OAK RIDGE National Laboratory

COVID-19 Efforts at OLCF: Summit vs. Coronavirus

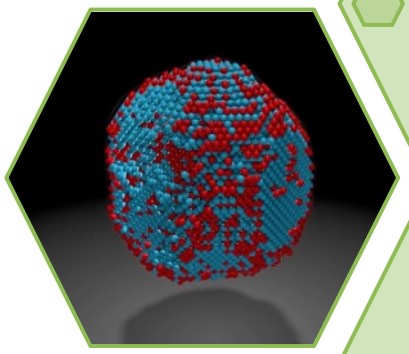
Georgia (Gina) Tourassi
National Center for Computational Sciences
Oak Ridge National Laboratory

Advanced Scientific Computing Advisory Committee
April 23, 2020

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

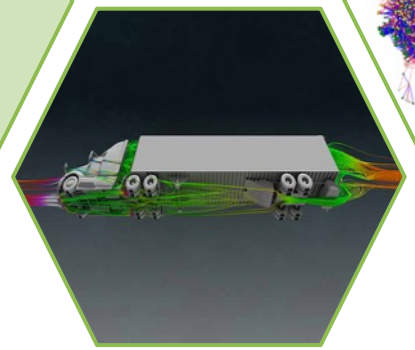
U.S. DEPARTMENT OF ENERGY

OLCF – A catalyst of scientific discovery



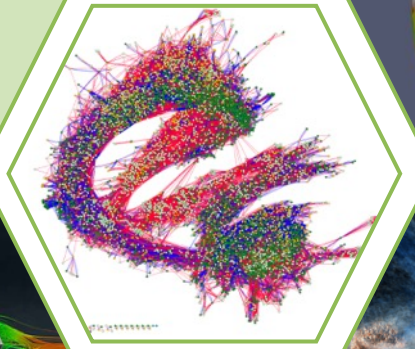
Nano-materials

