

## Automated molecular microscopy: The new Leginon system

Christian Suloway, James Pulokas, Denis Fellmann, Anchi Cheng, Francisco Guerra, Joel Quispe, Scott Stagg, Clinton S. Potter, Bridget Carragher\*

*The Scripps Research Institute, La Jolla, CA 92037, USA*

Received 14 December 2004, and in revised form 7 March 2005  
Available online 19 April 2005

### Abstract

We report here on the current state of our efforts in automated molecular microscopy. Our primary automated data acquisition software system, Leginon, has been completely redesigned over the past two years. The new distributed system has been developed using the Python programming language and is compatible with both Linux and Windows operating systems. The new flexible architecture was designed to allow for the development of customized data collection protocols, several of which are described here. The system has been used to acquire data for ~150 experiments and we have demonstrated the capacity for high throughput data acquisition by acquiring images of more than 100 000 particles in a single session at the microscope.  
© 2005 Elsevier Inc. All rights reserved.

*Keywords:* TEM; Cryo-electron microscopy; Automation; Pipeline

### 1. Introduction

Consider the following definition: “Automation [...] is the use of computers to control [...] machinery and processes, replacing human operators. It is a step beyond mechanization, where human operators are provided with machinery to help them in their jobs. [...] Some advantages are repeatability, tighter quality control, waste reduction, integration with [...] systems, increased productivity and reduction of labour. Some disadvantages are high initial costs and increased dependence on maintenance.<sup>1</sup>” This definition is intended as a description of industrial automation but it also provides a useful description for the efforts towards automation for macromolecular microscopy that have been ongoing over the past decade. One of the most successful results of these efforts has been the introduction of systems for automated acquisition of tomographic tilt series (Dierksen et al., 1992; Koster et al., 1992). About a dozen

automated tomography packages are now available from either commercial vendors or the academic community (Mastrorade, 1997; Zheng et al., 2004; Ziese et al., 2003) and very few microscopists would consider collecting a tomographic series without the aid of one of these systems.

Our own efforts over the past six years have been focused on developing an automated system capable of acquiring large numbers of high quality images of macromolecules under low dose conditions. Our initial goals when we began these developments were focused on providing a mechanized system (in the sense of the definition above) in that we had in mind that a computer controlled microscope might be of some help during the exacting and sometimes time-consuming task of acquiring images from a specimen embedded in vitreous ice (Kisseberth et al., 1997). More recently we have been focused on developing a truly automated system, one capable of operating without intervention by a human operator, with the goal of improving the throughput for macromolecular microscopy. An automated system has the potential to have an impact on throughput at various steps in the process of determining structures

\* Corresponding author. Fax: +1 858 784 9090.  
E-mail address: [bcarr@scripps.edu](mailto:bcarr@scripps.edu) (B. Carragher).

<sup>1</sup> <http://www.wordiq.com/definition/Automation>.

of macromolecules using EM. The most obvious applications are the need to image very large numbers of particles in a drive to improve structural resolution using single particle methods, and the need to reconstruct several variants of the same structure to shed light on the functional states of a macromolecular machine. En route to either of these goals, however, are often a long series of screening experiments required to identify appropriate preparation conditions for the specimen. Automated data acquisition can be integrated with robotic specimen handling systems to address this task (Potter et al., 2004). A related problem is the requirement to search for specimens which are extremely sparsely distributed on a grid, for example well preserved 2D crystals. Finding these “rare events” can be an extremely tedious task for a microscopist whereas automation software can be used either to directly find the feature of interest or to acquire and store images across the entire region of interest which can later be screened rapidly by eye. We have also for some time had as a goal to use our automated system to help define data collection protocols that result in data of the highest quality. We will address this particular goal in more detail in a later paper describing the collection of a dataset of a large number of single particles. An ancillary benefit of developing well defined and repeatable protocols for EM data collection and codifying these into automated algorithms is that it opens up the method beyond a limited group of practitioners who might require months or even years of experience to become experts.

The advantages of automation that we list here are quite similar to those mentioned in the definition of industrial automation with which we began. Unfortunately, some of the disadvantages also hold true; developing systems to replace human operators has high initial costs during the development effort and places increased dependence on the stability and reproducibility of the instruments being used. We will address these issues further below.

### 1.1. The Leginon system for automated data acquisition

Leginon was the first system to be developed that integrated control of the instrument and image acquisition with machine vision algorithms designed to replace an operator in the task of finding potential targets on a grid at a variety of scales (Carragher et al., 2000; Potter et al., 1999). The multi-scale imaging strategy that was developed started with the acquisition of an atlas of the entire grid built by montaging together images acquired at a magnification of  $\sim 50\times$ . This atlas is then automatically analyzed to determine the locations of grid squares and these are further assessed to select potential areas of interest based on simple image metrics like mean and variance. For each selected square, an image is then acquired at a magnification  $\sim 500\times$  and ana-

lyzed to identify further targets. For example, when examining specimens preserved in vitreous ice and supported using a holey carbon substrate, the targeting task is to search for the location of the holes and estimate the thickness of the vitreous ice across the hole. For each hole in a selected ice thickness range, the magnification can then be increased again ( $\sim 5000\times$ ) and the image analyzed to determine the final high magnification target positions. This analysis is dependent on the nature and distribution of the macromolecular specimen. For isolated single particles that are uniformly well distributed across a hole, the center of the hole can be identified and then high magnification image targets can be selected anywhere within the hole. Alternate targeting strategies were also developed for the system including the capability to find filaments and identify long straight areas along the filament axis and the ability to identify well defined 2D crystals (Potter et al., 1999). Once these final target locations have been specified, low dose drift measurement and focus routines are performed automatically and high magnification images or defocus pairs are collected at each target area. The only operator input required by the system is to setup the parameters required to define the targeting algorithms and to refill cryogens during data acquisition.

During any data collection session controlled by the automated system, all images, data acquisition parameters and user entered events are stored to a database that is an integral part of the Leginon software. A variety of web based user interfaces to the database were developed that provided the means to view and browse the images and keep track of the relationships between them (Fellmann et al., 2002).

The Leginon system was originally developed for the Phillips CM series microscopes but has since been adapted to run on FEI Tecnai instruments, supporting both Tietz and Gatan CCD cameras, and more recently to the JEOL 3100 (Glaeser, 2004a). The system as described above generally results in images of a quality equivalent to those acquired by a microscopist (Carragher et al., 2000; Carragher et al., 2004) and it has been used to image a variety of macromolecular specimens, including single particles (Mouche et al., 2003; Rouiller et al., 2001; Shih et al., 2004), helical filaments (Zhu et al., 2001) and 2D crystals (Cheng et al., 2003b; Potter et al., 1999).

Several other groups have also developed packages for mechanized or automated data collection for molecular microscopy. These include the Grace package developed by Brisson's group and designed for semi-automated data collection for electron crystallography (Oostergetel et al., 1998). The Grace package uses the Gatan scripting language for microscope control and image acquisition and runs on the Philips CM series microscopes. The AutoEM package, from the Subramaniam group at NIH, also uses the Gatan scripting

environment and runs on FEI Tecnai microscopes interfaced to Gatan CCD cameras. The system supports both semi-automated and automated data collection for single particle data collection (Zhang et al., 2001). For targeting, AutoEM relies principally on calibrations of the grid geometry which then allow it to visit every hole in a grid square systematically. The Frank group has recently modified and extended this software, now called AutoEMation (Lei and Frank, 2005), adding in an algorithm to identify holes in each grid square in a manner essentially identical to that described above for Leginon. On JEOL microscopes, the Chiu group has developed the JAMES package which provides for real-time data preprocessing and feedback on the quality of images acquired during a session. The JAMES system was used in collecting images used to reconstruct a virus to a resolution of 9 Å (Booth et al., 2004).

### 1.2. The new Leginon system

The original Leginon system was prototype software used to develop the methodology for automated molecular microscopy and demonstrate its potential. The software was written using a scripting language (tcl/tk) (Ousterhout, 1994) in combination with a tcl based image processing language (Potter and Moran, 1992) for application development. Using this software we demonstrated the feasibility of automated molecular microscopy and illustrated its potential in our laboratory for over four years. The prototype Leginon system did, however, have several limitations. Although the system worked rather well for the primary purpose for which it was designed (acquiring data from vitreous ice specimens supported on Quantifoil grids) it was not easy to adapt it to other data collection protocols which we were interested in developing. For example, targeting specimens prepared using alternative grid geometries, undertaking large screening studies, or combining the existing targeting algorithms with tomographic data collection. Having demonstrated the principals of automated molecular microscopy we wanted to generalize the Leginon system and extend it to other types of specimens and data collection protocols. At the same time we needed to optimize the system to achieve higher data throughput rates and improve the user interface so that we could make the software available to others outside our own lab. We found that, like many prototyping software projects, the software application had grown to the point that our ability to maintain it and continue adding features became very limited. We had learned many lessons from the original Leginon development effort and decided that a complete redesign and rewrite of the software was necessary to provide a system that could be readily supported, maintained and, most importantly, be extendable to a wide range of data collection applications.

The redesign of Leginon also coincided with a major redesign of the database to which the software is intimately linked. This was required to provide an automatically adaptive database structure that supports the flexibility in the new Leginon design. The original image database tracked all acquired images, the relationships between these images and a specified subset of data collection parameters (Fellmann et al., 2002). The database design used a straightforward system in which tables and fields are fixed and defined. Thus, whenever a new feature was developed which resulted in new parameters needing to be stored, the database table structure had to be manually modified to remain compatible with the software. This became particularly difficult when the relationship between images at multiple scales was modified or extended. The new Leginon database now uses an object oriented approach in which tables, fields, and data structures are updated dynamically. New applications and algorithms in Leginon can now be integrated into the database without requiring any manual updating of the structure of the database. This has had a major and critical impact on the ease with which we can extend and modify the automated system. The new database has also been extended to include storage of all information relating to the automated data collection. This includes a full set of parameters associated with every image and the results of the processing and analysis algorithms. In addition, the database is used to store information about the collection protocol and setup parameters for every application, including the overall design of the application itself.

In this paper, we will provide a detailed description of the architecture of the new system (which will from now on be referred to only as Leginon) and the design of some of the applications. We will also present some results of using the system over the past year and the current status of the software distribution.

