OAK RIDGE National Laboratory

The CoViD-19 High-Performance Computing Consortium on Summit

Bronson Messer

Director of Science

Oak Ridge Leadership Computing Facility

Oak Ridge National Laboratory

ASCAC Meeting – Sept. 2021

ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

OAK RIDGE National Laboratory

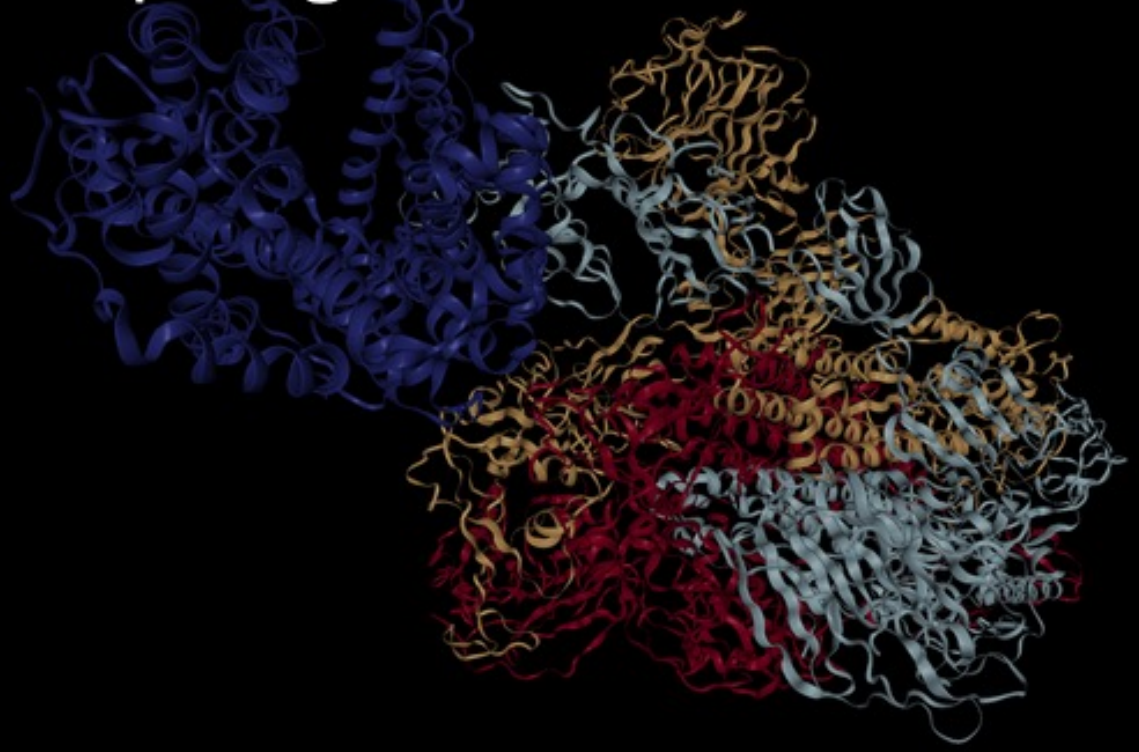
Oak Ridge Leadership Computing Facility (OLCF)

Mission: Deploy and operate the computational and data resources required to tackle global challenges

- Providing the resources to investigate otherwise inaccessible systems at every scale: from galaxy formation to supernovae to earth systems to automobiles to nanomaterials
- With our partners, deliver transforming discoveries in materials, biology, climate, energy technologies, and basic science

The COVID-19 HPC Consortium – A Model for Rapid Response

The COVID-19 High Performance Computing Consortium



Bringing together the Federal government, industry, and academic leaders to provide access to the world's most powerful high-performance computing resources in support of COVID-19 research.

