CDCS
CENTER FOR DATA AND COMPUTING
IN NATURAL SCIENCES

OPENING SYMPOSIUM 2022

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

TUHH

Contribution ID: **102**

Type: **Poster**

Metadata-based analysis of image quality for single particle cryo-EM

Single particle cryo-electron microscopy (cryo-EM) is an increasingly important method for determining the three-dimensional structure of proteins. As a single particle technique, it allows for the elucidation of large macromolecular complexes, provides information on protein dynamics and gives access to proteins that are difficult to crystallize.

For this purpose, molecules in aqueous solution are rapidly frozen and then analyzed by transmission electron microscopy. The resulting 2D images are processed in computationally intensive pipelines to finally reconstruct a 3D density map which can be used for building an atomic model. As a result of the low signal-to-noise ratio in the images, thousands to millions of 2D projection views are necessary to reconstruct a single density map. These images can be of varying quality due to several reasons that include beam-induced motion, structural defects and sample heterogeneity. Thus, selection procedures are required to create high-quality datasets.

State-of-the-art processing workflows employ cross-correlation-based classification algorithms in 2D and 3D for image selection. These rely on the assumption that bad images will cluster together in low-quality classes, which can then be discarded. In practice, seemingly good classes often contain low-quality images along with the high-quality ones, resulting in the need for classification cascades and finally a trade-off between discarding good images and keeping bad ones.

In this work, we investigate the potential of metadata collected in the processing pipeline for the selection of high-quality images. We process a dataset of fatty acid synthase (FAS) in state-of-the-art manner and divide the final dataset into subsets based on value ranges for meta-parameters that relate to different aspects of image quality. By comparing the gold-standard resolution achieved for reconstructions from these subsets to their expected resolution from a Rosenthal-Henderson plot, we determine which meta-parameters might be meaningful for image selection.

Primary authors: CAVASIN, Anna Theresa (Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research (HZI), Department of Structural Infection Biology, Hamburg, Germany & Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Hamburg, Germany); LUETTICH, Mario (MPI for Multidisciplinary Sciences, Department of Structural Dynamics, Göttingen, Germany); STARK, Holger (MPI for Multidisciplinary Sciences, Department of Structural Dynamics, Göttingen, Germany); KOLBE, Michael (Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research (HZI), Department of Structural Infection Biology, Hamburg, Germany & Universität Hamburg, Department of Chemistry, Hamburg, Germany); RAREY, Matthias (Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Hamburg, Germany)

Presenter: CAVASIN, Anna Theresa (Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research (HZI), Department of Structural Infection Biology, Hamburg, Germany & Universität Hamburg, ZBH - Center for Bioinformatics, Research Group for Computational Molecular Design, Hamburg, Germany)

Session Classification: Poster session with buffet

Track Classification: CDL3 (Systems Biology)