## 2. Design and architecture of the Leginon system

A microscopist readily adapts to changes in data collection protocols and the particular sequence of magnification steps (e.g., search, focus, and expose) that he/she uses will vary depending on the application and to some extent the quality of the specimen. For example, screening grids of specimens preserved in negative stain for appropriate crystallization conditions requires a different protocol for data collection than that used to acquire low dose images of vitreous ice embedded single particles. Developing software that has the same flexibility and capabilities as a human operator is a major challenge. However, although the list of potential data collection protocols is very large they almost all have in common several key tasks. These include basic microscope control and image acquisition, an assessment of

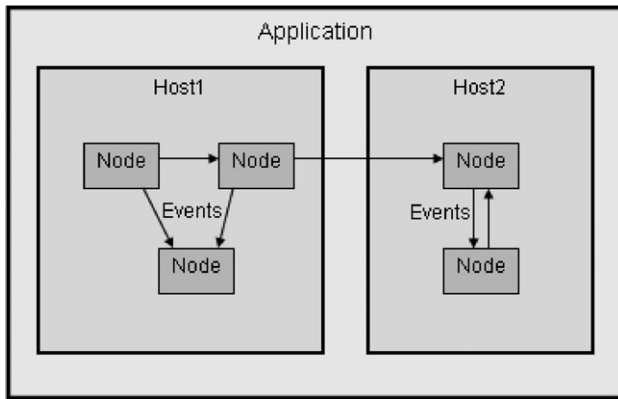


Fig. 1. Schematic diagram of a Leginon application. Common tasks are packaged into *nodes*, which can be connected into a self sustaining network called an *application*. Nodes communicate via *events* which can be used to signal other nodes to begin or resume performing an action. Nodes can be distributed across multiple computers on a network while still belonging to a single global application.

the acquired image to determine target selection and location, management of the targets to ensure low dose imaging conditions, checking and setting the defocus and astigmatism parameters, and checking and monitoring any drift in the specimen stage. The Leginon system packages up tasks like these into reusable modules, called *nodes*. While each node may be used individually and independently, the flexibility and power of the system arises from the ability to connect individual nodes into a self sustaining network called an *application* (see Fig. 1). The nodes in an application communicate with each other to determine which ones should currently be running tasks. A node communicates to another node by generating an *event*. The event may be used to signal other nodes to begin or resume performing an action. This event driven architecture provides a very flexible environment for application development. An overview of the node and event classes currently available in Leginon is outlined in Table 1. The architecture of the Leginon system is designed to allow nodes to be distributed across multiple computers on a network even though each still belongs to the same global running application (see Fig. 2). For example, the node that is acquiring an image can be running on a different processor than a node whose task is to analyze that image to identify target positions.

Leginon is controlled with a graphical user interface developed using the wxPython package.<sup>2</sup> The user interface provides facilities to manage the systems of nodes, as well as interact with the nodes themselves. Nodes can be created, bound with events, and removed using the interface. Likewise, applications can be started, edited, and stopped. Nodes have individual interfaces

embedded in the overall interface. Within each node's user interface, data is displayed, settings can be modified, and commands can be given (see Fig. 3 for an example). A logging facility allows users to monitor Leginon's activity through status icons and messages. Settings which are reused have their values stored in the database so they are remembered the next time Leginon is started. The graphical user interface allows simple and convenient interactivity when using Leginon.

A relational database manages all persistent data associated with Leginon. This includes (i) the design of the application in terms of the nodes which are used and the connections between them; (ii) all of the user preferences including microscope and camera settings, targeting preferences, focusing parameters, etc.; (iii) all data acquired during a session, including all images, data acquired from peripheral devices (e.g., temperature monitors), user annotations and notes, etc. Whenever a node produces data that is useful to the user or to other nodes, it is required to "publish" that data. This registers the data in the database and, if necessary, notifies other dependent nodes about the published data through the generation of an event. Once the data has been stored in the database it can be viewed using a variety of web based user interfaces which have been specifically developed for this purpose. This web based viewing system is particularly convenient in allowing off-site collaborators to monitor an experiment as well as transfer data and images to local systems.

The architecture of the connection between the database and the Leginon system is represented schematically in Fig. 4. A single python module, *data.py*, contains the definitions of all the data structures that need to be stored in the database. Each data structure is similar to a python dictionary type having key-value pairs. The *DBDatakeeper* module provides the high level interface through which Leginon nodes may call insert and query functions on these data structures. The *SQLDict* module performs the conversions between python dictionaries and records in the MySQL database. *SQLExpr* generates the correct SQL syntax for inserting data into tables and fields or modifying the database. *SQLDict* preserves relationships between different python structures by creating references between different database tables. The *SQLDict* module seamlessly updates the database using the *MySQLdb* python extension which maintains a persistent connection to the database. Using the *SQLDict* and *SQLExpr* modules, database table and field management is transparent to the python developer who is creating new nodes. For example, if a new data structure is needed in Leginon, it is simply added to the definitions in the *data.py* module. The first time the node makes an insert call, the table will be created and the data properly formatted and inserted. If the python structure is changed at some later stage, *SQLDict* will update the table automatically.

<sup>2</sup> <http://www.wxpython.org>.



Table 1  
Examples of Leginon node and event classes

	Description
<i>Node class</i>	
Acquisition	Acquires image(s) at each target it is given
Corrector	Performs flat field correction, spike removal on raw images
DriftManager	Monitors drift and adjusts targets following a drift
EM	Interface to the TEM and camera
FFTMaker	Calculates a power spectrum from an image
Focuser	Performs autofocusing at each target it is given
HoleFinder	Identifies holes in images and produces a set of targets
MosaicClickTargetFinder	Compiles a set of images into a mosaic (atlas of the grid)
MosaicTargetMaker	Generates a set of targets capable of producing a mosaic
Navigator	Allows movement (image shift or stage position) by clicking on an image
PresetsManager	Maintains and allows editing of all the presets for a session
<i>Event class</i>	
AcquisitionImagePublishEvent	An image has been acquired and stored in the database
ChangePresetEvent	Tell Presets Manager to change presets
DriftDeclaredEvent	Tell Drift Manager that drift has happened (e.g., after Z adjust)
DriftDetectedEvent	Tell Drift Manager that drift is in progress and should be monitored
DriftDoneEvent	Drift manager is done monitoring drift, because it has stopped
DriftWatchEvent	Tell Drift Manager to adjust some targets if necessary
ImageProcessDoneEvent	Image has been processed (e.g., HoleFinder done with square image)
ImageTargetListPublishEvent	Targets have been found and published in the database
PresetChangedEvent	Presets Manager has just changed to another preset
SetCameraEvent	Configure the camera
SetScopeEvent	Set parameters on the TEM
TargetListDoneEvent	Acquisition node is done acquiring a list of targets

Both SQLDict and SQLExpr are new tools that we have developed specifically for the purpose of transparently using a relational database from the Python programming environment.

The database is also used to manage the overall environment in which the data collection takes place. Thus, the database keeps track of the users of the system and the groups to which they belong, and the instruments and cameras available. This information is entered into the database using a web based user interface and is subsequently required by the Leginon system prior to beginning a data collection session.

An application is designed by selecting the nodes required for the application, deciding in which order they are to be executed and how they will communicate with each other. A graphical application editor is available as an aid in building new applications. Once an application has been built it is stored by the Leginon database and may be reused at any time. One of the primary applications will be described in detail below. This application, called MSI (MultiScale Imaging), is closely based on the prototype Leginon acquisition protocol which we have described previously (Carragher et al., 2000). Leginon applications can be exported as XML files which can then subsequently be imported into another Leginon database. In this way applications can be shared between users in different groups or at different institutions by simply making the XML file available, from a central web site for example. The XML file can also be directly edited and this is an alternative method for creating or

modifying an application. The size of the XML text file describing the MSI application is ~32 kb.

### 2.1. Hardware and software dependencies of the Leginon system

The overall architecture for Leginon is illustrated in Fig. 2. In the typical configuration used in our laboratory a client process controlling the microscope and camera runs on the local Windows based PC which is directly connected to the instrument. This process usually only includes a single node, to control the instrument, to conserve processor usage. The main process normally runs on a Linux or PC system located in a different room and connected to the microscope over a gigabit network. The database and webserver run on a separate Linux computer located in our server room.

Leginon has been implemented primarily in the Python programming language.<sup>3</sup> This language allows for cross platform compatibility and means that nodes can be distributed across computers of different types running different operating systems. Leginon makes use of the numarray python package<sup>4</sup> for image processing and the wxPython package for supporting the user interface. The Leginon database is implemented in MySQL<sup>5</sup>

<sup>3</sup> <http://www.python.org/>.

<sup>4</sup> [http://www.stsci.edu/resources/software\\_hardware/numarray](http://www.stsci.edu/resources/software_hardware/numarray).

<sup>5</sup> <http://www.mysql.com/>.

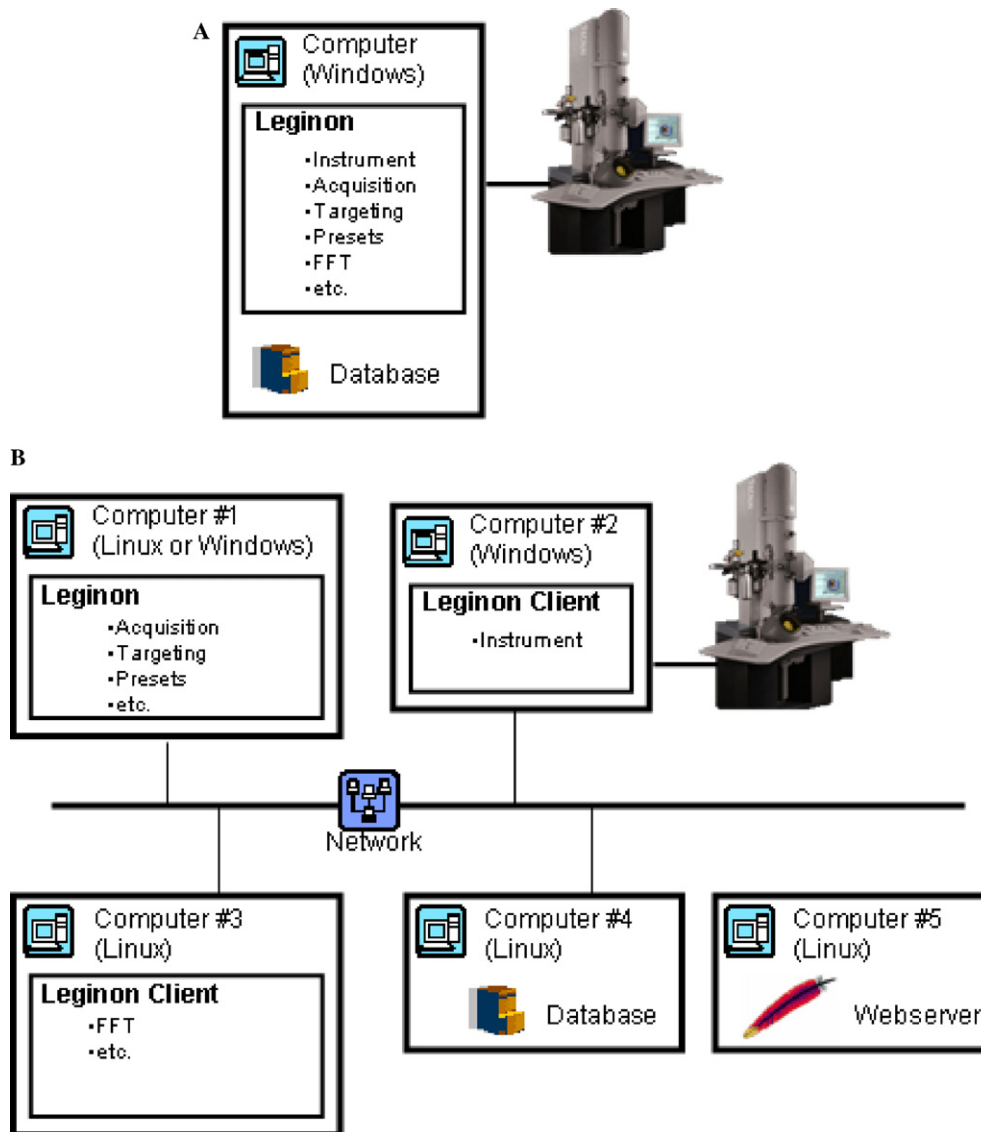


Fig. 2. Leginon may be configured to (A) run entirely on a single computer connected directly to the microscope or (B) be distributed across multiple computers.

and the web based graphical user interfaces are implemented using PHP<sup>6</sup> scripts. Leginon uses the Python extension packages, MySQLdb and XML, to support communication with the database.

