

VERSATILITY OF CRYO-ELECTRON MICROSCOPY AS A STRUCTURAL TECHNIQUE
INFORMS IRON MINERAL NUCLEATION AND GROWTH IN A MINI-FERRITIN

by

Colin Charles Gauvin

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2024

©COPYRIGHT

by

Colin Charles Gauvin

2024

All Rights Reserved

DEDICATION

This thesis is dedicated to my children, Charles and Eloise. There is no greater motivation than wanting to be the person your children think you are.

ACKNOWLEDGEMENTS

I would like to begin my acknowledgements by thanking my spouse, Mary, who has been everything I have not been for the past six years. This thesis simply would not exist without her, her unconditional support is easily the single greatest asset I have.

I would also like to thank my father, Charles Gauvin, who taught me the importance of doing a good job first and foremost, for yourself. To quote him: “If it’s worth doing, it’s worth doing right”. I’d also like to thank my mother Jane Gauvin, who embedded in me, at an early age, a general curiosity that persists to this day. She did not just answer the infinite questions I had growing up but also taught me to answer them myself.

I would like to thank my advisor – Martin Lawrence, who gave me a chance despite my non-traditional background, and quickly identified my strengths and helped me to discover a niche in which I could make real contributions.

I would also like to thank my undergraduate chemistry professor, Edward Neth, who taught me many things about chemistry and academic learning, but the most important lesson being that hard work alone is sufficient to accomplish anything.

I would like to thank Daniel Bollschweiler, who treaded a very similar path before me, proving that it could be done. Thank you for being my positive control and helping me along in your footsteps. I would also like to thank Jonathan Bouvette, who authored SmartScope and spent many hours helping me implement it at MSU. Additionally, both Daniel and Jonathan have helped to foster a great EM community and engage in endless conversations about all things microscopy.

TABLE OF CONTENTS

1. INTRODUCTION AND RESEARCH OBJECTIVES	1
Establishment of a Cryo-EM Facility for Structural Biology	1
Major Cryo-EM Methodologies	3
Single Particle Averaging.....	3
Cryo-Electron Tomography.....	7
Supporting Infrastructure	10
Single Particle Analysis of Biomineralization in <i>Pyrococcus furiosus</i> Dps-Like Protein.....	11
Research Objectives.....	13
Research Questions	13
2. CRYO-EM CORE FACILITY	15
Operational Infrastructure	15
Environmental Controls	16
Microscope Control Software	17
SerialEM.	17
SmartScope.	18
Computational Infrastructure	20
Networking & Connectivity.....	22
Data Storage & Management.....	26
Data Backups	29
CryoSPARC on Tempest.....	29
Using CryoSPARC.....	29
Managing CryoSPARC.	32
Single Particle Protocols	38
SerialEM	38
Screening.....	38
Data Collection.	62
SmartScope	88
Screening.....	89
Data Collection	94
Tomography Protocols	98
SerialEM Batch Mode.....	98
Selecting Targets.	98
Recording Tilt-Series.	106
PACETomo	107
Selecting Targets	108
Recording Parallel Tilt-Series.	112

TABLE OF CONTENTS CONTINUED

3. THE MECHANISM OF IRON NUCLEATION AND GROWTH IN A MINI-FERRITIN.....	113
Introduction.....	113
Results.....	116
Structure of apo Pf-Tftn at 1.86 Å resolution	116
Disordered N-Terminal Tail	119
Ordered C-Terminal Tail	119
Iron Loading.....	120
Nucleated Thioferritin.....	121
Nucleated Structure.....	122
Protein-Mineral Interactions I.....	125
Glutamate 50	125
Glutamate 54.....	126
Ferrihydrite	126
Nucleation Kinetics.....	127
Iron Loaded Structure	131
Role for the C-terminal Tail	133
Protein Mineral Interactions II.....	133
Mineral Growth from a Subset of C-terminal Tails	134
Dynamic Pore Structures	135
The Ferroxidase Center.....	135
Discussion	137
A Cryo-EM Standard	137
Strategies for Single Particle Analysis of Iron Loaded Ferritins	137
Tryptophan Fluorescence.....	138
Nucleation	139
Biom mineralization.....	140
Structure of the Mineralized Core.....	141
Methods.....	142
Expression and Purification of Pf-Tftn	142
Iron Loading.....	143
Vitrification	143
Cryo-TEM Data Collection.....	144
Single Particle Analysis for Apo-Pf-Tftn.....	144
Single Particle Analysis for Nucleated Mineral Dataset.....	145
Single Particle Analysis for Holo-Pf-Tftn.....	146
Fluorescence Lifetime.....	147
REFERENCES CITED	148

TABLE OF CONTENTS CONTINUED

APPENDICES.....	157
FACILITY MANAGEMENT SCRIPTS	158
Linux System Configuration.....	159
SSSD Configuration File	159
pam_mount Configuration File.....	160
CryoSPARC Scripts & Configuration.....	161
CryoSPARC Master Configuration File	161
SystemD Service File.....	162
Database Backup SystemD Service File.....	164
Tempest Lane Cluster Info File.....	164
Tempest Lane Cluster Submission Script	165
Delete Rejected Exposures Script.....	166
SerialEM Scripts & Troubleshooting.....	168
Startup Script	168
Atlas Script.....	169
Incorrect “View Mode” Defocus Offset.....	170
View-Record Alignment	171
SmartScope Scripts & Troubleshooting.....	172
Aperture Fix for Arctica Interface.....	172
Exposure Groups by Image Shift.....	173
Exposure Groups by MDOC.....	178
Talos Arctica Configuration.....	183
Parallel Illumination.....	183
OURRStore Tape Archival.....	184
THIOFERRITIN SUPPLEMENTAL INFORMATION.....	185

LIST OF TABLES

Table	Page
1. Table 1. Recommended starting configuration for SPA data collection, magnifications 8,500x-92,000x. Spot size is the recommended spot size for each magnification to achieve optimal camera dose at parallel illumination. C2 illumination settings available in Appendix A.	80
2. Table 2. SerialEM X/Y coordinate nomenclature. This nomenclature is appended to the end of the file name when saving multishot positions in the file name.....	87
3. Table 3. C2 lens intensity for parallel illumination at gun lens setting 3.7.....	183
4. Table 4. C2 lens intensity for parallel illumination at gun lens setting 4.....	183
5. Table 5. Structure validation report for Apo-Pf-Tftn.....	188
6. Table 6. Structure validation for nucleated Pf-Tftn.	189
7. Table 7. Structure validation for full mineral Pf-Tftn.....	191
8. Table 8. Structure validation for intermediate mineral Pf-Tftn.	192

LIST OF FIGURES

Figure	Page
1. Figure 1 Network diagram describing data-flow in the MSU Cryo-EM Facility. Data are recorded by the K3 camera and transferred by fiber optics to the camera controller (Megaera). Once there, data must pass through a proxy (hydra) to be accessible on the MSU network and via Globus. Globus can automatically move data to central storage such as Blackmore, or any other storage system, where it can then be operated on.....	23
2. Figure 2. An example of shared directories in the Globus web user interface. The Cryo-EM Facility shares are managed through the Globus UID of msucryoem and can be shared with any MSU NetID.....	27
3. Figure 3. Globus real-time synchronization options button.	27
4. Figure 4. Globus timer options.	28
5. Figure 5. Lane selection in the CryoSPARC web user interface. Each research group will have separate lanes created for all of their Tempest partitions, and each group's lane will only be visible to members of that group.	30
6. Figure 6. Custom cluster submission script variables. Note that "time_hr_min_sec" describes the format of the time entry. Command displays explicit submission commands to wrap the CryoSPARC job in – this can be useful for extending CryoSPARC functionality or wrapping the job in standard Unix software such as "time". The ram_gb_multiplier option is how much to multiply CryoSPARC's default memory allocation by – this is necessary as CryoSPARC chronically under-requests memory from the cluster for certain job types. The "gres" option allows selection of a "generic resource", on Tempest this describes the type of GPU. It defaults to A40, but A100 and H100 GPUs are also available depending on the requested lane.	31
7. Figure 7. Admin panel location in the CryoSPARC web user interface.	36
8. Figure 8. Cluster configuration settings in the CryoSPARC web user interface.....	37
9. Figure 9. User lane access management in the CryoSPARC web user interface. Marking a lane as "Available" to a user will let them submit jobs to that lane. The "Restricted" status will prevent the lane from being visible to that user.	37

LIST OF FIGURES CONTINUED

Figure	Page
10. Figure 10. Storage space on Gatan K3 data collection computer.	39
11. Figure 11. Auto-loader and temperature control panel. Note that temperatures will be displayed in red when above the recommended cryo-operation temperature.	40
12. Figure 12. Auto-loader inventory mechanism, which will check the status of each slot in the microscope cassette.	41
13. Figure 13. SerialEM current working directory dialog allows the user to select their base directory when using SerialEM.	42
14. Figure 14. SerialEM navigator menu full-montage selection.	43
15. Figure 15. SerialEM full montage setting, depicting the proper number of pieces, and low-dose mode selection.	44
16. Figure 16. SerialEM montage file-saving settings. MRC is recommended for grid atlas format.	46
17. Figure 17. Talos Arctica aperture control panel depicting available aperture options.	47
18. Figure 18. SerialEM navigator dialog window. Double click to load the desired grid map.	49
19. Figure 19. Talos Arctica auto-loader loading dialog. Gray numbers represent empty cassette slots, blue numbers represent full cassette slots, and the presence of a green checkmark denotes that the cartridge in that location has been loaded onto the microscope stage at some point.	50
20. Figure 20. Talos Arctica auto-loader turbo molecular pump controls. Also note the yellow cassette slow (#3), denoting the active cartridge currently on the stage.	51
21. Figure 21. SerialEM navigator dialog window demonstrating "add points" routine.	53
22. Figure 22. Low magnification image of an EM grid demonstrating added points location. Red crosses depict placed points, the yellow cross depicts the currently selected point in the navigator.	55

LIST OF FIGURES CONTINUED

Figure	Page
23. Figure 23. SerialEM “Low Dose Control” panel toggle.....	57
24. Figure 24. SerialEM "Camera & Script" control panel.	58
25. Figure 25. SerialEM "Tasks" menu showing "Eucentricity" options.	58
26. Figure 26. SerialEM "Low Dose Control" panel showing options to define position of "Focus" area.....	60
27. Figure 27. Medium magnification image demonstrating how to properly define the "Focus" area for data collection. The yellow box denotes the location of the focus position relative to the green box, which denotes the current “Record” position. The delta between the two positions will be preserved once set.	61
28. Figure 28. SerialEM "Montage Controls" panel.....	63
29. Figure 29. SerialEM "Imaging States" dialog window.....	64
30. Figure 30. SerialEM "Montaging & Grids" sub-menu.	65
31. Figure 31. SerialEM "Hole Finder" dialog window and settings.	66
32. Figure 32. Location of "Set-Multi-shot Parameters" dialog in the "Montaging & Grids" sub-menu.	67
33. Figure 33. SerialEM multi-shot parameters dialog window and settings. “Do records in multiple holes” will enable the multi-shot function of the microscope.	69
34. Figure 34. Medium magnification micrograph demonstrating proper hole centering for multishot setup. The red cross displays the current center of the microscope stage.	71
35. Figure 35. Medium-magnification micrograph depicting a green boxed hole to crop for template generation.	72
36. Figure 36. SerialEM template alignment settings dialog window.	73
37. Figure 37. Grid square map magnification image. The yellow cross denotes the currently selected navigator point.	74

LIST OF FIGURES CONTINUED

Figure	Page
38. Figure 38. Aligning the grid square magnification to the grid atlas coordinates. The green cross represents the user's selection of the feature to align to, while the yellow cross represents where SerialEM believes the feature should be, based on the previously placed point.	75
39. Figure 39. SerialEM "Shift to Marker" location in the "Navigator" menu.	76
40. Figure 40. SerialEM "Acquire at Items" dialog window, depicting settings for grid square maps and hole finding.	78
41. Figure 41. Low magnification image of an EM grid depicting a grid square with a hole sufficient for dose calculation.	82
42. Figure 42. Location of the "Calibrate Electron Dose" menu option in the "Tasks" menu.	83
43. Figure 43. Camera parameters for the "Record" setting for data collection in SerialEM.	84
44. Figure 44. Recommended file-saving options for data collection.	86
45. Figure 45. SerialEM "Acquire at Items" dialog box showing the proper settings for final, high-throughput data collection.	88
46. Figure 46. SmartScope recommended screening configuration.	90
47. Figure 47. SmartScope grid and autoloader options.	91
48. Figure 48. SmartScope annotated grid atlas depicting collected grid squares in green, good grid squares in blue, and cracked grid squares in purple.	92
49. Figure 49. SmartScope annotated grid-square map depicting collected holes in green, and identified holes in blue.	93
50. Figure 50. SmartScope user interface overview, depicting annotated grid atlas, annotated grid square map, annotated medium magnification image, and associated high-magnification records and CTF estimations.	94
51. Figure 51. SmartScope grid restart protocol.	95
52. Figure 52. SmartScope settings to enable data collection mode.	95

LIST OF FIGURES CONTINUED

Figure	Page
53. Figure 53. SmartScope collection statistics displaying both CTF estimations and ice thickness estimations from CTFFind 5, total (blue) and over the last 100 micrographs (red).....	97
54. Figure 54. SerialEM "Navigator" dialog box depicting the protocol for creating a medium-magnification polygon montage.	99
55. Figure 55. A grid atlas depicting the boundaries for a desired polygon montage in green.	100
56. Figure 56. SerialEM "Montage Settings" dialog window depicting proper settings for a medium-magnification polygon montage for tomography.....	101
57. Figure 57. SerialEM "Tilt Series Setup" dialog window depicting recommended starting configuration for dose-symmetric tilt-series acquisition.	103
58. Figure 58. SerialEM "Camera Parameters" dialog window depicting recommended settings for high-magnification tilt-series acquisition.	105
59. Figure 59. SerialEM "Navigator" dialog window depicting items marked for tilt-series acquisition. Note the "TSP" options under "Acq. Note". This indicates that the positions will be recorded as tilt-series, with a special focus position defined.	106
60. Figure 60. SerialEM "Acquire at Items" dialog window depicting the "Final Data" recommended settings for tilt-series acquisition.	107
61. Figure 61. A medium-magnification image of the selected target for PACETomo acquisition (blue box) as well as the total exposed area (yellow oval). The exposed area is larger than the acquisition area due to the beam size and shape, as well as stage tilt.	109
62. Figure 62. PACETomo "Targets" dialog window.	110
63. Figure 63. SerialEM "Navigator" dialog window depicting proper PACETomo targets selection.....	111

LIST OF FIGURES CONTINUED

Figure	Page
64. Figure 64 A) Stereo image of representative electron density and model. B) Cartoon representation of DPSL model. C) Single four-helix-bundle of DPSL with labeled helices. Black arrow depicts connecting BC helix. D) 90° rotation around Y axis of four-helix bundle depicting newly modeled C-terminal extension in dashed box. E) Cutaway view of DPSL model depicting hollow core and C-terminal extension (circled red on a single subunit).	117
65. Figure 65 Stick representation of the Pf-Tftn bacterioferritin-like active site, and conserved cysteine pair. Solvent channels annotated in brown.	118
66. Figure 66 Iron mineral model and associated density from nucleated structure. A) View of mineral density with 3-fold axis running vertically along the screen. B) 90° rotation around Y.	122
67. Figure 67. A) Pf-Tftn dodecamer showing the location of the iron minerals at the four acidic pores. Iron atoms are rust-colored, and oxygen atoms are colored red. A single subunit is highlighted in light green. The interacting region (Leu47-Glu54) is shaded dark green. B) A view of the acidic pore looking from the exterior of the protein cage to the interior. Protein chains are shaded purple, pink, and green. Interacting side chains and carbonyl groups are displayed. C) Stereo view of iron mineral contained in the acidic pore, rotated 90° about X and displaying a single of the three protein chains. The mineral is positioned so that the exterior surface of the cage is at the top of the mineral, and the interior of the cage is at the bottom of the mineral.	123
68. Figure 68 A) Location of Trp154 relative to iron deposition site, shown on a single subunit. B) Trp154 density from apo Pf-Tftn, depicting alternate locations A and B. Global map resolution 1.86 Å. C) Trp154 density from nucleated Pf-Tftn. Global map resolution 1.91 Å. Note the more ordered form and hole in the indole ring. D) Trp154 density from mineral loaded Pf-Tftn. The tryptophan remains in just a single conformation. Global map resolution 2.43 Å.....	128
69. Figure 69 Tryptophan fluorescence emission overlaid with iron oxide absorption spectra, demonstrating overlap between emission and absorption that could result in an inner filter effect.	129

LIST OF FIGURES CONTINUED

Figure	Page
70. Figure 70 Fluorescence lifetime (FLT) analysis of Pf-Tftn with Fe ²⁺ . A) 50 Fe/Tftn in the presence of O ₂ (blue) vs H ₂ O ₂ (orange). $\Delta\tau$ is ~65% of the maximum, regardless of substrate. B) 500 Fe ²⁺ /Tftn in the presence of oxygen (blue) vs peroxide (orange). Maximum $\Delta\tau$ is achieved almost instantaneously in the presence of H ₂ O ₂ , while the additional Fe ²⁺ seems to have little effect on O ₂ $\Delta\tau$ when compared to the 50 Fe ²⁺ , suggesting that the reaction may be limited by available dissolved O ₂ at higher Fe ²⁺ concentrations. C) Effect of iron:protein stoichiometry on $\Delta\tau$, demonstrating that even at ~25 iron atoms per cage, FLT is mostly composed of the nucleated waveform. Again, dissolved O ₂ concentrations appear to be limiting for higher Fe ²⁺ stoichiometry.	129
71. Figure 71 Iron mineral core formation in Pf-Tftn. A) Pf-Tftn dodecamer demonstrating density of large iron mineral in core. B) View down acidic pore of large mineral Tftn. Glu50 and Glu54 have rotated away from the pore. C) Single four-helix bundle and associated unsharpened mineral density depicting C-terminal interaction with large iron mineral. D) Single four-helix bundle and associated large iron mineral density, half-map sharpened, demonstrating specificity of interaction. E) Rotated view of single chain residues 170-176. Lys171 is buried deep within the core, while residues 170-176 may be forming electrostatic interactions via main chain carbonyl groups. F) View down acidic pore in intermediate iron mineral map, demonstrating no density near Glu50, but Glu54 interactions with a smaller iron mineral are preserved.....	133
72. Figure 72. Electron density occupying active site channel near in holo map, near conserved cysteine pair. Density is modeled as water for convenient representation, although no experimental evidence suggests it must be water.	136
73. Figure 73. Apo Pf-Tftn single particle analysis workflow.	186
74. Figure 74. Representative micrographs from the apo (A), nucleated (B), and mineralized (C) datasets. All micrographs at 88,000x magnification, dose of ~55 e/Å ² , and recorded -1.4 μ m defocus. Scale bars 50 nm.	186
75. Figure 75. Nucleated Pf-Tftn single particle analysis workflow.	187
76. Figure 76. Mineralized Pf-Tftn single particle analysis workflow	188

NOMENCLATURE

EM – Electron Microscopy

Cryo-EM – Cryo Electron Microscopy

SPA – Single Particle Analysis/Averaging

STA – Sub-Tomogram Averaging

BIS – Beam-Image Shift

AFIS – Aberration-Free Image Shift

FOC – Ferroxidase Center

kDa – Kilo-Dalton

PAM – Pluggable authentication modules

AD – Active Directory

DPS – DNA protection in starved cell protein

TFTN – Thioferritin

DPSL – DPS-like protein

ABSTRACT

Iron is an enigmatic element. While necessary for life, it also contributes to the generation of reactive oxygen species via the Fenton reaction. To mitigate this, cellular life has evolved the ferritin family of proteins, including the 24 subunit ferritins and bacterioferritins, and the 12 subunit DPSL “mini-ferritins”. Each of these catalyze the controlled oxidation and sequestration of iron as a hydrous ferric oxyhydroxide within their hollow protein cores. While there is a wealth of structural information on the unmineralized ferritins, little is known about the structures of the biomineralized forms, and the mechanism of ferric oxyhydroxide nucleation and growth. Here we report structural and biochemical characterization of a DPS-Like protein from *Pyrococcus furiosus*. This “thioferritin” utilizes a bacterioferritin-like ferroxidase center, but adopts the mini-ferritin quaternary structure, and is thus thought to sit at the evolutionary boundary between mini- and maxi-ferritins. In addition to the unmineralized structure, we report the 1.91 Å structure of *P. furiosus* thioferritin as it nucleates iron-oxyhydroxide distal to the ferroxidase site. In this very low iron form, a pair of conserved glutamate residues and unsaturated carbonyls at the 3-fold axis serve to template initial nucleation. We also determine structures of higher iron forms with a biomineralized ferrihydrite core, where C-terminal residues 170-176 interact directly with the initial mineral surface, which then grows towards the particle center. These studies provide important new insight into biological mechanisms for the controlled nucleation, growth and storage of ferric oxyhydroxide in this thioferritin specifically, and the ferritin superfamily as a whole.

CHAPTER ONE

INTRODUCTION AND RESEARCH OBJECTIVES

Establishment of a Cryo-EM Facility for Structural Biology

The structure-function relationship underpins nearly all modern understanding of biology. Fundamentally, the structure of a protein determines its ability to bind or interact with other molecules and perform its necessary biological or mechanical functions. Therefore, protein and other biological macromolecular structures have long been of interest to researchers. Historically, there have been a variety of ways to determine macromolecular structure, the most common of which is x-ray crystallography (1, 2). X-ray crystallography relies on using the regularity in the crystal as a means of signal amplification of the structural information of the macromolecule (3). This typically requires a well-behaved protein or macromolecular complex that will form crystals with dimensions of micron-size or larger. However, certain samples are not amenable to forming crystals, because they are too large, too dynamic, or not sufficiently soluble at the concentrations needed for crystal growth (3, 4). While transmission electron microscopy (TEM) is not a new technology, finding its roots in the early 20th century, recent hardware and software advances in the field of cryogenic transmission electron microscopy (cryo-EM) have made high-resolution determination of many structures that were otherwise not amenable to traditional structure techniques, at near atomic resolution (4-7). This is in large part due to the ability to prepare samples for analysis while hydrated in near-native conditions, and amplify signal from the macromolecule *in silico*, rather than relying on the crystallization process (4, 6, 8).

The fundamental method of structure determination by cryo-EM relies on imaging of a biological macromolecule, of sufficient size so as to have visible features in a micrograph (40 kDa+), that has been deposited onto a small metal grid, coated with some kind of thin support, typically either carbon or gold, and plunge frozen into liquid ethane or propane, at temperatures lower than -154 °C. These low temperatures are required to generate vitreous, electronically transparent ice (3, 6). Areas of the frozen grid containing sample are then exposed to electrons accelerated to a range between 120 kV-300 kV (3). The sample scatters the electrons, and the lens system of the microscope then reforms the image, magnifies it, and records it on the detector. While TEM has been an observation technique for more than 100 years, microscopes have historically not been able to generate 3D structural information at meaningful resolution for a variety of reasons (5-7, 9). Most microscopes traditionally used thermionic sources – LaB₆ crystals or tungsten filaments, that do not form a beam with sufficient spatial coherence to preserve high resolution details present in the sample. However, recent advances in field emission gun (FEG) technology have largely solved this problem (10). Additionally, TEMs have historically recorded the final magnified image onto film, or later, onto CCD cameras that are too slow to scale to large datasets and use photomultiplier layers that are not of sufficient detector quantum efficiency (DQE) to preserve high-resolution details, primarily due to coincidence loss. The emergence of metal oxide direct electron detectors with high DQE and capable of operating at extremely high frame rates has largely solved this problem (11, 12). Finally, the ubiquity of high-performance computers and high density storage has enabled signal amplification *in silico* via data averaging techniques that were previously difficult to efficiently scale to the number of particles often needed to reach near-atomic resolutions (13-15). The two major techniques

enabled by cryo-EM are single particle analysis (SPA) and cryo-electron tomography (cryo-ET). These techniques have led the way to what is commonly referred to as a “resolution revolution”, with significantly more cryo-EM structures being published each year (16).

Major Cryo-EM Methodologies

Single Particle Averaging. Single particle analysis of suitable samples produces high resolution 3D maps of electron density in the sample by relying primarily on a combination of mass data collection, followed by 2D and 3D alignment and averaging techniques for signal amplification *in silico* (5, 7, 9, 17). In a typical SPA experiment, high concentrations of sample are frozen into vitreous, electronically transparent ice, in presumably random orientations. Thousands to tens of thousands of different images of the sample are collected, each containing hundreds to thousands of copies of the macromolecule, hopefully capturing 2D projection observations from all possible angles. The individual copies of the macromolecule are then extracted from the image, and typically referred to as particles. Particles are then sorted in 2D into groups of similar orientation, and the signal is averaged. Provided that the particles are grouped and aligned correctly, signal from the electrons scattering off the sample will add constructively, while non-system noise will add destructively. This allows the formation of high-resolution images of each class of particles representing a unique projection angle (4, 7). An initial 3D map is then able to be constructed at low resolution, and particles are aligned to the 3D map by individually adjusting their rotation and translation relative to the initial low-resolution map, the sum of which is often referred to as a particle pose. The goal is ultimately to align the particles with as great of a spatial resolution as possible, to maximize constructive signal addition when averaging the signal across all particles. Once the pose is initially fit for each

particle, the particles can be re-averaged to generate a new, often higher-resolution map (7, 8, 14). As the signal increases, if the alignments of the 2D particles are of sufficient accuracy, the resolution of the map typically increases, allowing more precise particle poses to be calculated, and further improvement to the map signal (14, 15). This process, known as 3D refinement, is typically iterative and will proceed until further resolution gains are attenuated. Ultimately, each individual extracted particle will be assigned the best possible pose that describes the angle of the extracted particle image relative to the final 3D map. Due to the fact that, even in ideal conditions, signal is proportional to the square root of the number of observations, this process can take hundreds-of-thousands to millions of particle images to produce a map of 3 Å resolution or greater, typically with diminishing returns as more particles are added to the reconstruction (18).

Even with an amenable sample, other challenges remain in SPA. Electron dose is extremely destructive to biological samples, with dose greater than $\sim 20 \text{ e}^-/\text{\AA}^2$ beginning to destroy high spatial frequency features in the sample, and dose greater than $\sim 80 \text{ e}^-/\text{\AA}^2$ to $100 \text{ e}^-/\text{\AA}^2$ causing visible sublimation and destruction of the sample (19). Unfortunately, due to relatively weak interactions between high energy electrons and biological matter, a significant dose, typically greater than $40 \text{ e}^-/\text{\AA}^2$ is needed to generate sufficient contrast in the images for them to be interpretable (20). One workaround for this problem is to first defocus the image to introduce phase contrast. While phase contrast can often provide enough contrast to pick particles, it introduces delocalization of the signal in the final image, with high frequency details being delocalized furthest. This delocalization, termed the point spread function (PSF) when describing its effects in real-space, or contrast transfer function (CTF) when describing its effects

in Fourier space, produces oscillations in the sign of the intensity of the signal as a function of spatial frequency. If left untreated, these oscillations will cause destructive interference at specific spatial frequencies as a function of defocus, which will result in signal canceling when particles are averaged. However, this oscillation can be modeled and corrected *in silico* provided sufficient signal is present in the micrograph (21). To perform this correction, specifically at high spatial frequencies that are necessary to generate high resolution reconstructions, the precise defocus of each particle must be known. Due to imperfect microscope focusing, uneven grid topology, or even sample positioning in the ice, defocus can vary significantly, and must be independently estimated by fitting the oscillations of the CTF to calculate the defocus (21). Typically defocus estimates are first calculated coarsely, either for a single micrograph, or section of a micrograph. These estimates are used to apply a CTF correction to the extracted particles, at least to a resolution capable of creating a consensus map. Once a map of sufficient resolution is generated, defocus estimates can be refined on a per-particle basis (14, 15). The nature of the CTF oscillations also results in nodes where signal in the micrograph is fully cancelled, rather than just inverted, at specific spatial frequencies. The precise spatial frequency of these nodes is also a function of the defocus, so for this reason, datasets will typically be collected over a wide range of defoci, to prevent any one spatial frequency from being totally devoid of information in the final averaged structure.

Another problem with cryo-EM data, and also a result of the destructive nature of the electron source, is beam-induced-motion (BIM). BIM arises when the vitrified sample is exposed to the electron beam and begins to shift in the ice (22-24). BIM is typically most significant at the beginning of an exposure, where presumably the particles are still in a relatively high energy

state in the ice, due to the rapid nature of the plunge freezing process. However, this is exacerbated by the fact that at a dose rate of $15 \text{ e}^-/\text{s}$, the first 1-2 seconds also contain nearly all the high-resolution information in the micrograph (19). To resolve this conundrum, rather than write out a single micrograph that is the sum of the entire $\sim 50+ \text{ e}^-/\text{\AA}^2$ exposure, a shorter micrograph termed a “frame” can instead be written out every $\sim 0.8 \text{ e}^-/\text{pixel/s}$, provided that the detector is capable of recording at that framerate (22-24). The totality of frames from a single exposure is referred to as a “movie”, and a typical movie will be somewhere between 30 and 60 frames. All the frames in a movie can then be aligned with respect to each other, to track the beam induced motion and to create a single frame, motion-corrected micrograph. Like with the CTF correction, those frame alignments are typically done in patches, as a local solution is often more accurate than a global one (15). Once a 3D map of sufficient quality is achieved, motion can sometimes be corrected on a per-particle basis resulting in a resolution improvement, albeit at significant computational cost (24).

Once poses relative to the consensus 3D map have been calculated to high resolution, it is often possible to classify the individual particles based on their poses relative to each other – a process referred to as 3D variability or 3D classification (15, 25). These techniques can capture both composition and conformational differences that may be present between particles, but are averaged in the reconstruction, and therefore not visible in the structure. These techniques are often used to demonstrate in 3D the variety of conformational states available to a macromolecule (26, 27). These tools are unparalleled in providing insights to macromolecular function on a structural level.

While the versatility and near-native conditions of cryo-EM are noteworthy, not all samples are amenable to SPA techniques. This can be for a variety of reasons: the sample may be non-uniform, rendering it impossible to average. One example of this could be a cell, or enveloped virus, that is not always of consistent dimensions or internal arrangement. Furthermore, some samples adopt a preferred orientation when frozen in the ice, meaning that all projections will be identical or near identical, resulting in anisotropy in the final 3D reconstruction. If this anisotropy is severe enough, it can limit the resolution and ultimately the useful structural information available in the final map. Finally, it is often desirable to view biological macromolecules in their true native state, such as a ribosome inside of a cell. In this case, 3D information of a single particle, not a population average, is desired, and so single particle averaging techniques are not optimal.

Cryo-Electron Tomography. In some cases that are not amenable to SPA, cryo-electron tomography can still produce intermediate-to-high-resolution structural information. Rather than average together many observations of many different copies of the same sample in all different possible angles to generate 3D information, cryo-ET instead takes many observations of the same sample, but the sample is tilted between observations, resulting in discrete images of the sample from a wide range of angles, typically $\pm 60^\circ$ (28, 29). These images, collectively, are called a tilt series. Once recorded, the tilt series images are then aligned, typically to electron dense fiducial markers embedded in the sample, such as 5 nm – 10 nm gold beads. This alignment accounts for microscope stage drift between different images in the tilt series, which is frequently on the order of at least several nanometers between images in the tilt series. Once the tilt series is aligned, the images are then reprojected in 3D to create a volume of the observed

area, called a tomogram (30). This tomogram describes in 3D the sample that was exposed to the beam, with density in the tomogram representing electron density in the sample. However, due to some technical limitations to cryo-ET, the achievable resolution of a single tomogram is usually in the low-nanometer range.

The single greatest limitation is the cumulative electron dose the sample is exposed to. As previously described, electron exposure is destructive to biological samples. For SPA, this issue isn't as limiting, because after a sample is imaged once, the microscope stage moves to a fresh area of sample that has not been previously exposed to electrons, and the signal is simply added later *in silico*. However, with cryo-ET, all of the final signal must come from the same observation, and each feature imaged 25 or more times per tilt series. This means the maximum dose must be divided by the number of tilt images, resulting in an average dose-per-image of $\sim 1-3 \text{ e}^-/\text{\AA}^2$, and a final sample exposure often on the order of $80+ \text{ e}^-/\text{\AA}^2$. The end result of this low dose per image is that contrast is extremely weak, and features are not very observable in any given tilt. Additionally, the high total dose on the sample means that, while early images still preserve high-resolution details, later images in the sequence are typically devoid of any high-resolution information. The situation is exacerbated by the fact that the low dose typically requires samples to be imaged at greater defocus to provide sufficient contrast in any image. The high defocus in each image causes greater delocalization of signal, and the low dose makes modeling the CTF and correcting this more difficult, due to weaker signal in reciprocal space. Additionally, the tilt of the image introduces further defocus estimation problems, as, except for 0° tilt image, there will be a defocus gradient across the image. Finally, tilt angle is often not known precisely, as there may be pre-tilt in the stage and/or sample, so attempts to model the

defocus as a gradient may not be precise unless the tilt angle of each image can be accurately estimated (31).

There are other limitations of cryo-ET as well. As the sample tilts to a high angle, the thickness of ice that the electron must transmit through increases, resulting in more electrons being scattered at high angles and never reaching the detector, manifesting as even lower contrast at high tilt angles. Finally, due to the limited range of motion of microscope stage tilt (typically $\pm 60^\circ$ to $\pm 70^\circ$), the sample is not actually imaged from all possible sides. This results in a problem termed “the missing wedge”, describing a wedge of 3D Fourier space that is devoid of information in the final tomogram. This is similar to the preferred orientation problem in SPA, although often not as dramatic, and can result in 3D density appearing disproportionate or stretched along a specific axis (32). As a result, the resolution of any unprocessed tomogram is typically on the order of low nanometer, and nowhere near atomic resolution.

To mitigate some of these issues, several software techniques have been developed to extract repeating features that might exist inside of the 3D volume, and align and average them, increasing the signal, and recently allowing for near-atomic resolution of those features in some cases. This method is called “sub-tomogram-averaging” (STA), and while it is similar to SPA, due to higher levels of noise inherent to tomography, as well as the fact that the data are already in 3D, it introduces its own unique challenges (31, 33, 34). Still, despite these limitations, recent developments in cryo-ET have seen some larger, more electron dense particles approaching near-atomic resolution when averaged (35, 36).

Supporting Infrastructure

While the microscope itself must be of necessary quality to be capable of generating high resolution data, the resolution dependence of averaging techniques is ultimately proportional to the square root of the number of observations. Practically, this means that reaching high-resolution can require millions of direct observations, in the form of thousands of images. The quantity of data that is generated by the microscope makes the supporting computational infrastructure of near equivalent importance to that of the microscope hardware itself. Additionally, the interaction of several proprietary systems involved in generating the data means that the data flow and processing pipelines must be carefully considered at the conceptual stage of facility construction to ensure that the computational aspect does not bottleneck sample throughput.

Detailed in this thesis are the operational considerations of setting up a cryo-EM facility, as well as details specific to the implementation of the current cryo-EM facility at Montana State University. Additionally, the two major techniques (SPA and cryo-ET) are discussed in detail. Standard operating procedures (SOPs) and methods are given for each major technique, including microscope operation, data collection, data processing, and interpretation. Specifically, SPA is exemplified in the form of a project analyzing biomineral nucleation and growth in *Pyrococcus furiosus* DPS-like protein. Additionally, generic cryo-ET data collection methods are supplied.

Single Particle Analysis of Biomineralization in *Pyrococcus*
furiosus Dps-Like Protein

Iron and oxygen are both essential elements for life, yet together can form a lethal combination. The reaction of Fe(II) and molecular oxygen creates a dangerous hydroxyl radical (OH•), capable of causing significant damage to important cellular systems, such as Fe-S clusters (37, 38). As a result, cells have developed biochemical processes to mitigate this danger. Chief among them is the active sequestration of free Fe(II), which is also often coupled with the controlled oxidation of Fe(II) to the more chemically inert Fe(III). Proteins for such iron oxidation and sequestration represented in all domains of life by the ferritin family (39-44). The ferritin family consists of two major classes, the maxi-ferritins and mini-ferritins. The maxi-ferritins, which include the eponymously named ferritin, and bacterioferritin, are comprised of a core four-helix bundle that assembles into an octahedral 24-mer with a hollow core. In contrast, the mini-ferritins, comprised primarily of the “DNA protection in starved cells” (Dps) proteins utilize a similar four-helix bundle, but instead assemble into a dodecamer with tetrahedral quaternary arrangement, and exhibit an exterior diameter of 7-9 nm, with a smaller hollow core. All of these proteins operate in a similar manner – they both typically incorporate one copy of a ferroxidase center (FOC) per four-helix bundle that can use the high reduction potential of Fe(II) to transfer electrons to a powerful oxidant, such as O₂ or H₂O₂, while simultaneously oxidizing two atoms the dangerous Fe(II) to the more inert Fe(III) form. However, further differences between the maxi- and mini-ferritins abound when examining the arrangement of the FOC. In maxi-ferritins, the FOC is located in the center of the four-helix bundle, while in Dps, the FOC is present at the 2-fold interface between four-helix bundles. This FOC location as well as protein quaternary arrangement were thought to clearly define the boundary between maxi- and mini-

ferritins until the discovery of Dps-like (DPSL) protein in *Saccharolobus solfataricus* (45). While initially predicted from sequence to be a mini-ferritin, structural studies uncovered that, while similar to Dps in its dodecameric tetrahedral quaternary arrangement, DPSL's FOC is located in the center of its four-helix bundle, as is exhibited in maxi-ferritins (46). This places DPSL at the interesting evolutionary boundary between the maxi- and mini-ferritins, and it is easy to imagine how either class could have evolved from the other by transitioning through a DPSL intermediate state (46, 47).

While the ferritin family proteins have had their major structural features and ferroxidase activity thoroughly characterized, what is less well understood is the ultimate fate of the freshly generated Fe(III). While all the ferritin-family members have been shown to mineralize the Fe(III) product in their hollow cores, specifically as ferrihydrite in both ferritin and Dps, the precise structural mechanisms that catalyze nucleation and growth of this biomineral remain unclear (48-50). While generation of mineral-loaded proteins is possible, there are several complications with existing analysis techniques. Historically, most structural studies of ferritin have been performed with x-ray crystallography, this technique has not been amenable to structures of full particles, both because of the necessity for high concentrations of loaded particles, as well as heterogeneity in the loading particle that would not be visible in a population average. Additionally, while some cryo-EM studies of loaded ferritins have been attempted, these studies have encountered issues with the intense electron density and associated amplitude contrast contributed by the mineral core (44, 51). Existing alignment algorithms were not written while considering such intense amplitude contrast, and as such the mineral core tends to dominate alignments at low resolution, preventing proper alignment of the higher resolution

details of the surrounding protein. This results in calculated particle poses that remain trapped in low-resolution minima, resulting in an overall low-resolution structure, providing minimal biological insight.

Research Objectives

Cryo-EM is a nascent field of structural biology, and while accessibility is increasing, the high costs and complex nature of microscope operation and data analysis requires an established knowledge base to make efficient use of the technique. Additionally, most existing techniques and methodologies are not as mature as those in other structural fields. A successful project will need to integrate complex theoretical and practical knowledge applied in a dynamic manner towards the biological questions being asked. While the major projects presented in this thesis are not related biologically, they are united by the common theme of being initial applications of the two major cryo-EM methodologies at the nascent Montana State University Cryo-EM Facility. The research presented here attempts to lay the groundwork for broad application of both major cryo-EM methodologies, as well provide a starting point for a sustainable cryo-EM facility that can carry forward and grow the cryo-EM institutional knowledge at this university.

Research Questions

The 1st research question addresses how to reliably collect high-quality EM data for biological analysis. Considerations include how to ensure the quality of the data are sufficient for high resolution analysis, how to store and efficiently analyze the huge quantities of data, and how to ensure all necessary support systems are properly configured and accessible.

The 2nd research question: How do ferritin-family proteins nucleate and grow iron-oxide minerals in their core? To answer this question, we employed advanced single particle analysis techniques that provided high resolution structural insight into a Dps-like protein from *P. furiosus*, in its apo form, mineral-nucleated form, and fully loaded form. These structural details elucidate the precise protein-mineral interactions that enable the nucleation and growth of the iron-oxide mineral core in Dps-like proteins, demonstrating that nucleation occurs via precise interaction with main-chain carbonyl moieties and glutamate side chains, resulting in the growth of a ferrihydrite-like mineral in the core.

CHAPTER TWO

CRYO-EM CORE FACILITY

Operational Infrastructure

In March of 2020, Montana State University took delivery of a Talos Arctica G2 200 kV transmission electron microscope and Gatan K3 direct electron detector camera. The 12-grid autoloader and advanced computer scripting interface of the microscope, combined with the extremely high frame rate of the detector, results in a system that is capable of rapidly obtaining near-atomic resolution 3D structures of biological macromolecules and macromolecular complexes. Additionally, the precision of stage movement and sample holder with α -tilt capabilities allow the microscope to acquire both tomographic and micro electron diffraction (microED) data. However, due to the highly sensitive nature of the electromagnetic lenses present in advanced high resolution TEM systems, a number of environmental controls in the microscopy facility are required to attain the highest possible resolution images. Additionally, the pairing of the Talos Arctica microscope, made by Thermo Scientific, and K3 camera, made by Gatan, prevents the usage of Thermo Scientific's 'EPU' data collection software, and thus necessitates a 3rd party microscope operation software to be installed and configured. Finally, the Gatan K3 detector is 96 megapixels, theoretically capable of recording 1500 frames-per-second (FPS), with an 8-bit pixel depth, resulting in a max theoretical output of 144 gigabytes (GB) per second of data generation. While practical usage results in an attenuated rate of data collection, the typical ~500 GB/hour of data generated still presents significant burden to conventional computational resources. Care must be given to the computational infrastructure that must be in

place to both transfer and store the large quantities of data generated by the microscope, as a bottleneck at any point can severely reduce the cost efficiency of the entire microscope system.

Environmental Controls

The electromagnetic lenses present in high-resolution TEM systems are susceptible to a number of interferences, such as electromagnetic, acoustic, and vibrational interference.

Additionally, the conductivity of TEM lenses can change as a function of temperature.

Electromagnetic lenses in TEMs are typically made of carefully arranged electromagnetic coils that exert a highly tuned and calibrated magnetic field on passing electrons. These electromagnetic coils are kept at constant temperatures by a closed-loop chilled liquid cooling system. However, this cooling system may not accommodate large temperature increases in small amounts of time, which can allow lenses to heat up, leading to increased resistivity of the electromagnetic coils, thus altering the electromagnetic fields emitted by the lenses. This will frequently manifest in a misaligned or unstable electron beam. To mitigate this, the maximum temperature change in the microscope room over any 24-hour period should be ± 1 °F.

Additionally, chiller flow should be routinely checked, and the loop filter should be replaced monthly.

In addition to temperature, acoustic, vibrational and electromagnetic interference may result in degradation of data quality if left unchecked. Acoustic/vibrational interference typically manifests in the form of beam drift or jitter, and is clearly apparent in the Fourier transforms of the micrographs. The solution is simple – remove any loud or vibrating objects from the microscope room. For this reason, the microscope chiller, which runs a loud compressor, has been placed in an adjacent mechanical room. Electromagnetic interference can be more difficult

to assess and mitigate. The microscope room has been equipped with a Spicer SC24 Field Cancellation System, which detects electromagnetic interference across the microscope column, and applies an offset to the room to cancel out any detected interference. This system should be left operational when the microscope is collecting data, to achieve the greatest possible data quality.

Microscope Control Software

While the Talos Arctica microscope is manufactured by Thermo Fisher Scientific Inc., the primary K3 camera module is manufactured by Gatan, Inc. As a result, the default single particle data collection software installed on the microscope, “EPU” is unable to properly communicate with the K3 camera. However, there are several free software packages capable of being installed on the microscope or camera PC that can communicate with both, and effectively drive the system for high throughput data collection for both single particle and tomography. Packages available include SerialEM and SmartScope (52, 53).

SerialEM. SerialEM is free, open-source microscope control software that supports nearly all existing TEM and camera combinations (52). The main SerialEM user interface is installed on the K3 camera computer, but SerialEM also requires a server running on the microscope computer that interfaces with the advanced scripting package for the microscope. It is capable of fully driving the Arctica and the K3, with features such as loading and unloading grids, managing vacuum, changing apertures, controlling the electron beam, and performing many different alignments and calibrations. With SerialEM users can create grid maps, select grid squares for data collections, find holes in grid squares, collect and view both micrographs and tomograms, and perform basic post-processing such as frame alignment and CTF fitting to

estimate data quality. Additionally, many of these functions can be automated to varying degrees to allow for bulk data collection of both micrographs and tomograms.

Because of the many calibrations, settings, and configuration options available to SerialEM, many of which are specific to specific samples or data collection techniques, managing the SerialEM configuration can become quite time intensive. To mitigate this, users can save their own settings in separate files, and load them as needed. A graphical tool to easily manage the loading and saving of these configuration files, called Zuul, has been installed. SerialEM is highly extensible, supporting both its own internal scripting language, as well as a Python library that translates most SerialEM commands into Python functions. This Python library can be called internally by running Python code directly in SerialEM script files, or, additionally, it can be run externally by having SerialEM listen for Python commands on port 48000. When placed in external mode, direct user input to SerialEM is disabled (54).

SmartScope. SmartScope is a single-particle grid screening and data collection tool with a heavy emphasis on automation and ease of use¹. It is written primarily in Python, using Django and is installed on an auxiliary computer that must be capable of connecting to K3 computer. It runs as an external extension to SerialEM, and uses the SerialEM Python library for interfacing with the microscope and camera. The user interface can be accessed by any web browser on the same network, on port 48000. SmartScope is capable of screening grids for single particle data collection, as well as setting up data collection in a highly autonomous manner. It relies on a combination of simple algorithms and advanced computer vision for this automation.

To use SmartScope, users can tell SmartScope how many grids are loaded, what position they are in, and how many grid squares and holes-per-square to collect. SmartScope uses the

SerialEM Python library to queue the grids for screening and to collect grid maps. Once a grid map is obtained, integration with the Ptolemy U-net image recognition pipeline allows it to detect and classify grid squares as good, cracked, small, or out of stage movement range (55). Once grid squares have been identified and selected with Ptolemy, SmartScope will automatically create grid square maps, and in the process use traditional cross-correlation to re-center the grid squares, attenuating issues with large stage movements that typically manifest as a gradient of inaccurate stage movement from the center of the grid to the edges, which can lead to grid square maps that clip the edge of the camera field-of-view (FoV).

Once a grid square map has been created, SmartScope runs its own eucentric height routine that is faster than the default eucentric height routine in SerialEM, identifies and categorizes holes on the square based on the intensity of each hole, which is correlated to ice thickness. Then it organizes the holes into optimal groupings for beam-image-shift (BIS) collection and collects them as instructed. To collect data, SmartScope first takes a medium magnification image at the center of each BIS group, and then uses either traditional cross-correlation with a hole template image or identifies the multishot pattern with Ptolemy to align the microscope to the center of that hole for collection of the BIS group. Using Ptolemy for hole centering frequently results in better hole centering for challenging samples that prefer thicker ice, or add so much electron density to the images, such as ribosomes, that it disrupts cross-correlation/template matching algorithms.

In addition to collecting data, SmartScope also asynchronously aligns micrographs, calculates CTF fits, and presents data for display in the web UI. A user can select a screening session and see all the grids in that session. They can select a grid, and see the grid map, with all

the identified squares. They can then select a square and see all the holes. The users can instruct SmartScope to color-code the holes by intensity, BIS group, or if the holes have been collected and processed, they can also be color-coded by CTF fit as a rough indication of resolution and data quality.