100
Projects

6.8m
CPU cores

Consortium Members and Affiliates

Department of Energy National Laboratories

- Argonne National Laboratory
- Idaho National Laboratory
- Lawrence Berkeley National Laboratory
- Oak Ridge National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- Sandia National Laboratory

Academia

- Massachusetts Institute of Technology
- MGHPCC
- Rensselaer Polytechnic Institute
- University of Illinois
- University of California – San Diego
- Carnegie Mellon University
- University of Pittsburgh
- Indiana University
- University of Wisconsin – Madison
- Ohio Supercomputing Center
- UK Digital Research Infrastructure
- CSCS – Swiss National Supercomputing Centre
- SNIC PDC – Swedish National Infrastructure for Computing, Center for High Performance Computing

Industry

- IBM
- Amazon Web Services
- AMD
- BP
- D.E.Shaw Research
- Dell
- Google Cloud
- Hewlett Packard Ent.
- Intel
- Microsoft
- NVIDIA

Affiliates

- Atrio
- Data Expedition
- Flatiron
- Fluid Numerics
- Immortal Hyperscale InterPlanetary Fabrics
- MathWorks
- Raptor Computer Systems
- SAS
- The HDF Group

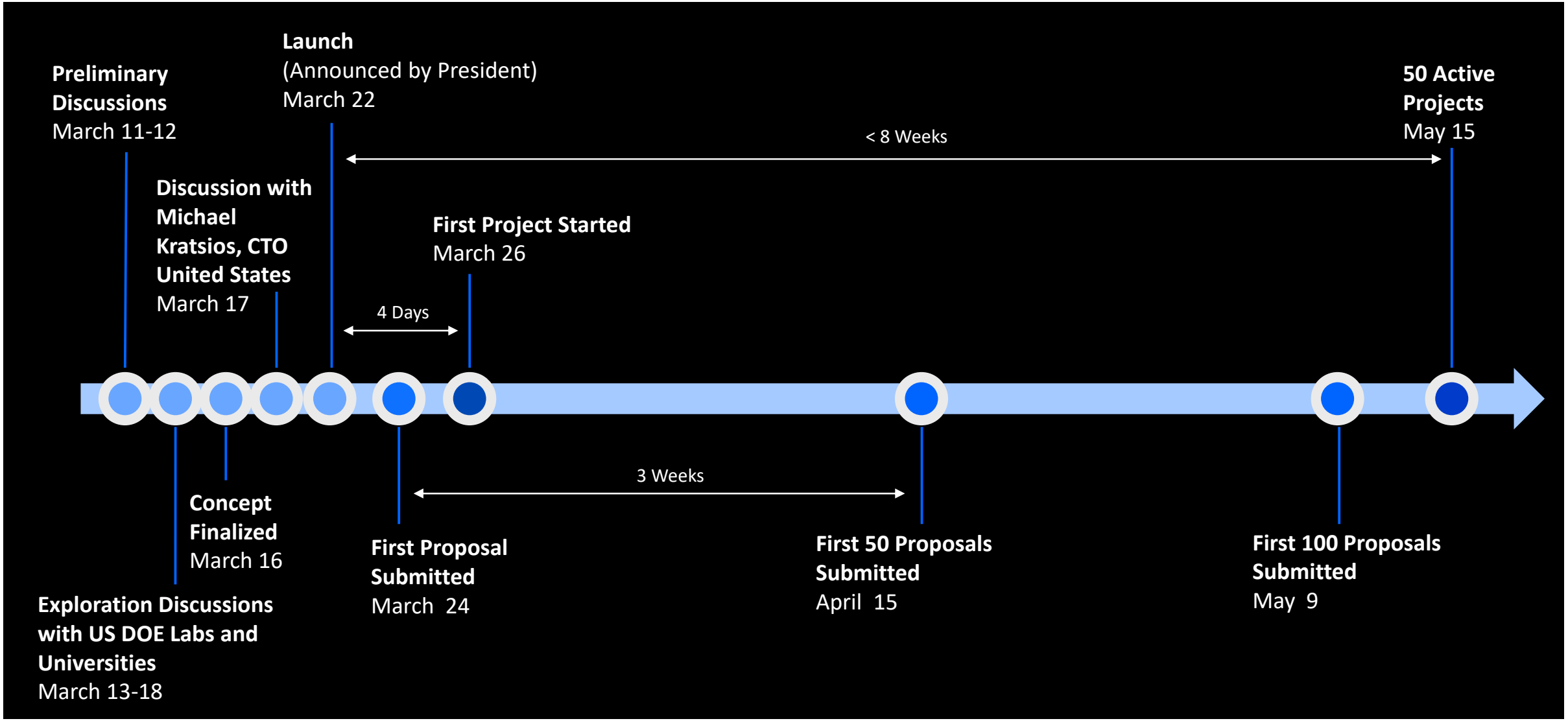
Federal Agencies

- NASA
- National Science Foundation
  - XSEDE
  - Pittsburgh Supercomputing Center
  - Texas Advanced Computing Center (TACC)
  - San Diego Supercomputing Center (SDSC)
  - National Center for Supercomputing Applications (NCSA)
  - Indiana University Pervasive Technology Institute (IUPTI)
  - Open Science Grid (OSG)
  - National Center for Atmospheric Research (NCAR)