Access to the microscope and CCD cameras is provided by the pyScope Python extension which has been developed and is maintained by our own group. The pyScope extension is a thin wrapper around various Application Programming Interfaces (API's) distributed by microscope and CCD camera manufacturer's. The pyScope extension is designed to provide a universal API for controlling different instruments. Using this extension, Leginon applications can control either a

Tecnai or a JEOL microscope and acquire images from both Gatan and Tietz CCD cameras. Other microscopes and CCD cameras can be added as needed. The effort involved in adding new instruments is not substantial but does depend on the specific instrument API.

## 2.2. Microscope calibrations and setup

Operating the microscope automatically and over a lengthy period of time while unattended by a human operator requires that the behavior of the instrument be predictable and stable. This behavior depends on a number of calibrations that must be accurately determined as well as the necessity for setting up relationships between magnification scales that are stable over time. During a long data acquisition session requiring multi-

<sup>6</sup> <http://www.php.net/>.

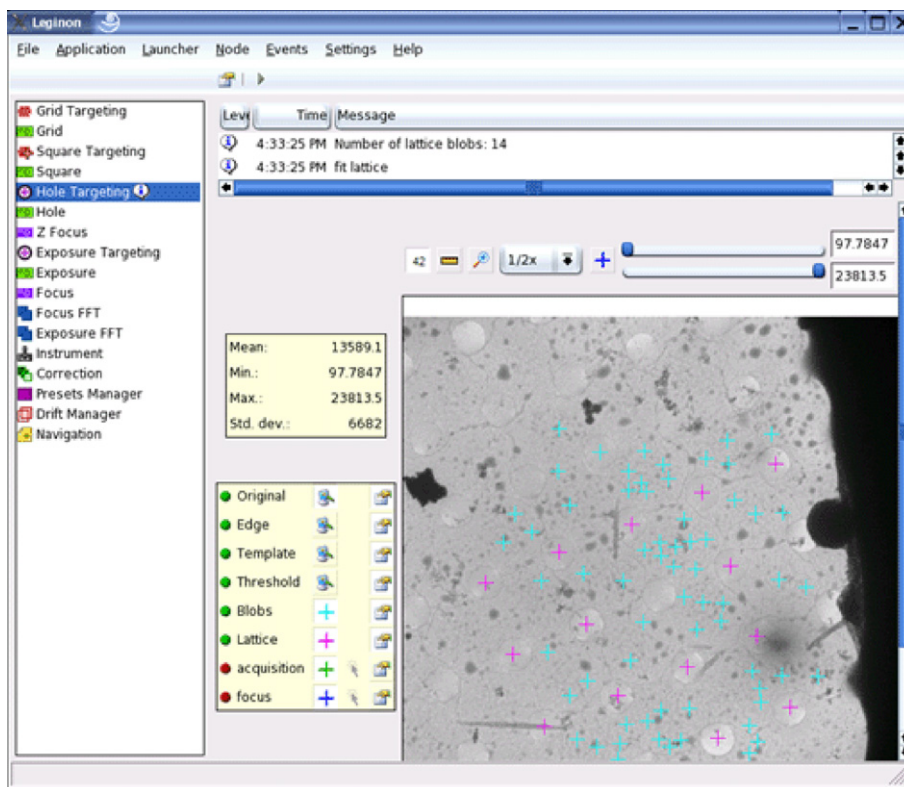


Fig. 3. A screen shot of the graphical user interface used to control Leginon. Each node has its own customized toolbar, control panel, and a message box which is used to display information regarding the status of the node as well as any warnings or errors. Icons are used to indicate the active status of a node; several of which may be simultaneously active.

scale targeting, the magnification of the instrument will be changed through a range of two orders of magnitude thousands of times over the course of the experiment and during this time the image and the beam must remain centered on the selected target at every magnification scale. The stability and predictable behavior of the system is dependent on the accuracy of the calibrations and preset alignments setup between magnification scales. Below we discuss a number of these calibration and setup procedures in more detail. We have also for the sake of completeness included a discussion of some general calibrations (e.g., pixel size) that, while not crucial to the stable behavior of the system, are convenient for later analysis. A Leginon application, called *Calibration*, provides a set of nodes designed to manage, and automate where possible, the measurement of the calibrations and the alignment of the images between magnification scales.

The *pixel size* at the specimen can be calibrated against the nominal microscope magnifications using the known crystalline spacings of Catalase crystals prepared using a protocol<sup>7</sup> modified from (Sumner and Dounce, 1955). CCD images of the crystals are acquired, power spectra computed, and the spacing of

the 2nd order diffraction spots measured. An average of at least three measurements is used to determine pixel size for magnification settings in the range 29k $\times$ –100k $\times$ . Pixel sizes for lower magnifications are estimated by scaling down from the accurately calibrated high magnification values. The pixel size calibration is not frequently re-measured as accuracy in this measurement is not critical to any of the other calibrations or the stable behavior of the microscope. Whenever accurate measures of the pixel size are required for a particular experiment, these are usually measured during data acquisition using an internal calibration standard like TMV. The calibrated pixel size values are very convenient to have generally available when displaying the images or at the beginning of analysis and reconstruction.

*Electron dose as a function of beam intensity* is calibrated by measuring the intensity on the CCD resulting from a measured beam current. The dose rate, measured at the specimen, is calculated from the beam current using:

$$\text{dose rate (e}^{-}/\text{\AA}^2 \text{s)} = (K * M^2 * B)/A, \quad (1)$$

where  $K = 6.25 \times 10^{18}$  (e<sup>-</sup>/coulomb);  $M$  is the magnification;  $B$  is the beam current (amperes or coulombs/s);  $A$  is the collection area on the screen ( $\text{\AA}^2$ ). This dose rate

<sup>7</sup> <http://ami.scripps.edu/publications/techreports/98-009/>.

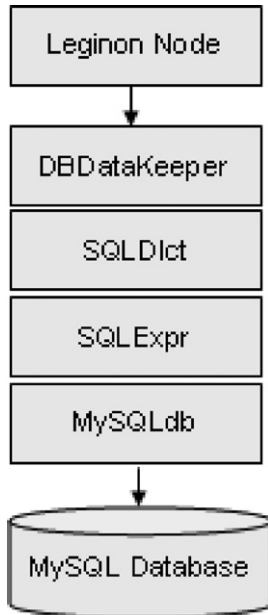


Fig. 4. A schematic representation of the connection between the Leginon system and the database. DBDataKeeper allows nodes to perform inserts and queries of Leginon data structures into the database. SQLDict performs the conversions between python dictionaries and records in the MySQL database and SQLExpr generates the correct SQL syntax for inserting data into tables and fields or modifying the database SQLDict updates the database using the MySQLdb python extension.

is then used to calibrate the sensitivity of the CCD in counts/electron by recording an image using the same beam configuration. The accuracy of the measured dose depends on the accuracy of the screen current measurement on the microscope. The value returned from the screen current measurement on our Tecnai instruments must be corrected to compensate for electron current lost due to charging of the phosphor screen. The correction factor can be determined using a Faraday cup or other independent device for accurately measuring electron current.

Calibrations which are critical to the Leginon system are the transformation matrices between the camera coordinate system and the positioning system used on the microscope. There are two basic methods available for repositioning the specimen relative to the imaging system: the goniometer and the image shift coils. The method chosen generally depends on the range of movement needed; we typically use the goniometer for movements  $> \sim 5 \mu\text{m}$  and the image shift coils for movements less than this.

The  $X$  and  $Y$  movements of the goniometer need to be precisely calibrated to accurately reposition targets to the center of the viewing area in a single movement. This requires that the non-linear movements of the Tecnai goniometer be modeled as described in (Pulokas et al., 2000). The model of the movement can be used to position the goniometer to within an accuracy of

$\sim 100 \text{ nm}$ , about a factor of 10 improvement over the accuracy attained assuming a linear behavior. The calibration is recorded by measuring the displacement of a specimen relative to known motions of the goniometer. These measurements are then used to calculate (i) a function (in the form of a harmonic series) that models the mechanical behavior of the stage, and (ii) a magnification adjustment (scaling and rotation) that allows the model function to be used at other magnifications. The harmonic series function needs to be determined at only one magnification, because the result will be normalized to use it at other magnifications. The magnification adjustment calibration must be determined for every magnification scale at which the goniometer will be used to reposition the specimen (typically the lower range of magnifications,  $500\times$ – $5000\times$ ). The required calibrations can be performed automatically using the goniometer modeling node in the Leginon Calibrations Application. The model function calibration requires  $\sim 50$  data points, ideally spread over at least two periods of the model function, to get a good fit and typically a fifth order harmonic series is used in the model. The magnification adjustment calibration requires only the constant terms of the model to be determined, as these are then used to scale the existing normalized model function from the base line calibration. The full calibration procedure with  $> 50$  measurements in each direction takes about 20 min to complete whereas the subsequent constant term calibrations requires only enough points to sample a single period and can be completed in under 5 min. Typical goniometer calibration curves are shown in Fig. 5 for the two Tecnai F20 instruments located in our laboratory. The goniometer calibration is generally stable over periods of several months and is only routinely checked or re-measured if the goniometer has undergone repair or replacement.

*Image shift calibrations* in the form of calibration matrices are required for every magnification that will be used in the Leginon application. These determine the image shifts which must be applied to move the specimen a specified distance measured on the CCD camera coordinate system. The procedure we use is based on that developed by Koster and de Ruijter (1992); three images are acquired at zero shift, and specified  $x$  and  $y$  shifts, and the resulting relative image movements are determined by phase correlation (Kuglin and Hines, 1975; Zhu et al., 2001). The transformation matrix is then calculated and stored into the database.

