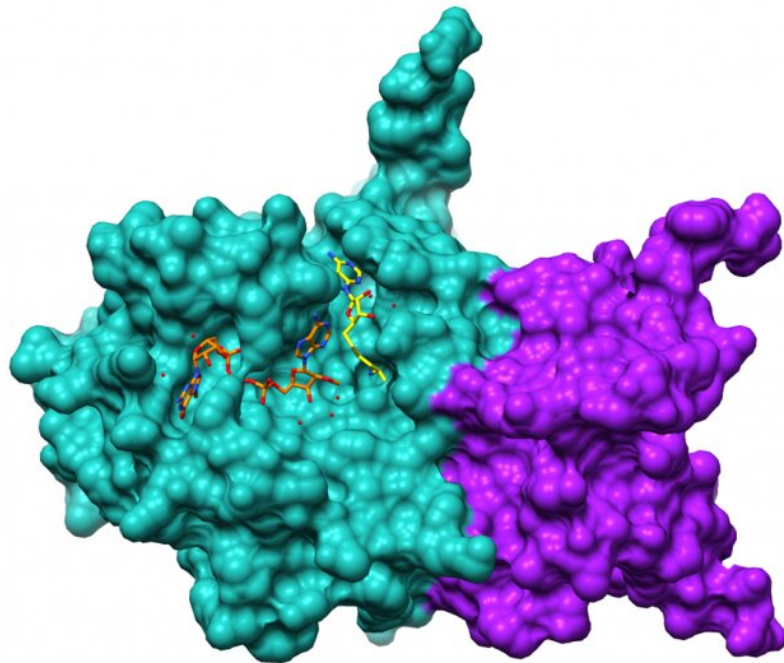# Researchers use Theta for real-time analysis of COVID-19 proteins

August 6 2020, by Nils Heinonen

Nsp10/16 surface with ligands. Researchers have developed a pipeline to connect ALCF supercomputers to APS experiments to enable real-time analysis of COVID-19 proteins, paving the way to elucidate important protein structural dynamics of the coronavirus. Credit: Mateusz Wilamowski, University of Chicago, Center for Structural Genomics of Infectious Diseases; George Minasov, Northwestern University, Center for Structural Genomics of Infectious Diseases

Argonne researchers have developed a pipeline between ALCF supercomputers and Advanced Photon Source experiments to enable on-demand analysis of the crystal structure of COVID-19 proteins.

As the coronavirus SARS-CoV-2 and its associated disease, COVID-19, developed and spread across the country and planet, the U.S. Department of Energy's (DOE) Argonne National Laboratory joined the global fight by beginning work to better understand and treat the pandemic. Several such lines of research have been launched at the Argonne Leadership Computing Facility, a DOE Office of Science User Facility, to take advantage of its considerable scientific resources; one of these lines has analyzed the crystal structure of a protein complex associated with the coronavirus.

Key to understanding the coronavirus is unraveling its structure. To this end, Argonne researchers have leveraged the ALCF's Theta supercomputer to analyze crystallographic images of a protein complex associated with the SARS-CoV-2. The images come from Argonne's Advanced Photon Source (APS), a DOE Office of Science User Facility, following experiments utilizing a technique known as serial synchrotron crystallography that is designed to elucidate the complex chemistry of viral proteins.

Serial synchrotron crystallography experiments employ high-intensity X-rays to reveal the structures of large molecules using only fractional radiation doses compared with the requirements of traditional crystallographic techniques. As a result, serial synchrotron crystallography permits researchers to image tens of thousands of microscopic crystals, with very short exposure lengths for each individual sample. The high speed of the technique leads to the generation of a vast array of data, the complexity and density of which necessitate sophisticated and computationally demanding analyses.

Massively parallel systems like Theta are unique in their ability to meet the demands that serial synchrotron crystallography poses for rapid, on-the-fly processing. Enabling Theta for use in on-the-fly processing is a data pipeline constructed around the supercomputer. This pipeline automates data acquisition, analysis, curation, and visualization, transporting results to a repository from which metadata can be extracted for publication.

The pipeline generates large image batches at a high rate, with [data transfers](#) achieving speeds of 700 megabytes per second thanks to Globus, a University of Chicago-run data management service.

"This pipeline's deployment between the APS and the ALCF for on-demand analysis has been a tremendous success," said Ryan Chard, a computer scientist at Argonne leading the image-processing efforts. "We achieved a processing rate of up to 95 images a second." This high speed made it possible to deliver instantaneous feedback to experimentalists at the APS.

The pipeline begins with Globus transferring images from the APS to the Theta system. The images are then analyzed and processed using FuncX, a function-as-a-service computation system that organizes the dispatch of individual tasks to available computing nodes. FuncX is subsequently also used to extract metadata about hits, identify crystal diffractions, and generate visualizations depicting both the sample and hit locations. After this the raw data, metadata, and related visualizations are published to a portal hosted at the ALCF, where they are indexed and made searchable for reuse.

Nineteen samples were analyzed across nearly 1,500 flows over the course of three ten-hour runs on the APS beam, during which over 700,000 images were processed on Theta. The resultant data were published to the data portal and used to further refine experimental work

and configurations. The orchestration required to facilitate research at this scale is enabled by research data automation services currently under development on the Globus platform, and underpinned by the reliable file transfer, and secure data sharing capabilities that are already widely used across APS beamlines. These capabilities will continue to improve with future planned enhancements to APS beamlines, ALCF supercomputers, Globus, and the APS-to-ALCF network. The forthcoming APS Upgrade, which will allow researchers to see things at scale they've never seen before with storage-ring based X-rays, will increase data rates by orders of magnitude. Combining these capabilities of the ALCF and APS Upgrade will greatly enhance the scientific discovery.

"The increasing biological relevance of serial synchrotron crystallography experiments has researchers preparing a number of further experiments in the coming weeks," said Darren Sherrell, a biophysicist and beamline scientist at the X-ray Science Division of the APS. "This work paves the way to elucidate important protein structural dynamics of the coronavirus."

Provided by Argonne National Laboratory