How supercomputing is taking on the COVID-19 pandemic

In "The Hitchhiker's Guide to the Galaxy" by Douglas Adams, the haughty supercomputer Deep Thought is asked whether he can find the answer to the ultimate question concerning life, the universe and everything. He replies that, yes, he can do it, but it's tricky and he'll have to think about it. When asked how long it will take him he replies, "Seven-and-a-half million years. I told you I'd have to think about it."

Real-life supercomputers are being asked somewhat less expansive questions but tricky ones nonetheless: how to tackle the COVID-19 pandemic. They're being used in many facets of responding to the disease, including to predict the spread of the virus, to optimize contact tracing, to allocate resources and provide decisions for physicians, to design vaccines and rapid testing tools and to understand sneezes. And the answers are needed in a rather shorter time frame than Deep Thought was proposing.

The largest number of COVID-19 supercomputing projects involves designing drugs. It's likely to take several effective drugs to treat the disease. Supercomputers allow researchers to take a rational approach and aim to selectively muzzle proteins that SARS-CoV-2, the virus that causes COVID-19, needs for its life cycle.

The viral genome encodes proteins needed by the virus to infect humans and to replicate. Among these are the infamous spike protein that sniffs out and penetrates its human cellular target, but there are also enzymes and molecular machines that the virus forces its human subjects to produce for it. Finding drugs that can bind to these proteins and stop them from working is a logical way to go.

I am a molecular biophysicist. My lab, at the Center for Molecular Biophysics at the University of Tennessee and Oak Ridge National Laboratory, uses a supercomputer to discover drugs. We build three-dimensional virtual models of biological molecules like the proteins used by cells and viruses, and simulate how various chemical compounds interact with those proteins. We test thousands of compounds to find the ones that "dock" with a target protein. Those compounds that fit, lock-and-key style, with the protein are potential therapies.

The top-ranked candidates are then tested experimentally to see if they indeed do bind to their targets and, in the case of COVID-19, stop the virus from infecting human cells. The compounds are first tested in cells, then animals, and finally humans. Computational drug discovery with high-performance computing has been important in finding antiviral drugs in the past, such as the anti-HIV drugs that revolutionized AIDS treatment in the 1990s.

World's most powerful computer

Since the 1990s the power of supercomputers has increased by a factor of a million or so. Summit at Oak Ridge National Laboratory is presently the world's most powerful supercomputer, and has the combined power of roughly a million laptops. A laptop today has roughly the same power as a supercomputer had 20-30 years ago.

However, in order to gin up speed, supercomputer architectures have become more complicated. They used to consist of single, very powerful chips on which programs would simply run faster. Now they consist of thousands of processors performing massively parallel processing in which many calculations, such as testing the potential of drugs to dock with a pathogen or cell's proteins, are performed at the same time. Persuading those processors to work together harmoniously is a pain in the neck but means we can quickly try out a lot of chemicals virtually.

Further, researchers use supercomputers to figure out by simulation the different shapes formed by the target binding sites and then virtually dock compounds to each shape. In my lab, that procedure has produced experimentally validated hits – chemicals that work – for each of 16 protein targets that physician-scientists and biochemists have discovered over the past few years. These targets were selected because finding compounds that dock with them could result in drugs for treating different diseases, including chronic kidney disease, prostate cancer, osteoporosis, diabetes, thrombosis and bacterial infections.

Billions of possibilities

So which chemicals are being tested for COVID-19? A first approach is trying out drugs that already exist for other indications and that we have a pretty good idea are reasonably safe. That's called "repurposing," and if it works, regulatory approval will be quick.

But repurposing isn't necessarily being done in the most rational way. One idea researchers are considering is that drugs that work against protein targets of some other virus, such as the flu, hepatitis or Ebola, will automatically work against COVID-19, even when the SARS-CoV-2 protein targets don't have the same shape.

The best approach is to check if repurposed compounds will actually bind to their intended target. To that end, my lab published a preliminary report of a supercomputer-driven docking study of a repurposing compound database in mid-February. The study ranked 8,000 compounds in order of how well they bind to the viral spike protein. This paper triggered the establishment of a high-performance computing consortium against our viral enemy, announced by President Trump in March. Several of our top-ranked compounds are now in clinical trials.

Our own work has now expanded to about 10 targets on SARS-CoV-2, and we're also looking at human protein targets for disrupting the virus's attack on human cells. Top-ranked compounds from our calculations are being tested experimentally for activity against the live virus. Several of these have already been found to be active.

Also, we and others are venturing out into the wild world of new drug discovery for COVID-19 – looking for compounds that have never been tried as drugs before. Databases of billions of these compounds exist, all of which could probably be synthesized in principle but most of which have never been made. Billion-compound docking is a tailor-made task for massively parallel supercomputing.

Dawn of the exascale era

Work will be helped by the arrival of the next big machine at Oak Ridge, called Frontier, planned for next year. Frontier should be about 10 times more powerful than Summit. Frontier will herald the "exascale" supercomputing era, meaning machines capable of 1,000,000,000,000,000,000 calculations per second.

Although some fear supercomputers will take over the world, for the time being, at least, they are humanity's servants, which means that they do what we tell them to. Different scientists have different ideas about how to calculate which drugs work best – some prefer artificial intelligence, for example – so there's quite a lot of arguing going on.

Hopefully, scientists armed with the most powerful computers in the world will, sooner rather than later, find the drugs needed to tackle COVID-19. If they do, then their answers will be of more immediate benefit, if less philosophically tantalizing, than the answer to the ultimate question provided by Deep Thought, which was, maddeningly, simply 42.

Jeremy Smith is the Governor's Chair, Biophysics, University of Tennessee

This article originally appeared at the Conversation and has been republished here with permission.