Journal of Structural Biology 198 (2017) 124–133

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Focus: The interface between data collection and data processing in cryo-EM

Nikhil Biyani [a], Ricardo D. Righetto [a], Robert McLeod [a], Daniel Caujolle-Bert [b], Daniel Castano-Diez [b], Kenneth N. Goldie [a], Henning Stahlberg [a,*]

[a] Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Basel, Switzerland
[b] BioEM Lab, C-CINA, Biozentrum, University of Basel, Basel, Switzerland

ARTICLE INFO

Article history:
Received 2 February 2017
Received in revised form 17 March 2017
Accepted 20 March 2017
Available online 23 March 2017

Keywords:
Cryo-EM
High throughput image processing
Automated data collection
Drift correction

ABSTRACT

We present a new software package called *Focus* that interfaces cryo-transmission electron microscopy (cryo-EM) data collection with computer image processing. *Focus* creates a user-friendly environment to import and manage data recorded by direct electron detectors and perform elemental image processing tasks in a high-throughput manner while new data is being acquired at the microscope. It provides the functionality required to remotely monitor the progress of data collection and data processing, which is essential now that automation in cryo-EM allows a steady flow of images of single particles, two-dimensional crystals, or electron tomography data to be recorded in overnight sessions. The rapid detection of any errors that may occur greatly increases the productivity of recording sessions at the electron microscope.

1. Introduction

Cryo-electron microscopy (cryo-EM) and cryo-electron tomography are employed in structural biology to determine the structure of vitrified biological samples, such as isolated proteins and protein complexes, two-dimensional (2D) protein/lipid crystals, or larger biological objects such as bacteria and nanocrystals. Until the development of direct electron detector cameras (DED) for electron microscopy, the resolution achieved in most cases was in the 5 - 30Å range, with few exceptions (Gonen et al., 2005). Unlike detectors that convert the electron signal to light via a scintillator, DEDs have radiation-hardened metal oxide semiconductor sensors that can be directly illuminated with the high-voltage primary electron beam. This architecture greatly reduces the point-spread function of the electron signal and results in superior detector quantum efficiency, enhanced readout speed and an excellent signal-to-noise ratio (SNR). The resolution required to count single electron events is achieved. Furthermore, the low read-out noise of DEDs makes it possible to record electron dose-fractionated multi-frame exposures, known as image stacks or movies. This means that movement of the sample during imaging can be corrected by software, as discussed further below. Averaging the aligned motion-corrected frames after they have been weighted by a dose-dependent B-factor to account for damage to the proteins by the electron beam, results in images that have superior resolution and a better SNR (Rawson et al., 2016). Together with improved image processing algorithms, such as maximum-likelihood (Sigworth, 1998; Lyumkis et al., 2013) and Bayesian (Scheres, 2012) methods, this has led to a resolution revolution in cryo-EM (Kühlbrandt, 2014) with the result that atomic models of the imaged biological structures can quite often be built (Crowther, 2016; Kühlbrandt, 2014; Liu et al., 2016; Merk et al., 2016; Yu et al., 2016).

The Gatan K2 Summit DED camera was the first to offer three recording modes. In the first mode, the linear mode, the energy deposited in the DED sensor by passing electrons is integrated over the entire exposure time. In the second mode, termed electron counting, single electron events are individually detected and counted. Finally, in the super-resolution mode, the camera driver registers the impact location of individual electrons on the detector with sub-pixel accuracy during the electron counting step, resulting in final images of twice the pixel resolution than the hard chip

Abbreviations: cryo-EM, Cryo-electron microscopy; DED, Direct electron detector; 2D, Two dimensional; 3D, Three dimensional; CTF, Contrast transfer function; FFT, Fast Fourier transform; GPU, Graphical processing unit; CPU, Central processing unit; MRC, MRC file format; SNR, Signal-to-noise ratio; GUI, Graphical user interface.

* Corresponding author at: C-CINA, Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland.
E-mail address: henning.stahlberg@unibas.ch (H. Stahlberg).

http://dx.doi.org/10.1016/j.jsb.2017.03.007
1047-8477/

of the DED detector. Image data are ideally recorded in super-resolution mode as multi-frame exposures of each region of interest. Use of the super-resolution algorithm means that the "4 k" chip (3838 × 3710 pixels) of a K2 Summit DED records "8 k" frames (7676 × 7420 pixels). Subsequently downsampling these frames to "4 k" by cropping in Fourier space, results in 4 k images with a SNR significantly superior to the SNR of 4 k images directly recorded in the "4 k" counting mode, due to the effect of anti-aliasing (see also Ruskin et al., 2013).

When multi-frame exposures are recorded using a DED, the total dose given to a region of interest is fractionated, allowing the images in the stack to be aligned by post-processing to eliminate specimen movement originating from physical drift of the specimen stage, electronic drift in the imaging system of the microscope, or physical movement of the sample under the electron irradiation. Several software packages are available to correct specimen motion, including Zorro (McLeod et al., 2016), MotionCor2 (Zheng et al., 2017), Unblur (Grant and Grigorieff, 2015), alignparts_lmbfgs (Rubinstein and Brubaker, 2015), MotionCorr (Li et al., 2013), and SerialEM (Mastronarde, 2005). Among them, Zorro, MotionCor2 and alignparts_lmbfgs perform both whole-frame and local drift-correction. In most cases, additional local drift-correction significantly improves the quality of the output image, but also increases the computational costs considerably. Optimally, the microscope operator should be able to view drift-corrected and averaged images directly on the camera computer during the microscopy session. Thus, the development of software tools that can achieve whole-frame 'on the fly' drift correction has gained interest (Li et al., 2015; Noble and Stagg, 2015; Fernandez-Leiro and Scheres, 2016). Local-drift correction cannot (yet) be computed during data collection on a conventional DED camera driver, due to limitations in computing speed.