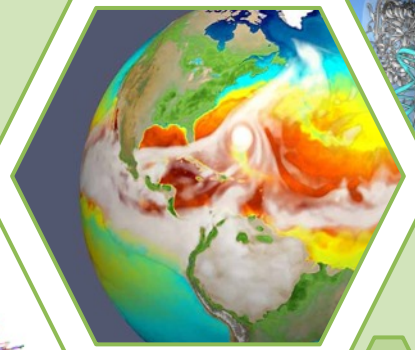
Fuel Efficient Vehicles



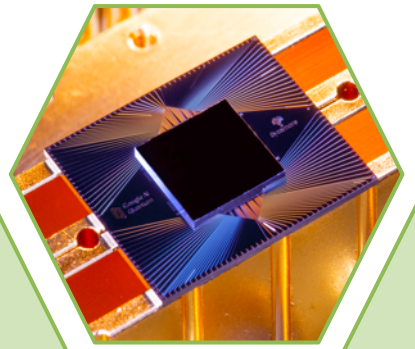
Biofuels

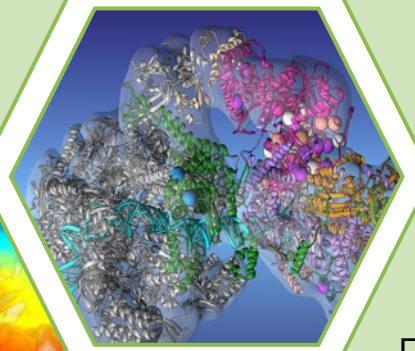
Earth System Modeling





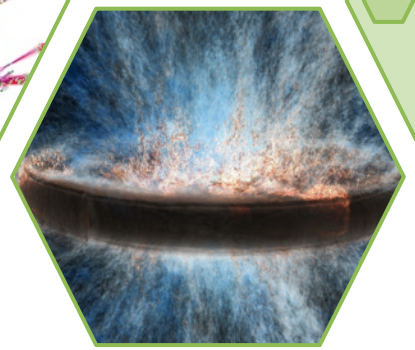
Genetic Disease





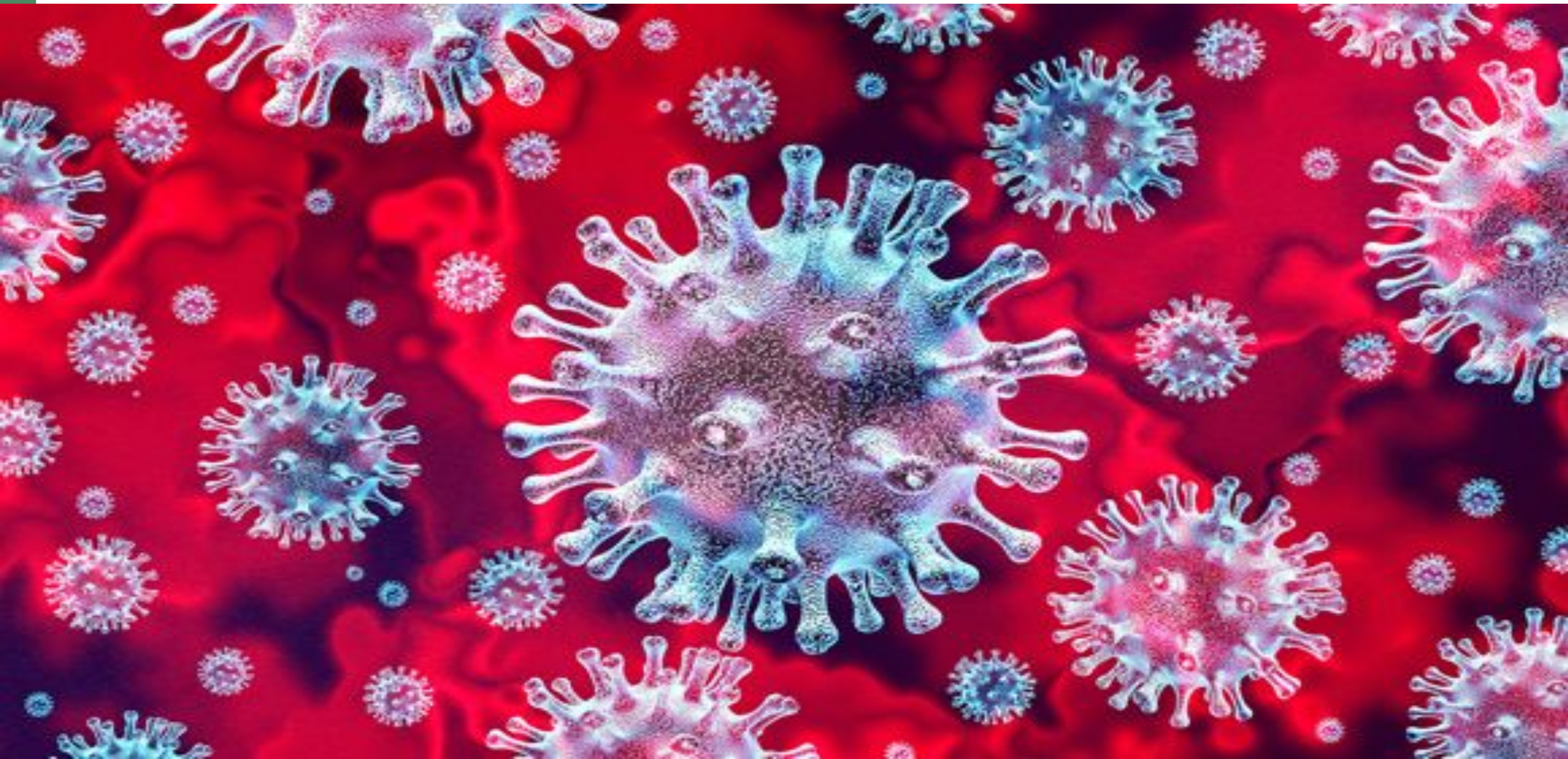
Quantum Computing

Galactic Winds



Human Exploration of Mars

Supercomputers – ally to biomedical discovery

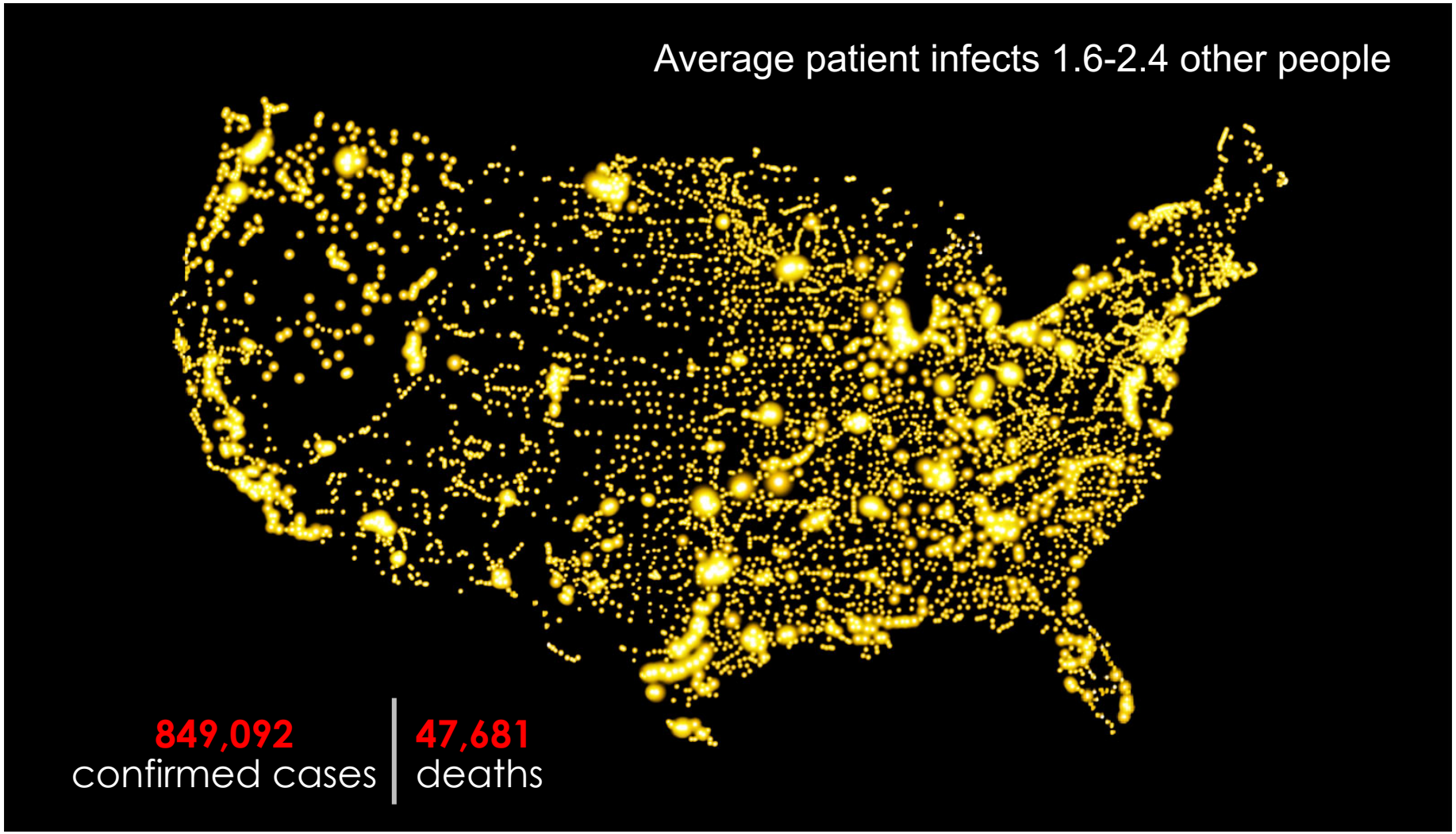


Mapping the Coronavirus Pandemic (April 22, 2020)