*Focus and astigmatism calibrations* are achieved using a very similar approach, again based on the method originally described by (Koster and de Ruijter, 1992). The response of the system to a specified set of focus and astigmatism adjustments is measured using phase correlations between the recorded images. Each of the matrix calibrations generally require less than 5 min to complete at each magnification. Although they are often stable



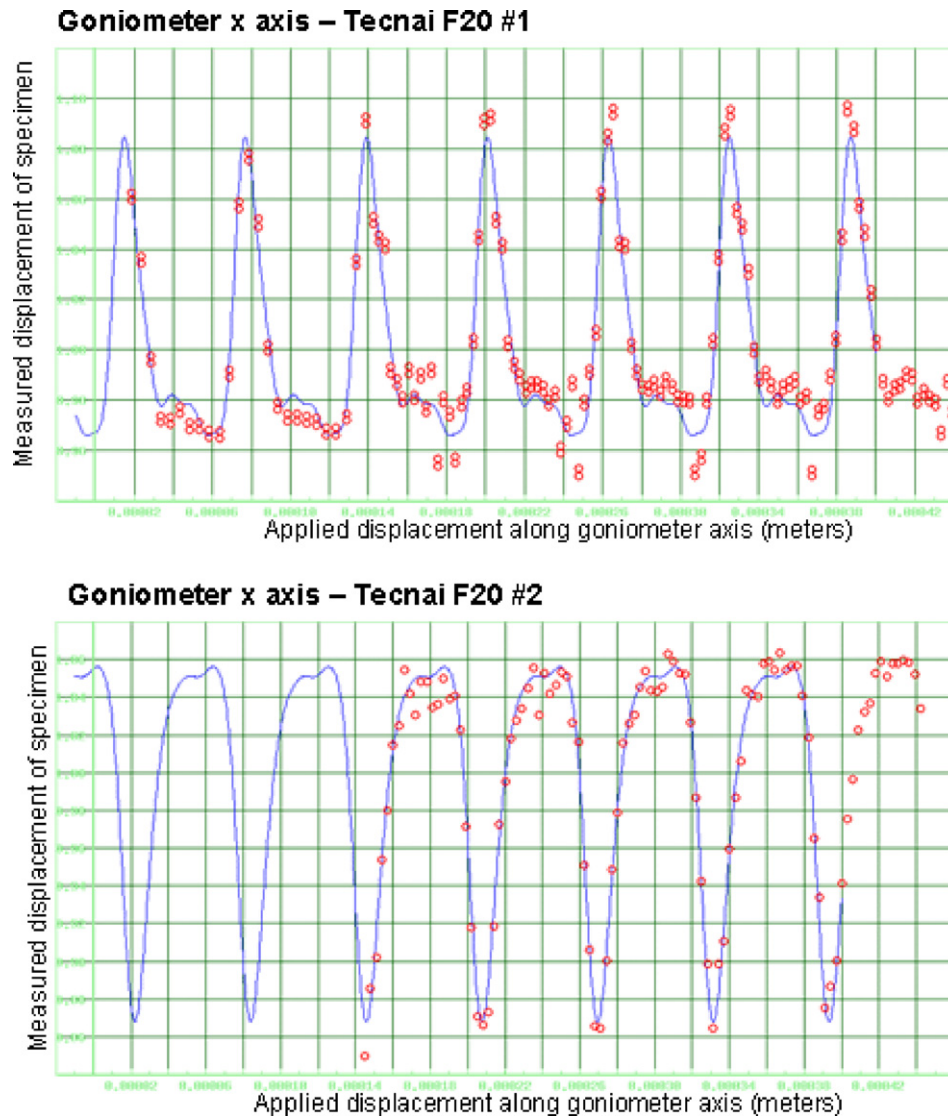


Fig. 5. Goniometer calibration curves for the  $x$ -axis of the goniometer for two separate Tecnai F20 instruments. The non-linear movements of the Tecnai goniometer are modeled by measuring the displacement of the specimen as a function of applied displacement along the goniometer axes. A harmonic series fit to these data points (solid curve) can then subsequently be used to calculate the goniometer displacement that must be applied to accurately relocate the specimen to a specified target.

over many weeks at a time we usually check and often re-measure these at all relevant magnifications before beginning any extended or critical data acquisition session.

The *Eucentric focus value* for the microscope is also recorded to the database in a routine that uses the auto-focus algorithm to measure the  $z$ -height of the specimen holder. While the *Eucentric focus value* changes with microscope alignment conditions, high accuracy for this measurement is not required by the software and thus the calibration only needs to be measured every few weeks. However, since it takes less than 1 min to complete, the standard practice in our laboratory is to re-measure it at the start of every experiment.

Creating *flat fields* to correct for irregularities of the CCD camera field of view, together with removing bias and dark current are part of the Calibration applica-

tion. In Legion, flat field images are obtained for each combination of camera settings (size, position, and binning) that will be required by the preset conditions. Flat field images are obtained by setting up flood beam conditions, removing the specimen from the field of view and acquiring an image using an exposure time that results in intensities approximately half the dynamic range of the CCD camera. Usually, three of these images are acquired and averaged together to form the final flat field image that contains the vignette of the optical system, any dirt on the surface of the CCD camera, and the variations in the sensitivity of the individual pixels. These calibrations usually take about 10–15 min to perform for a set of about 4–5 camera settings and they need to be repeated every few weeks.

The accuracy of the calibration measurements will depend on the state of the instrument. Thus, before beginning any calibration procedure we routinely check the alignments on the microscope, paying special attention to the rotation center and beam tilt pivot points, set the specimen to the eucentric height position, and adjust the objective lens to achieve Gaussian focus. These settings are checked at the beginning of any data collection session and tweaked during the session as a first response if the system appears to be behaving unpredictably.

A summary of the calibration measurement schedule that we normally follow in our lab is given in Table 2. Most of the calibrations are very stable and do not need to be performed on a daily basis. It is noticeable, however, that the calibrations will be much less stable if measured (or used) soon after the microscope has been in a low magnification mode, where the objective lens is essentially switched off. We believe that some of these instabilities arise from the fact that the column needs time to stabilize thermally when the objective lens is turned back on.

One further essential aspect in setting up a microscope for automated image acquisition and targeting is the need to establish a stable set of relative image and beam shifts that ensure that a target remains at the center of the CCD imaging area across the range of required magnification scales. The methods for setting up the required imaging conditions and the relationships between them are related to those used to setup an electron microscope for low dose imaging, e.g., in defining microscope settings for SEARCH, FOCUS, and EXPOSURE modes including the relative image and beam shift offsets between each of these conditions. In the case of a complex multistage imaging application like the one described below, at least five predefined imaging conditions must be setup and the image and beam shifts adjusted so as to keep a target centered when changing between them. The only real difficulty in setting up these imaging conditions is that hysteresis effects in the lenses mean that relative alignments are not normally achieved in one pass through the magnification scales. Instead, this is usually an iterative process that requires several cycles through all of the settings before a stable set of relative image shifts can be established. When several

magnification scales are required in an application this can be a time consuming task, particularly if the microscope is not perfectly well aligned or if the alignments have been substantially altered since the last time the settings were determined. Leginon provides several tools designed to assist in setting up the alignments between magnification scales. For example, using the *Navigator* node the user can identify a target in an image that should be centered in the field of view by a simple mouse click and the *Presets Manager* node can be requested to automatically cycle through the entire set of magnifications in a specified sequence to check for hysteresis. We are also considering developing a more automated approach to this problem as it remains one of the more tedious and time consuming aspects of preparing for automated data acquisition.

### 3. Leginon applications

#### 3.1. Leginon application example: multi-scale imaging

The Leginon MSI (multi-scale imaging) application is currently the workhorse application in our laboratory used for most routine data acquisition from either negative stain or vitreous ice specimens supported on either continuous or holey carbon. Below we will present a detailed description of a typical use of the MSI application: acquiring low dose images of vitreous ice specimens prepared over a Quantifoil holey carbon substrate. At the end of this section we will also briefly describe how the application can be modified for other data collection protocols.

The Leginon MSI application network is shown schematically in Fig. 6. The MSI application links together acquisition and target searching nodes at four magnification scales and is closely based on the prototype Leginon acquisition protocol which we have described previously (Carragher et al., 2000). The multi-scale imaging protocol requires the setup of six primary microscope imaging preset conditions. These are illustrated in Fig. 6 and include settings for acquiring images of an overall grid atlas, individual grid squares and holes

Table 2  
Schedule of calibrations

	No. of measurements required	Approximate time required per measurement (min)	Update frequency
Pixel size	1× for each magnification used by Leginon	10	Never
Modeled Stage Function	1× in mag range 300×–1500×	20	Every few weeks
Stage Mag Adjustment	1× for each mag in range 300×–1500×	5	Every few weeks
Image Shifts	1× for each magnification used by Leginon	5	Seldom
Beam Shifts	1× for each magnification used by Leginon	5	Seldom
Beam Tilt	1× at magnifications used for autofocus	5	Every few weeks
Flat Field	1× for every CCD camera configuration	10	Every few weeks
CCD sensitivity	1×	1	Never
Dose	1× for magnifications used for exposure images	1	Every session

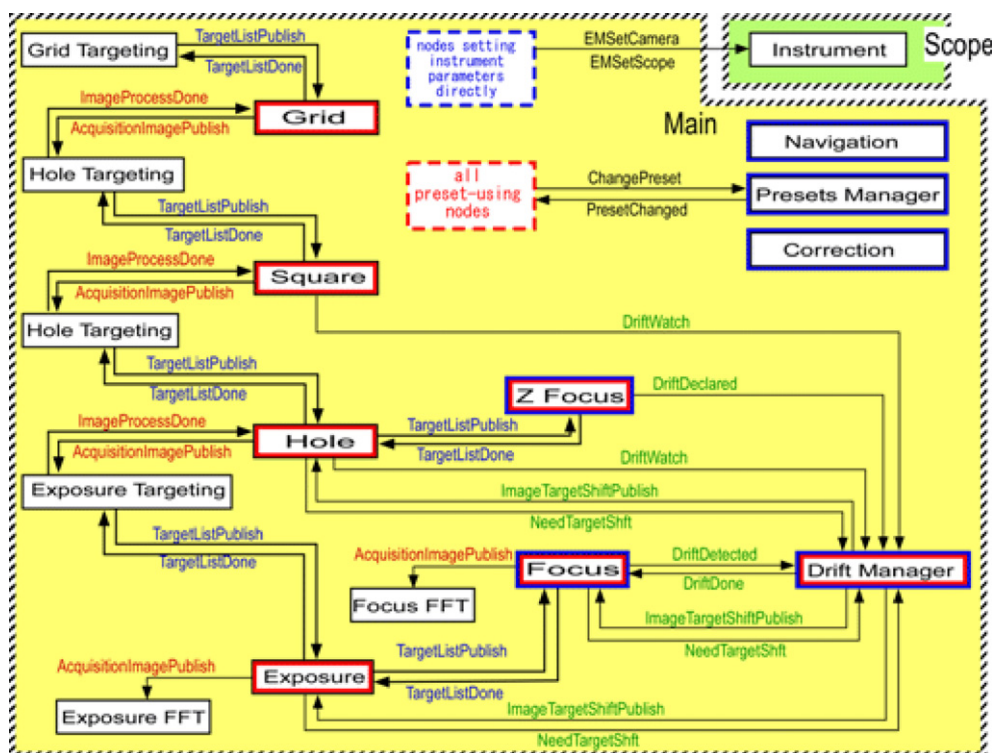


Fig. 6. A simplified schematic overview of the MSI application network. The MSI application links together acquisition and target searching nodes at four magnification scales (grid, square, hole, and exposure) and requires the setup of six primary microscope imaging preset conditions (grid, square, hole, exposure, Z-focus, and focus). Nodes requiring preset conditions (outlined in red) must be bound to the Presets Manager Node. Nodes which directly set microscope parameters are outlined in blue and must be bound to the Instrument Node. Events are denoted by the arrows connecting the nodes. Events associated with images are denoted in red, those associated with targeting are denoted in blue and events controlled by the Drift Manager Node are indicated in green. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

and the final exposures, as well as those required to set the defocus and measure drift, and to set the eucentric height. The event pathways which have been setup for this application follow a well defined priority protocol. A magnification acquisition node will acquire an image of a target (e.g., a square) and then wait for all higher magnification levels of targeting and acquisition (for example: holes, focus targets, high magnification focal series, etc.) to complete before moving on to the next target identified at the lower magnification scale. Some details on the individual nodes that make up the MSI application are presented below:

### 3.1.1. Instrument control node

The instrument control node runs on the computer directly connected to the microscope and camera and is responsible for transferring commands from other Leginon nodes to the devices, retrieving information on the state of the microscope and receiving and transferring images from the CCD.

