Hyper-Molecules: High Dimensional Maps of Molecular Conformations

R Lederman[1]
[1]The Department of Statistics and Data Science, Yale University
roy.lederman@yale.edu

One of the open problems in cryo-EM is mapping complex heterogeneity, such as continuous heterogeneity. We begin our discussion with the question what does it mean to recover a heterogeneous structure, compared to a homogeneous structure or several distinct structures? We introduce "hyper-molecules," a mathematical formulation which captures the continuum of states and the relationships between them, and a Bayesian framework for recovering these "hyper-molecules."

We present preliminary implementations and results, which demonstrates how the heterogeneous structures can be recovered from synthetic and real data, and discuss some of the practical challenges and solutions.

We discuss next steps in this work on a scalable framework which would map complex heterogeneous structures, and would optionally allow researchers to explicitly encode prior knowledge, such as general physical properties, and existing knowledge about the specific structure, when such knowledge is available, in order to resolve complex structures which could otherwise require unrealistic amounts of data.