210 countries
affected

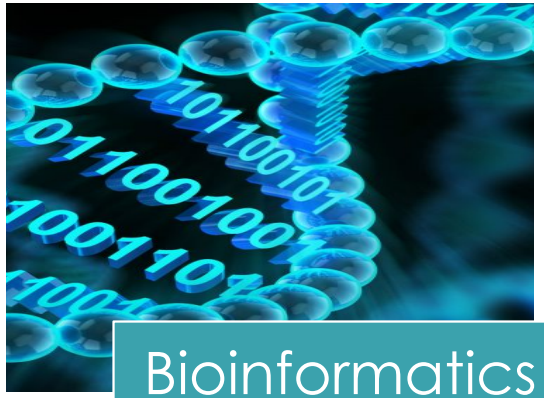
2,637,888
confirmed cases

184,235
deaths



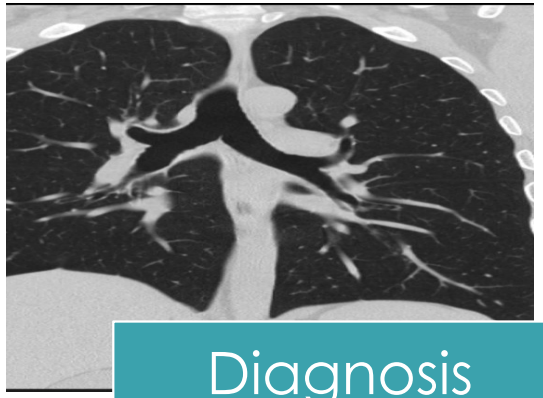
SOURCE: www.futurity.org

The Role of Supercomputing in COVID-19 Research



Bioinformatics

Drug Discovery

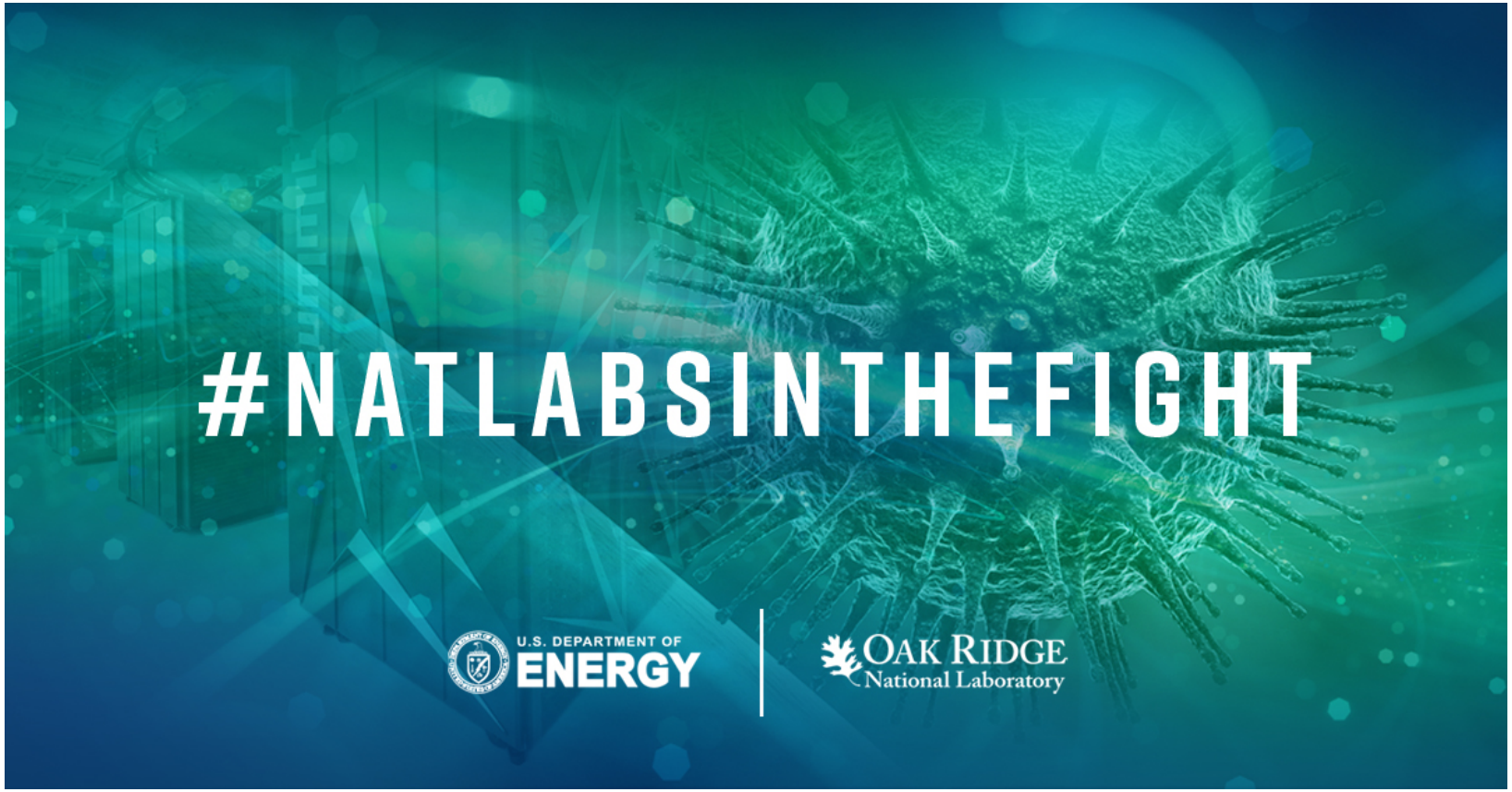


Diagnosis

Prognosis

Epidemiology

OLCF COVID-19 Response



COVID-19 HPC Consortium – *Announced March 22, 2020*

> **The COVID-19 High-Performance Computing (HPC) Consortium is a unique private-public effort spearheaded by the White House Office of Science and Technology Policy, the U.S. Department of Energy and IBM to bring together federal government, industry, and academic leaders who are volunteering compute time and resources on their world-class machines.**

