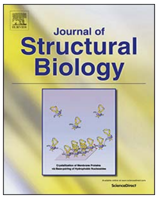
ARTICLE IN PRESS

Journal of Structural Biology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Biology

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



Exploring applications of crowdsourcing to cryo-EM

Jacob Bruggemann*, Gabriel C. Lander, Andrew I. Su

[a] *Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037 USA*

ARTICLE INFO

Keywords:
Structural biology
Cryo-EM
Crowdsourcing
Computation
Data processing
Single-particle analysis

ABSTRACT

Extraction of particles from cryo-electron microscopy (cryo-EM) micrographs is a crucial step in processing single-particle datasets. Although algorithms have been developed for automatic particle picking, these algorithms generally rely on two-dimensional templates for particle identification, which may exhibit biases that can propagate artifacts through the reconstruction pipeline. Manual picking is viewed as a gold-standard solution for particle selection, but it is too time-consuming to perform on data sets of thousands of images. In recent years, crowdsourcing has proven effective at leveraging the open web to manually curate datasets. In particular, citizen science projects such as Galaxy Zoo have shown the power of appealing to users' scientific interests to process enormous amounts of data. To this end, we explored the possible applications of crowdsourcing in cryo-EM particle picking, presenting a variety of novel experiments including the production of a fully annotated particle set from untrained citizen scientists. We show the possibilities and limitations of crowdsourcing particle selection tasks, and explore further options for crowdsourcing cryo-EM data processing.

1. Introduction

In the past several years cryo-electron microscopy (cryo-EM) has become a powerful tool for elucidating the structures of macromolecular complexes to near-atomic resolution, and has been effectively used to solve structures of membrane-bound and non-rigid proteins that are difficult to crystallize. Handling low signal-to-noise ratio cryo-EM data necessitates processing large amounts of data, involving thousands of individual micrographs each containing hundreds of particles. A crucial, early step in cryo-EM processing is the selection of individual protein particles from EM micrographs to be used in generating a 3D reconstruction. In the past, particles were hand-picked by a researcher after data collection, but since cryo-EM datasets can now consist of thousands of micrographs and hundreds of thousands of particles, manual picking has become viewed as an unnecessarily banal and time-consuming task for cryo-EM researchers (Scheres, 2015).

As a result, many algorithms have been developed to automate particle picking and reduce the time required for this crucial step in EM processing. Popular methods either identify features common to particles, such as particle size with DoG Picker, or use supplied templates to identify similar-looking subsections of a micrograph (Voss et al., 2009). Automatic methods are limited, however, in their ability to distinguish noise and contaminants from legitimate particles, and will sometimes misplace the center of particles in cases where they are closely packed. Inaccuracies in the collection of particle data can disrupt processing; in the challenging reconstruction of the HIV-1 envelope glycoprotein complex by Liao et al., Henderson noted that a lack of validation of the particle set picked via a template method begat a set of particles containing significant white noise, which nonetheless sufficiently matched the templates provided (Liao et al., 2013; Henderson, 2013). Indeed, manual selection by a trained microscopist is still viewed as an ideal strategy in many cases, especially when templates are not available or the protein particles are ill-defined in the micrographs. Implementing manual selection necessitates an immense amount of time and effort for this single processing step; as an example, Fan et al. manually boxed out 156,805 particles from 3743 micrographs when determining the structure of the InsP$_3$R ion channel. The time needed to produce a manually-picked set precludes its adoption as a regular procedure for particle picking, and a method that reduced the temporal investment could prove valuable for researchers. In addition, scientists seeking particular idealized structures can consciously or subconsciously impart their own biases into manual picking, preferring certain angular views of the particle or omitting subsets of particles that do not exhibit anticipated features (Cheng et al., 2015).

This work examines an increasingly popular method of data processing, crowdsourcing. A term coined in 2006, crowdsourcing opens a task normally assigned to a specific worker to a wider, more generalized userbase (Good and Su, 2013). In recent years, crowdsourcing initiatives have come to rely on the ability of the internet to quickly disseminate data and recruit users to perform the necessary processing.

There are many approaches to crowdsourcing, including scientific games (e.g., Foldit, Eterna) and paid microtask services (e.g., Amazon

---

* Corresponding author.
E-mail addresses: jbrugg@scripps.edu (J. Bruggemann), glander@scripps.edu (G.C. Lander), asu@scripps.edu (A.I. Su).

https://doi.org/10.1016/j.jsb.2018.02.006
Received 15 November 2017; Received in revised form 19 February 2018; Accepted 22 February 2018

Please cite this article as: Bruggemann, J., Journal of Structural Biology (2018), https://doi.org/10.1016/j.jsb.2018.02.006

Mechanical Turk, Crowdflower). Particularly intriguing is the emergence of 'citizen science' projects, which rely on community engagement and scientific intrigue to attract users to an otherwise menial task. Citizen science has proven extremely successful, with the project 'Galaxy Zoo' classifying over 1 million images from more than 100,000 users over nine months (Lintott et al., 2008). In this paper, we present and analyze the results from a citizen science project 'Microscopy Masters', which focused on crowdsourcing particle picking from single-particle cryo-EM micrographs. We examine the efficacy of crowdsourcing particle picking to lightly trained workers when compared to trained electron microscopists, and show that particle sets derived through crowdsourcing can yield robust and reliable 2D class averages and 3D reconstructions. The method presented here is shown as not only a viable time-saving option for datasets that confound automatic pickers, but also shows promise for future applications of crowdsourcing to cryo-EM data processing.