SmartScope also makes it very easy to make modifications to data collection in real-time. Users can manually add grid squares or even individual holes to a data collection. Additionally, they may alter data collection parameters such as BIS distance and stage tilt on-the-fly, while data collection is underway. Users can select from pre-set data collection protocols, or write their own, for granular control of every step of the data collection process. SmartScope is also extensible. Users can install their own image recognition models, as well as processing pipelines, and integrate them natively into SmartScope.

SmartScope is capable of extremely fast grid screening, as well as data collection. It can comprehensively screen 12 grids in less than a single work day, and can reach speeds of up to 500+ images/hour when in data collection mode. In addition to this, the automation and computer vision mean that very little staff time is necessary, when compared to traditional data collection methods such as SerialEM.

Computational Infrastructure

Due to user demand and the expensive nature of high resolution cryo-capable TEMs, both in up-front cost and service contracts, optimizing a microscopy facility to maximize throughput is essential for the long-term health and sustainability of the facility. While great effort has gone into both mechanical automation and reliability of the microscope, as well as data collection protocols and scripts, similar thought and effort must be given to the underlying computational

infrastructure, or data collection will be bottlenecked by the inability to move, store, process, and maintain the resultant data.

At the forefront of this problem is network connectivity. There are a variety of different computers involved in the microscopy facility, many of which run pieces of equipment. Some of these computers must be kept obfuscated to external networks for security considerations, or off the network entirely. Others must be connected to external networks in an efficient manner to facilitate maximum data transfer speeds, or real-time remote access. In addition to network construction, high framerate direct electron detectors such as the Gatan K3 generate massive amounts of data. A single, 12,000 micrograph, 24 hour data collection, even with state of the art image compression, will generate approximately 5 TB of initial data. Subsequent processing will often grow this dataset to the size of 15+ TB. In addition to just the sheer volume of data, it often needs to be accessible to high performance compute nodes in near real time, to allow pre-processing to inform data collection about the quality of data coming from the microscope. After data collection, data still need to be accessible to high performance computers, as full processing can take months or even years, with certain datasets. Furthermore, because federal granting agencies require data backup and archival for many projects, data needs to be available to be backed up. However, traditional methods of backup such as external USB thumb drives or hard drives may not be efficient, due to the large size of cryo-EM datasets, and slow USB transfer speeds. Therefore, a faster, more robust backup method geared toward larger datasets is needed.

All of these points have been considered in the construction of the computational infrastructure for the cryo-EM facility at MSU. Data may be written and moved to scalable storage at high speeds, where it is also made available for processing by central, scalable

compute resources, and easy transfer to external backup solutions. Finally, all of this is done in a secure manner that preserves user access controls and permissions.

Networking & Connectivity

Network connectivity is an important consideration for both the performance and security of a cryo-EM facility. The crux of the problem is that certain computers and instruments must be able to communicate with each other over the network, while also moving data on and off the network, all in a secure manner that protects user permissions and minimizes possible attack vectors, but without sacrificing performance. At the core of the microscope facility network is the microscope network itself. This is an external network, not connected directly to the MSU network. It connects the microscope PC to the microscope, as well as all the individual microscope parts that need to be controlled by the microscope PC. This includes items such as apertures and aperture controls, the stage, lens control boards, liquid nitrogen monitoring controllers, the flu screen, and the integrated Ceta-M camera and camera controller.

In addition to the microscope and microscope network, the facility also houses the Gatan K3 and associated computer (megaera), a control and documentation server (hydra), three network-connected temperature sensors, and a remote PC that enables Thermo Fisher engineers to remotely access the microscope PC. Of these, the remote PC, hydra, and megaera must all be able to interface directly with the microscope PC. Additionally, the remote pc, as well as hydra, must be accessible from the MSU network, while megaera and the temperature monitors must not have a direct connection to the MSU network, but instead proxy through hydra, which serves as the major access point to the facility from both on and off campus. Finally, megaera must be able to transfer data collected with the K3 through hydra to the campus centralized scalable

storage (blackmore), which must then be accessible to the campus centralized compute (tempest) as well as off-campus backup solutions.

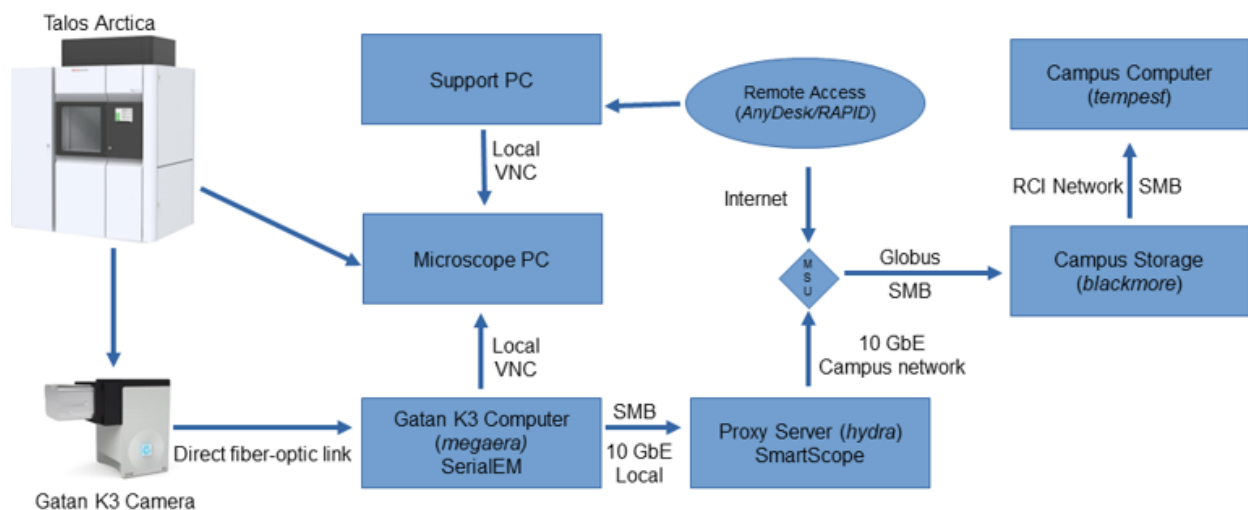


Figure 1. Network diagram describing data-flow in the MSU Cryo-EM Facility. Data are recorded by the K3 camera and transferred by fiber optics to the camera controller (Megaera). Once there, data must pass through a proxy (hydra) to be accessible on the MSU network and via Globus. Globus can automatically move data to central storage such as Blackmore, or any other storage system, where it can then be operated on.

Active directory (AD) on facility systems is primarily using the System Security Services Daemon (SSSD) and Pluggable Authentication Modules (PAM) management. The system is somewhat complex but an attempt to summarize it will be made. The University uses active directory for managing users. There are several ways to interface with active directory, including both SMB/winbind and LDAP. Active directory is used for mapping Unix user IDs to MSU NetIDs. Additionally, the university is using Kerberos for authentication of AD users. Kerberos is in addition to AD. Kerberos functions as “ticketing” software, so that the user doesn't need to actually provide NetID and password to a service or system (such as logging onto a computer). Instead, the AD system checks your NetID against a cache of directory accounts, and then checks

to see if it has any tickets on file for your account. If it does, authorization is granted. If not, it requests a new ticket through Kerberos. The purpose of Kerberos is to abstract authentication to a more secure entity than the individual service that may be using and collecting credentials. Rather than re-implement an AD/Kerberos system for every single service on a computer, Unix systems such as Linux and BSD have the eponymously named software, PAM (Pluggable Authentication Modules). PAM modules are software built on an existing backend compatible with an AD/Kerberos solution. These modules essentially bridge the user-facing programs with a secure AD authentication system, meaning that any user can use any software with their AD credentials. On Linux systems, System Security Services Daemon (SSSD) is the software that is responsible for centralized management of all these services.

When logging into a system running AD, either SSHD (the SSH daemon) or the graphical window manager listens for requests to log in. If a request is received, the login manager forwards that request to the PAM system. PAM looks up the AD backend, in this case, SSSD. SSSD handles authentication by looking up the AD user in the AD system it is registered with, and requesting a Kerberos ticket for them. SSSD forwards that ticket and any domain-wide or system-wide settings to PAM. These settings include things like where to copy the home folder from if none exists, what shell to default to, and other settings. PAM then checks for "optional" authentication modules, which, in the case of the stereo workstations, are SMB mounts of Blackmore storage using the same Kerberos ticket. PAM uses Kerberos ticket and a module called "pam_mount" to log into and mount network drives on those systems as the user is logging into the workstations. PAM then forwards the authorization status to sshd, which approves the login and places the user in their home folder with their network drives mounted.

openSUSE Linux ships with management software called YaST. This details the process for enrolling a computer to the domain using YaST. First, install the package “yast2-auth-client”. Run YaST as an administrator, and go to “Network Services --> User Logon”. Select “Change Settings”. Check off “Allow Domain User Logon” and “Create Home Directory” and then at the bottom, select “Add Domain”. Domain name should be set to Domain name: msu.montana.edu. Which service provides identity data should be set to “Microsoft Active Directory”. Which service handles user auth should be “generic Kerberos service”. Ensure that “Enable domain” is checked and click “OK.”. A new dialog box will pop up – enter the Kerberos realm as “msu.montana.edu” and the AD hostname as the hostname of the computer being enrolled, ending in “.msu.montana.edu”. The AD server hostname should be “obsidian.msu.montana.edu”, and the credential cache should be set to “iolite.msu.montana.edu”. At this point, additional packages may be downloaded.

A dialog box will appear, called “Active Directory Enrollment”. For the username, enter a NetID ending in @msu.montana.edu that is authorized to enroll new computers on the domain. Enter the user’s password and disable “Update AD’s DNS records as well”. If the enrollment is successful, confirmation will appear.

Finally, once enrolled, attempted login will fail as no default shell is configured. To configure a default shell, navigate to “Extended Options” for the domain. Search through extended options for “default_shell” and set “default_shell” to “/bin/bash”. SSSD may need to be restarted before changes go into effect – an admin user can do this with the “sudo systemctl restart sssd” command. The SSSD configuration file can be found in “/etc/sss/sss.conf” (Appendix A).

Data Storage & Management

Data collected using SerialEM will be located in the “movies” folder on the “X:\” drive on the Gatan computer, selected by the user. When using SmartScope, data will be saved to the “X:\SmartScope\movies\<session_name>\<grid_name>” directory. It should be noted that data saved here are not secure. Firstly, this data is not redundant. If a disk fails, it will be lost forever. Additionally, this is high-speed storage needed by the camera, and so the next user of the microscope may need space and data may be removed to make space, without warning. Therefore it is advisable to remove data from the “X:\” drive as quickly as possible. Data can be easily moved off the “X:\” drive in near-real time using the Globus software.

Log in to Globus using an institutional login, such as a .edu email account or MSU NetID. The facility staff can create a shared folder for users on the “X:\” drive using their MSU NetID so that they can see their data in Globus’s web user interface. Data can be moved to an arbitrary computer by installing “Globus Connect Personal” on that computer. Run Globus Connect Personal on that account, and login. Then, in a web browser, navigate to <https://www.globus.org> and login. In the menu bar on the left, select "File Manager". "Shared With You" should become visible. On this screen, the user should see your shared directory on the microscope computer.

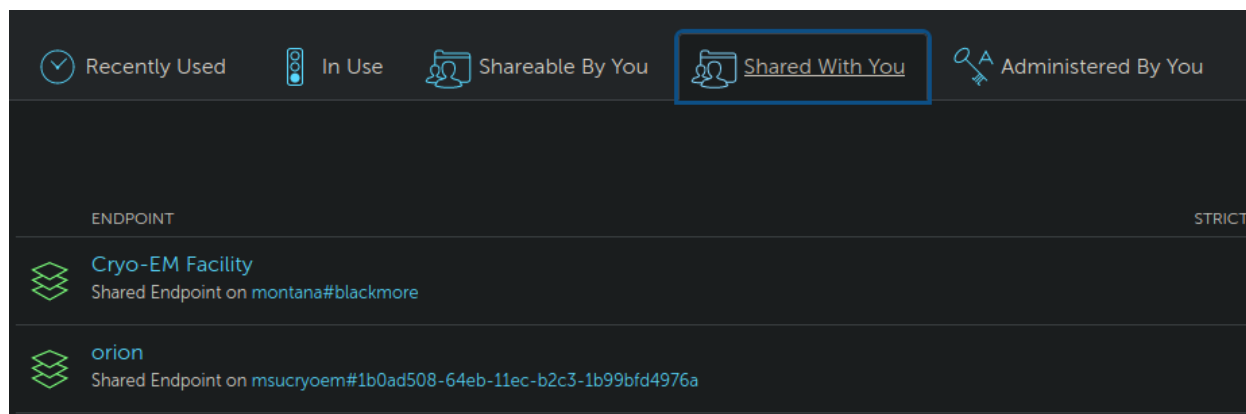


Figure 2. An example of shared directories in the Globus web user interface. The Cryo-EM Facility shares are managed through the Globus UID of msucryoem and can be shared with any MSU NetID.

Select that share, and then navigate to the file manager. In the file manager, select the files that are to be moved off of the microscope computer. Entire directories can also be selected. On the right hand side, select “Search” and navigate to “My Collections”. Select the computer that Globus Connect Personal is installed on and press the “Start” button. This will initiate the transfer, and the web browser can be safely closed. When moving data to Blackmore, Globus Connect Personal is not needed, as Blackmore has a dedicated Globus data-transfer node.

To move data in real-time, use the Globus Timers feature. This is useful when trying to use live processing tools such as cryoSPARC Live, or when the data transfer will take quite some time. To use Globus Timers, simply setup everything as listed above, except do not click start. Instead, select "Transfer and Timer Options" next to “Start”.

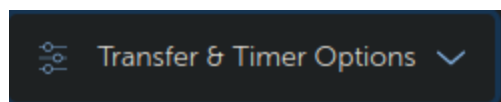
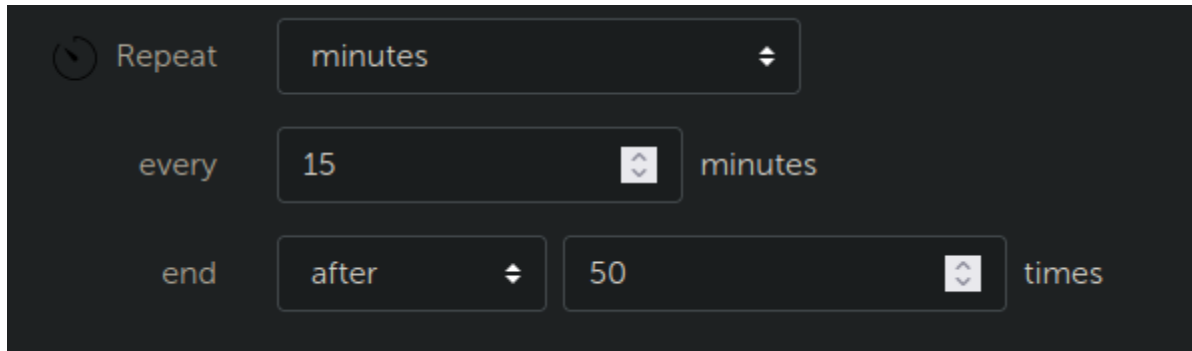


Figure 3. Globus real-time synchronization options button.

There are a number of options. Select the “Sync” option, and switch from “Checksums” to “File Size is Different” as calculating a checksum for every file will slow down the transfer considerably. Additionally, to prevent the transfer from hanging, select “Skip files on source with errors”. Finally, at the bottom, select the timing that is desired. For single particle, to an on-campus connection, every 15 minutes works well. Also select an end date, or total number of transfers. Start the transfer with the normal “Start” button, and a recurring transfer should begin.



The image shows a dark-themed user interface for configuring a recurring transfer. It features three rows of controls, each with a label on the left and a selection area on the right. The first row is labeled 'Repeat' and has a dropdown menu set to 'minutes'. The second row is labeled 'every' and has a numeric input field containing '15' followed by a 'minutes' label. The third row is labeled 'end' and has a dropdown menu set to 'after' followed by a numeric input field containing '50' and a 'times' label. Each input field and dropdown menu has a small double-headed arrow icon for adjustment.

Figure 4. Globus timer options.

To automatically mount network drives upon user login, any login service, such as a GUI login screen such as lightdm or SDDM, or in the case of terminal login, such as SSH, must be configured to use the “pam_mount” module upon user login. In addition, “pam_mount” itself needs specific configuration so that it knows the address of the share, and where to automatically mount it. PAM is configured using the software “pam-config”. There are two types of PAM modules, global modules which are inserted into “etc/pam.d/common--*” files which are then included by single service files. A single service file might be a user-facing service such as SSHD or SDDM or some other login service. Direct changes to those files will be overwritten by software updates, or by running the “pam-config” software, therefore it is recommended to make

configuration changes through the software, and not by hand. There are also single-service modules, which are options that can only be added directly to a single service. Rather than making a change globally, they must be made for whichever individual service is using PAM. This includes the “--mount” option, which is used to mount SMB shares. The configuration file for “pam_mount” is found in “/etc/security/pam_mount.config.xml” (Appendix A).

Data Backups

Montana State University is currently approved for usage of the “OU & Regional Research Store” (OURRstore) large scale tape archive storage facility. Space can be purchased through Research Cyberinfrastructure (RCI), and once adequately prepared, transferred via Globus for long-term archival. An example of this process is available in (Appendix A) (56).

CryoSPARC on Tempest

Tempest is a high-performance compute cluster operated by RCI, and contains a single, managed installation of CryoSPARC that can be used by any member of Tempest with their allocated Tempest resources (13). CryoSPARC on Tempest is feature complete, containing the latest, or near-latest release of CryoSPARC, 3D-Flex integration, Topaz and crYOLO integration (52).

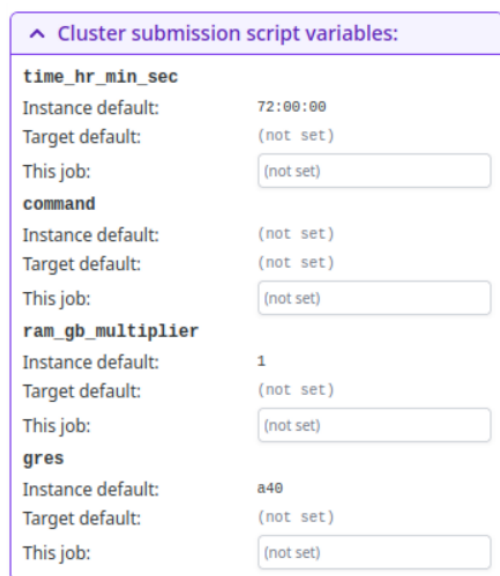
Using CryoSPARC. Because of permissions issues regarding the way CryoSPARC executes jobs, a service account run by RCI and the Cryo-EM Facility must be granted access to every group's data on Blackmore, and ability to run jobs on Tempest. For a new group that has never used CryoSPARC on Tempest, have the PI contact cryo-em_facility@montana.edu to have appropriate permissions arranged. Once access has been arranged, accounts must be registered on the CryoSPARC web server. To register accounts, contact cryo-em_facility@montana.edu with a list of first and last names, and associated emails for users you would like to have access

to CryoSPARC. That will initiate the registration procedure. Each user will be given a 4-character token. Navigate to the CryoSPARC homepage (REDACTED - Contact cryo-em_facility@montana.edu for address) and select "New Account". Fill out the form with the email that you submitted, and the token receive, and select "Create Account". Once the account is created, the user can login and begin to submit jobs.

To submit CryoSPARC jobs to Tempest, the user can use the normal GUI to queue them like normal. Each user will be given access to their group's associated "Lanes", one that submits jobs to the Tempest GPU Priority partition, using their group's reserved resources, and one that submits it to the Tempest GPU Unsafe partition, which can take advantage of unused capacity on the cluster, but with the risk that the job may be canceled to make room for other priority users should the need arise. There is also a "Live" partition that is available for those currently using the Arctica.

Figure 5. Lane selection in the CryoSPARC web user interface. Each research group will have separate lanes created for all of their Tempest partitions, and each group's lane will only be visible to members of that group.

Occasionally, a user may need to alter the time that the job is allotted, or the amount of system memory that is available. To change parameters, select "Cluster submission script variables:"



Cluster submission script variables:		
time_hr_min_sec	Instance default:	72:00:00
	Target default:	(not set)
	This job:	(not set)
command	Instance default:	(not set)
	Target default:	(not set)
	This job:	(not set)
ram_gb_multiplier	Instance default:	1
	Target default:	(not set)
	This job:	(not set)
gres	Instance default:	A40
	Target default:	(not set)
	This job:	(not set)

Figure 6. Custom cluster submission script variables. Note that “time_hr_min_sec” describes the format of the time entry. Command displays explicit submission commands to wrap the CryoSPARC job in – this can be useful for extending CryoSPARC functionality or wrapping the job in standard Unix software such as “time”. The ram_gb_multiplier option is how much to multiply CryoSPARC’s default memory allocation by – this is necessary as CryoSPARC chronically under-requests memory from the cluster for certain job types. The “gres” option allows selection of a “generic resource”, on Tempest this describes the type of GPU. It defaults to A40, but A100 and H100 GPUs are also available depending on the requested lane.

The variable “time_hr_min_sec” is in the format hr:min:sec and defaults to 72:00:00. The variable ram_gb_multiplier will multiply the amount of RAM that CryoSPARC reserves for a job by that value. For instance, if CryoSPARC reserves 40GB by default, setting this value to 2 will reserve 80GB. Most nodes on “gpupriority” have a maximum of 250 GB of memory that must be shared between two GPUs, although two nodes have 512 GB of memory. The Tempest

scheduling software (SLURM) will automatically determine which nodes to use based on the resources requested.

Topaz is available for use in CryoSPARC, as well as throughout Tempest. To select which version of Topaz to run, when setting up the Topaz job, press the directory icon below "Path to Topaz executable". By default, this should place the file picker back in \$HOME, but if it doesn't, type \$HOME into the path and press Enter. Once again, navigate by double-clicking into the topaz directory, select the desired version, ending in .sh.

Managing CryoSPARC. CryoSPARC is running on a virtual machine managed by RCI. The user account running CryoSPARC is the active directory service account “cryosparc”. This user's home folder contains the following directories:

/home/cryosparc

```

├── cryosparc_cluster
├── cryosparc_master
├── cryosparc_worker
├── slurm-examples -> /mnt/shared/slurm-examples/
├── test_project
└── topaz

```

CryoSPARC Master is installed under cryosparc_master. The configuration file for the Master process is detailed in Appendix A. The license key is registered to cryo-em_facility@montana.edu. CryoSPARC Worker is installed under the cryosparc_worker directory. The configuration file for the worker is detailed in Appendix A. The worker process is not installed in a traditional manner, due to the cluster setup. It is installed not as the

"Standalone", where the master and worker are the same install, but as the separate installation. However, no SSH access between the master and worker is needed because of the cluster configuration. One potential flaw in this setup is that the CUDA path is hardcoded into the configuration. This is unavoidable at the moment, but if CUDA functionality breaks, one thing to check is if that path still works. The path can be found in the CryoSPARC master config.sh file.

cryoSPARC can be managed via systemd. Currently, a systemd service file is installed at `"/home/cryosparc/.config/systemd/user/cryosparc-supervisor.service"`. It is detailed in Appendix. It is important to note that this is a user process, not a system process. Since systemd sometimes likes to kill user processes after logout, `"loginctl enable-linger"` was enabled for the cryosparc user. This sometimes still fails for unknown reasons, so if the cryoSPARC processes keep dying, check if systemd is killing them after logout via the aforementioned loginctl command. Do not use `"cryosparcm start"` and `"cryosparcm stop"` to manage CryoSPARC. Instead, use `systemctl` (example: `systemctl --user start cryosparc-supervisor.service`) as the 'cryosparc' user on the VM

CryoSPARC can be updated via the `"cryosparcm program"`. To do this, first shut down the systemd service by running `"systemctl --user stop cryosparc-supervisor"`. Then, update CryoSPARC by running `"cryosparcm update"`. This procedure should update CryoSPARC to the latest version. Sometimes the `"cryosparcm update"` command takes a very long time to download. In this case, run `"cryosparcm update --download-only"` while CryoSPARC is running, and stop CryoSPARC after the download is complete, and run `"cryosparcm update --skip-download"` to perform the actual installation.

CryoSPARC considers patches different than updates. Patches may solve critical bugs. To check for patches, run `"cryosparcm patch --check"`. If a patch is found, download the patch

manually with “cryosparcm patch –download-only”. Copy the downloaded “cryosparc_worker_patch.tar.gz” to “/home/cryosparc/cryosparc_worker”. Go to that directory and run “bin/cryosparcw patch”. Then, reload the interface and refresh the job types by running “cryosparcm cli "reload()” and “cryosparcm cli "refresh_job_types()”.

cryoSPARC uses MongoDB to manage users, projects, jobs, etc. This database is currently kept in the “/mnt/cryosparc_db” folder and is on a separate drive than the virtual machine (VM). Backups are enabled using systemd timers and the existing “cryosparcm backup” functionality, and the systemd timer files can be found in Appendix A. These services are installed under “/home/cryosparc/.config/systemd/user” and have been enabled and must be manually enabled every time the VM is restarted via the ‘systemctl enable <filename> command

CryoSPARC can integrate with SLURM. To add a SLURM cluster as a lane, two files are needed. A “cluster_info.json” file containing the cluster details and configuration, and a “cluster_script.sh” file to use as the template for a job submission script (Appendix A). These two files must be in the same directory, and to add them to CryoSPARC as a new lane, run “cryosparcm cluster connect” in that directory. A different directory is needed for each lane. Currently, the directory “/home/cryosparc/cryosparc_cluster/” contains directories by each Tempest group. Inside that directory are further directories, by lane name.

To add a group, simply duplicate one of the existing folders, and make changes to the files. Then run “cryosparcm cluster connect” in each new folder. Example configuration files are available in Appendix A. Things to modify when adding new lanes are the “name” which is what the lane will be named in CryoSPARC. Current syntax is “groupname_tempestpartition”. Additionally, “cache_path” should point to the group's group directory on Tempest. The

cryosparc user will need to have read and write permissions to this directory. Finally, "cache_reserve_mb" is how much free space CryoSPARC will always leave empty. This can be customized as required.

When adding new lanes, "--account" should be changed to the Tempest account of the group, and "--partition" may need to be changed to be the correct partition that this lane will submit jobs to. Additionally, there may be some optimization that can be performed here. For instance, should Tempest ever acquire > 2 GPU systems, the script at the bottom may need the devidx increased to 0 3 or whatever the number of GPUs is. However, it's currently untested if that script is even required, or if SLURM will handle everything itself just fine. Furthermore, things like RAM and nodes and ntasks might be able to be optimized further.

RCI must add the cryosparc user (uid=200248) to the Unix group for each lab/group using Tempest. Then, their Blackmore share must be mounted by RCI at /mnt/blackmore/<group_name>. However, CryoSPARC will be unable to access the files until it is fully restarted. To perform a full restart, one must first stop the cryosparc-supervisor.service and then log out of all instances with loginctl terminate-user cryosparc. Then, log back in, and the systemd service should automatically start. If there are issues getting access to the new group's folder, check that the cryoSPARC supervisor process is running as the correct groups by running "grep '^Groups' /proc/<csparc_procid>/status" where <csparc_procid> is the process ID number of the running CryoSPARC process. This should return a list of integers that correspond to the group IDs the process is running with. Ensure the new group is present. Note that restarting CryoSPARC will kill all running jobs.

In both the `cluster_info.json` and `cluster_script.sh` files, environment variables can be used by double curly bracketing them: `{{ variable }}`. For instance, `{{ time_hr_min_sec }}` is set to a variable. These variables can be set to an instance-wide default by administrators, but individual users can then modify them at submission time if needed. To set instance wide defaults, log into cryoSPARC with an administrator account. On the left-hand side of the screen, select the key icon.

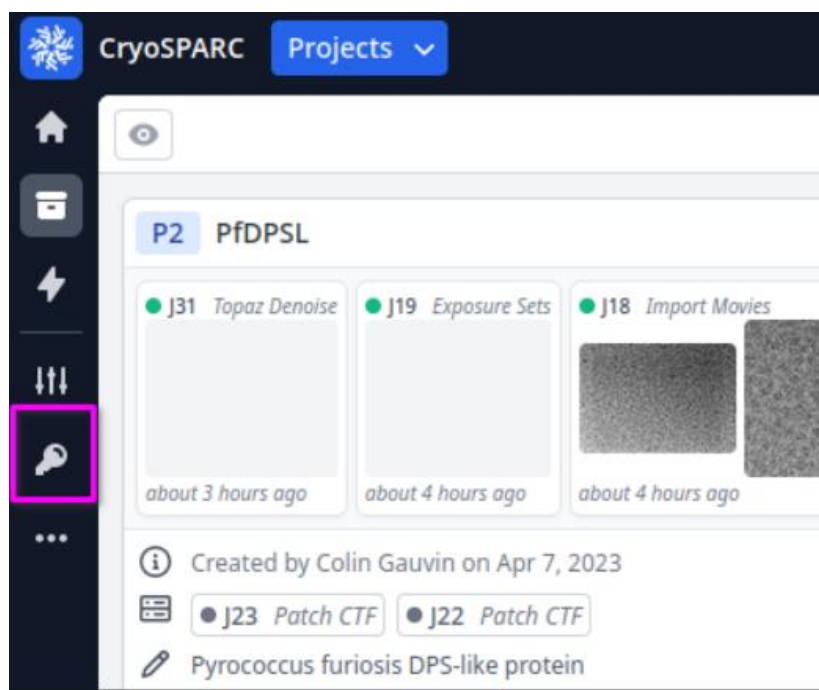
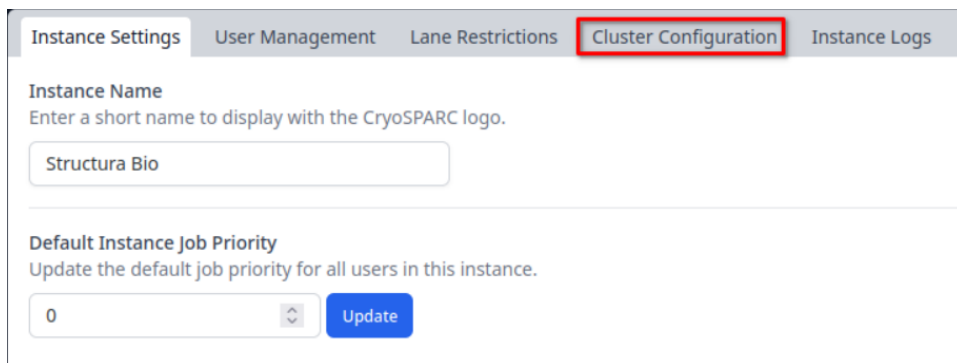


Figure 7. Admin panel location in the CryoSPARC web user interface.

In the top menu, navigate to “Cluster Configuration”. This is where custom cluster variables reside. Instance wide defaults can be set at the top. These will apply to all lanes on the instance. However, individual lane-wide defaults can be set below if desired. When users submit jobs, they can override these defaults on a per-job basis. This is helpful, for instance, if there is a

pending shutdown of Tempest and the job timer needs to be shortened to accommodate jobs in the meantime. See "Submitting Jobs to Tempest" for more information on submitting jobs.



Instance Settings User Management Lane Restrictions **Cluster Configuration** Instance Logs

Instance Name
Enter a short name to display with the CryoSPARC logo.

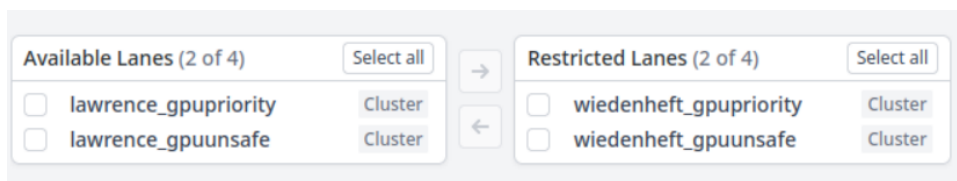
Structura Bio

Default Instance Job Priority
Update the default job priority for all users in this instance.

0 Update

Figure 8. Cluster configuration settings in the CryoSPARC web user interface.

Specific lanes can be restricted to specific users. Administrators should ensure that new users have access restricted to only their group's Tempest resources. This can be done by logging in with an administrator account, and navigating to the Admin page. In the top menu that opens, navigate to "Lane Restrictions". Select a user on the left. A new screen will open showing available and restricted lanes. Set these using the checkboxes and arrow icons accordingly.



Available Lanes (2 of 4)		Restricted Lanes (2 of 4)
☐ lawrence_gpupriority	Cluster	☐ wiedenheft_gpupriority
☐ lawrence_gpuunsafe	Cluster	☐ wiedenheft_gpuunsafe

Figure 9. User lane access management in the CryoSPARC web user interface. Marking a lane as “Available” to a user will let them submit jobs to that lane. The “Restricted” status will prevent the lane from being visible to that user.

Users must be added by an administrator. To do this, log into CryoSPARC with an administrator account, and select the key icon on the left-hand side of the screen. In the top

menu, select "User Management". At the top right of the screen, select "Create New User". Enter the details. The email entered here must be the one that is used when the user finishes the registration process. Once done, a token will display. Email that token to the user, who can then create a user account with the provided email and token.

Single Particle Protocols

This section will detail current best practices for SPA data collection using SerialEM, or SmartScope.

SerialEM

Screening. Before starting anything, first ensure that the microscope is setup to operate properly, and that SerialEM is configured as necessary. This is important so that one does not overwrite anyone else's settings, and also because SerialEM is configured to run a script upon starting that will place the microscope in a reasonable starting state (Appendix). On Magaera, i.e., the Gatan computer, close any open SerialEM instances. Then, examine the X:\ drive for available space.

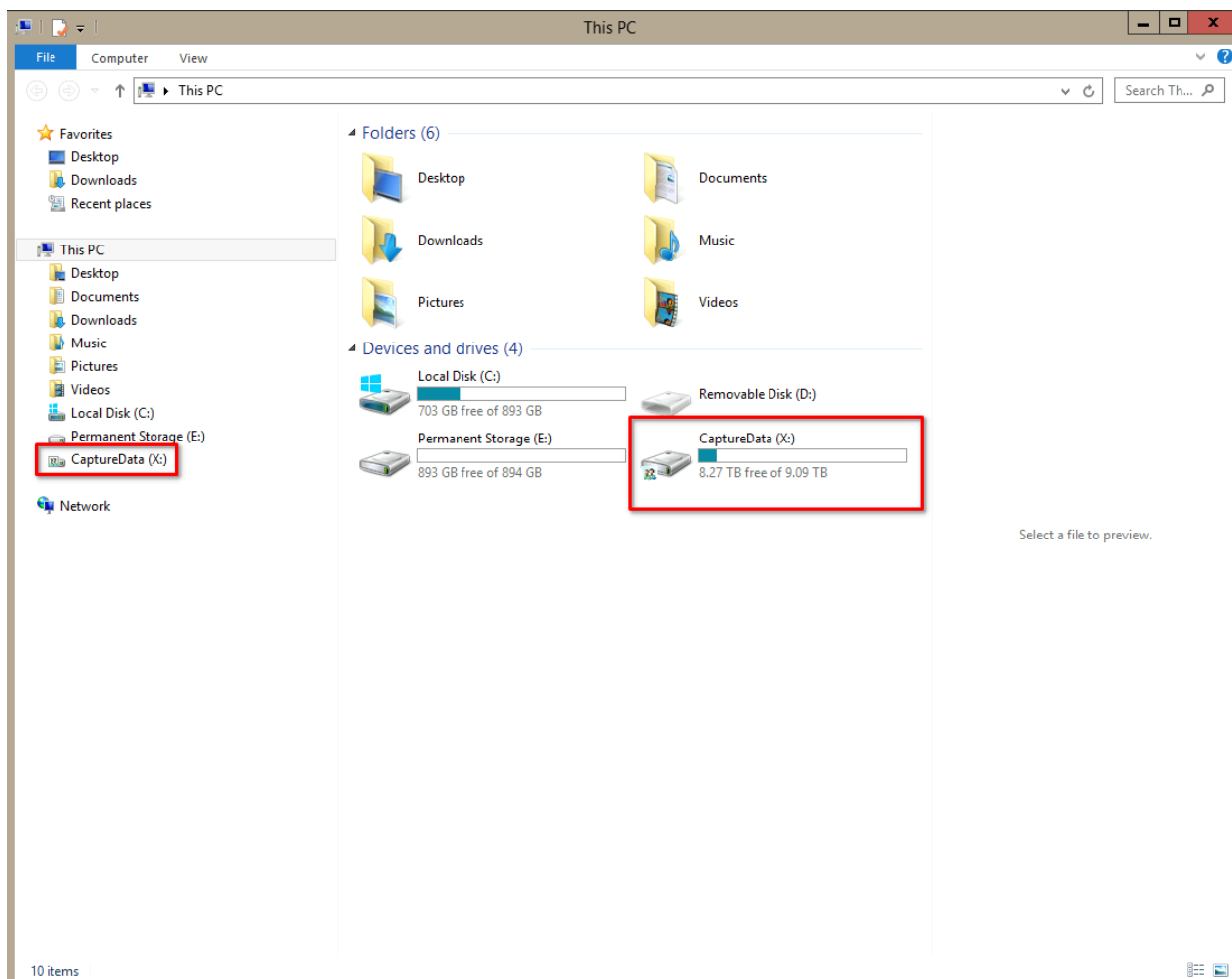


Figure 10. Storage space on Gatan K3 data collection computer.

There should be at least 3 TB per data collection overnight. Next, check the autoloader to ensure no other samples are loaded. If no other grids are loaded, ensure that the autoloader temperatures and vacuum are stable, and that the column valves are closed.

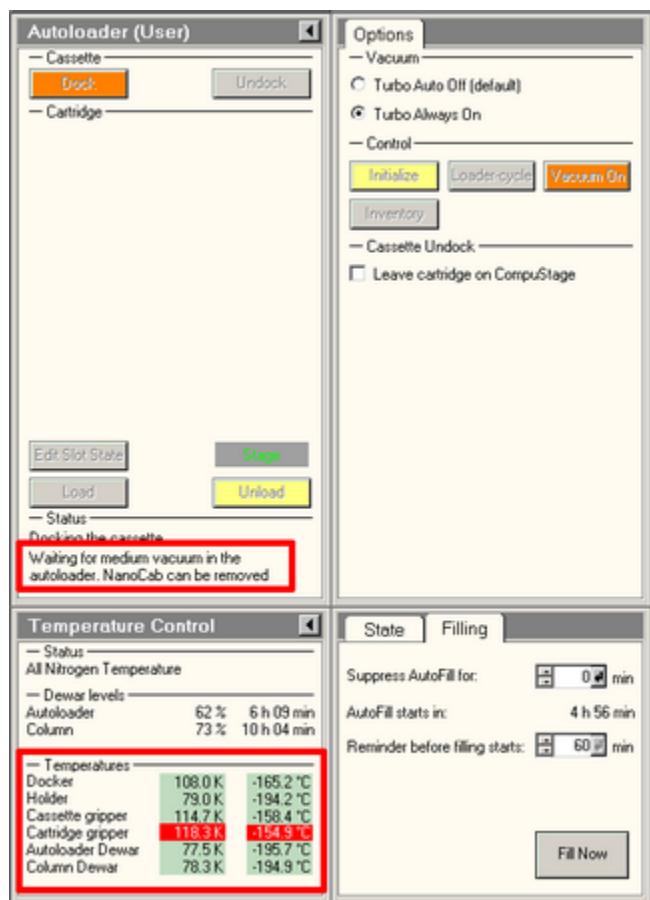


Figure 11. Auto-loader and temperature control panel. Note that temperatures will be displayed in red when above the recommended cryo-operation temperature.

Dock a NanoCab containing the cassette with your sample grids. Mount the NanoCab in the Autoloader brackets inside the TEM enclosure and press the green Dock/Undock button on the TEM digital display.

Before working with the grids, they must be imaged into a grid map that contains a corresponding Cartesian coordinate system that SerialEM can work with to identify grid squares and grid holes. Open the Autoloader tab on the Microscope PC. Run the grid inventory procedure for the cassette to allow the microscope to probe which cassette positions are occupied, and which are unoccupied. Optionally, you may enter a text description of each grid. To run the

inventory, select the flapout menu in the TEM UI (User Interface), and press the Inventory button.

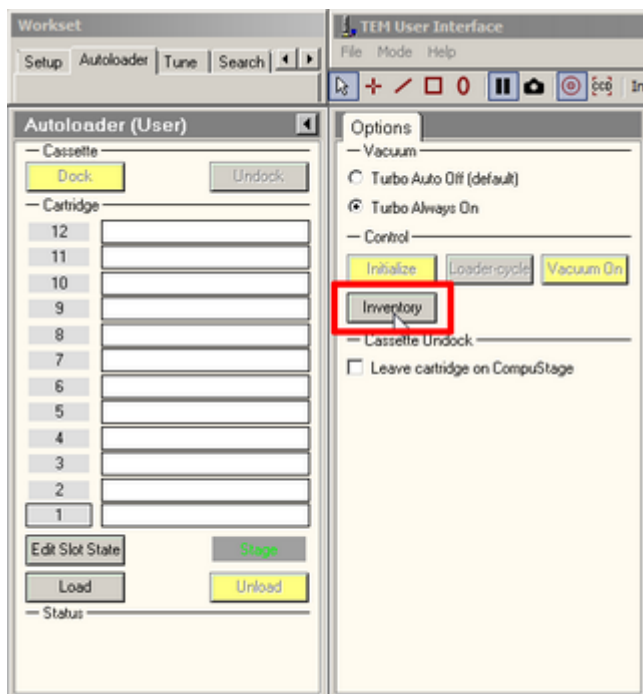


Figure 12. Auto-loader inventory mechanism, which will check the status of each slot in the microscope cassette.

After inventory, each occupied cassette slot will turn blue, and each unoccupied cassette slot will remain gray. Start SerialEM via the "Zuul" shortcut on the Windows desktop on the Gatan computer (the computer with the large monitor). Select your account if you have one. If you do not have an account, please use the Guest account. SerialEM will ask you to set your working directory. This should be a top-level directory on the X:\ drive of your name, or lab's name. If you do not have one, please make one.

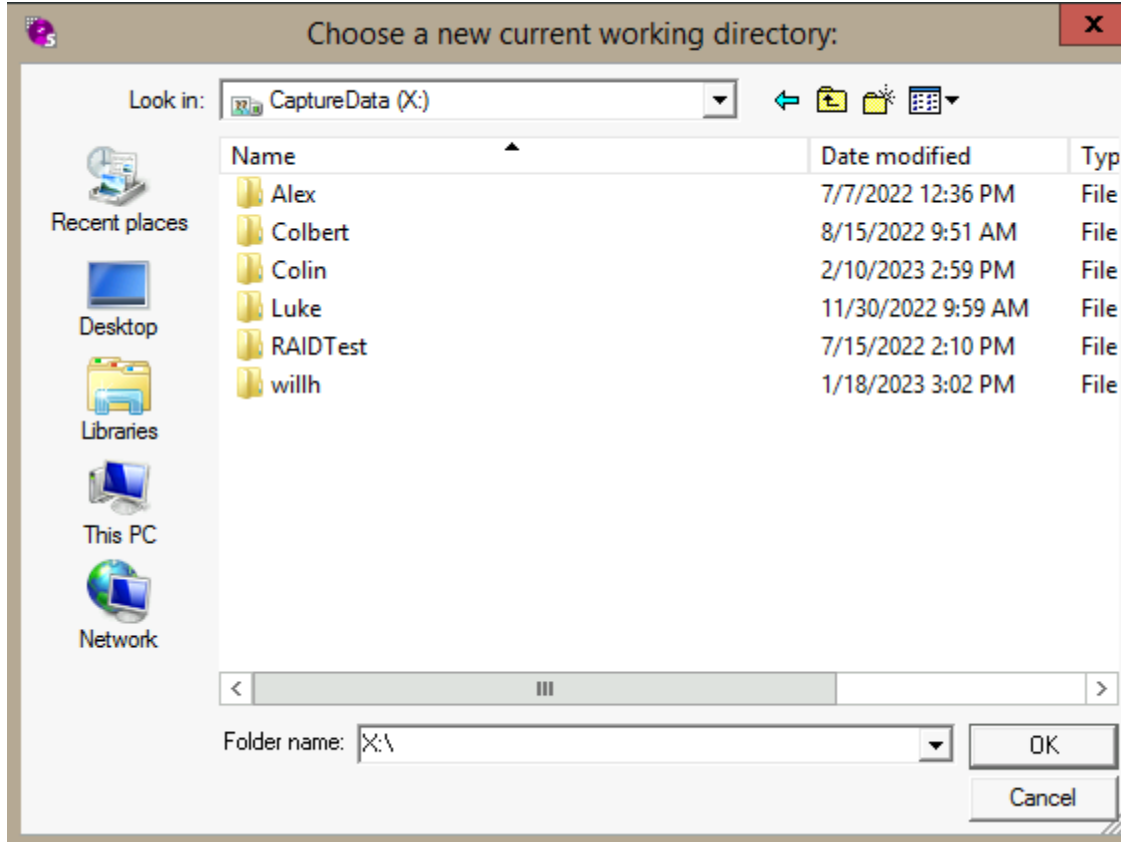


Figure 13. SerialEM current working directory dialog allows the user to select their base directory when using SerialEM.

Once set, the log will print out your current directory. In SerialEM, open the Navigator window, and then open the Imaging States window. If the Imaging States window does not automatically open with the Navigator window, go to the Navigator top menu, and select Open Imaging States. Select the Grid Map imaging state from the Imaging States menu by double left-clicking it. Setup a new montage by selecting the Navigator top menu, followed by “Montaging & Grids” and finally “Setup Full Montage.

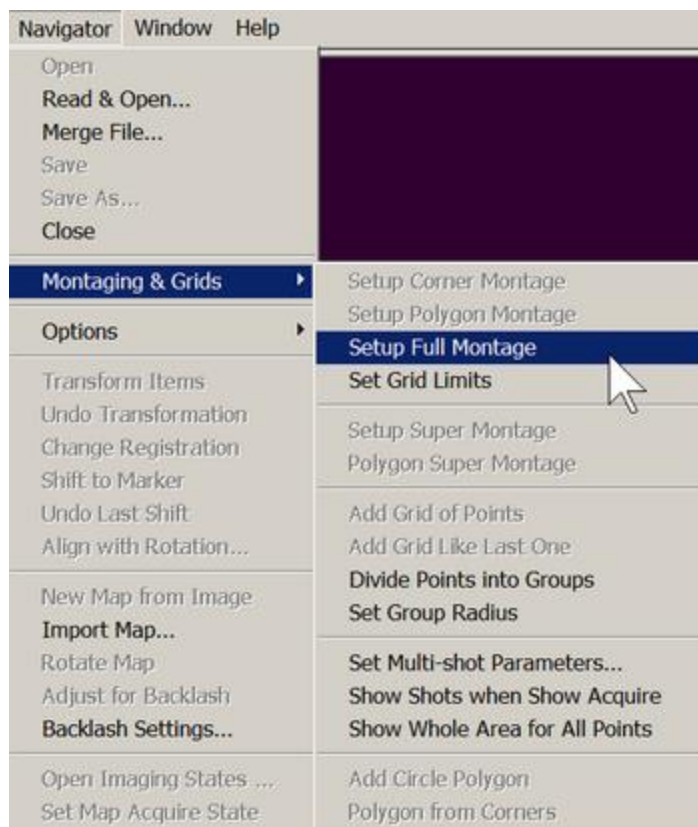


Figure 14. SerialEM navigator menu full-montage selection.

A new window will pop up asking for montage settings. Be sure to set Use Search parameters in Low Dose mode, and not select “Use View parameters” Search parameters should be magnification of 46x, the lowest magnification that still shows a full field-of-view. This should normally never change. The montage area should be 5x7.

