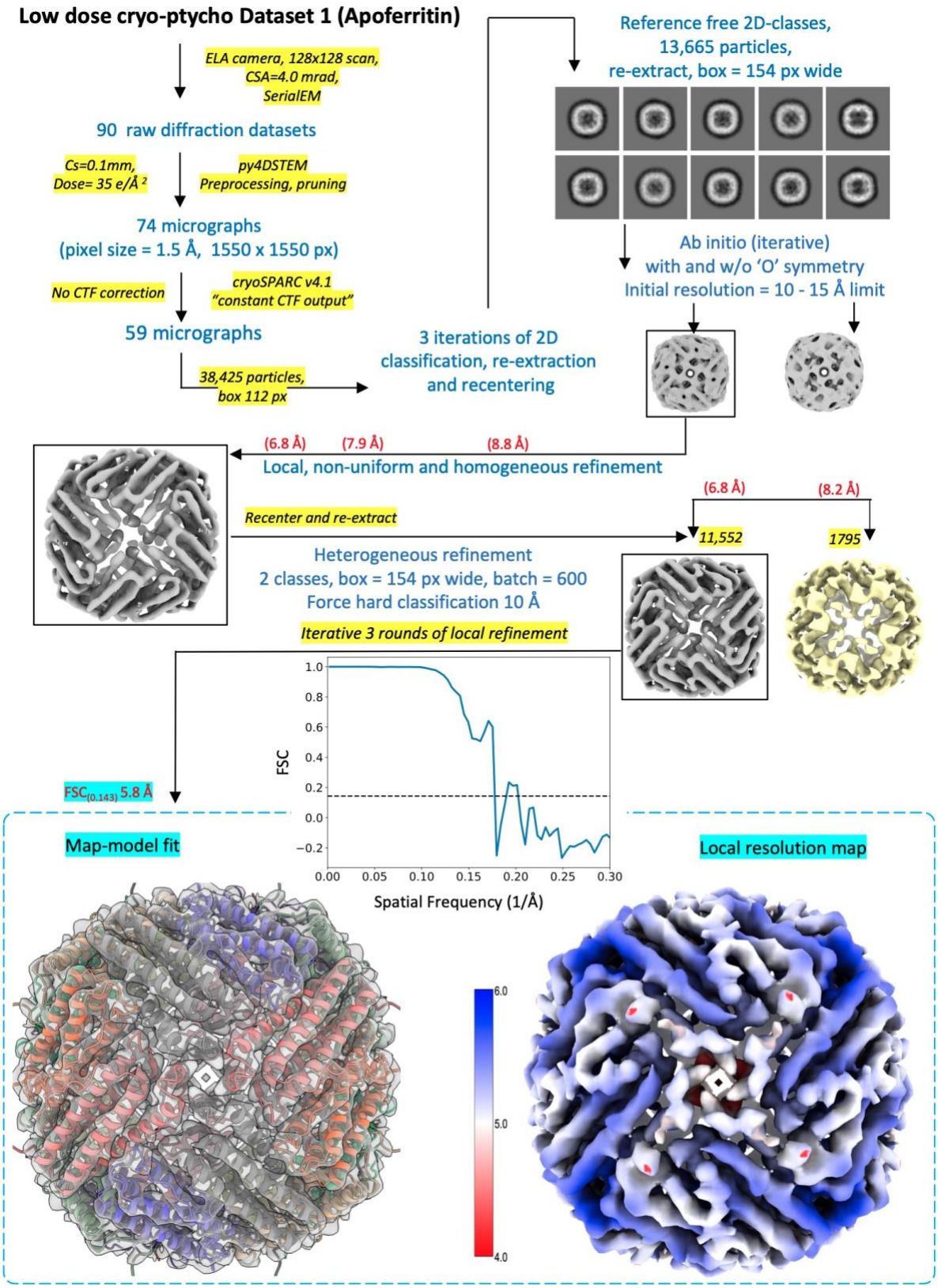
Extended Data Fig. 2: Real-space cryo-transmission electron microscopy structural analysis of Apoferritin at 1.09 Å resolution.