Once image stacks have been corrected for drift, they can be processed further using a reconstruction software package such as RELION (Scheres, 2012), FREALIGN (Grigorieff, 2007), CryoSPARC (Punjani et al., 2017), IMAGIC (van Heel et al., 2012) or EMAN2 (Tang et al., 2007; Ludtke, 2016) for single particles, IHRSR (Egelman, 2007) or Spring (Desfosses et al., 2014) for helical proteins, Dynamo (Castaño-Díez et al., 2012), PEET (Nicastro et al., 2006) or Jsubtomo (Huiskonen et al., 2010) for sub-tomogram averaging, or 2dx (Gipson et al., 2007a,b) for 2D electron crystallography. Further, software packages like Scipion (de la Rosa-Trevín et al., 2016) and Appion (Lander et al., 2009) aim at unifying the software available for single particle reconstructions, by allowing the output from one package to be used as the input for another in a transparent and manageable way.

As large data sets need to be recorded and processed, software packages such as SerialEM (Mastronarde, 2005), Leginon (Suloway et al., 2005), UCSF-Tomography (Zheng et al., 2007), Tom Toolbox (Nickell et al., 2005), EPU (Thermo Fisher Co.), Latitude in Digital Micrograph (Gatan Co.), or EMMenu (TVIPS, Germany), have been developed to extend or assist automated data collection at the microscope. In combination or alone (package dependent), they allow the automatic identification of specimen locations of interest and the collection of stacks of images or electron tomography tilt series comprised of stacks at each tilt angle based on initial manual input. This has greatly increased the output of electron microscopes, allowing them to record data for several days without interruption. However, because such sessions run day and night without the physical presence of an operator, a large number of unsuitable low quality images might be recorded for several hours, greatly reducing the productivity of the session. Ideally, the operator should be able to remotely monitor automated image acquisition and intervene from afar, if necessary. Software packages such as Leginon (Suloway et al., 2005) in combination with Appion (Lander et al., 2009), offer the possibility to remotely monitor some aspects of data acquisition. Here, we present an image-processing package called *Focus* that interfaces between cryo-EM data collection and data processing. *Focus* integrates software resources to create a user-friendly environment optimized to carry out electron microscopy image analysis tasks in a high-throughput manner, and runs several jobs in parallel on a batch-queue processor to achieve this. It can be used to remotely monitor image acquisition and aims at real-time drift correction. Statistics calculated for the recorded and processed images are summarized in a "*Project Library*" for user inspection and data pruning, and uploaded to a web server when *Focus* is set to remotely monitor image acquisition. *Focus* simplifies the execution of tasks, such as "Motion Correction" with various software (Zorro, Unblur, MotionCor2) or "CTF determination". It can be run in four different modes: "Drift correction only"; "Single particle", which implements particle picking and exports data in a format compatible with RELION; "2D electron crystallography", which implements fully automated 2D crystal image processing directly within *Focus*; "electron tomography", which offers drift-correction for dose-fractionated electron tomographs with consideration of incremental dose accumulation and rearrangement of recorded data from the so-called "Hagen Scheme" (Hagen et al., 2016) to "tilt-angle sorted" data. Additional functions and scripts can easily be added and integrated into the pipeline, by adding or editing C-shell or Python scripts, and by editing existing text files that define the set of available scripts or parameters. This allows to easily add scripts that call other third-party programs, as long as they can be reached via the command-line.

2. Implementation

2.1. Graphical user interface

The *Focus* graphical user interface (GUI) was developed using C++ (C++11 standards) and its user interface is based on Qt5.x. The creation of a user-friendly platform was of utmost priority. The workflow and controls are sectioned into various panels accessible via the top navigation bar (Fig. 1). These panels allow the user to perform processing tasks (in parallel if desired), manage images, view results and change settings. Graphical entities, such as icons, are used wherever possible. Processing tasks are accomplished either by precompiled executables or by python and shell scripts. *Focus* provides easy interaction with these executables and scripts (Fig. 2) via a combination of (i) parameter files and form containers that receive the input and pass it on to the required place, (ii) scripts that call external executables such as MotionCor2, (iii) a log container and progress bar to monitor the processing, and (iv) result containers that show the output values and the images produced. Images can be inspected using a full-screen image browser.

2.2. Project structure

In *Focus*, a dedicated project folder is created on disk for each research project using the *Project Wizard*, and assigned one of the following modes: (i) drift correction only, (ii) 2D electron crystallography, (iii) single particle, or (iv) electron tomography. When a new project is initialized, global parameters are established from a master parameter configuration file and stored in a project-level parameter file within this uppermost project folder (Fig. 3).

Each project contains groups of images encapsulated on-disk as subdirectories. Each image-group folder contains a unique subfolder for every image and its associated metadata. This image-oriented subfolder organization allows *Focus* to interface with other image processing packages. Before data are exported for use in a different follow-up processing environment, the export