Montage Setup

Camera

☐ Ceta

☒ K3

FITTING TO NAVIGATOR AREA: Change mag to adjust number of pieces. Changing mag, binning, overlap, or "Move stage" will refit to area

Magnification: 46 Binning: 2

Pixel size: 166 nm

Number of pieces in X: Y:

Piece size in X: Y:

Overlap in X: Y:

Minimum overlap: 10% and micron

Total Area: 12848 x 12720 pixels
2128 x 2107 microns

☒ Move stage instead of shifting image

☐ Image shift in blocks, IS up to microns

☐ Skip pieces outside Navigator item

☐ Do full rectangle; ignore list of pieces to skip

☒ Make map from each montage if Navigator open

☒ Close file when montage is done

☐ Use Montage Mapping, not Record parameters

☐ Use View parameters in Low Dose mode

☒ Use Search parameters in Low Dose mode

☐ Acquire with Multiple Record routine

☐ Use continuous mode with settling factor

☒ Turn off Drift Correction for stage montage

☐ Use settings for high-quality stage montage

Figure 15. SerialEM full montage setting, depicting the proper number of pieces, and low-dose mode selection.

SerialEM will then prompt the user to setup the montage file. At this time, the recommendation is to use the MRC file format for saving the montage file, as HDF file format occasionally causes a SerialEM crash when re-loading the file. The user may save the associated metadata if desired. Save the file to your folder on the X:\ drive. This will save your grid maps which you will likely re-use, so put it somewhere workable and give it a distinctive name.

File Properties

File type

Save images to

- ☒ MRC stack file
- ☐ HDF stack file
- ☐ TIFF file (one image per file)
- ☐ Series of TIFF files listed in an Autodoc file
- ☐ JPEG file (one image per file)

Type of compression in TIFF or HDF file

- ☐ None
- ☐ ZIP
- ☒ LZW
- ☐ JPEG

Image data treatment

Save non-float data as

- ☐ Bytes
- ☒ Integers

When saving 16 bit data

- ☒ Truncate above 32767
- ☐ Divide by 2
- ☐ Subtract 32768

Percent of pixels to truncate converting to bytes

As black (0): As white (255):

Metadata

Save in extended header

- ☒ Tilt angle
- ☒ Intensity
- ☒ Stage position
- ☒ Magnification
- ☒ Exposure dose

Maximum number of montages:
(Be generous)

☒ Save extra information in a '.mdoc' metadata file

☐ Skip this dialog in future (re-enable in File menu)

OK Cancel ?

Figure 16. SerialEM montage file-saving settings. MRC is recommended for grid atlas format.

Next, ensure that the apertures are set appropriately for low-magnification imaging. This means that the Condenser 2 aperture should be set to 150 μm , and the Objective aperture should

be removed. This can be accomplished by using the Microscope PC to navigate to either the Autoloader or Tune tabs and changing the aperture settings.

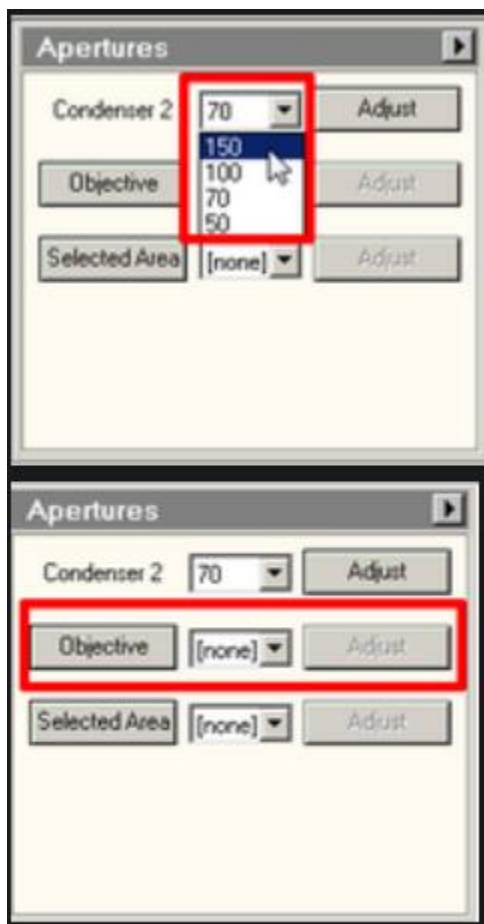


Figure 17. Talos Arctica aperture control panel depicting available aperture options.

Now, the grid maps can be recorded. For multiple samples, use the “Atlas Grids” script (Appendix). Navigate to the “Script” top menu and select “Run” and “Atlas Grids” (Script 2). Two windows will pop up, the first window will ask which grid number to start with. The second window will ask which grid number to stop with. For instance, if autoloader slots 1-4 are occupied, the user would enter 1 and 4 respectively. The script will subsequently montage all the requested grids. This operation will take approximately 10 minutes per grid.

Once grid atlases are created, to get an idea of what particles look like, one should record several images of holes in several grid squares of varying ice thickness and size. These images will provide the user with information about the quality of their ice and sample, and what approximate resolution can be expected to collect data out to in each grid square. To do this, first load a grid map. Check the recorded grid maps in the Navigator window for interesting areas, by double left-clicking on the navigator entry.

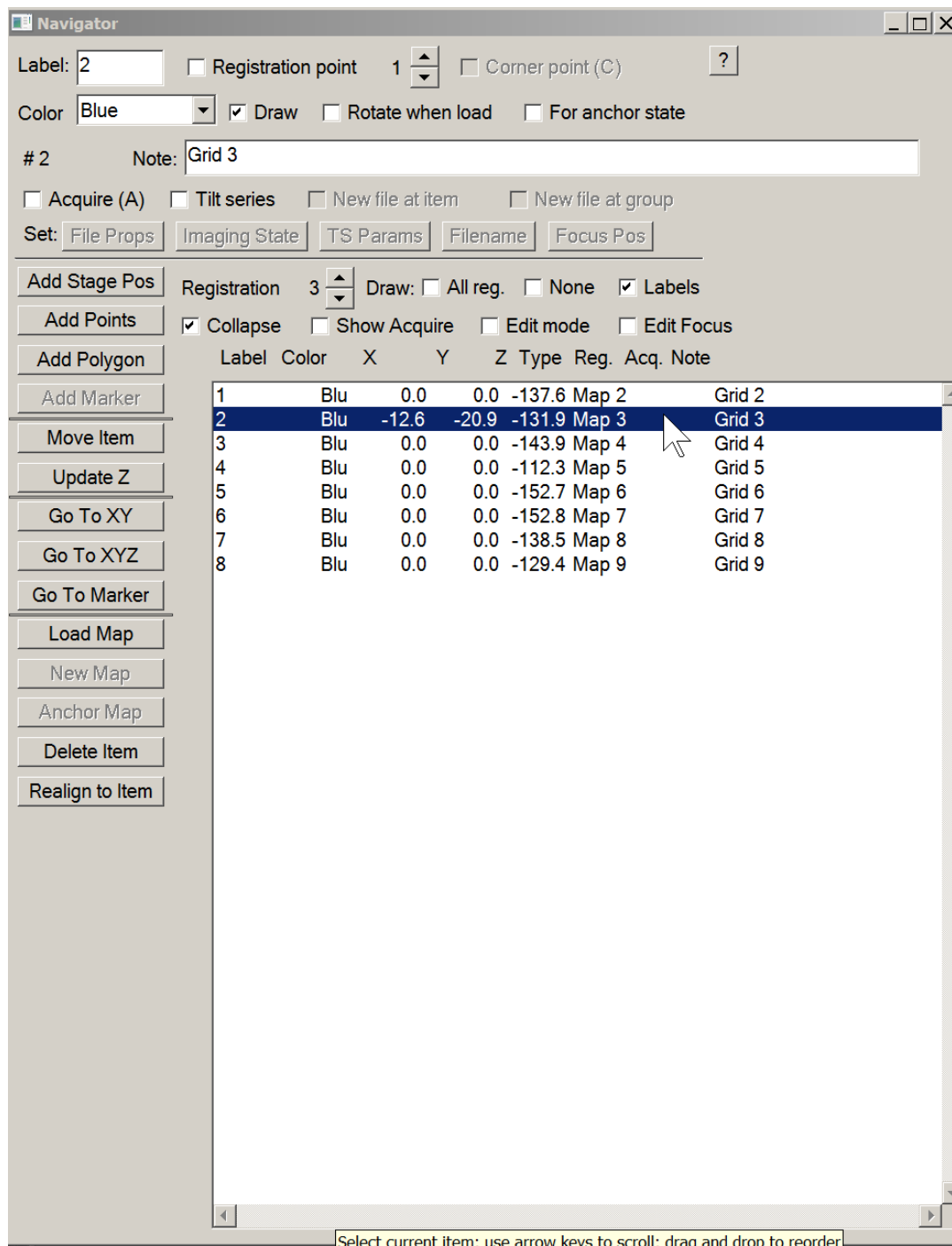


Figure 18. SerialEM navigator dialog window. Double click to load the desired grid map.

Ensure that the respective grid is loaded. If screening directly after creating grid atlases, the last grid will still be loaded. Therefore, it is typically a good idea to start with the last grid. To

ensure that the correct grid is loaded, check the “Autoloader” tab on the Microscope PC. The loaded grid will have a yellow highlight next to its name. If the incorrect grid is loaded, one may load the correct one by left-clicking the grid number and selecting “Load”.

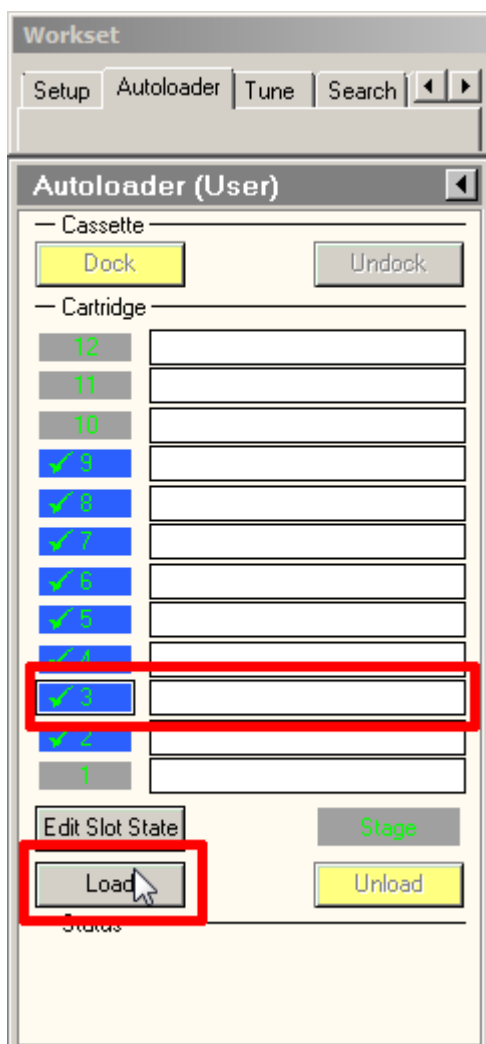


Figure 19. Talos Arctica auto-loader loading dialog. Gray numbers represent empty cassette slots, blue numbers represent full cassette slots, and the presence of a green checkmark denotes that the cartridge in that location has been loaded onto the microscope stage at some point.

Once the grid is loaded, be sure to disable the autoloader turbomolecular pump. This pump is useful when performing many sequential grid transfers, such as when recording grid

atlases, but vibrations from the pump will cause motion in the microscope stage that results in noticeable drift at high magnifications, resulting in reduced high-resolution features, and aberrations in the Fourier transform. To disable the autoloader turbomolecular pump, select the “Autoloader” tab on the Microscope PC, and select the top-most flap-out. Then, select “Turbo Auto Off”.

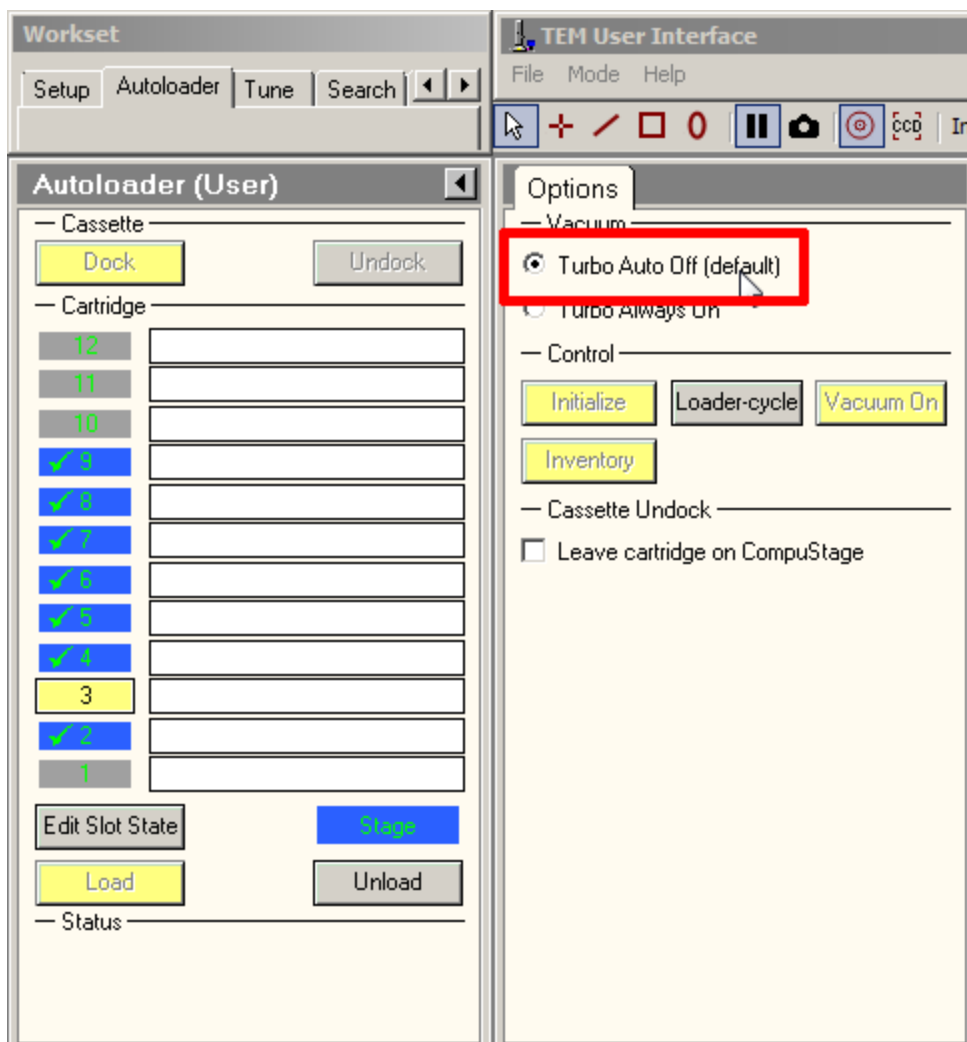


Figure 20. Talos Arctica auto-loader turbo molecular pump controls. Also note the yellow cassette slow (#3), denoting the active cartridge currently on the stage.

At this point, the column valves can be opened, either via the Microscope PC or SerialEM PC. Once the column valves have been opened, the user is free to navigate to areas of interest on the map, by placing points on the grid square map. To place points, first put the SerialEM Navigator into “Add Points” mode, by selecting “Add Points” on the Navigator window.

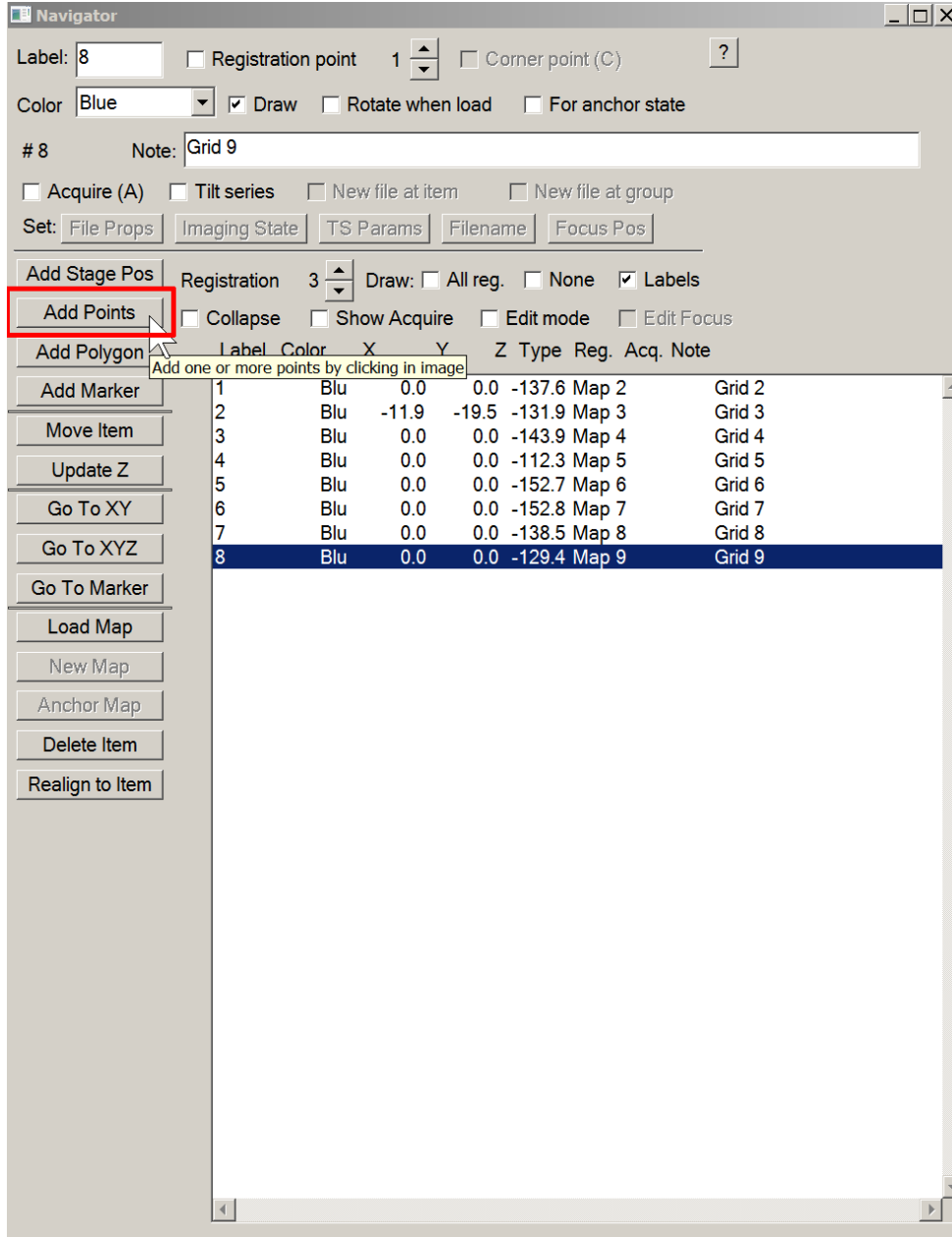


Figure 21. SerialEM navigator dialog window demonstrating "add points" routine.

Once the Navigator is in "Add Points" mode, left clicking anywhere on the open grid map will place an X and Y coordinate in the navigator at that location. It is worth noting that

points added together, in the same “Add Points” session, will be treated as a group. To stop adding points, select “Done Adding Points” in the Navigator window.

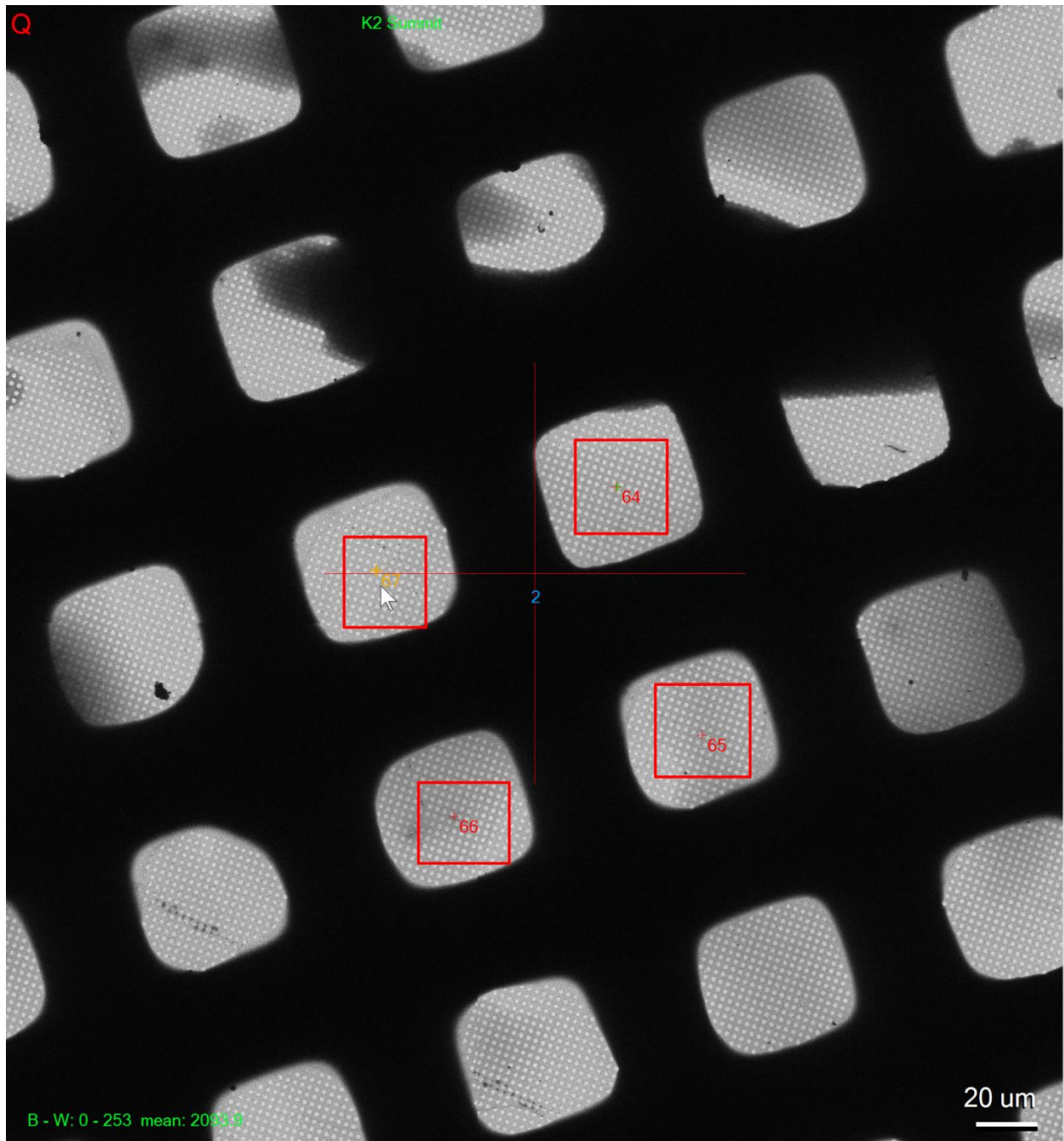


Figure 22. Low magnification image of an EM grid demonstrating added points location. Red crosses depict placed points, the yellow cross depicts the currently selected point in the navigator.

Once points are added, the grid squares can be screened. Screening involves navigating to each square, recording and saving a eucentric height for each grid square, and imaging the contents of the holes. To navigate to a freshly placed point, select the point in the Navigator window, and press either of “Go to XY” or “Go to XYZ”, as no Z height is yet record, the two buttons should be functionally identical. Once the stage has finished moving, insert the 50 μm Condenser 2 aperture. Then, ensure that the microscope is in low-dose mode by selecting “Low Dose Mode” in the “Low Dose Control Panel” if it is not already selected.

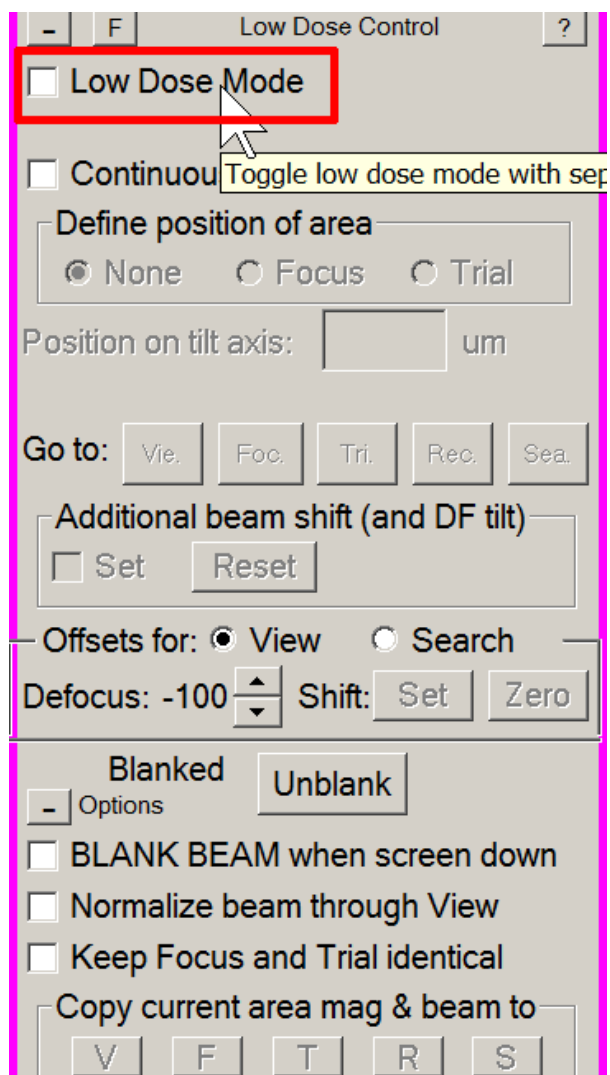


Figure 23. SerialEM “Low Dose Control” panel toggle.

To ensure that the stage is centered over the square, and there are no square edges visible, record a View image by selecting “View” under the Camera & Script control menu. This will take an intermediate magnification image (~1200x or 1600x) of the grid square.

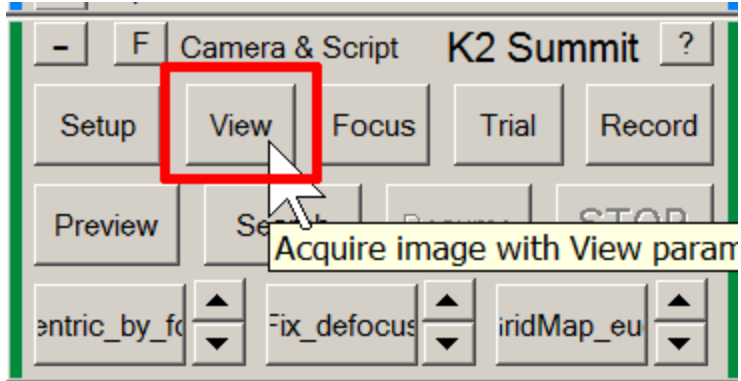


Figure 24. SerialEM "Camera & Script" control panel.

If the image looks good, that is, there is sufficient contrast and no grid square edges showing, then eucentric height should be recorded and saved for the square. To do this, select the top "Tasks" menu, then "Eucentricity" and finally "Rough Eucentricity". This will use the "View" mode to take a series of images while tilting the stage to calculate the Z-height offset necessary to bring the square to eucentric height.

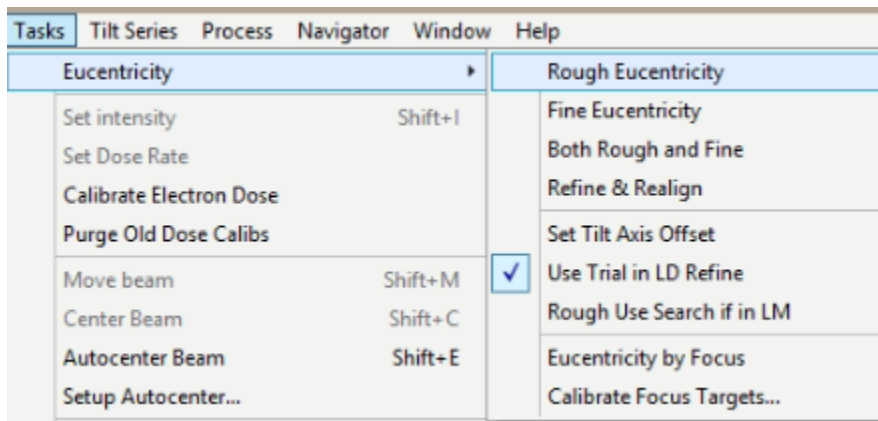


Figure 25. SerialEM "Tasks" menu showing "Eucentricity" options.

The eucentricity routine will take some time to complete. Once it is done, be sure to update the coordinates with the proper Z position by selecting "Update Z" on the entry in the

Navigator window. Proper eucentricity is required to enable proper focusing, which will be the next step. Typically, focusing is done on the carbon or gold support film between holes. The focusing area should be sufficiently far from the holes that it does not overlap with the hole when focusing, so that sample in the holes are not exposed to the electron beam twice. Additionally, there must be enough contrast in the area for the autofocus routine to work properly. To set the position of the focus area, take a new View image, as the image on your screen may have shifted due to the change in Z-height. Then, select “Define Position of Area > Focus”, and then left click on a suitable position on the View image in Buffer A. Select “None” again, before continuing past this step.

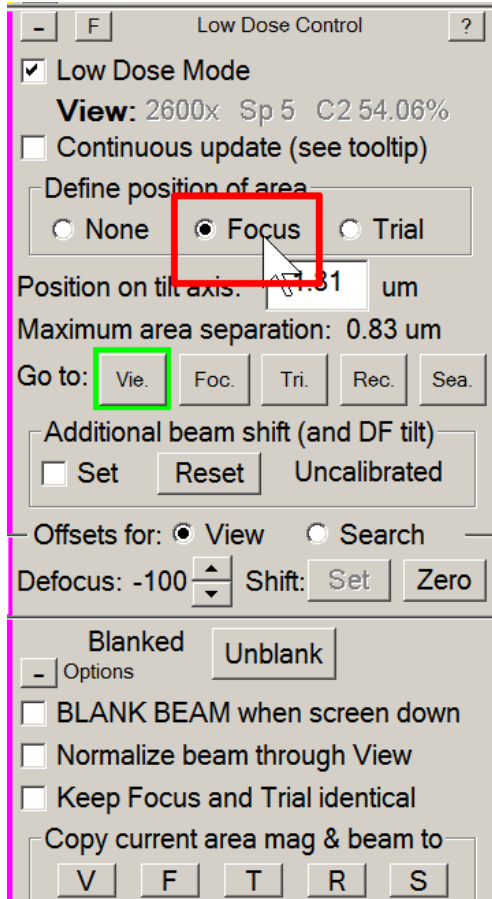


Figure 26. SerialEM "Low Dose Control" panel showing options to define position of "Focus" area.

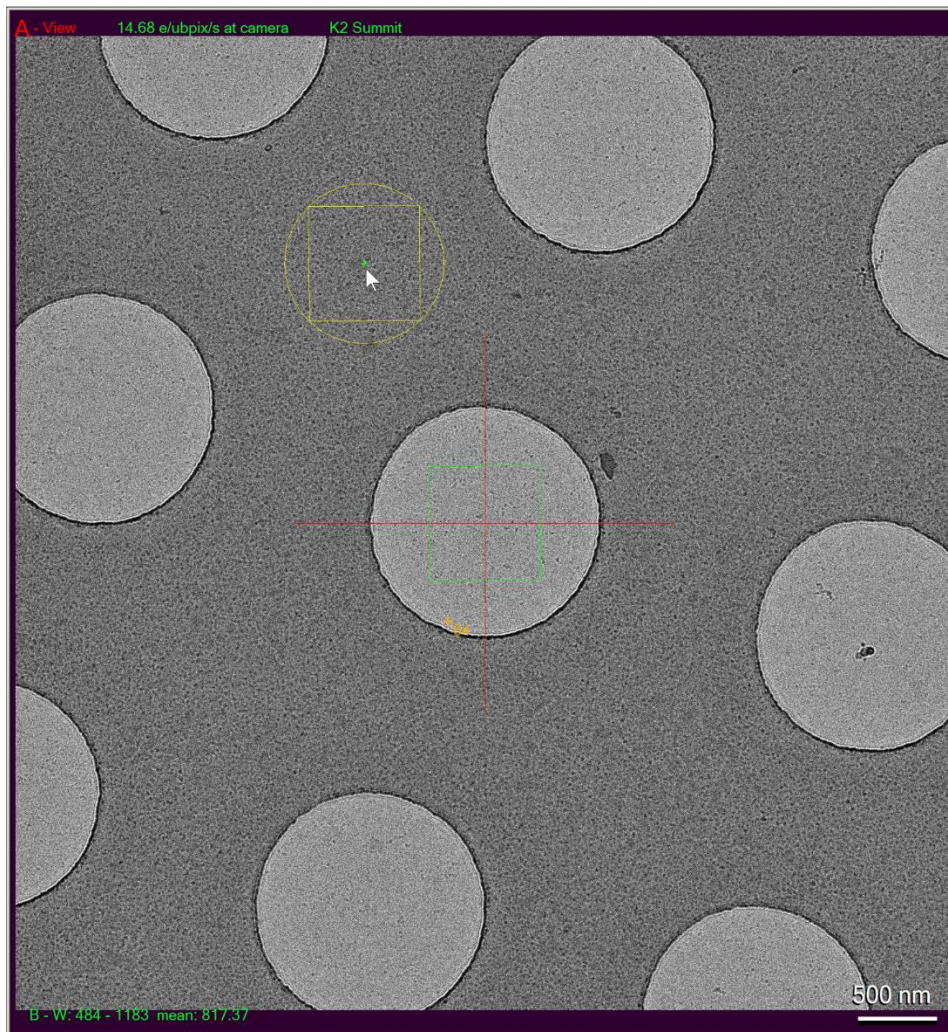


Figure 27. Medium magnification image demonstrating how to properly define the "Focus" area for data collection. The yellow box denotes the location of the focus position relative to the green box, which denotes the current "Record" position. The delta between the two positions will be preserved once set.

Once the focus position relative to the holes is set, the autofocus routine can be run. To set the target focus, select the "Focus/Tune" menu, and then "Set Target". Defocus targets are entered in μm . Once the defocus target is set to something reasonable, run the "Autofocus" routine by selecting "Autofocus" in the "Image Alignment & Focus" control panel, or by using the keyboard shortcut Ctrl + G. SerialEM will shift the beam to the target area, and then record a

series of images with different beam tilts, and calculate the existing microscope focus, and the offset needed to reach the target defocus. It will then automatically apply that offset and shift the beam back to where it started. Once the autofocus routine is finished, a micrograph can be recorded by selecting “Record” in the “Camera & Script” control panel. This will use the settings currently loaded into the “Record” low dose preset. Once the image is recorded, it can be saved by navigating to the “File” top menu and selecting “Save to Other”. Repeat this process as needed to screen all desired locations.

Data Collection. This tutorial covers setting up an automated data collection for single particle analysis, using SerialEM. Follow from start to finish to collect data for single particle analysis.

Before working with a grid to set up an automated data collection, one needs to have an accurate map of that grid. While there may exist a previously created grid map from the screening workflow, it is recommended to create a new map before setting up data collection, as when the grid is re-loaded onto the stage, it may rotate slightly, which will cause navigation coordinate inaccuracies. To create a new grid map, first ensure that low dose mode is active by checking the “Low Dose Control” panel. Additionally, on the Microscope PC under either the “Autoloader” or “Tune” tab, ensure that the C2 Condenser Aperture is set to 150 μm , and that the Objective aperture is removed. Also, ensure that the column valves are open. Then, in the top Navigator menu, select “Montaging & Grids” followed by “Setup Full Montage”. Ensure that the montage and file settings match those in Figure 15. At this point, rather than re-run the montage script, simply run single atlas of the current grid by navigating to the “Montage Controls” in the control panel. If it is not visible, collapse other control panels until it appears at the bottom.

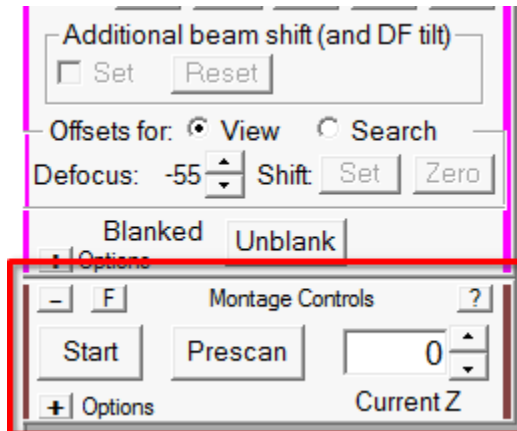


Figure 28. SerialEM "Montage Controls" panel.

You will collect the first gridmap shortly, but as you are doing that, you will also setup multishot imaging. Multishot greatly speeds up automated data collections by centering and focusing around one central grid square hole, and then using beam-image-shift (BIS) instead of stage movement to collect the surrounding holes. Common multishot configurations are 3x3 or 5x5 matrices, depending on the size of the grid holes and their spacing. To setup multishot, SerialEM needs to be made aware of both the regular hole spacing of the grid, as well as the rotation of the grid relative to the camera. Both pieces of information can be calculated automatically from a low-magnification image of a grid square. To begin, select the “Grid Square Map” imaging state, and ensure that the 50 μm C2 Condenser aperture is inserted, and the Objective aperture is removed.

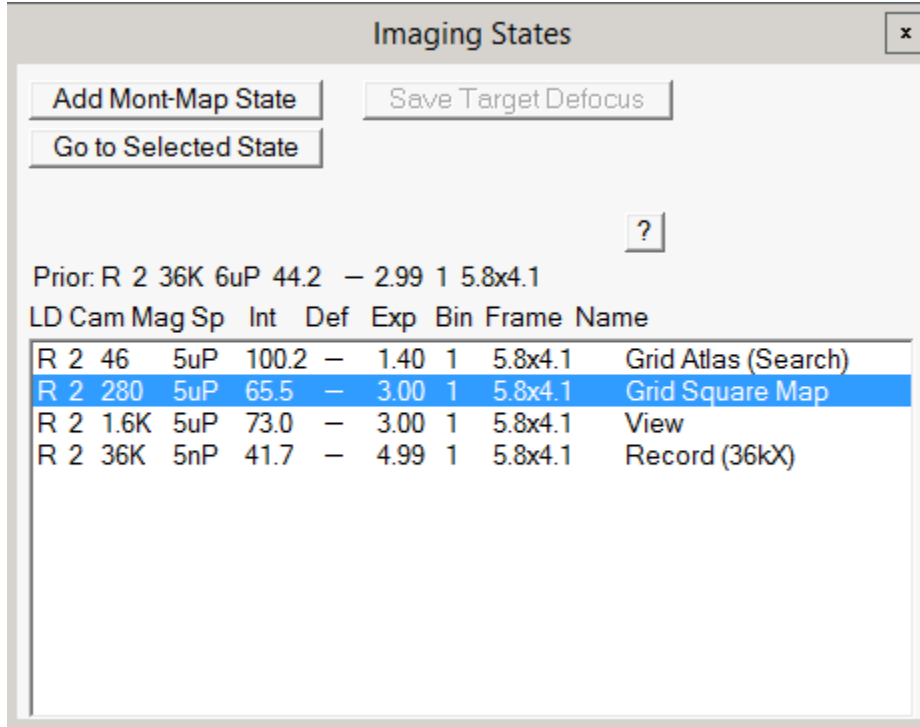


Figure 29. SerialEM "Imaging States" dialog window.

This will load a low-magnification imaging state into the “Record” low dose preset. Navigate to a grid square. Once that is loaded, take a Record image of the grid square. The grid square should be mostly visible within the image frame, and the holes should be identifiable. If the image looks good, in the “Navigator” top menu, open the “Montaging & Grids” sub-menu, and select “Find Holes in Regular Grid”.

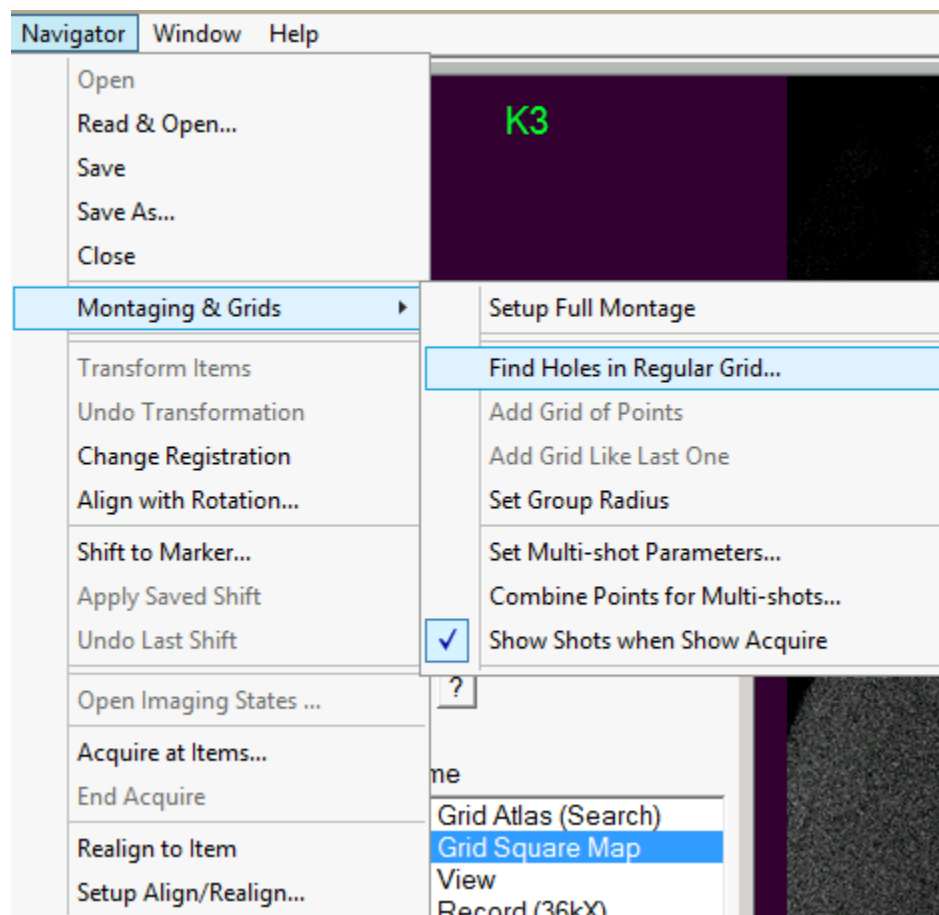


Figure 30. SerialEM "Montaging & Grids" sub-menu.

A new dialog window will appear that offers a variety of options for finding holes in a regular grid. If using holey carbon grids, typically all the options can be left at default, except for the spacing. The hole size should be set to the hole size of the grid. The periodicity should be equal to the sum of the hole size, and the spacing. For R 2/1 grids, hole size is 2, and periodicity is 3. For 1.2/1.3 grids, hole size is 1.2, and spacing is 2.5. In some cases, SerialEM may complain it cannot match the spacing when trying to detect holes. It may be worthwhile to measure the size of the holes on the grid in a View image, as they are not always accurate. Once the hole size and periodicity are entered, select "Find Holes".

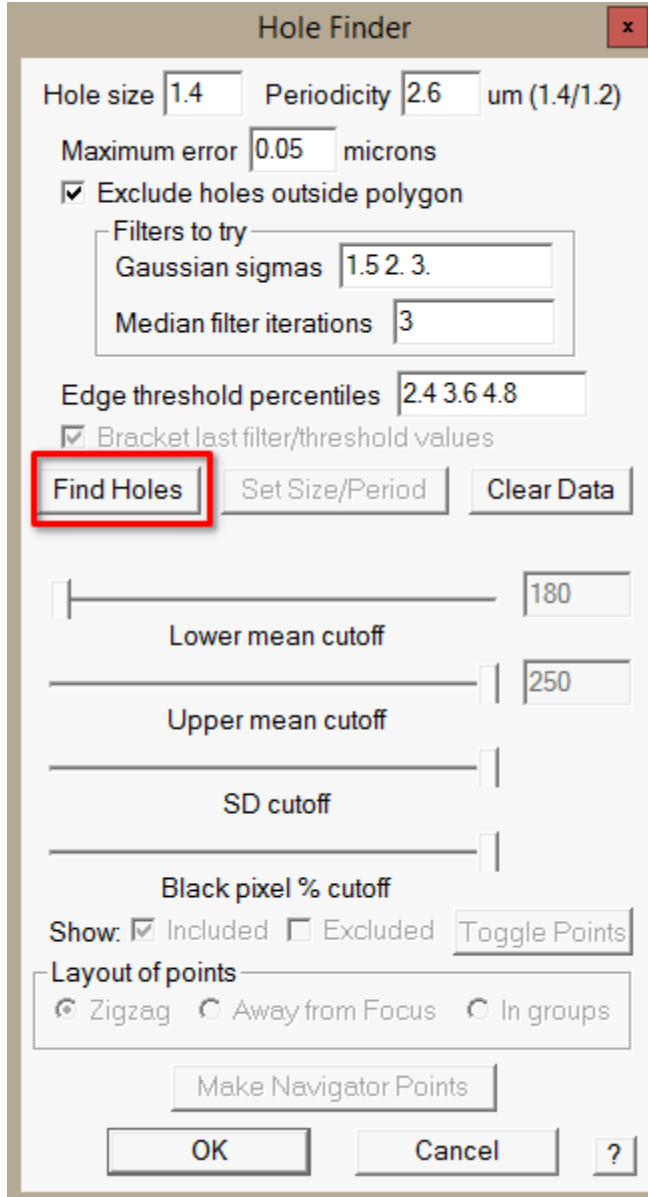


Figure 31. SerialEM "Hole Finder" dialog window and settings.

SerialEM's user interface will hang for a minute while it tries to identify the holes. Once the holes are identified, spend some time adjusting the sliders to select just the types of holes you would like to collect. These settings will carry over to further grids, so be sure to add some slight buffer space to the settings. When finished, select "OK" and close the dialog. Then, re-open the

top “Navigator” menu, select the “Montaging & Grids” sub-menu again, and this time choose “Set Multi-shot Parameters”

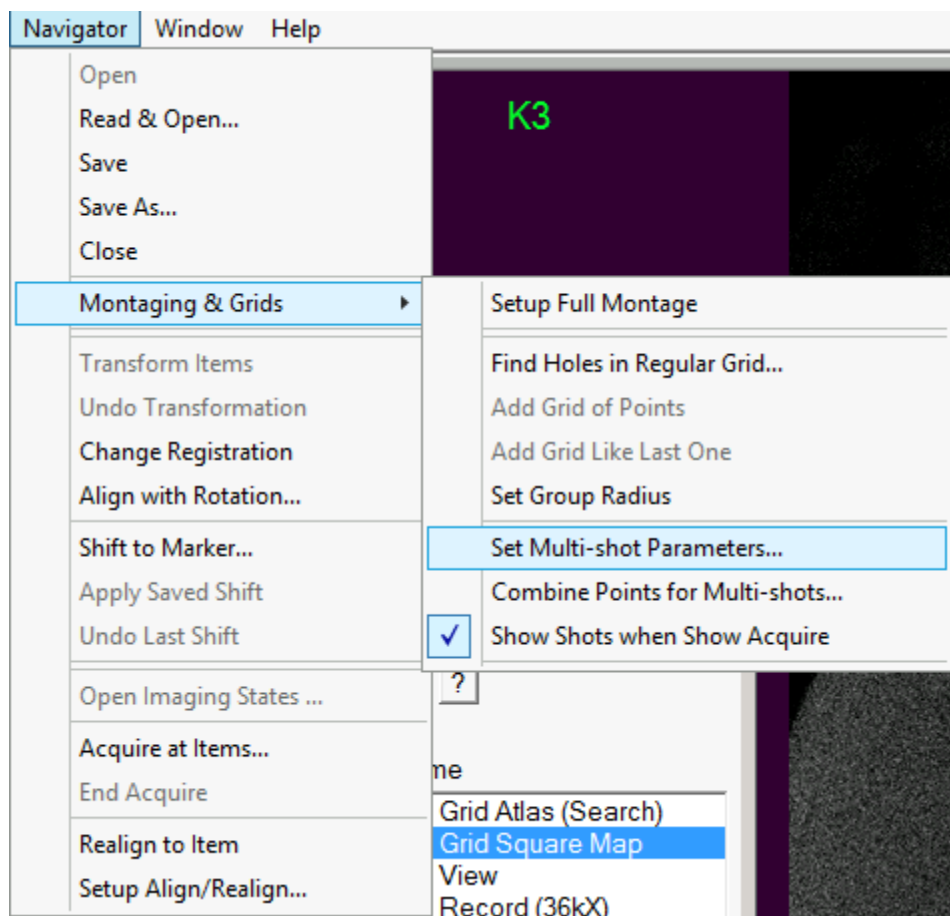


Figure 32. Location of "Set-Multi-shot Parameters" dialog in the "Montaging & Grids" sub-menu.

