BMC Methods

METHODOLOGY

Open Access

Efficient strategies and troubleshooting for single particle cryoEM data collection using EPU

Janarjan Bhandari[1†], Dmytro Kompaniiets[2†], Ajit Kumar Singh[1†], Carol Bator[1], Jason Porta[1] and Bin Liu[2*]

Abstract

Background CryoEM has gained popularity due to advancements in both hardware and software, making it possible to visualize biological molecular structures in detail. However, despite its growing use, cryoEM remains a costly technique, and limited access to cryo-capable TEMs is slowing research efforts. By refining experimental design and streamlining data collection approaches, researchers can significantly boost efficiency and reduce costs, making this technique more accessible and impactful.

Method This work presents a guide to some of the session setup parameters for automated cryoEM data collection using Thermo Fisher's EPU software.

Results The TIFF and MRC formats were compared, and the Faster acquisition mode was evaluated against the Accurate acquisition mode. Acquired TIFF files were significantly smaller than MRC files without any notable loss of resolution in the data. By minimizing stage movements, the Faster acquisition mode increases data collection speed, while Accurate acquisition provides precise centering on the optical axis. The models generated from all experiments exhibit similar resolutions, approximately 2.12 Å. These findings suggest that the Faster mode, coupled with recent advancements in hardware and software, can increase data collection speed by nearly 5 times.

Discussion This guide offers practical insights into optimizing data collection strategies based on the study's target and sample nature. Such optimization ensures the efficient use of resources, leading to reduced costs and time during both data collection and processing. Based on our findings, we recommend using the Faster mode with counted super-resolution, binning 2, and non-gain normalized TIFF output file format for all strategies outlined below.

Keywords Single particle cryoEM, Automated data collection, EPU, Parameters

Introduction

Transmission electron microscopy (TEM) allows for the visualization of structural elements of matter at the atomic scale and has numerous applications in materials science and biological research [1, 2]. Cryo-electron microscopy (cryoEM) is an extension of conventional TEM in which the specimen is vitrified in a thin layer of solution [3–5]. The cryoEM approach, along with recent advancements in hardware [6–8] and software [9, 10] development, has made this technique extremely popular among structural biologists. CryoEM enables

†Janarjan Bhandari, Dmytro Kompaniiets and Ajit Kumar Singh should be regarded as joint first authors.

*Correspondence:
Bin Liu
liu00794@umn.edu
[1] CryoEM Core Facility, The Hormel Institute, University of Minnesota, Austin, MN 55912, USA
[2] Section of Transcription & Gene Regulation, The Hormel Institute, University of Minnesota, Austin, MN 55912, USA

visualization of biological objects from the molecular level (single proteins) to the cellular level (single cell or tissue) in near-native conditions. CryoEM can be utilized in three major approaches: conventional imaging, cryo-electron tomography (cryoET), and cryo-electron scanning microscopy (cryoSEM). While cryoET and cryoSEM offer significant benefits for imaging samples at the cellular level to study surface topology and cellular or subcellular structures [11–15], conventional TEM imaging, followed by single particle analysis, reconstruction, and modeling, remains the primary method for study smaller targets, such as single proteins and protein complexes [13, 16]. Another significant advancement in structural studies using TEM was the development of direct electron detector (DED) cameras. Unlike traditional cameras, DED cameras directly detect electrons and convert them into a signal, which is then processed to create an image. Nowadays, one of the most widely used models is the K3 camera, manufactured by Gatan, which is part of the BioContinuum™ K3® system, consisting of the K3 camera and the GIF Continuum™ filter. The K3 camera can capture images at a rate of 1500 frames per second, enabling it to detect and record individual electron events (electron counting) as they reach the sensor. This capability not only enhances image quality but also reduces noise, improving sensitivity. Unlike the previous K2 detector, the K3 always acquires movies or images in super-resolution mode. Super-resolution imaging provides spatial resolution twice as fine as the physical pixel size. This is achieved by analyzing a single electron event across multiple pixels, allowing the detector to calculate the center of the electron event with sub-pixel precision. By evaluating the same electron multiple times from different pixels, the electron signal is significantly enhanced, enabling the capture of more detailed information and improving the signal-to-noise ratio [17].

High-resolution structural information is essential for understanding biological processes at the molecular level. Such precision enhances our understanding of the flow and regulation of molecular mechanisms, improving drug development and the design of highly specific and efficient therapeutics [18–20], as well as advancing our understanding of physiological processes. To obtain these details, a large volume of data must be collected and processed. Despite the rapid development of cryoEM microscopy, it remains an expensive high-end technique. Additionally, the physical availability of cryoEMs impacts data collection. Currently, available microscopes are struggling to meet the growing demand for structural information, which often leads to cost and availability constraints that hinder the speed of research. A correct experimental design and approach to data collection can save time and funds, significantly increasing the progress of research projects.

Significant advancements have been achieved in developing and utilizing automation for cryoEM data collection [21, 22]. Over the past decade, both academic groups and commercial entities have developed numerous data collection software packages [23, 24]. The most commonly used packages include Leginon [25] and SerialEM [26] from academia, JADAS [27] for JEOL microscopes, and EPU [28] for Thermo Fisher Scientific microscopes. Additionally, Latitude can be used with TEMs equipped with Gatan detectors. Thermo Fisher is the world's largest supplier of cryoEM instruments and equipment. Their cryoEM offerings include the Krios, Arctica, Glacios, and Tundra microscopes. EPU data acquisition software is usually included in the standard Thermo Fisher cryoEM package during microscope acquisition, making it available at the majority of cryoEM facilities. Optimizing the use of EPU data collection parameters will be beneficial for many cryoEM researchers.

