

B4 Autocollect: a Fully Automated Data Collection and Data Processing Pipeline at the Berkeley Center for Structural Biology

Daniil Prigozhin¹, Marc Allaire², Kevin Royal³, Jay Nix⁴, John Taylor⁵, Randall Cayford⁶, Anthony Rozales⁷, Jeff Dickert⁸

¹Lawrence Berkeley National Laboratory ²No affiliation given, ³No affiliation given, ⁴Molecular Biology Consortium, ⁵Lawrence Berkely National Lab, ⁶Lawrence Berkely National Lab, ⁷Lawrence Berkely National Lab, ⁸Lawrence Berkely National Lab
daniilprigozhin@lbl.gov

Ability to screen and process data from large numbers of crystals can be limited by the hands-on time required for data collection, reduction and dataset selection. This is especially true for highly repetitive large volume applications such as fragment-based drug design campaigns. To overcome this bottleneck, we have implemented and released to users a fully automated pipeline that starts with crystals and ends with reduced data and data summary statistics. Key components of the pipeline include task queue generation, diffraction-based crystal centering, and a data reduction pipeline with multiple resolution cutoff parameters. B4 Autocollect enables users to eliminate the time spent collecting data, and instead select among light, fully processed datasets that can be passed directly to automatic molecular replacement and ligand building pipelines such as phenix.ligand_pipeline for downstream processing. In the past year, both industry and academic users of Berkeley Center for Structural Biology beamlines have routinely used the B4 Autocollect pipeline to decrease operator time and cost per dataset and to increase overall productivity.