This will open the “Multiple Record Setup” dialog box. There are a number of options in this window that are of critical importance for data collection. First, ensure that “Do Records in multiple holes” is selected, and not “Do multiple Records in each hole”. This will configure SerialEM to collect a single image from each hole. Then, set the “Regular pattern of holes” to the desired array. The maximum beam-image-shift that the Arctica is capable of collecting is ~ 7.5

μm. For 1.2/1.3 grids, that corresponds to 5 by 5 pattern, but for 2/2 grids that corresponds to about 3x3 pattern. It may be desirable to omit the corners of a 3x3 when collecting for 2/2 grids. Once those options are configured, select “Use Last Hole Vectors”. SerialEM may prompt a warning – it is safe to ignore it. This will pull in the array that was calculated from the grid square map/find holes sequence. Then, ensure that “Save Record image” is disabled – instead, frames will be saved. Finally, set “Early Returns for K2/K3” to “Early return, except for full sum on first shot”. Typically, SerialEM will take a dose fractionated image of something like 30-50 frames for each hole. Then it will add the frames together, and display the average frame, before moving on to recording the next image. This slows down data collection, so selecting the “Early return, except for full sum on first shot” will instruct SerialEM to only calculate the average image for the first image of each multi-shot. This will speed up data collection, while still providing some feedback on the quality of the images as they are recorded.

Multiple Record Setup

☐ Do multiple Records within each hole

☒ Do Records in multiple holes

Multiplier of usual delay after shifting to hole: 1.5

Regular pattern of holes: 5 by 5

Spacing: 2.56 and 2.52 μm Angle to stage X: 31.4 deg

☐ Omit corners of 3 by 3 array (cross pattern)

☐ Use custom pattern (NONE DEFINED)

Save image shift values at holes

For Corners of Regular Pattern For Custom Pattern

Use Last Hole Vectors Step To and Adjust IS

IS to Nav Pt Save Image Shift End Pattern Abort

Image shift should be measured near focus

☐ Save Record image

Early Returns for K2/K3

☒ Full sum returned on all shots

☐ Early return only on last shot

☐ Early return on all shots

☐ Early return, except for full sum on first shot

Number of frames in return (-1 for all): 0

Extra delay after image shift: 0 sec

☒ Adjust beam tilt & astig to compensate for image shift

Coma versus image shift was calibrated with:
41.2% C2, spot 5, nanoprobe

Diameter of circles to draw: 1.3 μm

OK Cancel ?

Figure 33. SerialEM multi-shot parameters dialog window and settings. “Do records in multiple holes” will enable the multi-shot function of the microscope.

Finally, there is an option to adjust beam tilt and objective astigmatism to compensate for the beam-image-shift data collection. When using the BIS data collection features, the image

shift introduces beam tilt, because the beam-tilt pivot point calibration is never truly perfect. That beam tilt introduces coma at the various image shift positions. SerialEM can be calibrated to adjust for that coma at each image shift position, at a specific magnification and spot size. While the beam-tilt induced coma can be problematic for high resolution analysis, these factors can typically be compensated for in software, using either RELION particle polishing, or CryoSPARC's global CTF refinement, so collecting with this compensation enabled is not strictly necessary. However, if desired, the option is available during data collection. Typically, even with this compensation enabled, particle polishing/Global CTF refinement is needed regardless for highest resolution structures. The absolute resolution attainable does not seem to be impacted by whether the user chooses to enable this compensation or not.

Once the multi-shot parameters are set, re-load the high-magnification Record imaging state you are intending to use. Then, with the multi-shot window still open, left click somewhere on the open grid square map image. The multi-shot settings should be displayed – ensure that they match the hole configuration of the grid.

While SerialEM will have estimated stage coordinates for each identified hole in the grid square map images, the coordinates are not sufficiently accurate to move to the exact location of each hole. Therefore, before it records the multi-shot pattern, SerialEM will first take a View image, and center the hole using image shift. To do this, SerialEM must know what a hole *should* look like. This hole image can be generated directly in SerialEM. Take a View image, and center a carbon hole. This hole should be relatively clean of debris and uniform-looking.

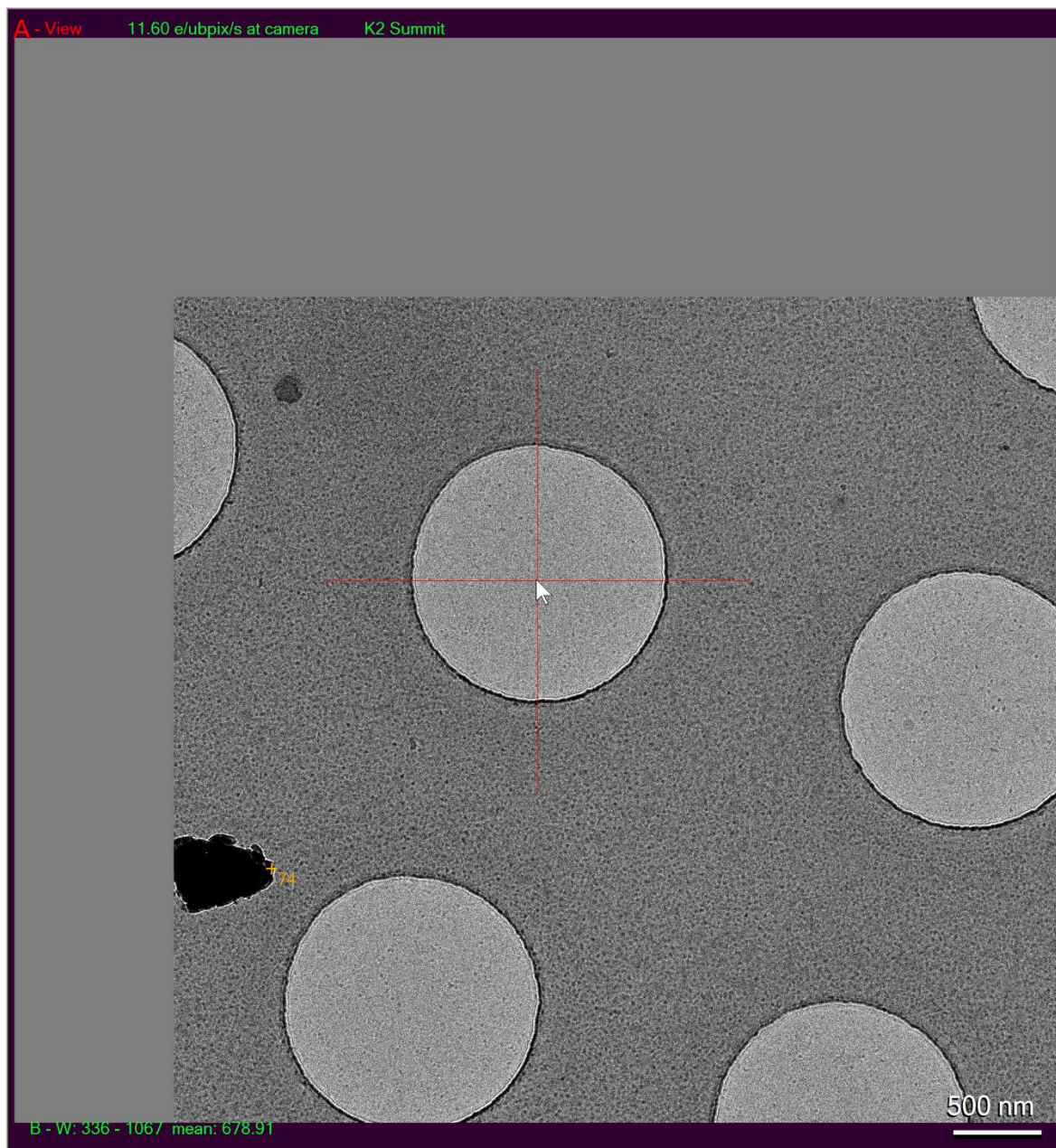


Figure 34. Medium magnification micrograph demonstrating proper hole centering for multishot setup. The red cross displays the current center of the microscope stage.

Then, crop out the surrounding area by holding down Ctrl + Shift and dragging the mouse to create a box around the desired hole. Once the box is drawn and suitable, in the top “Process” menu, select “Crop Image”. That will crop the image to the size of the drawn box.

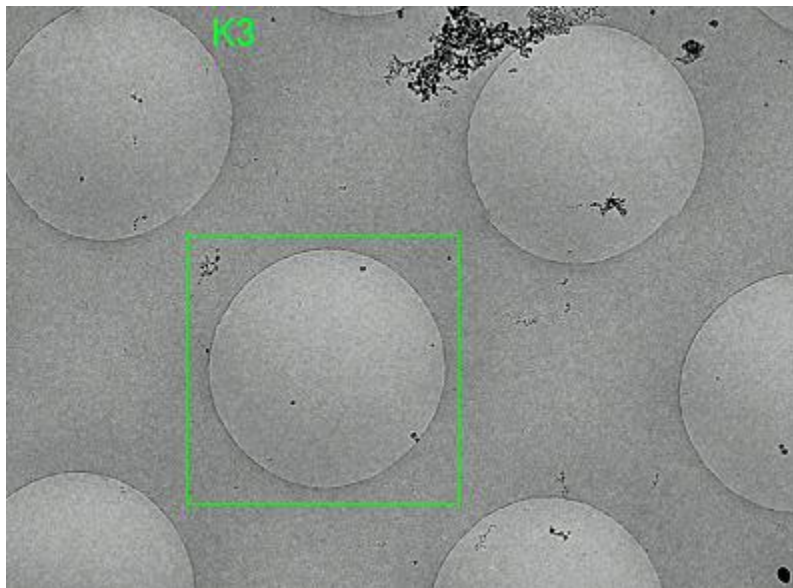


Figure 35. Medium-magnification micrograph depicting a green boxed hole to crop for template generation.

Then, open the “Navigator” top menu and select “Setup Align/Realign”. This will open the “Navigator Align Setup” dialog box. Give the template a name, and select “Make Map from Image in A”, and follow the prompts. The maximum alignment shift may require some empirical adjustments depending on the grid, to work properly. If so, the data collection can be paused and this value adjusted at any time.

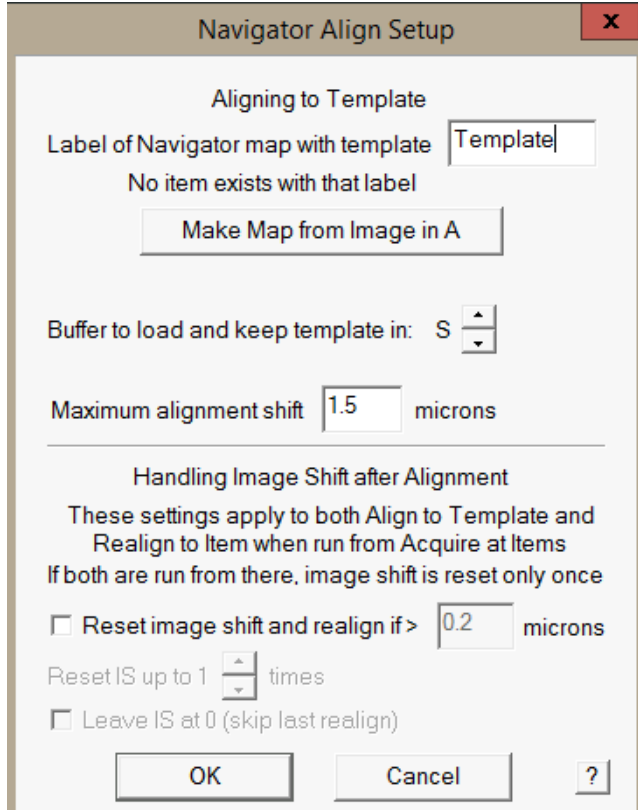


Figure 36. SerialEM template alignment settings dialog window.

The next step is to select all the desired grid squares to be collected. First, ensure that low dose mode is enabled by selecting the “Low Dose Mode” check box in the “Low Dose Control” panel. Then, load the “Grid Square Map” imaging state into the Record pre-set. Now, the new imaging state must be aligned to the grid atlas. There may be some X/Y translation when moving between the atlas magnification, and the grid square map magnification, that must be compensated for. In the grid atlas, find a grid square with an obvious feature, such as ice contamination, and place a marker on a spot you can see (green cross) by left clicking on it. Turn that marker into a point (orange cross) by selecting Add Marker in the Navigator window.

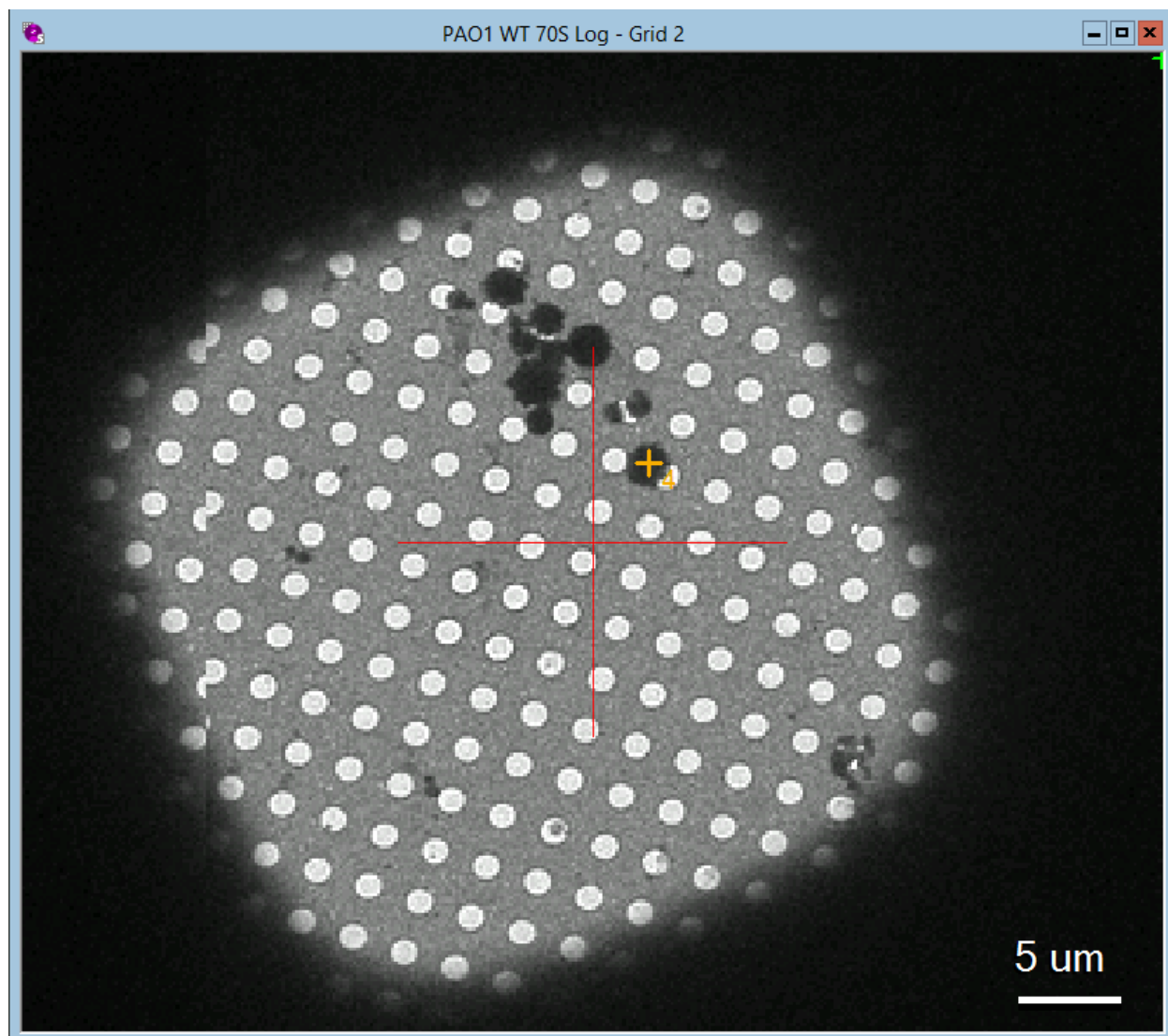


Figure 37. Grid square map magnification image. The yellow cross denotes the currently selected navigator point.

Navigate to the new point by selecting “Go To XYZ” with the point selected, in the Navigator window. Take a Record image at that point. Place a marker (green cross) on that feature in the image by left clicking the feature.

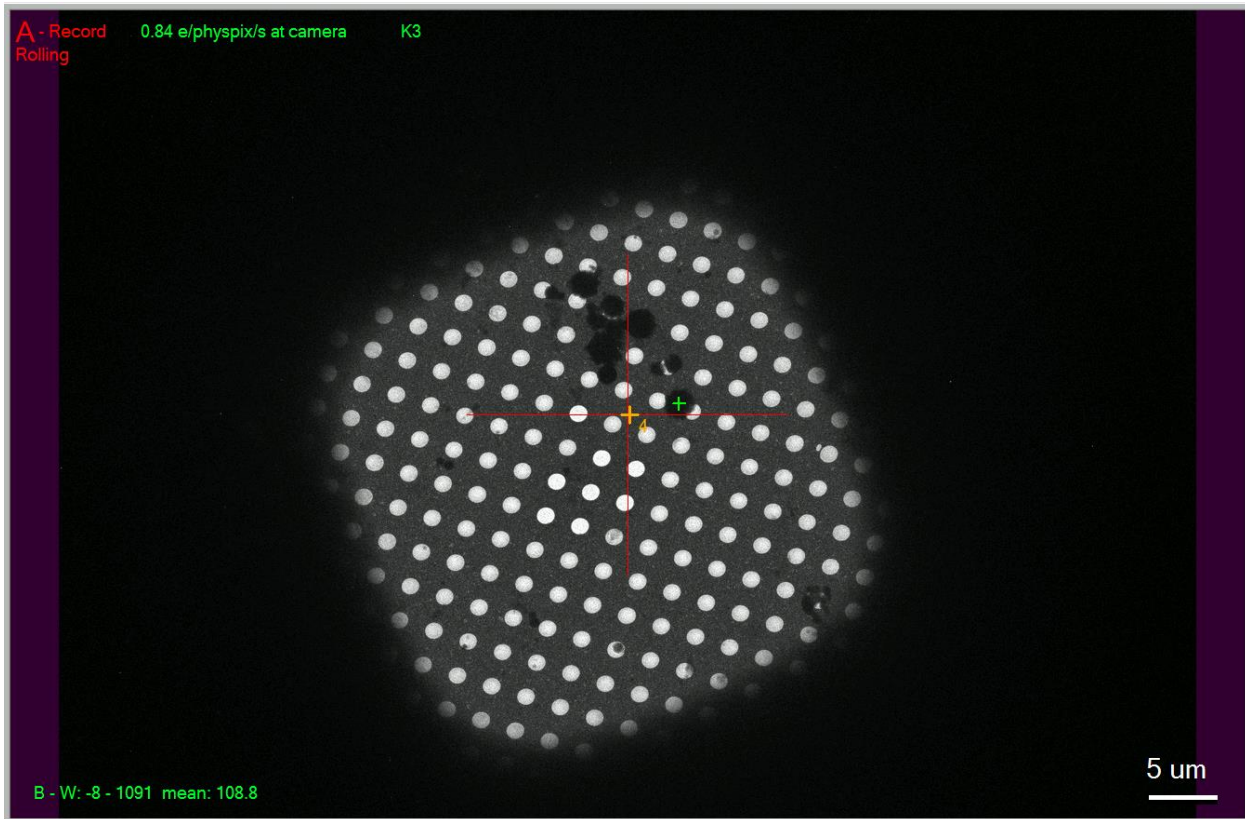


Figure 38. Aligning the grid square magnification to the grid atlas coordinates. The green cross represents the user's selection of the feature to align to, while the yellow cross represents where SerialEM believes the feature should be, based on the previously placed point.

Shift the point (orange cross) to the marker (green cross) by going to the Navigator menu, and selecting "Shift to Marker".

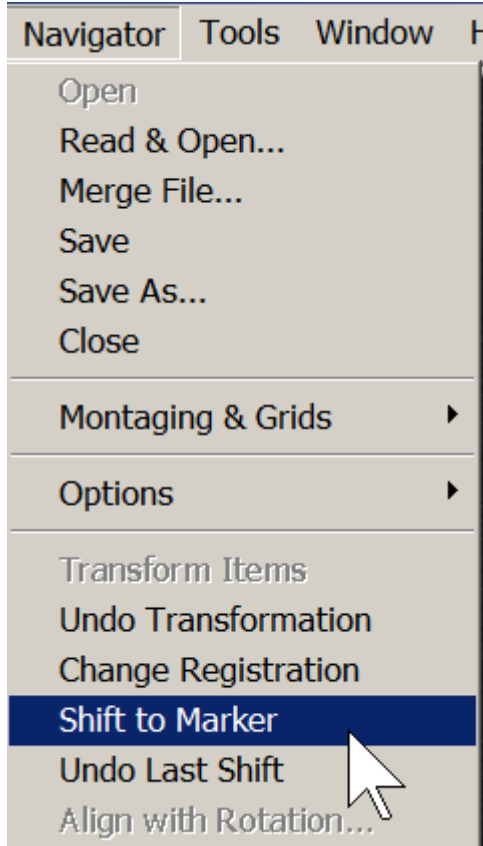


Figure 39. SerialEM "Shift to Marker" location in the "Navigator" menu.

This lets SerialEM know that the spot on the atlas at 46x translates to the point (green cross) that was placed in the higher magnification image. Now when you click somewhere in the atlas, it should go to the correct spot on the grid. You can now safely delete the marker from the navigator, if desired.

Once the grid square map imaging state is aligned properly with the grid atlas, the next step is to select the desired squares for data collection, record grid square maps of them, and find and group all the holes for multishot. This process can be automated. First, select the desired grid squares by Re-load the grid map in the Navigator window by double left-clicking it. In the Navigator, select Add Points and left click on the center of the desired grid squares. Once done

adding, mark added points for acquisition in the Navigator window by checking the “Collapse” and the “Acquire (A)” boxes. Then, check the “New file at group check box”, and click OK in the two following pop-up windows (the default settings are fine). Define a file name (e.g. grid_square_maps.mrc). If multiple groups are shown, the Navigator option “Add Points” was de-selected before point adding. In this case, set the “Acquire (A)” and “New file at group” flags for each individual group. File names will be incremented automatically.

Next, acquire the eucentric grid square maps. Open the Navigator top menu and select “Acquire At Items”. A new window should open up. At the top left, select “Mapping”. Select “Primary Action: Acquire and save image or montage” option. Under Task-related options select "Relax stage after final movement". Under *General options*, select "Close column valves at end. In the column on the right, Tasks before or after Primary, select "Rough Eucentricity" and nothing else. Ensure that for “General controls for task: Rough Eucentricity” are set to determine eucentric height for every grid square. Ensure that the “Hole Finder & Combiner” is below “Primary Action” and enabled. Select “Setup” and ensure that the hole finder and combiner are properly setup. The hole finder will use the settings previously set. The combiner will combine them according to the specified multishot parameters.

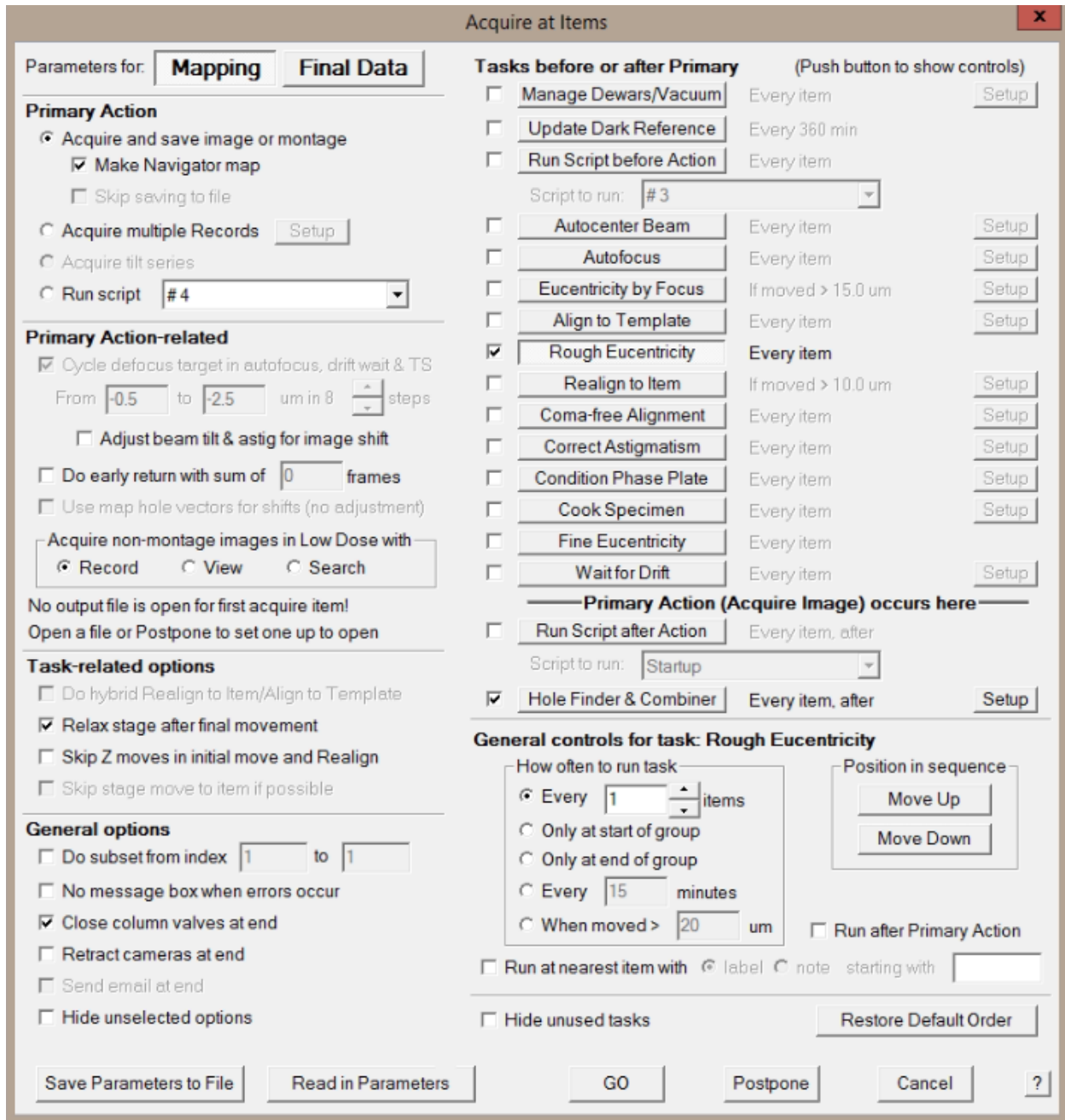


Figure 40. SerialEM "Acquire at Items" dialog window, depicting settings for grid square maps and hole finding.

Once everything is setup, select "Go". This will navigate to each requested X/Y position, determine and save eucentric height, and then take a low magnification image of each square.

Once the low-mag image is recorded, it will identify holes, and combine them into the multi-shot pattern, adding them as navigator points. This process can take several hours, depending on the number of grid squares selected.

Once all the grid square maps are recorded, the next step is to configure the camera and exposure settings. The Gatan K3 camera is a complex instrument, and ideal settings can be difficult to ascertain. This section will provide some reasonable defaults, and ways to make required alterations. Typically, for SPA, the camera should be operated in counting mode, which will provide a higher detector quantum efficiency than operating in linear mode. In counting mode, the optimal dose rate is 15 electrons/pixel/second. To get as close as possible to that dose rate at any given magnification on a 2-condenser lens microscope, the spot size must be adjusted. Do not adjust C2 intensity, as that must be set to a specific value at each spot size to preserve parallel illumination. In general, it is better to have a dose rate below 15 electrons/pixel/second than above, to avoid coincidence loss. Therefore, if 15 cannot be achieved, aim for the closest, but lower value.

First, decide which magnification to select. There are a number of tradeoffs involved at this step. In general, DQE increases as a function of spatial frequency, which means that individual particles recorded at higher magnification contain more signal. However, field-of-view (FoV) decreases as a function of magnification, which means that fewer particles are available to be averaged together. While lower magnification will contain more particles, there is a threshold of spectral signal that must be met before signal is considered additive. In general, a pixel size between 0.5 Å and 1.5 Å is a reasonable target – exactly which magnification will depend on the sample and desired resolution. The dose rate in pixels is defined by the detector

peak DQE. However, the dose rate in $\text{\AA}^2$ will change depending on the magnification. In general, higher magnifications require lower spot sizes to reach the 15 electrons/pixel dose rate, which will translate to a higher dose rate in terms of electrons/ $\text{\AA}^2$, which will translate to shorter exposure times, and also shorter frame times. Reasonable starting points have been provided. (Table 1).

Table 1. Recommended starting configuration for SPA data collection, magnifications 8,500x-92,000x. Spot size is the recommended spot size for each magnification to achieve optimal camera dose at parallel illumination. C2 illumination settings available in Appendix A.

Mag (nominal)	Pixel Size ($\text{\AA}$)	SR Pixel Size ($\text{\AA}$)	Spot Size	Flux at Camera (e/p/s)	Frame Time (s)	Exposure Time (s) ($\sim 50 \text{ e}/\text{\AA}^2$)	C2 (μm)	Notes
8500	4.934	2.467	9	14.87	0.5	85	150	170 frames - use with caution!
11000	3.840	1.920	9	9.9	0.5	85	150	170 frames - use with caution!
13500	3.033	1.517	8	16.02	0.25	30	70	120 frames - use with caution!
17500	2.397	1.199	7	18.62	0.25	15	70	
22000	1.852	0.926	7	11.24	0.25	15	50	
28000	1.454	0.727	6	13.37	0.1	10	50	
36000	1.104	0.552	5	13.69	0.1	6	50	Typical SPA
45000	0.9061	0.453	4	15.11	0.08	3	50	Typical SPA
57000	0.687	0.343	3	14.54	0.08	2	50	Typical SPA
73000	0.5637	0.282	2	14.76	0.08	1.5	50	
92000	0.4483	0.225	2	9.46	0.1	1.5	50	

One additional consideration with the camera is the resolution. The K3 camera provides a super resolution mode. This mode determines a more precise location for each electron by using neighboring pixels to triangulate the quadrant of the pixel where the primary electron impact occurred. This effectively quadruples the total area of the sensor. However, super resolution works better at lower magnifications. At magnifications finer than 1 Å pixel size, it is not generally recommended to use super resolution mode because of the larger file size and slower data collection.

To set the camera settings, it is recommended to first select the desired imaging magnification. Then, navigate to an empty area of the grid (i.e. a hole), and record an image to measure the dose.

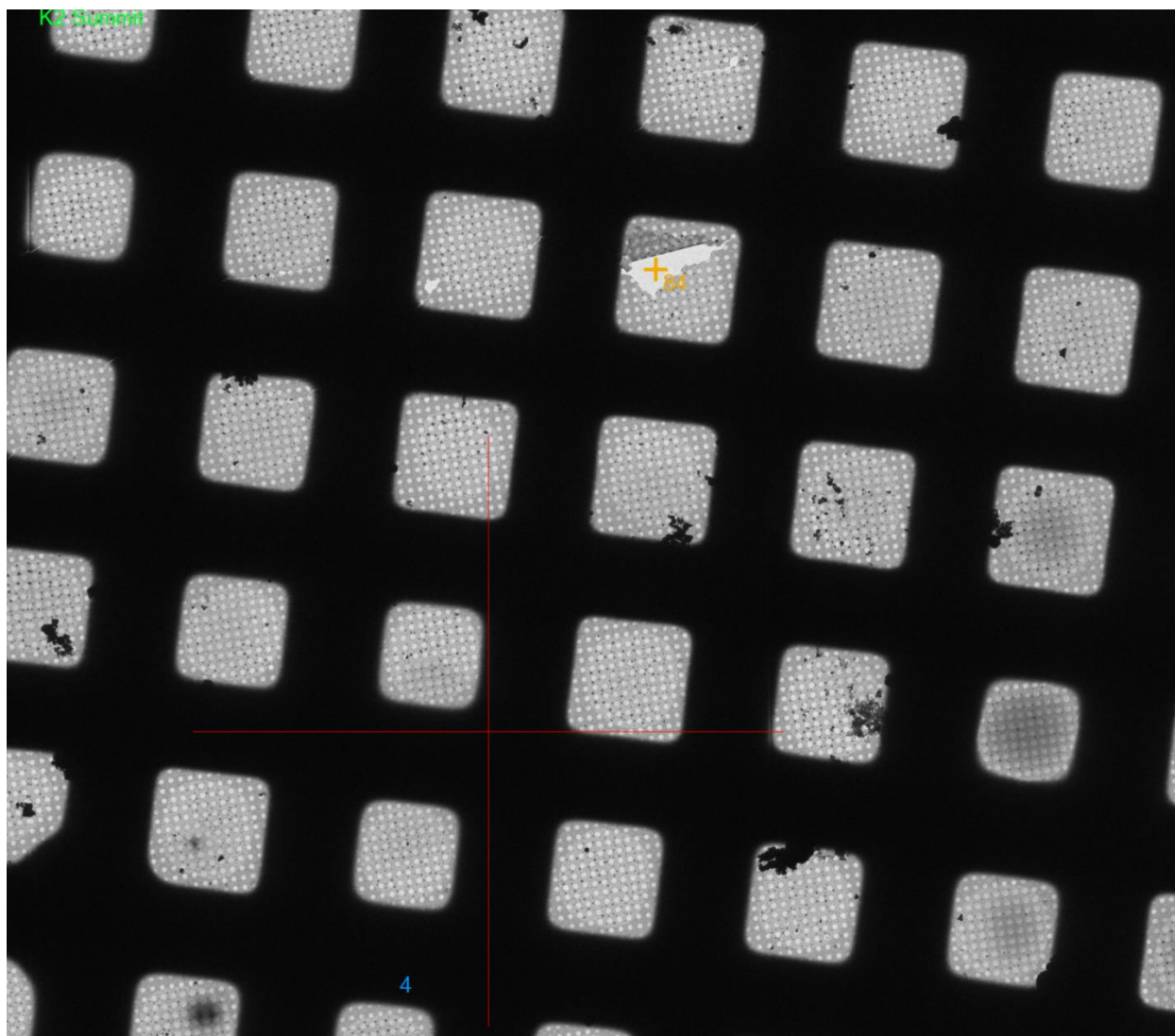


Figure 41. Low magnification image of an EM grid depicting a grid square with a hole sufficient for dose calculation.

The Record image should not contain any features, it should be entirely gray. To calculate the dosage, select the top “Tasks” menu, and select “Calibrate Electron Dose” and follow the prompts.

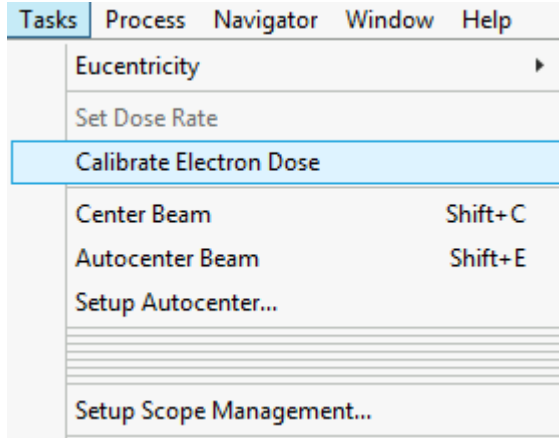


Figure 42. Location of the "Calibrate Electron Dose" menu option in the "Tasks" menu.

SerialEM will then report the total dose at the sample. When imaging, expect the dose on the camera to be slightly lower than the dose at the sample. To make fine adjustments, in the top menu, select “Camera” and “Parameters” – this will open the Camera settings menu. Ensure you have selected the “Record” settings, and then make sure that both “Save Frames” and “Dose Fractionation” are enabled. Set the total exposure time to provide sufficient dose (~ 50 electrons/ $\text{\AA}^2$) on the sample and set the dose fractionation time so that each frame will have approximately 0.8 electrons/pixel.

Camera Parameters -- K3

Camera -- Selects active camera
☐ Ceta ☒ K3
☐ Match pixel size ☐ Match intensity when changing camera

Parameter set
☐ Search ☐ View ☐ Focus ☐ Trial ☒ Record ☐ Preview ☐ Mont-map

Parameters for **Record** Copy from: Ceta

Acquisition
☐ Continuous ☒ Single Image

Processing
☒ Unnormalized ☐ Gain Normalized
Sum will be gain-normalized

Binning
☐ 0.5 ☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 8

Positioning
Top 0
Left 0
Bottom 4092
Right 5760
Recenter
Swap X & Y

Exposure time 2.956 sec
Drift settling 0.0000 sec
Minimum 0.0007 if not 0.0

Shutter mode
☒ Beam blanking only

Binned size: 5760 x 4092
0.52 x 0.37 um @ 0.091 nm

Area size
Quarter Half Full
Wide Quarter Wide Half
10% Less 10% More Square

Dose: 55.39 e/A2 at spot 4, C2 42.12%
1.50 e/A2 per frame Update 15.39 e/phys pixel/s at specimen

Operating mode
☐ Linear ☒ Counting

☒ Dose Fractionation mode
Frame time 0.08 sec
☐ Align frames Set Up

☒ Save frames
37 raw to TIF-LZW, binned
Set File Options
Set Folder

☐ Use correlated double sampling (CDS)
☒ Bin counting frames by 2

Anti-aliasing
☐ Save variable frame sums Set Up

OK Acquire Cancel ?

Figure 43. Camera parameters for the "Record" setting for data collection in SerialEM.

At this point, the file saving should be setup. There are multiple considerations when setting up the file options. First, it is recommended to disable gain normalization. Typically,

counting frame images contain 2-3 bits of information per pixel. This means that when applying compression, 5/8ths of each pixel is easily compressible. However, after gain normalization, the image is rescaled to use all 8 bits of information. This results in an image that is less easily compressed. Therefore, it is recommended to apply the gain normalization on-the-fly when writing out each motion corrected micrograph, not to each individual frame. Furthermore, the file format is an important consideration for large datasets. MRC does not support native compression, so it is instead recommended to save images as TIF stacks using LZW compression. Finally, consider binning. The K3 camera always records in super-resolution mode, but will bin the images by two when selecting the “1” binning option. By default, SerialEM will handle the binning in real-space. This can result in aliasing of high resolution information into low resolution information, and is also slow. Therefore, it is recommended to select the “Bin counting frames by 2” option, which will bin the frames directly in hardware by first cropping them in Fourier space, which also will mitigate aliasing effects. Finally, set the file name options.

Frame File Options [X]

File type

☐ MRC file ☐ Use extension .mrcs
☒ TIFF (LZW compression)
☐ TIFF (ZIP compression)

Current settings will save norm to TIF-LZW stack

☐ Save one frame per file
☐ Save mdoc file for each frame stack

☐ Save unnormalized frames even if Gain Normalized is selected
☐ Reduce normalized super-resolution frames by 2 with antialiasing

☐ Pack unnormalized data as 4-bit
 ☐ Pack unnormalized binned by 2 data as 4-bit, not 8-bit
 ☐ Use non-standard mode 101 for 4-bit MRC files
☐ Save frames without rotation/flip to standard orientation

Components for Folder and File Names

Use component in name of:

Folder	File
☐	☒ Base name: mysample
☐	☐ File name of current open file (minus extension)
☐	☐ Label of Navigator item ☒ Only when Acquiring at Items
	☒ Sequential number starting from: 1 Digits: 5
	☐ Tilt angle
	☐ Month and day
	☐ Hour, minute, second
	☐ Numeric date (including year) and time at start of name
	☒ Multiple Record hole number and position in hole

Format: Base_#####_Hole-pos

Figure 44. Recommended file-saving options for data collection.

It is recommended to include the “Multiple Record hole number and position in hole” in the file name, so that SerialEM will append each frame with its X/Y coordinate in the multi-shot

pattern. This will allow later separation and refinement of global CTF parameters/particle polishing by grouping all exposures at each specific multishot position. Below is a table of the exposure group X/Y coordinate positions. (Table)

Table 2. SerialEM X/Y coordinate nomenclature. This nomenclature is appended to the end of the file name when saving multishot positions in the file name.

X-2, Y+2	X-1, Y+2	X+0, Y+2	X+1, Y+2	X+2, Y+2
X-2, Y+1	X-1, Y+1	X+0, Y+1	X+1, Y+1	X+2, Y+1
X-2, Y+0	X-1, Y+0	X+0, Y+0	X+1, Y+0	X+2, Y+0
X-2, Y-1	X-1, Y-1	X+0, Y-1	X+1, Y-1	X+2, Y-1
X-2, Y-2	X-1, Y-2	X+0, Y-2	X+1, Y-2	X+2, Y-2

First, take a View image and ensure that the Focus position is properly set. Then open the top “Navigator” menu and select “Acquire at Items”. Ensure that “Final Data” mode is selected, **not** “Mapping”. The “Primary Action” should be “Acquire multiple Records”, and “Primary Action-related” should indicate “NO SAVING”. This is because the frames will be saved instead of the summed micrographs. This is also a good time to make sure your Camera Parameters for Record do NOT have the "Align Frames" option checked.

The exact sequence of events prior to the “Primary Action” is important. First, check vacuum management, then hole should be auto-centered. Once auto-centered, the beam should be auto-centered on the carbon. Then, focus should be done, and finally, the multiple record. Doing this out of order can result in runaway focus or missed images. Once those options are configured, ensure that defocus is set to cycle properly.

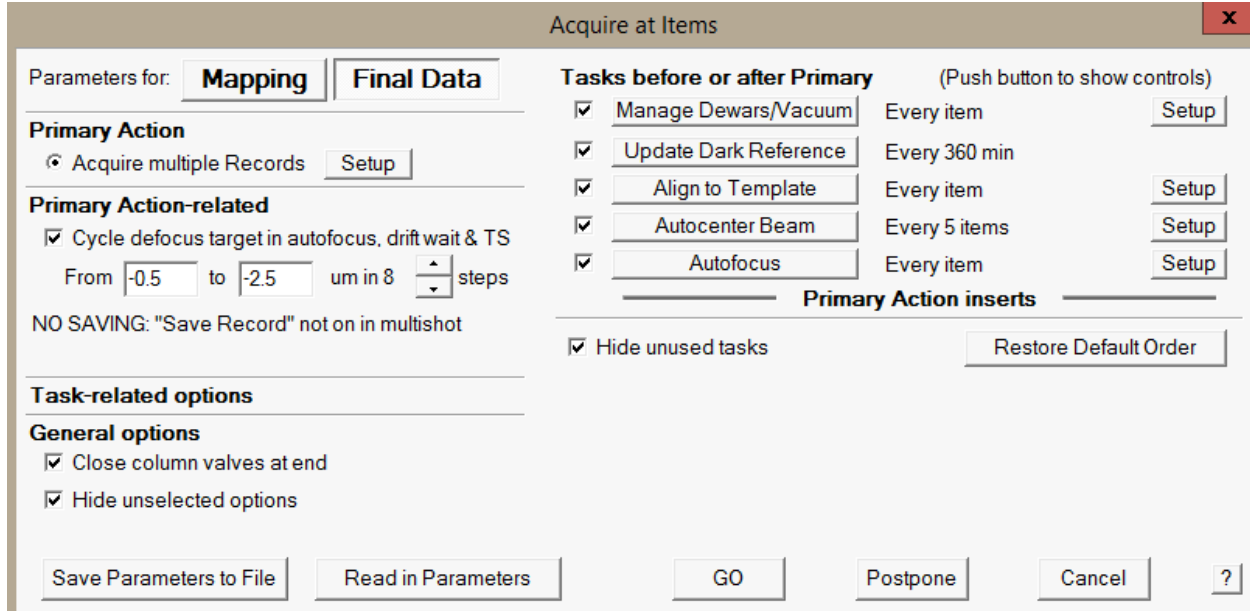


Figure 45. SerialEM "Acquire at Items" dialog box showing the proper settings for final, high-throughput data collection.

To run the acquisition, select “Go”. It is a good idea to monitor several things at the beginning of the data collection. Check that it is aligning to the template properly. Watch a few align to templates, and make sure it's not creeping back to the best match instead of finding the proper hole closest to the navigator point. Make sure the autofocusing is actually working and the targets are right. Make sure the multiple record is finding the center of each hole as it goes, instead of clipping edges. Make sure files are being saved properly, where you set to save them. If there are issues with any of these things, you can stop the navigation acquire and fix them, and restart it.

SmartScope

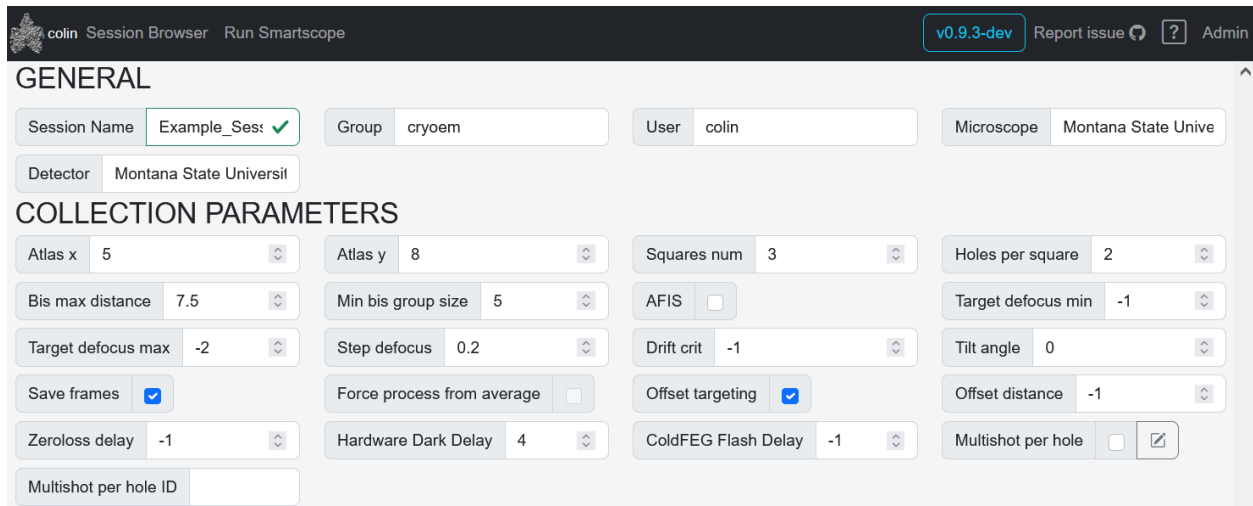
Before running SmartScope, SerialEM requires a couple of small configuration changes. To begin, close any open instances of SerialEM and open the “SmartScope” preset in Zuul (53).