### 3.1.2. Acquisition nodes

The basic task of an acquisition node is to acquire an image at a given target position using preset microscope

and camera settings. There are four acquisition nodes in the MSI application responsible for acquiring images at scales relevant to imaging the entire *grid*, individual *grid squares*, individual *holes* within a square and the final high magnification *exposures*. An acquisition node normally is activated by receiving a target from a targeting node. Its first task is to move the target to the center of the viewing area. Repositioning the target can be achieved by using either standard goniometer stage movements, stage movements adjusted for the modeled behavior of the goniometer, or the image shift coils. The method used to reposition is an option that can be selected by the operator but in general we use the normal movements of the goniometer when large movements ( $> \sim 50 \mu\text{m}$ ) across the grid are required, for example when repositioning a selected grid square to the center for further imaging. We use modeled stage movements when we need to move over smaller distances ( $\sim 5\text{--}50 \mu\text{m}$ ) more accurately, for example when centering a hole selected from a square image. And finally we use the image shift coils when a high level of accuracy is required, for example in centering a target for final high magnification imaging. Once the target has been centered, the acquisition node sets up one of the

preset microscope conditions by passing a request to the Instrument Control Node and then acquires an image.

### 3.1.3. Grid targeting

An atlas of the entire grid is generated by passing a raster of targets to an image Acquisition node. The targets are generated so as to acquire a mosaic of images by moving back and forth across the grid, to minimize the required goniometer movements. The atlas is a quasi-optimal set of images needed to cover a defined circular area. Typically, the grid acquisition node is associated with a magnification of  $\sim 120\times$ , and  $\sim 25$  images are required to provide coverage of the complete 2 mm grid holder aperture as shown in Fig. 7A. The collection of the atlas usually requires about 10 min to complete if the entire grid is imaged.

### 3.1.4. Square targeting

Grid square targets can be selected from the grid atlas using a combination of automated and manual methods. Initially an automated algorithm identifies the grid squares in the atlas by computing a binary threshold image and finding contiguous regions in the image. The mean and standard deviation of the contiguous regions is then calculated and the region is selected if these values are within a user specified range (see Fig. 7B). The user then has the option of modifying the selection by either deleting the automatically identified square target or adding additional targets to the selected set. Once a set of grid squares has been selected, the data acquisition process is started and, unless halted by the operator or by an error, can continue completely automatically until all squares have been processed.

### 3.1.5. Hole targeting

The goniometer is positioned to the center coordinates of each selected grid square target and an image of the grid square is acquired using an Acquisition node associated with a preset condition suitable for acquiring an image of an entire grid square. An example of a grid square image acquired at a nominal magnification of  $500\times$  is illustrated in Fig. 7C. The acquired images are passed along to the hole targeting node which allows for manual or automated selection of holes. The automated algorithm first enhances edges using a Sobel filter, then cross correlates the edge enhanced image to a ring template where the inner and outer diameter are parameters which can be modified by the user. The phase correlation map is subsequently thresholded and the resulting peak locations are tested against the expected spacing and geometry of the hole lattice. Our experience has been that the parameters required in the various steps in the hole finder algorithm will require some optimization, particularly when the holes are not circular or if there is minimal contrast between the hole and the background carbon. This optimization is part of the

initial calibration and setup procedures and we are usually able to settle on suitable parameters after only a few minutes of experimentation.

After the location of all potential holes in a square is established, the ice thickness of the holes is estimated by measuring the average intensity,  $I$ , and standard deviation within a circular region of the image located at the center of the hole. The intensity value is then used to estimate ice thickness as

$$t \approx C \log(I/I_0),$$

where  $I_0$  is the intensity of the beam measured over an empty hole. This method provides only an estimate of ice thickness and the value of  $C$  is dependent on specific imaging conditions (keV, defocus, size of the holes in the carbon, etc.) It does, however, provide a predictable method for choosing holes with consistent ice thickness over time once a baseline has been established by an initial manual inspection of the acquired images. Thus, the user normally adjusts the limits set for acceptable ice thickness values after inspecting the first few acquired images. The standard deviation measured for each hole is used to reject holes that contain contamination or other large variations in ice quality.

### 3.1.6. Setting eucentric focus

Once suitable target holes have been identified on the square, the next step is to check, and if necessary reset, the height of the specimen to the eucentric position (i.e., the position of the goniometer  $z$ -axis at which tilt does not couple to sample translation). Maintaining the specimen at the eucentric height is critical to the stability of the image and beam shifts required to keep targets centered while changing magnification scales. To perform this task we normally use an area of the grid centered on a hole that was previously identified as having ice not suitable for high magnification imaging. In the event that all holes in the square are designated as acceptable targets, one of them is sacrificed for the procedure and then deleted from the list of targets. The eucentric height is adjusted by the  $Z$ -focus node which is similar to an acquisition node in that it has the ability to center a target and setup a preset microscope condition. However, in the  $Z$ -focus node the simple image acquisition step is replaced by a set of procedures to measure the  $z$ -axis offset of the stage and reset it to the eucentric position. These procedures are based on measuring the image shift associated with a known beam tilt and are similar to those used for auto focusing. The procedure for measuring and adjusting eucentric height typically requires about 50 s to complete.

### 3.1.7. Exposure targeting

At the completion of the eucentric height procedure the hole targets selected from the square images are passed to the hole acquisition node. This node uses the



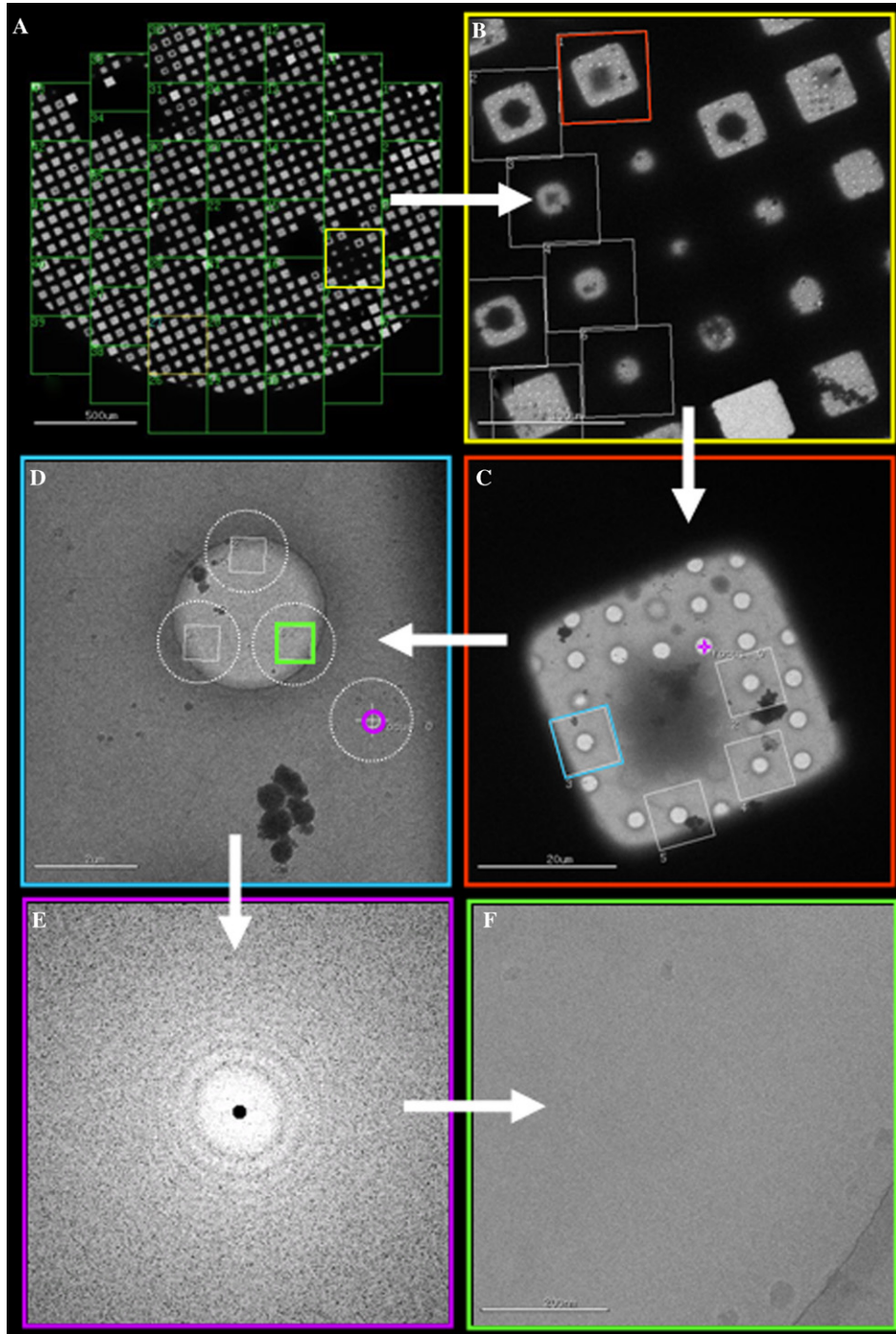


Fig. 7. Targeting and acquisition steps in the MSI application. Color codings indicate the target/acquisition connections. The first step is always the acquisition of an overall grid atlas (A) constructed from a mosaic of images, acquired at a nominal magnification of 120 $\times$ , one of which is shown in (B). Squares selected from the atlas are imaged at 500 $\times$  (C) and used to select holes, imaged at 5000 $\times$  (D). One of the holes, marked with a purple cross in (C) is used as a Z-focus target to reset eucentric height. From the hole image (D) final high magnification targets are selected along with a low dose focus position marked by a purple circle. The approximate size of the electron beam used to acquire the focus image, shown as a power spectrum in (E), and the final exposure image in (F) are indicated by dashed circles in (D). The focus and exposure images were both acquired at a nominal magnification of 50 000 $\times$ .

modeled goniometer movement to move one of the targeted holes to the center of the viewing screen, set the microscope to a preset magnification ( $\sim$ 5000 $\times$ ) and acquires an image of the hole as illustrated in Fig. 7D.

The mechanical motion of the goniometer is not sufficiently accurate to locate the center of the hole precisely at the center of the field of view even when using the modeled goniometer movement. As a result the center

of the hole will not normally coincide with the center of the image and thus it is necessary to determine the location of the center of the hole to provide a precise reference point from which to setup high magnification targets and the low dose focus position(s). The exposure targeting node is thus essentially identical to the hole targeting node described previously and differs only in the parameters used to select targets. The precise number and locations of the high magnification targets relative to the hole center are set by the operator and may be changed at any time during the experiment. For example, a single target may be selected at the center of the hole or multiple targets may be selected in some set geometry relative to the hole size and center (Fig. 7D). The exposure targeting node also controls the selection of the low dose focusing positions. These can be arranged in a simple fixed geometry relative to the center of the hole (Fig. 7D) or, more typically, the node will automatically select a suitable low dose target by selecting one of a set of potential targets at a fixed radius from the center of the hole. This automatic selection procedure takes into account the mean and standard deviation of the image at the focus target position and is very useful in avoiding grid bars and discriminating against thick ice or contaminated areas which might complicate the auto focus procedure.