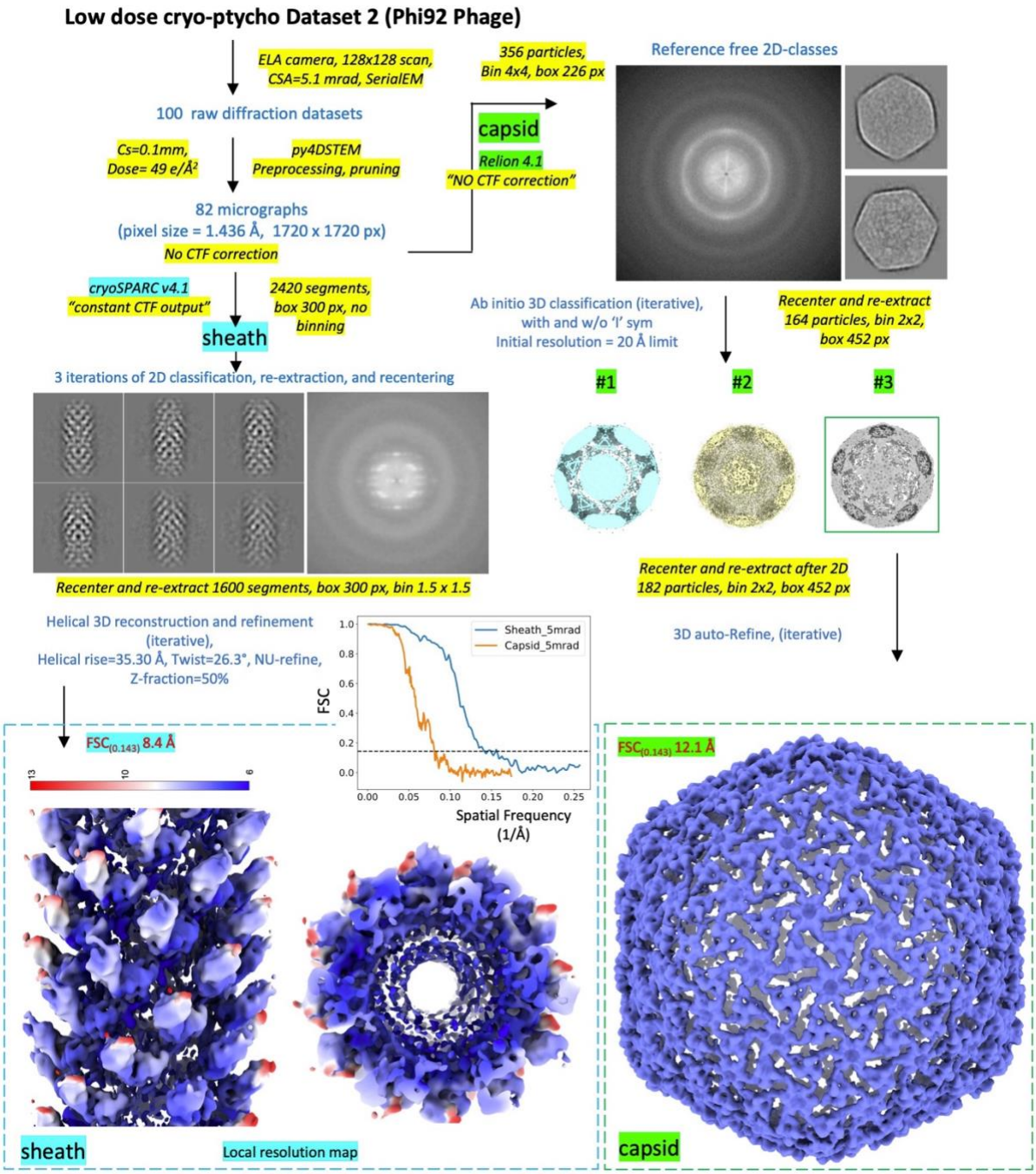
Extended Data Fig. 3: Real-space cryo-transmission electron microscopy of apoferritin and its data processing details.



