Journal of Structural Biology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms

Daniel Castaño-Díez [a,b,*], Mikhail Kudryashev [b,c,d], Henning Stahlberg [b]

[a] BioEM Lab at C-CINA, Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland
[b] Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland
[c] Max Planck Institute for Biophysics, Max-von-Laue-Straße 3, 60438 Frankfurt am Main, Germany
[d] Buchman Institute for Molecular Life Sciences, Goethe University of Frankfurt, Max-von-Laue-Straße 15, 60438 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 27 February 2016
Received in revised form 31 May 2016
Accepted 7 June 2016
Available online xxxx

Keywords:
Cryo electron tomography
Subtomogram averaging
Data management

ABSTRACT

Cryo electron tomography allows macromolecular complexes within vitrified, intact, thin cells or sections thereof to be visualized, and structural analysis to be performed *in situ* by averaging over multiple copies of the same molecules. Image processing for subtomogram averaging is specific and cumbersome, due to the large amount of data and its three dimensional nature and anisotropic resolution. Here, we streamline data processing for subtomogram averaging by introducing an archiving system, *Dynamo Catalogue*. This system manages tomographic data from multiple tomograms and allows visual feedback during all processing steps, including particle picking, extraction, alignment and classification. The file structure of a processing project file structure includes logfiles of performed operations, and can be backed up and shared between users. Command line commands, database queries and a set of GUIs give the user versatile control over the process. Here, we introduce a set of geometric tools that streamline particle picking from simple (filaments, spheres, tubes, vesicles) and complex geometries (arbitrary 2D surfaces, rare instances on proteins with geometric restrictions, and 2D and 3D crystals). Advanced functionality, such as manual alignment and subboxing, is useful when initial templates are generated for alignment and for project customization. *Dynamo Catalogue* is part of the open source package *Dynamo* and includes tools to ensure format compatibility with the subtomogram averaging functionalities of other packages, such as Jsubtomo, PyTom, PEET, EMAN2, XMIPP and Relion.

1. Introduction

1.1. Subtomogram averaging from cryo-electron tomograms

Cryo-electron tomography (cryo-ET), with its unique capacity for the three-dimensional (3D) visualization of macromolecular complexes in a close-to-native state, is a rapidly developing technology (Harapin et al., 2013; Lucic et al., 2013). Subtomogram averaging (STA) recovers the structure of a given macromolecule by locating multiple noisy copies of the object of interest in one or several tomograms, and integrating them into one or multiple structures with a higher signal-to-noise ratio (SNR) (Briggs, 2013). This integration typically consists of an *alignment* step that accurately locates and imparts a common orientation to the initially differently oriented particles, followed by *averaging* of the aligned particles and *classification* of the particles into more homogeneous groups. The procedure has been widely used, with applications ranging from large macromolecular complexes inside the native context (Beck et al., 2007; Kudryashev et al., 2013; Pigino et al., 2011) to membrane proteins on membranes (Faini et al., 2013; Pfeffer et al., 2012) or isolated protein complexes (Dudkina et al., 2011). Currently, 11% of the structures deposited in the electron microscopy data bank (EMDB) were solved by STA. Generally, there is a correlation between the number of particles and the resolution obtained (Kudryashev et al., 2012), suggesting that a larger amount of processed particles results in a higher final resolution. Several structures determined by STA have sub-nanometer resolution, revealing the secondary structure of the proteins present

Abbreviations: 2D, two dimensions/dimensional; 3D, three dimensions/dimensional; EM, electron microscope; cryo-EM, cryo-electron microscopy; cryo-ET, cryo-electron tomography; EMDB, electron microscopy data bank (www.emdatabank.org); FSC, Fourier shell correlation; SNR, signal-to-noise ratio; STA, subtomogram averaging.

* Corresponding author at: BioEM Lab, C-CINA, Biozentrum, University of Basel, Mattenstrasse 26, 4058 Basel, Switzerland.
E-mail address: daniel.castano@unibas.ch (D. Castaño-Díez).

http://dx.doi.org/10.1016/j.jsb.2016.06.005
1047-8477/

(Bartesaghi et al., 2012; Schur et al., 2013, 2015); in each case, a large number of 3D subtomograms was required.

Several software packages are available for the 3D processing of multiple subtomograms. These include *AV3* (Forster et al., 2005), *Jsubtomo* (Huiskonen et al., 2010), *PyTom* (Hrabe et al., 2012), *EMAN2* (Galaz-Montoya et al., 2015), *PEET* (Nicastro et al., 2006), *Relion* (Bharat et al., 2015), and *Dynamo* (Castano-Diez et al., 2012).

1.2. Particle picking

Subtomogram analysis starts with the extraction of subtomograms from one or several tomographic reconstructions. These 3D subtomograms are referred to as *particles* in the following. The location, orientation and exact number of particles of interest in tomograms are not defined *a priori*. In STA, "particle picking" commonly refers to a set of actions carried out automatically or with input from the operator, with the goal of approximately determining the positions and possibly also the orientations of subtomograms, to produce a set of particles, each *particle* containing a unique single copy of a macromolecule of interest, accompanied by adequately formatted metadata. Compared to two-dimensional (2D) particle picking in micrographs for single particle analysis, picking 3D particles from cryo-ET volumes has two specific difficulties. First, the data are 3D volumes which usually suffer from anisotropic resolution, making their visualization more complex. Second, particle identification is more challenging for both automated methods and human operators, since tomography typically images 3D particles in their native context as opposed to isolated single particles.

