Journal of Structural Biology 189 (2015) 114–122

Contents lists available at ScienceDirect

Journal of Structural Biology

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

# Semi-automated selection of cryo-EM particles in RELION-1.3

Sjors H.W. Scheres

*MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK*

ARTICLE INFO

*Article history:*
Received 28 October 2014
Received in revised form 20 November 2014
Accepted 30 November 2014
Available online 6 December 2014

*Keywords:*
Electron cryo-microscopy
Single-particle analysis
Automated particle picking

ABSTRACT

The selection of particles suitable for high-resolution cryo-EM structure determination from noisy micrographs may represent a tedious and time-consuming step. Here, a semi-automated particle selection procedure is presented that has been implemented within the open-source software RELION. At the heart of the procedure lies a fully CTF-corrected template-based picking algorithm, which is supplemented by a fast sorting algorithm and reference-free 2D class averaging to remove false positives. With only limited user-interaction, the proposed procedure yields results that are comparable to manual particle selection. Together with an improved graphical user interface, these developments further contribute to turning RELION from a stand-alone refinement program into a convenient image processing pipeline for the entire single-particle approach.
© 2014 The Author. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Recent advances in electron cryo-microscopy (cryo-EM) single-particle analysis have made it possible to obtain near-atomic resolution structures for a much wider range of specimens and from much fewer particles than before. Previously, cryo-EM maps with sufficient detail to see amino acid side chains could only be obtained for hundreds of thousands asymmetric units of large icosahedral viruses (Grigorieff and Harrison, 2011). However, last year a ribosome reconstruction with details of around 4 Å was reported from 35 thousand (asymmetric) particles (Bai et al., 2013), and a 20S proteasome structure to 3.3 Å was reported from 1.8 million asymmetric units (Li et al., 2013). More recently, a 3.2 Å map for the yeast mitochondrial large ribosomal subunit was reported from 47 thousand particles (Amunts et al., 2014), a 3.4 Å structure of the $F_{420}$-reducing [NiFe] hydrogenase from 319 thousand asymmetric units (Allegretti et al., 2014), and a 3.4 Å structure of the TRPV1 ion channel from 142 thousand asymmetric units (Cao et al., 2013).

Two developments play an important role in these advances. The first is the development of direct-electron detectors, which are much more efficient at detecting electrons than conventionally used photographic film or charged-coupled devices (CCDs) (McMullan et al., 2009). The higher detection quantum efficiency (DQE) of the new detectors yield images with much improved signal-to-noise ratios (SNRs). This has a "double effect" on the resolution of 3D reconstructions: not only need one average over fewer particles to obtain a given resolution, but one can also align and classify each particle better, so that reconstructions are blurred to a much smaller extent than before. This then also relates to the second development: that of powerful new image processing algorithms. In particular, unsupervised image classification algorithms may be used to separate projections of distinct 3D structures, so that relatively impure or structurally heterogeneous samples may still lead to high-resolution structure determination, e.g. see (Fernndez et al., 2013; Voorhees et al., 2014). Moreover, as the new detectors are also very fast, one can now record multiple images during irradiation of the sample in the microscope. Since interactions with the incoming electrons cause movement of the sample, movie processing algorithms that correct for these beam-induced movements may further increase resolution (Campbell et al., 2012; Bai et al., 2013; Li et al., 2013). These developments have opened up the possibility to apply high-resolution cryo-EM structure determination to a much wider range of samples than before, which will attract many new researchers to this exciting field.

Together with increased interest in the technique, the call for high-throughput, easy-of-use and automation will also grow. One step in the data processing pipeline of high-resolution structure determination that may take considerable amounts of time and user-input is the selection of particles that are suitable for 3D reconstruction. In the past this process was typically done manually by the researcher, who would sit in front of a computer screen and click on each individual particle. Over the last 15 years, many algorithms to automate this often tedious procedure have been proposed, see (Nicholson and Glaeser, 2001; Zhu et al., 2004) for earlier reviews. More recently, implementations of automated particle picking algorithms were made available in EMAN2 (Tang et al., 2007), SIGNATURE (Chen and Grigorieff, 2007), DOGPICKER

*E-mail address:* scheres@mrc-lmb.cam.ac.uk

http://dx.doi.org/10.1016/j.jsb.2014.11.010
1047-8477/© 2014 The Author. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(Voss et al., 2009), XMIPP (Sorzano et al., 2009), and ARACHNID (Langlois et al., 2014) among others. These approaches may broadly be divided into two categories: feature-based and template-based approaches. In the feature-based approaches, different characteristics of the particles are expressed in some numerical manner (features) and features calculated from local areas in the micrographs are compared to a set of expected features. In the template-based approaches, images that express the expectation how the particles look like are correlated against the micrographs, often using fast Fourier-transform (FFT) accelerated algorithms (Roseman, 2003, 2004). The distinction between the two types of approaches is not always clear, as sometimes expected features are calculated from template images themselves. In general, template-based approaches introduce a higher degree of prior information into the picking process than feature-based approaches, which may be both an advantage and a disadvantage. The advantage of using more prior information is that it allows to detect weaker signals. However, the high levels of noise in the micrographs also make the picking task extremely prone to reference bias. Thereby, relying heavier on prior information becomes a disadvantage in cases where this information is incorrect.

