Journal of Structural Biology 192 (2015) 235–244

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Single-particle cryoEM analysis at near-atomic resolution from several thousand asymmetric subunits

Dario Oliveira Passos, Dmitry Lyumkis *

Laboratory of Genetics and Helmsley Center for Genomic Medicine, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, United States

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form 29 September 2015
Accepted 1 October 2015
Available online 22 October 2015

Keywords:
Single-particle analysis
Ribosome
Direct detectors
Over-fitting

ABSTRACT

A single-particle cryoEM reconstruction of the large ribosomal subunit from *Saccharomyces cerevisiae* was obtained from a dataset of ~75,000 particles. The gold-standard and frequency-limited approaches to single-particle refinement were each independently used to determine orientation parameters for the final reconstruction. Both approaches showed similar resolution curves and nominal resolution values for the 60S dataset, estimated at 2.9 Å. The amount of over-fitting present during frequency-limited refinement was quantitatively analyzed using the high-resolution phase-randomization test, and the results showed no apparent over-fitting. The number of asymmetric subunits required to reach specific resolutions was subsequently analyzed by refining subsets of the data in an *ab initio* manner. With our data collection and processing strategies, sub-nanometer resolution was obtained with ~200 asymmetric subunits (or, equivalently for the ribosomal subunit, particles). Resolutions of 5.6 Å, 4.5 Å, and 3.8 Å were reached with ~1000, ~1600, and ~5000 asymmetric subunits, respectively. At these resolutions, one would expect to detect alpha-helical pitch, separation of beta-strands, and separation of Cα atoms, respectively. Using this map, together with strategies for *ab initio* model building and model refinement, we built a region of the ribosomal protein eL6, which was missing in previous models of the yeast ribosome. The relevance for more routine high-resolution structure determination is discussed.

1. Introduction

The last several years have witnessed dramatic improvements to the technological infrastructure revolving around single-particle cryo-electron microscopy (cryoEM). The advent of direct detectors has revolutionized the field (Bai et al., 2013; Campbell et al., 2012; Li et al., 2013). Using state-of-the-art instrumentation, near atomic resolution reconstructions of megadalton-sized complexes are now becoming relatively routine (Bai et al., 2015), at least in the most structurally homogeneous and stable regions, and an increased number of small particles are becoming amenable to the technology (Cao et al., 2013; Liang et al., 2015; Liao et al., 2013; Lu et al., 2014). With continuous improvements to the imaging and image processing protocols, it is becoming increasingly more feasible to analyze particles that are smaller and more heterogeneous (Kühlbrandt, 2014). Nevertheless, practical possibilities have not reached theoretical predictions made several decades ago (Henderson, 1995), and many areas of improvement still exist.

Several aspects make single-particle cryoEM a particularly attractive technique for structural biologists studying macromolecules and macromolecular assemblies. These include, but are not limited to, the absence of a need to crystallize the sample, the requirement of less sample for structural analysis (often an order of magnitude), and the ability to decipher structurally heterogeneous populations within the sample (Cheng, 2015). A typical disadvantage is that lower bounds on particle size exist, currently limiting the technique to relatively large (>100 kDa) macromolecules and macromolecular assemblies (often referred to as "particles" in single-particle analysis). The reason for this is because cryoEM images are noisy (Baxter et al., 2009), and a minimal mass is therefore required to produce sufficient scattering contrast for the alignment of images of individual particles (Henderson, 1995). Thus, different sized particles will require varying amounts of data to reach a given resolution. Other factors that may affect the outcome include particle purity (and/or state of heterogeneity) and distribution, imaging conditions, processing procedures, among other reasons (Rosenthal and Henderson, 2003; Saad et al., 2001; Stagg et al., 2008, 2014). An understanding of how much data is required to reach defined resolutions in single-particle cryoEM is of general interest when trying to extract the maximum amount of information from a given dataset.

* Corresponding author.
E-mail address: dlyumkis@salk.edu (D. Lyumkis).

http://dx.doi.org/10.1016/j.jsb.2015.10.002
1047-8477/

In this work, we determined the cryoEM structure of the large ribosomal subunit from *Saccharomyces cerevisiae*. Both the gold-standard and frequency-limited approaches to single-particle refinement, as implemented within the Frealign package (Grigorieff, 2007), are utilized for structure determination, and both provide nearly identical resolution curves and nominal resolution values. Through implementation of the high-resolution noise substitution test (Chen et al., 2013), we provide additional support that the resulting reconstruction is not biased by overfitting. Using the dataset as a test sample, we then systematically analyze the amount of asymmetric subunits that are required to reach different resolutions by refining subsets of the data in an *ab initio* manner and show that near-atomic resolution reconstructions can be obtained with as few as one to two thousand individual asymmetric subunits; because the ribosomal subunit is asymmetric, this corresponds to an equivalent amount of single particles in the data. Last, we use recently described approaches implemented for model building and refinement in Rosetta to provide atomic-level information for a portion of a ribosomal protein that was previously omitted in crystallographic refinement. We expect these results to provide a relevant resource for future cryoEM studies.

