Leginon: New features and applications

Anchi Cheng, Carl Negro, Jessica F. Bruhn, William J. Rice, Sargis Dallakyan, Edward T. Eng, David G. Waterman, Clinton S. Potter, Bridget Carragher

Published version information

This is the peer reviewed version of the following article:

Citation: A Cheng et al. 'Leginon: new features and applications.' Protein Science, vol. 30, no. 1 (2020): 136-150.

DOI: 10.1002/pro.3967

Protein Science
DOI10.1002/pro.3967

Leginon: New Features and Applications

Anchi Cheng[1], Carl Negro[1], Jessica F. Bruhn[2], William J. Rice[3,4], Sargis Dallakyan[1], Edward T. Eng[1], David G. Waterman[5,6], Clinton S. Potter[1,7], Bridget Carragher[1,7]

[1] The National Resource for Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, 89 Convent Ave, New York, NY, 10027, USA.
[2] NanoImaging Services, San Diego CA.
[3] Cryo-Electron Microscopy Core, New York University School of Medicine, New York, NY 10016, USA.
[4] Department of Cell Biology, New York University School of Medicine, New York, NY 10016, USA
[5] STFC, Rutherford Appleton Laboratory, Didcot OX11 0FA, England.
[6] CCP4, Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, England.
[7] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

For correspondence: Bridget Carragher (bcarr@nysbc.org)

Short Title: Leginon: New Features and Applications

Keywords: cryoEM, automation, microED

Received: Aug 17, 2020;Revised: Sep 28, 2020;Accepted: Sep 29, 2020

Abstract

Leginon is a system for automated data acquisition from a transmission electron microscope. Here we provide an updated summary of the overall Leginon architecture and an update of the current state of the package. We also highlight a few recent developments to provide some concrete examples and use cases.

1. Introduction

Leginon is a modular system for automatically acquiring images from a transmission electron microscope (TEM).[1-4] This software has been used for many years (the primary papers have been cited ~800 times) to collect data for either single particle cryo electron microscopy (cryoEM) or cryo electron tomography (cryoET). Recently, functionality has also been added to provide for data collection for microcrystal electron diffraction (microED)[5] an approach of interest to organic chemists and structural biologists.

High-resolution electron microscopy requires a series of targeting and image acquisition steps designed to obtain the highest quality high magnification images in the shortest amount of time. CryoEM single particle analysis and cryoET projects require imaging vitrified samples suspended as a thin layer over a fenestrated (holey) substrate, typically carbon or gold, supported by an electron microcopy grid consisting of a metal mesh (see Fig. 1). The goal of imaging is to locate holes where the vitrified ice is of optimal thickness and the particles are well distributed in the ice, adjust certain microscope parameters, and then acquire a final high magnification image or movie tilt series. For microED, the targets of interest are potentially well-ordered microcrystals with suitable dimensions for data collection (see section 3.1). Note that the overall quality and distribution of sample across a typical EM grid does not allow for a brute force approach to targeting where, for example, high magnification images would be systematically acquired as a raster across the EM grid; this would lead to a very high number of unsuitable images and be unacceptably inefficient given the high cost of using these instruments. Leginon thus manages and automates the task of selecting targets, at a series of increasing magnifications, with the goal of selecting optimal targets to provide high-quality high-magnification movies; once set up it can

run uninterrupted and unattended for many days at a time.

There are a number of other academic and commercial automated, or semi-automated, software systems that perform similar data collection (reviewed in[6]), most notably SerialEM[7] and EPU (Thermo Fisher Scientific). Leginon is unique among these in its centralized approach to multiple microscopes and project management. Leginon is closely integrated with a web server that provides users with live feedback and the opportunity for remote collaborations during data collection. In addition, the Leginon database and webviewer also serve as the basis for live analysis of the data using Appion,[8] software that "wraps" a large number of other packages. The close links with a database, a webviewer and integration with Appion means that preliminary analysis allows data filtering and feedback to be concurrent with the data collection.

The current Leginon framework was published in 2005.[3] Updates to support new microscopes, cameras and peripheral services are included in version releases on an ongoing basis[1]. Other functionalities added since the initial publication have been scattered in various technical publications and/or with papers focused on the biological results. These include support for tomography,[4] robotic grid screening,[9,10] live processing with Appion,[8] random conical tilt data collection,[11] specialized schemes for tilted image collection,[12] ice thickness measurement,[13] and various use cases.[14-17] The goal of this paper is to provide an updated summary of the overall Leginon architecture and an update of the current state of the package. We also highlight a few recent developments to provide some concrete examples and use cases.

2. Design

[1] Leginon.org

Leginon architecture (Fig. 2) has not fundamentally changed from that described in the original publication in 2005. At its base is a database for storing and accessing metadata, and a file system for storing acquired raw images and movies as well as processed data. Leginon accesses these two systems to perform two main functions:

- Instrument controls, written in python, using an object-relational mapping (ORM) system developed specifically for the project. These controls also require scripting access to the instrument, usually provided by the instrument manufacturer, or sometimes by third-party middleware, e.g. in the case of the Gatan camera access is via SEMCCD.dll from SerialEM.
- Data display as a web service through the software bundle LAMP (Linux, Apache, MySQL, PHP). Originally, the operating system used for the web service software bundle was not limited to Linux. However, to reduce support overhead, we now only officially support the more stable version of CentOS.

At an institution like ours[2], multiple microscopes, multiple projects, and multiple user groups means that a high level of organization is critical. The centralized database underlying Leginon provides the organization required to support large scale management and also makes possible live-processing through Appion. Appion is distributed with Leginon as an add-on functionality. Appion uses the same database, file storage system, and the same ORM in its python wrapper scripts, and the web service acts as its user interface. If processing is desired as an add-on functionality, Appion uses the same systems in its python wrapper scripts. With so many rapid developments of processing software, we have found it difficult to keep Appion current

[2] semc.nysbc.org

and we no longer recommend it as a full workflow package that proceeds all the way to 3D reconstruction and refinement. Instead, we focus on offering real time feedback during data collection, including full functionality and analysis for motion correction, ice thickness evaluation, CTF fitting, and particle picking (Fig. 3). We also offer tools for importing and exporting data to and from Appion. For example, users can export a star file including CTF estimates from Appion or import 2D classification and 3D refinement jobs from cryoSPARC,[18] and can import the final results from a cryoSPARC run for archiving. For each data collection session, we have also added an option to display an overall summary of the experiment in a report form (Fig. 4). This report can be sent as a PDF file that can be shared with collaborators or archived for future reference.

From an end user's point of view, the instrument control python programs are workflows made from functional modules called "nodes". The position of the nodes in the workflow in relation to others are defined by bindings. The basic workflow of Leginon requires multi-scale imaging (MSI) (Fig. 1). At each scale, an *acquisition* node outputs child images taken at a given set of microscope and camera conditions based on input target positions defined on an existing parent image. A *targeting* node follows each imaging node that accepts an image as its input and outputs the targets chosen within the node either automatically, or optionally, manually as selected by the user. Images are acquired at progressively higher magnifications; at the highest magnification where the final desired data is recorded, a specific subclass of acquisition nodes is used to perform specialized tasks such as acquiring direct detector movies, tomography tilt series, or diffraction tilt series. In addition, each node class instance can have unique settings in the workflow to give it a more refined behavior. A number of accessory nodes handle management

of the microscope and camera preset values, autofocusing, adjustment of the targets to account for specimen drift when returning to targets queued up previously, etc. A recently developed accessary node provides ice thickness measurements for every high magnification image.[13] Together, the nodes, bindings, and the corresponding settings can be saved as an application. While any user can create their own applications, most users use the many applications that are provided standard with Leginon (see Table 1 for some of the more common applications).

Leginon is distributed as open-source under the Apache2 license. The code and documentation are freely available[3] and a Docker container image is available[4] that provides simulated data collection so as to provide an opportunity to learn the overall operation, set up of automated target finding, and the general feel of running an automated data collection system.

3. Major New Features

Since 2005, approximately 17000 revisions have been made to the Leginon and Appion code base. These correspond to numerous enhancements and advances; only a few of the most illustrative or important ones will be described here. A more comprehensive list is available at leginon.org.

3.1 Micro Crystal Electron Diffraction Workflow

[3] leginon.org
[4] github.com/nysbc/leginon-tutorial

The philosophy behind Leginon is both to provide a full workflow (from project selection, to data collection, and data display), and to provide good quality assurance to maximize efficiency. The recently developed microcrystal electron diffraction application, MSI-Diffr, workflow provides an illustrative example (Fig. 5).

Continuous rotation electron diffraction using a transmission electron microscope (TEM), also known as microcrystal electron diffraction (MicroED), continuously records rotation data in diffraction mode from three-dimensional crystals without applying a beam tilt. Due to the much stronger interactions between electron and matter, significantly smaller crystals, typically 0.2-1 µm in length, can be used compared to the much larger crystals required for X-ray crystallography (~5-10 µm for a synchrotron source and ~50 µm for a lab source). The method has been developed over several years for protein crystals[19] but recently gained attention when it was demonstrated that the method could also be applied to small organic crystals, to produce data at sub-Angstrom resolution which can be phased using *ab initio* methods.[20,21] This application to small molecules (limited to <1000 Da) in particular makes this a valuable tool for medicinal chemistry, chemical biology, natural product research, organic semiconductor research and synthetic organic chemistry. We developed a Leginon implementation of the MicroED workflow for a TFS Glacios equipped with Ceta-D camera, i.e. not requiring a frame saving image server. The application is as simple to use as the tomography application, only requiring the user to manually target potential crystals of interest. It includes accurate crystal position recall from a queued list of targets through iterative movement in image search mode, a eucentric height correction for each target, extra lens normalization required by the diffraction mode setup, automated aperture selection, coordination of imaging to diffraction mode switching with

automatic beamstop insertion and retraction, and synchronized stage rotation with continuous data collection. It also uses the true rolling-shutter mode of the camera which avoids gaps in data collection. The recorded diffraction series are uploaded into the Leginon database along with other metadata, and also automatically converted to crystallographic SMV format.