2. Results

2.1. Production of gold standard

A 'gold standard' or 'ground truth' for evaluating annotated subjects is crucial for beginning any classification study. In the case of algorithmic particle picking, evaluation is typically performed relative to a set of manually picked micrographs. Although manually picked datasets are available from previous studies examining particle picking, they generally contain a small number of images and are only annotated by a single individual (Scheres, 2015). In order to create a richer gold standard for evaluating our crowdsourcing protocol, micrographs were chosen from a single-particle cryo-EM dataset of the 26S proteasome lid complex (Dambacher et al., 2016). Out of the 3,446 micrographs used in the published refinement, 190 were marked by at least two randomly-assigned cryo-EM experts, with a total of nine contributing participants.

In addition, intra-expert agreement was measured by requiring each expert to mark five randomly chosen images twice. The complete union of all marks by all experts totaled to 13,028 particles and was used as the ground truth for all following accuracy measurements in this paper.

Agreement between two annotations was calculated using the Jaccard index, defined as the proportion of the size of the intersection of the particles picked in two annotations to the size of the union. Intra-expert agreement was found to be surprisingly low and only slightly higher than inter-expert agreement, intra-agreement between all experts averaged 0.56 and inter-agreement averaged 0.45. This indicates that less than three quarters of particles picked by a single individual are picked again on reannotation by the same annotator on the same image. Among those annotators who completed all assigned micrographs, agreement was consistently higher among intra-expert comparisons (Table 1).

Table 1
Intra-expert and inter-expert agreement calculated using Jaccard index for experts who created the gold dataset. Of the nine experts who participated, one did not complete all assigned micrographs and so was not included in this table.

User	Inter	Intra
Expert 1	0.45	0.64
Expert 2	0.42	0.46
Expert 3	0.54	0.65
Expert 4	0.40	0.51
Expert 5	0.44	0.64
Expert 6	0.44	0.58
Expert 7	0.45	0.61
Expert 8	0.48	0.51

2.2. Initial testing

A chief concern for crowdsourcing, especially citizen science, is building a reliable userbase, either through accessing pre-existing groups of users or attracting users through social media and community engagement. Since we desired to annotate a large, fully manually-picked particle set, we hosted our experiment on an established crowdsourcing platform, Panoptes, a Zooniverse-run initiative for citizen science projects.

Initial testing for the crowdsourcing system was performed by paid workers recruited through Amazon Mechanical Turk (AMT). Workers were recruited, trained, and paid through AMT, while Panoptes hosted the particle selection tasks and stored the results (Fig. 1).

Testing produced 16,562 particles chosen by 42 unique workers using the same set of 190 images in the gold standard. Based off of feedback from AMT workers, additional instructions, shown in (Fig. S2), were added to the picking interface.

Importantly, this initial testing was used to determine the optimal number of people to assign to each image, as well as to establish a voting mechanism. To this end, at least 10 workers annotated each micrograph and accuracy statistics were derived for randomized subsets of those workers using various voting thresholds (Fig. 2a). Limited returns after five annotators at union led to the choosing of this threshold as optimal; in all subsequent experiments, each image was shown to five annotators and the “crowd” output was defined as the union of users' annotations.

2.3. Zooniverse

After testing in Amazon Mechanical Turk, the project, dubbed Microscopy Masters, was launched on Panoptes in March 2016. A total dataset of 209,696 particle picks was produced over a year from 3,446 micrographs, with 2,108 unique volunteers. The parameters established in our initial testing were utilized, with each image being classified by five different users and the 'voting threshold' set to one, meaning the total union of all classifications performed on an image were used to generate the final data set of picks. For individual users, we observed a marked decrease in F-score and recall in the Zooniverse set, as shown in Fig. 2, which we attributed to differing incentives between paid testing on AMT and unpaid volunteers on Zooniverse. In particular, the number of particles selected in each image by Zooniverse volunteers is highly variable; a peak at zero in the distribution of particles picked per-user per-image resulted in a corresponding peak at zero for recall and F-score, as well as a peak at one for precision (Fig. 2b). Association of low recall and low-cardinality annotations implied a body of “low-effort” annotations, where a user did not fully complete the image before submission. Aggregation of the five user annotations per image mitigated the low individual accuracy, yielding an average aggregate F-score comparable to that of the AMT-annotated data, as shown in (Fig. 2d).

Average agreement between the voted crowd annotations and individual expert annotations was found to be slightly less than inter-expert agreement, with the mean for inter-expert agreement at 0.45 and mean between the crowd and experts at 0.40 (Fig. 3).

2.4. Reconstruction

Refinement of the proteasome lid complex structure was performed using the crowdsourced dataset collected through Zooniverse, as well as the dataset used in Dambacher et al., which was picked using a template-based method (Dambacher et al., 2016). Since the resolution of a refined single-particle reconstruction generally correlates with the number of particles in the dataset, a reconstruction was also generated using a random subset of the template-picked particles with the same cardinality as the crowdsourced data, called the 'normalized template' set (Cheng, 2015). Particle stacks were extracted from the micrographs,