### 3.1.8. Drift check and focus and astigmatism adjustment

The low dose focus position(s) are passed along to the focus node which is identical to the Z-focus node described earlier, again differing only in the settings within the node. This node moves the low dose target to the center, adjusts the microscope to the preset magnification setup for focusing, and then passes control to an algorithm responsible for checking drift and measuring and resetting defocus and astigmatism. An optional feature allows for setting a time interval during which the beam is used to burn off the ice and expose the carbon layer. The drift check procedure acquires pairs of images from the low dose target position and measures the displacement between them by phase correlation to estimate the drift velocity of the specimen movement. The time points for each image are based on the system clock at the time each image was acquired and thus provide a reasonably accurate estimate of the time interval. If the drift is detected to be greater than a user specified tolerance, control is passed along to a drift manager node which will be described in detail below. Only after the drift is determined to be within tolerance does control pass back to the focus node which continues with the defocus and astigmatism correction procedure. The autofocus algorithm we use mimics the well established method of measuring displacements as a function of induced beam tilt as described in (Koster and de Ruijter, 1992). When the focus procedure is completed an image is usually acquired at the low dose focus position and a

power spectrum of the image is calculated (Fig. 7E). These images are frequently used to provide an accurate estimate of the contrast transfer function of the instrument during post processing procedures.

### 3.1.9. Final exposure imaging

The final step is to acquire high magnification images at the target positions selected by the Exposure Targeting node. When acquiring multiple targets from a single hole, the footprint of the beam is adjusted so that there is no overlap between the exposed areas targeted while also ensuring that part of the background carbon support film is illuminated, as illustrated in Fig. 7D. Defocus pairs or series of images can then be acquired from each exposure target position using a standard acquisition node associated with appropriate preset conditions. The first image is typically acquired close to focus ( $\sim 3\text{--}1\ \mu\text{m}$  underfocus) with a dose of  $\sim 10\ \text{e}^-/\text{\AA}^2$  and the second further from focus ( $\sim 1\text{--}3\ \mu\text{m}$  underfocus) with a total accumulated dose of  $\sim 20\ \text{e}^-/\text{\AA}^2$  (Fig. 7F). The exposure node can also optionally be connected to an FFT node so that the power spectrum of each image is automatically calculated and stored to the database.

There are several additional nodes incorporated into the MSI application. The *presets manager* node is used to define the microscope and camera settings that will be used by the various acquisition and focus nodes. These settings are stored in the database and thus can be shared by a number of different applications; for example the calibration application. The *navigation* node provides some capabilities for manual navigation around the grid and is most often used during the initial setup as an aid to aligning the relative image and beam shifts between the presets. The *correction* node is responsible for modifying acquired images by performing flat field corrections, removing any bad rows and columns in the CCD array, and suppressing spurious spikes in the image resulting from stray X-rays.

The *drift manager* node is responsible for measuring and adjusting for specimen drift. The node is typically called from the focus node when drift is detected above the set tolerance. The drift manager node measures the drift in the same way as the focus node but provides a number of additional features, for example an option for waiting a set interval between consecutive drift measurements. More importantly, the drift manager provides the capability to adjust targets to compensate for a change in location resulting from the drift. Thus, once drift has been detected by the drift manager, any acquisition node which receives a target will request an adjusted set of targets from the drift manager node. The drift manager node will then acquire a new image at the magnification appropriate to the targeting operation, measure the offset from the previously acquired image and make available a list of targets that have been adjusted to compensate for the drift. The essential role

of the drift manager node is thus to provide a method for drift correction that is transparent to the operator. The drift manager is also used to correct targets after a eucentric height adjustment.

### 3.2. Other application examples

The MSI application can be readily adapted to other related data collection protocols. For example, when acquiring images from a continuous carbon grid the targeting node used for holes and exposures as described above is replaced by a node that instead of finding holes applies a user defined raster across the selected area. Targets are selected from this raster based on the mean and standard deviation of the image at the raster point. These values can be used to select ice of a defined thickness across the carbon when using vitreous ice specimens or areas of reasonably uniform stain from negatively stained specimens. A variety of other applications have also been developed using some subset of the basic nodes described above. For example, a very simple Manual Application has been developed as an aid for users who want to manually control the microscope and acquire images but who would like to use the Legion database as an archive for the images and parameters associated with them. The application network consists only of the Instrument control node, the Correction node and a simple Manual Acquisition type node. The user operates the microscope as usual and simply uses Legion to manually acquire an image whenever desired. The Manual Acquisition node allows for integration with the microscope “Low Dose” positions in that it will switch between the microscope FOCUS setting to the EXPOSURE setting when an image acquisition is requested. Images can be acquired and displayed either singly or continuously and optionally recorded to the database along with all microscope and CCD settings (for example, KeV, magnification, pixel size, nominal defocus, binning, etc.) The Manual Acquisition Node also includes an option for tracking grid numbers which is useful when screening large numbers of grids. The value of the manual acquisition application to a user is in providing a rational record of an experiment and in providing the web based viewing tools which allow the images to be browsed and readily shared with collaborators.

A further example of a Legion application is provided by a system that has been developed to screen large numbers of grids by integrating a robotic grid loading system into the Legion architecture (Potter et al., 2004). The application includes several of the nodes already described in the MSI application which are used to acquire a grid atlas and select targets on squares. It also includes an additional node which is used to communicate with the robotic system and synchronize the actions of the robot and the microscope.

## 4. Legion examples

We started using the new Legion system in our laboratory approximately one year ago but the software has only recently been brought to a level where we consider it stable and ready for general use and distribution to other laboratories. Nevertheless, we have collected a large number of datasets with the system and we will briefly describe some of these experiments to provide an idea of the practical results that can be achieved with the Legion system.

Over the last year we have used Legion to acquire data for about 150 experiments resulting in about 2 Tbytes of image data. In regard to the large size of the associated image storage it is worth noting that the accumulated total size of the entire database table structures is still less than 100 Mbytes and this is the only data that is kept permanently online. All images acquired using Legion are automatically stored to an Institute wide data management archive system which combines a high performance disk system and a large high speed tape library to provide for essentially unlimited storage. The Legion administration tools include a web interface to the archive system that provides a convenient, searchable system for monitoring the datasets, determining their online or offline status, and requesting retrieval from the archive if they are offline.

On average, data collection sessions over the past year tended to last about 12 h (with an average of ~400 total images acquired); the longest experiment on a vitreous ice specimen lasted ~35 h (~1700 images acquired) and the longest robotic screening experiment lasted 120 h (~3200 images acquired). The total time for the experiment includes time for the initial calibration checks and adjustments, setting up the preset alignments, adjusting the parameters of the targeting algorithms and a built in wait time of about 20 min to allow the column to readjust after the grid atlas is acquired. The setup procedures are typically completed within about an hour by an experienced user and even relative novices can operate the system with reasonable efficiency. This was evident during a recent weeklong workshop to train new users of the Legion system. As part of the course the trainees installed a full working version of Legion on their own laptop computers and were then required to use this system to start up Legion and acquire data using the MSI application during an allocated session of 3 h. All four groups participating were able to complete this task and successfully acquire data from the test grid. The only instance when the setup time can become extended is when the microscope has been substantially realigned, in which case the Legion presets and alignments might need to be completely readjusted, a procedure that in the worst case can take up to about an hour to complete.



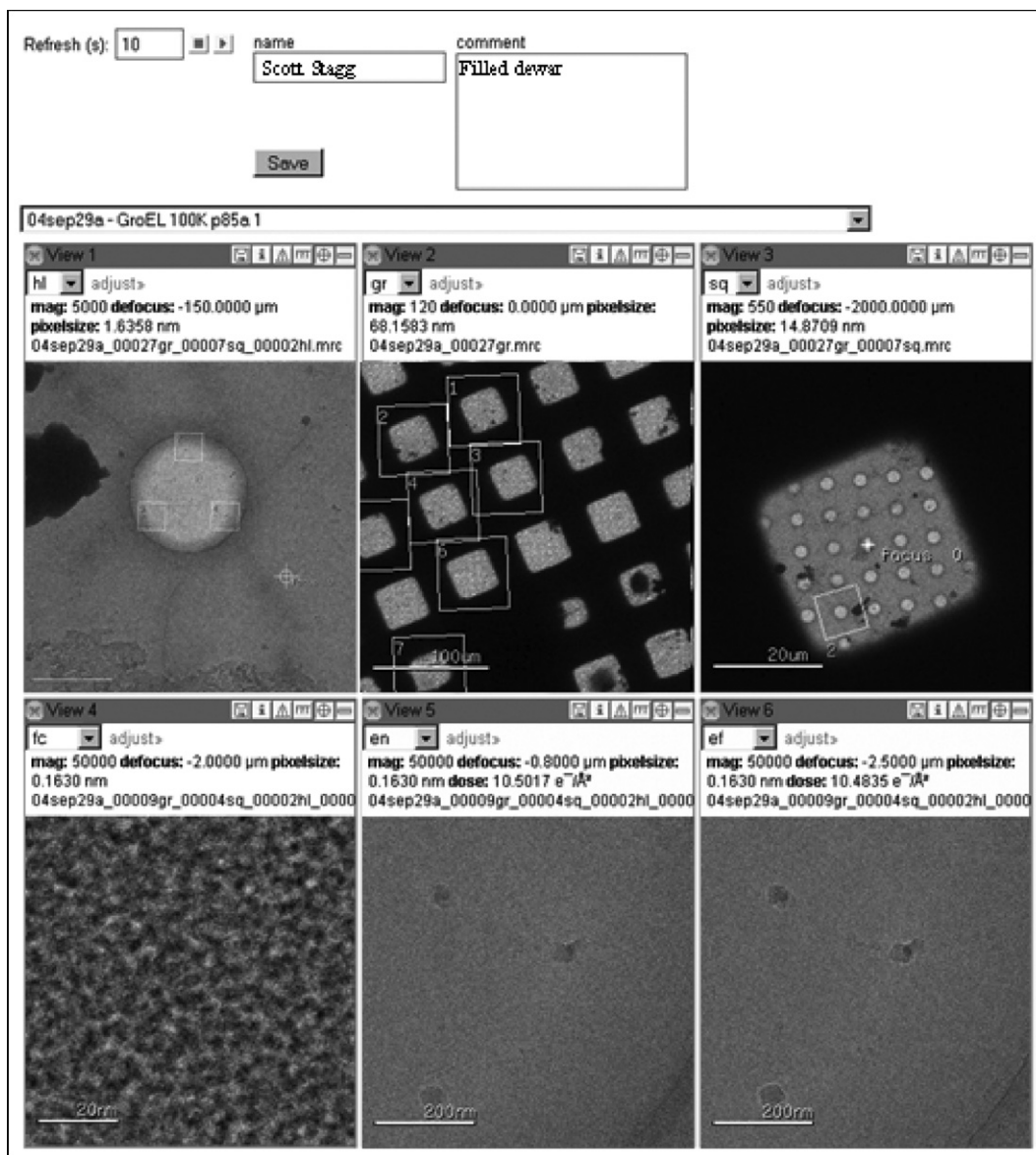


Fig. 8. The Legion Observer Interface is a web based viewer that displays the most recently acquired images at a selected refresh rate. The viewing windows may be set to display any of the images acquired by Legion and can be closed if they are not required. The Legion Observer interface can be used to monitor the status of an experiment from any computer with a network connection and a web browser. It also provides the capability for entering user comments which are time stamped and stored to the database.