Set the “Working Directory” to “X:\SmartScope”. Select the desired magnification from the listed magnifications in the “Imaging States” window and load it into the Record low dose setting. Open the “Camera” top menu and select “Parameters”. Copy the exposure time and frame time from the “Record” setting to the “Preview” setting. This is necessary because SmartScope uses the “Preview” setting for acquiring high magnification images, due to SerialEM limitations. No file saving settings need to be configured – SmartScope will handle those automatically. Finally, in the “Low Dose Control” panel, press the + box to show advanced options. Load the “Record” area settings and copy them to “F”. This will place the Record settings into Focus, which will speed up data collection by minimizing the magnification changes.

Screening. To begin screening, login to the SmartScope interface, and select “Run Smartscope” from the top menu. Give the session a unique name, although beware of length, if the name is too long it may crash. Select the necessary group and user, as well as the microscope, and camera.

Atlas magnification in SmartScope is 62x, as opposed to the normal 48x. Typically, a 5x8 atlas will work well. Set “Squares num” to the number of squares to be screened. Set “Holes per square” to the number of multishot groups to record on each screened square. Set “Bis max distance” (Beam Image Shift max distance) to 7.5 (μm) for 5x5 on 1.2/1.3 grids, or to a lower number for smaller multishot groupings. Set “Min bis group size” to the minimum number of holes that SmartScope should require the group to have to be considered worth collecting. This can save time when doing large collections. Enable “AFIS” (Aberration-Free Image Shift) to use SerialEM’s “Image shift vs coma” calibration, if desired. Note that this will require the

calibration to be done in SerialEM. Set the “Target defocus min/max” to the minimum and maximum desired defocus, and the step size to the desired step. “Drift crit” and “Tilt angle” should be left at default, unless a tilt is desired. Leave “Save frames” as enabled and ignore “Force process from average”. If particle location inside the holes is unknown, consider enabling “Offset targeting”, and setting the “Offset distance” to -1. This will cause SmartScope to add incremental shifts of 0.1 μm - 0.5 μm to the high-mag images, in the event that particles prefer thicker ice or the carbon hole edge. Ignore “Zero loss delay”, set “Hardware Dark Delay” to something like 4 or 6 – this is the hours between new dark references. Ignore “ColdFEG Flash Delay” and “Multishot per hole”.



The screenshot shows the SmartScope configuration interface. At the top, there is a header bar with the user name 'colin', 'Session Browser', 'Run Smartscope', and a version indicator 'v0.9.3-dev'. Below the header, the 'GENERAL' section contains fields for 'Session Name' (Example_Ses), 'Group' (cryoem), 'User' (colin), and 'Microscope' (Montana State Unive). The 'COLLECTION PARAMETERS' section contains various settings for data collection, including 'Atlas x' (5), 'Atlas y' (8), 'Squares num' (3), 'Holes per square' (2), 'Bis max distance' (7.5), 'Min bis group size' (5), 'AFIS' (unchecked), 'Target defocus min' (-1), 'Target defocus max' (-2), 'Step defocus' (0.2), 'Drift crit' (-1), 'Tilt angle' (0), 'Save frames' (checked), 'Force process from average' (unchecked), 'Offset targeting' (checked), 'Offset distance' (-1), 'Zero loss delay' (-1), 'Hardware Dark Delay' (4), 'ColdFEG Flash Delay' (-1), and 'Multishot per hole' (unchecked).

Figure 46. SmartScope recommended screening configuration.

Once setup, enter the grid information. Grids should be given unique names, and their parameters such as their location in the autoloader, grid type, spacing, and material, should be entered accurately. SmartScope has a number of protocols that are available to be selected for each grid. There are two major differences: SPA protocols work similarly to SerialEM and need a

template supplied for hole centering. They also support collection with stage tilt. The Ptolemy protocols use the Ptolemy image recognition library that has been trained on cryo-EM data to center holes. The “precise” variant of each uses intermediate magnification for rough eucentric estimations, while the non-precise variants use the faster, but less accurate low magnification for eucentric height estimations. If the user selects “auto”, SmartScope will use Ptolemy non-precise for screening, and Ptolemy precise for data collection mode. Auto is therefore the recommended protocol selection. Once the grids are setup, select “Submit”.

AUTOLOADER

Grid_1 ✓

✕

Position 1

Hole Type R1.2/1.3

Mesh Size 200

Mesh Material Carbon

Protocol auto

Grid_2 ✓

✕

Position 2

Hole Type R1.2/1.3

Mesh Size 200

Mesh Material Carbon

Protocol auto

+

Submit

Figure 47. SmartScope grid and autoloader options.

Upon clicking “Submit” the user will be redirected to the “Session Controls” page. Selecting “Start” will begin the session. In SerialEM, the status should display “External Control”. SmartScope will montage the grid, select grid squares to image, record high magnification images, and move on to the next grid to repeat the process with no user input required. Along the way, SmartScope collects a wealth of data, annotating cracked or small or contaminated grids, as well as estimating CTF fits and ice thickness.

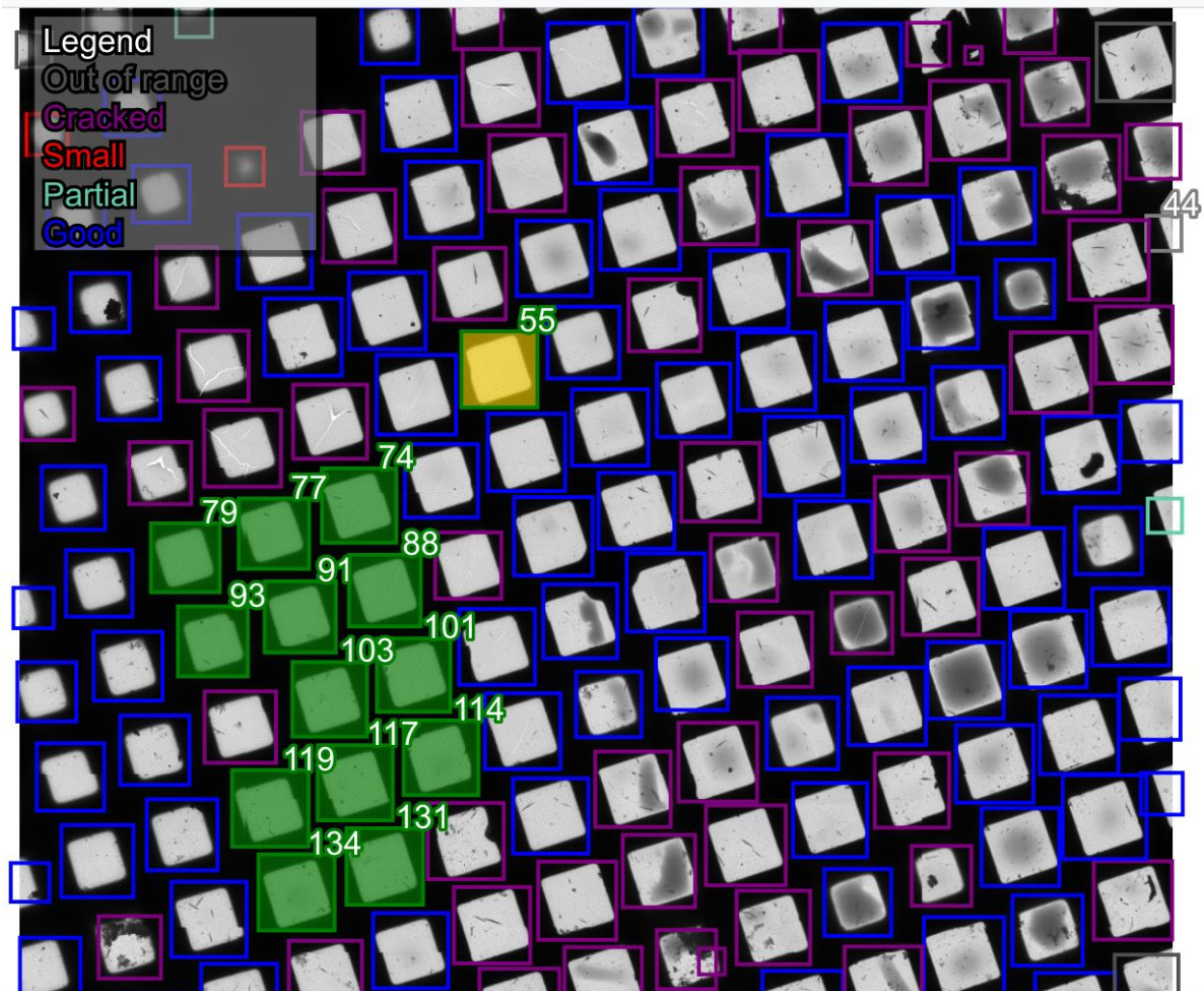


Figure 48. SmartScope annotated grid atlas depicting collected grid squares in green, good grid squares in blue, and cracked grid squares in purple.

In the atlas, users can select squares that have been imaged, indicated by their green border. Once the grid square is selected, SmartScope will display the square and all the identified and imaged holes. Blue indicates available holes to be imaged, while green indicates holes that have been imaged and had pre-processing completed.

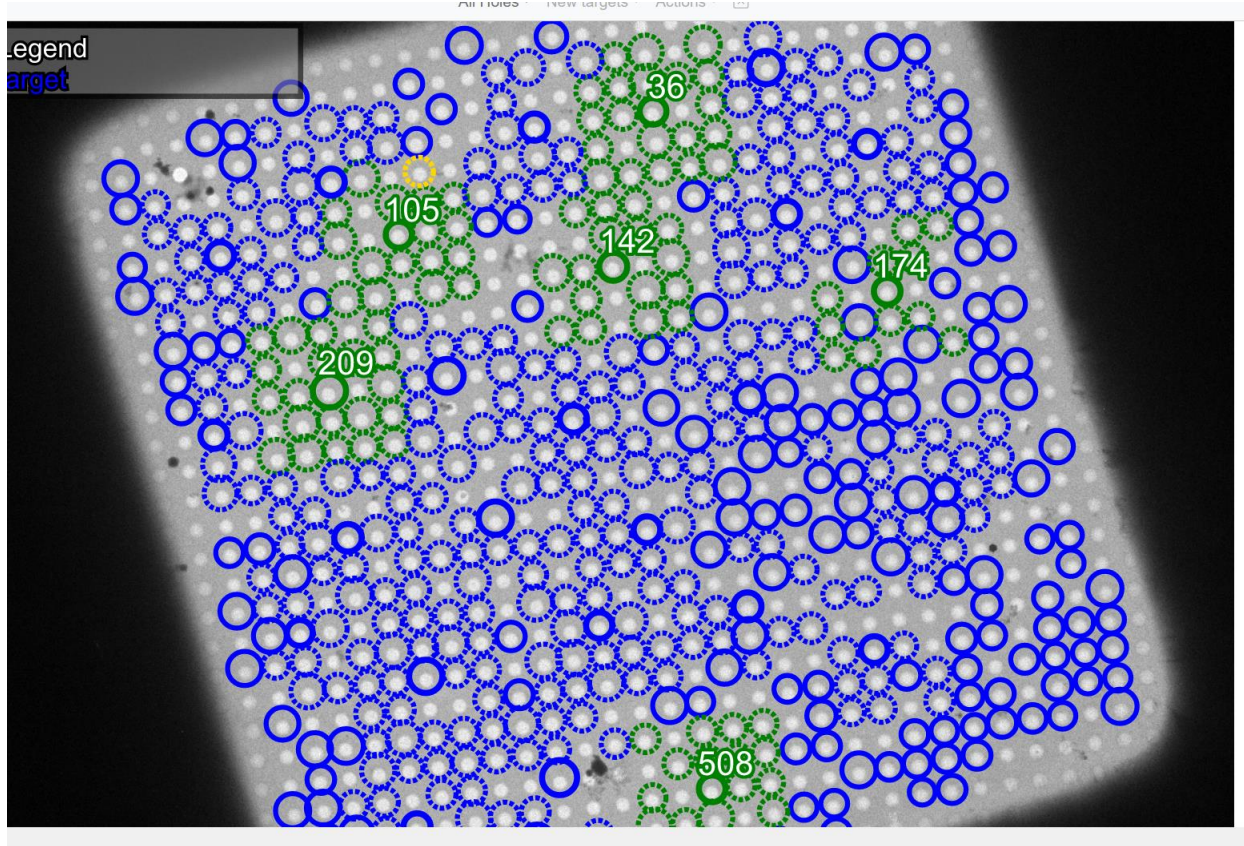


Figure 49. SmartScope annotated grid-square map depicting collected holes in green, and identified holes in blue.

Users can select additional holes or squares to image by left clicking, and then under the “Actions” menu, selecting “Add to queue”. Left clicking a green hole will show the high magnification image.

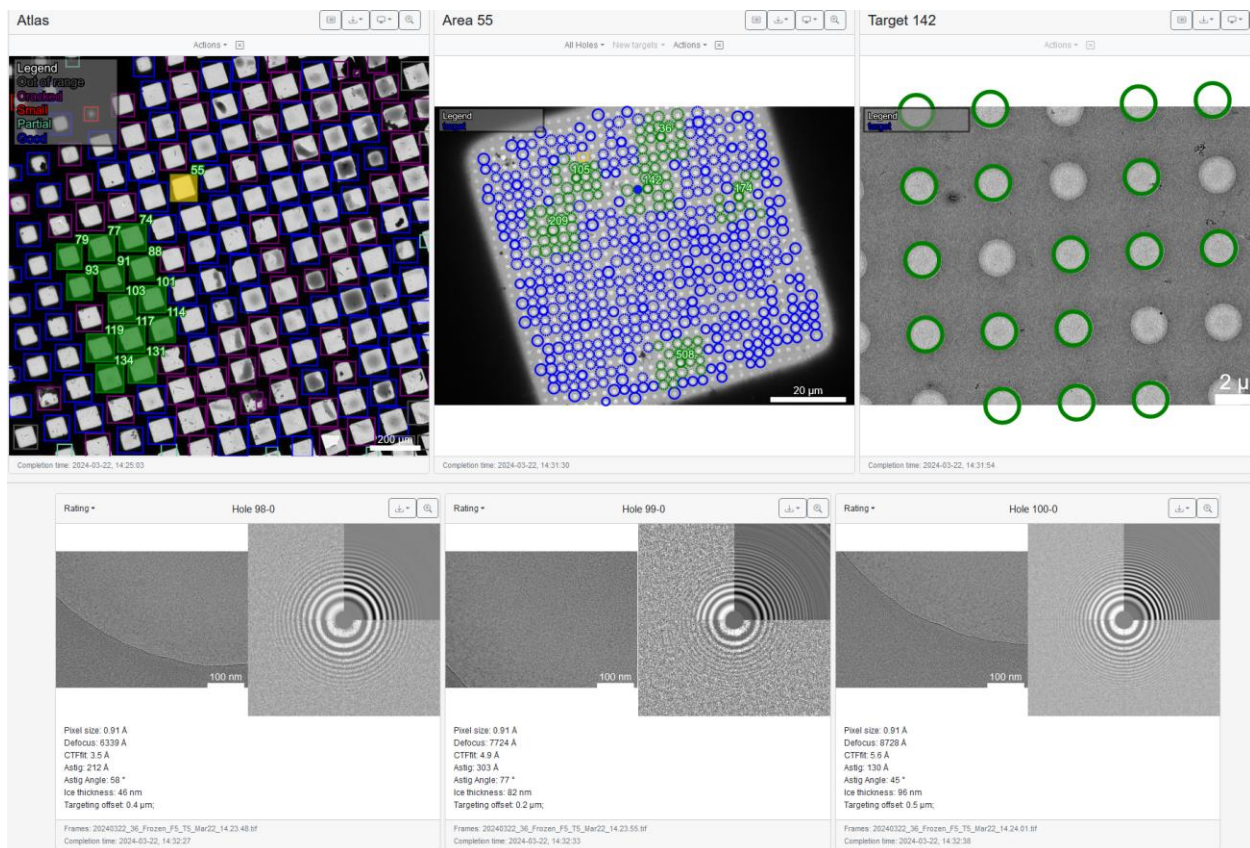


Figure 50. SmartScope user interface overview, depicting annotated grid atlas, annotated grid square map, annotated medium magnification image, and associated high-magnification records and CTF estimations.

Data Collection. Extended data collection in SmartScope is very similar to screening, with just a few important differences. Once SmartScope finishes screening a grid, it will set that grid status to “Completed”. The grid must be restarted by selecting “restart” on the grid page.

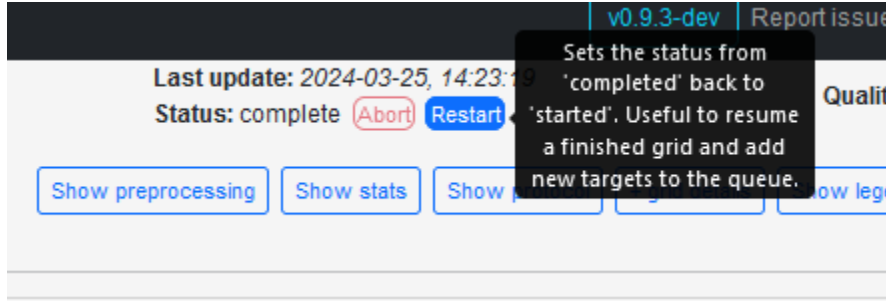


Figure 51. SmartScope grid restart protocol.

Once the grid is restarted, select grid squares in the atlas and add them to the queue. It is not recommended to add holes to squares that have already been partly imaged, as the coordinates may mismatch when the grid is re-loaded. Instead, it is recommended to add entirely new squares. Then, edit the grid details. Ignore the “Squares num” parameter but set “Holes per square” to 0. That will enable data collection mode for SmartScope, and collect all reasonable holes in the square. If no offset targeting is desired, be sure to uncheck the box. If a static offset target is desired, leave the box checked, but set the offset target distance to the desired number. Select “Save” when finished.

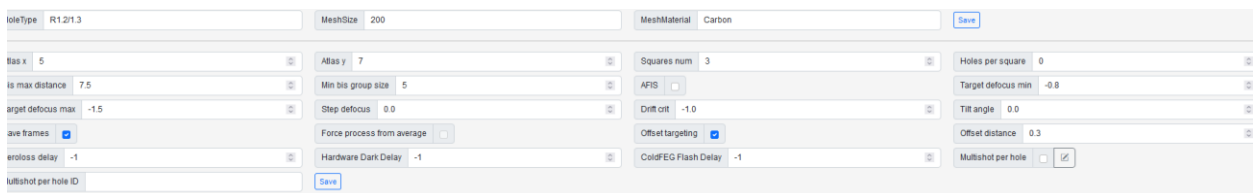


Figure 52. SmartScope settings to enable data collection mode.

Once properly configured, it is recommended to disable the built-in preprocessing, as this will generate quite a bit of redundant data. Finally, go to “Session Controls” and start the session. The grid will be re-loaded if it is not already on the stage, and SmartScope will begin collecting all the grid square maps, identifying holes, and recording high magnification images. In data

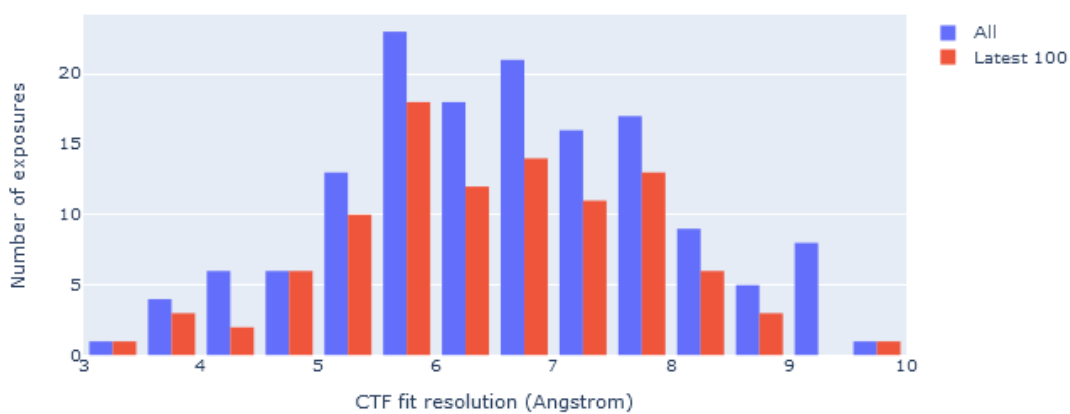
collection mode, SmartScope first collects all the grid square maps before beginning high mag imaging. Data collection statistics and rates can be viewed by clicking on the “Show stats” button for the grid:

Collection statistics



Holes in queue:	0
Total exposures in queue:	0
Holes acquired:	148
Holes per hour:	236.0
Holes in the last hour:	148
Total time spent:	0:37:37
Estimated time remaining:	0:00:00

CTF fit distribution



Ice thickness distribution

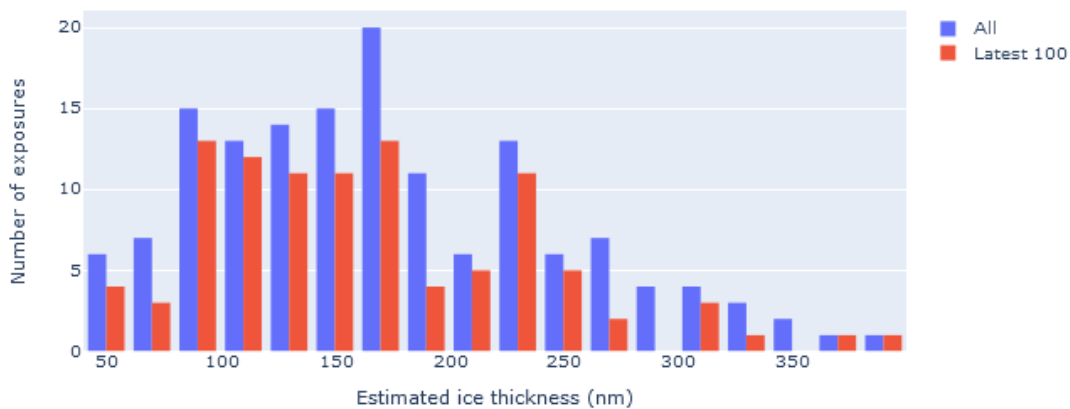


Figure 53. SmartScope collection statistics displaying both CTF estimations and ice thickness estimations from CTFFind 5, total (blue) and over the last 100 micrographs (red).

By default, SmartScope will save movies to “X:\SmartScope\movies”, and are arranged in directories by session, and grid name. It is still recommended to move the movies via Globus, similar to SerialEM SPA data collection.

Tomography Protocols

This section will detail current best practices for recording tomograms with SerialEM’s built in batch tilt-series functionality, or the more advanced PACETomo that features multi-shot tilt-series collection. To collect just a single tomogram, it is recommended to skip the “Selecting Targets” section and jump straight to “Recording Tilt-Series”.

SerialEM Batch Mode

Selecting Targets. The first step for selecting a large number of targets for tilt-series collection is to create an intermediate (View mode) magnification map. This can be done easily by drawing a polygon on the grid atlas, around the area of interest. To do this, open a grid atlas and select “Add Polygon” in the Navigator window.

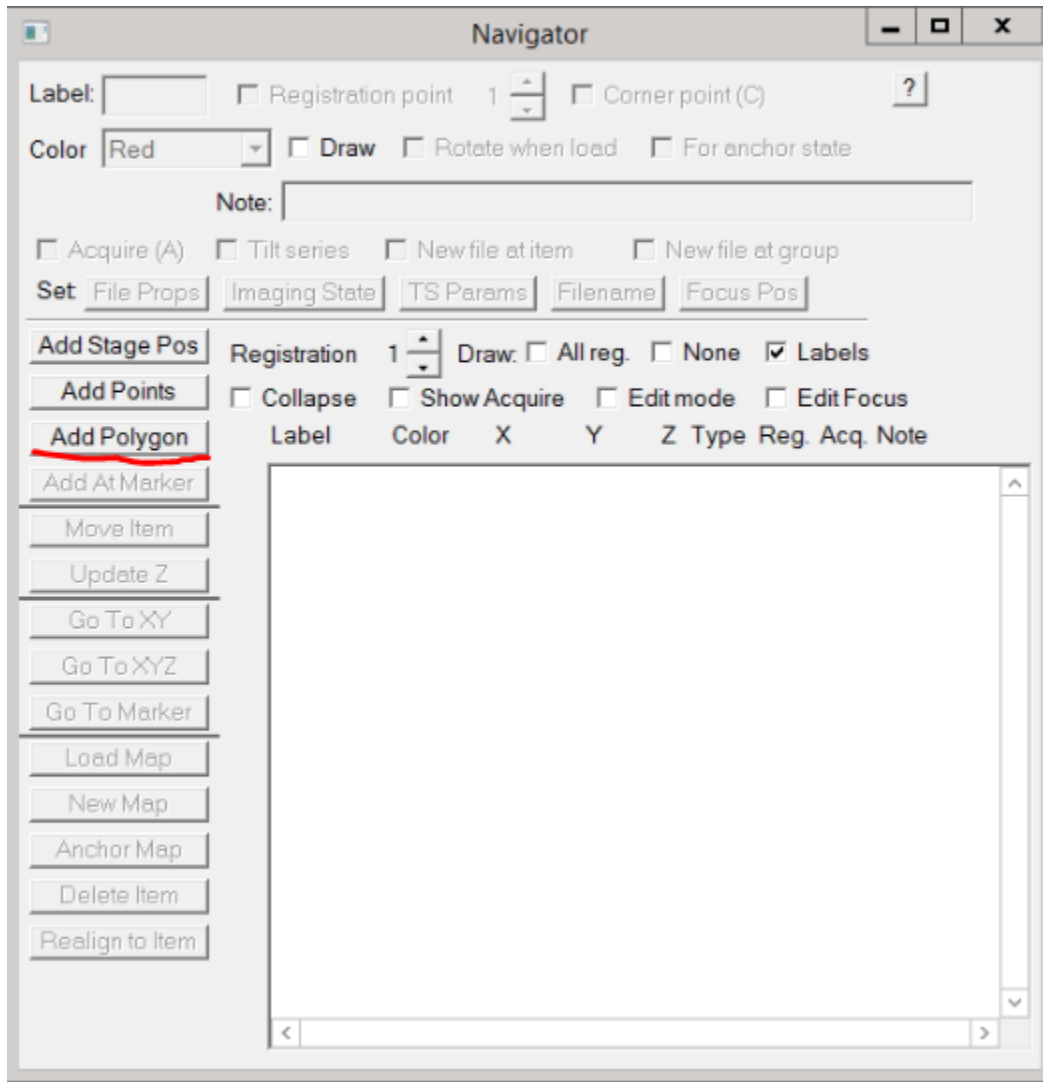


Figure 54. SerialEM "Navigator" dialog box depicting the protocol for creating a medium-magnification polygon montage.

Then, draw the polygon onto the grid atlas area that is to be montaged. This can be done by clicking the left mouse button to draw line segments.

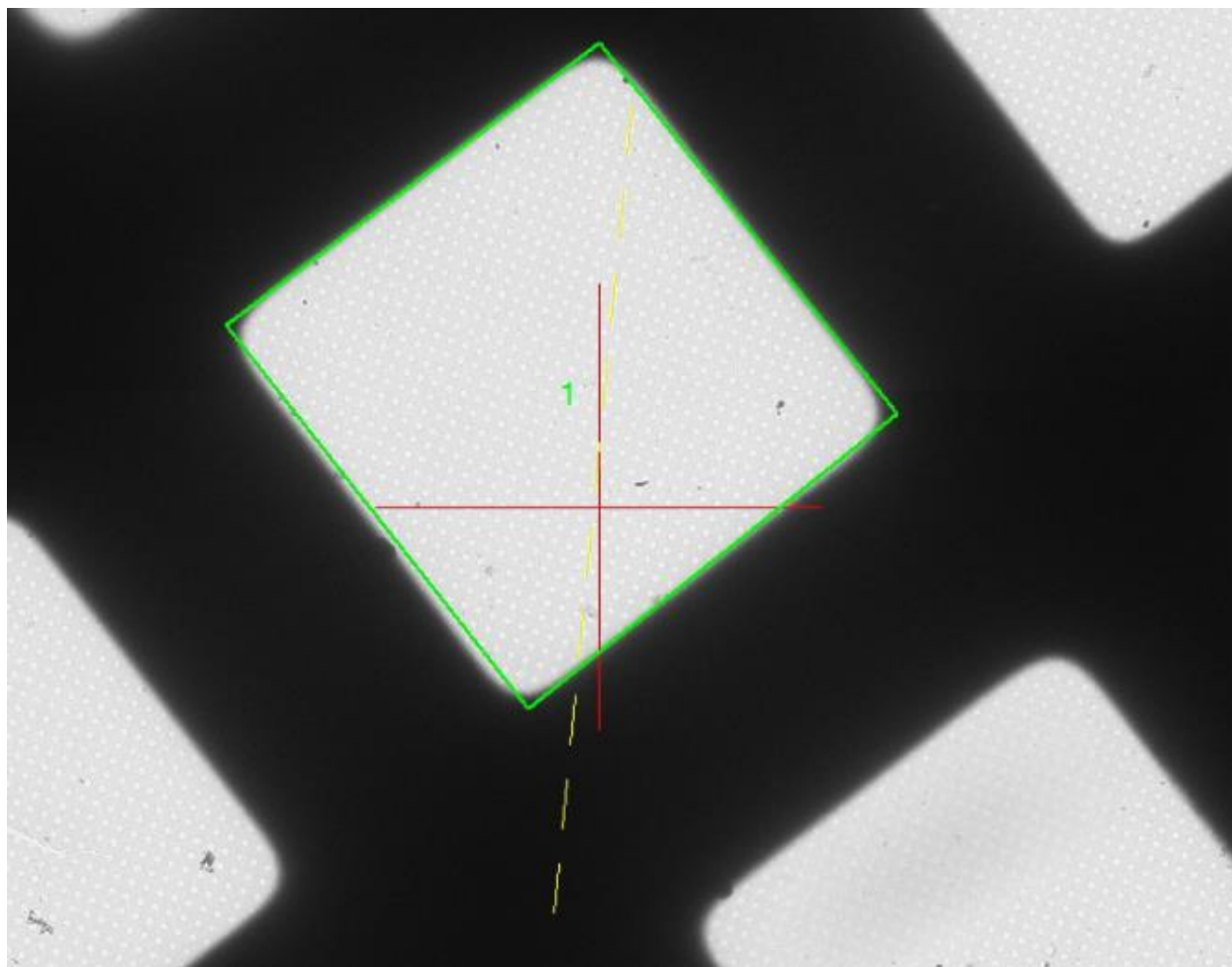


Figure 55. A grid atlas depicting the boundaries for a desired polygon montage in green.

Once the polygons are drawn, setup a new polygon montage by selecting the top “Navigation” menu, then “Montaging & Grids”, then “Setup Polygon Montage”. Select “Use View parameters in Low Dose mode” and SerialEM should automatically calculate the polygon size. When done, select “OK” and follow the prompts to save the polygon atlas file. Then, start the polygon atlas via the “Montage Control” panel in the lower left corner of the screen. Some other control panels may need to be collapsed for this to become visible.

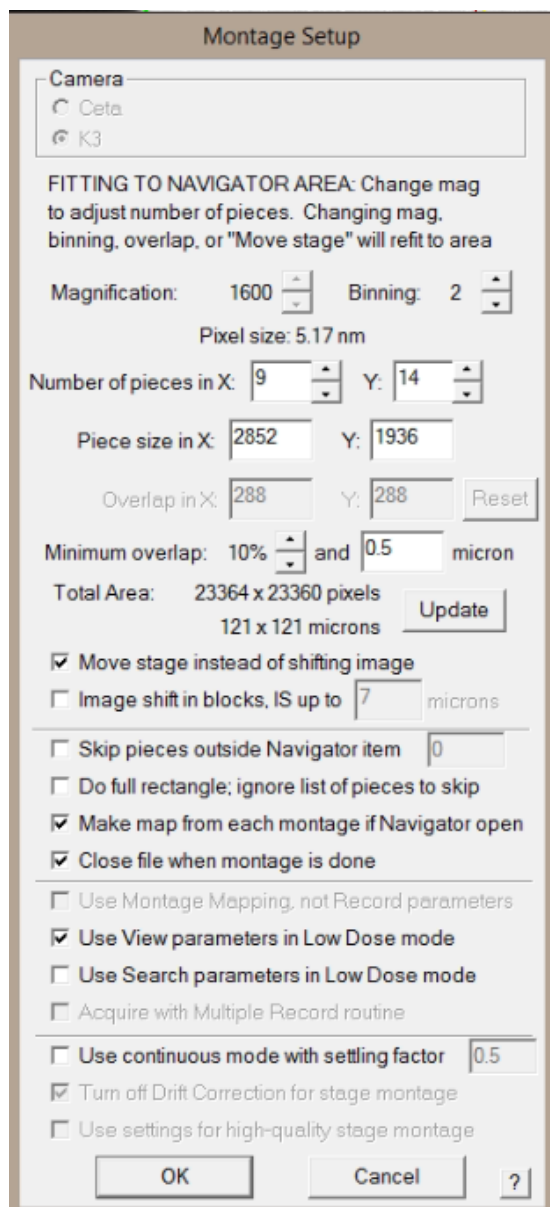


Figure 56. SerialEM "Montage Settings" dialog window depicting proper settings for a medium-magnification polygon montage for tomography.

Once the polygon atlas is created, load the atlas by double left clicking it in the Navigator window. First, load the desired magnification imaging state, and calibrate the dose over an empty area. Then, in the "Record" camera parameters, set the exposure to something small, such as 2-3 electrons/Å².

Load the desired magnification imaging state, and calibrate the dose over an empty area. Then, in the “Record” camera parameters, set the exposure to something small, such as 2-3 electrons/Å². In the Navigator window, select “Add Points” and place points over all the desired collection locations. Collapse the group, set its status to “Acquire (A)”, and select “Tilt series”. Follow the prompts to generate a new file. If the file name ends with a number, SerialEM will automatically increment the number for each new item. Once the files are setup, SerialEM will open the tilt-series setup page. First, set the desired tilt angle limits, and the base increment. It is recommended to leave the collection scheme as “dose-symmetric” unless there is a good reason to use a different scheme. Then, check the bottom left under “Beam Intensity or Exposure Control”. Enable variable exposure time if desired, or leave disabled. Examine the total dose. If the total dose is too high, reduce either the number of tilts, or the “Record” exposure time. If it is too low, increase the number of tilts or “Record” exposure time. Under “Autofocus Control”, ensure that the “Defocus target” is reasonable. Set the “Repeat autofocus if focus differs from prediction by” to a tolerable range – something like 0.5 μm. Do the same for “Check autofocus, stop if measured < X actual defocus”. Importantly, ensure that “Stop if autofocus change exceeds n um”, as this will prevent scenarios where runaway defocus occurs. It may be desirable to enable “Refine eucentricity”. This will use the camera settings in “Trial” mode and the autofocus position to refine the eucentric height at high magnification. This is typically required for high-magnification tomography, as any small inaccuracy in eucentric height will cause the image to translate outside of the detector area at high tilt. Finally, it is recommended to set “Repeat Record if percentage of field lost is more than” at something reasonable, such as 15%.

Tilt Series Setup

Tilt Angle Specifications

Tilt to: 60 End at: -60 Swap Base increment: 4 Total # of tilts: 31

☐ Increment proportional to cosine of angle

☒ Run series in two directions from: 0 ☐ Return with Walk Up

Magnification for anchor: 2000 11 um ☒ Use View for anchor

☒ Use dose-symmetric scheme **Set Up**

Delay time after tilting by basic increment: 10 seconds

☐ Change some settings as a function of angle **Set Changes**

Size, Image Shift, and Recentering Specifications

Magnification: 36000 Binning: 1 Pixel size: 0.110 nm

☐ At start and when needed, track at magnification: 5300

Limit image shift to: 3 um at start

☒ Recenter beam from Low Dose tracking images

☐ Autocenter beam every: 10 minutes

☐ Autocenter beam in second half of series at: 50 degrees

☐ Refine the zero-loss peak every: 15 minutes

Beam Intensity or Exposure Control

☒ Change Record exposure time instead of beam intensity

☐ Keep exposure time constant

☒ Vary exposure to be 2.6 x higher at 60 deg (1/cos to 1/0.73 power)

☐ Set exposure to keep mean counts of Record images at: 5000

The image in buffer A implies a Record zero-tilt mean of 0

☐ Taper counts down to: 4000 above +/- 50 degrees

☐ Keep exposure below current value (use if exposure already set up)

☐ Do not increase exposure above value for first saved image

☐ Use current exposure at zero tilt (increase to start series)

Dose: 1.92 e/A2 at 0 deg; 4.98 e/A2 at 60 deg; Total: 85.9 e/A2

☒ Change other exposures proportional to Record change

☒ Change drift settling proportional to exposure change

Camera and Acquisition Parameters

☐ Ceta ☒ K3

☐ Average Record dark references

☐ If taking frames, get initial partial sum of: 1 frames

Autofocus Control

Defocus target: -5 um Beam tilt: 1.6 milliradians

Autofocus offset: 0 um Iteration threshold: 8 um

Do autofocus at least every: 3 degrees ☐ Skip Autofocusing

☐ Focus every time above: 50 degrees

☒ Repeat autofocus if focus differs from prediction by >: 0.5 um

Do autofocus when error in focus prediction exceeds: 0.2 um

☒ Check autofocus, stop if measured <: 0.5 X actual defocus

☐ Stop if autofocus exceeds absolute focus limits

☒ Stop if autofocus change exceeds: 5 um

Initial and Final Actions

☒ Align to image now in Buffer A

☐ Refine eucentricity

☒ Leave mid-tilt anchor near plus or minus: 40 degrees

☐ Use mid-tilt anchor that you have left in buffer: T

☒ Close column valves at end of series

Tracking Control Parameters

☐ Wait for drift to settle **Set Up**

☒ Repeat Record if percentage of field lost is more than: 15

☐ Stop and ask before repeating Record in Low Dose mode

☐ Stop if Autoalign shifts more than: 30 % of image size

Get tracking image when error in X/Y prediction is >: 1.5 % of field

Track: ☐ Before ☐ After ☒ Before & after Autofocusing

☐ Stop after aligning tracking image to allow manual adjustment

☐ Align only with tracking images, not Record images

☐ Align with Preview before getting new track reference

Get new track reference if Record alignment differs by >: 2 %

Previous Next First Half Second Half Full

Set Up Extra Output OK Cancel ?

Figure 57. SerialEM "Tilt Series Setup" dialog window depicting recommended starting configuration for dose-symmetric tilt-series acquisition.

By default, the settings for the first tilt-series will be inherited to each subsequent tilt-series. If special changes are required, such as target defocus, select "TS Params" on the "Navigator" entry to edit it. Next, the focus parameters must be configured. Focus is an important consideration when collecting tilt-series, because the tilt axis is nearly perfectly vertical, which means that at any given tilt angle, the left and right sides of the image are

different focal distances. That means only the very center of the image should be considered for focus estimations. SerialEM can crop out the sides of the image when in Focus mode, automatically. Select the “Camera” top menu, and then “Parameters”. Navigate to the “Focus” tab and ensure that “Dose Fractionation mode” is disabled. Once disabled, the “Area size” section should become available. Select “Wide Half” to reduce the size of the detector area to half the height, but running the length of the X axis. Then, select “Swap X & Y” to rotate the detector area 90°. Finally, select “Recenter” to center the X and Y coordinates of the detector area. This will limit the “Focus” camera mode to only a strip of the detector running roughly parallel to the tilt axis.

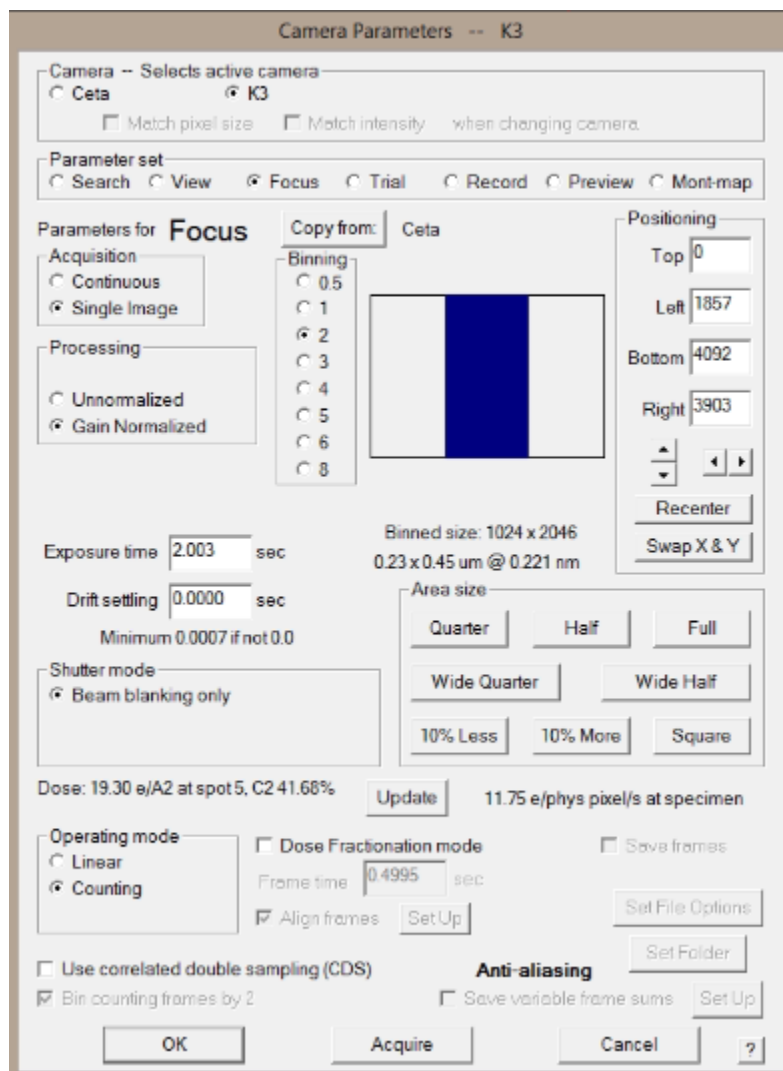


Figure 58. SerialEM "Camera Parameters" dialog window depicting recommended settings for high-magnification tilt-series acquisition.

Once camera focus parameters are set, it is time to set the focus position. Focus position when recording tilt-series requires special consideration, especially when in batch mode. The focus position should always be along the tilt-axis. However, since the location of the feature in each hole may change, the default focus position often times winds up in an area that is not appropriate for the sustained dosage when recording a tilt-series. To work around this, set the focus position manually for each point, un-collapse the group and select a single point, and click

“Focus Pos”. The focus position for that item should now be displayed. Ensure that it is in an appropriate area, before selecting the next point and doing the same. Iterate until all focus areas are properly set. SerialEM will display a “P” in the “Acq” column of the Navigator window if a custom focus position is set.

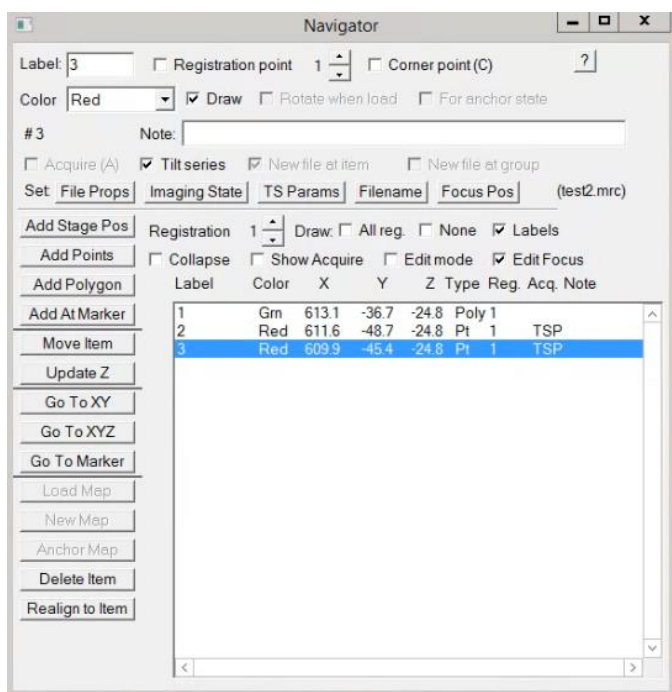


Figure 59. SerialEM "Navigator" dialog window depicting items marked for tilt-series acquisition. Note the "TSP" options under "Acq. Note". This indicates that the positions will be recorded as tilt-series, with a special focus position defined.

Recording Tilt-Series. Once configured, open the “Navigator” top menu. Select “Acquire at Items” and choose “Final Data” mode. Set the “Primary Action” to “acquire and save image or montage” and ensure that “Make Navigator map” is disabled. Set the defocus cycle as desired, and then set the tasks before primary to first “Realign to Item” for every item, and then autocenter beam. This will move the microscope to the stage location and realign to the image in

the polygon montage. Then it will autocenter the beam in the focus area and begin the tilt-series, saving it to the specified file.

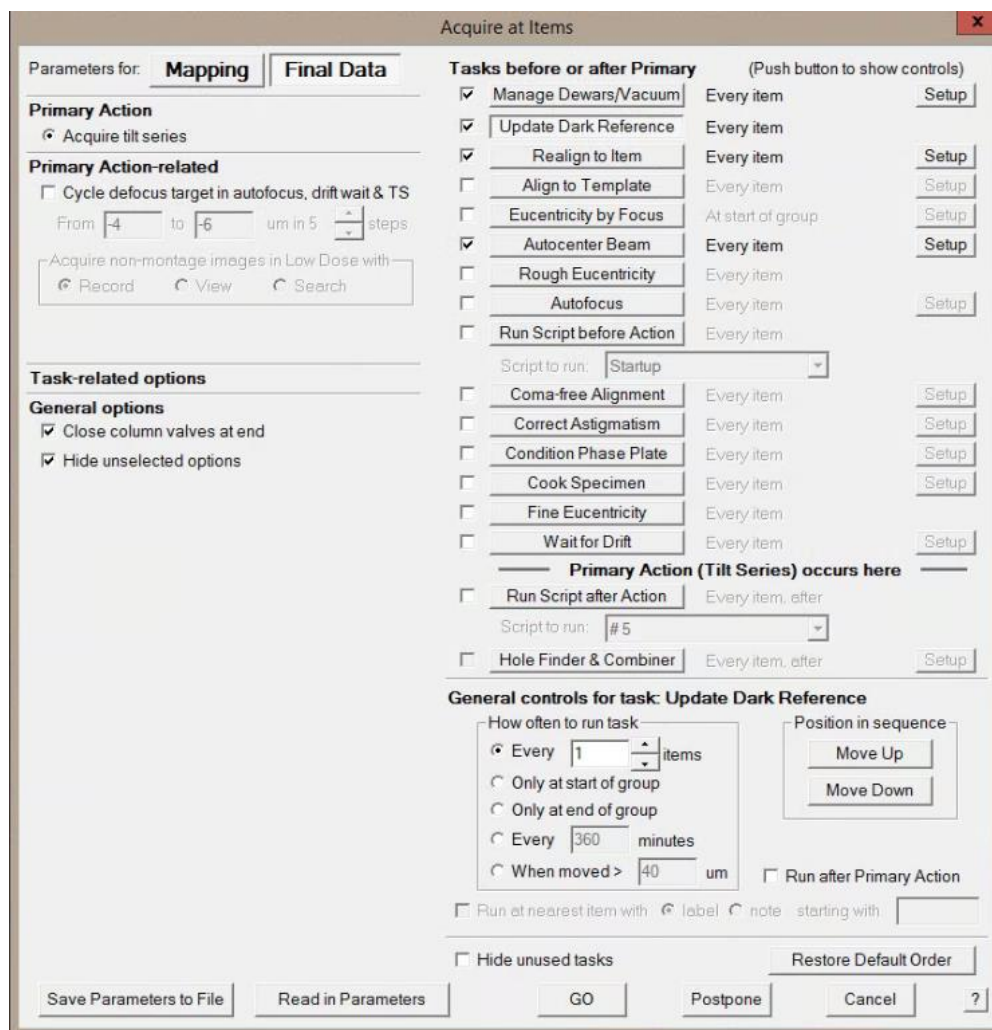


Figure 60. SerialEM "Acquire at Items" dialog window depicting the "Final Data" recommended settings for tilt-series acquisition.