International

- Korea Institute of Science and Technology Information (KISTI)
- Ministry of Education, Culture, Sports, Science and Technology (MEXT) – Japan
  - RIKEN Center for Computational Science (R-CCS)

Early Timeline of the Consortium



Topical Overview of HPC Consortium Projects

- Basic Science
  - Viral structure, viral-human interaction, viral evolution, environmental effects, tools
- Therapeutics
  - Target discovery, small molecule design, protein design, drug repurposing, technologies for development
- Patients
  - Trajectory and outcomes, medical tech, supply chain, epidemiology, detection, etc.

COVID-19 HPC Consortium Projects at OLCF

- 11 projects, 1.42M Summit node-hours allocated to date
  - This represents a bit more than half of all the projects and allocated time on Summit for COVID-19-related research since March 2020.
- Several projects also made use of resources offered by other Consortium members (e.g. Google, TACC, IBM, etc.)
- COVID-19 research projects enjoy queue priorities equal to INCITE projects.

Summit COVID-19 HPC Consortium Projects

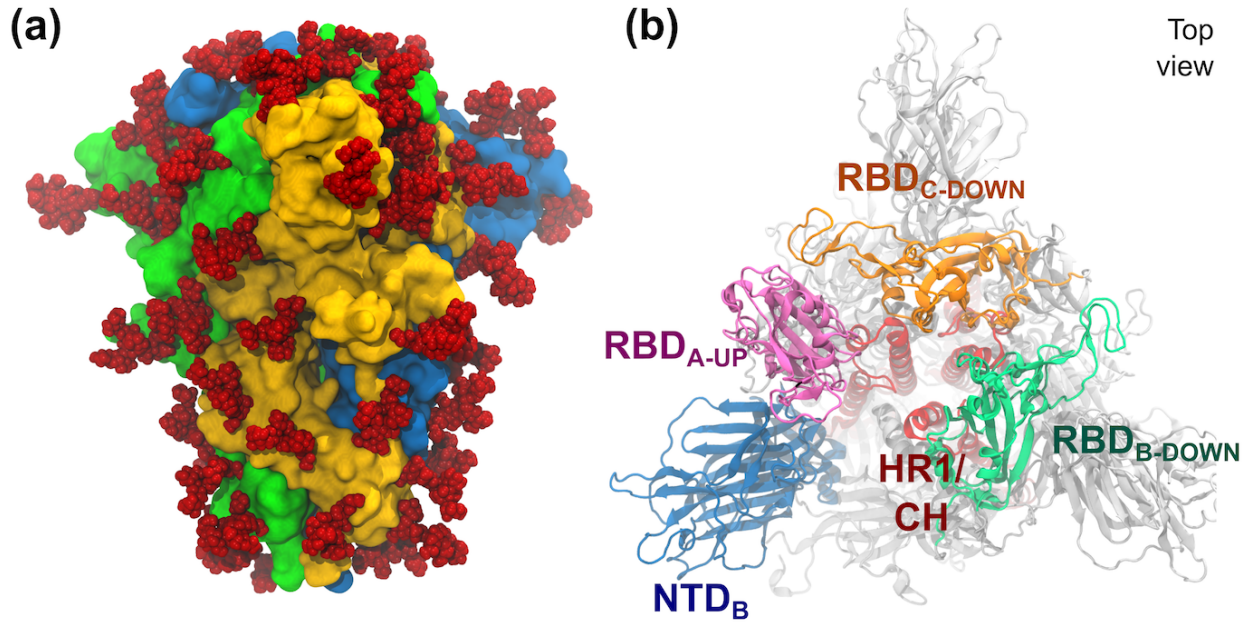
- **Rick Stevens (Co-I Arvind Ramanathan), Argonne National Laboratory**
  "Targeting the SARS-CoV-2 Proteome with Artificial Intelligence (AI) Driven Small Molecule Design and Screening"
- **Debsindhu Bhowmik and John Gounley, Oak Ridge National Laboratory**
  "Structural Modeling of COVID-19 with HPC"
- **Jeremy Smith, Univ. of Tennessee and Oak Ridge National Laboratory**
  "Drug Discovery for COVID-19"
- **Andrés Cisneros, University of North Texas**
  "Using MD and QM/MM to improve drug candidates for nCoV-19 targets"
- **Jin Yu, University of California – Irvine**
  "Dissecting inhibitor impacts on viral RNA polymerase and fidelity control of RNA synthesis in SARS-CoV-2"
- **Ken Dill, State University of New York at Stony Brook**
  "Physical Models of COVID-19 Related Proteins"
- **Jason Crain, IBM**
  "COVID-19: Molecular Simulations and Underexplored Drug Targets"
- **Jennifer Diaz, Icahn School of Medicine at Mount Sinai**
  "Prediction of Synergistic Drug Combinations for Treatment of COVID-19"
- **James Gumbart, Georgia Institute of Technology**
  "Determining The Contribution of Glycosylation To SARS-CoV-2 S-protein Conformational Dynamics"
- **Gregory Voth, University of Chicago**
  "Multiscale Modeling of SARS-CoV-2 Virion and Structural Proteins"
- **Karan Kapoor, University of Illinois at Urbana-Champaign**
  "Simulation of Full-scale, Membrane-Bound SARS-CoV2 Spike Proteins in Crowded Viral Envelope"