2. Materials and methods

2.1. Specimen purification and preparation for cryoEM

Sample containing 60S ribosomal subunits and associated cofactors was purified similar to what has been previously described (Lyumkis et al., 2014). The protein Ltn1$^{\Delta R}$ in our case was tagged with 5× Flag. 3 μL of sample was applied onto a freshly plasma treated (6s, Gatan Solarus plasma cleaner) holey carbon C-flat grid (Protochips, Inc.), allowing the sample to adsorb for 30-s, and then plunge-freezing into liquid ethane using a manual cryo-plunger in an ambient environment of 4 °C.

2.2. Cryo-EM data collection

Data were acquired using the Leginon software (Suloway et al., 2005) installed on an FEI Titan Krios electron microscope operating at 300 kV, with a dose of 25 e−/Å^2 and an estimated underfocus ranging from 0.5–2.5 μm (distributed in an approximately Gaussian manner and centered at 1.75 ± 0.38 μm). The dose was fractionated over 25 raw frames collected over a 5-s exposure time (200 ms per frame) on the Gatan K2 Summit direct detection device, with each frame receiving a dose of ~8.5 e$^-$/pixel/s. We used the super-resolution imaging mode for recording movies. 1833 such movies were automatically collected and recorded at a nominal magnification of 22,500, corresponding to a super-resolution pixel size of 0.655 Å at the specimen level (physical pixel size of 1.31 Å). The individual super-resolution frames were aligned using a GPU-enabled frame-alignment program (Li et al., 2013). After aligning and summing the frames, the micrographs were binned by 2 in Fourier space to a pixel size of 1.31 Å. Summed and Fourier-binned micrographs were used for all subsequent image processing operations, with the exception of the final model refinement.

2.3. Pre-processing of single-particles

All pre-processing operations described below were performed within the Appion pipeline (Lander et al., 2009). Individual programs incorporated within Appion and used for the data analysis are cited below, as appropriate. The CTF for all micrographs was estimated using the CTFFind3 package (Mindell and Grigorieff, 2003). All micrographs were also manually masked with the "manual masking" tool in Appion to remove regions that were over carbon and/or containing large amounts of aggregated particles. A set of 16 class averages was generated after manually selecting ~300 representative particles from the micrographs, creating a small stack, and running the CL2D algorithm (Sorzano et al., 2010). Once initial 2D templates were obtained, particles were then selected from the full dataset using the template-picking strategy implemented within FindEM (Roseman, 2004). A phase-flipped, contrast-inverted, and 4× binned stack was created from these picks using a box size of 64 pixels and pixel size of 5.24 Å. The stack was subjected to reference-free 2D alignment and clustering using CL2D (Sorzano et al., 2010) to obtain 256 2D classes. The classes were inspected visually, and any classes that did not contain the features of a 60S ribosomal subunit, or those that did not display identifiable features altogether, were discarded. The remaining stack contained 99,167 particles. An *ab initio* model was generated from reference-free 2D classes obtained from the 4× binned particles using OptiMod (Lyumkis et al., 2013c), which resulted in a low-resolution map of the 60S ribosomal subunit. This map was low-pass filtered to 60 Å prior to the assignment of particle orientations. After assigning initial orientations to this 60S map (see below), 3D classification in Frealign using 10 classes (Lyumkis et al., 2013a) helped identify an additional ~25k particles, which did not produce high-resolution reconstructions. After discarding the above particles, 75,653 particles remained for the final high-resolution refinements.

2.4. High-resolution frequency-limited refinement of single-particles

For the frequency-limited refinement, a single stack was refined against the initial model obtained in the pre-processing step above. To obtain initial orientations, the 60S volume was used as input to the projection-matching protocol implemented within the Xmipp processing package (Sorzano et al., 2004). Data was low-pass filtered to 30 Å, and ten iterations of projection-matching were performed. The above steps produced an input stack that contained initial orientation parameters for high-resolution refinement in Frealign (Grigorieff, 2007). An unbinned stack that had not been phase-flipped or contrast-inverted was used as input to the final orientation parameter refinement. We performed five initial cycles of single-model refinement using the remaining 75,653 particles. The frequency range used for assigning particle orientations was set to between 250 → 20 Å for the first cycle, 250→10 Å for cycles 2–3, and 250 → 8 Å for cycles 4–5. Subsequently, for the next five cycles, the high-resolution frequency of refinement was dropped by 1 Å per cycle, to 250 → 7 Å, 250 → 6 Å, 250 → 5 Å, 250 → 4 Å, and 250 → 3 Å, for cycles 6, 7, 8, 9, and 10, respectively. To assess resolution at each cycle of refinement, a soft-edged mask that adopts the shape of the 60S subunit (raised cosine edge with a width of 6 pixels) was created and applied to the half-maps using the 'postprocess' command in Relion (Scheres, 2012). After the first cycle of refinement, the resolution of the reconstruction had already reached ~5 Å. After 3 cycles of refinement, the resolution of the reconstruction had reached ~3.0 Å. The final resolution of 2.9 Å was reached after 10 cycles of refinement (Fig. 2(C)).

2.5. Gold-standard refinement of single-particles

An identical procedure to that described above was employed, except that the 75,653-particle dataset was split randomly into two subsets prior to the determination of any orientation parameters, including the initial orientation search in Xmipp. The orientations for each particle within each 37,826-particle subset were assigned and refined independently of one another, except for a