COVID-19 HPC Consortium Members – 33 to date

Industry

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

National Labs

- Argonne National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- Oak Ridge National Laboratory
- National Energy Research Scientific Computing Center
- Sandia National Laboratories

Academia

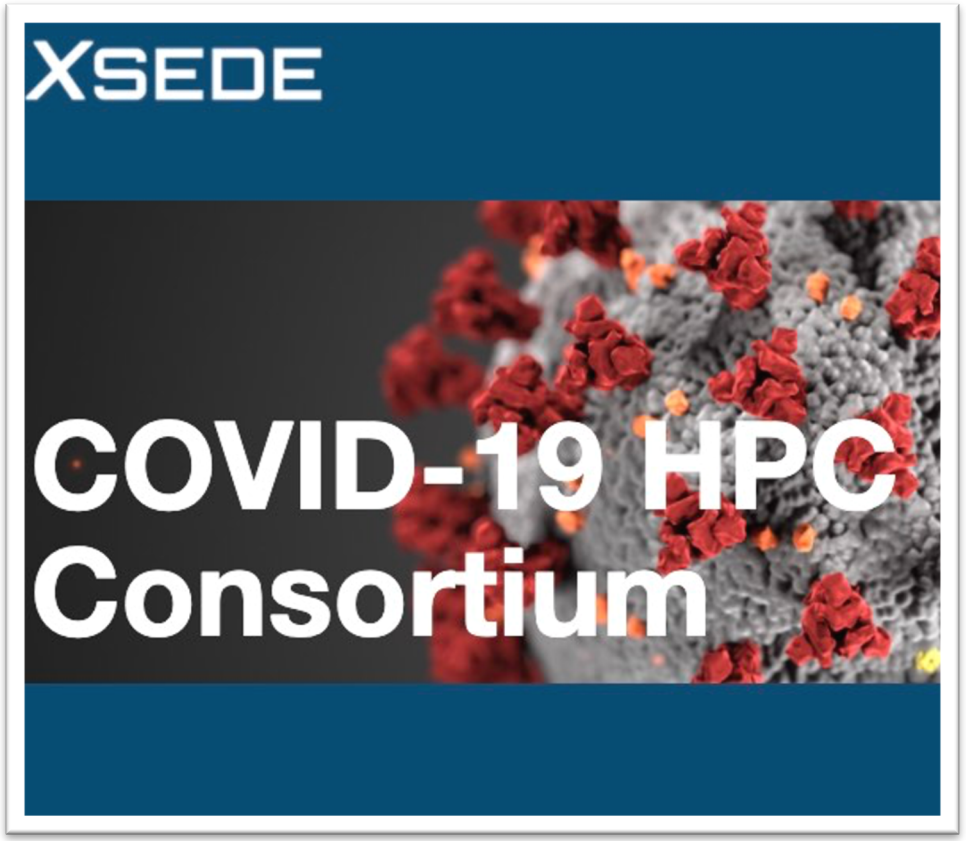
- Massachusetts Institute of Technology
- Rensselaer Polytechnic Institute
- University of Illinois
- University of Texas at Austin
- University of California - San Diego
- Carnegie Mellon University
- University of Pittsburgh
- Indiana University
- University of Wisconsin-Madison

Federal Agencies

- NSF
- NASA

COVID-19 HPC Consortium – Available resources

- 105K nodes
- 3.8M CPU cores
- 41K GPUs
- 418 petaflops



TO DATE:
27 projects selected and triaged

https://covid19-hpc-consortium.org/

COVID-19 Proposal Management at OLCF

PI initiates request

- All proposals are routed through our Director's Discretionary (DD) program.
- Individuals apply for time directly to our DD program and via the COVID-19 High Performance Computing Consortium.

Proposal Review

- **DD:**
  - reviewed at ORNL,
  - user agreement executed with PI's institution,
  - PI agrees to monitor project,
  - export control review
- **HPC Consortium:**
  - reviewed by Consortium Technical Committee,
  - user agreement executed with PI's institution,
  - PI agrees to monitor project,
  - export control review

Allocation and Expectations

- Currently, COVID-19 projects are enabled through September 30.
- Typical award amount is 50-100k Summit node-hours.
- Consortium PIs are asked to provide weekly progress reports.
- OLCF requires quarterly reporting or more often if results are obtained and are to be published.

OLCF COVID-19 Dashboard

- **More than 1,251,200 Summit node-hours have been allocated to 8 projects to date (287,000 hours have been consumed)**