Topical Overview of HPC Consortium Projects

- Basic Science
  - **Viral structure**, viral-human interaction, viral evolution, environmental effects, tools
- Therapeutics
  - Target discovery, **small molecule design**, protein design, **drug repurposing**, technologies for development
- Patients
  - Trajectory and outcomes, medical tech, supply chain, epidemiology, detection, etc.

Determining The Contribution of Glycosylation To SARS-CoV-2 S-protein Conformational Dynamics (PI: James Gumbart, Ga. Tech)

- Study of the conformational transition of the SARS-CoV-2 spike protein from a closed, noninfectious state to an open one that can bind to its receptor ACE2 on human cells.
- One particular antibody's (BD-368-2) epitope is exposed regardless of whether the S-protein is in the up or down state (Yu et al. https://doi.org/10.1101/2021.08.12.456168)



Summit Node Architecture Upgrade

Added 54 larger-memory nodes for COVID research and other applications needing more memory

- Add 54 nodes, each with double the HBM and 4x the DDR4 and NVMe
- Allows jobs that need larger on-node memory to run on up to 54 nodes
- Applications that benefit
  - Computational chemistry and MD
  - AI/DL for medical image analysis
  - Others…

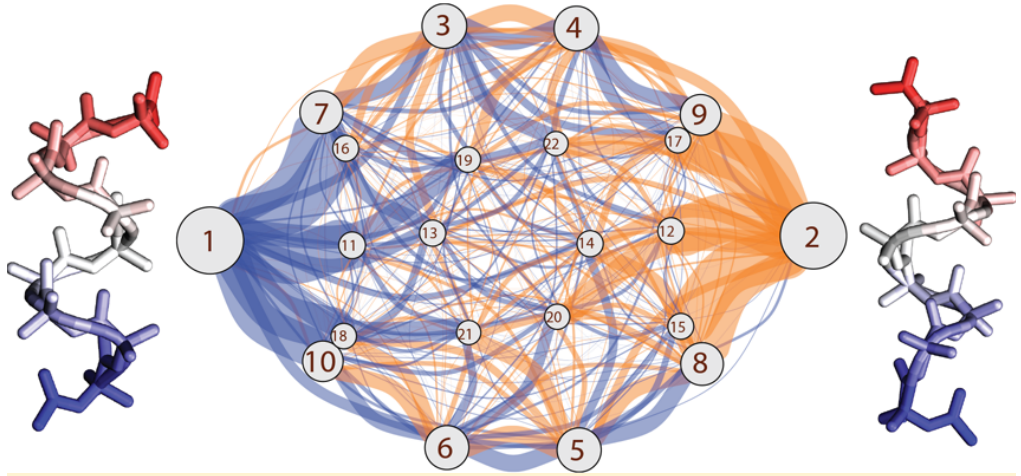
Feature	Current Summit Nodes	Large Memory Nodes
Peak $FLOPS_{64}$	**200 PF**	**203 PF**
Number of Nodes	**4,608**	**54**
Node Performance	**43 TF**	**43 TF**
Memory per Node	**512 GB DDR4 + 96 GB HBM2**	**2048 GB DDR4 + 192 GB HBM2**
NV Memory per Node	**1.6 TB**	**6.4 TB**
Total System Memory	**2.8 PB + 7.4 PB NVM**	**2.9 PB + 7.7 PB NVM**
System Interconnect	**Dual Rail EDR-IB (25 GB/s)**	**Dual Rail EDR-IB (25 GB/s)**
Interconnect Topology	**Non-blocking Fat Tree**	**Non-blocking Fat Tree**
Bi-Section Bandwidth	**115.2 TB/s**	**115.2 TB/s**
Processors on Node	**2 IBM POWER9™ 6 NVIDIA Volta™**	**2 IBM POWER9™ 6 NVIDIA Volta™**
File System	**250 PB, 2.5 TB/s, GPFS™**	**250 PB, 2.5 TB/s, GPFS™**

Prediction of Synergistic Drug Combinations for Treatment of COVID-19 (PI: Jennifer Diaz, Icahn [now UCLA])

