Journal of Structural Biology 178 (2012) 139–151

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

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

Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments

Daniel Castaño-Díez *, Mikhail Kudryashev, Marcel Arheit, Henning Stahlberg

Center for Cellular Imaging and Nano Analytics (C-CINA), Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland

ARTICLE INFO

Article history:
Available online 8 January 2012

Keywords:
Subtomogram averaging
Single Particle Tomography
High-performance computing
GPU computing
Classification

ABSTRACT

Dynamo is a new software package for subtomogram averaging of cryo Electron Tomography (cryo-ET) data with three main goals: first, *Dynamo* allows user-transparent adaptation to a variety of high-performance computing platforms such as GPUs or CPU clusters. Second, *Dynamo* implements user-friendliness through GUI interfaces and scripting resources. Third, *Dynamo* offers user-flexibility through a plugin API. Besides the alignment and averaging procedures, *Dynamo* includes native tools for visualization and analysis of results and data, as well as support for third party visualization software, such as Chimera UCSF or EMAN2. As a demonstration of these functionalities, we studied bacterial flagellar motors and showed automatically detected classes with absent and present C-rings.

Subtomogram averaging is a common task in current cryo-ET pipelines, which requires extensive computational resources and follows a well-established workflow. However, due to the data diversity, many existing packages offer slight variations of the same algorithm to improve results. One of the main purposes behind *Dynamo* is to provide explicit tools to allow the user the insertion of custom designed procedures – or plugins – to replace or complement the native algorithms in the different steps of the processing pipeline for subtomogram averaging without the burden of handling parallelization. Custom scripts that implement new approaches devised by the user are integrated into the *Dynamo* data management system, so that they can be controlled by the GUI or the scripting capacities.

Dynamo executables do not require licenses for third party commercial software. Sources, executables and documentation are freely distributed on http://www.dynamo-em.org.

1. Introduction

1.1. Cryo Electron Tomography and subtomogram averaging

Cryo Electron Tomography (cryo-ET) images the three-dimensional (3d) structure of cells (Frank, 2006; Lucic et al., 2005). Samples are kept as close as possible to native conditions during their imaging in the Transmission Electron Microscope (EM), and the final computational model that describes the three dimensional structure of the cell is constructed by integrating different views of the same sample.

Although at typical working conditions, electron microscopy allows the resolution of features at the nanometer scale, image quality of the constructed 3d model is strongly degraded by different artifacts, such as the missing wedge, and low signal-to-noise due to limited dose to avoid radiation damage. As a consequence, while coarse features of the cell – as membranes and filaments – can be recognized with the naked eye, a quantitative description of the structure of individual macromolecules cannot be reliably recovered without further processing.

The *subtomogram averaging* technique (Forster and Hegerl, 2007; Winkler, 2007) aims at overcoming these limitations by identifying and averaging the common signal that can be found in different noisy copies of the same macromolecular compound. This conceptual similarity with Single Particle techniques also shows in the use of cross-correlation as main tool to align data particles to references of increasing quality. Subtomogram averaging is therefore also often referred to as "Single Particle Tomography", when the technique is applied on isolated particles (Schmid, 2011).

1.2. Software for subtomogram averaging

Subtomogram averaging is becoming an increasingly common tool for cryo-ET studies (Al-Amoudi et al., 2007; Beck et al., 2004; Brandt et al., 2009; Briggs et al., 2009; Liu et al., 2008; Medalia et al., 2002; Ortiz et al., 2010). Comprehensive software packages that cover a wide range of aspects in the analysis of tomographic data – such as acquisition, reconstruction, visualization, and particle picking – often incorporate support for subtomogram averaging, giving users the possibility to perform the full processing pipeline in the same environment. In this sense, the TOM package (Nickell et al., 2005) provides a framework for the tool AV3 (Forster and

* Corresponding author.
E-mail address: daniel.castano@unibas.ch (D. Castaño-Díez).

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2011.12.017

Hegerl, 2007), and subtomogram averaging utilities of the package PEET (Nicastro et al., 2006) are available from Etomo, part of the IMOD package (Kremer et al., 1996).

The jsubtomo package (Huiskonen et al., 2010) has been developed on the basis of Bsoft (Heymann et al., 2008), which incorporates its own subtomogram averaging utility (bfind), as also does the Protomo package (Winkler, 2007). SPIDER's technique RAMOS (Rath et al., 2003; Renken et al., 2009) is also available.

Other packages are specific for subtomogram averaging and address new algorithmic approaches. Bartesaghi et al. (2008) use a spherical harmonics formulation to accelerate numerically demanding computations, and Amat et al. (2010) use noise statistics models to increase the attained resolution.

