Testimony of Rick Stevens
Associate Laboratory Director
for Computing, Environment and Life Sciences
Argonne National Laboratory

Before the
Subcommittee on Investigations and Oversight
Committee on Science, Space and Technology
U.S. House of Representatives

"Repurposing Therapeutic Drugs for COVID-19: Research Challenges and Opportunities"

June 19, 2020

Chairman Foster, Ranking Member Norman, and Members of the Subcommittee, thank you for the opportunity to participate in today's discussion about Drug Repurposing for COVID-19 and the role of the U.S. Department of Energy (DOE) National Laboratories.

I am Rick Stevens, the Associate Laboratory Director responsible for Computing, Environment and Life Sciences research at Argonne National Laboratory and a Professor of Computer Science at the University of Chicago. My research focuses on finding new ways to advance science and health outcomes using computation. This includes helping to develop the Exascale computing initiative and the emerging AI for Science initiative in the DOE. Currently I lead research projects that are developing and using High-Performance Computing (HPC) and Artificial Intelligence (AI) in infectious diseases, cancer and in other areas of science.

I have been associated with the DOE national laboratory system for nearly forty years. My testimony today is on behalf of my work at Argonne National Laboratory and does not represent the views of the Department of Energy.

Let me start by telling you a bit about the virus we are fighting.

The virus is called **SARS-CoV-2,** and the disease it causes in humans is called **COVID-19**.

The SARS-CoV-2 virus is a coronavirus closely related (> 90% similar) to the SARS and MERS coronaviruses that emerged in 2003 and 2012 and is more distantly related (50% similar) to the coronaviruses that cause the common cold.

Its natural host is a bat.

There is no evidence to suggest it was engineered.

The virus was probably transmitted to humans via an intermediate host similar to SARS (civet) and MERS (camel), but we currently don't know precisely what species that intermediate animal was.

Globally, the biomedical community has sequenced and analyzed over 30,000 genome sequences of the SARS-CoV-2 virus, and while there are some differences between samples (isolates), they are minor and do not seem to give rise to significant differences in virulence or mortality.

The SARS-CoV-2 virus (like all viruses) is not alive.

It cannot replicate outside of the host and is dependent entirely on the host to complete its replication cycle.

The particle of the virus is quite small, on the order of 100nm in diameter or about 1000th the thickness of a human hair and is not visible to the human eye.

In this very small volume, the virus packs a single-stranded RNA genome that codes for about 30 proteins.

Of those 30 proteins, about two-thirds of them are used to commandeer the host cell into helping it replicate. The other proteins are used to package the viral RNA and build the virus structure. Some proteins have multiple functions.

Each infected host cell produces about 1000 new virus particles in about 10 hours.

Of the 30 proteins in the virus genome, about 10 of them are plausible drug targets.

Thanks to the work at the light-sources of the DOE national laboratories and elsewhere, we have good atomic level structures for most of these proteins. In fact, most of these proteins are highly similar to the proteins from the SARS 1 virus. The SARS 1 virus has been studied by a small global research community for over a decade and these research results provided a critical head-start to understanding SARS-CoV-2.

The virus proteins also interact with many human proteins. Some virus proteins turn off functions in the host that would normally help recognize the virus and would then activate aspects of the immune system. In other cases, the viral proteins regulate host processes to help create a more suitable environment for the virus to replicate.

All told, perhaps 200-300 host proteins are involved in some form of interaction with the virus proteins, and some of those (perhaps 10-20) are plausible drug targets.

Globally, there are about 2500 drugs that are approved by national authorities and are generally available in the marketplace.

These existing "potentially repurposable drugs" are obvious candidates for consideration for therapeutics for COVID-19 since they have already passed through numerous tests for safety and side effects, and the dosing for human use is reasonably well understood.

What is not deeply understood (certainly not at the beginning of this pandemic) is which of these existing “on the market” drugs are active against targets in the virus or host to treat COVID-19.

I am personally involved in research projects that are actively working on discovering small molecule antiviral drugs for COVID-19, which include repurposing existing drugs. Resources for these efforts come from NIH and the Department of Energy, including crucial support under the CARES Act.

My research partners include teams at the University of Chicago and a nine-laboratory consortium of DOE laboratories (Argonne, Oak Ridge, Berkeley, Brookhaven, Livermore, Los Alamos, Pacific Northwest, Sandia, and SLAC).

In these collaborations, my role has been to lead the efforts to using large-scale computing and artificial intelligence to identify existing drugs and new compounds for treating COVID-19.

I have a wonderful team of about 100 people across many institutions deeply engaged in different aspects of the computational effort.

To address the challenge of therapeutics development for COVID-19 requires the creation of interdisciplinary teams of scientists that include physicians, virologists, synthetic and medicinal chemists, molecular biologists, biochemists, structural biologists, computational biologists, bioinformaticians, computer scientists, data scientists, drug designers, molecular engineers, and AI researchers.

These teams started forming in March 2020 when it became clear that COVID-19 was becoming a pandemic.

Our early goal was to use computational screening of existing drugs to quickly determine candidates for experimentation, and to get those priority hits into the experimental pipelines as fast as possible. To give you a sense of the urgency, in some cases, our UChicago colleagues synthesized molecules in house since it was faster than ordering them.

Once those initial screens were underway, we started focusing on and are continuing a longer-term effort to investigate new compounds prioritized specifically for effectiveness on COVID-19 related targets.