Two specialized nodes were added to Leginon in this implementation: (i) a preview of potential crystals and (ii) acquisition of rolling-shutter movie series in diffraction mode while the specimen stage is continuously tilted. Furthermore, we added workarounds to address the lack of official scripting control of the beamstop as well as the initiation and saving of true rolling shutter movie acquisition. This was achieved using AutoIt[5] scripting called within Leginon functions. AutoIt interacts with the graphical interface of the microscope control system running on the Windows PC to control required functions. AutoIt can access most Windows objects from their class instance rather than the primitive display position used by other programs; this has the advantage of higher reliability. Overall, these features deliver a high success rate and high quality for automated microED data collection.

The crystal preview node is an acquisition node subclass. It is a good example of the quality assurance tools built into Leginon. Users can choose to collect a single diffraction image, over a small continuous tilt range, using this node to help decide whether a potential crystal should be included in the queue of crystals destined for full automated data collection. A single diffraction image can be used to estimate resolution, identify pathologies such as streaky spots (high mosaicity) and the presence of multiple lattices, determine if a crystal is too thick to penetrate, provide a rough estimation of unit cell dimensions (important for distinguishing contaminants from

[5] https://www.autoitscript.com/site/)

the target of interest), and can guide the user to select the most optimal data collection strategy (particularly camera length selection and dose per frame). Using this node ensures that the operator only queues the highest quality crystals for final data collection with an optimal collection strategy, improving the overall efficiency and quality of the data collected. A small continuous tilt range is used because electron diffraction can be very sparse at a fixed angle. The default tilt range for a preview image is 5°, but any value can be specified by the operator.

The second specialized node performs the main tasks of continuous rotation electron diffraction: centering each crystal in X, Y and Z, setting up diffraction imaging conditions, inserting the beamstop, and continuously tilting the crystal while acquiring diffraction images in rolling shutter mode. In the microED implementation in SerialEM[22] centering the crystal and beamstop position were performed manually for each target during a one-time set up for microED collection; this was required because the microscopes manufacturer did not provide external control of the beamstop. Following this initial setup in SerialEM, each crystal was centered by recalling and setting a saved stage position, and a diffraction movie was acquired using a multithreaded global shutter mode. Since digital cameras do not allow for acquiring exposures during the readout phase of this shutter mode, gaps in the "continuous" movie occur and can be up to 30% of the cycle. This is not the case in rolling shutter mode. In the Leginon implementation, each crystal is iteratively centered to an accuracy of 0.1 µm, monitored by acquiring images in imaging mode. This level of accuracy is especially important when operating with a 20 µm C2 aperture which on our Glacios corresponds to a beam diameter of ~0.6 µm, which is often about the same size as, or smaller than, the dimensions of the crystals of interest. A very small beam is advantageous because it reduces the background signal from the grid substrate and other crystals/debris that

might be in the beam path and eliminates the need for a selected area aperture.

Once the crystal has been adequately centered, Leginon switches out of imaging mode and into diffraction mode. It is very important that the beam remains stably centered behind the beamstop after this switch to diffraction mode to protect the camera and ensure the quality of the data. A series of lens normalizations were added to achieve the required reproducibility and reduce beam movement during collection. Next, the stage is tilted to the selected start angle and then the stage is rotated at the selected tilt speed for the selected tilt range and diffraction data is acquired. To ensure that a constant speed is achieved, a two second delay was added before the beam is un-blanked and data collection is initiated. As with other acquisition nodes in Leginon, many targets can be selected and then a large batch can be submitted for automatic data collection. Overnight data collection can collect hundreds of datasets from a single grid without the need for user intervention. Overall, these features reliably deliver high quality data in a highly automated fashion.

The Leginon full workflow approach does not end when the data is written onto the local disk. For example in the microED application, services are made available to automatically transfer the raw movies to a central storage system, and upload these, and their associated metadata, to the Leginon database, where they may be viewed and assessed by anyone with permission and access to a web browser [see Fig. 3(b)]. This viewer displays diffraction images paired with images of the targeted crystal and includes a ruler to measure the resolution of a given reflection. The movies are also converted to SMV crystallographic format with all of the necessary metadata required for data processing in standard crystallographic software packages. As SMV format does not allow for negative pixel values, an offset value is added to all pixel values. This

value is reported in the image header as the LEGINON_OFFSET and corresponds to the most negative value recorded. Additionally, a suggested pedestal value (IMAGE_PEDESTAL) is determined by examining the minimum pixel values across the dataset after applying the offset and selecting the largest of these. For crystallographic software packages that allow for negative pixel values, such as DIALS,[23] this pedestal can be applied to better approximate the true baseline. Additionally, a DIALS format class was created for data collected using a Ceta-D camera operated in rolling shutter mode[6]. In the future we plan to incorporate DIALS into Appion to provide on-the-fly data processing to help the operator collect the best possible data as efficiently as possible.

3.2 Ice Thickness Measurement

A critical aspect of acquiring useful cryoEM images is selecting targets with optimal vitreous ice thickness. Ice that is too thin can exclude particles, whereas ice that is too thick results in both addition of unnecessary noise and increases the possibility that particles will be overlapping in the final projection image. Somewhat surprisingly, measurement of ice thickness during data collection has become routine only recently.[13] Leginon includes two tools to measure ice thickness of final high-magnification images. For microscopes with an energy filter, it uses the ratio of intensity of images with and without the slit inserted. By calibrating the apparent mean free path of electrons through ice and providing this value in the Leginon settings, Leginon can automatically determine the ice thickness. For microscopes without an energy filter, a comparison of intensity over vacuum with intensity over ice allows determination of ice thickness using Beer's

[6] https://raw.githubusercontent.com/dials/dxtbx_ED_formats/master/FormatSMVCetaD_TUI.py

Law. Again, a scattering coefficient must be provided to the Leginon interface. Mean-free paths and scattering coefficients for several microscope configurations are listed in[13]. We have also since added:

Glacios or TFS Talos Arctica (200 keV), no energy filter; 100 µm objective; ALS = 1055;

TFS Talos Arctica (200 keV), no energy filter; 70 µm objective; ALS = 745.

While these are good starting values for Leginon users they should be verified using tomography for a specific instrument. Using the energy filter provides the most accurate method as it does not require a measured fixed intensity and so it is unaffected by changes in beam intensity. It does, however, add several seconds of time to each image, since two additional images need to be taken and the energy filter slit needs to be removed and inserted. In practice, the ice thickness can be measured using both methods during screening to establish that the ALS method is providing accurate results, and then the EF method need only be done once in a while (e.g. every 100 images), to ensure that the beam intensity is holding steady.

While it is advantageous to measure the ice thickness for all images, it is even more valuable to correlate this final ice thickness measurement to the nominal ice thickness estimated from the intensity in the holes using lower magnification images. This is useful as it is often difficult to estimate ice thickness by eye, particularly on gold grids. The Leginon Summary page provides a plot of the hole thickness estimation value versus the final, measured thickness in the high magnification images (Fig. 6). The graph is linear, with some scattering, and correlation coefficients are generally above 0.9. Using this correlation, it is straightforward to have Leginon only acquire images of targets which are in the desired thickness range. Users at both New York Structural Biology Center (NYSBC) and New York University (NYU) find this feature exceptionally

valuable, as it allows them to focus on taking images with optimal ice thickness using the limited microscope time they have available. We are currently exploring methods to predict ice thickness at an even earlier step in the process with the goal of indicating estimated thickness values in the atlas view.

3.3. Phase plate characterization

An extensive array of Volta phase tools have been developed for Leginon over the last few years and continue to be available despite the current relatively low interest in this once highly anticipated hardware option. The features include a fully automated workflow for MSI imaging, charging of the phase plate spot, and moving to the next spot after a user-defined number of images has been acquired. The carbon areas of the phase plate do not behave uniformly, and thus some areas are unsuitable for imaging. A Leginon node was created for characterizing all phase plate spots and unsuitable spots are automatically avoided during automated data collection. We also control the direction of the beam tilt used in the autofocusing routine to direct the focusing position to a spent phase plate spot rather than a fresh phase spot.

3.4. Leginon Remote

The Leginon main user interface (Fig. 7) is written in wxPython. It is feature rich and comes with several automated algorithms for selecting targets. However, we have found a significant number of users that require very specific targeting, for example to identify sparse filaments or during initial screening to identify which hole, or region inside a hole, yields the best ice and particle density. Our facility, based at the Simons Electron Microscopy Center (SEMC),

serves the New York cryoEM and cryoET community through nine institutional memberships to the NYSBC. We also serve the national community through the National Resource for Automated Molecular Microscopy[7] and the National Center for CryoEM Access and Training[8]. Many users would like at least limited remote control of grid targeting. This can be provided through VNC or some other remote-desktop software, which is used routinely by our staff. However, many external users are not seasoned microscopists and would face a steep learning curve in using the powerful Leginon GUI, as well as potentially doing some harm to the microscope if provided with full control of all options. In addition, there are potential security concerns in providing VPN access to a very large number who are widely geographically dispersed.