This paper describes recent developments in the RELION software (Scheres, 2012a,b) that are centred around a new template-based particle picking algorithm. The choice for a template-based approach was motivated by its larger potential to select particles from noisy data. The workflow proposed is a semi-automated one. The researcher manually picks particles from a low number of micrographs; uses reference-free 2D class averaging inside RELION to calculate average images of these particles; performs template-based automated particle picking with those class averages on all micrographs; and then relies on a new sorting algorithm and further 2D class averaging to remove false positive from the data. To facilitate this process an improved graphical user interface (GUI) was also implemented. Whereas RELION was originally proposed as a stand-alone refinement program, these developments continue the evolution of RELION into a software package that provides a convenient pipeline for most of the single particle analysis tasks. Together with an improved movie-processing approach to correct for beam-induced motion in samples of relatively small particles (Scheres, 2014), the developments presented here represent the main improvements in the latest (1.3) release of RELION.

## 2. Approach

### 2.1. Particle picking

The template-based particle picking algorithm proposed employs an additive model with white Gaussian noise in real-space. A micrograph $\mathbf{X}$ that contains $N$ individual particles $i$ at coordinates $\vec{t}_i$ in the micrograph is described as follows (also see Fig. 1A):

$$\mathbf{X}(\vec{r}) = \mu(\vec{r}) + \sigma(\vec{r}) * \left\{ \mathbf{N}(\vec{r}) + \sum_{i=1}^{N} \mathbf{A}_{k_i}^{\phi_i}(\vec{r} - \vec{t}_i) \right\}, \quad (1)$$

where:

- $\mathbf{X}(\vec{r})$ is the micrograph, i.e. a two-dimensional image that was recorded in the electron microscope, and $r$ describes the two-dimensional position in that image.
- $\mathbf{N}(\vec{r})$ is an image of independent (or white) Gaussian noise with mean zero and standard deviation one.
- $\mu(\vec{r})$ and $\sigma(\vec{r})$ are position-dependent additive and multiplicative *normalisation* factors that bring the recorded noise levels in the micrograph to mean zero and standard deviation one. Variations in $\mu$ and $\sigma$ with position typically describe experimental variations in ice thickness, electron dose, *etc.*

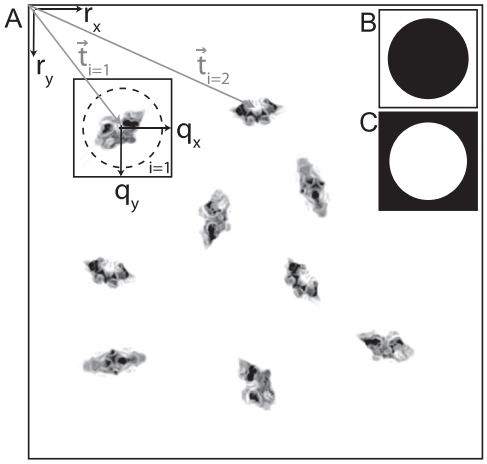


**Fig.1.** Schematic representation of the data model. (A) Representation of a micrograph, with coordinate vectors $r = (r_x, r_y)$ inside the micrograph, and coordinate vectors $q = (q_x, q_y)$ inside each particle image. Vectors $t_i = (t_x, t_y)$ place the $i$th particle inside the micrograph with an unknown in-plane rotation $\phi_i$ with respect to a common frame of reference. Inset (B), mask $\mathbf{M_o}$ which is used for normalisation of the particle images: average and standard deviation of the background pixels are calculated in the white area of this mask. Inset (C), mask $\mathbf{M_i}$ which is used for the particle sorting algorithm: all statistics on the difference images between each particle and its corresponding template are calculated in the white area of this mask.

- $\mathbf{A}_{k_i}^{\phi_i}$ is one of $K$ known, two-dimensional *template* images $\mathbf{A}_k$ with internal positions $\vec{q}$. Typically, the template images are much smaller than the micrograph (in the summation above, the template image is zero outside the defined box size). Therefore, for any given $\vec{r}$ and template $\mathbf{A}_{k_i}$ at position $\vec{t}_i$, the internal position will be $\vec{q} = \vec{r} - \vec{t}_i$. The $K$ different template images may describe projections in different directions of the same molecule, or they may describe projections of different molecules; $k_i$ describes which of the template images corresponds to the $i$th particle; and $\phi_i$ describes the relative in-plane rotation between the particle and that template image.

Given $\mathbf{X}(\vec{r})$ and $K$ template images $\mathbf{A}_k$, the task at hand is to identify all $N$ combinations of $\vec{t}_i$, $\phi_i$ and $k_i$. Based on positive experiences with maximum-likelihood approaches, e.g. see (Scheres et al., 2007; Scheres, 2012a), the choice was made to implement a probability-based similarity metric for this task. The assumption of Gaussian noise in Eq. (1) naturally leads to a Gaussian similarity measure. Unlike the cross-correlation coefficient, as used for example in the template-based picking program findEM (Roseman, 2003, 2004), the squared difference term inside the Gaussian metric is not invariant to multiplication with or addition of a constant. This means that one needs to account for the varying intensity levels in the recorded micrographs, and one needs to determine the normalisation factors $\mu(\vec{r})$ and $\sigma(\vec{r})$ to bring all particles on the same intensity level.

Upon extraction of individual particles from the micrographs, RELION relies on a normalisation procedure that uses a circle (with a user-defined radius $R$, see Fig. 1B) to divide each extracted particle image in a background area (outside the circle) and a particle area (inside the same circle) (Sorzano et al., 2004). By subtracting the average value of the pixels in the background area from the entire particle image, and subsequently dividing the entire image by the standard deviation of the pixels in the background area, noise levels with zero-mean and unity-standard deviation are obtained for all particles, independent of variations in ice thickness, exposure or other uncontrolled experimental factors.

Download English Version:

https://daneshyari.com/en/article/5913940

Download Persian Version:

https://daneshyari.com/article/5913940

Daneshyari.com