

Processing Structurally Heterogeneous Cryo-EM Data Using Atomic Models

Gabriella Reggiano¹ and Frank DiMaio¹

¹ Department of Biochemistry, University of Washington, Seattle, WA, USA.

Macromolecules often must undergo both large and small conformational changes to perform their biological functions. This dynamic behavior poses a challenge to single particle cryo-electron microscopy (cryo-EM) as it can be difficult to isolate a single conformation within a sample. The current favored method to address conformational heterogeneity is using three-dimensional classification, to sort the images into a user-defined number of classes based on the three-dimensional volume they reconstruct. Three-dimensional classification has difficulty identifying subtly different states because particles are often missorted; an image is correctly assigned only if the difference between the classes is large enough to overcome the noise caused by averaging projections from different three-dimensional structures [1].

More complex analyses of conformational heterogeneity are also limited. In multi-body refinement, the user defines a discrete number of independently moving bodies and then performs iterative signal subtraction to simultaneously improve the resolution of the domains [2]. Because the different domains are treated as rigid bodies moving independently of each other, the method is not able to capture changes that may occur for the amino acids at the interface between bodies. In addition, multi-body refinement is only amenable to large bodies (100kD). The movement of small domains or small regions of the protein cannot be resolved as they do not provide enough signal to stabilize the alignment [3]. A more complex analysis of continuous motion is manifold embedding, which can deduce the lowest energy trajectory of the protein and create three-dimensional movies by pulling images associated with a conformation. However, it doesn't address the issue of map reconstruction and only interpolates in density space [4,5].

We are developing a method that utilizes atomic models to guide particle sorting and address continuous motion of biological macromolecules (Fig. 1). Density maps generated from atomic models contain less noise than averages of projections of a heterogeneous sample so particles can be sorted more accurately and more precisely, producing classes with more subtle changes in conformation. By using atomic models, we are not limited by the size of the region of interest, nor are we treating the system as rigid bodies. Information at the interface between domains will not be lost. In cases of continuous motion, atomic models can be used to cross-average density maps and pool particles into discrete bins.

Starting with an initial model and density map, we generate hypothetical conformational states that the protein may adopt in two different ways. In the first approach, we generate conformational states using existing methods within Rosetta, such as relaxing the protein in the density map or rebuilding the protein with fragment-based methods [6,7]. Both techniques generate ensembles of atomic models in areas with low resolution density, which can then be clustered based on RMS. In the second method, we use the normal modes to generate different atomic models and select for models that have good correlation with the consensus density map. The normal modes, vectors of motion where neighboring alpha carbons move at the same frequency, have been shown to describe large scale conformational changes for many proteins [8]. As a result, this method may better approximate the trajectory of the protein. After calculating possible conformations the protein may adopt, we generate density maps of those states and feed them to three-dimensional classification in existing image processing suites. In this way, we can

guide three-dimensional classification to choose conformational states that are most representative of the data. The classes can then be eliminated or cross-averaged using atomic models and particles can be re-sorted. Individual classes are then sent to refinement, and more accurate models are built into the refined classes. We then iterate through this process, generating conformational states and density maps from these new models and re-sort particles.

References:

- [1] C Sorzano et al., *J Struct Bio* **171** (2010), p. 197.
- [2] T Nakane et al., *eLife* **7** (2018), e36861.
- [3] C Sorzano et al., *Acta Cryst* **F75** (2019), p. 19.
- [4] A Dashti et al., *PNAS* **111** (2014), p.492.
- [5] J Frank and A Ourmazd, *Methods* **100** (2016), p.61.
- [6] F DiMaio et al., *J Mol Bio* **392** (2009), p.181.
- [7] R Wang et al., *eLife* **5** (2016), e17219.
- [8] W Krebs et al., *PROTEINS* **48** (2002), p.682.

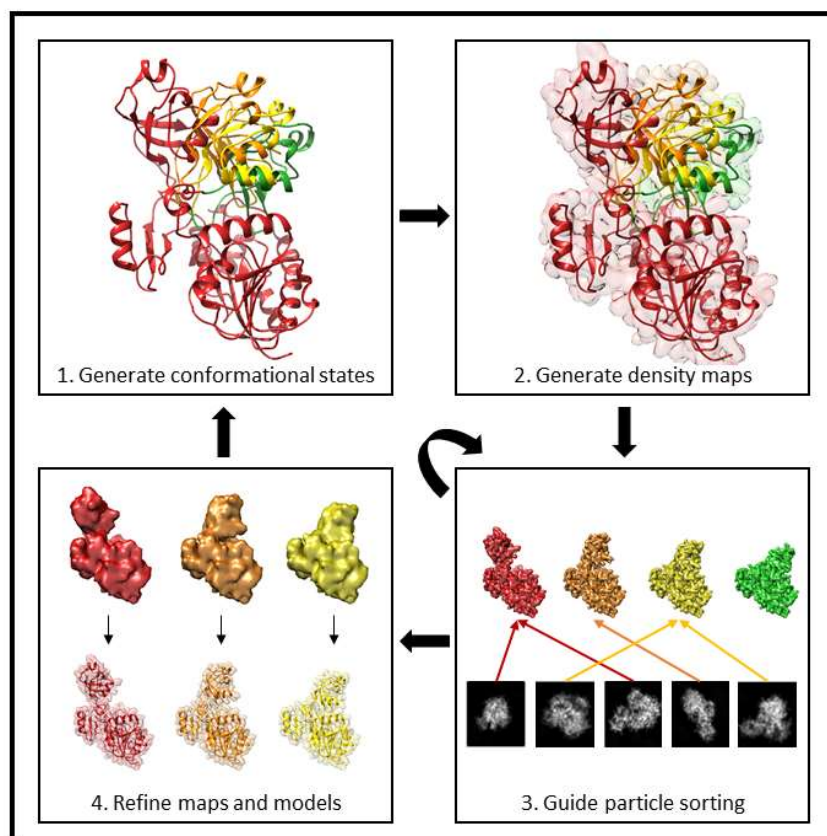


Figure 1. Overview of method to process cryo-EM data with atomic models. The method involves four steps: (1) generate possible conformational states (yellow, orange, green) from a density map and initial model (red), (2) generate density maps from those states, (3) feed these states to three-dimensional classification to guide particle sorting, and (4) select classes to send on to refinement. Step 3 can be repeated, eliminating or cross averaging classes and resorting particles before sending classes on to refinement. After refining maps and fitting models into the new maps, we generate new conformational states and iterate through the process.