PACETomo

PACETomo is a series of Python scripts for SerialEM that can collect multiple tilt-series simultaneously with multi-shot, similar to that used for single particle data collection. In addition

to multi-shot features, PACETomo also has various algorithms to compensate for things such as stage drift, stage pre-tilt, and Z-axis offset.

Selecting Targets. PACETomo has two ways to record data. One is by using a regular array of holes. The other is to manually place nearby points for collection. Starting with a View image, or polygon map as detailed in the “SerialEM Batch Mode” chapter, navigate to the “Scripts” top menu, select “Run”, and run “PACETomo: Enable”. This will enable the Python scripting interface for PACETomo. Then, in the “Image Alignment & Focus” control panel, expand the options and ensure that “Move stage for big mouse shifts” is unchecked. Back in the scripts menu, run the “PACETomo: Select Targets” script. The script will prompt you to choose a directory, and then will ask you to provide a rootname for the points it will place. It will then ask you to center your target by dragging the image, and pressing the “b” key when finished. Once centered, it will ask you if you want to take a preview image at the point. Select “Yes” and it will take a preview. You may then refine the position at high magnification if desired. Alternately, you can just accept the position. PACETomo will then add the point, and show the total area exposed to the beam in a yellow circle.

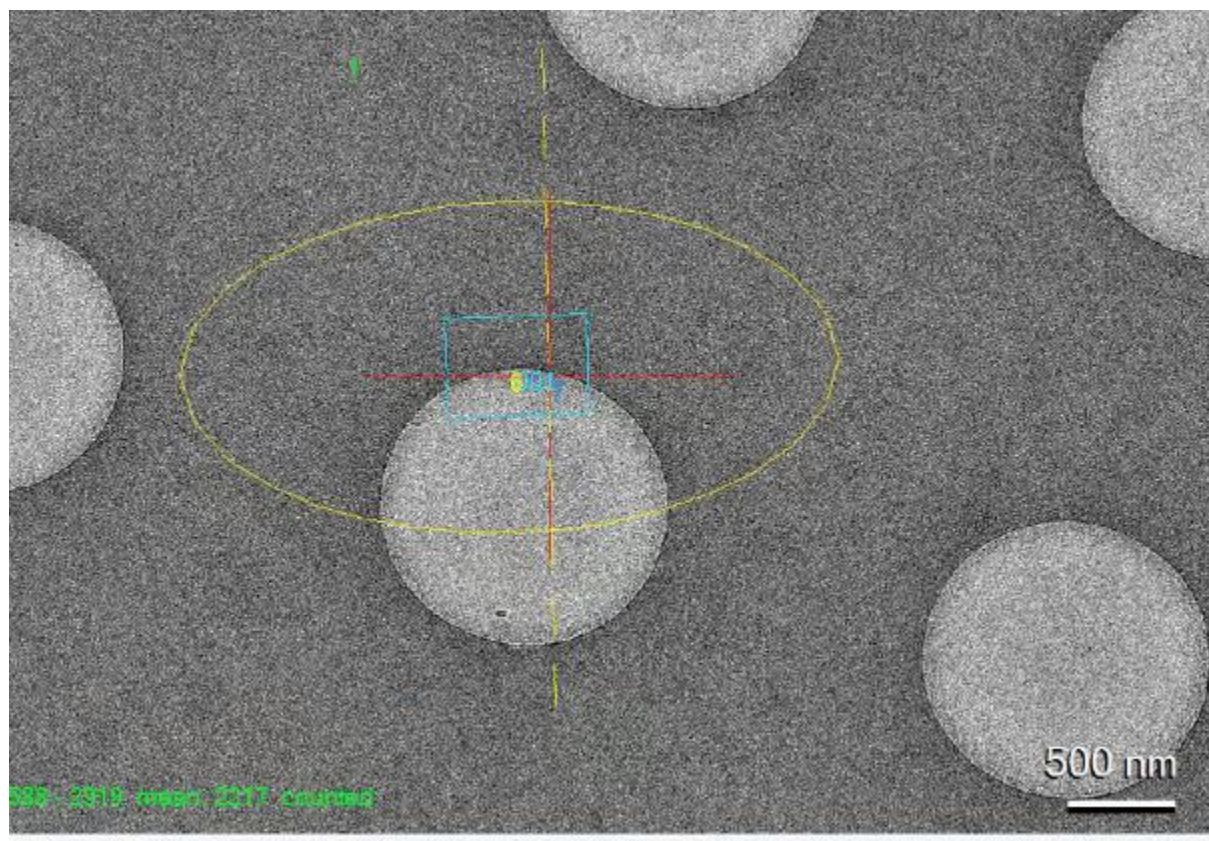


Figure 61. A medium-magnification image of the selected target for PACETomo acquisition (blue box) as well as the total exposed area (yellow oval). The exposed area is larger than the acquisition area due to the beam size and shape, as well as stage tilt.

PACETomo will then ask you to place more points. Iterate until done. When done, simply press “b” and select “No” for a preview image. It will then ask if you want to keep looking for a target. Select “No”. At this point, the main PACETomo interface will open. This tutorial will not cover all of the features for PACETomo, but there are a number of options including setting specific targets for tracking, or the order in which to collect, that can be customized here. Alternately, more targets can also be added from this window. Once satisfied, simply select “Save” and then close the window.

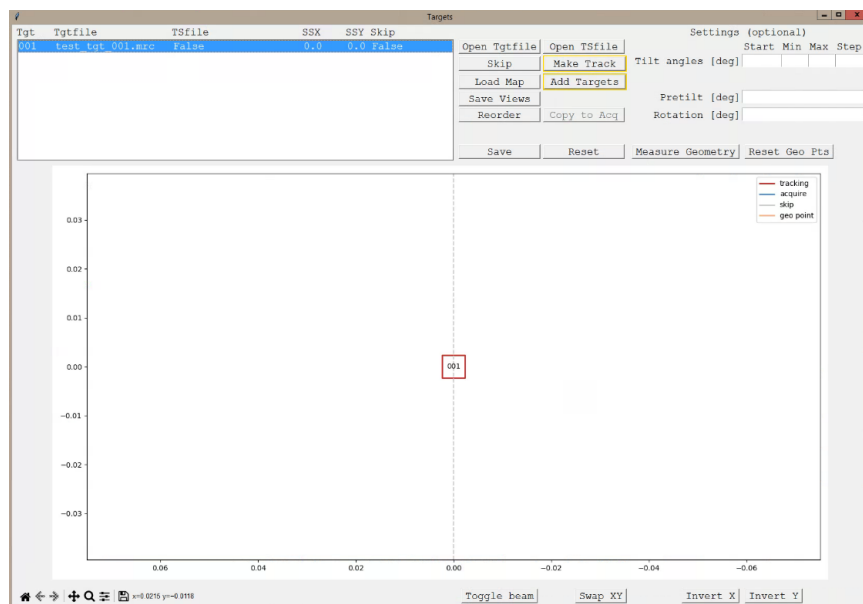


Figure 62. PACETomo "Targets" dialog window.

In the navigator, there will now be two entries for each target added – one entry ending in “.txt” and the other entry ending in “.mrc”. Select the top, “.txt” entry by left clicking.

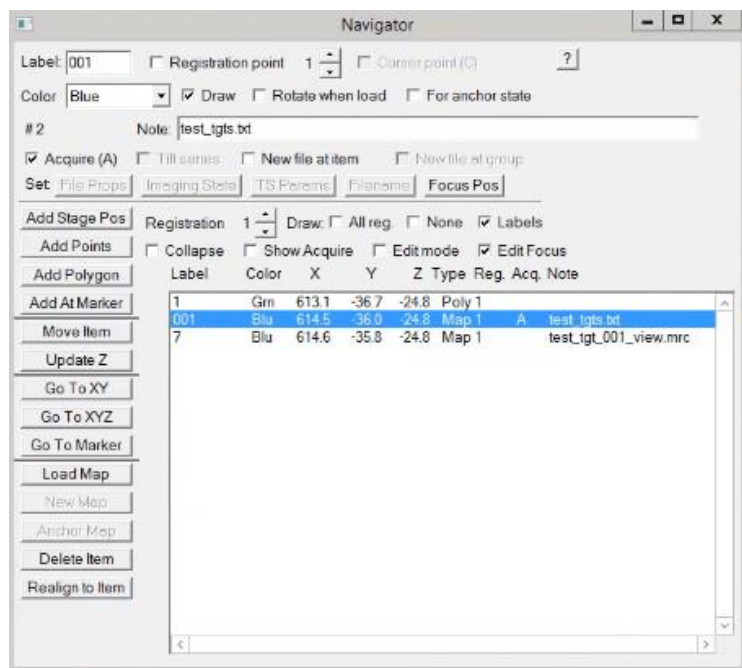


Figure 63. SerialEM "Navigator" dialog window depicting proper PACETomo targets selection.

PACETomo can also be configured for collection of a regular array of holes, such as 3x3 or 5x5 multi-shot, and target selection can be automated. To do this, first take a Record image using the “Grid Square Map” imaging state. Use the “Find Holes” routine on the image to find the holes. Then, open the “Scripts” menu, and select “Edit/Run One Line...”. Enter “ReportLastHoleVectors 0” and run it. This will print out four hole vectors in the log. Enter those vectors in the “Select Targets” script, under “#Advanced Settings” for “vecA” and “vecB”. Above that, set “targetPattern” to True, and “alignToP” to true. Set “size” to 1 if using 3x3, and “2” if using 5x5. Then, take a View image of a hole. Using Ctrl + Shift, draw a box around a hole, and then open the top “Process” menu. Select “Crop Image” and crop the image to contain just the hole and some surrounding support. Go to the “Buffer Controls” control panel, and select “Copy active image to buffer” and copy the active image to buffer P.

Now, open the grid atlas. Place one point on the center of each desired square on which to center and collect a multishot pattern, and set them to acquire. Now, run the “Select Targets” script and follow the prompts. Once the main target selection window opens, select “Copy to Acquire” to set all the navigator points marked “Acquire (A)” to inherit that same pattern.

Recording Parallel Tilt-Series. Once targets are setup, edit the “PACE: Collect Tilt Series script.” Ensure that the parameters make sense for your collection. The most important setting is “tgtPattern”. Set this to false for positions placed by hand, or true to collect a regular array of holes. When satisfied, run the script and it will collect the targets as set up.

CHAPTER THREE

THE MECHANISM OF IRON NUCLEATION AND GROWTH
IN A MINI-FERRITINIntroduction

Iron is a necessary element for life and plays a key role in a variety of metabolic processes across all known life forms. However, free iron, in its ferrous form (Fe(II)) is susceptible to oxidation via hydrogen peroxide, which results in generation of the hydroxyl radical ($\text{OH}\cdot$), a highly toxic chemical and threat to cellular life (38). Therefore, careful regulation of free Fe(II) inside cells is necessary to prevent free-radical damage (37, 38).

Ferritin family proteins, nearly ubiquitous throughout the tree of life, have evolved specifically to address this problem (57). Ferritins oxidize Fe(II) to Fe(III) and store it within the ferritin nanocage as hydrated ferric oxide. The ferritin family is frequently recognized to consist of three major classes; i) ferritin, ii) bacterioferritin and iii) DPS (DNA Protection in Starved cells) (49, 57-60). The eponymously named ferritin is a hollow nanocage 10-12 nm in diameter, composed of 24 subunits that assemble with octahedral (432) symmetry. The subunit is a 4-helix bundle that coordinates a bi-nuclear iron center. The iron binding sites are coordinated by glutamate and histidine residues that are buried in the middle of the helical bundle, and serves as the site of ferroxidase activity (41, 42, 45, 46, 48, 60). Like ferritin, bacterioferritin shares the 24 subunit assembly and the ferroxidase center (FOC), but incorporates a heme group at each two-fold symmetry axis. DPS also shares the 4-helix bundle fold, but in contrast, this “mini-ferritin” instead forms a smaller, dodecameric 9 nm nanocage with tetrahedral (23) symmetry. Further,

rather than a ferroxidase site sequestered within the 4-helix bundle, the FOC is instead found on the inside of the shell at the 2-fold subunit interface.

Ferritin, bacterioferritin and DPS are each capable of oxidizing Fe(II) to Fe(III) via coupled reduction of O_2 to H_2O_2 , or H_2O_2 to H_2O . However, ferritin and bacterioferritin (maxi-ferritins prefer O_2 , while DPS generally prefers H_2O_2 . The oxidized iron product, $Fe^{3+}OOH$, is extremely insoluble. While typical iron concentrations are on the order of 10^{-4} M, Fe(III) is insoluble at concentrations above 10^{-18} M (49). Ferritins once again contribute a solution to this problem. In addition to safe oxidation of Fe(II) to Fe(III), they are also capable of storing the mineral form of $Fe^{3+}OOH$ inside of their main cavity (58, 61). While the exact composition of the mineral varies, it is typically thought to be some form of ferrihydrite (61). The larger ferritins can store between 2500-5000 iron atoms in total, while the mini-ferritins range between 250-500 iron atoms.

An additional member of the ferritin family, overlooked at times, is DPSL. First identified in hyperthermophilic archaea, DPSL forms a 12-subunit (dodecameric) nanocage that preferentially utilizes H_2O_2 as a substrate(45, 46, 62, 63). For these reasons, DPSL was first perceived as a DPS-Like mini-ferritin. However, structural studies soon revealed a surprise; DPSL lacks the DPS ferroxidase site and instead retains a bacterioferritin-like ferroxidase center in the middle of the 4-helix bundle (46). For these reasons, DPSL is proposed to lie at the evolutionary interface between the mini- and maxi-ferritins (46, 47, 57, 64). Importantly, DPSL further differentiates itself from both the (bacterio)ferritins and DPS by the presence of two strictly conserved cysteine residues that are juxtaposed between the ferroxidase center and the surface of the particle. Together, the conserved cysteine and the bacterioferritin-like ferroxidase

center serve as hallmarks of the archaeal DPSL proteins that are easily identified in the primary sequence as the thioferritin motif, and differentiate DPSL as a distinct member of the ferritin superfamily (47). “DPS-Like” is thus a misnomer for this member of the ferritin superfamily, and for that reason we henceforth refer to *Pyrococcus furiosus* DPSL as a “thioferritin” (Tftn) to better differentiate it from DPS proteins. While the cysteine residues in the thioferritin motif are potentially redox active, whether they play a role in catalysis, or gating access to the ferroxidase site is not yet understood. Finally, in addition to archaea, thioferritins also participate in the oxidative stress response in anaerobic bacteria like *Bacteroides fragilis*, the most commonly isolated bacteria from anaerobic infections (47). Thioferritins are thus of significant clinical interest as well (46, 47).

Crystallographic studies have provided a wealth of structural information on the apo forms of (bacterio)ferritin and DPS. However, crystallographic studies of mineralized ferritins have proven more challenging, and there are few available structures of ferritin family proteins with mineralized cores. Thus, the detailed molecular interactions that result in biomineral nucleation and growth are currently unclear (44, 51). We hypothesized such structures might be significantly more amenable to cryo-EM. Indeed, here we report not only the high-resolution (1.86 Å) apo structure of thioferritin from *Pyrococcus furiosus* (Pf-Tftn), but the structure of an early mineralization intermediate, and a fully mineralized form as well. These structures provide unique insight into the mechanisms of ferritin iron mineralization and storage across three distinct stages of the biomineralization process; i) mineral free, ii) initial nucleation and early deposition of iron oxide, and iii) the transition to the fully mineralized form. We also report a fluorescence lifetime assay that reports on nucleation and initial iron deposition.

Results

Previous work has demonstrated that the Tftn from both *Saccharolobus solfataricus* (Ss-Tftn) and *Pyrococcus furiosus* (Pf-Tftn) will mineralize Fe(II) using either H₂O₂ or O₂, though peroxide is preferred (45, 62). Prompted by our earlier crystallographic work with Ss-Tftn (46), we first screened apo-Ss-Tftn for cryo-EM studies. However, initial single particle work gave maps with only modest resolution. For this reason, we turned instead to Pf-Tftn, for which there was no structural information.

Structure of apo Pf-Tftn at 1.86 Å resolution

The Pf-Tftn structure was determined at 1.86 Å resolution by single particle analysis using our in-house 200 kV Talos Arctica and Gatan K3 camera. Details on the structure determination and workflow are presented in the Methods and Appendix B - Figure 73, and representative density is shown in Figure 64, panel A.

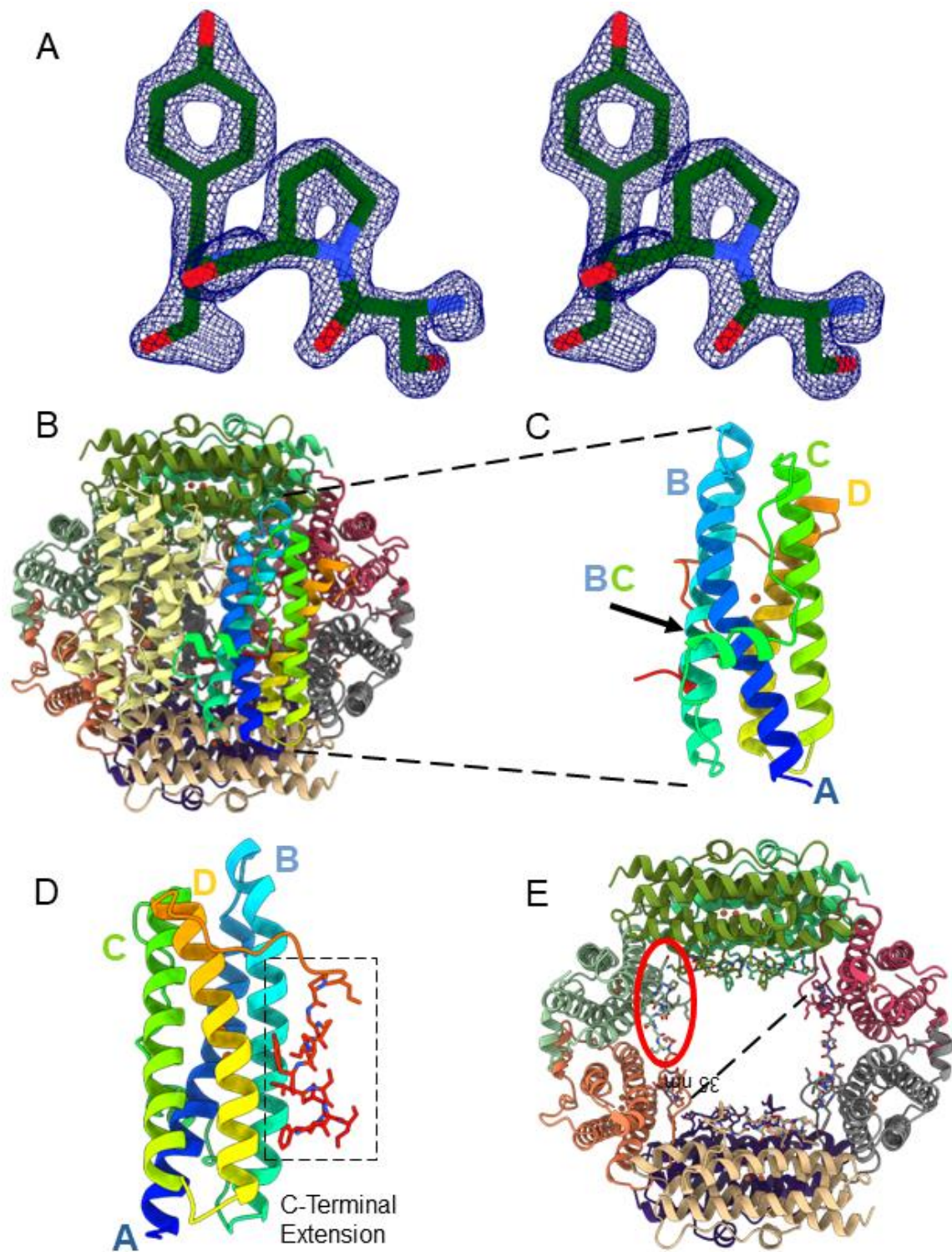


Figure 64 A) Stereo image of representative electron density and model. B) Cartoon representation of DPSL model. C) Single four-helix-bundle of DPSL with labeled helices. Black

arrow depicts connecting BC helix. D) 90° rotation around Y axis of four-helix bundle depicting newly modeled C-terminal extension in dashed box. E) Cutaway view of DPSL model depicting hollow core and C-terminal extension (circled red on a single subunit).

As expected, the structure revealed a dodecameric particle with tetrahedral (23) point group symmetry, with subunits that exhibit a 4-helix bundle fold that incorporates a bacterioferritin-like ferroxidase center with each iron coordinated by a histidine and pair of acidic residues. In addition, two conserved cysteine residues reside adjacent to solvent channels leading into the ferroxidase center (Figure 65).

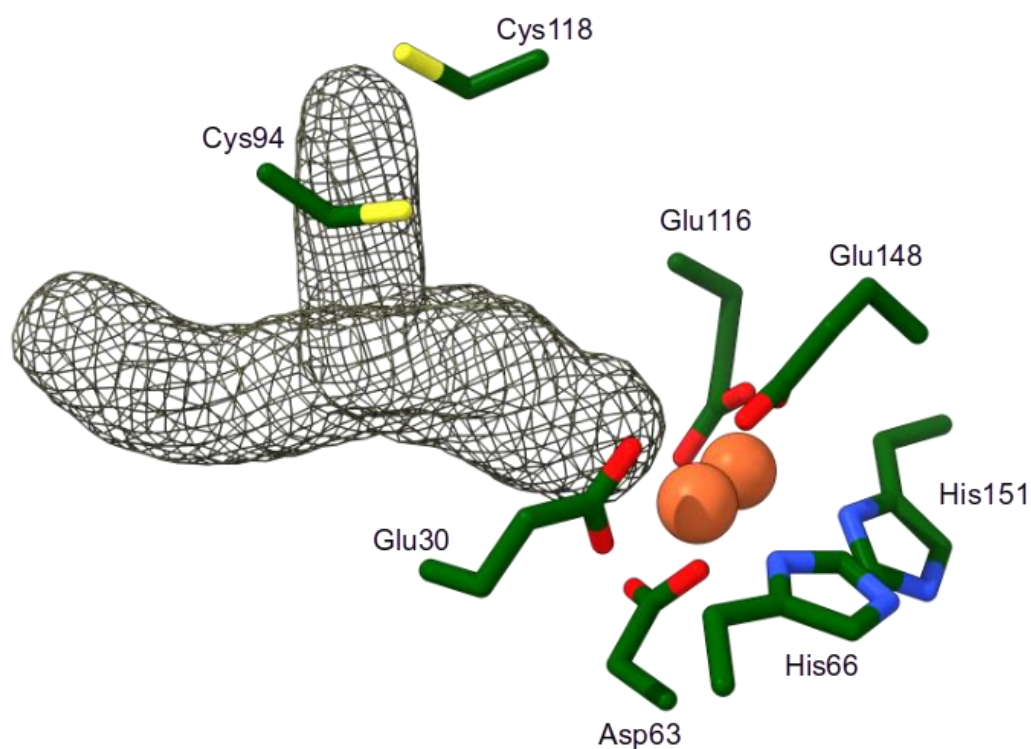


Figure 65 Stick representation of the Pf-Tftn bacterioferritin-like active site, and conserved cysteine pair. Solvent channels annotated in brown.

To our knowledge, the only other thioferritin structures solved are those from *Saccharolobus solfataricus* (Ss-Tftn, 2CLB) and *Bacteroides fragilis* (Bf-Tftn, 2VZB) (46, 47).

While Ss-Tftn and Bf-Tftn are quite similar, they differ significantly with regard to their C-terminal tails. C-terminal residues 174-188 are unmodeled in the Ss-Tftn structure, but the C-terminus enters the interior of the protein shell before becoming disordered. In contrast, in Bf-Tftn the C-terminal residues are present as an extension of helix-D, and remain on the exterior surface of particle where they occlude the acidic 3-fold pore (46, 47). Consistent with its hyperthermophilic, archaeal origins, structural superposition of these structure show that Pf-Tftn is more similar to Ss-Tftn ($C\alpha$ RMSD = 0.69 Å) than to Bf-Tftn ($C\alpha$ RMSD = 2.72 Å).

Disordered N-Terminal Tail

While many DPS proteins, as well as both Ss- and Bf-Tftn exhibit an N-terminal helix in the apo structure, Pf-Tftn does not – instead showing disorder in this region. In fact, it is the disorder at N-terminus of Pf-Tftn that appears to effectively limit resolution to 1.86 Å. The local resolution histogram actually shows a significant number of voxels distributed between 1.86 and 1.6 Å resolution (Appendix B, Figure 73). In some DPS proteins, the N-terminal helix is implicated in DNA-binding activity, likely via insertion of the N-terminal helix into the major groove of double stranded DNA, although other DPS proteins lack this DNA binding activity (65, 66).

Ordered C-Terminal Tail

The major difference between Ss-Tftn and Pf-Tftn lies in the C-terminal tail. In both structures, following helix D on the exterior surface, the C-terminal tail transits the particle. In Ss-Tftn it then becomes disordered. In Pf-Tftn, however, the polypeptide then turns and runs parallel to the four-helix bundle along the inside of the shell, with 13 additional ordered residues (Figure 64). Interestingly, not only is this tail rich in positively charged residues (3 Lys, 3 Arg), but the added

structure also results in 7 unsaturated main chain carbonyl groups (Pro172, Glu174, Pro176, Lys180, Phe181, Leu182 and Lys183) that project into the interior solvent. Multiplied by 12-fold symmetry, the C-terminal tails contribute 84 carbonyls that line the interior surface of Tftn, that might potentially interact with ferrihydrite. With the addition of these ordered residues at the C-terminus, the measured interior diameter is approximately 30 Å, with a calculated interior volume of 11,860 Å³. Based on the stoichiometry and structural models for the expected ferrihydrite mineral core (67) observed in ferritin (61), this implies a theoretical capacity of approximately 345 to 430 iron atoms, depending upon the structure of the mineralized core (67, 68).

Iron Loading

One of the hallmark functions of ferritin-family proteins is their ability to oxidize Fe(II) to Fe(III), while completing a simultaneous reduction of molecular oxygen (O₂) to hydrogen peroxide (H₂O₂), or hydrogen peroxide to water (H₂O), in a controlled manner (49, 65). Additionally, the insoluble iron-oxide product, often thought to be in the form of ferric hydroxide (FeOOH) is then mineralized inside the cavity of the protein shell, typically as ferrihydrite ((Fe³⁺)₂O₃·0.5H₂O) (42, 50, 57). Both ferritin and DPS can use either O₂ or H₂O₂ for oxidation, although ferritin prefers O₂, and DPS prefers H₂O₂. While previous studies on Ss-Tftn and Pf-Tftn show a preference for H₂O₂, in our hands attempts to full load Pf-Tftn using H₂O₂ resulted in the formation of visible precipitates and loss of protein from the soluble fraction. In contrast, the addition of a single aliquot of 50 Fe(II) per cage, followed by addition of stoichiometric H₂O₂ lessened the problem, with significant Pf-Tftn remaining in solution after 10 minutes. Because a structure of a partially loaded particle could provide mechanistic insight into the nucleation and

initial deposition of iron (hydr)oxide, we pursued the structure of this potentially informative intermediate by cryo-EM.

Nucleated Thioferritin

Following incubation with 50 iron atoms per cage and stoichiometric amounts of H_2O_2 for 10 minutes, Pf-Tftn was vitrified on cryo-EM grids and imaged for single particle analysis (Appendix B, Figure 74). Notably, with the exception of 3,000 particles, which were excluded from the reconstruction, particles exhibiting iron density in the core were not visible in the micrographs, or upon 2D classification. Ultimately, a total of 1,742,467 particles were used in homogenous refinement, yielding a map at 2.20 Å resolution that was largely similar to that for the apo protein (Appendix B – Figure 74). However, clear density at the 3-fold acidic pores suggested partial iron occupancies at this position. For this reason, the particle set was D2 symmetry expanded to give poses for ~7,000,000 unique 3-fold pores. A mask was then placed around the acidic 3-fold interface, and the resulting particles were subjected to 3D classification, which gave two classes of roughly equal size, corresponding to iron free (3,385,345 particles) and iron rich pores (3,584,523 particles). Unfortunately, an asymmetric (C1) reconstruction of the iron rich 3-fold particles, with or without homogenous refinement, did not resolve a high-resolution structure, although the strongest density in the entire map was at this location. Therefore, the symmetry expansion on the iron rich data set was reversed, giving ~1,600,000 particles with iron present in at least one 3-fold, and these were then used for homogenous tetrahedral refinement. With subsequent particle polishing, this gave a 1.91 Å structure. Critically, despite averaging over the tetrahedral symmetry, which decreased the occupancy of

the iron-(hydr)oxide structure to ~50% (3,584,523/6,969,868), this gave nicely resolved iron atoms, allowing modeling of the nascent iron-oxyhydroxide structure (Figure 66).

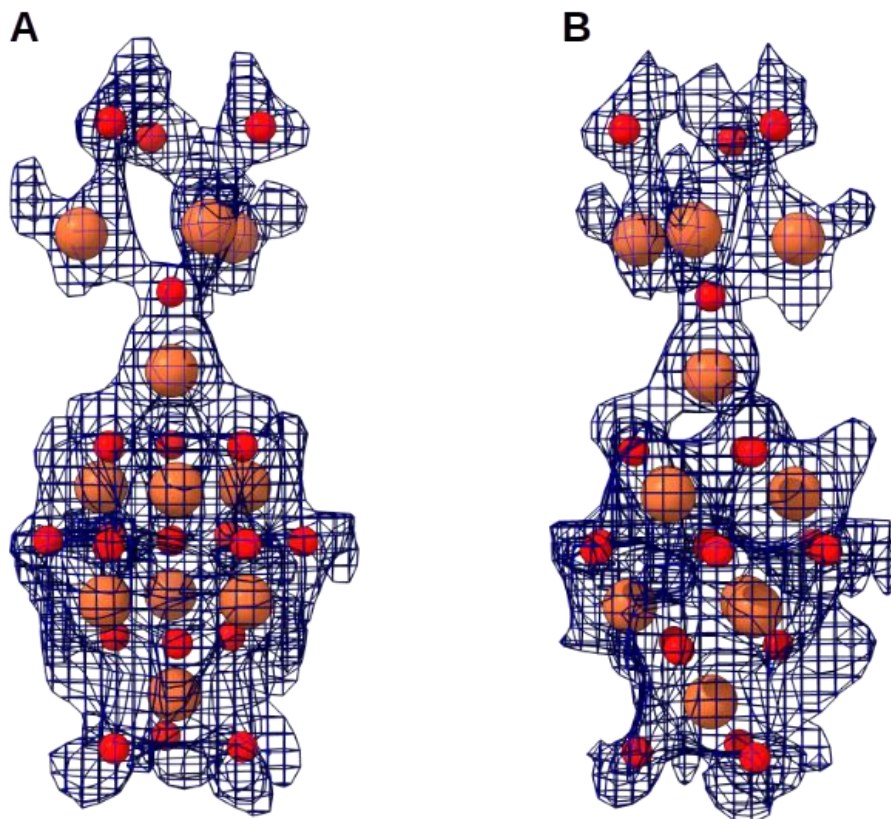


Figure 66 Iron mineral model and associated density from nucleated structure. A) View of mineral density with 3-fold axis running vertically along the screen. B) 90° rotation around Y.

Nucleated Structure

With the exception of the acidic pore, the structure of this iron nucleated particle is largely similar to the apo structure, with $C\alpha$ R.M.S.D of 0.37 Å. Within the pore, 11 iron atoms and 19 oxygen atoms were modeled in a 3-fold symmetric structure (Figure 67).

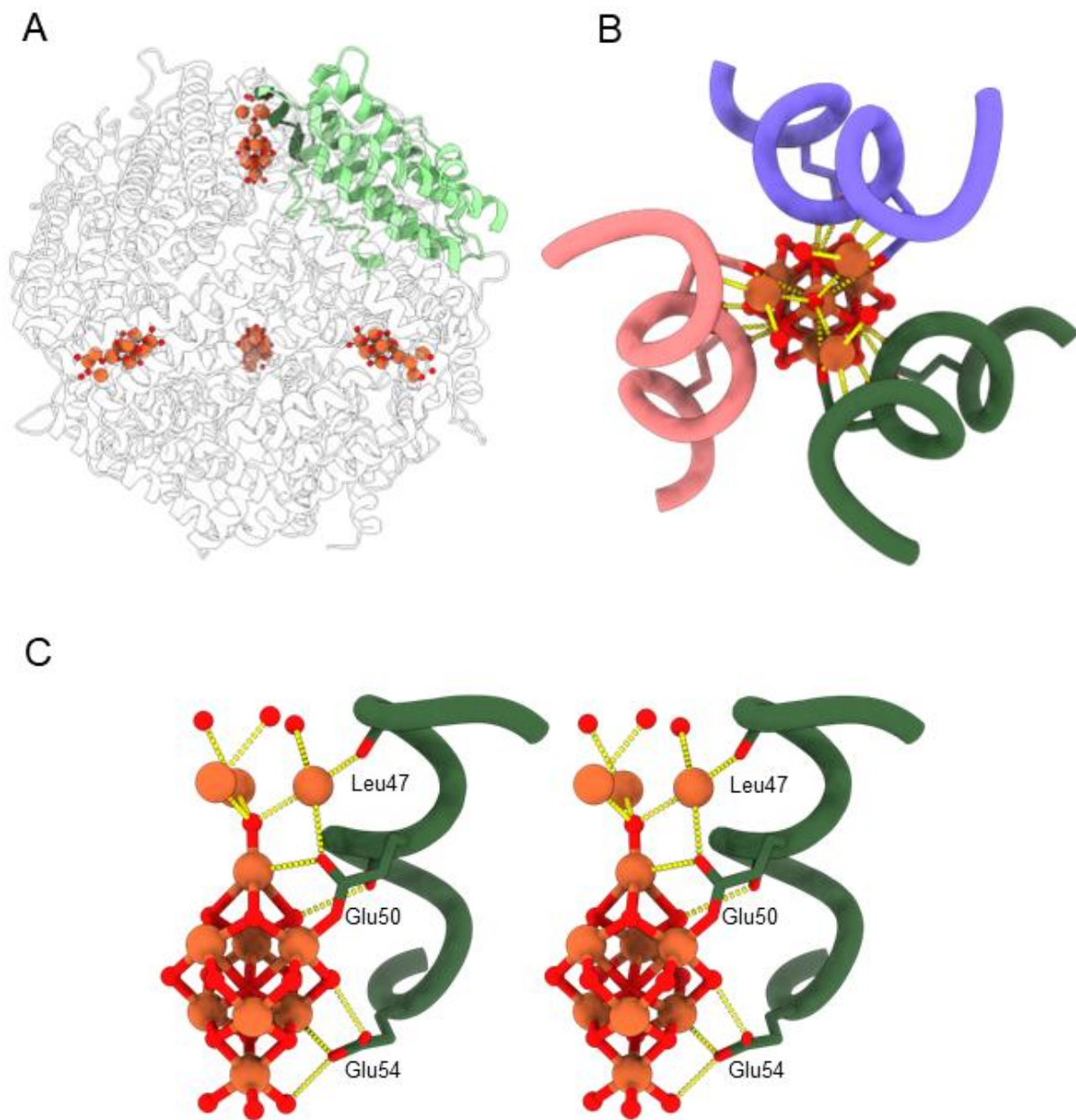


Figure 67. A) Pf-Tftn dodecamer showing the location of the iron minerals at the four acidic pores. Iron atoms are rust-colored, and oxygen atoms are colored red. A single subunit is highlighted in light green. The interacting region (Leu47-Glu54) is shaded dark green. B) A view of the acidic pore looking from the exterior of the protein cage to the interior. Protein chains are shaded purple, pink, and green. Interacting side chains and carbonyl groups are displayed. C) Stereo view of iron mineral contained in the acidic pore, rotated 90° about X and displaying a single of the three protein chains. The mineral is positioned so that the exterior surface of the cage is at the top of the mineral, and the interior of the cage is at the bottom of the mineral.

Iron oxide and iron oxyhydroxide mineral structures are commonly formed of planes of close packed oxygen atoms, with iron atoms occupying various interstitial positions between the planes. Examples include the hexagonal close packed (ABABAB) structure of hematite, the cubic close-packed structure of maghemite (ABCABC), as well as two competing models for ferrihydrite (67, 68). In this context, the locations of the oxygen atoms within the 3-fold pore are indeed consistent with planes of close-packed structures. Moving along the three-fold axis from the interior of the particle outwards (bottom to top in Figure 67), the six oxygen layers are consistent with planar stacking in an ABCACA manner, with the four innermost layers (ABCA...) resembling a cubic close-packed structure (ccp), while the 4 outer layers (..CACA) are instead hexagonal close-packed (hcp). In this model, the two middle layers (..CA..) represent an overlap between these two different packing modes. The number of oxygens in each layer is quite small. Each of the A and B layers have only 3 oxygens, symmetrically arranged around the 3-fold axis, while the C layers have a single oxygen atom coincident with the 3-fold, which in the case of the third layer, is surrounded by a hexagonal distribution of six additional in-plane oxygens.

Iron present at each of the interstitial vacancies between the oxygen planes, giving a close packed iron atom lattice as well. The eight iron atoms in the innermost 4 layers show clear octahedral coordination to six oxygens, while the last 3 irons at the top of the structure, while ordered, appear under-coordinated. Measuring between oxygens in the top and bottom layers, the mineralized iron structure extends nearly 15 Å along the 3-fold axis, but is only 6 Å at its widest point (O to O across layer 3). Many of the oxygen atoms on the surface of the mineralized structure also appear undercoordinated, with only 2 or 3 bonds to Fe, suggesting

they're present as hydroxyl moieties, rather than O^{2-} . This is consistent with the expected iron oxyhydroxide nano-mineral, as are the observed O-Fe bond distances which range from 1.95 to 2.15 Å for the 17 oxygens in the bottom 5 layers. The outermost layer of 3 "oxygens" are an exception, where average O-Fe bonds distances are 2.54 Å, suggesting these are instead strongly ordered waters.

Protein-Mineral Interactions I

Residues in the 3-fold channel that interact with the mineralized iron include the main chain carbonyl of Leu47, Glu50 and Glu54 (Figure 67). The role of the Leu47 carbonyl is the most straightforward, where it coordinates iron in the exterior iron layer. In contrast to the ionic interactions between the iron and the mineralized oxygen atoms, at 2.5 Å the carbonyl oxygen-iron bond distance is significantly greater.

Glutamate 50

The role of Glu50 is more complex, as the negative charge appears to be delocalized over both side chain oxygens, allowing interactions with 3 different iron atoms. The strongest interaction is with iron in the middle, or third iron layer of the complex, where the Oε1-Fe bond distance is 2.0 Angstroms. In addition, Oε2 lies 2.9 Å from the single iron in layer 4, and 3.0 Angstroms from the nearest iron in layer 5. Importantly, Glu50 also contributes its main chain carbonyl group. But unlike the Leu47 carbonyl, the Glu50 carbonyl oxygen lies within H-bonding distance (3.1 Å) of an oxygen atom in layer 4, again consistent with an oxyhydroxide nano-mineral.

Glutamate 54

Both oxygens in Glu54 also interact with the mineral structure. As might be expected, Oε1 again interacts with iron, with a 2.65 Å bond distance to the nearest iron in layer 2. Unlike Glu50, however, Oε1 also interacts with oxygen in the bottommost layer, and Oε2 is 2.45 Å from oxygen in the 3rd layer, suggesting hydroxide in the first and third oxygen layers as well. Thus, the carboxylate and carbonyl moieties present within the "acidic" pore not only accommodate the positively charged iron, but the hydroxide at the surface of the nascent mineral as well. For these reasons, the inorganic oxygens with fewer than 4 iron bonds are modeled as hydroxide.

Ferrihydrite

While the outermost layer of iron and oxygen exhibit extended bond distances and less regular geometry, the first 5 oxygen layers and intervening iron atoms have a notably crystalline appearance. In this light, ferritin and DPS each mineralize iron as ferrihydrite (50, 61, 69), suggesting Pf-Tftn should as well. Currently, a generally accepted model for ferrihydrite is lacking, with two different models currently in the literature (67, 68), although a recent *ab initio* thermodynamics study suggests they may actually coexist (67, 68, 70). Regardless, each model is in a hexagonal space group that includes a crystallographic 3-fold axis. For this reason, we aligned the ferrihydrite 3-fold axes with the 3-fold axis in Pf-Tftn to compare these ferrihydrite models to the nucleated Tftn structure. While the nucleated structure is not a perfect match to either ferrihydrite model, we note greater similarity to the model of Michel et al. (67). In this model, 4/5th of the iron atoms are octahedrally coordinated with the remainder in tetrahedral coordination (50). In the unit cell, this gives rise to a cluster of 4 octahedrally coordinated irons sitting on each of the 3-fold axes, with a 5th tetrahedrally coordinated iron

immediately above them. We find similar 3-fold symmetric, 4-iron clusters in nucleated thioferritin. Specifically, the 4-iron cluster and associated oxygens in ferrihydrite superpose on the bottom 4 irons and surrounding oxygens that are coordinated by Glu54. In addition, when rotated 180° about the horizontal axis, the same cluster also superpositions on the middle 4 irons coordinated by Glu50, although the top iron on the 3-fold (Fe2 in ferrihydrite) lacks obvious octahedral coordination. The two 4-iron clusters in the nucleated Tftn mineral thus share a central plane of hexagonally arranged oxygens (layer 3). Notably, the central oxygen in this plane is not as well resolved. While this might be due to Fourier ripples from the six surrounding irons, an oxygen at this position would require octahedral coordination, as opposed to tetrahedral coordination seen for the remaining oxygens, and the iron-oxygen bond lengths are substantially longer (2.5 Å).

Overall, we see that thioferritin utilizes a 3-fold symmetric surface, rich in carboxylate groups and unsaturated main chain carbonyls, to template nucleation of the iron oxyhydroxide core, resulting in the construction of a nascent, 3-fold symmetric iron oxyhydroxide structure whose bottom most 4-iron cluster superpositions well on the ferrihydrite structure. Thus, the 3-fold symmetry of the thioferritin cage is not a casual consequence of a hollow dodecameric cage, but is functionally critical to catalyze nucleation of the iron oxyhydroxide nanoparticle.

Nucleation Kinetics

Several smaller structural features are noted in the remainder of the Tftn structure, in particular rearrangement of C-terminal residues Lys162 to Pro172 as they transit the protein shell. This rearrangement appears to be responsible for locking Trp154, which is poorly ordered in the apo form, into a highly ordered side chain in the nucleated structure (Figure 68).

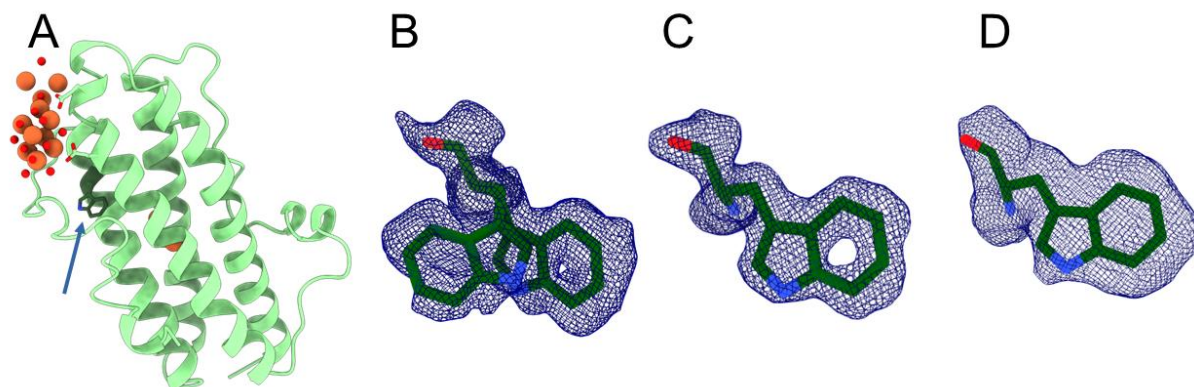


Figure 68 A) Location of Trp154 relative to iron deposition site, shown on a single subunit. B) Trp154 density from apo Pf-Tfn, depicting alternate locations A and B. Global map resolution 1.86 Å. C) Trp154 density from nucleated Pf-Tfn. Global map resolution 1.91 Å. Note the more ordered form and hole in the indole ring. D) Trp154 density from mineral loaded Pf-Tfn. The tryptophan remains in just a single conformation. Global map resolution 2.43 Å.

This suggested tryptophan fluorescence could potentially report on early events in the mineralization reaction. However, because the broad iron oxyhydroxide absorption spectrum falls on top of the tryptophan fluorescence emission (Figure 69), we instead investigated changes in fluorescence lifetimes (71-73).

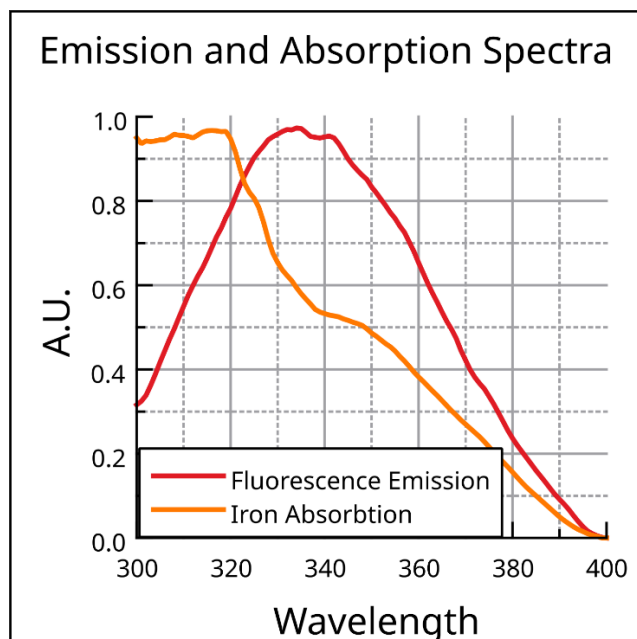


Figure 69 Tryptophan fluorescence emission overlaid with iron oxide absorption spectra, demonstrating overlap between emission and absorption that could result in an inner filter effect.

Tryptophan fluorescence lifetime (FLT) was monitored as a function of time at two different iron concentrations, for both H_2O_2 and O_2 (Figure 70).

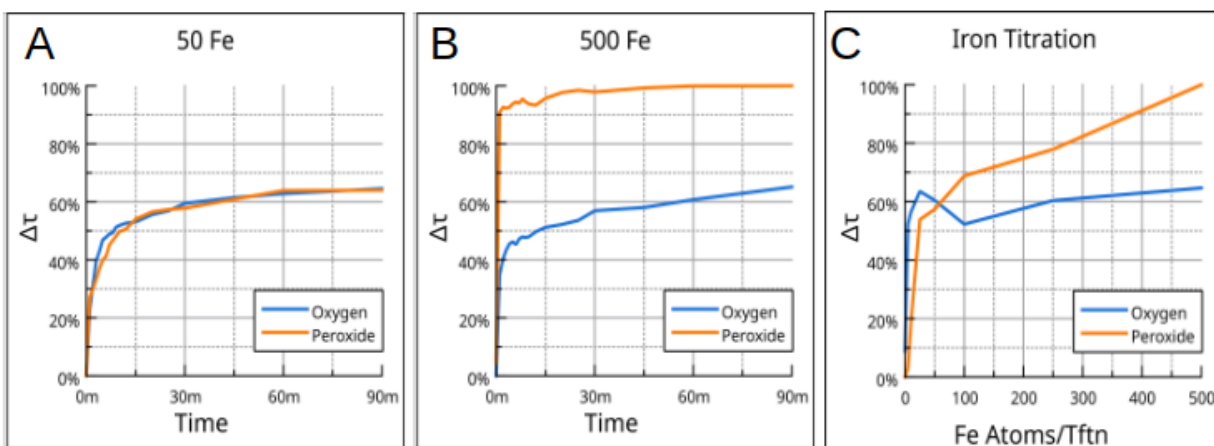


Figure 70 Fluorescence lifetime (FLT) analysis of Pf-Tftn with Fe^{2+} . A) 50 Fe/Tftn in the presence of O_2 (blue) vs H_2O_2 (orange). $\Delta\tau$ is ~65% of the maximum, regardless of substrate. B) 500 Fe^{2+} /Tftn in the presence of oxygen (blue) vs peroxide (orange). Maximum $\Delta\tau$ is achieved almost instantaneously in the presence of H_2O_2 , while the additional Fe^{2+} seems to

have little effect on $O_2 \Delta\tau$ when compared to the 50 Fe^{2+} , suggesting that the reaction may be limited by available dissolved O_2 at higher Fe^{2+} concentrations. C) Effect of iron:protein stoichiometry on $\Delta\tau$, demonstrating that even at ~ 25 iron atoms per cage, FLT is mostly composed of the nucleated waveform. Again, dissolved O_2 concentrations appear to be limiting for higher Fe^{2+} stoichiometry.