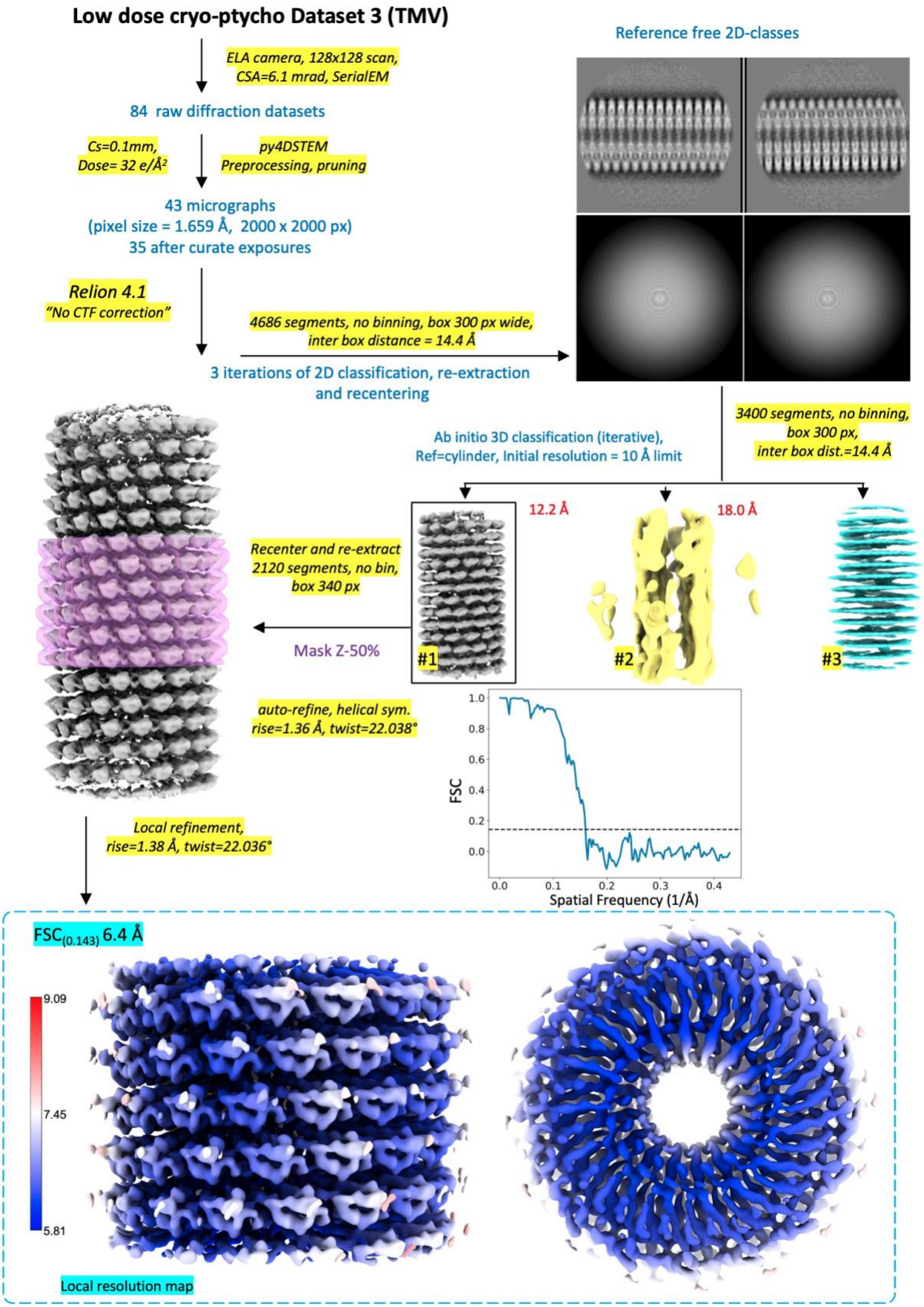
Extended Data Fig. 4: Ptychography structure determination of Apoferritin, using 4D-STEM data. The atomic model of Apoferritin used for the fitting was PDB 8J5A.