- Aim to identify pairs of drugs that synergistically perturb the human cell network that interacts with the COVID virus and possibly impact the way a cell responds to a COVID infection
- Used a Random Forest machine learning algorithm, developing a classifier to learn how the molecular profiles of individual drugs transform into the molecular profiles of pairs of drugs. **New CARES Act high-memory Summit nodes were essential for this work.**
- Pilot set of about 22,000 drugs—ones that were most like the drugs used in the classifier's training set, aiming for the most reliable results.
- Top five drug pairs being investigated in the Garcia-Sastre Lab at the Icahn School of Medicine at Mount Sinai. 1M drug screen being planned

Physical Models of COVID-19 Related Proteins
(PI: Ken Dill, Stony Brook)

- The SARS-CoV-2 virus has 29 proteins. In April 2020, only 8 of them had experimentally determined structures. This was (is) an impediment to understanding mechanisms and drug discovery.
- MELD (Modeling Employing Limited Data)used input from various public servers (PSIPRED, trRosetta) to guide molecular dynamics (MD) simulations using ML to enable a Bayesian reduction of the search space by focusing the sampling to only structures that are compatible with the extant data.
- The most challenging part of the project was the lack of predicted residue-residue contacts from the machine learning server (trRosetta). This is important because such contacts are input in the MELD x MD to reduce the configurational space of protein chains to be explored. But, the ML models had, of course, never seen these proteins before. (They existed in no databases.)



Prediction results made public
https://github.com/roynassar89/MELDxMD-COVID19-related-structure-predictions

Structural Modeling of COVID-19 with HP
(PI: Debsindhu Bhowmik, ORNL)

COVID19 RESEARCH: A COMPREHENSIVE NOVEL AI-BASED QUANTITATIVE BIOLOGY APPROACH

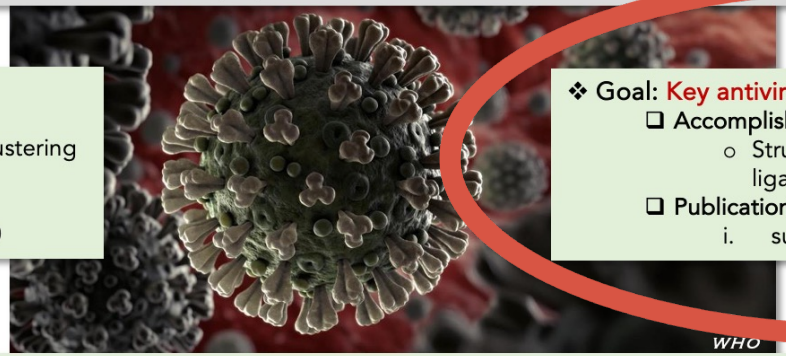
Continued

❖ Goal: Potential Therapeutic Targets (*through S protein*):
- ❑ Accomplishments:
  - o Structural Flexibility within Spike (S) Protein
  - o Built AI-based clustering for molecular design and analysis
- ❑ Publications:
  - i. 10.1101/2020.04.17.047548
  - ii. In preparation