In conducting this work, we are using three significant classes of national user facilities:

1. **The DOE and National Science Foundation (NSF) supercomputers:** Argonne Leadership Computing Facility (ALCF), Oak Ridge Leadership Computing Facility (OLCF), National Energy Research Scientific Computing Center (NERSC), Lawrence Livermore National Laboratory (LLNL), Texas Advanced Computing Center (TACC), San Diego Supercomputer Center (SDSC), Pittsburgh Supercomputing Center (PSC), National Center for Supercomputing Applications (NCSA), etc.

These supercomputers are being used to model virus proteins, protein interactions, virtual drug screening and to build and run AI models to predict binding and other properties of potential drugs.

2. **The DOE supported x-ray light sources and neutron sources:** Advanced Photon Source (APS) at Argonne, Advanced Light Source (ALS) at Lawrence Berkeley, Stanford Synchrotron Radiation Light Source (SSRL) at SLAC, National Synchrotron Light Source II (NSLS II) at Brookhaven, and Spallation Neutron Source (SNS) at Oak Ridge National Laboratories.

The light-sources and neutron source are used to determine structures of proteins, protein complexes and the position and orientation of drugs bound to the active sites in proteins.

And critically the:

3. **National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases (NIAID) supported Biocontainment Laboratories (Biosafety Level 3):** the Howard T Ricketts Laboratory operated by the University of Chicago and the University of Tennessee Health Science Center Regional Biocontainment Laboratory, where experimental screening on active "live" virus can take place.

The biocontainment laboratories are critical since without them we are limited to working only on biochemical assays and not full virus replication assays.

At Argonne National Laboratory we have developed and published a database of over 4 billion known compounds from many sources and the associated information to support Artificial Intelligence methods to accelerate computational screening. Over 2 million hours of computer time were used to build this database.

We have also identified a priority library of about 10 million molecules that are readily available for experimental validation. This library has been used to screen drug targets and to identify priority compounds that can be quickly acquired for experimental validation.

My team is using the world's fastest supercomputers housed at DOE Leadership Computing Facilities at Argonne and Oak Ridge National Laboratories, and other machines across the country, to virtually screen these 4 billion compounds on dozens of COVID-19 drug targets. This is something that would not have been possible just a few

years ago and is not possible without the advances in AI and machine learning, supported by the DOE's Office of Science

These computational efforts have been ongoing for several months now and hundreds of top scoring molecules (including many drug repurposing candidates) have been experimentally screened for antiviral activity, and thousands more are in the pipeline for experimental screens.

An important point to make here is that while repurposing is a fast route to possible treatments, in general, the repurposing candidates are not the ideal molecules to bind to COVID-19 specific targets.

Long-term, it is highly likely that the most effective drugs for treating COVID-19 will need to be purpose built, and this will take time and longer-term investment.

Our goal in the short term is to try to do as much up-front work as possible on these molecules so that the private sector—that would ultimately need to produce new or repurposed drugs—has reasonable places to start.

We want to provide candidate lead compounds that have demonstrated activity on important COVID-19 targets.

A few observations on how well this is going and some recommendations for the future:

- The national research community is working well together and has quickly identified and integrated capabilities from DOE, NIH, and many universities to fight COVID-19.
- The DOE has created a National Virtual Biotechnology Laboratory effort, supported by CARES Act funding, that is coordinating COVID-19-related efforts at all the DOE labs; this is working well.
- Interagency communication paths are open and senior leaders at the agencies are coordinating.
- The national laboratory efforts are coordinating with Pharma and have ongoing conversations about methods and potential hand offs on drug candidates.

As these national collaborations continue and we look to continue to make an impact in addressing COVID-19 and future potential pandemics, I offer a few considerations and recommendations for going forward:

- Experimental assay development needed to validate drug repurposing candidates is difficult and time consuming and is a bottleneck for antiviral drug development efforts. These assays should be a high-priority part of long-term research programs and considered critical national infrastructure.

- Once assays exist for a new target, baseline experimental screens of existing drugs can be done relatively quickly, inexpensively, and in parallel with

computational work and would provide important datasets for AI efforts. Standing libraries of repurposing compounds are available and should be part of standard inventory at regional biocontainment laboratories.

- The national network of research capabilities that has been assembled to address the COVID-19 challenge should be maintained and tuned for future epidemics and emerging pathogens. Modern therapeutics development needs tight integration of high-throughput experiments, structure determination, assay and methods development, data analysis, AI, and simulation.

- Standing interagency agreements should be put in place to streamline funding transfers, materials transfers, personnel exchange, and data sharing. Certainly DOE, NIH, NSF, DOD all have capabilities that need rapid coordination and access. A collaboration network is needed to exercise these connections on an ongoing basis.

Finally, SARS-CoV-2 and COVID-19 is not the first pandemic and it is not likely the last one we will see. Zoonotic diseases (those that originate in animals and infect humans) are likely to become more common as the pressures of economic and social development and land use changes push larger human populations deeper into traditionally unsettled areas and cause more human contact with diverse animal species.

I believe we need a systematic approach to sampling and understanding the diversity of pathogens carried by animals in natural environments and to use this information to understand the likelihood of transition to human hosts, their pathology, and to get a head start on therapeutics, and in general to be more prepared for the next outbreak.

We should be mounting internationally coordinated expeditions to sample the many thousands of known carrier species for microbial pathogens and use this information to manage risk.

The cost of proactive risk management for emerging pathogens will be orders of magnitude less than the cost of another pandemic and the tragic loss of life.

The DOE laboratories exist to do the science and to create technology for the national grand challenges. As organizations they are specifically designed to support fast moving, interdisciplinary science, precisely the kind of science that is needed to respond to national challenges like COVID-19.

Looking forward I believe we should consider how to structure a permanent role for the National Laboratories to augment and support the national response to future infectious disease challenges.

Thank you. I would be happy to answer any questions you or other members of the committee may have.