To address the need for limited off site control, we have developed the "Leginon-Remote" application, which provides users with the ability to remotely select targets, pause and continue data collection sessions, and chat with on-site staff, all through a standard web browser (Fig. 8). The application is written in Python and Javascript using the Django web development framework. It uses a Nginx web server, a Postgres database to store metadata, Slack for live user-staff communication, and Docker for deployment, data, and service management. While a full Leginon installation is required to use "Leginon-Remote", the source code and deployment are isolated from Leginon and the two applications communicate via REST HTTP requests. "Leginon-Remote" is open source distributed under the MIT license and instructions for installation, configuration, and operation are available via github[9].

[7] nramm.nysbc.org
[8] nccat.nysbc.org
[9] https://github.com/nysbc/leginon-remote

3.5 Aberration correction of large beam-image shift

Large beam-image shifts are used to target many holes for every stage move and defocus setting.[24] This approach significantly speeds up data collection at the cost of introducing aberrations into the images. As an averaging technique, cryoEM does not appear to require every image to be perfect as long as the average converges to the correct answer.[17] However, significant aberration affects CTF estimation as well as FSC-based half-maps assessment of resolution.[25] Fortunately, the strongest aberrations induced by beam-image shift, mainly coma and astigmatism, can be corrected both during and after data collection (Relion 3.1, CryoSPARC 2). The calibrations necessary to achieve coma correction during data collection are semi-automated; implementation of this feature in Leginon is similar to that in SerialEM (Xu and Mastronarde, personal communication). In general, these calibrations are sufficient to treat all images as being in the same beam-tilt group when final beam tilt corrections are performed in software using Relion or CryoSPARC. However, if these calibrations are somewhat off, or if the coma-correction was accidentally turned off in Leginon, then it is critical to correct for these effects to obtain the highest resolution. Since images can be collected with image shifts of up to 8 µm in various directions, they need to be divided into similar image shift groups prior to software correction. As all data collection parameters are stored in the Leginon database a simple program, tiltgroupwrangler.py, is provided to sort the tilt groups. It requires downloading the star file from Leginon that contains CTF values and beam tilts for all groups. After this data is read into tiltgroupwrangler.py, a plot is made of all x and y image shift values and a slider is used to determine the number of groups (Fig. 9), with 50-100 being sufficient. Grouping is achieved using k-means clustering, and the program outputs a Relion 3.0 file with images divided into groups.

Beam tilt corrections can then be performed in software by treating the particles from images in these groups as separate sets. A module is also available to do the same grouping for particles processed with CryoSPARC, though this requires the availability of a CryoSPARC python library to write the required cs files.

4. Quality Assurance

Leginon focus on the quality of data collection by using an intelligent, rather than brute-force, approach to selecting targets for imaging. Various hole-finding algorithms already described in the original publication and in the on-line manual are aimed at being selective in the targeting process. Overall our goal is to make the use of the microscope more efficient by monitoring its activity, and to improve image quality by analyzing the quality during the data collection.

***Appion Image Rejector*:** Image rejector analyzes the movie drift speed, CTF estimation confidence scores, and presence of ice crystal and automatically hides movies that fail user specified criteria. The ice crystal detection checks the radial average of power spectra for the presence of a sharp peak at the expected resolution. The list can also be used to remove poor quality data to reduce storage usage.

Error and idle notifications: Despite best efforts by the microscope and camera manufacturers, as well as of the automation software developers, hardware and software problems may still arise

during an automated data collection session. Microscope operators would like to monitor data collection even during off-hours to ensure data quality. To reduce stress on the operators while also minimizing idle time, Leginon actively reports unusual events through text messages or emails. We currently use the Slack Application[10] to post these notifications to a specific channel that microscope operators subscribe to. Leginon also monitors active usage of the microscope and the camera and if the instrument is idle for a user specified set time, a Slack notification is posted. These idle notifications catch user errors as well as instrument errors that stall the workflow and have been very valuable in improving the overall efficiency of data collection.

Black Stripe Detector: In our experience using the Gatan K2 Summit detector, we have noticed that images will occasionally include large areas that are blank. The camera is divided into a series of 8 "stripes", and data from each stripe is sent to the digitizer independently. Software issues, which are still not well characterized, occasionally cause one or more of these channels to lose connection, resulting in a "black stripe" of all zero data from this channel. As we operate our microscopes on a 24/7 schedule, this problem quite often occurs during hours when staff are not actively observing data collection. The "Black Stripe" node was created to monitor this problem. It actively scans all high magnification images for the presence of a completely blank channel. Upon detection, it sends an error message to Leginon, and if the Leginon error notification is turned on, all EM staff receive this message via Slack. Quite often someone is available to pause the imaging and address the issue. Fortunately, we have found that in most cases, it can be solved by using the Gatan camera software to turn the camera off and on again.

[10] slack.com

Typically new gain references must then also be acquired. Both actions can be performed remotely and are generally completed within an hour. If someone does not attend to the matter the default option is to continue data collection, since occasionally the black stripes will disappear over time. Should we notice the same problem occurring on the Gatan K3 cameras, we will update this node to include these cameras.

***Report Generation*:** All information associated with data collection and analysis is stored in a database, which we query to generate and display summary tables and histograms of various parameters in the form of a "report" (Fig. 4). In particular, we include imaging and defocus summary tables, histograms of ice thickness, an image of the grid atlas, and a table providing experimental details including TEM parameters (magnification, high tension, defocus range, spot size, intensity, Cs and energy filter settings) and camera parameters (dimension, binning, exposure time, pixel size, dose rate, total dose, frame rate and total frames). We also provide Experimental Methods, Acknowledgment, and Reference sections. To include other information computed outside of Appion, we have implemented an option to import 2D classification and 3D refinement jobs from cryoSPARC into the report. Users can select which cryoSPARC server to connect to by indicating an IP address and entering the project and job ID of the 2D classification or 3D refinement jobs. The image of the 2D classes and the FSC curve computed by cryoSPARC is then also included in the report (Fig. 4). These reports are a nice way to observe the overall progress of the experiment and share this information with remote users or collaborators.

Benchmarking workflows: To ensure the instrumentation and software meets the needs of a

cryoEM or cryoET experiment, our workflows are subjected to regular benchmark tests.[17] While the information limit of the microscope is checked during daily routine alignments, we also run a full workflow benchmark on an annual basis and also after any major instrument upgrades or repairs. Test samples, such as apoferritin or aldolase, are used to collect data under standard conditions to identify any bottlenecks and track instrumentation performance. During these annual microscope benchmark tests, the full Leginon and Appion pipelines are used with on-the-fly feedback to verify that the IT infrastructure can handle continuous data collection. If issues are found, then preventative maintenance may be scheduled to maintain the capacity and efficiency of high-end instrumentation. The results of one of these benchmark tests are shown (Fig. 10).

Discussion

Advances in camera and microscope technology have allowed cryoEM data collection to become fairly routine. Since high-end microscopes and cameras are both very expensive and in very high demand, we try to provide the most efficient data collection and ensure that it is of high quality, while reducing the burden on the microscopists managing the instruments. Highlighted above are some of the tools we have added in the last few years to provide new features, improve efficiency, and strengthen quality assurance. With these enhancements, combined with the underlying database that provides and organizes records of all previous data collections, we believe Leginon is a good platform to use to improve future data quality.

One of our goals for the future is to routinely collect ~10,000 movies during a 24 hour user session, including screening time. This translates to over 400 movies per hour, or < 9 seconds

per movie. There are several limiting factors that must be addressed to achieve this aim. Firstly, since grid and specimen quality are still the most variable parameters that determine the success of cryoEM projects, screening of suitable areas on the grid is still a necessary step and often the most time consuming period for the microscope operator. It is not unusual that 4-6 hours are spent on this process as the operator and researcher collaborate to find the best grid and the best imaging strategy. The currently typically quoted data collection speeds usually ignore the time spent on this step. It would be highly desirable to automate this process by implementing a pipeline to manage grid exchanges unattended by an operator assisted by machine-learning algorithms to rapidly optimize targeting strategies based on previous experience on similar samples.

The core of the data acquisition speed limit is the time required to acquire and save consecutive movies. As a result, acquisition parameters can have a significant impact and for fast cameras such as the Gatan K3, can change the speed of data collection significantly. For example, the upper limit of the Gatan K3 camera data collection speed, using the SEMCCD.dll, changes depending on collection strategies, including super-resolution mode and using a multithreading acquisition known as “early-return” feature. A 50 frame movie collection can cycle within 3 seconds using the LZW Tiff format in standard image counting mode with 39 ms per frame. In super-resolution mode, twice the time, i.e., 6 seconds, is required for the same operation. The gains provided by super-resolution data quality have to be considered against the loss in number of images that can be acquire in a fixed period. Another popular strategy used on the K2 camera is the use of the SEMCCD.dll “early-return” feature where the acquisition request thread is returned before file saving is completed. When targeting time is dominated by stage

movements, this is a highly desirable method since the stage settling can occur while movies are being saved. However, we observed on the K3 camera, that even though the first acquisition of the super-resolution movie takes only 3 seconds, later movies have to wait for the previous movie to be saved, paying the penalty of buffering the movie in memory, and thus the time spent on subsequent movie acquisitions becomes 8 seconds. This time difference and the fact that it is common now to take many beam-image shift targeting that requires negligible settling time before any stage movement, means that the early-return multithreading option is now usually a disadvantage.