Drug candidates

Completed

❖ Goal: Drug Discovery Pipeline:
- ❑ Accomplishments:
  - o AI-based conformational clustering tool for Ensemble Docking
- ❑ Publications:
  - i. 10.1021/acs.jcim.0c01010



Drug targets

❖ Goal: Key antiviral Drug Targets (*through M^{pro}*):
- ❑ Accomplishments:
  - o Structural dynamics of main protease (M^{pro}) ligand binding sites
- ❑ Publications:
  - i. submitted

New drugs molecule

❖ Goal: New Drug molecules:
- ❑ Accomplishments:
  - o Built Genetic Algorithms based Generative Adversarial Networks (GANs) with adaptive training
- ❑ Publications:
  - i. 10.1186/s13321-021-00494-3

Follow on

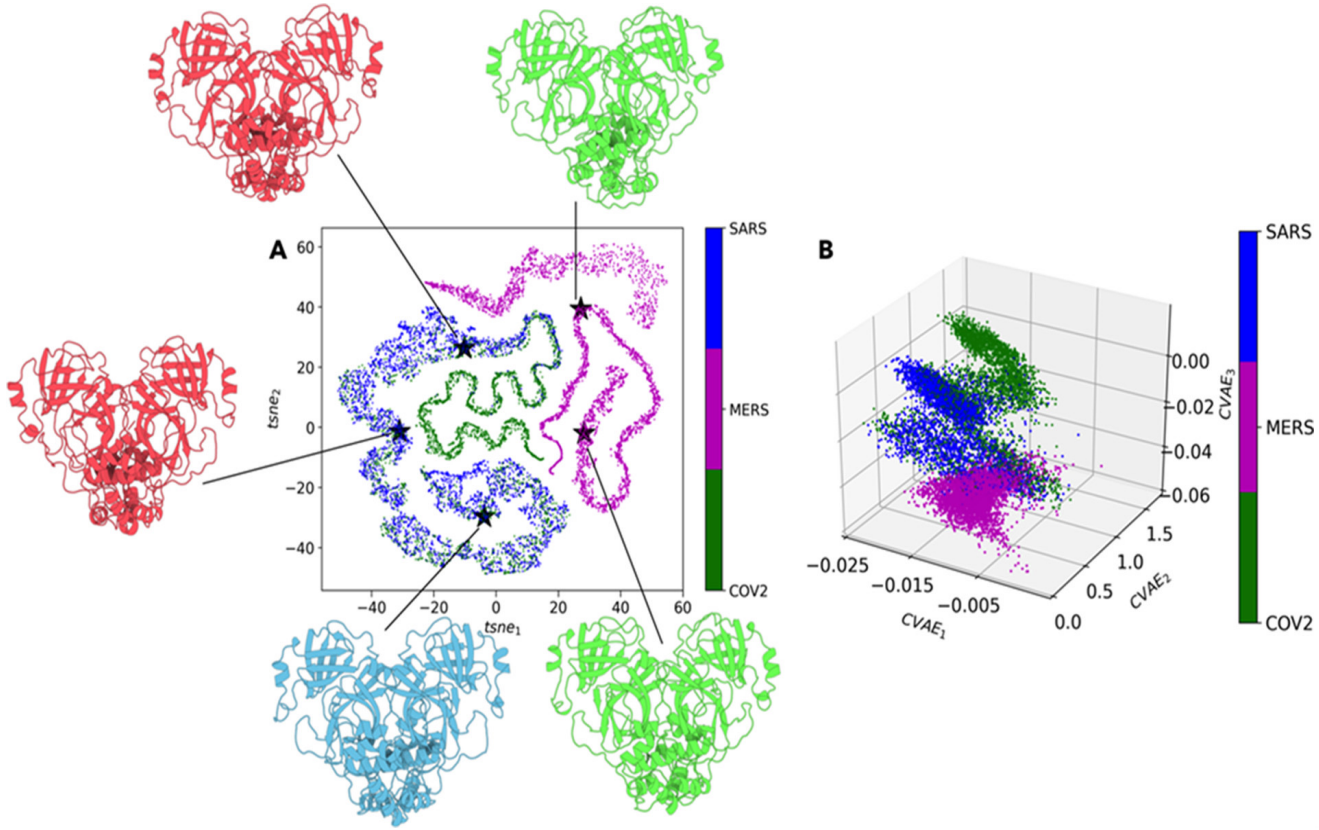
AI-based drug design through active learning on HPC