Once the system has been setup, unless the operator requests manual targeting or user confirmation after one of the automated algorithms, the only operator intervention explicitly required is the replenishment of liquid nitrogen (LN2) in the side entry cold stage when imaging vitreous ice specimens. The cold trap on the microscope is automatically replenished using a system provided by the Norhof company<sup>8</sup> which allows for precision controlled filling of LN2. The system is controlled by two sensors, providing an upper and lower level for LN2 in the cold trap dewar. The system can run contin-

uously for many days at a time and is essential for maintaining the cold trap during extended sessions required by the robotic screening application. We hope to implement a similar system to automate the task of filling the cryostage, which on our systems maintains vitreous ice specimens at a temperature of  $\sim -170$  C, but currently we are still doing this by hand. Thus every 1–2 h, the cold stage dewar is topped off and a small amount of LN2 is rapidly boiled off to eliminate bubbling in the dewar.

During automated data collection the system is typically monitored using a specifically adapted version of the web based viewing interface to the database, the Legion observer interface (LOI) (see Fig. 8). This inter-

<sup>8</sup> <http://www.norhof.com/>.



face may be setup by the operator to display any set of images acquired by the acquisition or focus nodes. The LOI then manages the display so that only the most current images are displayed. The LOI also allows the user to enter a comment at any time and these comments are time stamped and stored into the database so that they can later be linked to an image or set of images. The LOI is very useful when running long experiments as it allows the operator to monitor the experiment from his/her office or from home. The data served from the database is scaled and compressed (by factors which can be selected by the user) and can be viewed quite efficiently using a broadband network or even a modem if the large  $4k \times 4k$  images are reasonably compressed. Any image acquired to the database is also immediately available for viewing using any of a variety of other web based viewers (see Fig. 9). These viewers are typically used in post experiment assessment and to transfer data to remote collaborators.

To assess the capabilities of the automated system for acquiring high throughput data we set ourselves the goal of acquiring more than 100 000 single particles during a single experimental session. The significance of the number 100 000 arises from the view, generally accepted by the cryoEM community (Glaeser, 2004b; Glaeser,

1999; Henderson, 1995), that approximately this number of asymmetric units of a macromolecule will be required to reconstruct a map to a resolution where it becomes possible to discern secondary protein structure ( $\sim 6\text{--}10 \text{ \AA}$ ). While there have been certain dramatic exceptions (Yonekura et al., 2003), the experimental evidence so far has tended to support this rule of thumb. As a test sample we used GroEL, a molecular chaperone derived from *Escherichia coli*, that has been extensively studied using both X-ray (Braig et al., 1994) and EM methods (Ludtke et al., 2004; Ranson et al., 2001; Roseman et al., 1996, 2001). We have now completed several experiments in which we have acquired more than 100 000 images of individual GroEL particles using the MSI application of Legimon. Some details of one of these experiments are provided in Table 3; data was acquired during a 20 h session in which a total of 354 defocus pairs ( $4k \times 4k$  CCD frames; magnification  $50k\times$ ;  $2.26 \text{ \AA}/\text{pixel}$ ) were acquired from 157 holes automatically selected from 26 grid squares. An automated particle selection algorithm (Roseman, 2003, 2004; Zhu et al., 2004) was then used to select a total of  $\sim 130\ 000$  particles from these images; illustrated in Fig. 10. A complete description of this work is currently being prepared for publication.

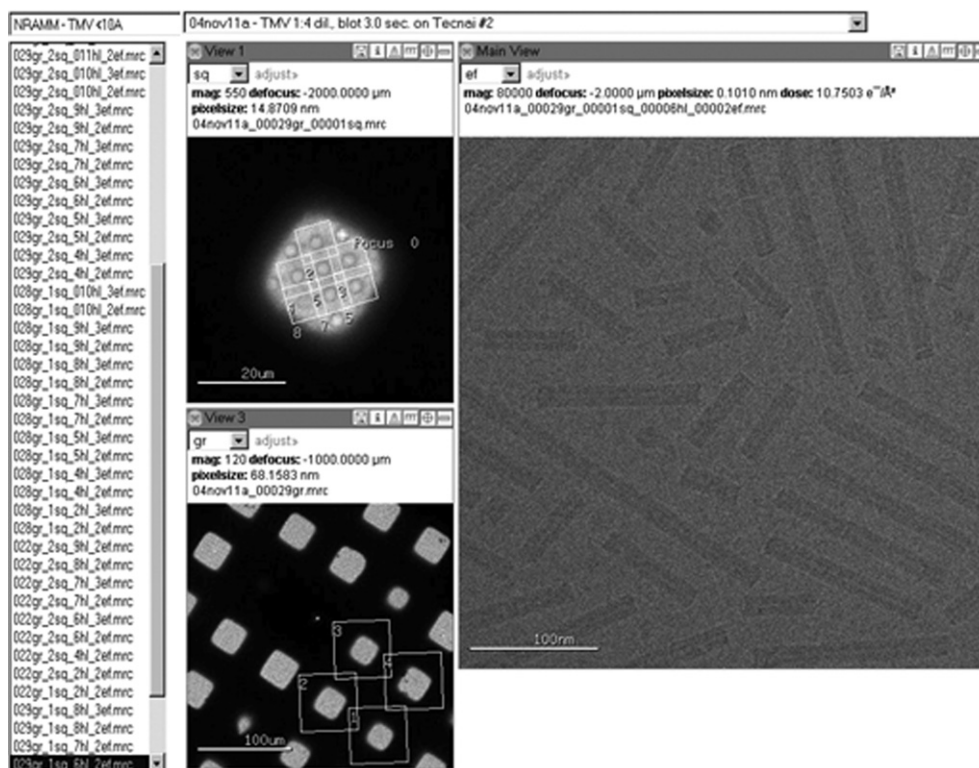


Fig. 9. Several web based viewing tools may be used to inspect the images stored in the Legimon database. In the example shown here the high magnification, far from focus image is being displayed in the main window and the corresponding square and grid atlas regions are shown in the smaller windows. The user can browse through the set of images selected in the main window using the file list shown to the left and the appropriate images will be loaded into the smaller windows if they are related. Icons along the top, right hand image frame are used to turn the display of targets or scale bars on or off, compute a Fourier transform, save the image to a file, or generate a new page providing a detailed description of the image.

Table 3  
Leginon experiment examples

Example experiment	Specimen	Preservation	No. of grids	No. of squares	No. of holes	No. of exposures	Total No. of images	No. of particles selected	Total time (h)
1	GroEL (~600 kD)	Vitreous ice	1	26	157	354 * 2	1042	130 000	20
2	Lobster CP (~450 kD)	Vitreous ice	1	133	183	374 * 2	1080	11 000	28
3	Virus-like particle (~2 mD)	Vitreous ice	1	4	62	181	353	8 000	8
4	Physical surface feature	Stain	1	4	38	292	380	—	6

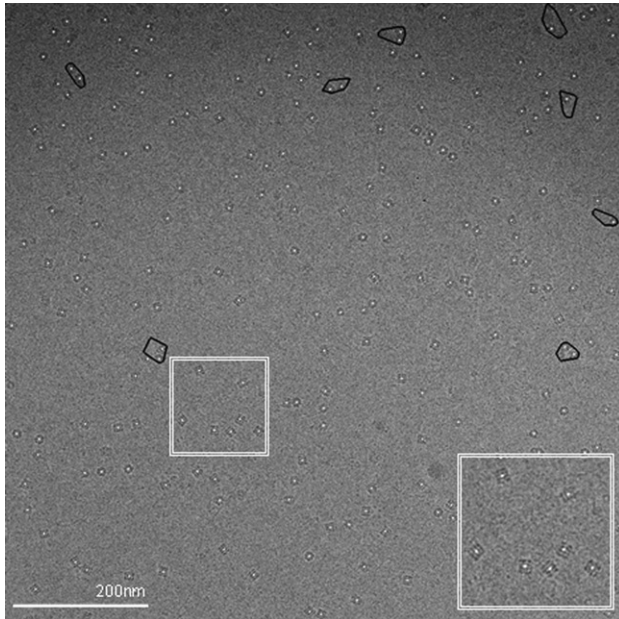


Fig. 10. An example of one of the 354 far from focus images (2.0  $\mu\text{m}$  underfocus, 50 000 $\times$  nominal mag., pixel size 2.263 nm) acquired of a GroEL test specimen. Particles automatically selected from these images are indicated with white dots. Particles within the black outlined areas were excluded. The area inside the double white line has been enlarged in the inset to provide a more detailed view. A total of 133 216 particles were automatically selected from the complete set of far from focus images.

We have also provided in Table 3 some details of other fairly typical results of Leginon experiments. For example, in collaboration with Justin Kollman of UCSD, we have used Leginon to collect several large datasets of a clottable protein (~420 kD) of decapod crustaceans (Fuller and Doolittle, 1971a; Fuller and Doolittle, 1971b). In the example illustrated in Table 3, 375 high magnification image pairs (50k $\times$ , 1.63  $\text{\AA}$ /pixel) were acquired in a session lasting ~28 h. From these images approximately 11 000 individual macromolecules were manually selected for single particle processing. The third example shows results from a collaborative project to characterize the structure of virus like particles (diameter ~50 nm). In a session lasting ~8 h, 191 high magnification images (50k $\times$ , 1.63  $\text{\AA}$ /pixel) were acquired from which ~8500 individual particles were automatically selected. A final example is shown for data collection from a very sparse specimen prepared over

continuous carbon and which did not require vitrification. The experiment in this case was completely automated, as no dewar filling was required, and ran unattended overnight. A total of 292 high magnification images were acquired in ~6 h.

## 5. Leginon distribution

We recently released the first official version of the software, Leginon 1.0, which is freely available to academic and non-profit institutions (see <http://nramm.scripps.edu> for details regarding the academic licensing agreement). The distribution includes the complete source code for the Leginon system and the applications discussed in this paper as well as all of the software required to use the database viewing tools. All prerequisite packages (e.g., python, php) for installing Leginon are available for download on our website. Linux packages are in rpm format for use with the SuSE Linux operating system. Windows packages are distributed as executable installers. The Leginon system itself is packaged in the same way.

The system includes detailed documentation, so far about 150 pages of documentation have been written and include installation instructions, training examples and detailed descriptions of the applications. The documentation is publicly available at the Leginon website (<http://nramm.scripps.edu/software/leginon/>). In principle, the system could be remotely uploaded and installed and the documentation could provide sufficient information to get a new group started. In practice however, we believe that at this early stage of development and given the complexity of the applications, a more effective method for distribution is to provide on site training and we have thus initiated a series of week long training workshops designed for groups who want to transfer the system to their own instruments at their home institutions. In the future we hope that groups other than our own will develop new nodes and application networks for Leginon and make these available to all Leginon users.