COVID-19 PROJECT UTILIZATION

PROJECT		NUMBER OF JOBS RUN TO-DATE	NODE HOURS USED TO-DATE	NUMBER OF JOBS RUNNING NOW	NODES IN USE NOW
Cleveland Clinic COVID-19 Computational Science Projects Requiring Additional Compute Needs PI: Feixiong Cheng, Cleveland Clinic		12	67	1	1
Computational Systems Biology for Viral Detection and Classification of Emerging Infections and Rapid Identification of Causal Agents and Mechanisms PI: Daniel Jacobson, Oak Ridge National Laboratory (ORNL)		479	41,913	1	1
COVID-19: Submodular Optimization & Graph Clustering Approaches for Designing Effective Vaccination Strategies PI: Mahantesh Halappanavar, Pacific Northwest National Laboratory		159	1,751	1	256
Dissecting inhibitor impacts on viral RNA polymerase and fidelity control of RNA synthesis in SARS-CoV-2 PI: Jin Yu, University of California - Irvine		0	0	0	0
Drug Discovery for COVID-19 PI: Jeremy Smith, University of Tennessee	✓	2,245	195,970	1	16
Molecular Simulations of Complexes of SARS-CoV Spike Proteins with Human Receptor ACE2 PI: Sameer Varma, University of South Florida		516	1,812	2	4
Structural Modeling of COVID-19 with HP PI: Debsindhu Bhowmik, Oak Ridge National Laboratory (ORNL)	✓	1,126	31,759	1	1
Targeting the SARS-CoV-2 Proteome with Artificial Intelligence (AI) Driven Small Molecule Design and Screening PI: Rick Stevens, Argonne National Laboratory	✓	456	12,475	5	12
Using MD and QM/MM to improve drug candidates for nCoV-19 targets PI: Gerardo Cisneros, University of North Texas	✓	6	9	1	1
		4,999	**285,757**	**13**	**292**

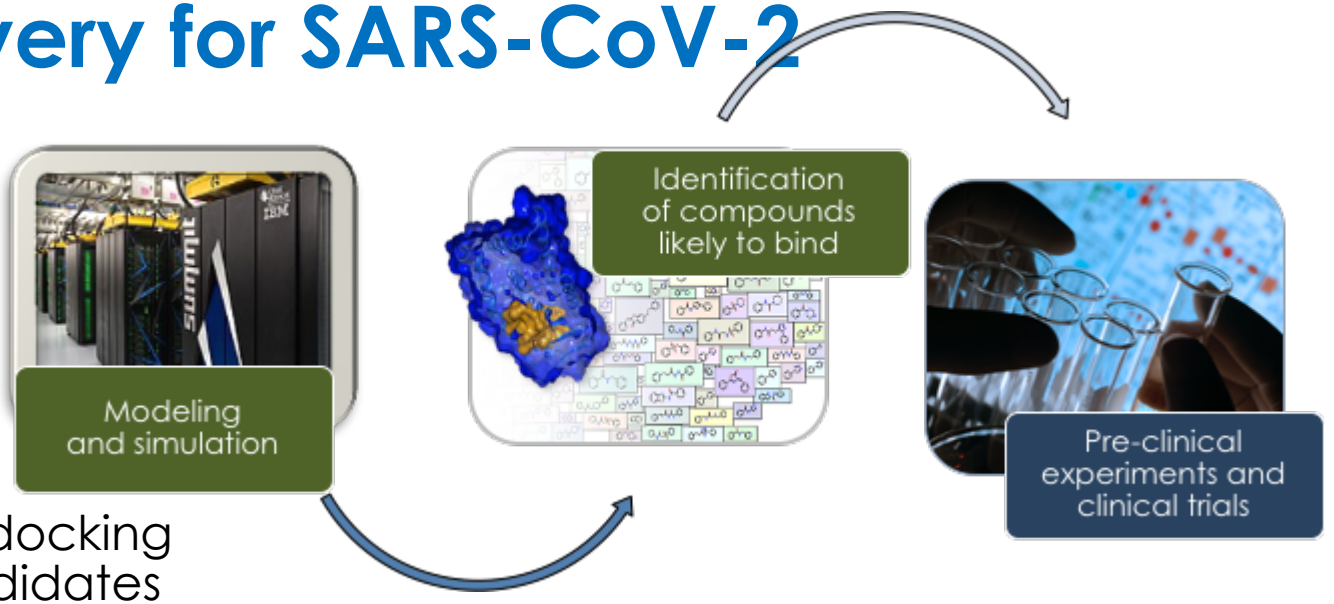
✓ *Indicates COVID-19 HPC Consortium Projects*

Rapid Antiviral Drug Discovery for SARS-CoV-2

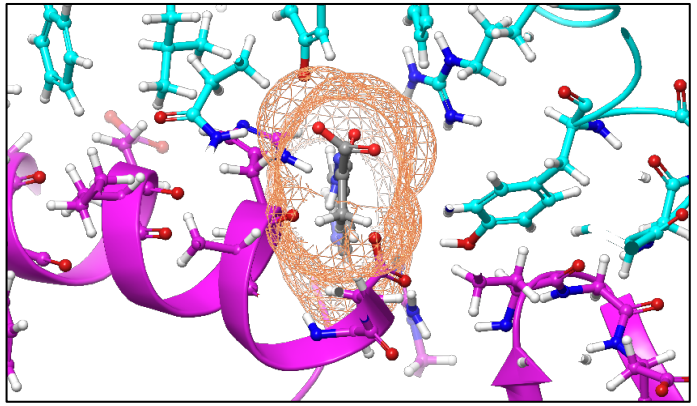
PI: Jeremy Smith,
ORNL/University of Tennessee

Status: Published on ChemRxiv

Source: HPC Consortium



- Virtual high-throughput screening ensemble docking campaigns to find the best-ranked drug candidates
- More than 8,000 compounds screened to identify those that are most likely to bind to the main "spike" protein.
- 77 compounds of interest were found that might have value in experimental studies of the virus.
- Colleen Jonsson (UT Health Science Center, Memphis) directs one of the few labs permitted to perform live virus tests and is testing the efficacy of drugs from the ORNL list on the novel coronavirus.