Recent software updates for most data acquisition software, including EPU, now support modes of automatic data acquisition based on either accurate stage movement to center each imaging area on the optical axis or image/beam shift for imaging multiple areas with each stage movement (Fig. 1). The difference between the two modes of data acquisition is that stage movements are used to locate and center specific regions to illuminate by the electron beam; or the stage movement is used to locate a center of the region and remain stationary, while the beam itself will be moved to illuminate numerous spots around the center. The minimization of stage movements significantly increases the rate of data acquisition [21].

EPU offers two methods (Fig. 1), based on the principles described, each consisting of three preparatory steps. The first step involves setting up the general session settings (Fig. 1A). The second step allows for selecting squares and holes within these squares that are suitable for data collection (Fig. 1B). The third step involves setting up the acquisition template by defining the acquisition areas within the hole and selecting the location for autofocus (Fig. 1C). Once preparation is complete, you can view the running session screen, where you can monitor the areas where images are being taken and see the latest acquired image (Fig. 1D). Optical centering with Image/Beam Shift is called Faster acquisition. In Faster acquisition mode, EPU processes the Foil Holes in groups. The Faster acquisition procedure uses stage movements to center multiple Foil Holes at once as a group. This reduces the number of stage movements, magnification changes, and lens normalizations. Mechanical centering with stage movement is called

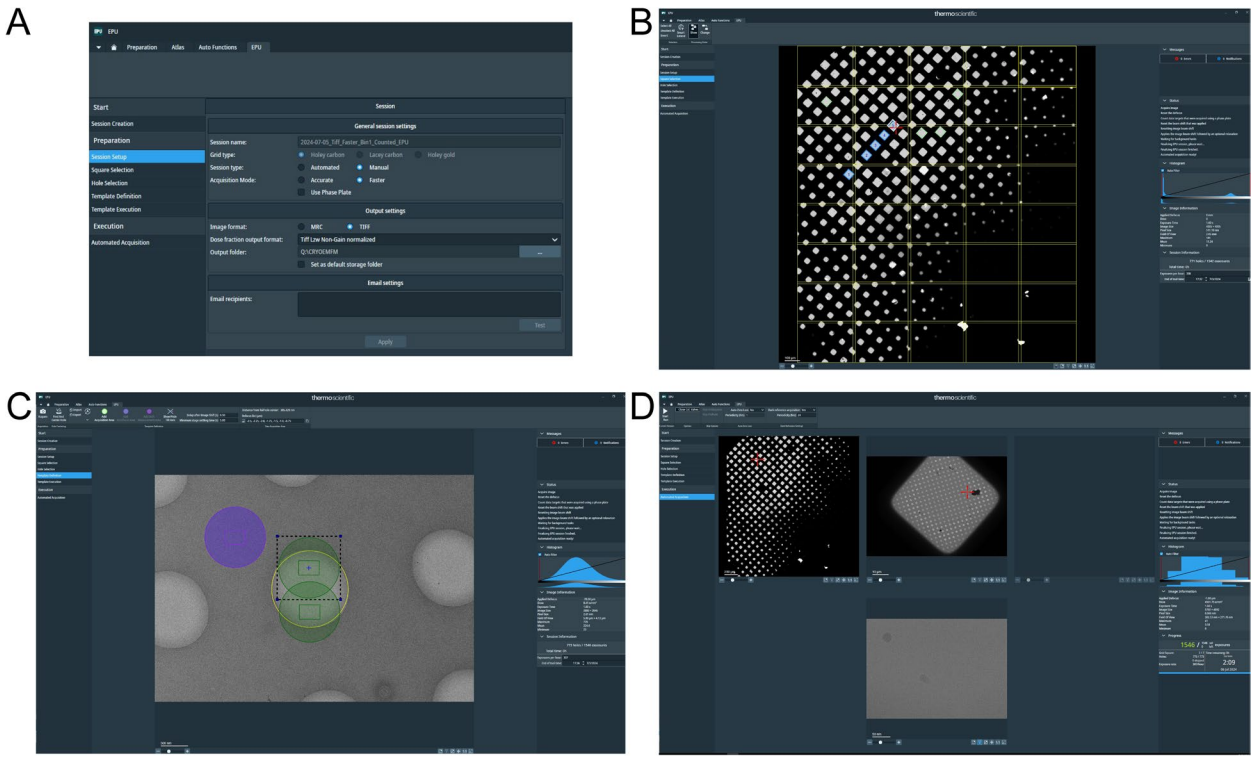

Fig. 1 Example of the automated data collection setup within the EPU software. **A** Initial session setup; **B** Grid square selection; **C** Template setup; **D** Running session

Accurate mode. In Accurate mode, EPU processes each individual Foil Hole independently from its surrounding Foil Holes. EPU first locates the individual Foil Hole, and then accurately centers it in the camera field of view. This procedure requires magnification changes, lens normalizations, and uses small stage shifts to mechanically adjust the position of the specimen.

Here we compare these two automated data acquisition modes and solved cryoEM maps obtained from the processing of collected micrographs. We utilized the VitroEase Apoferritin Standard protein sample available from Thermo Fisher Scientific, which is one of the most used protein standards for microscope quality assurance. We describe the advantages of each mode and provide a comprehensive guide on how experimental design can be approached by utilizing these modes and adjusting specific parameters.

Materials and method

Cryo-EM grid preparation

VitroEase Apoferritin Standard protein (3.5–4.0 mg/ml), purchased from Thermo Fisher Scientific, was used for testing the different modes of data acquisition. Quantifoil holey carbon grids (1.2/1.3; 300 mesh Copper) were glow-discharged for 60 s at 15 mA in a glow discharger (PELCO easiGlow). A 3.5 µL aliquot of Apoferritin protein was applied to the freshly glow-discharged grid using an FEI Vitrobot Mark IV plunger (Thermo Fisher Scientific, USA) under conditions of > 90% humidity at 4 °C, with a blot time of 4 s and zero blot force. The grids were then plunge-frozen in liquid ethane, clipped in autogrid rings, and stored in liquid nitrogen (Supplementary Table 1).