Finally, well-established packages that originated from the Single Particle field are now including extensions for subtomogram averaging tasks. XMIPP includes the ml_tomo tool (Scheres et al., 2009), which implements a maximum likelihood approach for simultaneous alignment and classification of subtomograms (an approach also followed in Stolken et al., 2011), and EMAN2 (Tang et al., 2007) is now distributed with a specific module for Single Particle Tomography based on the algorithms described in Schmid and Booth (2008).

1.3. Dynamo overview

Many of the packages mentioned implement the same or a very similar pipeline to perform subtomogram averaging. However, they are not always easy to use, or they are not ready to benefit from available high-performance computing architectures. Finally, the best performing processing approach may differ from sample to sample, so that it would be helpful to be able to try different algorithms at once.

With this in mind, we created *Dynamo* as a new tool. *Dynamo* performs subtomogram averaging computations, and supports the development of new algorithms. *Dynamo* combines graphical interfaces with command line capacities to address three main goals:

1.3.1. User friendliness

Dynamo allows users an easy way to design a full subtomogram averaging experiment. First-time users will get their first average in minutes. Experienced users can concentrate on the physics aspects of the design of the experiment, as the cumbersome operational and technical aspects are transparently managed by the package.

1.3.2. High performance

Dynamo provides a compact platform to design or run experiments in the whole range of high-performance computing (HPC) environments that nowadays are becoming available to an increasingly larger number of users. The software package allows easy installation and operation in clusters of parallel processors, multicore desktops or servers with a single or several graphics processing units (GPUs) and also clusters of GPUs. The algorithm implementation has been adapted in each case in order to adjust the resource management to the different computing architectures and device communication protocols.

1.3.3. Flexibility

Dynamo intends to be used not only as closed package for subtomogram averaging, but also as a development tool. The package is structured in a modular form, allowing an easy insertion or replacement of new algorithmic steps. Specific software tools are provided to ensure that this adaptability does not come at the price of a reduced performance.

Each of the following sections is devoted to one of these goals. Section 2 presents the software package, leaving for Section 3 and the Supplementary materials a more mathematically oriented introduction of the particularities of the used algorithm. Section 4 discusses the performance of the package in different computing environments. Section 5 details then the mechanisms that allow a natural embedding of algorithms designed by the user into the native structure of the software. Finally, a case study on real data is presented in Section 6.

2. Software concept

Fig. 1 depicts the data processing steps that are required before beginning a subtomogram averaging experiment: alignment of tilt series, tomographic reconstruction and identification of approximate positions of particles of interest in the tomograms. The additional use of CTF correction procedures might be necessary towards high resolution (Fernandez et al., 2006; Kudryashev et al., 2012; Xiong et al., 2009).

Dynamo assumes that these steps were undertaken previously by the user, preparing a folder with all particle data, i.e., the subvolumes cropped from the tomograms.

2.1. Procedure through the GUI

The main GUI presents a unified view on all aspects of a project. It is shown in Fig. 2 along with a sketch of the basic procedure to compute an average from a set of particles. *Dynamo* divides user intervention into three operational steps (project setup, experiment design, and execution), which we briefly discuss in the following. Concrete guidelines and data formats are detailed in the manual, along with a walkthrough with an example synthetic data set intended to get the user started with *Dynamo* operations.

2.1.1. Project setup

The minimal input required from the user is the location in the file system of a folder that contains the particle data. Optionally, files containing other basic data (such as an initial reference, masks, or an initial guess for the alignment parameters) can be indicated by the user. If no such files are defined, the corresponding information will automatically be taken from default values.

2.1.2. Experiment design

The user indicates the number of desired iterations, and specifies parameters that determine exactly, which operations will be carried out at each group of iterations. These parameters might control the refinement loop on all particles (angular range and sampling to scan, frequency filtering, symmetrization or resizing of particles), or might modify the posterior operations (particle selection, classification, averaging) required to construct an updated reference.

Additionally, parameters for user-defined functions inserted in the *Dynamo* flow can also be passed through the main GUI.

2.1.3. Execution

The user selects the computing environment intended to perform the calculations, detailing how many CPUs or GPUs will be dedicated to the project. With this information, *Dynamo* automatically parses the project designed by the user into a command script, which can then be submitted to the computing environment for execution, or stored as part of a more complex project to be launched later.

2.2. Command line procedure

The main GUI is the first contact point of new users with the software, providing a simple way to design and supervise an