We also need to address some data acquisition overhead in Leginon. As a distributed and multi-instrument system, Leginon requires more overhead in transferring information than similar packages like SerialEM, EPU, and Lattitude. For example, to transfer a single Gatan K3 super-resolution dark-corrected image as an array (11520x8184 pixels; 180 MB) through a non-dedicated 10G network, followed by loading and applying a gain correction, and then saving the resulting image can take up to 6 seconds. We work around this limitation by transferring a reduced size image. However, there are almost certainly more time savings to be discovered.

Another obvious limiting factor to data collection speed is the time required for lens normalization. Electro-magnetic lenses and deflectors in electron microscopes exhibit strong hysteresis requiring elaborate lens normalizations to ensure reproducible results when changing settings such as magnification and spot size etc. When beam size, magnification, and defocus are all changed, as required in the MSI workflow, multiple lens normalization become necessary, and it can take 6-9 seconds to complete them all. These time-consuming processes will need to be consolidated without losing the benefits to further improve data collection speed.

Currently, Leginon still has a 2-3 second overhead, in the best case, from each beam-image shifted movie as a result of calibration queries, calculating the required changes, changing the instrument parameters, and waiting for instrument stabilization. The rate of data collection, including overhead from autofocusing and nitrogen filling averages between 240-400 movies per hour. The rate is determined by the exposure time, the total number of frames and the number of beam-image shifted images targeted per stage move.

Other factors which affect collection speed are out of control of software but could be addressed by future hardware improvements. Nitrogen refills take about 10 minutes every 8 hours on the TFS microscopes. A larger dewar inside the microscope would reduce this frequency. Energy filters in our experience need to have slit centering checked every 1-2 hours, which takes about 3 minutes. More stable filters or better room temperature control would help reduce this frequency. Dark references on the direct detectors need to be updated several times per day, and ideally this requirement could be reduced. Finally, grid quality is often not ideal. If there are many holes which are too thick or too thin, the number of good targets per stage movement is reduced, which effectively slows down the collection. We generally aim to collect a high proportion of good images rather than collecting the absolute maximum number possible and letting the user sort out the bad ones post-collection.

CryoEM data collection automation has come a long way since the last paper describing Leginon was published in 2005. We will continue to update it to use new instruments and add features to improve the efficiency of high-quality data collection and assessment.

Acknowledgements

We are grateful to the staff of the Simons Electron Microscopy Center at the New York Structural Biology Center for input and patient testing. The work presented here was conducted at the National Resource for Automated Molecular Microscopy located at the New York Structural Biology Center, supported by grants from the NIH (GM103310, OD019994) and the Simons Foundation (SF349247).

Table 1: Standard MSI applications in Leginon

Name	Primary use
MSI-T	Single particle data collection (typically from holey C or Au EM grids)
MSI-Tomography	Cryo electron tomography data collection
MSI-Raster-Screen	Data collection from multiple grids (typically for negative stain screening)
MSI-PP-Stitch	Data collection using phase plates
MSI-RCT	Data collection using random-conical tilt geometry
MSI-T Tilt	Data collection using a tilted stage
MSI-Diffr	Data collection for micoED

References

1. Carragher B, Kisseberth N, Kriegman D, Milligan RA, Potter CS, Pulokas J, Reilein A **(**2000) Leginon: an automated system for acquisition of images from vitreous ice specimens. J Struct Biol 132:33-45.
2. Potter CS, Chu H, Frey B, Green C, Kisseberth N, Madden TJ, Miller KL, Nahrstedt K, Pulokas J, Reilein A, Tcheng D, Weber D, Carragher B (1999) Leginon: a system for fully automated acquisition of 1000 electron micrographs a day. Ultramicroscopy 77:153-161.
3. Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, Stagg S, Potter CS, Carragher B (2005) Automated molecular microscopy: the new Leginon system. J Struct Biol 151:41-60.
4. Suloway C, Shi J, Cheng A, Pulokas J, Carragher B, Potter CS, Zheng SQ, Agard DA, Jensen GJ (2009) Fully automated, sequential tilt-series acquisition with Leginon. J Struct Biol 167:11-18.
5. Nannenga BL, Shi D, Leslie AGW, Gonen T (2014) High-resolution structure determination by continuous-rotation data collection in MicroED. Nat Methods 11:927-930.
6. Cheng A, Tan YZ, Dandey VP, Potter CS, Carragher B (2016) Strategies for automated cryoEM data collection using direct detectors. Methods Enzymol 579:87-102.
7. Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152:36-51.
8. Lander GC, Stagg SM, Voss NR, Cheng A, Fellmann D, Pulokas J, Yoshioka C, Irving C, Mulder A, Lau PW, Lyumkis D, Potter CS, Carragher B (2009) Appion: an integrated, database-driven pipeline to facilitate EM image processing. J Struct Biol 166:95-102.
9. Cheng A (2013) Automated grid handling and image acquisition for two-dimensional crystal screening. Methods Mol Biol 955:297-305.
10. Hu M, Vink M, Kim C, Derr K, Koss J, D'Amico K, Cheng A, Pulokas J, Ubarretxena-Belandia I, Stokes D (2010) Automated electron microscopy for evaluating two-dimensional crystallization of membrane proteins. J Struct Biol 171:102-110.
11. Yoshioka C, Pulokas J, Fellmann D, Potter CS, Milligan RA, Carragher B (2007) Automation of random conical tilt and orthogonal tilt data collection using feature-based correlation. J Struct Biol 159:335-346.
12. Tan YZ, Baldwin PR, Davis JH, Williamson JR, Potter CS, Carragher B, Lyumkis D (2017) Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nat Methods 14:793-796.
13. Rice WJ, Cheng A, Noble AJ, Eng ET, Kim LY, Carragher B, Potter CS (2018) Routine determination of ice thickness for cryo-EM grids. J Struct Biol 204:38-44.
14. Stagg SM, Lander GC, Pulokas J, Fellmann D, Cheng A, Quispe JD, Mallick SP, Avila RM, Carragher B, Potter CS (2006) Automated cryoEM data acquisition and analysis of 284742 particles of GroEL. J Struct Biol 155:470-481.
15. Cheng A, Leung A, Fellmann D, Quispe J, Suloway C, Pulokas J, Abeyrathne PD, Lam JS, Carragher B, Potter CS (2007) Towards automated screening of two-dimensional crystals. J Struct Biol 160:324-331.

16. Stagg SM, Lander GC, Quispe J, Voss NR, Cheng A, Bradlow H, Bradlow S, Carragher B, Potter CS (2008) A test-bed for optimizing high-resolution single particle reconstructions. J Struct Biol 163:29-39.
17. Kim LY, Rice WJ, Eng ET, Kopylov M, Cheng A, Raczkowski AM, Jordan KD, Bobe D, Potter CS, Carragher B (2018) Benchmarking cryo-EM single particle analysis workflow. Front Mol Biosci 5:50.
18. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290-296.
19. Nguyen C, Gonen T (2020) Beyond protein structure determination with MicroED. Curr Opin Struct Biol 64:51-58.
20. Gruene T, Wennmacher JTC, Zaubitzer C, Holstein JJ, Heidler J, Fecteau-Lefebvre A, De Carlo S, Muller E, Goldie KN, Regeni I, Li T, Santiso-Quinones G, Steinfeld G, Handschin S, van Genderen E, van Bokhoven JA, Clever GH, Pantelic R (2018) Rapid structure determination of microcrystalline molecular compounds using electron diffraction. Angew Chem Int Ed Engl 57:16313-16317.
21. Jones CG, Martynowycz MW, Hattne J, Fulton TJ, Stoltz BM, Rodriguez JA, Nelson HM, Gonen T (2018) The cryoEM method MicroED as a powerful tool for small molecule structure determination. ACS Cent Sci 4:1587-1592.
22. de la Cruz MJ, Hattne J, Shi D, Seidler P, Rodriguez J, Reyes FE, Sawaya MR, Cascio D, Weiss SC, Kim SK, Hinck CS, Hinck AP, Calero G, Eisenberg D, Gonen T (2017) Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. Nat Methods 14:399-402.
23. Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID, Sauter NK, Evans G (2018) DIALS: implementation and evaluation of a new integration package. Acta Cryst D74:85-97.
24. Weis F, Hagen WJH (2020) Combining high throughput and high quality for cryo-electron microscopy data collection. Acta Cryst D76:724-728.
25. Bromberg R, Guo Y, Borek D, Otwinowski Z (2020) High-resolution cryo-EM reconstructions in the presence of substantial aberrations. IUCrJ 7:445-452.

Figures

Figure 1. *Multi-Scale Imaging and Leginon accessory functions*: Images at higher magnification are acquired by defining targets on the parent images. Acquisition and Targeting node classes string these processes together. Additional targets can be added to perform other automated tasks. Focusing targets (blue) are used to adjust eucentric height and perform low-dose focusing. Reference targets (violet) are used to perform periodic alignments such as the energy filter zero-loss slit. The reference node may also pass data to an accessory node to do additional tasks such as ice thickness measurement.