The compound, shown in gray, was calculated to bind to the SARS-CoV-2 spike protein, shown in cyan, to prevent it from docking to the Human Angiotensin-Converting Enzyme 2, or ACE2, receptor, shown in purple. Credit: Micholas Smith/Oak Ridge National Laboratory, U.S. Dept. of Energy

Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface

Micholas Dean Smith[1,2] and Jeremy C. Smith[1,2,3*]

* Corresponding author: smithjc@ornl.gov

[1] Center for Molecular Biophysics, The University of Tennessee/Oak Ridge National Laboratory, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831

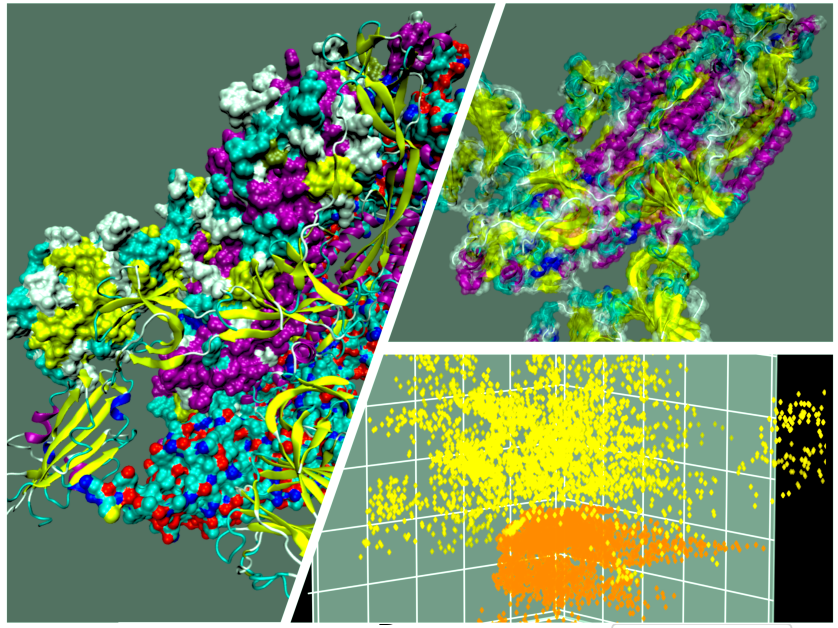
Discovering Molecular Mechanisms of the Human Coronavirus

PI: Debsindhu Bhowmik, Oak Ridge National Laboratory

Status: Published on BioARxiv

Source: ECP, HPC Consortium

- Deploying an AI-driven computational approach on Summit to provide complete molecular view of the SARS-CoV-2 S protein for assisting both vaccine and drug design efforts.
- SARS-CoV-2 S protein protomer and trimer structures displayed clear separation of clusters in the latent space, with structural transitions in specific flexible regions identified.
- These regions are promising natural targets of immune recognition suggesting therapeutic action for protein destabilization.



bioRxiv preprint doi: https://doi.org/10.1101/2020.04.17.047548. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

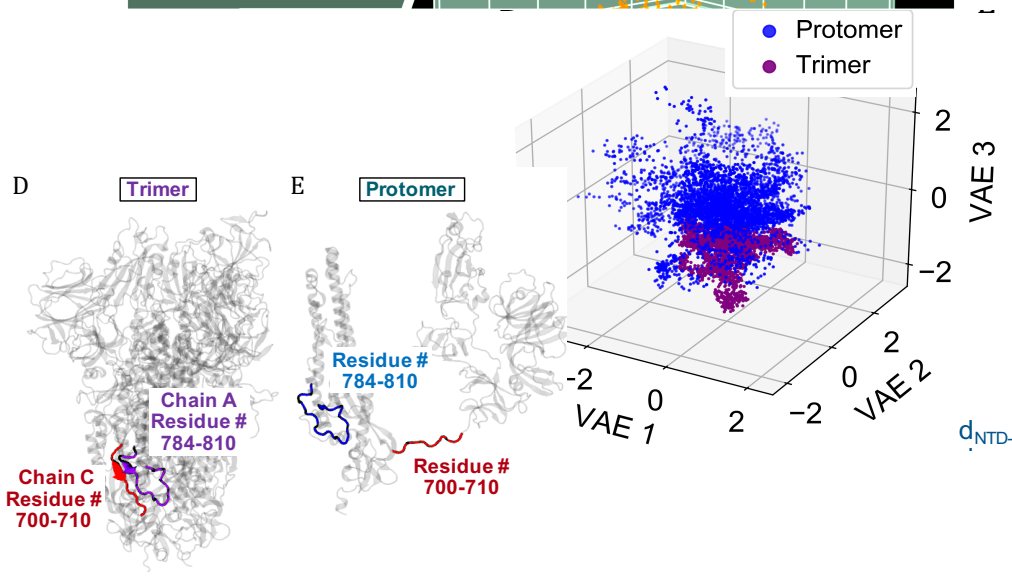
Distinct Structural Flexibility within SARS-CoV-2 Spike Protein Reveals Potential Therapeutic Targets

Serena H. Chen*, M. Todd Young*, John Gounley*, Christopher Stanley*, and Debsindhu Bhowmik*

*Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA
{chens, youngmt1, gounleyjp, stanleycb, bhowmikd}@ornl.gov

***Abstract*—The emergence and rapid worldwide spread of the novel coronavirus disease, COVID-19, has prompted concerted efforts to find successful treatments. The causative virus, severe**

understanding of the molecular pathways of SARS-CoV-2 and constituent structures. The structural biology community



OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Systems Biology of COVID-19 on Summit

PI: Dan Jacobson, Oak Ridge National Laboratory

Status: Published on BioARxiv

Source: DD

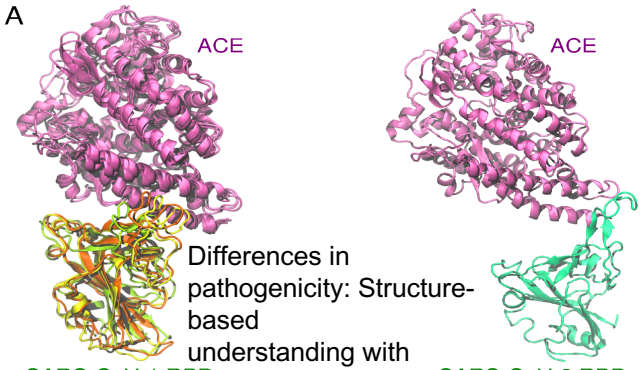
- Structure prediction and mutational analysis of the entire SARS-CoV-2 proteome.
  - Determined functions of viral proteins that appear to be key to virulence and pathogenicity
- Gene expression analysis of healthy individuals and COVID-19 patients.
  - Analysis of over 17,000 expression datasets
  - New model for routes of infection and transmission
  - Discovery of a Functional Immune Deficiency Syndrome in COVID-19 patients
  - Discovery of key molecular mechanisms of pathology
  - Determining key targets for therapeutic intervention
  - Explaining the causes for the variation in response across the population