Data collection

Vitrified grids were imaged using a Titan Krios microscope (ThermoFisher, USA) at the Hormel Institute, University of Minnesota cryoEM facility, operating at 300 kV. The microscope was equipped with a K3 direct electron detector and a Gatan BioContinuum energy filter. Six datasets were collected in different modes: (i) Counted super-resolution mode (non-gain normalized TIFF, accurate, with binning 2), (ii) Counted super-resolution mode (non-gain normalized TIFF, faster, with binning 2), (iii) Counted super-resolution mode (non-gain normalized MRC, accurate, with binning 2), (iv) Counted super-resolution mode (non-gain normalized MRC, faster, with binning 2), (v) Counted super-resolution mode (non-gain normalized TIFF, faster, with binning 1), and (vi) Counted mode (non-gain

normalized TIFF, faster, with binning 1). All data sets were collected on the same grid, with different squares selected for each dataset. Automated data collection was performed using the EPU 2.8.0.1256REL software package (Thermo Fisher Scientific) at a slit width of 20 eV and a nominal magnification of 130,000×, with a pixel size of 0.664 Å for five datasets and 0.332 Å for one data set. For all data sets, the exposure was set with an electron dose of 50 e-/Å^2 distributed over 40 frames. A total of ~ 1,200 movies were collected for each data set, with nominal defocus values ranging from −2.5 to −0.75 µm.

Image processing

Cryo-EM data were processed using cryoSPARC v4.5.3 [29], and the procedure was applied consistently across all data sets. Dose-fractionated movies underwent Patch motion correction and Patch CTF estimation. Initial particle picking was performed using the automated Blob picker, with particle diameters set between 100–150 Å. The particles picked by Blob picker were extracted from micrographs using a box size of 336 Å. The extracted particles were subjected to several rounds of 2D classification, and the best 2D classes were selected for automated template-based particle picking. The maximum particle diameter for the template particle picking was set to 150 Å, and the particles were extracted using the same box size of 336 Å. These particles then underwent additional rounds of 2D classification. The best 2D classes from this step were used for Ab-initio reconstruction into three classes, employing octahedral symmetry. Octahedral symmetry was chosen based on the previously solved single-particle cryo-EM structure of Apoferritin. The best Ab-initio classes were selected based on visual inspection of the volume maps in ChimeraX [30], with preference given to the class containing the highest number of particles. This class was used for homogeneous refinement under octahedral symmetry. To ensure consistency across all data sets, ~ 180,000 particles were selected for further processing from each dataset. These particles were subjected to local motion correction, using a box size of 336 Å, followed by homogeneous refinement with octahedral symmetry applied (Supplementary Figs. 1–7). Additionally, the datasets were reprocessed with an extra step: we included .XML files during the movie import stage to input beam shift information. This, along with the Exposure Group Utilities, allowed us to correct any effects of potentially mis-calibrated AFIS beam shifts (Supplementary Fig. 8A and 8B). Fitting of the apoferritin model into the reconstructed maps and figure creation were performed using ChimeraX [30].

Result

Experimental sample and setup

Apoferritin protein has been widely used as a standard for optimizing cryo-electron microscopy (cryoEM) data collection strategies due to its symmetry, stability, and well-characterized structure. To determine the most effective strategies for single-particle cryo-EM data collection using the EPU software on a Titan Krios microscope, we purchased apoferritin protein commercially from Thermo Scientific. This ensured consistency in the protein samples used across various data collection strategies, leading to more reliable and comparable results**.** Our goal was to evaluate the impact of different data collection strategies on total collection time, computational storage requirements, and the quality of the final cryoEM map reconstruction. For this purpose, we collected six datasets using the same vitrified grid, selecting squares of similar ice thickness for each dataset. Key parameters were kept constant across all experiments, including magnification (130,000x), defocus range (−2.5 µm to −0.75 µm), and the number of frames per movie (40) (Supplementary Tables 1–2). The primary variations between the datasets lay in three areas: 1) Acquisition mode: counted mode vs. counted super-resolution mode. 2) Data collection approach: accurate "mechanical centering with stage shift" vs. faster "optical centering with image/beam shift", and 3) Movie file format: MRC vs. TIFF (Fig. 2, Supplementary Figs. 1–8, Table 1).

Comparison of two modes and file formats

Each dataset comprised a similar number of movies, ranging from 1,354 to 1,546. The data collection rate, however, differed significantly between the Faster and Accurate modes. For instance, the counted super-resolution mode with binning 2 and faster optical centering achieved a high collection rate of 736–758 movies per hour. In contrast, the counted super-resolution mode with binning 2 and accurate mechanical centering had a slower rate, collecting between 134 and 160 movies per hour. This difference in collection rates highlights the trade-off between speed (Faster mode) and precision (Accurate mode). Data storage requirements also varied depending on the format of the movie files. Movies saved in TIFF format required approximately 166 MB per file, while those saved in MRC format required significantly more storage, averaging around 920 MB per file. All datasets were processed using cryoSPARC v4.5, following identical workflows. Around 180,000 particles were selected for homogeneous refinement in each experiment to maintain consistency, applying octahedral symmetry. In terms of resolution, the best cryoEM map was obtained using the counted super-resolution accurate mode with binning