Figure 2. *Leginon architecture:* The microscope, cameras and other instruments (phase plates, energy filters, etc.) are controlled by facility staff or local users through the Leginon user interface (see Fig. 7). Database and File Servers store the data and metadata which is viewed using a web browser; Appion also uses the web viewer as a user interface for live preprocessing (see Fig. 3). A new web interface was recently added to provide remote manual targeting and on-site support requests (see Fig. 8).

Figure 3. *Web viewers:* (a) Integrated Leginon and Appion web tools provide real-time display of acquired images and results from on-the-fly particle picking, CTF estimation, motion correction and ice thickness. (b) Similar tools are available for observing and monitoring microED data collection.

Figure 4: *Summary report:* Data summarized includes (a) image statistics; (b) ice thickness; (c) defocus and CTF estimates; (d) overall grid atlas; (e) experimental setup; (f) experimental methods; (g) image processing results uploaded by the operator.

Figure 5. *Leginon microED workflow:* First, an atlas is collected to locate good grid squares with crystals. The user then selects grid squares and areas of interest within grid squares for higher magnification images. These images are fed into a queue of images for user targeting. The user can then target suitable crystals from these images and add them to the exposure collection queue. The user has the option to test the diffraction of any crystal with Diffraction Preview Mode. This immediately takes a single diffraction image from the targeted crystal so that the user can decide whether to add that crystal to the queue or omit it based on a visual inspection of the diffraction pattern displayed in the Web Viewer [Fig. 3(b)]. Once a sufficient number of crystals have been targeted, the user submits the queue. Leginon then automatically moves to each crystal target, and performs eucentric focusing. The stage is then set to the starting tilt angle and rotated at a constant speed to the desired end point while the camera records diffraction data at a set camera length in rolling shutter mode. The image files are saved, and converted to MRC format, for the web viewer, and SMV format for import into standard crystallographic software. This process repeats until all crystal targets have been processed. Note that the data collection parameters can be altered by the user at any time during the execution of the queue of crystal targets.

Figure 6. *Ice thickness monitoring***:** Plot of predicted hole ice thickness versus measured high

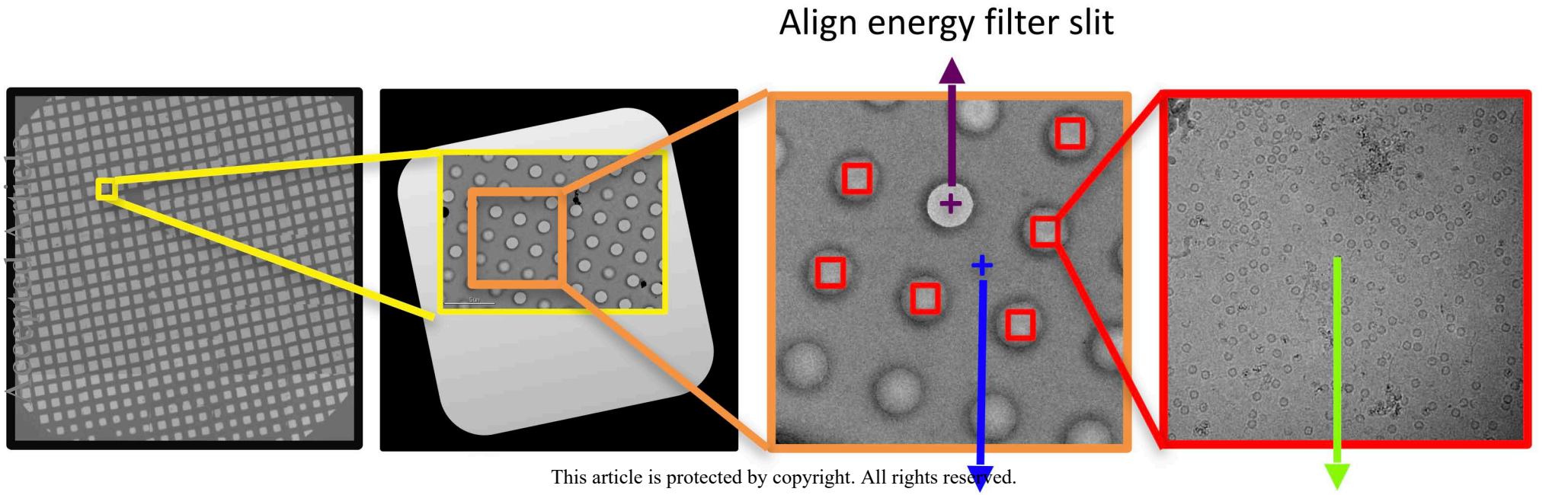
magnification target ice thickness indicates close correlation. To obtain ice thickness between 20 and 80 nm, the hole thickness values should be constrained to values from 0.05 to 0.14 or more precise values can be calculated from the fit.

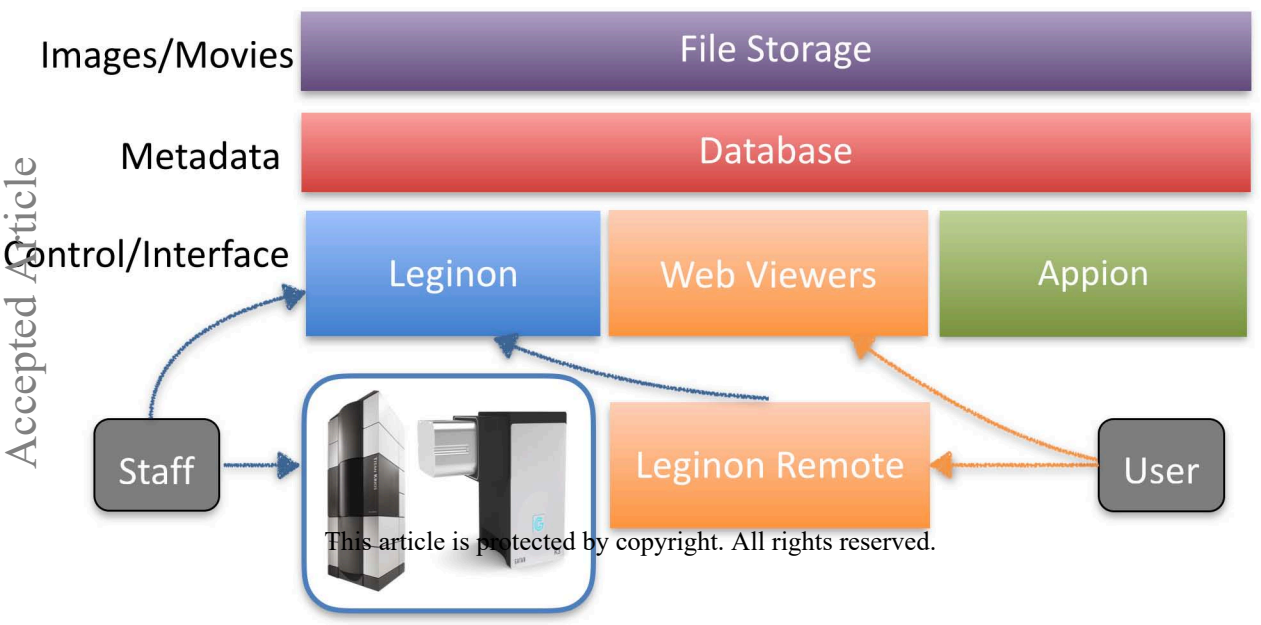
Figure 7. *Leginon GUI interface:* (a) Workflow list and node selection. (b) Toolbar for the selected node in (a), showing Exposure_Targeting in this case. (c) Continuous log of events including warning and errors. (d) Node information panel, in this case providing image statistics. (e) Auto-targeting types and settings arranged by processing steps.