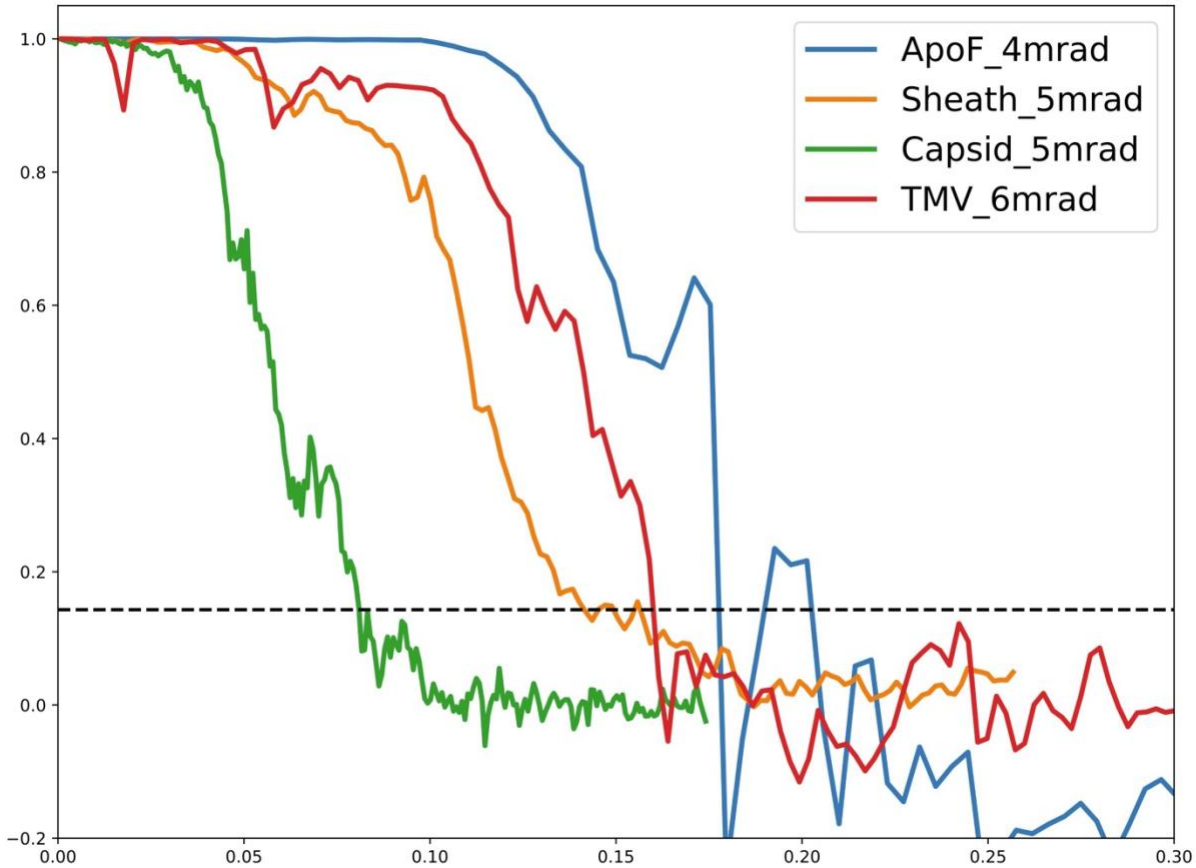
Extended Data Fig. 5: Ptychography structure determination of the bacteriophage Phi92, using 4D-STEM data.



Extended Data Fig. 6: Ptychography structure determination of tobacco mosaic virus (TMV), using 4D-STEM data.



Extended Data Fig. 7: FSC curves belonging to each individual dataset.



Supplementary Information

	Apoferritin	Apoferritin	Phi92 Sheath	Phi92 Capsid	TMV
Microscope	TFS Titan Krios, CFEG, Falcon 4i	TFS Titan Krios, CFEG, Probe corrector, TVIPS USG, Dectris ELA detector			
Voltage [kV]	300				
Nominal Magnification	250'000	—			
Electron exposure [e-/A2]	40	35	49		32
CSA [mrad]	—	4.0	5.1		6.1
Detector outer angle [mrad]	—	6.58	6.85		8.61
Step size [Å]	—	20.0			
Probe positions per micrograph	—	128 x 128			
Micrographs	9'846	90	83		43
Defocus [µm]	0.5 - 0.9	1.4 - 2.3	1.3 - 2.1		0.8 - 2.0
Pixel size [Å]	0.3	1.50	1.44		1.15
Initial particle numbers	1'899'133	13’665	2’420	356	3’400
Final particle numbers	411'705	11'552	1’600	182	2’120
Symmetry imposed	O	O	Helical	I	Helical
Guinier Plot B-factor [Å^2]	18	642	350	—	283
FSC threshold	0.143				
Map resolution [Å]	1.09	5.8	8.4	12.1	6.4
EMPIAR ID	TBD	TBD	TBD	TBD	TBD
EMDB ID	EMD-19436	EMD-19425	EMD-19430	EMD-19432	EMD-19413
PDB ID	8RQB	—	—	—	—
ResLog B-factor [Å^2]	28	174	739	—	358
Predicted resolution if expanding to 1'000'000 particles [Å]	1.05	3.11	6.08	—	4.38

Supplementary Table 1: Cryo-EM and cryo-electron ptychography data collection and map statistics