# Interrogating macromolecular complex assembly by systematically analyzing the composition of highly heterogeneous structural ensembles

**Laurel Kinman[1], Jingyu Sun[2], Joaquin Ortega[3], Joey Davis[4]**

*[1]Massachusetts Institute of Technology [2]Center for Structural Biology, McGill University, [3]Center for Structural Biology, McGill University, [4]MIT*

*lkinman@mit.edu*

Cryo-EM represents a unique and powerful opportunity to structurally characterize biomolecules at the single-particle level, and to draw biological insights from the heterogeneity observed within structural ensembles. Doing so, however, represents a significant computational challenge, and necessitates improved methods for studying extremely heterogeneous datasets. Here, we present an approach that combines our recently-published cryoDRGN method to reconstruct highly heterogeneous structural ensembles with a high-throughput compositional analysis that allows us to quantify the presence and absence of individual domains or whole proteins across hundreds-to-thousands of cryo-EM density maps. This analysis produces a highly interpretable representation of the compositional heterogeneity present within a dataset. Using this representation, we can identify cooperative and mutually-exclusive occupancy relationships between various subunits, extract subsets of particles for traditional high-resolution refinement, and define pathways of structural change including complex assembly. We have applied this approach to understand the role of a universally-conserved methyltransferase in biogenesis of the 30S ribosomal subunit. By comparing the structural ensembles observed in the presence and absence of this factor, we have uncovered that this factor performs a novel proof-reading role in ribosome assembly. In sum, this work establishes a framework for systematically interrogating compositionally heterogeneous structural ensembles produced by tools such as cryoDRGN, and it highlights the value of this framework in illuminating underlying biological mechanisms.