❖ Goal: Data-centric Approach for Novel Therapeutics Design Against COVID-19

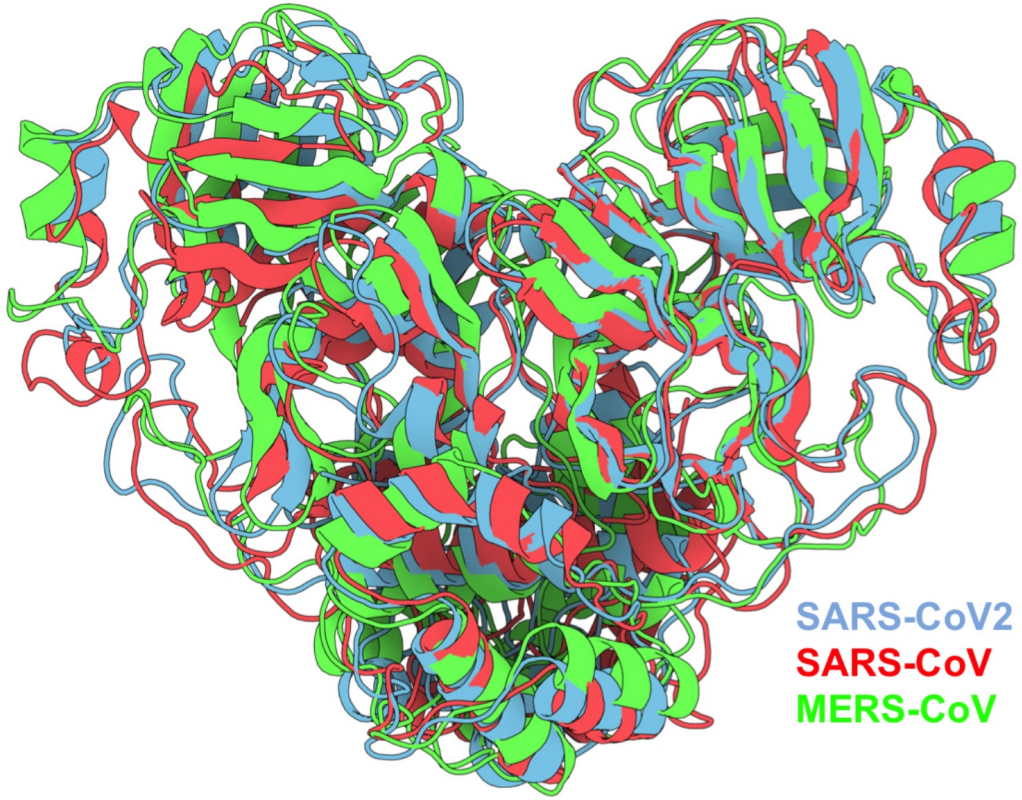
Bhowmik et al. | Funding Acknowledgements: DOE/ ASCR/ ECP/ NVBL

Structural Modeling of COVID-19 with HP (PI: Debsindhu Bhowmik, ORNL)