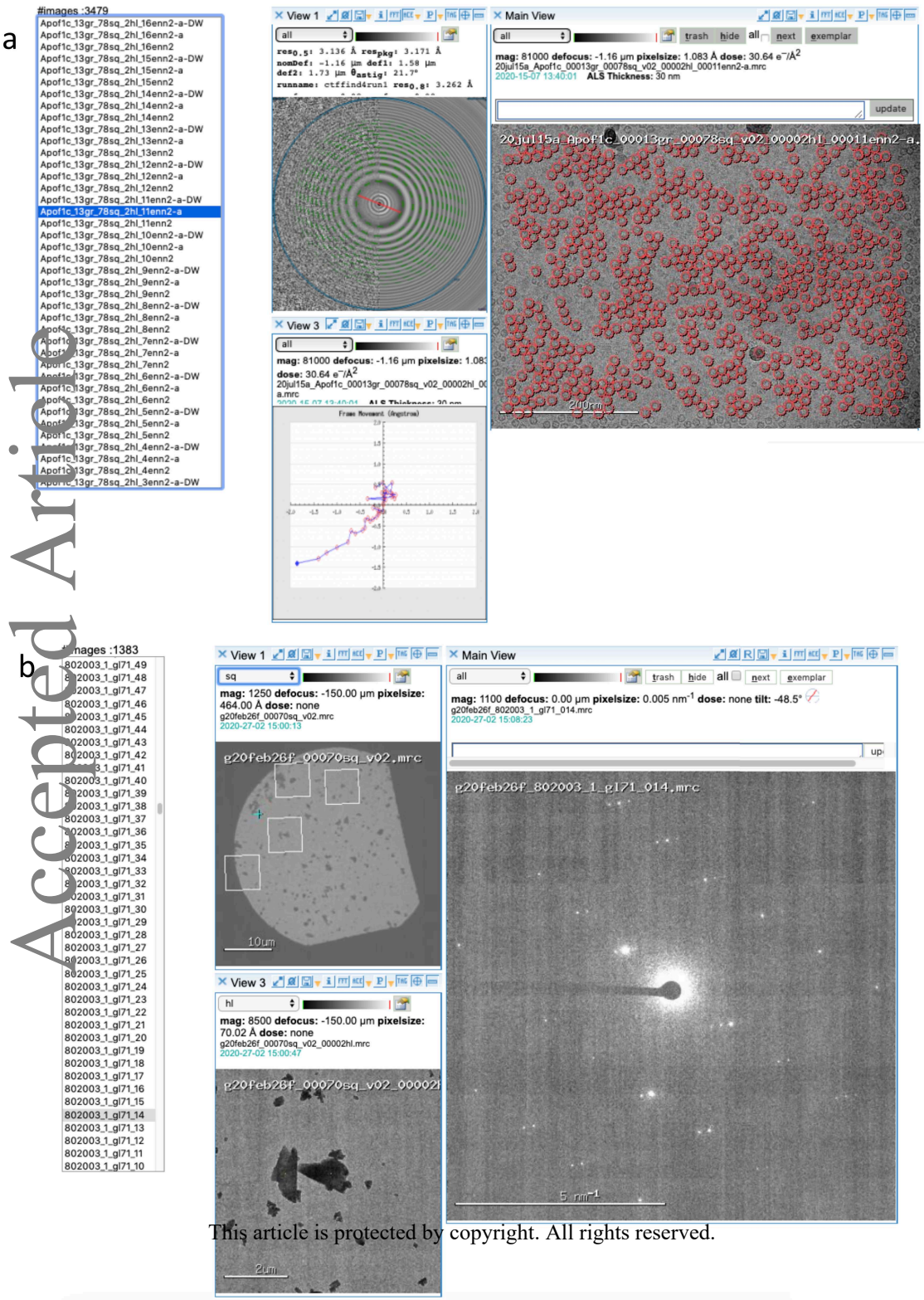
Figure 8. *Leginon-remote interface:* (a) Interactive target selection; (b) Leginon operation status showing nodes requiring remote intervention and nodes that are busy; (c) remote tools available include: grid atlas refresh, queue submission, pause/resume, and column valve closing; (d) session progress; (e) live chat with staff. [need some targets on this image]

Figure 9. (a) Main interface of tiltgroupwrangler.py. Using the star file provided by Leginon as input, various tools allow examination and modification of the information; a slider is used to select the number of groups to determine. (b) Plot of the tilt groups found; raw tilt values are displayed as blue dots, centroids of each group as red dots, and grouping is visualized by the map colors.

Figure 10. *Annual microscope checkout results:* (a) Isosurface representation of the apoferritin map using 325,636 particles from 564 movies taken on a TFS Titan Krios with Gatan GIF/K3 at 0.825Å/pixel and 54e^-/Å^2 total dose. (b) Representative 2D class averages; box size 27.4 x 27.4 nm. (c) An α-helical segment from one β subunit (PDB: 6V21) is shown in heteroatom representation docked into the corresponding region of the reconstruction. (d) FSC Plot. (e) Guinier Plot.







a

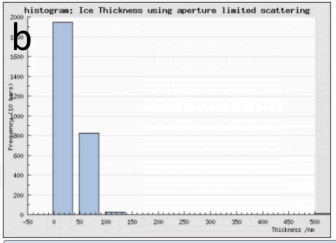
Imaging Summary

Preset label	Mag (X)	Dose (e⁻/Å²)	Pixel size (Å)	Dimensions	Binning	Image count
gr	1550		4,160.0	1024x1024	4	23
sq	2250	0.011	158.4	1440x1023	8	259
hln	3600	0.000	99.0	1440x1023	8	317
enn	81000	65.978	1.083	5760x4092	2	2804
enn-a	81000	65.978	1.083	5760x4092	2	2803
enn-a-DW	81000	65.978	1.083	5760x4092	2	2803

Total images: 9009 **Duration:** 1 days, 03 hrs, 37 min

b

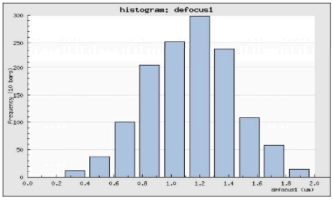
Ice thickness summary



c

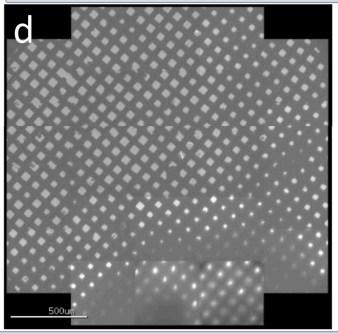
Defocus summary

	nb	min	max	avg	stddev
defocus1	3462	0.17 µm	3.91 µm	1.19 µm	0.37 µm
defocus2	3462	0.21 µm	4.10 µm	1.29 µm	0.38 µm
angle_astigmatism	3462	-89.946	89.888	46.057	43.234
extra_phase_shift	3462	0	0	0	0
resolution_80_percent	3462	2.949 Å	100.000 Å	5.740 Å	8.884 Å
resolution_50_percent	3462	2.697 Å	100.000 Å	4.525 Å	4.286 Å
package resolution	3462	2.704 Å	1433.230 Å	25.645 Å	173.824 Å



d

Atlas



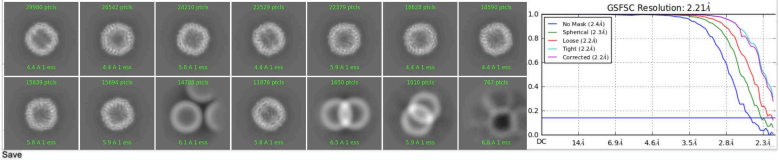
e

Experimental setup

TEM:	EF-Krios
Magnification:	81000
High tension (kV):	300
Defocus range (μm):	0.5 - 2.1
Spot size:	6
Intensity:	1.28920718487037e-06
Cs (mm):	2.700
Energy filtered:	yes
Energy filter width:	20

Camera:	GatanK3
Dimension:	5760 x 4092
Binning:	2 x 2
Exposure time (ms):	2500
Pixel size (Å):	1.0825
Dose rate (e⁻/Å²/s):	26.39
Total dose (e⁻/Å²):	65.98
Frame rate (ms):	50.00
Total frames:	50

g



f

Experimental Methods

EF-Krios operated at 300 kV with a GatanK3 imaging system collected at 81,000X nominal magnification. The calibrated pixel size of 0.5413 Å was used for processing. Movies were collected using Leginon (Suloway et al., 2005) at a dose rate of 26.39 $e^-/Å^2/s$ with a total exposure of 2.50 seconds, for an accumulated dose of 65.98 $e^-/Å^2$. Intermediate frames were recorded every 0.05 seconds for a total of 50 frames per micrograph. A total of 2804 images were collected at a nominal defocus range of 0.5 – 2.1 µm.

