Advances using machine learning and computational tools for crystal growth and detection

Sarah Bowman[1], Miranda Lynch[2]
[1]HWI [2]Hauptman-Woodward Medical Research Institute
sbowman@hwi.buffalo.edu

While there is much focus on how machine learning (ML) can assist analysis of structural datasets from diffraction and microscopy experiments, less recognized is the contribution of ML methods to upstream stages of the structure determination pipeline. A critical bottleneck in biomolecular crystallography is identifying conditions that will generate diffraction quality crystals. Imaging methods are key to the process of screening for successful crystallization hits and produce mineable data regarding the crystallization process. When crystallization screening is carried out in a high-throughput manner, the enormous number of images requires a large commitment of time and effort for locating successful crystallization conditions. To address this obstacle, machine learning approaches can be used to empower better detection of crystals. One successful example includes the MAchine Recognition of Crystallization Outcomes (MARCO) scoring algorithm, a deep learning convolutional neural network which was trained and tested on brightfield crystallization images in a collaborative effort between the National Crystallization Center at HWI, GoogleBrain, major pharmaceutical companies, and other crystallization facilities worldwide [1]. To further enhance crystal detection, we are working to integrate standard brightfield imaging with additional imaging modalities, including Ultraviolet Two-Photon Excited Fluorescence (UV-TPEF) and Second Harmonic Generation (SHG). We specifically are pursuing machine learning approaches to achieve combining the different information contributions of each imaging modality. Combining UV-TPEF and SHG enables detection of submicron crystals and crystals under precipitate, as well as distinguishes between protein and salt crystals. We have recently developed a Graphical User Interface called MARCO-Polo, which uses the MARCO algorithm on brightfield images coupled to a viewing platform for examining multiple imaging modalities, such as UV-TPEF and SHG [2]. Our current efforts are focused on ways to improve image processing and analysis for crystallization screening, with specific emphasis on machine learning approaches to combine the information from brightfield images and alterative imaging modalities to facilitate steps towards automated imaging to streamline crystal detection and acquisition.

[1] Bruno AE, Charbonneau P, Newman J, Snell EH, So DR, Vanhoucke V, Watkins CJ, Williams S, Wilson J. Classification of crystallization outcomes using deep convolutional neural networks. PLOS one. 2018 Jun 20;(6):e0198883.

[2] Holleman ET, Duguid E, Keefe LJ, Bowman SEJ. Polo: an open-source graphical user interface for crystallization screening. Journal of applied crystallography. 2021 Apr 1;54(2).