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.06.028712. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

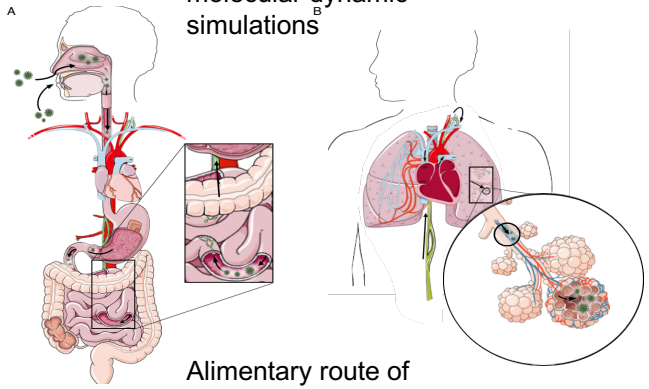
Confronting the COVID-19 Pandemic with Systems Biology

Erica Teixeira Prates, Ph.D.[1‡], Michael R. Garvin, Ph.D.[1‡], Mirko Pavicic, Ph.D.[1], Piet Jones, M.S.[1,2], Manesh Shah, M.S.[1], Christiane Alvarez[1], David Kainer, Ph.D.[1], Omar Demerdash, M.D.[1], B Kirtley Amos, M.S.[5], Armin Geiger, M.S.[1,2], John Pestian[1,3,4], Kang Jin, B.S.[3,4], Alexis Mitelpunkt, M.D.[3,4], Eric Bardes, B.S.[3,4], Bruce Aronow, Ph.D.[3,4], Daniel Jacobson, Ph.D.[1,2,6*]

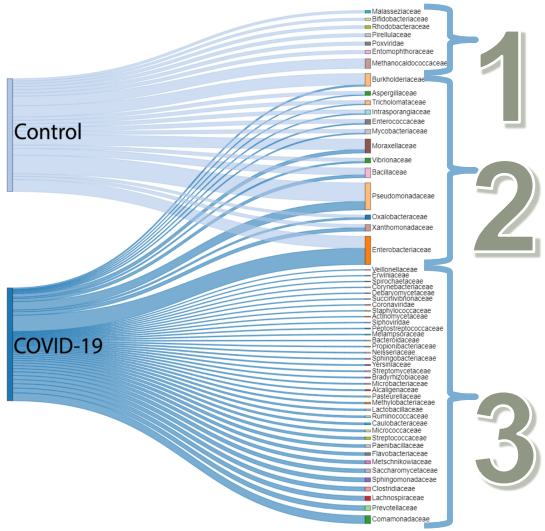
[*]Corresponding author: Daniel Jacobson, Ph.D.[1,2,6], jacobsonda@ornl.gov, 865 574 6134



Differences in pathogenicity: Structure-based understanding with molecular dynamic simulations



Alimentary route of



Shifts in the COVID-19 Microbiome

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Using MD and QM/MM to improve drug candidates for nCoV-19 targets

PI: Gerardo Cisneros, University of North Texas

Source: HPC Consortium

Status: In progress

- Perform classical molecular dynamics (MD) and hybrid quantum mechanics/molecular mechanics (QM/MM) simulations on two different drug targets of COVID-19, the SARS-CoV main protease and the NSP12 RNA-directed RNA polymerase.

AI-driven integrative biology for accelerating therapeutic discovery against SARS-CoV-2

PI: Rick Stevens / Arvind Ramanathan, Argonne National Laboratory

Source: HPC Consortium

Status: In progress

- Our proposed work seeks to address both the fundamental biological mechanisms of the virus and the disease, while simultaneously targeting the entire viral proteome to identify potential therapeutics.

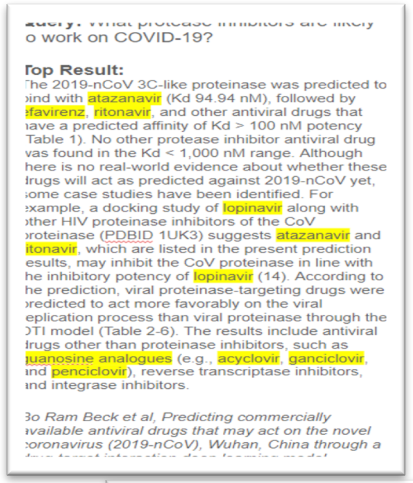
AI Text Mining of Coronavirus Scientific Publications

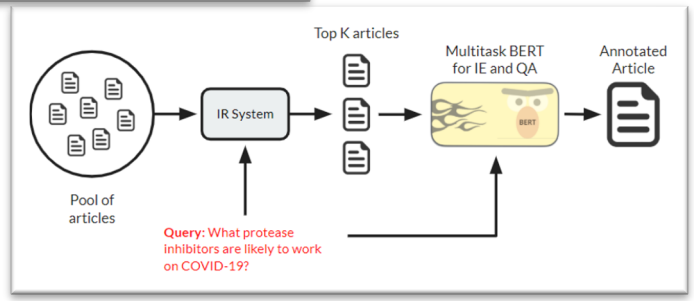
PI: Gina Tourassi, Oak Ridge National Laboratory