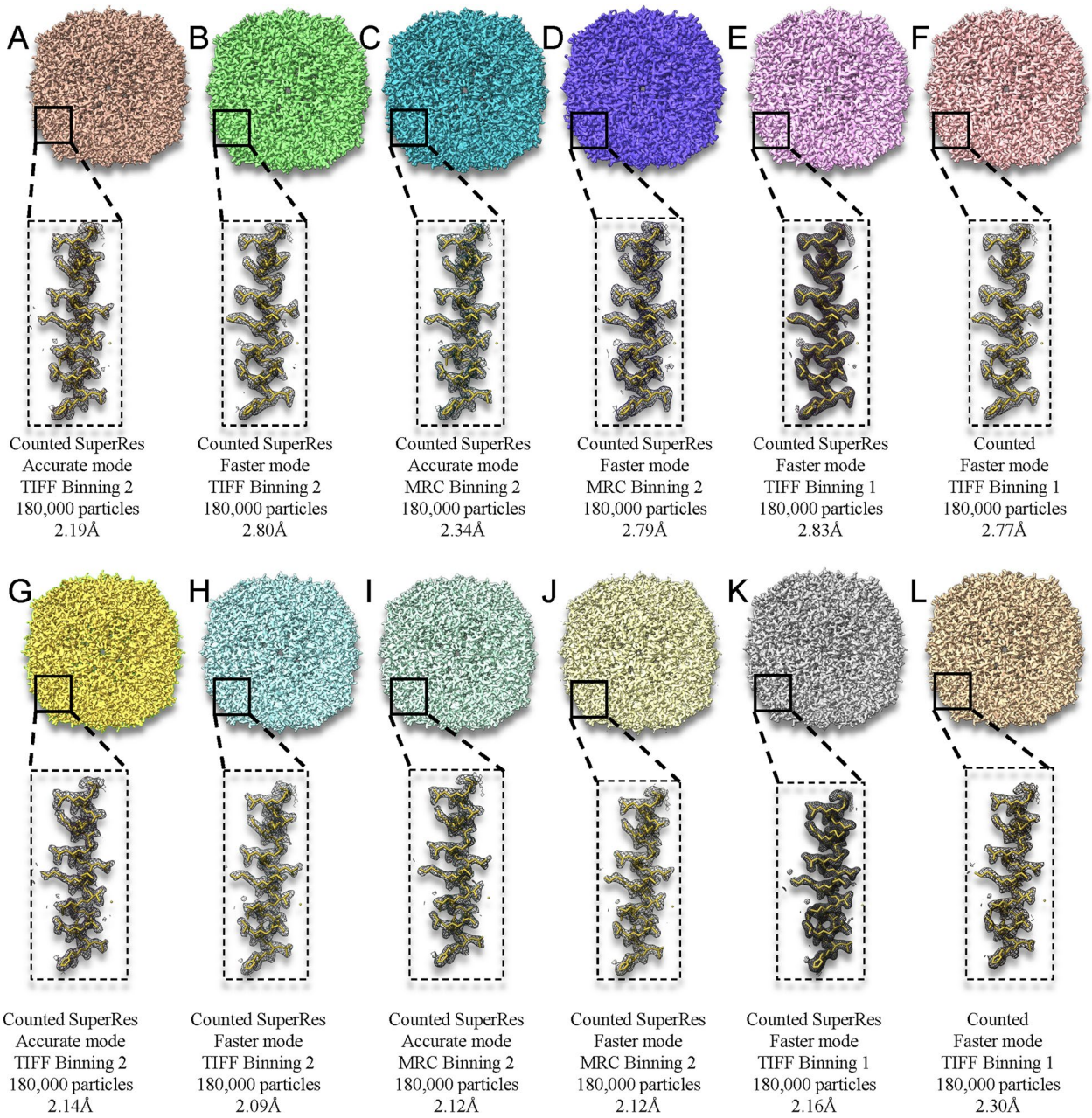

Fig. 2 Overview of data processing results. The final cryoEM maps for all six experiments. **A**-**F** Experiment 1–6; **G**-**L** Experiment 1–6 with Beam Shift corrected. The inserts show the local maps with models and experimental summaries

2 in TIFF format (Experiment 1), yielding a resolution of 2.19 Å (Fig. 2A). In comparison, the counted super-resolution faster mode with binning 2 in TIFF format (Experiment 2) resulted in a lower resolution of 2.80 Å (Fig. 2B), despite the faster data collection rate. A similar trend was observed with the data collected in MRC format. The counted super-resolution accurate mode with binning 2 (Experiment 3) produced a resolution of 2.34 Å (Fig. 2C), while the counted super-resolution faster mode with binning 2 (Experiment 4) yielded a lower resolution of 2.79 Å (Fig. 2D). Additionally, we collected two datasets (Experiment 5 and 6) in counted super-resolution and counted modes, using both Faster mode and binning 1 for comparison. These datasets were processed using the same workflow, resulting in similar resolutions (Fig. 2E and F). However, the file sizes differed by a factor of two, with the counted super-resolution files being larger.

Table 1 Data collection and processing parameters

Parameters	Experiment 1 (EMD-48840)	Experiment 2 (EMD-48841)	Experiment 3 (EMD-48842)	Experiment 4 (EMD-48843)	Experiment 5 (EMD-48844)	Experiment 6 (EMD-48845)
Pixel Size (Å)	0.66	0.66	0.66	0.66	0.33	0.66
Acquisition mode	Counted Super resolution	Counted Super resolution	Counted Super resolution	Counted Super resolution	Counted Super resolution	Counted
Mode	Accurate	Faster	Accurate	Faster	Faster	Faster
Image format	TIFF	TIFF	MRC	MRC	TIFF	TIFF
Binning	2	2	2	2	1	1
Bit depth	8	8	8	8	8	8
Images per hour	134	736	160	758	338	307
Total number of images	1354	1540	1582	1546	1542	1546
CTF average (Å)	~2.5	~2.5	~2.5	~2.5	~2.5	~2.5
Number of squares selected	5	5	5	5	5	7
Number of holes selected	677	770	791	773	771	773
Each movie size	~166 MB	~166 MB	~920 MB	~920 MB	~288 MB	~166 MB
Final number of particles	~180,000	~180,000	~180,000	~180,000	~180,000	~180,000
Symmetry	Octahedral	Octahedral	Octahedral	Octahedral	Octahedral	Octahedral
Resolution	2.19 Å	2.80 Å	2.34 Å	2.79 Å	2.83 Å	2.77 Å
Resolution after beam shift correction	2.14 Å	2.09 Å	2.12 Å	2.12 Å	2.16 Å	2.30 Å