Atlas Collection

Targeting Crystals at Low Magnification

Test Diffraction with Preview Mode

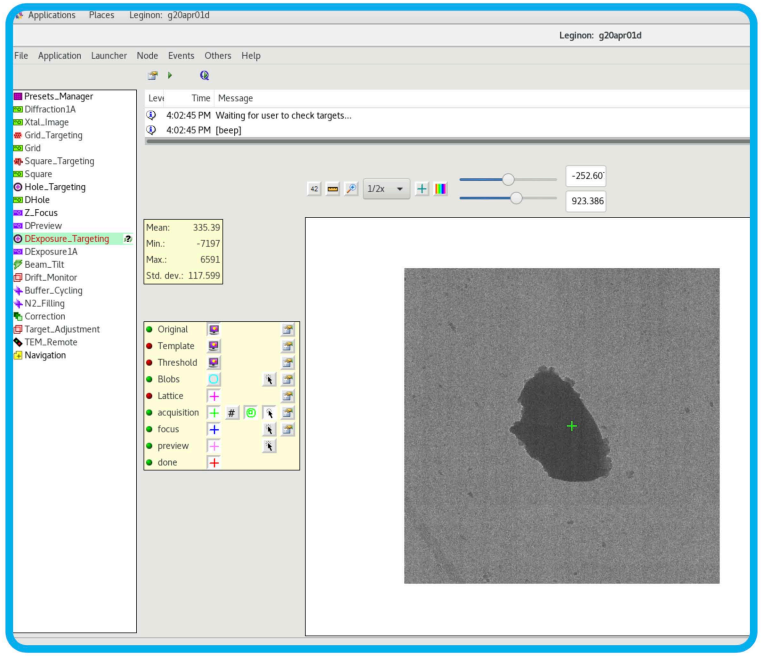
Queue Crystals for Data Collection

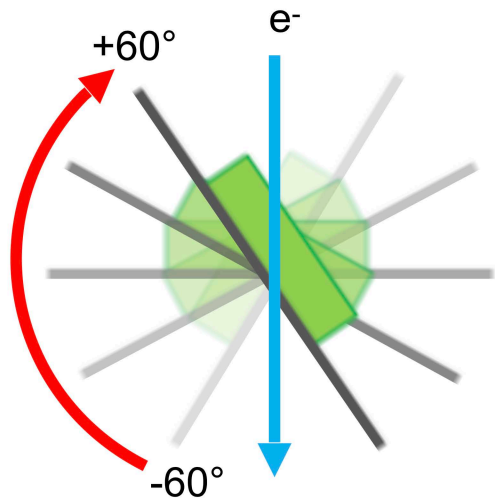
Move to Crystal and Eucentric Focus

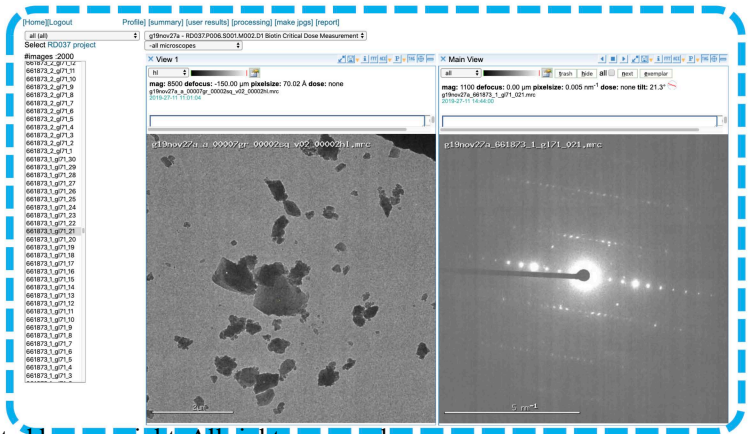
Continuous Stage Tilt Synchronized with Data Collection

Image Saving and File Type Conversion

Monitor Collection in Web Viewer



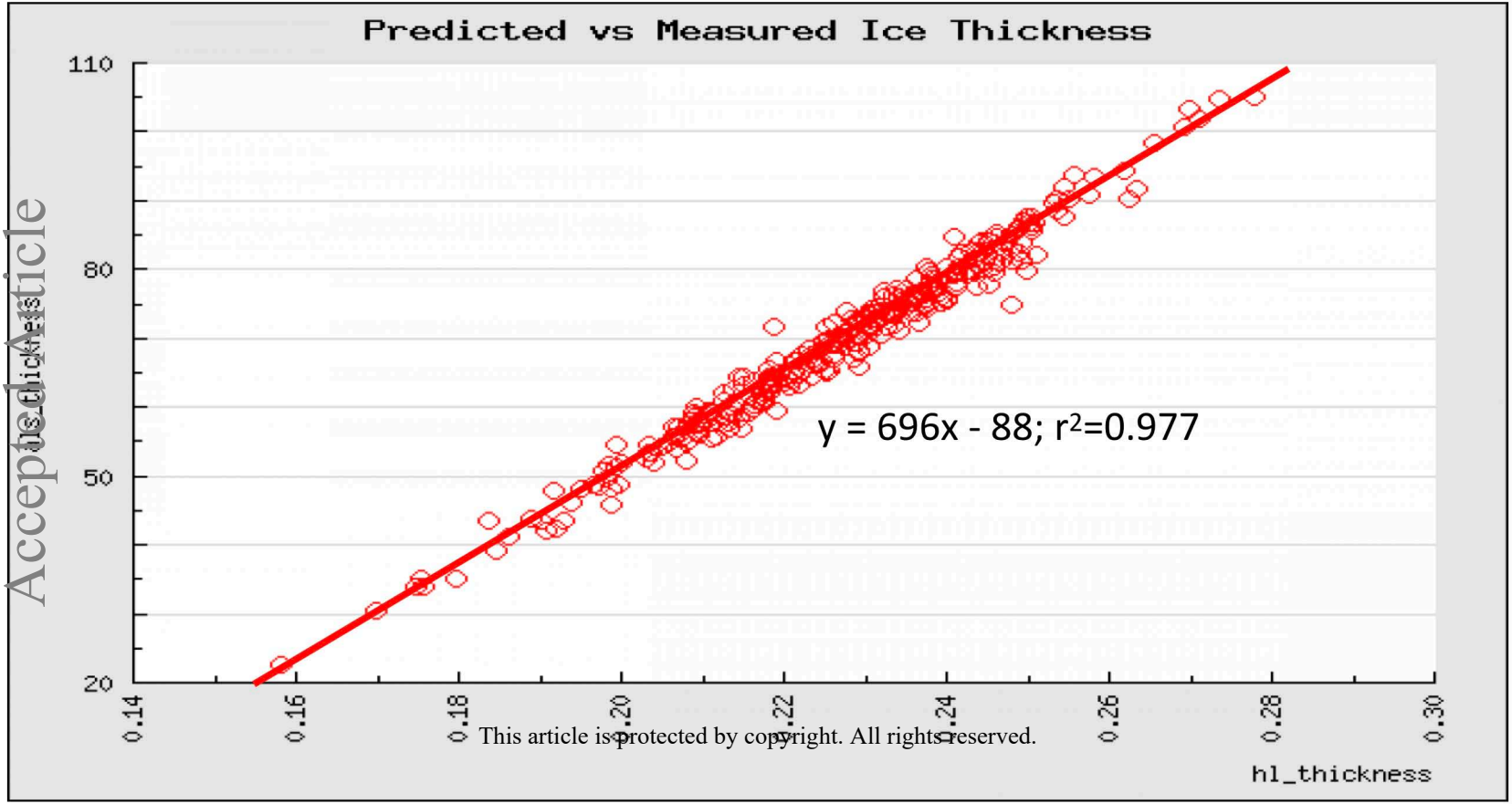




Fully Automated

User Guided

Optional



Leginon: 20jul08a

a

- Presets_Manager
- Beam_Tilt_Image
- Grid_Targeting
- Grid
- Square_Targeting
- Square
- Hole_Targeting
- Hole
- Exposure_Targeting
- Exposure
- Exposure_FFT
- Preview
- Focus
- Z_Focus
- Beam_Tilt
- Drift_Monitor
- Instrument
- Buffer_Cycling
- N2_Filling
- Correction
- Target_Adjustment
- TEM_Remote
- IceT
- Navigation

b

c

Level	Time	Message
	4:01:09 PM	Waiting for targets from remote 25
	4:01:08 PM	[beep]
	4:01:08 PM	[beep]
	4:00:54 PM	Publishing targets

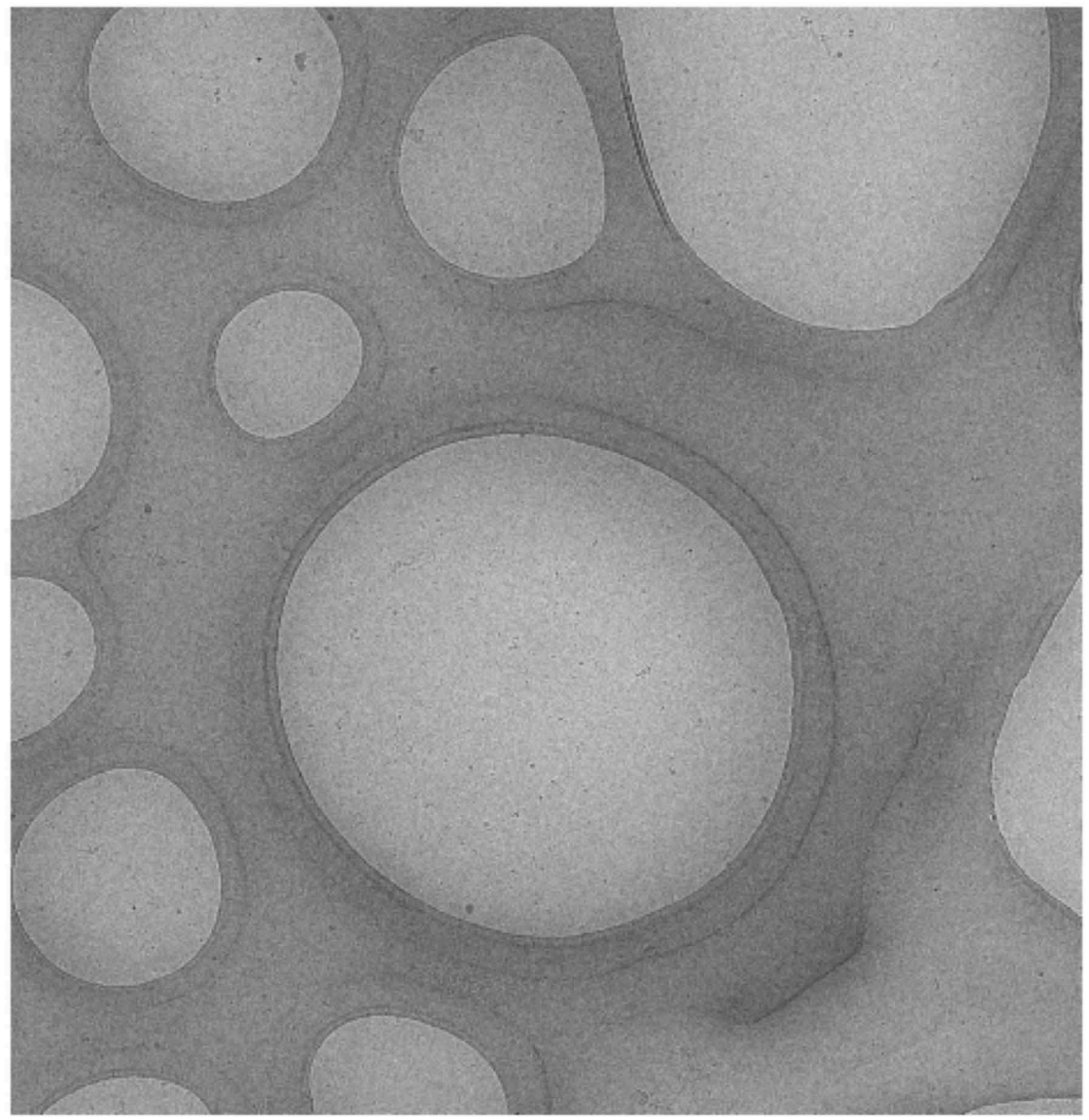
d

Mean:	378.251
Min.:	0
Max.:	765
Std. dev.:	79.5851

e

- Original
- Template
- Threshold
- Blobs
- Lattice
- acquisition
- focus
- preview
- done