Indeed, in the presence of iron and either H_2O_2 or O_2 , we found a rapid, time dependent decrease in fluorescence lifetime (τ) of Tftn relative to the iron free controls ($\tau = 3.1$ ns). While at high iron concentrations (500 per cage) the change in FLT was greater for H_2O_2 than O_2 , at 50 irons per cage the FLT curves were similar, suggesting that O_2 concentrations become limiting at greater than 50 irons per cage.

At 500 irons per cage with H_2O_2 , the change in FLT was essentially 100% complete within a minute ($\tau = 1.6$ ns, $\Delta\tau = 1.5$ ns). Relative to this, the 50-iron experiment shows an approximate 50% change ($\Delta\tau = 0.75$ ns) at the 10-minute time point used for the structural studies above. This is consistent with the 3D classification experiment that indicated nucleated mineral at half of the 3-fold pores (above). Thus, rather than bulk mineralization, changes in fluorescence lifetime clearly reports on early events in mineralization.

We next titrated iron concentrations and measured FLT at the 90-minute time point, where the change in FLT plateaued in the above experiments. With oxygen alone, the FLT approaches the maximal change in FLT with only 25 irons, and even at half this value (12.5 per cage) is nearly 80% complete. Notably, with less than 25 irons per cage, O_2 alone shows a greater change in FLT than with addition of H_2O_2 . On the surface this might suggest O_2 is a preferred substrate, however it is likely that the addition of peroxide induces some iron precipitation in solution, reducing the iron available for mineralization within the particle at these

low iron concentrations. In this light, at low iron concentrations, the iron and peroxide curves are roughly similar.

Iron Loaded Structure

To circumvent problems associated with the precipitation of both Fe and Pf-Tftn when we attempt to more fully mineralize Tftn with peroxide, we turned to an overnight incubation with O₂. Native mass spectrometry indicated a broad distribution of iron:Tftn stoichiometry, with an average of 100 irons per cage. Visual inspection of the micrographs as well as the results of 2D classification revealed relatively heavy iron density in the center of roughly half the particles, with the remaining particles exhibiting a spectrum of decreasing iron densities, in which these smaller iron densities were no longer centered. We proceeded through initial homogenous refinement without discriminating, and then used a mask at the center of nanocavity for 3D classification to select a subset of 1,503,830 iron loaded particles. Using poses from the initial homogenous refinement, the most heavily iron loaded structure was further refined with tetrahedral averaging, to give a final map at 2.4 Å resolution (Appendix B – Figure 75).

The largest ultrastructural change is the appearance of a large, dense, roughly spherical core in the center of the cavity, approximately 30 Å in diameter. However, at higher contour levels, the strongest density is not present at the center, but instead found as 3-fold symmetric density adjacent to the inner Tftn surface in the vicinity of the acidic 3-fold pore (Figure 71).

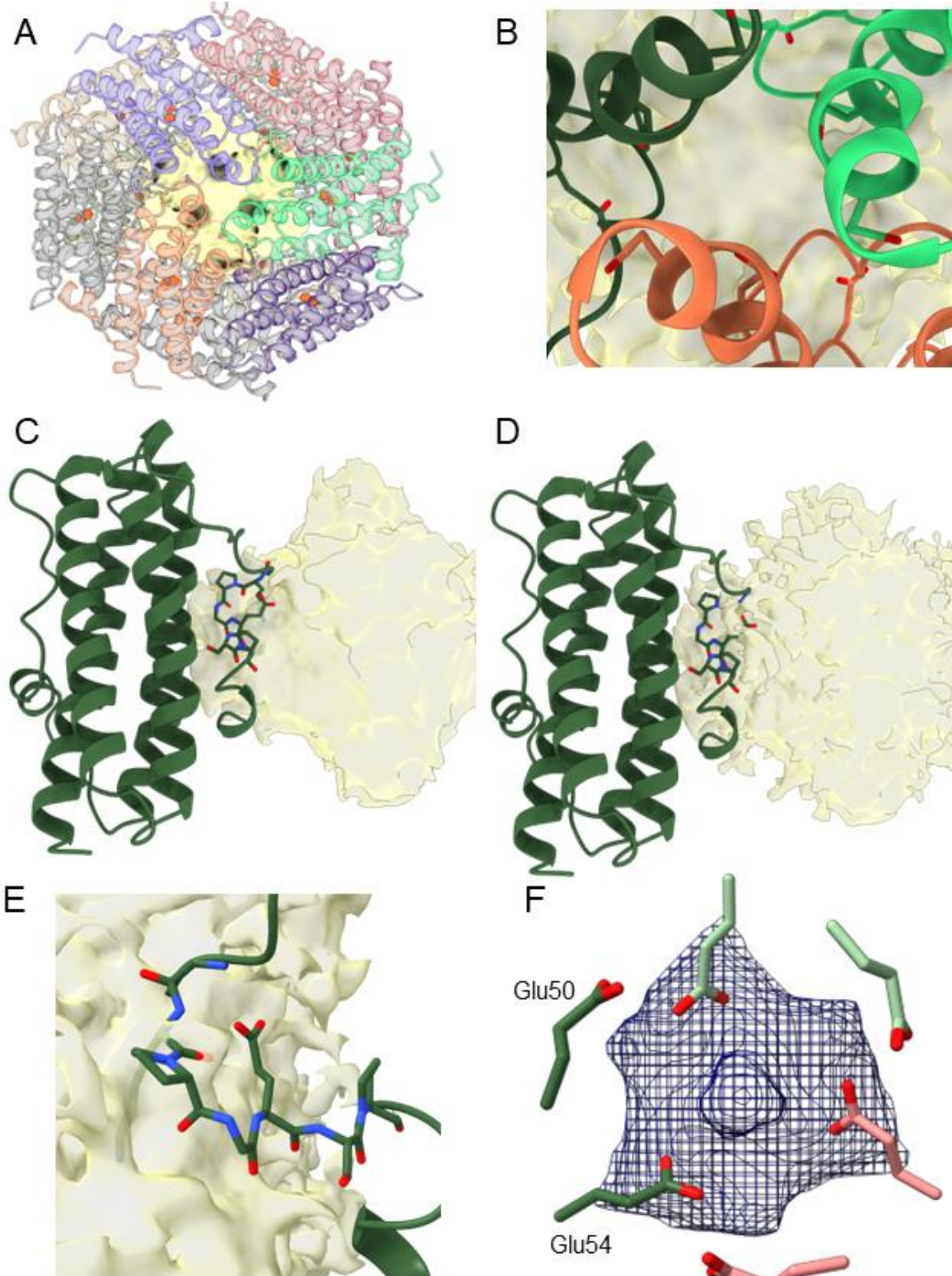


Figure 71 Iron mineral core formation in Pf-Tftn. A) Pf-Tftn dodecamer demonstrating density of large iron mineral in core. B) View down acidic pore of large mineral Tftn. Glu50 and Glu54 have rotated away from the pore. C) Single four-helix bundle and associated unsharpened mineral density depicting C-terminal interaction with large iron mineral. D) Single four-helix bundle and associated large iron mineral density, half-map sharpened, demonstrating specificity of interaction. E) Rotated view of single chain residues 170-176. Lys171 is buried deep within the core, while residues 170-176 may be forming electrostatic interactions via main chain carbonyl groups. F) View down acidic pore in intermediate iron mineral map, demonstrating no density near Glu50, but Glu54 interactions with a smaller iron mineral are preserved.

Further, the density at the acidic 3-fold pore that was so prominent in the nucleated structure is now absent, indicating the dynamic nature of iron that is initially mineralized in the 3-fold pore.

Role for the C-terminal Tail

Overall, with a few critical exceptions, the structure of the protein is largely identical to the structure of the nucleated form. This includes the C-terminal tail, which remains ordered along the interior surface of the particle, where it interacts with the mineralized core. Specifically, as the polypeptide turns away from helix D on the outside of the particle and transits the protein shell near the acidic 3-fold pore, residues 170-176 emerge on the inside of the particle and interact with the mineralized core.

Protein Mineral Interactions II

The most dramatic interaction is mediated by the Lys171 side chain, which projects directly into the mineral, where the positive charge of the ϵ -amino group presumably serves as an iron mimetic to interact with the O^{2-} anions in the mineral (Figure 71). Similarly to the nucleated structure, the negatively charged Glu173 side chain and the main chain carbonyl groups of Pro170, Gly173 and Pro176 provide additional interactions. Because the C-terminal tails emerge near the 3-fold axis, residues 170-176 are present in triplicate about the 3-fold, with 3 copies of

Lys171 lying near each other. From Lys171, the C-terminal tails then radiate away from the 3-fold. Collectively, these 21 residues coordinate the mineralized core on the interior surface of thioferritin. Further, as the contour levels are reduced, additional density radiating laterally and especially towards the center of the particle becomes apparent. This suggests subsequent mineral growth away from the C-terminal coordinated mineral towards the center of the particle.

Mineral Growth from a Subset of C-terminal Tails

To confirm this, we returned to the mineralized cryo-EM data set and selected 2D class averages with lower iron density and an asymmetric iron distribution (Appendix B – Figure 75). Following initial reconstruction and homogenous refinement to 2.15 Å with 3,313,933 particles, we used a mask to classify particles based on the presence or absence of the core iron mineral. We identified 1,504,967 particles with significant mineral density in the core. Homogeneous refinement of this dataset resulted in spurious poses, possibly due to the intensity from the mineral core. Instead, the particles were reconstructed with poses from the original, mixed species refinement, resulting in a 2.43 Å map. This particle subset underwent several rounds of 2D classification to select 125,248 particles with smaller, asymmetric mineral core density. These particles were symmetry-expanded by D2, and a mask was placed around the end of the C-terminal tail for subsequent 3D classification, identifying 167,240 particles with significant iron density at a specific 3-fold. Homogenous refinement with 3-fold (C3) symmetry then gave a 3.4 Å map with strong density for the C-terminal coordinated mineral. In support of the above hypothesis, this map also lacked density at the center of the particle and the remaining 3-folds.

Dynamic Pore Structures

This structure, with its intermediate iron content relative to the nucleated and iron loaded structures described above, revealed one additional surprise. While there was strong density for the C-terminal mineral, and the map lacked density for the irons coordinated by Leu47 and Glu50 in the nucleated structure, density for the Glu54 coordinated irons was still present (Figure 71, Panel F). Further, significant solvent space separates the C-terminal coordinated mineral from the Glu54 coordinated irons. Thus, in the initial nucleated structure, iron is present predominantly within the 3-fold acidic pores. In this intermediate structure, iron must move out of the pores, as density for the iron-oxo clusters at Leu47 and Glu50 is lost, but still present for Glu54. Concomitantly, mineralized iron coordinated by C-terminal residues 170-176 becomes the dominant feature. As iron content increases further, the Glu54 iron-oxo cluster is also lost, while the mineralized core grows into the center of the particle from one or more constellations of the 3-fold symmetric C-terminal tails. From this perspective, Trp154 remains ordered in the intermediate iron, high iron structure. Thus, while Trp154 initially reports on the earliest events in iron mineralization, once it adopts the ordered, shorter lifetime form, it remains there regardless of iron content in the acidic 3-fold pore. This is consistent with the appearance of the reordered C-terminal residues during nucleation, which are then be locked into place as residues 170-176 subsequently coordinate the mineral.

The Ferroxidase Center

The iron loaded structure also shows significant change around the ferroxidase center. Density is still strong for the two oxo-bridged irons and their coordinating side chains, but additional strong, spherical density is now apparent above them in the solvent channel leading

into this site from the exterior surface. In addition, there are several breaks in the main chain density along the solvent channels, suggesting a conformationally dynamic structure. Despite this, density for the cysteine side chains in the thioferritin motif is still clear. The residues are reduced, and Cys118 is present in two different rotamers (Figure 72).

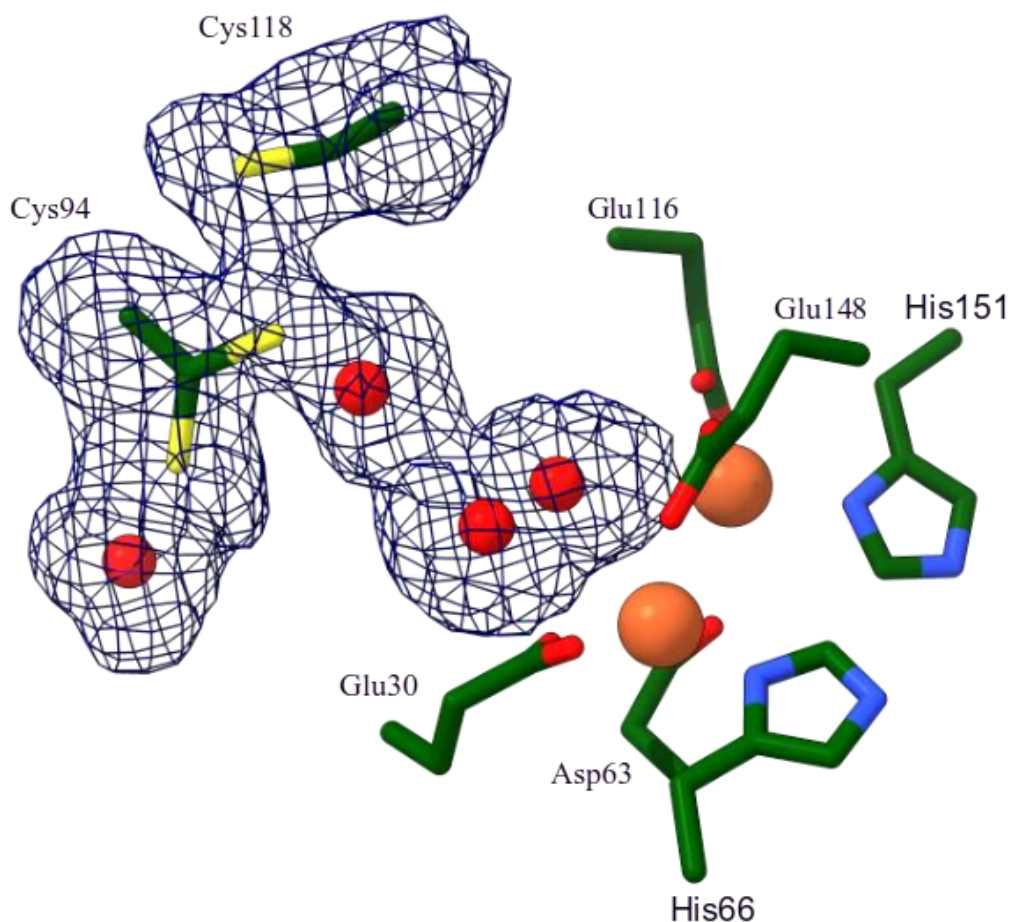


Figure 72. Electron density occupying active site channel near in holo map, near conserved cysteine pair. Density is modeled as water for convenient representation, although no experimental evidence suggests it must be water.

The first has been seen before, with the side chain extending towards Cys94, hydrogen-bonded to the sulfur atom. However, a new conformation also appears, in which the Cys118 side

chain instead points down towards the di-iron site, where it coordinates unmodeled spherical density, potentially an oxo-species or mobile iron juxtaposed between Cys118 and the ferroxidase center. In addition, Asp63 in the ferroxidase center is also present in a second conformation, in which the side chain swings away from the irons, further indicating a dynamic situation at the ferroxidase center itself.

Discussion

A Cryo-EM Standard

Pf-Tftn cloned into the pET-30 vector shows significant leaky expression, allowing Pf-Tftn to be expressed in simple LB media without induction (62). Purification is also straight forward, with heat denaturation at 85 °C for 10 minutes followed by size exclusion chromatography on the cleared lysate, yielding ~10 mg of purified Tftn per gram of cell pellet (62). In addition, purified Pf-Tftn can be stored at -80 °C if needed, with ~90% recovery when thawed. Pf-Tftn is thus an extremely convenient protein to express and purify for structural studies. Given the extremely high-resolution structure presented here, determined on a 200 kV instrument no less, Pf-Tftn might make for a convenient standard for cryo-EM, much like the often used horse spleen H-chain ferritin, but one that is significantly easier to handle.

Strategies for Single Particle Analysis of Iron Loaded Ferritins

Notably, *ab initio* alignment and refinement of iron-loaded particles was difficult, likely due to the high amplitude contrast of the iron core, and heterogeneity in the size, shape and location of the mineral within the particles. Even the largest datasets routinely failed to give high resolution details for the protein main chain. Instead, alignments seemed to over-weight the

mineral core, and higher density regions along the 2-fold axes. This problem could be mitigated in a variety of ways, although not without some tradeoffs. In a mixed particle dataset with roughly equal proportions of empty and mineralized particles, the two species refined nicely together to high resolution. At that point, particles could be classified in either 2D or 3D, and the mineral-containing particles alone could be reconstructed using the poses from the original mixed refinement, giving maps to intermediate or high resolution. Alternatively, the mineral core could be subtracted prior to refinement, although strong residual density along the 2-fold axes of the particles still lead to spurious alignments in some cases. Subtracted particles, once refined, can then be re-extracted with iron, and reconstructed using the refined particle poses from the particle subtracted structure. However, this typically resulted in lower-than expected resolution, relative to reconstruction with an identical number of empty particles. In any event, and perhaps contrary to expectations, an initial reconstruction using a heterogenous mix of both empty and full particles gave better starting poses for the iron containing particles, which could then be selected and taken forward for the iron loaded structures. We believe the inclusion of the empty particles indirectly focuses alignment on the protein shell, as it is the common feature across all particles, whereas ab initio solutions using only the iron containing particles are dominated by the high amplitude, heterogenous features of the mineralized iron core.

Tryptophan Fluorescence

We also noted that Trp154 remains well ordered in the iron loaded structure. Thus, while Trp initially reports on the earliest events in iron mineralization, once it adopts the shorter lifetime, ordered form, it remains there regardless of iron content in the acidic 3-fold pore. This is consistent with the conformational reordering C-terminal residues 162-170 during nucleation,

which is then apparently locked into place as residues 170-176 subsequently coordinate the mineral. For this reason, as iron departs the acidic three-fold pore, the fluorescent life time is not expected to return to the shorter lifetime species seen in apo-Pf-Tftn.

Nucleation

While a multitude of ferritin, bacterioferritin and DPS structures show iron density in the di-iron ferroxidase center, a high-resolution structure that resolves the ferric oxyhydroxide mineral have been elusive. Indeed, to our knowledge the nucleated mineral structure described here represents the first such 3-dimensional mineral structure for any member of the ferritin family. Our structure implicates the 3-fold symmetry of the pore, with its symmetry match to ferrihydrite, in catalyzing nucleation of the ferric oxyhydroxide mineral. And because the acidic 3-fold pore is a conserved feature in the thioferritins (46, 47), this mechanism for biomineral nucleation is likely to be conserved across the thioferritin family. It may also extend to mini-ferritins in general, as acidic residues are also present in the 3-fold “B-type” or “ferritin-like” pores in DPS (41, 58, 74, 75). However, in DPS, the 3-fold pore is clearly implicated in iron entry (76-78), and whether the acidic 3-fold pore serves in both capacities in DPS (or Tftn) remains an open question. Regarding nucleation in the larger 24-mer ferritin and bacterioferritin particles, a recent cryo-EM study with *S. coelicolor* bacterioferritin reported the presence of small Fe-biomineral assemblies located close to the fourfold channels on the inner surface of the shell (44). Although the structure of the biomineral was not resolved in this work, nor was a mechanism discussed, it was suggested that the fourfold channel serves to seed nucleation in bacterioferritin. In addition, crystallographic studies have identified locally ordered 2-D iron clusters on the interior surfaces of several ferritins and DPS (79-82), and these

have also been suggested as possible sites of nucleation, although ordered metals might also arise through interaction with an interior surface that is naturally designed as a structural and electrostatic complement to the biomineral.

Biomineralization

Using particles from the iron loaded data set, we were also able to resolve structures with an intermediate amount of iron, and a more fully loaded particle. Both structures show the appearance of mineralization coordinated by a constellation of three C-terminal tails arranged about a single 3-fold axis. Importantly, the intermediate iron loaded structure refined with C1 symmetry strongly suggests that this density is present at only a subset of the 4 possible locations in each particle, i.e, that growth with respect to each of the 4, C-terminal clusters is asymmetric. In contrast, the C1 reconstruction also suggests the mineral appears symmetrically within a given cluster. Thus, once again, the 3-fold symmetry of the particle appears integral to the process of mineralization.

In addition, as the mineralized iron coordinated by C-terminal residues 170-176 becomes the dominant feature, mineral content in the pores is diminished, first with loss of Leu47/Glu50 coordinated iron in the intermediate form, followed by loss of the Glu54 coordinated mineral in the loaded form. Again, the C1 and C3 reconstructions are valuable, showing that the Glu54 iron is present at the same 3-fold as the C-terminally coordinated mineral in the intermediate iron particles, and that while they are in close proximity, they appear to be separated by solvent. Finally, as we transition from the intermediate to more fully loaded structures, we see mineral growth in the lateral and axial directions, especially toward the center of the particle.

The use of negatively charged side chains and main chain carbonyls to coordinate the mineral are a common theme for the acidic 3-fold pore and the C-terminal tail. However, the C-terminal tail also utilizes positively charged lysine and arginine side chains as well, to help coordinate the checkerboard pattern of polar hydroxyl groups, as well as the positive and negative charge on the mineral surface. With local resolution at the protein/mineral interface at ~ 2.5 Å, the structure of the protein is quite well resolved and these interactions are very clear, and in the case of Lys171, quite dramatic. Thus, we also see that the C-terminal tails do not randomly insert themselves into the interior of the iron mineral, but are instead largely excluded and present in well-ordered structure at the protein/mineral interface.

Interestingly, the C-terminal tail was found on the exterior of the particle in the *B. fragilis* thioferritin structure (47), suggesting an alternative mechanism for biomineralization in the bacterial thioferritin subclass. In contrast, is the cryo-EM study on *S. coelicolor* bacterioferritin discussed above (Nucleation), which also determined the structure of an iron loaded particle at 3.27 Å resolution when refined with octahedral (432) symmetry, and examined structures refined with C1 symmetry as well. Like our work with Pf-Tftn, they too find asymmetric growth of the mineral from the particle surface toward the center, and also report an interaction with the C-terminal tail. While the interaction is apparently not well resolved, and details on the role of specific C-terminal residues are not reported, it seems the C-terminal tail may be expected to play an important role in mineralizing other members of the ferritin family as well.

Structure of the Mineralized Core

The local resolution of the mineral surface however, quickly falls off to ~ 4 -5 Å. For this reason, we have not attempted to model the bulk of the mineral core itself. However, we note

that if the mineralized core is indeed present as microcrystalline ferrihydrite, and the three-fold axes of thioferritin align with a 3-fold symmetry element in the ferrihydrite core, refinement of the mineral structure with C3 or tetrahedral symmetry could potentially reinforce the local ferrihydrite structure in the neighborhood of the C-terminal tails, especially if the particles were only partially full, with limited iron content near the center of the particles. For this reason, careful crystallization of ferrihydrite in thioferritin might potentially allow the structure of the core itself to be visualized in the future.

Methods

Expression and Purification of Pf-Tftn

Pf-Tftn was expressed and purified as previously described (62). Briefly, *E. coli* BL21 cells were transformed with 5 ng of vector, and plated onto agar plates with 30 mg/mL kanamycin, and grown overnight at 37 °C. A fresh colony was picked and used to inoculate 5 mL of lysogeny broth (LB) media (1% tryptone w/v, 0.5% yeast extract w/v, 0.5% NaCl w/v), in a 15 mL plastic centrifuge tube with the lid loosened, in a shaking 37 °C incubator for 12 hours. 750 µL of the overnight culture was used to inoculate 750 mL cultures of LB media with 30 mg/mL kanamycin, which were grown in 3.5 L baffled flasks at 37 °C for 18 hours, with no induction. The cells were harvested via centrifugation in a Beckman Model J-6B centrifuge using a JS-5.2 swinging bucket rotor at 3000 RCF for 30 minutes. The pellets were stored at -20 °C until subsequent purification.

As needed, pellets containing expressed Pf-Tftn were thawed, resuspended in 100 mM NaCl and 50 mM MES at pH 6.5 with, at a ratio of 5 mL buffer per 1 g of cell pellet, and then lysed in a French press. The lysate was immediately spun at 30,000 RCF for 45 minutes in a

Sorvall RC-2B equipped with a Fiberlite F21 fixed angle rotor. The soluble fraction was then heated at 85 °C for 10 minutes, before being spun again. After centrifugation, the cleared lysate was divided into 1 mL aliquots and frozen at -80 °C. When needed, aliquots were thawed, filtered with a 0.22 µm PVDF filter, and passed over a Superose 6 size exclusion column equilibrated with 100 mM NaCl in 50 mM MES at pH 6.5. Tftn eluted at a volume of 14.5 mL, corresponding to a molecular weight of approximately 280 kDa. Purified Tftn was concentrated to 2.5 mg/mL (measured via Bradford assay (83)), before being frozen again at -80 °C. For all downstream assays, Pf-Tftn aliquots were thawed and centrifuged at 16,000 RCF for 5 minutes in a microcentrifuge, and the supernatant was drawn off for subsequent use.

Iron Loading

For iron loading assays, protein was diluted to a final concentration of 0.2 mg/mL in 50 mM MES, 100 mM NaCl, pH 6.5 buffer. FeSO₄ was dissolved into 0.1% HCl, and stoichiometric amounts were mixed with the diluted protein. Final concentrations were always calculated in ratio of Fe²⁺ to protein dodecamer. For assays including hydrogen peroxide, ratios were 0.5 H₂O₂: 1.0 FeSO₄.

Vitrification

Upon vitrification, Pf-Tftn exhibits a preference for carbon grid supports when vitrified in low ionic strength conditions. However, increasing the salt concentration to 100 mM causes Pf-Tftn to partition into the ice, where at sufficient concentrations it may form close-packed arrays with a 9 nm particle spacing, reminiscent of the hexagonal lattices previously observed in DPS-DNA biocrystals (49, 84) (Appendix B – Figure 73). At high salt concentrations, 4 µL of protein at 2 mg/mL protein was applied to either Quantifoil 300 Mesh Cu/Carbon R 1.2/1.3 or

Au-flat 300 Mesh 1.2/1.3 (EMS AUFT313-50) grids. The samples were blotted stringently and plunge frozen into liquid ethane using a Vitrobot Mark IV (Thermo Scientific). Upon freezing, the samples were clipped into autogrids and stored in liquid nitrogen until analyzed.

Cryo-TEM Data Collection

The prepared grids were loaded into our home 200 kV Talos Arctica G2 (Thermo Scientific) and imaged with the K3 camera (Gatan). Data collections were performed using SerialEM (apo, nucleated), and SmartScope (holo), using a 5x5 multi-shot scheme, or beam-image-shift (BIS) distance of 7.5 μm , where appropriate (52, 53). The apo data were collected at 0.345 $\text{\AA}/\text{pixel}$ in super resolution mode. The holo and nucleated datasets were collected at 0.69 $\text{\AA}/\text{pixel}$, without using super resolution. The dose for all datasets was 55 $\text{e}/\text{\AA}^2$. A defocus range of -0.6 to -1.5 microns was used for all data collections.

Single Particle Analysis for Apo-Pf-Tftn

A total of 27,973 movies were recorded, and underwent patch motion correction and patch CTF estimation in CryoSPARC Live (13). Movies were discarded based on quality and resolution of the CTF fit, calculated defocus, total motion, and relative ice thickness, resulting in 13,996 movies kept. After data collection, 100 micrographs were picked using the CryoSPARC blob picker, using an 80-100 $\text{\AA}$ blob size. Particles were extracted, and 2D classification was performed in CryoSPARC. Particles were selected that contributed to 2D classes that exhibited visual secondary structure. Those particles were used to train a crYOLO model that was then used to pick particles in all remaining movies (85). Particles were extracted again and re-classified in 2D. Obvious junk classes were removed, and 200,000 particles were used to perform a 2 class Ab initio reconstruction with enforced tetrahedral symmetry, generating a good

map, and a junk map. Heterogeneous refinement of all particles against the two maps was performed iteratively in CryoSPARC, each time selecting just the particles that aligned well to the good map for the next iteration. After three iterations, the remaining 2,583,454 particles were put into homogeneous refinement with enforced tetrahedral symmetry. The aligned particle stack then underwent two iterations of global CTF refinement in CryoSPARC, followed by local CTF refinement, before a final homogeneous refinement that enforced tetrahedral symmetry.

An initial homology model was generated using Phyre2 (86). The homology model was symmetry expanded onto the existing *S. solfataricus* Tftn structure (PDB 2CLB) using ChimeraX (46, 87) and docked into the experimental half maps using the Phenix ChimeraX module, `emplace_local` (88, 89). The consensus map was then sharpened using Phenix `b_iso_to_d_cut`, and the docked model was iteratively rebuilt and refined against the sharpened map using Coot and Phenix `real_space_refine` (88-90).

Single Particle Analysis for Nucleated Mineral Dataset

The workflow for the analysis of the 50 Fe dataset was identical to the apo dataset as far as the first model from homogenous refinement. A total of 3915 exposures were recorded. 3094 of these exposures were selected based on quality metrics, with 1,742,467 particles going into the homogenous refinement, giving an initial structure at 2.33 Å resolution by gold standard FSC. This map showed clear density in the 3-fold acidic pores corresponding to the presence of mineralized iron. The particles were then aligned to the D2 symmetry axes (x, y, z) and symmetry expanded by 222 point groups symmetry, giving 4 virtual particles per real-particle, for a total of 6,969,868 virtual particles, each one representing a unique 3-fold interface. A mask was then placed around the 3-fold interface near the Glu50 residues on the z axis, and

CryoSPARC 3D Classification was performed on that region. Particles were divided into two classes based on the presence or absence of the mineral in the masked region. Density was detected in the masked region for 3,584,523 virtual particles, indicating that roughly half the acidic 3-fold pores contained mineral. Symmetry expansion was reversed, and 1,619,884 particles remained. Those particles were then refined with T symmetry, for a final structure at 1.91 Å by GSFSC.

Single Particle Analysis for Holo-Pf-Tftn

A total of 13,577 exposures were recorded. 9094 of these exposures were selected based on quality metrics. 3,153,592 particles were used in the consensus reconstruction which refined to 2.15 Å resolution by gold standard Fourier shell correlation (GSFSC). The work-flow up through initial homogenous reconstruction was identical to the apo dataset. Protein density was then masked out, and CryoSPARC 3D Classification was used to sort particles based on the presence (1,504,967) or absence (1,650,473) of a strong mineral core. Particles with mineral cores were then reconstructed using poses from the initial homogenous refinement. The resolution of this reconstruction was 2.43 Å by GSFSC. For the intermediate iron density reconstructions, the mineralized particles were then classified in 2D, and classes that exhibited only partial mineral density in the core were selected, giving 125,320 particles. The mineral core was then subtracted, and the poses were refined using tetrahedral symmetry to a resolution of 3.14 Å by GSFSC. Once refined, the original particles were re-extracted and re-constructed using the updated poses and C1 symmetry, to a resolution of 3.54 Å by GSFSC.

Fluorescence Lifetime

Iron at the indicated concentrations was titrated across a 96-well quartz microplate (Hellma) containing 0.1 mg/mL Pf-Tftn in 50 mM MES, pH 6.5 and 100 mM NaCl (final concentrations) with total well volumes of 100 μ l, with three technical replicates per plate for each experiment. Wells were excited at λ =280 nm and time-resolved emission of individual wells followed at λ =345 nm using a NovaFluor fluorescence lifetime spectrometer over the time-course of the experiment (Fluorescence Innovations) (91, 92). The waveform at each specific experimental condition and time point (0, 1, 2 ... minutes) was recorded over a total decay time of 128 ns at 2-ps intervals. The data were then fit to an exponential decay function with GraphPadPrism [$I(t) = I_{oe}-t/\tau$] to determine tau for each specific time point. The Δ -tau was determined by subtracting tau for each experimental condition from the tau for the iron free data. 100% Δ -tau was taken as Δ -tau for the 500 Fe data at 90 minutes.

REFERENCES CITED

1. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis *Nature* 181, 662-666 10.1038/181662a0
2. Jones, N. (2014) Crystallography: Atomic secrets *Nature* 505, 602-603 10.1038/505602a
3. Wang, H. W., and Wang, J. W. (2017) How cryo-electron microscopy and X-ray crystallography complement each other *Protein Sci* 26, 32-39 10.1002/pro.3022
4. Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013) Structure of the TRPV1 ion channel determined by electron cryo-microscopy *Nature* 504, 107-112 10.1038/nature12822
5. Frank, J. (1975) Averaging of low exposure electron micrographs of non-periodic objects *Ultramicroscopy* 1, 159-162 10.1016/s0304-3991(75)80020-9
6. Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, A. W. et al. (1988) Cryo-electron microscopy of vitrified specimens *Q Rev Biophys* 21, 129-228 10.1017/s0033583500004297
7. Bai, X. C., Fernandez, I. S., McMullan, G., and Scheres, S. H. (2013) Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles *Elife* 2, e00461 10.7554/eLife.00461
8. Fernandez-Leiro, R., and Scheres, S. H. (2016) Unravelling biological macromolecules with cryo-electron microscopy *Nature* 537, 339-346 10.1038/nature19948
9. Frank, J., Goldfarb, W., Eisenberg, D., and Baker, T. S. (1978) Reconstruction of glutamine synthetase using computer averaging *Ultramicroscopy* 3, 283-290 10.1016/s0304-3991(78)80038-2
10. Hewat, E. A., and Neumann, E. (2002) Characterization of the performance of a 200-kV field emission gun for cryo-electron microscopy of biological molecules *J Struct Biol* 139, 60-64 10.1016/s1047-8477(02)00515-4
11. McMullan, G., Faruqi, A. R., and Henderson, R. (2016) Direct Electron Detectors *Methods Enzymol* 579, 1-17 10.1016/bs.mie.2016.05.056
12. Peng, R., Fu, X., Mendez, J. H., Randolph, P. S., Bammes, B. E., and Stagg, S. M. (2023) Characterizing the resolution and throughput of the Apollo direct electron detector *J Struct Biol X* 7, 100080 10.1016/j.jsbx.2022.100080
13. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions *J Struct Biol* 128, 82-97 10.1006/jsbi.1999.4174

14. Kimanius, D., Forsberg, B. O., Scheres, S. H., and Lindahl, E. (2016) Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2 *Elife* 5, 10.7554/eLife.18722
15. Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination *Nat Methods* 14, 290-296 10.1038/nmeth.4169
16. Chiu, W., Schmid, M. F., Pintilie, G. D., and Lawson, C. L. (2021) Evolution of standardization and dissemination of cryo-EM structures and data jointly by the community, PDB, and EMDb *J Biol Chem* 296, 100560 10.1016/j.jbc.2021.100560
17. Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy *J Mol Biol* 213, 899-929 10.1016/S0022-2836(05)80271-2
18. Heymann, J. B. (2022) The progressive spectral signal-to-noise ratio of cryo-electron micrograph movies as a tool to assess quality and radiation damage *Comput Methods Programs Biomed* 220, 106799 10.1016/j.cmpb.2022.106799
19. Li, X., Zheng, S. Q., Egami, K., Agard, D. A., and Cheng, Y. (2013) Influence of electron dose rate on electron counting images recorded with the K2 camera *J Struct Biol* 184, 251-260 10.1016/j.jsb.2013.08.005
20. Frank, J. (2006) Three-dimensional electron microscopy of macromolecular assemblies : visualization of biological molecules in their native state, 2nd Ed., Oxford University Press, Oxford ; New York
21. Rohou, A., and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs *J Struct Biol* 192, 216-221 10.1016/j.jsb.2015.08.008
22. Scheres, S. H. (2014) Beam-induced motion correction for sub-megadalton cryo-EM particles *Elife* 3, e03665 10.7554/eLife.03665
23. Ripstein, Z. A., and Rubinstein, J. L. (2016) Processing of Cryo-EM Movie Data *Methods Enzymol* 579, 103-124 10.1016/bs.mie.2016.04.009
24. Zivanov, J., Nakane, T., and Scheres, S. H. W. (2019) A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis *IUCrJ* 6, 5-17 10.1107/S205225251801463X
25. Chen, M., and Ludtke, S. J. (2021) Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM *Nat Methods* 18, 930-936 10.1038/s41592-021-01220-5

26. Marino, J., and Schertler, G. F. X. (2021) A set of common movements within GPCR-G-protein complexes from variability analysis of cryo-EM datasets *J Struct Biol* 213, 107699 10.1016/j.jsb.2021.107699
27. Baid, P., and Sengupta, J. (2023) Cryo-EM captures a unique conformational rearrangement in 23S rRNA helices of the Mycobacterium 50S subunit *Int J Biol Macromol* 253, 126876 10.1016/j.ijbiomac.2023.126876
28. Grunewald, K., Desai, P., Winkler, D. C., Heymann, J. B., Belnap, D. M., Baumeister, W. et al. (2003) Three-dimensional structure of herpes simplex virus from cryo-electron tomography *Science* 302, 1396-1398 10.1126/science.1090284
29. Briggs, J. A., Grunewald, K., Glass, B., Forster, F., Krausslich, H. G., and Fuller, S. D. (2006) The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions *Structure* 14, 15-20 10.1016/j.str.2005.09.010
30. Mastronarde, D. N., and Held, S. R. (2017) Automated tilt series alignment and tomographic reconstruction in IMOD *J Struct Biol* 197, 102-113 10.1016/j.jsb.2016.07.011
31. Ludtke, S. J. (2016) Single-Particle Refinement and Variability Analysis in EMAN2.1 *Method Enzymol* 579, 159-189 10.1016/bs.mie.2016.05.001
32. Zhao, C., Lu, D., Zhao, Q., Ren, C., Zhang, H., Zhai, J. et al. (2023) Computational methods for in situ structural studies with cryogenic electron tomography *Front Cell Infect Microbiol* 13, 1135013 10.3389/fcimb.2023.1135013
33. Cope, J., Heumann, J., and Hoenger, A. (2011) Cryo-electron tomography for structural characterization of macromolecular complexes *Curr Protoc Protein Sci Chapter 17, Unit17* 13 10.1002/0471140864.ps1713s65
34. Tegunov, D., Xue, L., Dienemann, C., Cramer, P., and Mahamid, J. (2021) Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells *Nat Methods* 18, 186-193 10.1038/s41592-020-01054-7
35. Ni, T., Frosio, T., Mendonca, L., Sheng, Y., Clare, D., Himes, B. A. et al. (2022) High-resolution in situ structure determination by cryo-electron tomography and subtomogram averaging using emClarity *Nat Protoc* 17, 421-444 10.1038/s41596-021-00648-5
36. Chaillet, M. L., van der Schot, G., Gubins, I., Roet, S., Veltkamp, R. C., and Forster, F. (2023) Extensive Angular Sampling Enables the Sensitive Localization of Macromolecules in Electron Tomograms *Int J Mol Sci* 24, 10.3390/ijms241713375
37. Henle, E. S., Luo, Y., and Linn, S. (1996) Fe²⁺, Fe³⁺, and oxygen react with DNA-derived radicals formed during iron-mediated Fenton reactions *Biochemistry* 35, 12212-12219 10.1021/bi961235j

38. Wardman, P., and Candeias, L. P. (1996) Fenton chemistry: an introduction *Radiat Res* 145, 523-531, <https://www.ncbi.nlm.nih.gov/pubmed/8619017>
39. Ford, G. C., Harrison, P. M., Rice, D. W., Smith, J. M., Treffry, A., White, J. L. et al. (1984) Ferritin: design and formation of an iron-storage molecule *Philos Trans R Soc Lond B Biol Sci* 304, 551-565 10.1098/rstb.1984.0046
40. Andrews, S. C., Arosio, P., Bottke, W., Briat, J. F., von Darl, M., Harrison, P. M. et al. (1992) Structure, function, and evolution of ferritins *J Inorg Biochem* 47, 161-174 10.1016/0162-0134(92)84062-r
41. Grant, R. A., Filman, D. J., Finkel, S. E., Kolter, R., and Hogle, J. M. (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA *Nat Struct Biol* 5, 294-303 10.1038/nsb0498-294
42. Ilari, A., Stefanini, S., Chiancone, E., and Tsernoglou, D. (2000) The dodecameric ferritin from *Listeria innocua* contains a novel intersubunit iron-binding site *Nat Struct Biol* 7, 38-43 10.1038/71236
43. Haikarainen, T., and Papageorgiou, A. C. (2010) Dps-like proteins: structural and functional insights into a versatile protein family *Cell Mol Life Sci* 67, 341-351 10.1007/s00018-009-0168-2
44. Jobichen, C., Ying Chong, T., Rattinam, R., Basak, S., Srinivasan, M., Choong, Y. K. et al. (2023) Bacterioferritin nanocage structures uncover the biomineralization process in ferritins *PNAS Nexus* 2, pgad235 10.1093/pnasnexus/pgad235
45. Wiedenheft, B., Mosolf, J., Willits, D., Yeager, M., Dryden, K. A., Young, M. et al. (2005) An archaeal antioxidant: characterization of a Dps-like protein from *Sulfolobus solfataricus* *Proc Natl Acad Sci U S A* 102, 10551-10556 10.1073/pnas.0501497102
46. Gauss, G. H., Benas, P., Wiedenheft, B., Young, M., Douglas, T., and Lawrence, C. M. (2006) Structure of the DPS-like protein from *Sulfolobus solfataricus* reveals a bacterioferritin-like dimetal binding site within a DPS-like dodecameric assembly *Biochemistry* 45, 10815-10827 10.1021/bi060782u
47. Gauss, G. H., Reott, M. A., Rocha, E. R., Young, M. J., Douglas, T., Smith, C. J. et al. (2012) Characterization of the *Bacteroides fragilis* bfr gene product identifies a bacterial DPS-like protein and suggests evolutionary links in the ferritin superfamily *J Bacteriol* 194, 15-27 10.1128/JB.05260-11
48. Su, M., Cavallo, S., Stefanini, S., Chiancone, E., and Chasteen, N. D. (2005) The so-called *Listeria innocua* ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties *Biochemistry* 44, 5572-5578 10.1021/bi0472705

49. Theil, E. C., Matzapetakis, M., and Liu, X. (2006) Ferritins: iron/oxygen biominerals in protein nanocages *J Biol Inorg Chem* 11, 803-810 10.1007/s00775-006-0125-6
50. Michel, F. M., Hosein, H. A., Hausner, D. B., Debnath, S., Parise, J. B., and Strongin, D. R. (2010) Reactivity of ferritin and the structure of ferritin-derived ferrihydrite *Biochim Biophys Acta* 1800, 871-885 10.1016/j.bbagen.2010.05.007
51. Chesnokov, Y., Mozhaev, A., Kamyshinsky, R., Gordienko, A., and Dadinova, L. (2022) Structural Insights into Iron Ions Accumulation in Dps Nanocage *Int J Mol Sci* 23, 10.3390/ijms23105313
52. Mastronarde, D. N. (2005) Automated electron microscope tomography using robust prediction of specimen movements *J Struct Biol* 152, 36-51 10.1016/j.jsb.2005.07.007
53. Bouvette, J., Huang, Q., Riccio, A. A., Copeland, W. C., Bartesaghi, A., and Borgnia, M. J. (2022) Automated systematic evaluation of cryo-EM specimens with SmartScope *Elife* 11, 10.7554/eLife.80047
54. Mastronarde, D. N. Script commands
55. Kim, P. T., Noble, A. J., Cheng, A., and Bepler, T. (2023) Learning to automate cryo-electron microscopy data collection with Ptolemy *IUCrJ* 10, 90-102 10.1107/S2052252522010612
56. OU & Regional Research Store (OURRstore)
57. Andrews, S. C. (2010) The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor *Biochim Biophys Acta* 1800, 691-705 10.1016/j.bbagen.2010.05.010
58. Bozzi, M., Mignogna, G., Stefanini, S., Barra, D., Longhi, C., Valenti, P. et al. (1997) A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in *Listeria innocua* *J Biol Chem* 272, 3259-3265 10.1074/jbc.272.6.3259
59. Chiancone, E., Ceci, P., Ilari, A., Ribacchi, F., and Stefanini, S. (2004) Iron and proteins for iron storage and detoxification *Biometals* 17, 197-202 10.1023/b:biom.0000027692.24395.76
60. Bevers, L. E., and Theil, E. C. (2011) Maxi- and mini-ferritins: minerals and protein nanocages *Prog Mol Subcell Biol* 52, 29-47 10.1007/978-3-642-21230-7_2
61. Chasteen, N. D., and Harrison, P. M. (1999) Mineralization in ferritin: an efficient means of iron storage *J Struct Biol* 126, 182-194 10.1006/jsbi.1999.4118
62. Ramsay, B., Wiedenheft, B., Allen, M., Gauss, G. H., Lawrence, C. M., Young, M. et al. (2006) Dps-like protein from the hyperthermophilic archaeon *Pyrococcus furiosus* *J Inorg Biochem* 100, 1061-1068 10.1016/j.jinorgbio.2005.12.001

63. Maaty, W. S., Wiedenheft, B., Tarlykov, P., Schaff, N., Heinemann, J., Robison-Cox, J. et al. (2009) Something old, something new, something borrowed; how the thermoacidophilic archaeon *Sulfolobus solfataricus* responds to oxidative stress PLoS One 4, e6964, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19759909
64. Lundin, D., Poole, A. M., Sjoberg, B. M., and Hogbom, M. (2012) Use of structural phylogenetic networks for classification of the ferritin-like superfamily J Biol Chem 287, 20565-20575 10.1074/jbc.M112.367458
65. Hong, Y., Wang, G., and Maier, R. J. (2006) *Helicobacter hepaticus* Dps protein plays an important role in protecting DNA from oxidative damage Free Radic Res 40, 597-605 10.1080/10715760600618882
66. Bhattacharyya, G., and Grove, A. (2007) The N-terminal extensions of *Deinococcus radiodurans* Dps-1 mediate DNA major groove interactions as well as assembly of the dodecamer J Biol Chem 282, 11921-11930 10.1074/jbc.M611255200
67. Michel, F. M., Ehm, L., Antao, S. M., Lee, P. L., Chupas, P. J., Liu, G. et al. (2007) The structure of ferrihydrite, a nanocrystalline material Science 316, 1726-1729 10.1126/science.1142525
68. Manceau, A., Skanthakumar, S., and Soderholm, L. (2014) PDF analysis of ferrihydrite: Critical assessment of the under-constrained akdalaite model Am Mineral 99, 102-108 10.2138/am.2014.4576
69. Kauko, A., Pulliainen, A. T., Haataja, S., Meyer-Klaucke, W., Finne, J., and Papageorgiou, A. C. (2006) Iron incorporation in *Streptococcus suis* Dps-like peroxide resistance protein Dpr requires mobility in the ferroxidase center and leads to the formation of a ferrihydrite-like core J Mol Biol 364, 97-109 10.1016/j.jmb.2006.08.061
70. Sassi, M., Chaka, A. M., and Rosso, K. M. (2021) Ab initio thermodynamics reveals the nanocomposite structure of ferrihydrite Commun Chem 4, 134 10.1038/s42004-021-00562-7
71. Callis, P. R. (2011) Predicting fluorescence lifetimes and spectra of biopolymers Methods Enzymol 487, 1-38 10.1016/B978-0-12-381270-4.00001-9
72. Bernhard, S. P., Goodman, C. K., Norton, E. G., Alme, D. G., Lawrence, C. M., and Cloninger, M. J. (2020) Time-Dependent Fluorescence Spectroscopy to Quantify Complex Binding Interactions ACS Omega 5, 29017-29024 10.1021/acsomega.0c03416
73. Brown, S., Gauvin, C. C., Charbonneau, A. A., Burman, N., and Lawrence, C. M. (2020) Csx3 is a cyclic oligonucleotide phosphodiesterase associated with type III CRISPR-Cas that degrades the second messenger cA(4) J Biol Chem 295, 14963-14972 10.1074/jbc.RA120.014099

