Refining low-resolution Cryo-EM structures with Bayesian inference driven integration of multiscale simulations

Hagstrom, A.¹, Ma, H.¹, Bhowmik, D.¹, Stanley, C.B.,¹ Mitchell, J.C.,² O'Neill,³ H.M., Ramanathan, A.^{4*}

¹Computational Science and Engineering Division, ²Center for Molecular Biophysics, Biosciences Division, ³Neutron Scattering Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

⁴Data Sciences and Learning Division, Computing, Environment and Life Sciences Directorate, Argonne National Laboratory, Lemont, IL, USA. *Corresponding author: ramanathana@anl.gov

1. Abstract

Despite significant progress in structure determination techniques such as X-ray crystallography/nuclear magnetic resonance (NMR), our ability to obtain atomistic insights into complex biological phenomena remains challenging. Understanding the relationship between conformational flexibility and function is an issue that experimental structural biology continues to struggle with and flexible regions in proteins are often absent from the structures obtained using X-ray crystallography and NMR techniques. Techniques such as neutron/X-ray scattering and cryo-electron microscopy (cryo-EM) can potentially provide structural insight into flexible and heterogeneous biomolecular assemblies, but their inherent low resolution makes it difficult to elucidate atomistic/mechanistic details.

Although single particle cryo-EM is projected to displace traditional techniques for structure determination of large proteins and complexes ($\geq \sim 150$ kDa), flexible and heterogeneous biological assemblies present obvious challenges in approaching atomic resolution with this technique. On the other hand, small-angle scattering (SAS) with X-rays (SAXS) and neutrons (SANS) has been very informative for studying the structures of flexible systems such as multi-domain proteins and membrane proteins. When combined with computational modeling, it is possible to determine not only the ensemble of structures that reflect the conformational state of a protein but also the dynamical properties at relevant temporal scales of the system.

Currently, biologists must rely on 'piecemeal' data drawn from these diverse analysis techniques, which makes it an immense challenge to obtain comprehensive and quantitative insights into biomolecular function from cryo-EM techniques. We posit that multiscale molecular simulations integrated with novel statistical inference tools can serve as an effective 'glue' to fuse partial information from cryo-EM observations. However, integrating simulations with low resolution structure determination techniques is not straightforward since the ratio between the number of conformational degrees of freedom (tracked in MD) and the number of experimental constraints (derived from SAXS/SANS/cryoEM) remains unfavorably large.

In this paper, we present our preliminary results in developing and validating a multiscale molecular simulation framework that allows us to directly simulate Cryo-EM volumetric data using continuum approaches, namely using the fluctuating finite element analysis (FFEA) technique. The FFEA generated data is mapped onto atomistic resolution using traditional molecular dynamics (MD) simulations – while simultaneously using Bayesian methods to maximize the agreement between experimentally observed constraints and simulations generated constraints. We demonstrate our approach in simultaneously refining and simulating conformational ensembles of several benchmark protein systems for which cryo-EM experimental data is available at different resolutions. By simulating the protein at successive low resolutions of the cryo-EM maps, we show that our approach can successfully reconstruct observed classes of conformations in the cryo-EM data, while mapping the conformational transitions leading to the observed data. Together, our integrative modeling approach could be used for refining low resolution cryo-EM data by combining multiscale molecular simulations and Bayesian inference strategies.



Figure 1 Workflow for reconstructing atomistic ensembles from low resolution Cryo-EM experimental data, combining Bayesian inference with multiscale molecular simulations and small angle scattering datasets.

Figure 1 shows an outline of our workflow. We start with a cryo-EM map to instantiate the FFEA simulation; the simulation can be either set up at experimentally determined resolution or successively mapped to lower resolutions, in which case a coarse-grained simulation may result. Although our use of FFEA does not incorporate the full physical description of a protein (such as electrostatics, solvent interactions, etc.), it still possesses sufficient granularity to generate conformations that are viable. These conformations can be mapped using internal tools of FFEA to generate atomistic configurations of the protein system, which are then simulated using OpenMM toolkit, under explicit solvent conditions. Note that it is possible to transition between all-atom and continuum resolutions dynamically, depending on agreement between experimental and simulated data. The snapshots generated from the FFEA simulations are used to generate 2D images (which are representative of the cryo-EM data), as well as X-ray/neutron scattering profiles. We then use our Bayesian inference approach to maximize the agreement between experimental and simulation datasets (for both the cryo-EM and scattering data).