42 1/2x 0 765



Application MSI-T started

Main menu

Square_Targeting 1

Hole_Targeting 0

Exposure_Targeting 1

b

Leginon Status

need_remote_input@Square_Targ

need_remote_input@Exposure_Ta

c

Leginon Tools

Square_Targeting

Exposure_Targeting

1

TEM_Remote

d

Time

Session start:

07/22/2020 12:07 AM UTC

Session stop:

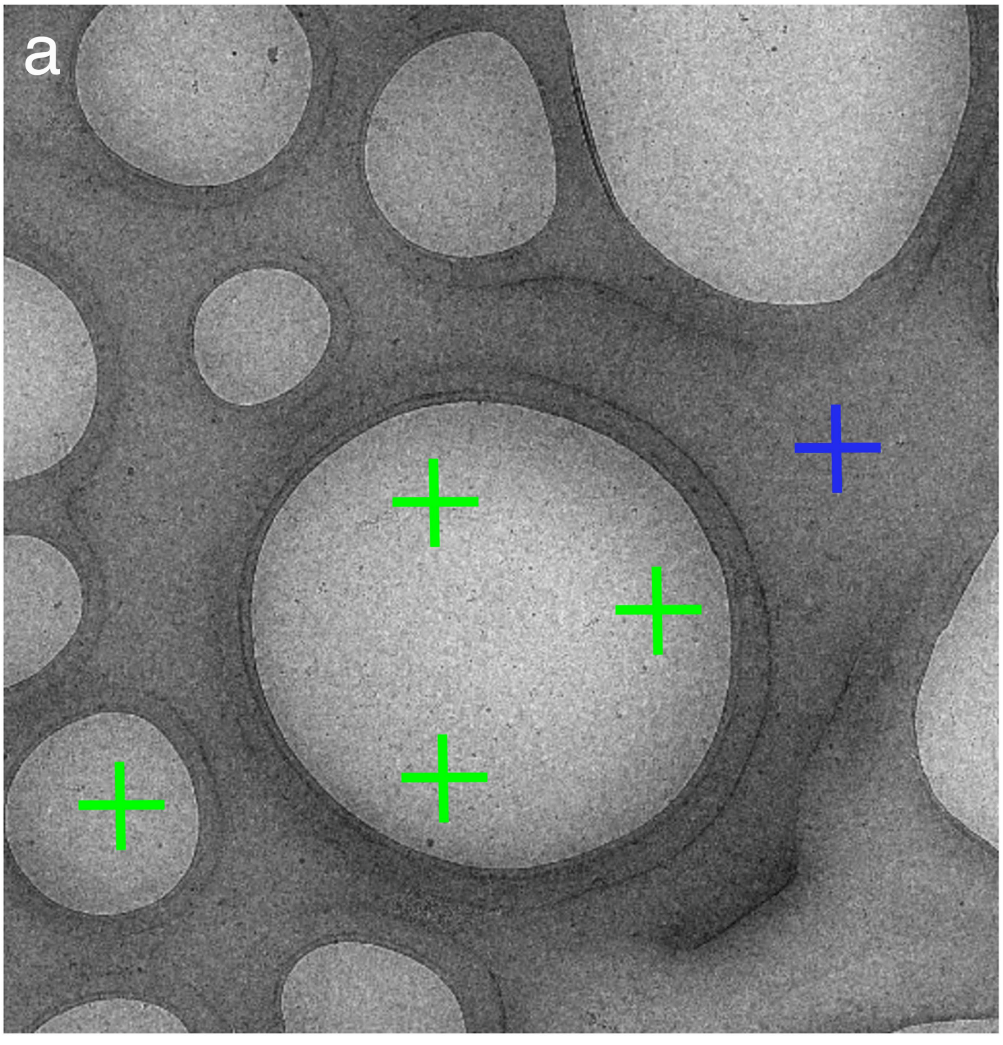
07/24/2020 12:07 AM UTC

Session progress

36.5%

Leginon - Exposure_Targeting

a



acquisition target mode is currently active.

acquisition target

focus target

reference target

Zoom Reset

Clear Targets

Toggle Crosshairs

Submit

Chat with Staff

e

Test User: Hello, I need some help

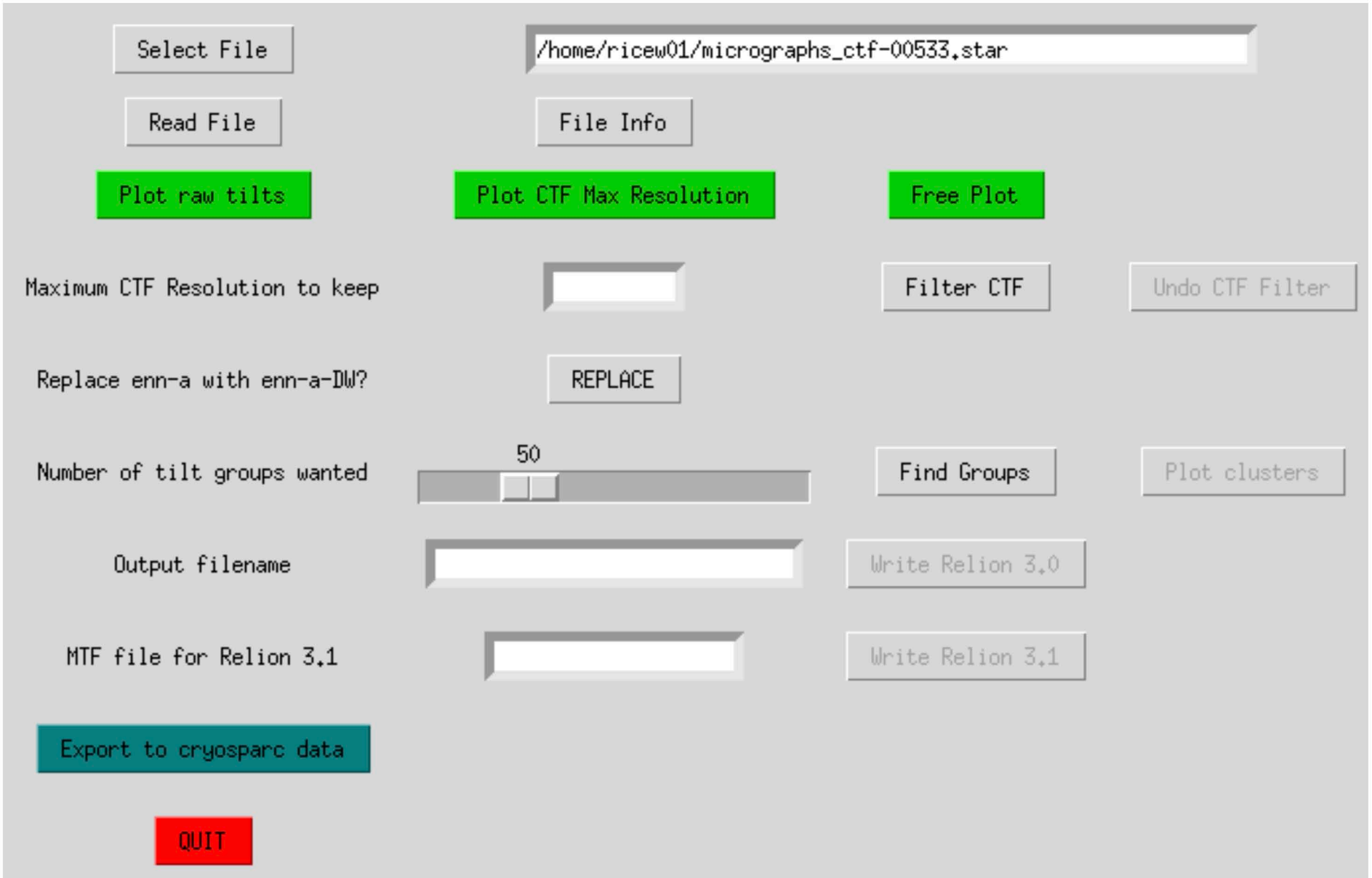
10:47:20 | July 22

Staff: What can we do for you ?

10:47:40 | July 22

Type a message

a



b