Building Ground Truth Benchmarks of Structural Heterogeneity for Cryo-EM

Dr. Andrew V. Grassetti[1], Laurel F. Kinman[1], Dr. Joseph H. Davis[1]
[1]Massachusetts Institute of Technology
setti@mit.edu

Rapid advances in cryo-EM homogeneous reconstruction algorithms have contributed significantly to the so-called "resolution revolution", and were enabled by the existence of well-characterized benchmark datasets. In contrast, despite the recent surge in heterogeneous classification and reconstruction algorithms, which have the potential to contribute substantial biological insight, analogous experimental benchmark datasets are lacking. By leveraging the targeting power of the RNA-guided DNA endonuclease Cas9, we have generated a series of 13 related macromolecular complexes with defined structural features to serve as such benchmarks. Specifically, in each complex, catalytically inactivated Cas9 is bound to a unique DNA scaffold, differing only in the number of base pairs extending from the Cas9 active site. With this system, we have introduced defined structural changes as small as a ~600 Da base pair extension, or as large as a ~10 kDa DNA helix. To our knowledge, these data represent the only catalog of cryo-EM structures with prescribed, biochemically defined, and subtle variability specifically generated to benchmark and improve cryo-EM heterogeneity analysis software. Here we present our strategies for system design, complex formation and biochemical characterization, cryo-EM data collection, and image processing of each construct in the series. Additionally, we demonstrate the utility of the ground truth datasets by combining individual particle stacks in silico and challenging heterogeneous reconstruction algorithms to accurately analyze encoded heterogeneity either in aggregate or on a per-particle basis.