Upon observing discrepancies in the results, we decided to reprocess our data by including the .XML files, which contain information about beam shifts. This approach corrects potential AFIS miscalibration or beam misalignment in both the Accurate and Faster modes. The results after beam shift correction showed minimal improvement for the Accurate mode but significant enhancement in the resolution of models reconstructed from data acquired in Faster mode. Specifically, after correction, Experiments 1 and 3 achieved resolutions of 2.14 Å and 2.12 Å, respectively (Fig. 2G and I), while Experiments 2 and 4 improved to 2.09 Å and 2.12 Å, respectively (Fig. 2H and J). This suggests that Faster mode, with its much faster acquisition speed and comparable movie quality, may now offer greater benefits. Following reprocessing, Experiments 5 and 6 showed resolution improvements similar to those of Experiments 2 and 4 (Fig. 2K and L). Additionally, the choice of file format (TIFF vs. MRC) significantly affects data storage requirements, with the MRC format demanding substantially more space without significantly improving resolution.

Discussion

Utilization of cryo-EM has greatly increased over the past 10 years, as evidenced by the year-over-year trend of rising numbers of cryoEM structures deposited in public databases and published in peer-reviewed journals. However, the technical development of hardware and software, coupled with the need for appropriate facilities to house the equipment and expensive service contracts for maintenance, creates a high overhead cost for cryoEM research. While national cryoEM research centers may offer free data collection, access is typically limited to applicants who provide strong preliminary evidence and detailed experimental annotations, reasoning, and planning. Additionally, there is often a long waiting time for free data collection at these facilities. Other facilities may pass on at least some of their costs to the researchers utilizing their cryoEM resources. Therefore, it is essential to know how to efficiently acquire the data needed for a cryoEM project.

Acquisition speed and its effect on experiments

The speed of data collection is one of the main factors affecting a user's time and associated fees. Thermo Fisher Scientific's EPU software offers two modes of data acquisition known as Faster and Accurate. The Faster mode utilizes the Aberration Free Image Shift (AFIS) software to enable imaging of areas that are not on the optical axis without requiring stage movement. Cryo-EM grids with an array of holes can have images collected from multiple holes with a single stage movement for each grouping

of holes using this mode. However, the Image/Beam Shift may induce coma and astigmatism under sub-optimal conditions, which may decrease the final image quality [22, 31, 32]. Now with the introduction of AFIS in a data collection workflow, the faster mode should produce data of comparable quality. Like any other component, AFIS calibration requires regular maintenance. However, if the AFIS calibration is not ideal, it can still be addressed during the image processing stage. For example, cryoSPARC software offers Exposure Group Utilities. Acquired movies can be imported along with .XML files that contain information about the image/beam shifts in X and Y coordinates. This allows for the grouping and correction of micrographs based on this data. Additionally, this approach can help to potentially improve the quality of data acquired with Accurate mode by correcting potential beam shifts or misalignments. An improvement in the presented structure is marginal but could be more significant depending on a particular dataset or data collection.

In Accurate mode, the stage is moved to center each Foil hole on the optical axis, and images are acquired from only one hole per stage movement. The more frequent stage movements in Accurate mode, along with changes in magnification and lens normalization, significantly slow down the rate of data acquisition. In fact, Accurate data acquisition may be up to five times slower than Faster data acquisition (Table 1). Consequently, the differences in data acquisition modes provide distinct advantages in terms of the quantity of data collected within a specific timeframe.

For example, if the sample has a preferential orientation problem, more images may be required to adequately sample the less-represented views. Faster speed can also be beneficial when coupled with the online data processing approach using cryoSPARC Live. The higher volume of data allows for further assessment of the structure. For example, if the solved structure indicates non-optimal sample conditions (such as chemical compound not binding to the target protein or an incomplete or damaged protein structure), appropriate decisions can be made to optimize conditions or switch to different samples for the remainder of data collection period. Additionally, the more practical side of faster data collection is that it can accommodate more user slots in a facility shared by numerous users.

File format and space requirements

Another parameter tested was the format of stored data. The main difference between MRC and TIFF images, from our perspective, is their size since Experiments 1 and 3 (Movies collected in Accurate mode), or 2 and 4 (Movies collected in Faster mode) show similar resolutions during processing (Table 1). TIFF images take roughly 5 times less space, making storage management significantly easier. The quality of information stored in both formats is comparable (Fig. 2G-L), with the main difference being how metadata is stored within the file. Both formats also have an 8-bit pixel depth. As our data processing results show, the quality of solved models is nearly identical, with negligible variation between 2.12 and 2.09 or 2.14 Å for MRC and TIFF format stored data respectively. Such a difference does not contain any significant information that affects the final structural model. Different file formats may impact compatibility with various software used for data processing, but there are usually methods to convert TIFF to MRC if necessary. Storage space is another factor that influences the potential cost of an experiment, especially since movies recorded in TIFF format under LZW file compression would be the optimal choice.