74. Orban, K., and Finkel, S. E. (2022) Dps Is a Universally Conserved Dual-Action DNA-Binding and Ferritin Protein J Bacteriol 204, e0003622 10.1128/jb.00036-22
75. Williams, S. M., and Chatterji, D. (2023) Dps Functions as a Key Player in Bacterial Iron Homeostasis ACS Omega 8, 34299-34309 10.1021/acsomega.3c03277
76. Pulliainen, A. T., Kauko, A., Haataja, S., Papageorgiou, A. C., and Finne, J. (2005) Dps/Dpr ferritin-like protein: insights into the mechanism of iron incorporation and evidence for a central role in cellular iron homeostasis in Streptococcus suis Mol Microbiol 57, 1086-1100 10.1111/j.1365-2958.2005.04756.x
77. Bellapadrona, G., Stefanini, S., Zamparelli, C., Theil, E. C., and Chiancone, E. (2009) Iron translocation into and out of Listeria innocua Dps and size distribution of the protein-enclosed nanomineral are modulated by the electrostatic gradient at the 3-fold "ferritin-like" pores J Biol Chem 284, 19101-19109 10.1074/jbc.M109.014670
78. Williams, S. M., Chandran, A. V., Vijayabaskar, M. S., Roy, S., Balaram, H., Vishveshwara, S. et al. (2014) A histidine aspartate ionic lock gates the iron passage in miniferritins from Mycobacterium smegmatis J Biol Chem 289, 11042-11058 10.1074/jbc.M113.524421
79. Zeth, K., Offermann, S., Essen, L. O., and Oesterhelt, D. (2004) Iron-oxo clusters biomineralizing on protein surfaces: structural analysis of Halobacterium salinarum DpsA in its low- and high-iron states Proc Natl Acad Sci U S A 101, 13780-13785 10.1073/pnas.0401821101
80. Pan, Y. H., Sader, K., Powell, J. J., Bleloch, A., Gass, M., Trinick, J. et al. (2009) 3D morphology of the human hepatic ferritin mineral core: new evidence for a subunit structure revealed by single particle analysis of HAADF-STEM images J Struct Biol 166, 22-31 10.1016/j.jsb.2008.12.001
81. Pozzi, C., Ciambellotti, S., Bernacchioni, C., Di Pisa, F., Mangani, S., and Turano, P. (2017) Chemistry at the protein-mineral interface in L-ferritin assists the assembly of a functional ($\mu(3)$ -oxo)Tris[($\mu(2)$ -peroxo)] triiron(III) cluster Proc Natl Acad Sci U S A 114, 2580-2585 10.1073/pnas.1614302114
82. Ciambellotti, S., Pozzi, C., Mangani, S., and Turano, P. (2020) Iron Biomineral Growth from the Initial Nucleation Seed in L-Ferritin Chemistry 26, 5770-5773 10.1002/chem.202000064
83. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding Anal Biochem 72, 248-254 10.1006/abio.1976.9999
84. Frenkiel-Krispin, D., Levin-Zaidman, S., Shimoni, E., Wolf, S. G., Wachtel, E. J., Arad, T. et al. (2001) Regulated phase transitions of bacterial chromatin: a non-enzymatic pathway for generic DNA protection EMBO J 20, 1184-1191 10.1093/emboj/20.5.1184

85. Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A. et al. (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM *Commun Biol* 2, 218 10.1038/s42003-019-0437-z
86. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis *Nat Protoc* 10, 845-858 10.1038/nprot.2015.053
87. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I. et al. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers *Protein Sci* 30, 70-82 10.1002/pro.3943
88. Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A. et al. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography *Acta Crystallogr D Struct Biol* 74, 531-544 10.1107/S2059798318006551
89. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I. et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix *Acta Crystallogr D Struct Biol* 75, 861-877 10.1107/S2059798319011471
90. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot *Acta Crystallogr D Biol Crystallogr* 66, 486-501 10.1107/S0907444910007493
91. Muretta, J. M., Kyrychenko, A., Ladokhin, A. S., Kast, D. J., Gillispie, G. D., and Thomas, D. D. (2010) High-performance time-resolved fluorescence by direct waveform recording *Rev Sci Instrum* 81, 103101 10.1063/1.3480647
92. Petersen, K. J., Peterson, K. C., Muretta, J. M., Higgins, S. E., Gillispie, G. D., and Thomas, D. D. (2014) Fluorescence lifetime plate reader: resolution and precision meet high-throughput *Rev Sci Instrum* 85, 113101 10.1063/1.4900727

APPENDICES

APPENDIX A

FACILITY MANAGEMENT SCRIPTS

Linux System ConfigurationSSSD Configuration File

[sssd]

config_file_version = 2

domains = msu.montana.edu

services = pam,nss

[pam]

[nss]

[domain/msu.montana.edu]

id_provider = ad

auth_provider = ad

enumerate = false

cache_credentials = false

case_sensitive = true

override_homedir = /home/%u

fallback_homedir = /etc/skel

default_shell = /bin/bash

[domain/default]

auth_provider = krb5

chpass_provider = krb5

krb5_realm = MSU.MONTANA.EDU

krb5_server = iolite.msu.montana.edu

```
krb5_validate = False
```

```
krb5_renewable_lifetime = 1d
```

```
krb5_lifetime = 1d
```

pam_mount Configuration File

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<!DOCTYPE pam_mount SYSTEM "pam_mount.conf.xml.dtd">
```

```
<!--
```

```
    See pam_mount.conf(5) for a description.
```

```
-->
```

```
<pam_mount>
```

```
    <debug enable="2"/>
```

```
    <mntoptions allow="nosuid,nodev,loop,encryption,fsck,nonempty,allow_root,allow_other"/>
```

```
    <mntoptions require="nosuid,nodev"/>
```

```
    <logout wait="0" hup="no" term="no" kill="no"/>
```

```
    <mkmountpoint enable="1" remove="true"/>
```

```
<volume options="sec=ntlmv2,domain=MSU,vers=3" server="blackmore.msu.montana.edu"
```

```
user="*" path="data" fstype="cifs" mountpoint="/home/%(DOMAIN_USER)/blackmore"/>
```

```
</pam_mount>
```

CryoSPARC Scripts & ConfigurationCryoSPARC Master Configuration File

Instance Configuration

```
export CRYOSPARC_LICENSE_ID="REDACTED"

export CRYOSPARC_MASTER_HOSTNAME="tempest-cryosparc.msu.montana.edu"

export CRYOSPARC_DB_PATH="/home/cryosparc/cryosparc_db"

export CRYOSPARC_BASE_PORT=39000

export CRYOSPARC_DB_CONNECTION_TIMEOUT_MS=50000

export CRYOSPARC_HOSTNAME_CHECK="tempest-cryosparc.msu.montana.edu"
```

Security

```
export CRYOSPARC_INSECURE=true

export CRYOSPARC_DB_ENABLE_AUTH=true
```

Cluster Integration

```
export CRYOSPARC_CLUSTER_JOB_MONITOR_INTERVAL=10

export CRYOSPARC_CLUSTER_JOB_MONITOR_MAX_RETRIES=1000000
```

Project Configuration

```
export CRYOSPARC_PROJECT_DIR_PREFIX='CS-'
```

Development

```
export CRYOSPARC_DEVELOP=false
```

Other

```
export CRYOSPARC_CLICK_WRAP=true
```

Instance Configuration

```
export CRYOSPARC_LICENSE_ID="REDACTED"

export CRYOSPARC_MASTER_HOSTNAME="tempest-cryosparc.msu.montana.edu"

export CRYOSPARC_DB_PATH="/home/cryosparc/cryosparc_db"

export CRYOSPARC_BASE_PORT=39000

export CRYOSPARC_DB_CONNECTION_TIMEOUT_MS=50000

export CRYOSPARC_HOSTNAME_CHECK="tempest-cryosparc.msu.montana.edu"
```

Security

```
export CRYOSPARC_INSECURE=true

export CRYOSPARC_DB_ENABLE_AUTH=true
```

Cluster Integration

```
export CRYOSPARC_CLUSTER_JOB_MONITOR_INTERVAL=10

export CRYOSPARC_CLUSTER_JOB_MONITOR_MAX_RETRIES=1000000
```

Project Configuration

```
export CRYOSPARC_PROJECT_DIR_PREFIX='CS-'
```

Development

```
export CRYOSPARC_DEVELOP=false
```

Other

```
export CRYOSPARC_CLICK_WRAP=true
```

SystemD Service File

[Unit]

Description=Process Monitoring and Control Daemon for CryoSPARC

[Service]

Type=forking

Environment="MODULEPATH=/mnt/shared/modulefiles:/mnt/shared/ebmodules/eb/all/"

Environment="MODULEPATH_modshare=/mnt/shared/modulefiles:1"

Environment="MODULESHOME=/usr/local/Modules/4.6.1"

Environment="MODULES_CMD=/usr/local/Modules/4.6.1/libexec/modulecmd.tcl"

Environment="BASH_ENV=/usr/local/Modules/4.6.1/init/bash"

Environment="ENV=/usr/local/Modules/4.6.1/init/profile.sh"

Environment="PATH=/home/cryosparc/cryosparc_master/deps/external/mongodb/bin:/home/cryosparc/cryosparc_master/bin:/home/cryosparc/cryosparc_master/deps/anaconda/envs/cryosparc_master_env/bin:/home/cryosparc/cryosparc_master/deps/anaconda/condabin:/usr/local/Modules/4.6.1/bin:/usr/local/bin:/usr/bin:/usr/local/sbin:/usr/sbin:/opt/slurm/20.02.3/bin:/opt/slurm/20.02.3/sbin:/home/cryosparc/cryosparc_master/bin"

ExecStart=/home/cryosparc/cryosparc_master/bin/cryosparcm supervisord -c
/home/cryosparc/cryosparc_master/supervisord.conf

ExecStartPost=/home/cryosparc/cryosparc_master/bin/cryosparcm start --systemd

PIDFile=/home/cryosparc/cryosparc_master/run/supervisord.pid

WorkingDirectory=/home/cryosparc/cryosparc_master

[Install]

WantedBy=default.target

Database Backup SystemD Service File

[Unit]

Description=cryosparc_backup

[Timer]

OnCalendar=daily

Persistent=true

[Install]

WantedBy=timers.target

[Unit]

Description=cryosparc_backup

[Service]

Type=oneshot

ExecStart=/home/cryosparc/cryosparc_master/bin/cryosparcm backup --

dir=/mnt/blackmore/cryoem/cryosparc_db_backups

WorkingDirectory=/home/cryosparc

Tempest Lane Cluster Info File

cluster_info.json

{

"qdel_cmd_tpl": "scancel {{ cluster_job_id }}",

"worker_bin_path": "/home/cryosparc/cryosparc_worker/bin/cryosparcw",

"qinfo_cmd_tpl": "sinfo --format='%0.8N %0.6D %0.10P %0.6T %0.14C %0.5c %0.6z %0.7m %0.7G
%0.9d %20E'",

```

"qsub_cmd_tpl": "sbatch {{ script_path_abs }}",
"qstat_cmd_tpl": "squeue -j {{ cluster_job_id }}",
"send_cmd_tpl": "{{ command }}",
"name": "lawrence_gpupriority",
"cache_path": "/home/group/cryoem",
"cache_reserve_mb": 10000
}

```

Tempest Lane Cluster Submission Script

```

#!/bin/bash

#SBATCH --account=priority-cryoem

#SBATCH --job-name=cryosparc_{{ project_uid }}_{{ job_uid }}

#SBATCH --partition=gupriority

#SBATCH --output={{ job_log_path_abs }}

#SBATCH --error={{ job_log_path_abs }}

#SBATCH --nodes=1

#SBATCH --mem={{ (ram_gb|float * ram_gb_multiplier|float)|int }}G

#SBATCH --ntasks-per-node=1

#SBATCH --cpus-per-task={{ num_cpu }}

#SBATCH --gres=gpu:{{ gres }}:{{ num_gpu }}

#SBATCH --time {{ time_hr_min_sec }}

module load CUDA/11.1.1-GCC-10.2.0

available_devs=""

```

```

for devidx in $(seq 0 1);
do
    if [[ -z $(nvidia-smi -i $devidx --query-compute-apps=pid --format=csv,noheader) ]] ; then
        if [[ -z "$available_devs" ]] ; then
            available_devs=$devidx
        else
            available_devs=$available_devs,$devidx
        fi
    fi
done
export CUDA_VISIBLE_DEVICES=$available_devs
srun {{ run_cmd }}

```

Delete Rejected Exposures Script

This script can be run on Tempest using CryoSPARC Tools to delete exposures that were collected, but later rejected in CryoSPARC. To use, paste the script into a file ending in .py, fill out the variables, and execute ‘python <filename>.py’.

```

import os

from pathlib import Path

from getpass import getpass

from cryosparc.tools import CryoSPARC

cs_license = input("Enter your CryoSPARC license, or leave blank for Tempest: ") or
"REDACTED"

```

```

cs_hostname = input("Enter your CryoSPARC web URL, or leave blank for Tempest") or
"tempest-cryosparc.msu.montana.edu"

cs_email = input("Enter your CryoSPARC account email: ")

cs_password = getpass("Enter your CryoSPARC account password: ")

cs = CryoSPARC(license=cs_license, email=cs_email, password=cs_password,
host=cs_hostname, base_port=39000)

project = cs.find_project(str('P' + str(input("Which cryoSPARC project number? "))))

job = project.find_job(str('J' + str(input("Which cryoSPARC job number contains the rejected
exposures? "))))

exposures_rejected = job.load_output('rejected_exposures')

project_dir = Path(project.dir())

movies_list = []

for group in ('movie_blob',):

    field = group + "/path"

    if field in exposures_rejected:

        print(f"Found {len(exposures_rejected)} rejected exposures. Please wait, this may take
some time. Wait for program to tell you it's done.")

        for rel_path in exposures_rejected[field]:

            abs_path = project_dir / rel_path

            resolved_path=Path.resolve(abs_path)

            movies_list.append(str(resolved_path))

```

```

print("Done.")

# Run this if you want to save a list of filenames to be deleted. Otherwise, skip down to delete

# with open("rejected_exposures.txt", "a") as output_file:

#   print(str(movies_list).replace(',', '\n'), file=output_file)

# output_file.close()

print("Beginning file removal, please wait, this may take some time. Wait for program to tell you
it's done.")

for i in range(len(movies_list)):

    try:

        os.remove(str(movies_list[i]))

    except FileNotFoundError:

        print("File", str(movies_list[i]), "not found")

print("Done.")

```

SerialEM Scripts & Troubleshooting

Enclosed in this appendix are tips and common failure mode troubleshooting guides for using SerialEM.

Startup Script

MacroName Startup

SetLowDoseMode 1

SetDirectory X:\

UserSetDirectory

Echo -----

ReportDirectory

Echo -----

OpenChooserInCurrentDir

Echo Loading Low Dose Record

GoToLowDoseArea R

Echo Opening Navigator...

OpenNavigator

Echo Opening Imaging States...

OpenImagingStateDialog

Echo Done

Atlas Script

MacroName AtlasGrids

EnterOneNumber Which grid slot to start with?

offset = \$reportedValue1

EnterOneNumber Which grid slot to stop with?

remain = \$reportedValue1

Echo Will montage grids from \$offset to \$remain

offset = \$offset

remain = \$remain

SetApertureSize 1 150

Loop \$remain index

```

cart = $index + $offset
cart = $cart - 1

ReportSlotStatus $cart

if $RepVal1 == 1

    ScreenUp

    Echo Loading cartridge $cart

    LoadCartridge $cart

    SetColumnOrGunValve 1

    MoveStageTo 0 0 0

    M

    NewMap 0 Grid $cart

        ChangeItemRegistration $reportedvalue1 $cart

    endif

endloop

CloseFile

SetApertureSize 1 50

SetColumnOrGunValve 0

Echo Grid maps finished at

ReportDateTime

```

Incorrect “View Mode” Defocus Offset

Typically, it is desirable to have a larger defocus when in View mode than in record mode, to make objects more visible at low magnification. Typically, this defocus offset is

approximately -50 microns. However, sometimes due to SerialEM bugs or glitches, this offset sometimes doubles -100 microns. This can drastically affect View-Record alignment. To reset it, follow this procedure.

- 1) In the “Low Dose Control” panel, make sure that “Record” mode is loaded.
- 2) Using the “Microscope Control” panel, set the defocus in record mode to 0.
- 3) Load “View” mode, and set the defocus offset to 0.
- 4) In the “Microscope Control” panel, set the microscope defocus to 0.
- 5) In the “Low Dose Control” panel, re-load “Record” mode, and set the “Defocus Offset” back to -50.
- 6) Re-load “View” mode and verify that the microscope defocus is now -50.

View-Record Alignment

Because “View” mode is typically a much lower magnification, and an entirely different FEG-probe mode than “Record”, sometimes, where the operator centers in “View” is not actually what gets imaged in “Record”. Additionally, a high defocus offset in “View” mode can cause further drift. If “View” and “Record” are not aligned, the first thing the operator should do is check that the “View” mode defocus offset is correct. If it is correct, or if has intentionally been changed, but “View” and “Record” are still not aligned, follow the procedure below to re-align them.

- 1) Find and center (via right-mouse click and drag) an object in either “Record” or “Preview” mode that is large enough to see from “View” mode as well. (Hint: “Preview” mode behaves like “Record” but without saving/aligning frames and is typically a shorter exposure).

- 2) Take a “View” image and see where the crosshair goes.
- 3) Right-click and drag the mouse to center the crosshair over the object centered in “Record”.
- 4) In the “Low Dose Control” panel, select “Shift: Set”. SerialEM will take a new “View” image. The crosshair should now be centered over the same area it was when centered in “Record”. If not, repeat step 4 until it is.
- 5) Find a new object in “View” mode, center it in “View” mode, and take a “Record” or “Preview”. The “Record” should now be centered (or very close) where the crosshairs were in “View” mode.

SmartScope Scripts & Troubleshooting

These scripts and configuration files complement SmartScope data collection.

Aperture Fix for Arctica Interface

The Talos Arctica is not currently configured with Thermo Fisher’s “UTAPI” that replaces the advanced scripting interface. As a result of this, SerialEM has only indirect automated control of the apertures, using accessibility tools to select the aperture size on the microscope PC. This causes automated aperture changes to experience occasional issues, specifically when a large number of changes occur over a relatively short period of time. Typically, waiting a few seconds and running the aperture change command again will solve the issue. However, this presents a problem when using software that has automated aperture changes that can’t be manually controlled, such as SmartScope. Therefore, it is desirable to

manually edit the SmartScope microscope interface Python code to add a wait and duplicate aperture change.

The aperture interface file can be found on hydra at “/home/cryoem/SmartScope/SmartScope-dev/Smartscope/core/interfaces/tfsserialem_interface.py”. There are multiple functions for aperture controls. Of specific interest are “change_aperture_temporarily” and “remove_objective_aperture”. For each instance of “sem.SetApertureSize(aperture,aperture_size)”, add a “time.sleep(wait)” followed by another “sem.SetApertureSize(aperture,aperture_size)” command underneath. Do the same for the “sem.RemoveAperture” and “sem.ReInsertAperture” commands as well. This will essentially call each aperture change command, wait 10 seconds, and then call the command again. Note that each time SmartScope is updated, this file will be overwritten.

Exposure Groups by Image Shift

```
from pathlib import Path

import re

import numpy as np

import matplotlib

from matplotlib import pyplot as plt

import matplotlib.cm as cm

from sklearn.cluster import KMeans

from cryosparc.tools import CryoSPARC

from getpass import getpass
```

```
##### Instructions #####
```

```
# Place a copy of this file in each directory you have .mdoc
```

```
# files in, load the cryosparc-tools conda environment (if using)
```

```
# and run this script: python3 cs_exposure_groups.py
```

```
#
```

```
# The major difference between this script and cs_mdoc_image_shift.py is
```

```
# that this script will do the kmeans clustering itself and update the groups
```

```
# while cs_mdoc_image_shift.py will update the image shift values of
```

```
# micrograph so that CryoSPARC can do the clusering
```

```
##### Variables #####
```

```
expected_groups = 25
```

```
cs_license = 'asdf1348-replace-me-123lkj12'
```

```
cs_hostname = 'mycryosparcinstance.mysupercoolinstitution.com'
```

```
cs_port = 39000
```

```
cs_email = 'myemail@mydomain.com'
```

```
cs_project = 'P83'
```

```
cs_workspace = 'W1'
```

```
cs_job = 'J56'
```

```
cs_job_output = 'exposure'
```

```
showplot = True
```

```
##### (il)Logic #####
```

```
#Set directory and get list of mdocs
```

```

mdoc_directory = Path('.')

mdoc_list = list(mdoc_directory.glob('**/*.mdoc'))

#Initialize list that will hold dictionary with exposure information
exposures = []

#initialize numpy array for clustering and plotting
image_shifts = np.empty(shape=(len(mdoc_list),2))

#Iterate through list of mdoc files
for i in range(len(mdoc_list)):

    #Open individual mdoc file
    with open(mdoc_list[i]) as mdoc:

        #Read each line in mdoc file
        for line in mdoc:

            #if the line matches, store the name of the file and also the image shift values into a list
            if re.search("ImageShift",line):

                shift_data = {

                    "index": i,

                    "exposure": str(mdoc_list[i]).rsplit(sep='.',maxsplit = 1)[0],

                    "x_shift": line.rsplit()[2],

                    "y_shift": line.rsplit()[3],

                    "group": 1

                }

                exposures.append(shift_data)

```

```

#Convert dictionary x and y shift into numpy array

for i in range(len(exposures)):

    image_shifts[i,0] = exposures[i]['x_shift']

    image_shifts[i,1] = exposures[i]['y_shift']

#Kmeans clustering - probably not optimized

kmeans = KMeans(init='k-means++', n_clusters=expected_groups, n_init=10)

kmeans.fit(image_shifts)

#Assign an exposure group to each exposure in the dictionary

for i in range(len(exposures)):

    exposures[i]['group'] = kmeans.labels_[i]

#Plot the exposures and ask to proceed

if showplot == True:

    plt.scatter(image_shifts[:,0],image_shifts[:,1])

    plt.scatter(kmeans.cluster_centers_[0],kmeans.cluster_centers_[1],c='red')

    plt.show()

proceed = input("Would you like to use these exposure groups? y/n: ")

if proceed != "y":

    print("Did not confirm - exiting unceremoniously")

    quit()

#Get CS password

print("Please enter CryoSPARC password: ")

cs_password = getpass()

```

```

#Connect to CS

cs = CryoSPARC(license=cs_license, email=cs_email, password=cs_password,
host=cs_hostname, base_port=cs_port)

assert cs.test_connection()

project = cs.find_project(cs_project)

job = cs.find_job(cs_project, cs_job)

cs_exposures = job.load_output(cs_job_output)

grouped_exposures = cs_exposures.copy()

for row in grouped_exposures.rows():

    for i in range(len(exposures)):

        if exposures[i]['exposure'] == row['movie_blob/path'].rsplit(sep='/')[-1]:

            row['mscope_params/exp_group_id'] = exposures[i]['group']

            row['ctf/exp_group_id'] = exposures[i]['group']

            print(row)

            break

#for row in grouped_exposures.rows():

#    print(row['movie_blob/path'], row['mscope_params/exp_group_id'])

print(grouped_exposures.fields())

project.save_external_result(

    workspace_uid=cs_workspace,

    dataset=grouped_exposures,

    type="exposure",

```

```

name="exposure",

slots=["mscope_params","ctf"],

passthrough=(cs_job,cs_job_output),

title="Grouped Exposures"

)

```

Exposure Groups by MDOC

```

from pathlib import Path

import re

import sys

from cryosparc.tools import CryoSPARC

from getpass import getpass

from dotenv import dotenv_values

import argparse

from typing import Iterable, Dict, List

import numpy as np

##### Instructions #####

# Place a copy of this file in each directory you have .mdoc

# files in, load the cryosparc-tools conda environment (if using)

# and run this script: python3 cs_mdoc_image_shift.py

##### Variables #####

CONFIG = dotenv_values(".env")

CS_JOB_OUTPUT_TYPE = 'all_exposures'

```

```
def parser(args:List) -> argparse.Namespace:
```

```
    parser = argparse.ArgumentParser()
```

```
    parser.add_argument('--email','-e',type=str, help="Email or account used in cryosparc")
```

```
    parser.add_argument('--project', '-p',type=str, help='Cryosparc project. i.e. P83')
```

```
    parser.add_argument('--workspace','-w',type=str, help='Workspace number, i.e. W1')
```

```
    parser.add_argument('--job','-j',type=str,help='Input job number, i.e. J56')
```

```
    parser.add_argument('--mdoc_dir','-d',type=str, help='Directory where the exposures mdoc  
files are located')
```

```
    parser.add_argument('--recursive', '-r', action='store_true', default=False, help="Whether to  
look for the mdoc file recursively from the mdoc_dir.")
```

```
    args = parser.parse_args(args)
```

```
    if any([args.email is None, args.project is None, args.job is None, args.workspace is None]):
```

```
        parser.print_help()
```

```
        sys.exit(1)
```

```
    return args
```

```
def recursively_find_mdoc(mdoc_dir:Path, filename:'str') -> Iterable:
```

```
    ##### (il)Logic #####
```

```
    #Set directory and get list of mdocs
```

```
    mdoc_directory = Path(mdoc_dir)
```

```
    if not mdoc_dir.exists():
```

```
        raise NotADirectoryError(f'Directory {mdoc_dir} does not exist.')
```

```
    mdoc_list = mdoc_directory.glob(f'**/{filename}')
```

```

    return next(mdoc_list)

#Iterate through list of mdoc files

def parse_mdoc(mdoc_file:Path) -> Dict:

    #Open individual mdoc file

    mdoc = mdoc_file.read_text()

    pattern = r"ImageShift = (-?\d+\.\d+)\s+(-?\d+\.\d+)"

    matches = re.findall(pattern, mdoc)

    if matches:

        return np.array(matches[0]).astype(float)

def get_exposures_from_job(project:str,job) -> np.ndarray:

    cs_project = cs.find_project(project)

    cs_job = cs.find_job(project,job)

    cs_exposures = cs_job.load_output(CS_JOB_OUTPUT_TYPE)

    return cs_project, cs_job, cs_exposures.copy()

def progress(i,total,interval=2):

    completion = round(i/total*100)

    if completion % interval == 0:

        print(f'{completion}% completed. {i}/{total}', end='\r')

if __name__ == "__main__":

    args = parser(sys.argv[1:])

```

```

cs_password = getpass("Please enter CryoSPARC password: ")

#Connect to CS

cs = CryoSPARC(license=CONFIG['cs_license'], email=args.email, password=cs_password,
host=CONFIG['cs_hostname'], base_port=int(CONFIG['cs_port']))

assert cs.test_connection(), "Connection to cryosparc failed"


#Find dataset and copy it to a new dataset

cs_project, cs_job, grouped_exposures = get_exposures_from_job(args.project,args.job)

#For every row in the dataset, search entire mdoc data and see if it fits a known movie and
image shift. There's probably a much faster way to do this.

total = len(grouped_exposures)

for ind,row in enumerate(grouped_exposures.rows()):

    exposure = row['movie_blob/path'].rsplit(sep='/')[-1]

    mdocfile = Path(args.mdoc_dir,exposure+'.mdoc')

    if args.recursive:

        mdocfile = recursively_find_mdoc(Path(args.mdoc_dir),filename=exposure+'.mdoc')

    mdoc = parse_mdoc(mdocfile)

    # if exposures[i]['exposure'] == row['movie_blob/path'].rsplit(sep='/')[-1]:

    row['mscope_params/beam_shift'] = mdoc

    row['mscope_params/beam_shift_known'] = 1

    progress(ind,total)

```

#Make a new external job with the updated values, and passthrough the remaining, non-updated values

```
cs_project.save_external_result(  
    workspace_uid=args.workspace,  
    dataset=grouped_exposures,  
    type="exposure",  
    name="exposure",  
    slots=["mscope_params"],  
    passthrough=(args.job,CS_JOB_OUTPUT_TYPE),  
    title="ImageShift Updated Exposures"  
)
```

Talos Arctica ConfigurationParallel Illumination

Gun Lens 3.7

Table 3. C2 lens intensity for parallel illumination at gun lens setting 3.7.

Spot Size	C2
1	46.78
2	45.23
3	43.35
4	42.27
5	41.20
6	39.78
7	38.22
8	37.62
9	37.00
10	NA
11	NA

Gun Lens 4

Table 4. C2 lens intensity for parallel illumination at gun lens setting 4.

Spot Size	C2
1	46.8%
2	45.6%
3	43.9%
4	42.3%
5	41.1%
6	39.6%
7	38.6%
8	37.3%
9	36.9%
10	NA
11	NA

OURRStore Tape Archival

To utilize the OURRStore tape archival program, files must be first assembled into one archive, and that archive must then be broken up into 200 gigabyte files. This process can be combined into a single shell command.

```
tar cvzf - <directory_name> | split --bytes=200GB - <filename>.backup.tar.gz.
```

Once the operation is complete, the .tar.gz files can be uploaded to OURRStore via Globus.

APPENDIX B

THIOFERRITIN SUPPLEMENTAL INFORMATION

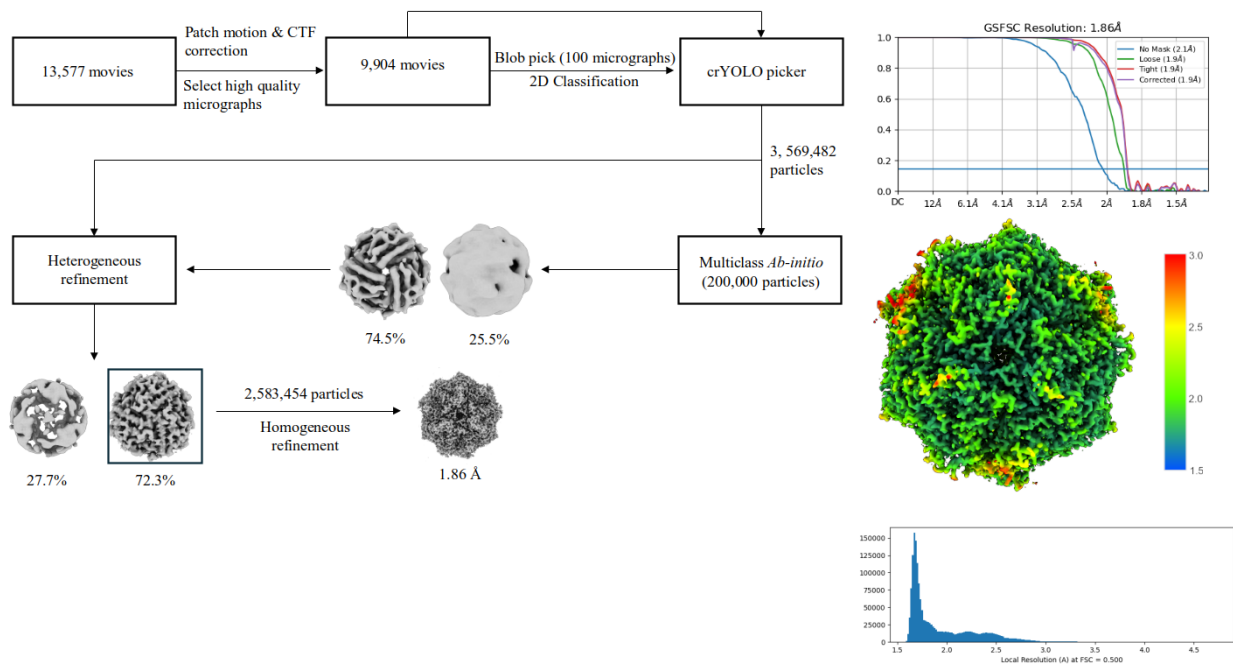


Figure 73. Apo Pf-Tftn single particle analysis workflow.

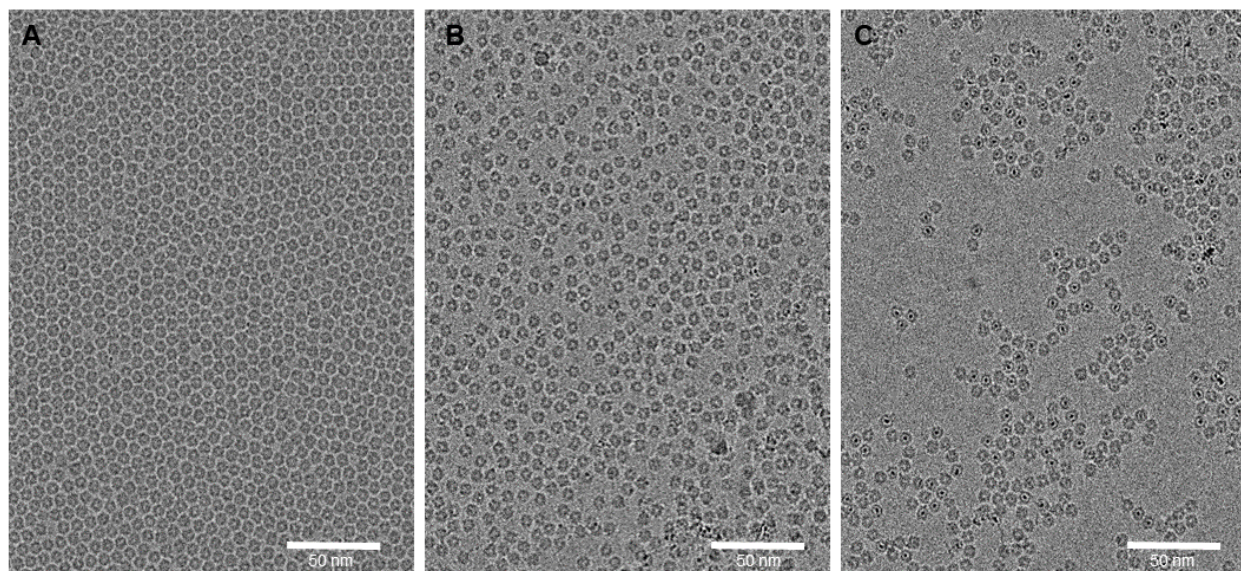


Figure 74. Representative micrographs from the apo (A), nucleated (B), and mineralized (C) datasets. All micrographs at 88,000x magnification, dose of $\sim 55 \text{ e}/\text{\AA}^2$, and recorded -1.4 μm defocus. Scale bars 50 nm.

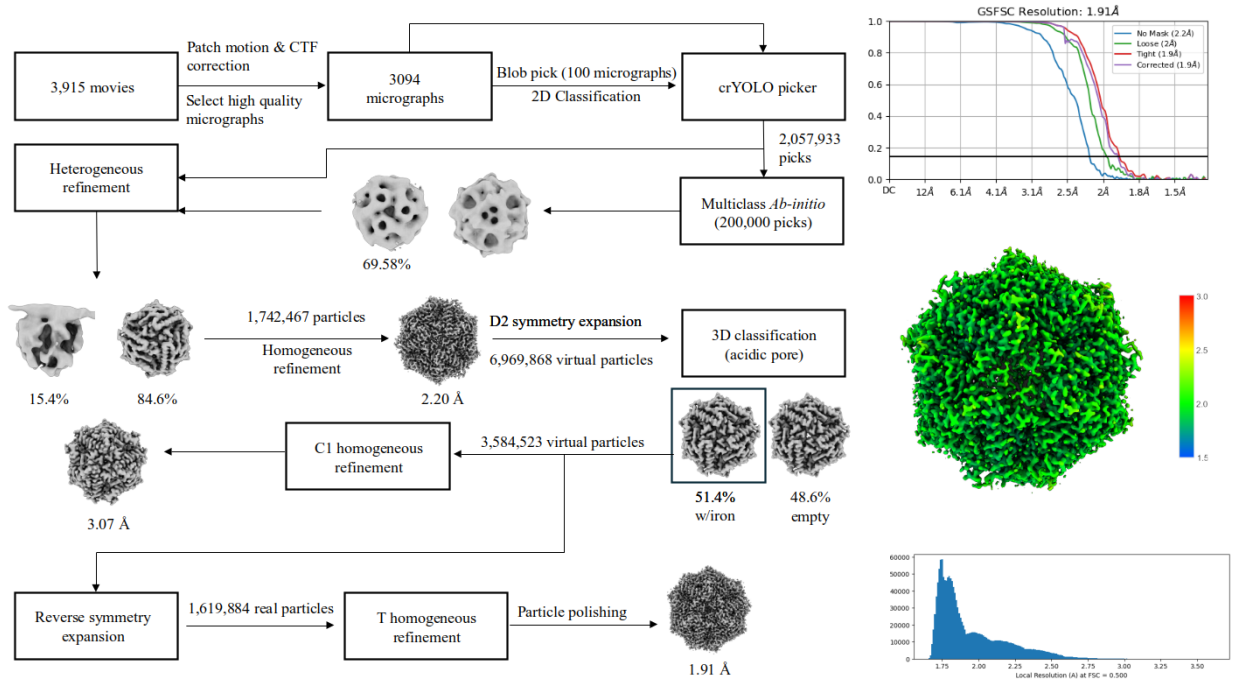


Figure 75. Nucleated Pf-Tftn single particle analysis workflow.

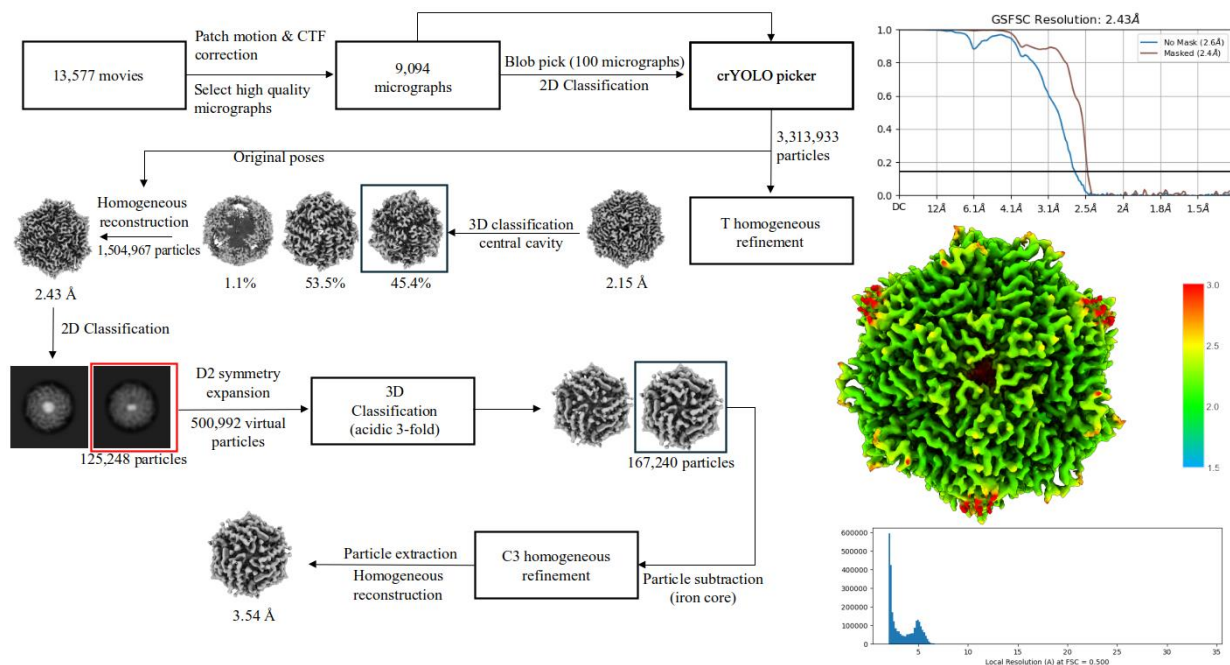


Figure 76. Mineralized Pf-Tftn single particle analysis workflow

Table 5. Structure validation report for Apo-Pf-Tftn

Model	Apo
Composition (#)	
Chains	12
Atoms	16956 (Hydrogens: 0)
Residues	Protein: 2040 Nucleotide: 0
Water	252
Ligands	FE: 24
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.009 (0)
Angles (°) (# > 4σ)	1.048 (0)
MolProbity score	1.3
Clash score	5.53
Ramachandran plot (%)	
Outliers	0
Allowed	1.19
Favored	98.81
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 2028)	1.32 (0.18)
helix (N = 1452)	0.99 (0.13)
sheet (N = 0)	--- (---)

loop (N = 576)	0.90 (0.27)
Rotamer outliers (%)	0
Cβ outliers (%)	NA
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	0.6
ADP (B-factors)	
Iso/Aniso (#)	16956/0
min/max/mean	
Protein	2.98/77.92/21.16
Nucleotide	---
Ligand	28.42/43.55/35.98
Water	0.64/33.76/5.00
Occupancy	
Mean	0.99
occ = 1 (%)	98.23
0 < occ < 1 (%)	1.77
occ > 1 (%)	0
Data	
Box	
Lengths (Å)	105.57, 109.02, 101.43
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	1.9
Resolution Estimates (Å)	Masked Unmasked
d FSC (half maps; 0.143)	--- ---
d 99 (full/half1/half2)	2.1/---/--- 2.1/---/---
d model	2.1 2.1
d FSC model (0/0.143/0.5)	1.8/1.9/1.9 1.9/1.9/1.9
Map min/max/mean	-28.58/52.59/-0.01
Model vs. Data	
CC (mask)	0.88
CC (box)	0.78
CC (peaks)	0.8
CC (volume)	0.86
Mean CC for ligands	0.67

Table 6. Structure validation for nucleated Pf-Tftn.

Model	Nucleated
Composition (#)	
Chains	12
Atoms	17044 (Hydrogens: 0)
Residues	Protein: 2040 Nucleotide: 0

Water	516	
Ligands	FE: 68	
Bonds (RMSD)	O: 68	
Length (Å) (# > 4σ)		
Angles (°) (# > 4σ)	0.009 (0)	
MolProbity score	1.134 (0)	
Clash score	1.15	
Ramachandran plot (%)	3.62	
Outliers		
Allowed	0	
Favored	0.6	
Rama-Z (Ramachandran plot Z-score, RMSD)	99.4	
whole (N = 2028)		
helix (N = 1452)	1.45 (0.18)	
sheet (N = 0)	1.13 (0.12)	
loop (N = 576)	--- (---)	
Rotamer outliers (%)	0.86 (0.28)	
Cβ outliers (%)	0.7	
Peptide plane (%)	NA	
Cis proline/general		
Twisted proline/general	0.0/0.0	
CaBLAM outliers (%)	0.0/0.0	
ADP (B-factors)	0.6	
Iso/Aniso (#)		
min/max/mean	17044/0	
Protein		
Nucleotide	6.30/84.60/23.47	
Ligand	---	
Water	0.01/100.33/58.06	
Occupancy	24.77/52.36/34.20	
Mean		
occ = 1 (%)	1	
0 < occ < 1 (%)	99.65	
occ > 1 (%)	0.35	
Data		
Box		
Lengths (Å)	105.57, 110.40, 102.81	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	1.9	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	2.2/---/---	2.2/---/---
d model	2.2	2.2
d FSC model (0/0.143/0.5)	1.9/1.9/2.0	1.9/1.9/2.0
Map min/max/mean	-28.48/63.39/0.05	

Model vs. Data	
CC (mask)	0.87
CC (box)	0.74
CC (peaks)	0.77
CC (volume)	0.85
Mean CC for ligands	0.71

Table 7. Structure validation for full mineral Pf-Tftn.

Model	Holo – Full
Composition (#)	
Chains	12
Atoms	17016 (Hydrogens: 0)
Residues	Protein: 2040 Nucleotide: 0
Water	540
Ligands	FE: 24
Bonds (RMSD)	
Length (Å) (# > 4 σ)	0.009 (0)
Angles (°) (# > 4 σ)	0.791 (0)
MolProbity score	1.35
Clash score	6.35
Ramachandran plot (%)	
Outliers	0
Allowed	1.19
Favored	98.81
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 2028)	-3.68 (0.15)
helix (N = 1452)	-2.94 (0.09)
sheet (N = 0)	--- (---)
loop (N = 576)	-0.38 (0.26)
Rotamer outliers (%)	0
C β outliers (%)	NA
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	0.6
ADP (B-factors)	
Iso/Aniso (#)	17016/0
min/max/mean	
Protein	0.34/55.72/11.68
Nucleotide	---
Ligand	4.43/9.42/6.93
Water	7.29/45.01/18.39
Occupancy	

Mean	1	
occ = 1 (%)	99.01	
0 < occ < 1 (%)	0.99	
occ > 1 (%)	0	
Data		
Box		
Lengths (Å)	105.57, 112.47, 104.19	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	2.4	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	2.5/---/---	2.5/---/---
d model	2.5	2.5
d FSC model (0/0.143/0.5)	2.4/2.4/2.5	2.4/2.4/2.5
Map min/max/mean	-22.93/43.12/0.58	
Model vs. Data		
CC (mask)	0.84	
CC (box)	0.8	
CC (peaks)	0.77	
CC (volume)	0.82	
Mean CC for ligands	0.74	

Table 8. Structure validation for intermediate mineral Pf-Tftn.

Model	Holo – Intermediate
Composition (#)	
Chains	3
Atoms	4107 (Hydrogens: 0)
Residues	Protein: 510 Nucleotide: 0
Water	0
Ligands	FE: 6
Bonds (RMSD)	
Length (Å) (# > 4 σ)	0.003 (0)
Angles (°) (# > 4 σ)	0.522 (0)
MolProbity score	1.76
Clash score	7.14
Ramachandran plot (%)	
Outliers	0
Allowed	5.36
Favored	94.64
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 2028)	0.94 (0.38)
helix (N = 1452)	1.83 (0.28)
sheet (N = 0)	--- (---)

loop (N = 576)	-2.14 (0.45)	
Rotamer outliers (%)	0.7	
Cβ outliers (%)	NA	
Peptide plane (%)		
Cis proline/general	0.0/0.0	
Twisted proline/general	0.0/0.0	
CaBLAM outliers (%)	6.02	
ADP (B-factors)		
Iso/Aniso (#)	4107/0	
min/max/mean		
Protein	13.60/26.62/18.75	
Nucleotide	---	
Ligand	14.47/22.87/18.67	
Water	---	
Occupancy		
Mean	1	
occ = 1 (%)	100	
0 < occ < 1 (%)	0	
occ > 1 (%)	0	
Data		
Box		
Lengths (Å)	96.60, 99.36, 57.27	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	3.5	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	3.7/---/---	3.7/---/---
d model	3.8	3.8
d FSC model (0/0.143/0.5)	3.4/3.5/3.8	3.5/3.6/4.3
Map min/max/mean	-22.12/27.56/0.49	
Model vs. Data		
CC (mask)	0.78	
CC (box)	0.42	
CC (peaks)	0.17	
CC (volume)	0.68	
Mean CC for ligands	0.77	