Cho, et al. (2021)
https://doi.org/10.1021/acs.jcim.1c00449



Clustering using variational auto-encoder



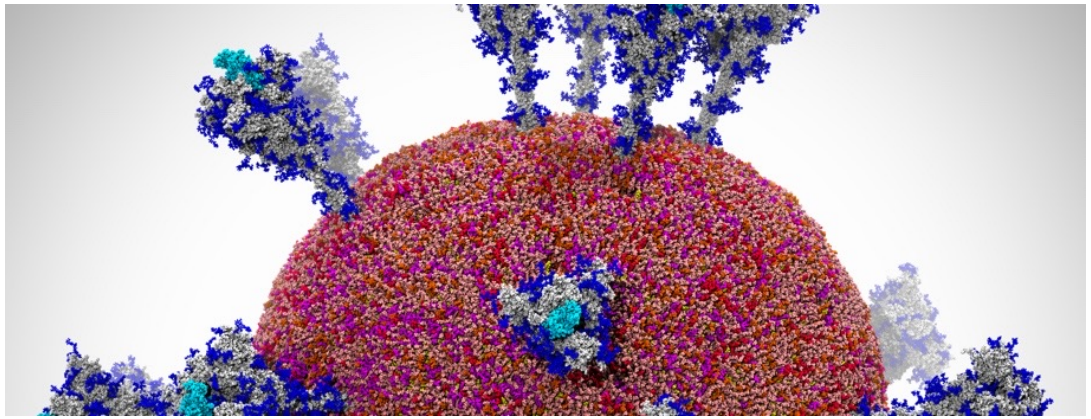
Representative structure of the most populated metastable state in SARS-CoV2 (cyan), SARS-CoV(red), and MERS-CoV (green)

Gordon Bell Special Prize-Winning Team Reveals AI Workflow for Molecular Systems in the Era of COVID-19 (PI: Rommie Amaro, UCSD)

Research by team at ANL and UC San Diego leads to a novel understanding of SARS-CoV-2 and a new method for studying disease

The Science

Imaging techniques, such as X-ray imaging and cryogenic electron microscopy, can provide snapshots of viruses such as SARS-CoV-2, but these fall short of capturing the dynamic movements of viral proteins. Computer simulations can help scientists capture the movements of these structures virtually. Now, a team led by Rommie Amaro at the University of California San Diego and Arvind Ramanathan at ANL have built a first-of-its-kind workflow based on AI and have run it on OLCF's Summit supercomputer to simulate the virus's spike protein in numerous environments, including within the SARS-CoV-2 viral envelope comprising 305 million atoms—the most comprehensive simulation of the virus performed to date. The accomplishment has earned the ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research.



A snapshot of a visualization of the SARS-CoV-2 viral envelope comprising 305 million atoms. Image Credit: Rommie Amaro, UC San Diego; Arvind Ramanathan, ANL

The Impact

The team was able to successfully scale NAMD to 24,576 of Summit's NVIDIA V100 GPUs. The results of the team's initial runs on Summit have led to discoveries of one of the mechanisms that the virus uses to evade detection as well as a characterization of interactions between the spike protein and the protein that the virus takes advantage of in human cells to gain entry—the ACE2 receptor. The team is now integrating their scientific code, NAMD, into their workflow pipeline to fully automate the transition from simulation to AI for data processing without gaps.

PI(s)/Facility Lead(s): Rommie Amaro
ASCR Program/Facility: INCITE / OLCF
ASCR PM: Christine Chalk
Publication(s) for this work: Lorenzo Casalino, Abigail Dommer, Zied Gaieb, Emilia P. Barros, Terra Sztain, Surl-Hee Ahn, Anda Trifan, Alexander Brace, Anthony Bogetti, Heng Ma, Hyungro Lee, Matteo Turilli, Syma Khalid, Lillian Chong, Carlos Simmerling, David J. Hardy, Julio D. C. Maia, James C. Phillips, Thorsten Kurth, Abraham Stern, Lei Huang, John McCalpin, Mahidhar Tatineni, Tom Gibbs, John E. Stone, Shantenu Jha, Arvind Ramanathan, and Rommie E. Amaro. "AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics." *International Journal of High-Performance Computing Applications, SC20*(2020).

Summary

- The COVID-19 HPC Consortium is a public-private consortium that provides COVID-19 researchers worldwide access to the world's most powerful high-performance computing resources to significantly advance the pace of scientific discovery in the fight to stop the virus.
- OLCF has been a partner in this endeavor from the inception of the Consortium and has provided a unique set of capabilities for a variety of projects.
- The Consortium allowed us to open our aperture to a number of projects that were *uniquely* enabled by the power of Summit. We anticipate more projects on a wide variety of topics in the coming months.