Binning and its effect

Binning of the collected data affects both data collection speed and file size. Binning changes how pixels are represented between the sensor and the image: binning 1 corresponds to a single sensor pixel per image pixel while binning 2 groups four sensor pixels into one final image pixel. Higher binning enhances the contrast of naturally low-contrast biological samples. The K3 always collects images in super-resolution mode, producing output that closely corresponds to images collected in super-resolution counting mode with binning 1 (unbinned), resulting in a pixel size of 0.33 Å in our case. During data collection with EPU and the K3 camera, binning can be performed in two ways. When data is collected using counted super-resolution mode with binning 1, the K3 collects movies in its default super-resolution format and directly transfers them to EPU without binning at super-resolution. If data is collected using counted super-resolution mode with binning 2, the K3 collects movies in its default super-resolution format, bins them twofold, and transfers them to EPU with a physical resolution. When data is collected in counted mode with binning 1, the K3 collects movies in its default super-resolution format and transfers them to EPU, where twofold binning is applied to achieve a physical resolution. Therefore, movies collected in counted mode with binning 1 and those collected in super-resolution mode with binning 2 should have the same physical pixel size of 0.66 Å. However, binning in EPU is slower than binning on the camera (EPU vs. Gatan Microscopy Suite (GMS), respectively). This difference results in approximately twice the speed of data acquisition and a twofold reduction in file size between counted binning 1 and counted super-resolution binning 2 (Table 1). Although higher binning decreases

the resolution of the final image [33], the quality of the final maps from our experiments remains similar for counted super-resolution mode with binning 2 or binning 1 (Table 1, Fig. 2G-K). However, counted with binning 1 shows slightly lower resolution compared to the others (Table 1, Fig. 2L). The maximal physical resolution provided by the hardware is often not achievable for the biological samples, so trade-offs between binning, data acquisition speed, and expected maximum quality of the final image may be considered to optimize time and computational resources.

Counted super-resolution vs counted

The difference between counted and super-resolution counted modes theoretically allows for a twofold increase in maximum achievable resolution. Data in experiments 5 and 6 were collected in counted super-resolution and counted modes, respectively. Both data collections were performed with binning set to 1, under faster collection mode, and with output movies saved in the non-gain normalized TIFF format. We observed that the data collection speed for both attempts decreased by 2.3 times, from ~ 750 movies per hour to ~ 325 movies per hour, but remained similar for both counted super-resolution and counted modes. Movies saved in counted mode required ~ 166 MB per file, while counted super-resolution movies ~ 288 MB per file. Data processing followed similar methods as in previous experiments, resulting in two cryoEM maps with nearly identical resolutions of 2.83 Å and 2.77 Å or 2.16 Å and 2.30 Å after AFIS correction (Fig. 2E, F, K and L) for the counted super-resolution and counted modes, respectively. Following the binning logic described earlier, counted super-resolution and counted modes with binning set to 1 will have similar data collection speed. However, counted super-resolution could theoretically yield better results due to the smaller pixel size of 0.33 Å. While counted super-resolution with binning set to 2 and counted mode with binning set to 1 will have similar pixel sizes but a twofold difference in data collection speed. The lack of significant differences between the cryoEM maps may be attributed to the excellent performance of the sample and the imposed symmetry, which could smooth out any variations.

Experimental design suggestions

Maintaining a balance between cost and efficiency can be quite challenging, and defining the desired experimental output is crucial in determining this balance. Aiming for the highest resolution where it is not needed can lead to significant cost increases. On the other hand, if the obtained resolution doesn't provide an answer, it may lead to significant delays in the experiment and cost increases due to the necessity of repeating data collection. Under comparable conditions, such as an equivalent number of collected images and particles used during processing, the Accurate mode yielded a 2.19 Å resolution map, while the Faster mode produced a 2.8 Å resolution map. This difference in resolution is attributed to the suboptimal AFIS calibration, as all datasets achieved ~ 2.12 Å resolution maps after image aberration correction during data processing. When beam alignment and AFIS calibration are correct, no significant resolution difference may be observed between Accurate and Faster modes. It is also important to note that beam shift grouping and correction can be performed during data processing, even if AFIS calibration is not perfectly aligned.

Depending on the goals of the experiment and available resources, the following points should be considered (Fig. 3). Based on our findings, we recommend acquiring data in Faster mode using counted super-resolution with binning 2 and non-gain normalized TIFF output file format for all strategies described below.

The first strategy is screening, which primarily aims to assess the initial quality and properties of the sample. Screening is typically performed using mid-range equipment, such as 100 kV or 200 kV microscopes (e.g., Tundra, Glacios, or Arctica), for preliminary sample evaluation. These microscopes are all compatible with EPU software. Typically, up to one hundred movies can be collected, with the number of movies depending on factors such as particle density and sample concentration. Screening helps estimate particle quality and quantity, and optimize biochemical and freezing conditions, such as buffer composition, complex assembly stoichiometry, and grid preparation conditions. A key consideration during screening is selecting the appropriate magnification and pixel size. Generally, the pixel size should be chosen based on the desired final resolution, typically about one-third of it, to achieve ~ 67% of the Nyquist frequency. For larger objects, a larger pixel size is recommended, as the magnification is lower to accommodate their size. These parameters do not directly influence the data collection settings but should be tailored to individual samples for optimal downstream processing.

The second one is initial (trial) data collection. This can be performed on both mid-range and high-end microscopes, such as Glacios, Arctica, or Titan Krios, depending on availability. Based on the results from screening, larger datasets—up to a thousand images—are typically collected. For certain experimental purposes, trial data collection may provide sufficient information to answer the research questions. For example, if your target is well-characterized biochemically and high resolution is not required, such as for studying complex assembly or the overall architecture of large molecular complexes, this approach can significantly reduce both costs and time

