What is a single cryo-EM image worth? A survey of high-resolution cryo-EM model system datasets

Vinh Truong[1], Tiffany Chen[2], Elizabeth Kellogg[3]

[1]No affiliation given, [2]Kellogg Lab, [3]NA

vht3@cornell.edu

Significant advancements in both cryo-EM hardware and software have, remarkably, achieved true atomic resolution (~1 Å) cryo-EM reconstructions. Given that this was achieved using reasonably sized particle datasets, this suggests that the improved resolution arises from increased signal contained within individual cryo-EM particle images. Here, we ask the question: what structural features can be reliably identified from small sets of aligned cryo-EM images using various microscope setups? Remarkably, 10 pre-aligned particle images of a high-symmetry structure (i.e apoferritin) from the cold-FEG (CFEG) particle data-set was sufficient to obtain a ~10 A reconstruction, suggesting that substantial structural information may be present at the fold-level within small sets of cryo-EM images. We next compared small pre-aligned sets of particle images (ranging from 1 to 100 particles per set) against the experimental reconstruction (obtained using the full set of images) using established metrics such as the Fourier Ring Correlation (FRC). Unsurprisingly, we discovered that individual particles were too noisy to reliably assess information content using FRC, even when using improved metrics (such as the masked Wiener filter) to assess signal-to-noise ratios (SSNR). However, analysis of 100 pre-aligned particle images was sufficient to identify distinct signatures in the FRC profiles corresponding to secondary structure elements, such as α-helices. We compared FRC profiles from CFEG datasets and conventional XFEG datasets and discovered differences in the FRC profiles that suggest that more high-resolution spectral signal is contained within CFEG compared to XFEG images. We next compared the FRC profiles obtained from experimental reconstructions and atomic models and found that atomic models are mostly insensitive to simulation details. However, alterations to the atomic model result in sharp drops in FRC profiles, indicating that such metrics are highly sensitive to the atomic model being used to compare to experimental projection images. These computational analyses constitute our first attempts towards analyzing small sets of particle projection images. Ultimately, such techniques may be used to build thermodynamic landscapes describing protein structure and function without relying on massive datasets or extensive averaging to extract structural features.