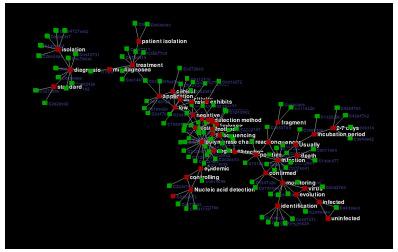
Status: In Development

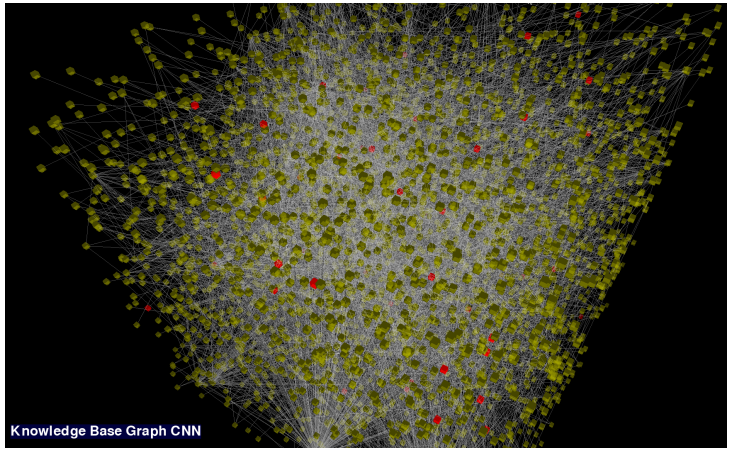
Source: ECP

- Automated information extraction and question answering system to assist knowledge summarization and knowledge discovery from COVID-19-related scientific publications.
- Given a query, return the most relevant article segments annotated with key entities, relations, and spans that address the query.
- Approach 1: Multitask deep transformer language models (BERT)
  - Semantic MEDLINE
  - PubMed Database
- Approach 2: Knowledge Base Graph Convolutional Neural Networks (KB-GCNN)
  - UMLS Concept Relation Knowledge Graph
  - GeneOntology Term Relation Knowledge Graph
- Scalable Training on Summit Supercomputer









OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Summit Node Architecture Upgrade

Adding 54 "larger memory" nodes for apps needing larger 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
  - AI/DL for medical image analysis

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™

Preparing OLCF to handle sensitive data according to national data protection standards (FISMA, HITECH and NIST SP 800-66)

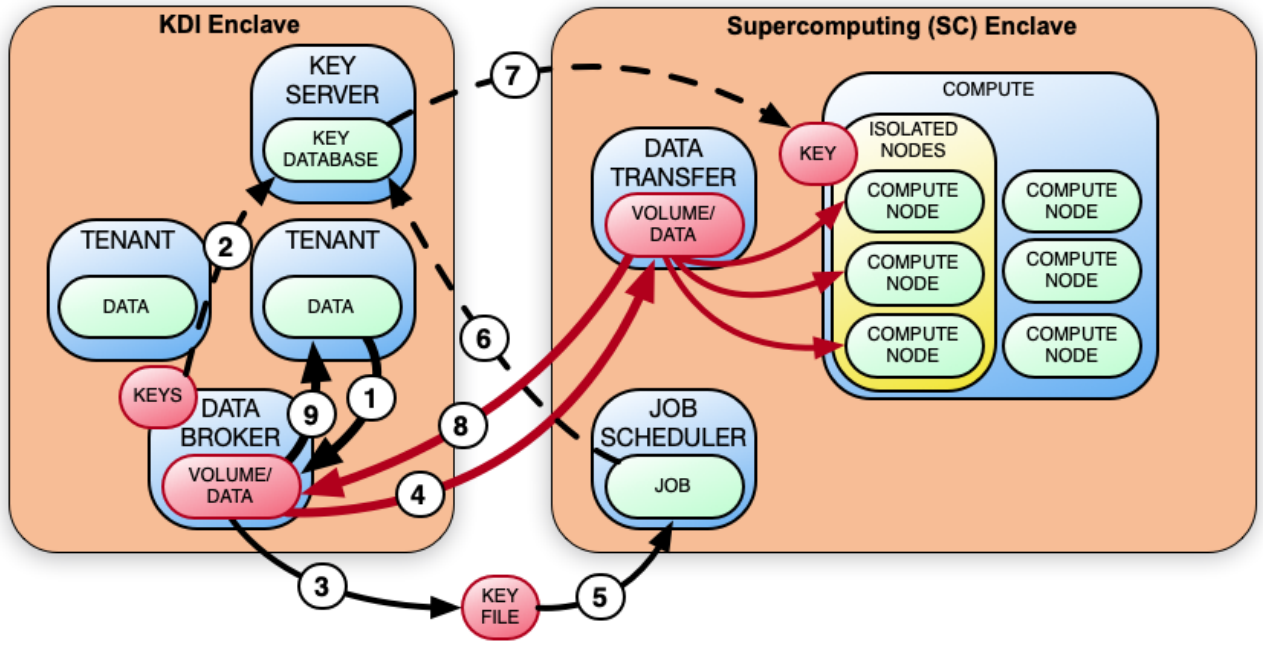
Access to sensitive data assets (e.g., VA, NCI) can help us gain COVID-19 insights

- Predict adverse events
- Predict effectiveness and safety of treatments
- Model more efficient / effective use of medications
- Model to mitigate supply shortages and optimize resource allocation

Preparing OLCF to handle sensitive data

CITADEL: Enable leadership-scale computing with protected data

- A new framework for transferring and computing PII and PHI data on Summit
  - From Knowledge Discovery Infrastructure (KDI) to Summit
- Keep the data encrypted until the last moment in a just-in-time decryption/encryption capability.
- The key for the encryption and decryption of that data is separated from the data itself to further reduce the risk of the data being accessed by unauthorized individuals.
- The ATO is handled as an addendum to the NCCS/Summit ATO. It is undergoing review by the CISO.



Questions?