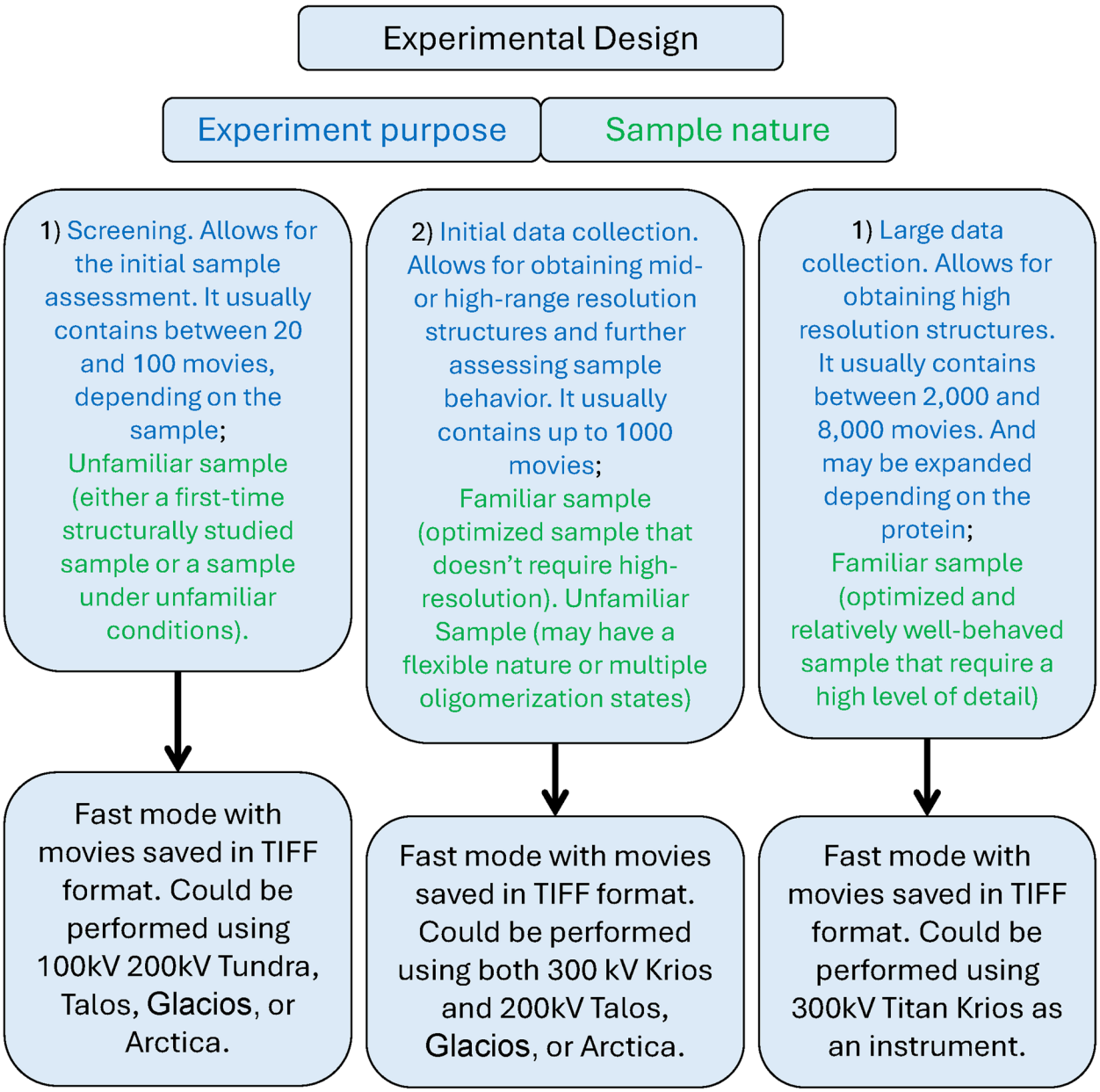

Fig. 3 CryoEM session design. Three general routes that can be used to optimize data collection cost and time for different samples and target resolution

required for data collection, without aiming for the highest possible resolution.

The third one is large (final) data collection, typically performed using high-end 300 kV microscopes like the Titan Krios. The dataset size generally ranges from 2,000 to 8,000 movies, but it can be adjusted based on the sample requirements or desired resolution. For example, when working with a previously characterized protein or sample and aiming to improve resolution to visualize specific structures, this step is essential. In contrast, for novel or uncharacterized targets, or for dynamic and flexible proteins or unstable binding partners, it may be beneficial to collect even larger datasets. This allows for particle sorting into subsets during data processing, which can provide insights into dynamic molecular interactions. Additionally, larger datasets may aid in refining focus in specific regions.

Concluding remarks

In conclusion, based on our experience and results, EPU automated data collection should generally be performed using the Faster acquisition mode, which typically

includes AFIS correction. Output movies should be recorded in counted super-resolution mode with binning 2, using the TIFF file format with LZW compression. For optimal results, data processing should incorporate the import of .XML files and optical grouping.

Many steps in cryo-EM can be performed on lower-end microscopes, and the data obtained may suffice, potentially eliminating the need for more expensive equipment or preparing the sample efficiently for cost- and time-effective data collection. Researchers should clearly define their experimental questions, as this will determine the required resolution, resources, and time needed for data collection. It is essential to assess whether higher resolution is truly necessary to answer these questions. For example, studying specific protein–protein interactions may only require a resolution of around 3 Å. While a 2.2 Å resolution could provide clearer details, the additional time and cost to achieve it may not yield new insights.

Additionally, the dynamics and flexibility of the sample can affect the final quality of cryo-EM maps, and these factors may present biochemical constraints that limit resolution. Therefore, optimizing the sample prior to data collection is crucial to avoid delays and maximize the efficiency of microscope usage. Future improvements in data collection strategies will require analyzing a broader range of samples and adjusting processing workflows accordingly. A critical aspect of designing a cryo-EM experiment is recognizing that different samples may not behave similarly at all times. While similar samples may show comparable behaviors, it is important to start with tested conditions while remaining flexible to adjust as needed in response to unexpected outcomes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44330-025-00025-8.

Supplementary Material 1.

Acknowledgements
Not applicable

Authors' contributions
J.B. and B.L. initiated, designed, and supervised the experiments. C.B. performed protein sample and grids preparations, J.B., A.K.S., J.P. performed screening, and optimization, A.K.S. and D.K. performed image processing and map reconstruction. J.B. conducted high throughput data collection on Titan Krios. All authors contributed to the analysis of the data and the interpretation of the results. D.K., A.K.S., and B.L. wrote the manuscript with contributions from the other authors.

Funding
CryoEM core (J.B, C.B, A.K.S and J.P), The Hormel Institute, University of Minnesota (B.L.)