The Leginon documentation has been implemented using Docbook (Walsh and Muellner, 1999). DocBook is an XML vocabulary particularly well suited to technical documentation. The DocBook vocabulary allows

documents to be stored in a “presentation-neutral” form that can readily be reformatted into HTML pages, PDF files, or any of a wide variety of presentation formats, including Microsoft HTMLHelp, UNIX man pages, JavaHelp, TeX, TeXinfo, RTF, etc. We have modified docbook to add the capability for annotating the documentation pages using a web browser. The note system is implemented using PHP and allows notes to be entered by any user reading the documentation using a standard web browser. The additional notes are stored into a database from where they are retrieved for viewing with the appropriate documentation page. Users of the software can thus annotate the documentation to reflect personal usage preferences or additional useful directions which could at some later stage be incorporated into the official documentation pages.

The Leginon software source and documentation is managed using CVS,<sup>9</sup> a version control system which tracks the history of source files and documents and takes care of merging files which have been modified by multiple developers. CVS has tags and branches for maintaining and revising different release versions (1.0, 1.0.1, 1.1, etc.) Bug tracking and reporting in Leginon is managed using Bugzilla.<sup>10</sup> This web based system allows any registered user to enter a bug which can be marked with a specific severity (minor, major, normal, enhancement) and priority (low, normal, and high). The software development team uses the Bugzilla web pages to track and prioritize bug fixes and new feature development in Leginon.

## 6. Conclusions and future directions

The development of the new Leginon system took place over a period of about two years. During the first year the software infrastructure and new database were designed, the basic nodes were implemented and some preliminary applications were developed. At the beginning of the second year the system was put into general use in our laboratory and the software development team worked very closely with an application testing team to refine the main applications for maximum efficiency and ease of use, develop a graphical user interface and write the necessary documentation. At least 10 people have been involved in the development of the system, several of them working almost full time on the project for most of the two year period. The time and effort involved in this project correspond to the “high initial costs” associated with the development of any automated system as defined in the opening paragraph of this paper. Nevertheless, we are satisfied that

the system as it now exists more than justifies the effort involved in the development. Over the next year we will develop a number of new nodes and application networks. For example we plan to develop a node to calibrate the beam intensity and spread so that dose can be automatically set and controlled by the application rather than by precise setting of the preset acquisition parameters. This will be the first step towards a more generalized application to automate the calibrations and setup, and monitor and adjust these during the course of an experiment. New application networks that we plan to develop include capabilities for acquiring tilted image pairs and sequences of images as required for random conical or tomographic reconstruction geometries.

Our ultimate goal is to develop an automated “pipeline” for molecular microscopy that will integrate, and automate as far as possible, the entire process, including specimen preparation, specimen screening, data acquisition, data evaluation, and image processing and reconstruction. In this sense we follow in the footsteps of the recent advances of macromolecular X-ray crystallography to develop high throughput structure determination pipelines (Abola et al., 2000). Finally, we should note that our goal in automation is not simply to replace human operators by machines but instead to free the intelligence and creativity of people to concentrate on the much more difficult and challenging tasks of choosing the interesting problems and interpreting the final results.

The contributions of each of the authors to the work described in this paper were as follows: James Pulokas, Christian Suloway, and Denis Fellmann designed and developed the Leginon system. James Pulokas led the overall development effort, Christian Suloway led the effort for the instrument interface and the graphical user interface, Denis Fellmann led the database design and integration effort. Anchi Cheng worked closely with the developers in testing the system and suggesting refinements and improvements. Anchi Cheng and Francisco Guerra wrote the documentation. Joel Quispe and Scott Staggs tested the system and suggested many improvements. Clint Potter and Bridget Carragher managed the overall effort and authored the paper.

## Acknowledgments

We thank Dr. Arthur Horwich for supplying us with the GroEL test specimen and Dr. Justin Kollman for the data on lobster clotting protein. This project was primarily funded by a grant from the National Center for Research Resources at NIH (RR17573) with additional funding from NIH (GM61939) and NSF (DBI-0352386).

<sup>9</sup> <http://www.gnu.org/software/cvs/>.

<sup>10</sup> <http://www.bugzilla.org/>.

## References

- Abola, A., Kuhn, P., Earnest, T., Stevens, R.C., 2000. Automation of X-ray crystallography. *Nat. Struct. Biol.* 7, 973–977.
- Booth, C.R., Jiang, W., Baker, M.L., Hong Zhou, Z., Ludtke, S.J., Chiu, W., 2004. A 9A single particle reconstruction from CCD captured images on a 200 kV electron cryomicroscope. *J. Struct. Biol.* 147, 116–127.
- Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A., Sigler, P.B., 1994. The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* 371, 578.
- Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R.A., Potter, C.S., Pulokas, J., Reilein, A., 2000. Legion: an automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.* 132, 33–45.
- Carragher, B., Fellmann, D., Guerra, F., Milligan, R.A., Mouche, F., Pulokas, J., Quispe, J., Suloway, C., Zhu, Y., Potter, C.S., 2004. Rapid, routine structure determination of macromolecular assemblies using electron microscopy: current progress and remaining challenges. *J. Synchrotron Rad.* 11, 83–85.
- Cheng, A., Schweissinger, D., Dawood, F., Kumar, N., Yeager, M., 2003b. Projection structure of full length connexin 43 by electron cryo-crystallography. *Cell Commun. Adhes.* 10, 187–191.
- Dierksen, K., Typke, D., Hegrl, R., Koster, A.J., Baumeister, W., 1992. Towards automatic electron tomography. *Ultramicroscopy* 40, 71–87.
- Fellmann, D., Pulokas, J., Milligan, R.A., Carragher, B., Potter, C.S., 2002. A relational database for cryoEM: experience at one year and 50 000 images. *J. Struct. Biol.* 137, 273–282.
- Fuller, G.M., Doolittle, R.F., 1971a. Studies of invertebrate fibrinogen. I. Purification and characterization of fibrinogen from the spiny lobster. *Biochemistry* 10, 1305–1311.
- Fuller, G.M., Doolittle, R.F., 1971b. Studies of invertebrate fibrinogen. II. Transformation of lobster fibrinogen into fibrin. *Biochemistry* 10, 1311–1315.
- Glaeser, R., 2004a. Personal Communication.
- Glaeser, R.M., 1999. Review: electron crystallography: present excitement, a nod to the past, anticipating the future. *J. Struct. Biol.* 128, 3–14.
- Glaeser, R.M., 2004b. Historical background: why is it important to improve automated particle selection?. *J. Struct. Biol.* 145 15–18.
- Henderson, R., 1995. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* 28, 171–193.
- Kisseberth, N., Whittaker, M., Weber, D., Potter, C.S., Carragher, B., 1997. emScope: a tool kit for control and automation of a remote electron microscope. *J. Struct. Biol.* 120, 309–319.
- Koster, A.J., Chen, H., Sedat, J.W., Agard, D.A., 1992. Automated microscopy for electron tomography. *Ultramicroscopy* 46, 207–227.
- Koster, A.J., de Ruijter, W.J., 1992. Practical autoalignment of transmission electron microscopes. *Ultramicroscopy* 40, 89–107.
- Kuglin, C.D., Hines, D.C., 1975. The phase correlation image alignment method. *Proc. Int. Conf. Cybern. Soc.*, 163–165.
- Lei, J., Frank, J., 2005. Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* 150, 69–80.
- Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., Chiu, W., 2004. Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure (Camb)* 12, 1129–1136.
- Mastrorarde, D.N., 1997. Dual-axis tomography: an approach with alignment methods that preserve resolution. *J. Struct. Biol.* 120, 343–352.
- Mouche, F., Zhu, Y., Pulokas, J., Potter, C.S., Carragher, B., 2003. Automated three-dimensional reconstruction of keyhole limpet hemocyanin type 1. *J. Struct. Biol.* 144, 301–312.
- Oostergetel, G.T., Keegstra, W., Brisson, A., 1998. Automation of specimen selection and data acquisition for protein crystallography. *Ultramicroscopy* 40, 47–49.
- Oosterhout, J.K., 1994. Tcl and the Tk Toolkit. Addison-Wesley, Reading, MA.
- Potter, C.S., Chu, H., Frey, B., Green, C., Kisseberth, N., Madden, T.J., Miller, K.L., Nahrstedt, K., Pulokas, J., Reilein, A., Tcheng, D., Weber, D., Carragher, B., 1999. Legion: a system for fully automated acquisition of 1000 micrographs a day. *Ultramicroscopy* 77, 153–161.
- Potter, C.S., Moran, P.J., 1992. Viewit: A Software System for Multi-Dimensional Biomedical Image Processing, Analysis, and Visualization.
- Potter, C.S., Pulokas, J., Smith, P., Suloway, C., Carragher, B., 2004. Robotic grid loading system for a transmission electron microscope. *J. Struct. Biol.* 146, 431–440.
- Pulokas, J., Green, C., Kisseberth, N., Potter, C.S., Carragher, B., 2000. Improving the positional accuracy of the goniometer on the Philips CM series TEM. *J. Struct. Biol.* 128, 250–256.
- Ranson, N.A., Farr, G.W., Roseman, A.M., Gowen, B., Fenton, W.A., Horwich, A.L., Saibil, H.R., 2001. ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell* 107, 869–879.
- Roseman, A.M., 2003. Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy* 94, 225–236.
- Roseman, A.M., 2004. FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* 145, 91–99.
- Roseman, A.M., Chen, S., White, H., Braig, K., Saibil, H.R., 1996. The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* 87, 241–251.
- Roseman, A.M., Ranson, N.A., Gowen, B., Fuller, S.D., Saibil, H.R., 2001. Structures of unliganded and ATP-bound states of the *Escherichia coli* chaperonin GroEL by cryoelectron microscopy. *J. Struct. Biol.* 135, 115–125.
- Rouiller, I., Pulokas, J., Butel, V.M., Milligan, R.A., Wilson-Kubalek, E.M., Potter, C.S., Carragher, A.B., 2001. Automated image acquisition for single particle reconstruction using p97 as the biological sample. *J. Struct. Biol.* 133, 102–107.
- Shih, W.M., Quispe, J.D., Joyce, G.F., 2004. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* 427, 618–621.
- Sumner, J.B., Dounce, A.L., 1955. Liver catalase. *Methods Enzymol.* 2, 775–781.
- Walsh, N., Muellner, L., 1999. DocBook: The Definitive Guide. O'Reilly.
- Yonekura, K., Maki-Yonekura, S., Namba, K., 2003. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* 424, 643–650.
- Zhang, P., Beatty, A., Milne, J.L.S., Subramaniam, S., 2001. Automated data collection with a Tecnai 12 electron microscope: applications for molecular imaging by cryomicroscopy. *J. Struct. Biol.* 135, 251–261.
- Zheng, Q.S., Braunfeld, M.B., Sedat, J.W., Agard, D.A., 2004. An improved strategy for automated electron microscopic tomography. *J. Struct. Biol.* 147, 91–101.
- Zhu, Y., Carragher, B., Kriegman, D.J., Milligan, R.A., Potter, C.S., 2001. Automated identification of filaments in cryoelectron microscopy images. *J. Struct. Biol.* 135, 302–312.
- Zhu, Y., Carragher, B., Potter, C.S., 2004. Improving template matching based particle selection for cryo-electron microscopy, IEEE ISBI04 conference proceedings.
- Ziese, U., Geerts, W.J., Van Der Krift, T.P., Verkleij, A.J., Koster, A.J., 2003. Correction of autofocusing errors due to specimen tilt for automated electron tomography. *J. Microsc.* 211, 179–185.