Quick Take

Coronavirus Reveals Its Atomic Makeup Through Massive Supercomputer Simulation

Michael Feldman, Melissa Riddle, and Steve Conway
August 2020

HYPERION RESEARCH OPINION

A digital simulation of the COVID-19 (SARS-CoV-2) spike protein that was run on the Frontera supercomputer at the Texas Advanced Computing Center (TACC) may provide needed insights for the design of potential vaccines and drugs being developed against the virus. Understanding the structure of the spike protein is crucial because the virus uses this component to infect human cells. Since this simulation provides a level of structural detail beyond that which is achievable by experiment, it promises to help researchers figure out how the virus is able to infect humans so effectively.

The Frontera work also demonstrates the importance of building a body of computational research directed at virus pathogens. In this case, the researchers were able to leverage existing HPC-mediated efforts that examined H1N1 influenza virus proteins, significantly speeding up the development of the SARS-CoV-spike protein model. The current work on Frontera will likely be used to drive additional HPC research focused on this current virus as well as those for future pathogens.

The Frontera work is an example of how rapidly supercomputing resources around the world are being directed toward the current pandemic. The TACC system is not alone in this regard, the latest example being the use of the recently installed Fugaku supercomputer for SARS-CoV-2 research at RIKEN in Japan.

SITUATION OVERVIEW

A research team at the University of California, San Diego (UCSD) led by Professor Rommie Amaro has developed a 200-million-atom model of the SARS-CoV-2 spike protein. This protein is a critical component of the virus since it is used to infiltrate the host cell membrane, enabling the SARS-CoV-2 RNA to be injected into the cell.

The structural composition of the spike protein was derived from cryogenic electron microscopy and X-ray crystallography data. The resulting simulation on the Frontera supercomputer produced a three-dimensional physical model of the protein and enabled researchers to explore its time-dependent dynamics. As a result, the model was able to provide a level of detail that could not otherwise be revealed from physical observations and experiments.

According to Amaro, the NAMD molecular dynamics simulations used in developing and testing the model was run on TACC's Frontera supercomputer, using up to 4,000 of the system's nodes (or about 250,000 of its processing cores). The model is computationally demanding since the simulation has to process the dynamic interactions between the protein's atoms.

The spike protein model was developed in a matter of months following the rapid spread of the virus in early 2020, with the first complete production run of the simulation performed in mid-March. The rapid implementation was enabled by a similar effort of the UCSD research team directed at simulating surface proteins of the H1N1 influenza virus. That model, which encompassed 160 million atoms, was aimed at understanding the atomic behavior of the surface proteins, with the idea of using the information to develop drug candidates. The result of that research was published in February 2020. According to Amaro, the SARS-CoV-2 research work on Frontera is ongoing, currently amounting to about 2.7 million node hours.

The 4,000 nodes used for the spike protein simulation amounts to about half of the Frontera's computing resources. The system, a Dell EMC PowerEdge cluster powered principally by Intel Xeon Platinum 8280 processors, is currently ranked as the eighth most powerful supercomputer in the world, with a High Performance Linpack (HPL) result of 23.5 petaflops. The NSF-funded system was installed at TACC in 2019.

FUTURE OUTLOOK

The Frontera modeling work on SARS-CoV-2 spike protein is adding to the base of knowledge the medical community is collecting to understand how the virus is able to spread so effectively throughout human populations. Work in this area is certain to continue and will drive future models of the viral envelope. This will be especially critical as the virus mutates and variants of the spike protein are formed.

Due to the spike protein's function in infecting host cells, it is one of the most crucial elements of the virus and a prime target for a vaccine for drug. But there are other proteins on the viral envelope as well, some of which could also be potential vaccine or drug targets. The UCSD research team intends to model some of these different SARS-CoV-2 proteins in future investigations.

About Hyperion Research, LLC

Hyperion Research provides data driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user and vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors, and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com