Data availability
The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank with accession numbers from EMD-48840 to EMD-48845 for experiment 1–6, respectively.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 8 October 2024 Accepted: 10 February 2025
Published online: 17 February 2025

References

1. Saibil HR. Cryo-EM in molecular and cellular biology. Mol Cell. 2022;82(2):274–84.
2. Single-particle Cryo-EM of Biological Macromolecules: IOP Publishing; 2021. Available from: https://doi.org/10.1088/978-0-7503-3039-8.
3. Dandey VP, Budell WC, Wei H, Bobe D, Maruthi K, Kopylov M, et al. Time-resolved cryo-EM using Spotiton. Nat Methods. 2020;17(9):897–900.
4. Naydenova K, Jia P, Russo CJ. Cryo-EM with sub-1 A specimen movement. Science. 2020;370(6513):223–6.
5. Levitz TS, Weckener M, Fong I, Naismith JH, Drennan CL, Brignole EJ, et al. Approaches to Using the Chameleon: Robust, Automated, Fast-Plunge cryoEM Specimen Preparation. Front Mol Biosci. 2022;9: 903148.
6. McMullan G, Naydenova K, Mihaylov D, Yamashita K, Peet MJ, Wilson H, et al. Structure determination by cryoEM at 100 keV. Proc Natl Acad Sci U S A. 2023;120(49): e2312905120.
7. Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown P, et al. Single-particle cryo-EM at atomic resolution. Nature. 2020;587(7832):152–6.
8. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Atomic-resolution protein structure determination by cryo-EM. Nature. 2020;587(7832):157–61.
9. Kimanius D, Dong L, Sharov G, Nakane T, Scheres SHW. New tools for automated cryo-EM single-particle analysis in RELION-4.0. Biochem J. 2021;478(24):4169–85.
10. Mendez JH, Chua EYD, Paraan M, Potter CS, Carragher B. Automated pipelines for rapid evaluation during cryoEM data acquisition. Curr Opin Struct Biol. 2023;83: 102729.
11. Baumeister W. Cryo-electron tomography: A long journey to the inner space of cells. Cell. 2022;185(15):2649–52.
12. Turk M, Baumeister W. The promise and the challenges of cryo-electron tomography. FEBS Lett. 2020;594(20):3243–61.
13. Chua EYD, Mendez JH, Rapp M, Ilca SL, Tan YZ, Maruthi K, et al. Better, Faster, Cheaper: Recent Advances in Cryo-Electron Microscopy. Annu Rev Biochem. 2022;91:1–32.
14. Lazic I, Wirix M, Leidl ML, de Haas F, Mann D, Beckers M, et al. Single-particle cryo-EM structures from iDPC-STEM at near-atomic resolution. Nat Methods. 2022;19(9):1126–36.
15. Huang C, Kim JS, Kirkland AI. Cryo-electron ptychography: Applications and potential in biological characterisation. Curr Opin Struct Biol. 2023;83: 102730.
16. Guaita M, Watters SC, Loerch S. Recent advances and current trends in cryo-electron microscopy. Curr Opin Struct Biol. 2022;77: 102484.
17. Feathers JR, Spoth KA, Fromme JC. Experimental evaluation of super-resolution imaging and magnification choice in single-particle cryo-EM. J Struct Biol X. 2021;5: 100047.
18. Renaud JP, Chari A, Ciferri C, Liu WT, Remigy HW, Stark H, et al. Cryo-EM in drug discovery: achievements, limitations and prospects. Nat Rev Drug Discov. 2018;17(7):471–92.
19. Lees JA, Dias JM, Han S. Applications of Cryo-EM in small molecule and biologics drug design. Biochem Soc Trans. 2021;49(6):2627–38.
20. Zhu KF, Yuan C, Du YM, Sun KL, Zhang XK, Vogel H, et al. Applications and prospects of cryo-EM in drug discovery. Mil Med Res. 2023;10(1):10.
21. Cheng A, Yu Y. Recent advances in data collection for Cryo-EM methods. Curr Opin Struct Biol. 2024;86: 102795.

22. Cheng A, Eng ET, Alink L, Rice WJ, Jordan KD, Kim LY, et al. High resolution single particle cryo-electron microscopy using beam-image shift. J Struct Biol. 2018;204(2):270–5.
23. Tan YZ, Cheng A, Potter CS, Carragher B. Automated data collection in single particle electron microscopy. Microscopy (Oxf). 2016;65(1):43–56.
24. Bepler T, Borst AJ, Bouvette J, Cannone G, Chen S, Cheng A, et al. Smart data collection for CryoEM. J Struct Biol. 2022;214(4): 107913.
25. Suloway C, Shi J, Cheng A, Pulokas J, Carragher B, Potter CS, et al. Fully automated, sequential tilt-series acquisition with Leginon. J Struct Biol. 2009;167(1):11–8.
26. Schorb M, Haberbosch I, Hagen WJH, Schwab Y, Mastronarde DN. Software tools for automated transmission electron microscopy. Nat Methods. 2019;16(6):471–7.
27. Zhang J, Nakamura N, Shimizu Y, Liang N, Liu X, Jakana J, et al. JADAS: a customizable automated data acquisition system and its application to ice-embedded single particles. J Struct Biol. 2009;165(1):1–9.
28. Thompson RF, Iadanza MG, Hesketh EL, Rawson S, Ranson NA. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. Nat Protoc. 2019;14(1):100–18.
29. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14(3):290–6.
30. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. 2023;32(11): e4792.
31. Glaeser RM, Typke D, Tiemeijer PC, Pulokas J, Cheng A. Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. J Struct Biol. 2011;174(1):1–10.
32. Wu C, Huang X, Cheng J, Zhu D, Zhang X. High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift. J Struct Biol. 2019;208(3): 107396.
33. Guo F, Jiang W. Single particle cryo-electron microscopy and 3-D reconstruction of viruses. Methods Mol Biol. 2014;1117:401–43.

Publisher's Note