Further, in spite of the common goal, in tomography the generic term "particle picking" expresses different procedural approaches in different scenarios. Determination of particle locations might involve purely automatic methods relying on image analysis of the tomograms, such as a pattern matching-based identification, as implemented in *molmatch* (Frangakis et al., 2002), regular picking from a surface, e.g., a bacterial membrane (Amat et al., 2010), or picking from a tubular crystal (Bharat et al., 2012). Alternatively, particle picking might be based on visual inspection and manually targeting the structure of interest on a computer monitor. Integration of *a priori* information arising from geometric constraints is a frequent requirement: the particles to average might lie on cellular membranes (Kudryashev et al., 2013), on virus capsids (Huiskonen et al., 2010), on lipid vesicles (Faini et al., 2013), along the axial path of tubular structures (Nicastro et al., 2006; Pigino et al., 2011), or might be the repeating subunits that generate such structures. Sometimes, the particle coordinates might be additionally constrained by the presence of crystalline order (2D or 3D) or symmetry, like the single vertices of an icosahedral virus (Gil-Carton et al., 2015), and the symmetry assumption might need to be partially weakened to accommodate the actual behavior of the observed data (Peralta et al., 2013). Also, related geometric surfaces might be fully or partially populated with particles (Maurer et al., 2013). The various possible particle arrangements require the use of different combinations of automated and manual particle picking procedures

Here, we present an extension to the *Dynamo* package, called *Dynamo Catalogue*, which facilitates automated or semi-automated subtomogram particle picking from various geometries used in STA, such as isolated particles, filaments, vesicles, arbitrary 2D surfaces and 2D and 3D crystals. This toolbox is an integral part of the *Dynamo* workflow and allows scripting of repetitive tasks through the command line. *Dynamo Catalogue* features a graphical user interface (GUI)-enabled toolbox that assists in the management of the data and metadata, including geometry and metadata conversions, and offers pipelines for the STA workflow from particle picking to STA and classification. *Dynamo Catalogue* is available within the *Dynamo* package as open source at http://www.dynamo-em.org.

2. Results and discussion

2.1. Data management system in Dynamo

In order to streamline STA projects we have implemented a database that integrates particle picking and extraction into the *Dynamo* workflow. The database is organized as a catalogue system with a user-accessible GUI (`dynamo_catalogue_manager`; here and below the *Dynamo* commands are stylized in courier font) and a set of command-line tools for scripting. The objects of the database are (1) links to tomograms and (2) models describing particle geometry (Fig. 1). Properties of the tomograms include a link to the file location, defocus, magnification, and to the orientation of the missing wedge. *Dynamo* catalogues can be generated and administered using the GUI or a command line. They provide the following functionalities:

1. Tomogram browsing. Catalogues generate a gallery of thumbnails providing a quick overview of the data. GUIs for 3D tomogram viewing include tools for navigation of volumes designed for 3D particle picking (`dynamo_slicer`, `dynamo_tomoview`). `Dynamo_preview` allows large volumes to be viewed without loading them fully into memory. In the other viewers, a region of interest can be specified (in `dynamo_preview`), archived, loaded into memory, and fully annotated.
2. Geometric modeling. The user can choose a model describing the geometry of the particle distribution, *i.e.*, depending on whether they are isolated particles, filaments, vesicles, etc. An extensive library of geometric models is available, adapted to the typical data collection geometries encountered in life-

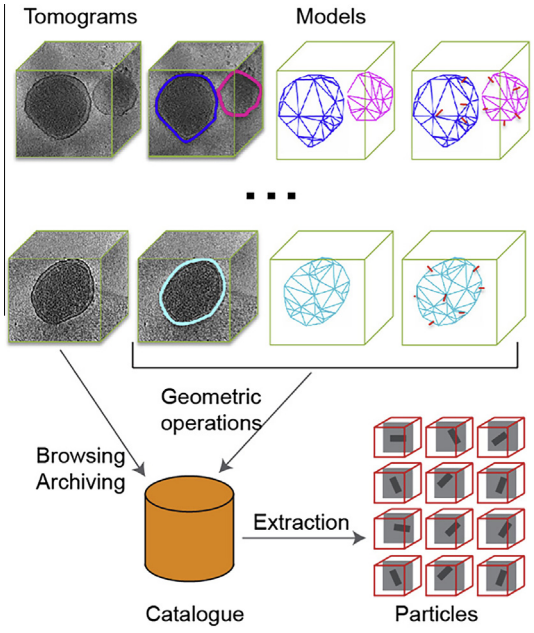

Fig. 1. The *Catalogue* database. Sketch of the STA workflow driven by *Dynamo Catalogue*. The raw data delivered by the microscope are several tilt series of projections. A reconstruction is performed for each series (top left). The resulting tomograms are then analyzed within *Dynamo Catalogue* through manual or semiautomatic methods, using models that ultimately define the approximate positions and orientations of particles. The *Catalogue* database records all the parameters, allowing the particles to be extracted and formatted adequately for subtomogram averaging in *Dynamo* or other packages whenever required.

Please cite this article in press as: Castaño-Díez, D., et al. *Dynamo Catalogue*: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. J. Struct. Biol. (2016), http://dx.doi.org/10.1016/j.jsb.2016.06.005