CLINICAL IMPLICATIONS OF BASIC RESEARCH

Elizabeth G. Phimister, Ph.D., *Editor*

How to Discover Antiviral Drugs Quickly

Jerry M. Parks, Ph.D., and Jeremy C. Smith, Ph.D.

We urgently need effective drugs for coronavirus disease 2019 (Covid-19), but what is the quickest way to find them? One approach that sometimes seems akin to a "Hail Mary" pass in American football is to hope that drugs that have worked against a different virus (such as hepatitis C or Ebola) will also work against Covid-19. Alternatively, we can be rational and specifically target proteins of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) so as to interrupt its life cycle.

The SARS-CoV-2 genome encodes approximately 25 proteins that are needed by the virus to infect humans and to replicate (Fig. 1). Among these are the notorious spike (S) protein, which recognizes human angiotensin-converting enzyme 2 in the initial stage of infection; two proteases, which cleave viral and human proteins; the RNA polymerase, which synthesizes viral RNA; and the RNA-cleaving endoribonuclease. Finding drugs that can bind to the viral proteins and stop them from working is a logical way forward and the priority of many research laboratories.

One approach toward this goal involves mimicking nature with the use of computational structure-based drug discovery (Fig. 2). In this process, computers "dock" trial compounds into binding sites in three-dimensional models of the protein targets. The binding affinities of the compounds are calculated with the use of physics-based equations that quantify the interactions between the drug and its target. The top-ranked compounds are then tested experimentally to see if they do indeed bind and have the required downstream effects (such as stopping viral infectivity) on cells and in animal models.

Structure-based drug discovery has been important in finding antiviral drugs, an example being nelfinavir, discovered in the 1990s, to treat human immunodeficiency virus (HIV) infection. Unfortunately, though, at that time the process was relatively inefficient: calculations were inaccurate and computers so feeble that only about 100 compounds could be docked at a time. Moreover, both the target and the drug had to be held rigid in the docking process in a lock-and-key approach. Rigid docking does not often take place in real life, because proteins undergo thermally driven internal motions that lead to fluctuating binding-site shapes.

Since the 1990s, the power of supercomputers has increased by a factor of a million or so. Rigid docking of over a billion compounds can now be performed in a few days. Thus, virtual high-throughput screening is outperforming equivalent experimental high-throughput screening and can rapidly identify very tightly binding compounds.[1] Furthermore, molecular-dynamics simulations can be performed to calculate internal protein motions, and candidate drugs can be screened through a process that uses the different shapes formed by the binding site in a procedure known as "ensemble docking."[2] This approach is more realistic than rigid docking and has been successful, for example, in serving the HIV drug-discovery efforts from the 2000s onward. In our own laboratory, ensemble docking has produced experimentally validated hits against each of the 16 protein targets presented to us over the past few years.

Modern supercomputers such as the Summit supercomputer at Oak Ridge National Laboratory, which is currently the world's most powerful, perform massively parallel processing in which many calculations are performed at the same time. This enables molecular-dynamics simulations of many replicas of the target to be run in

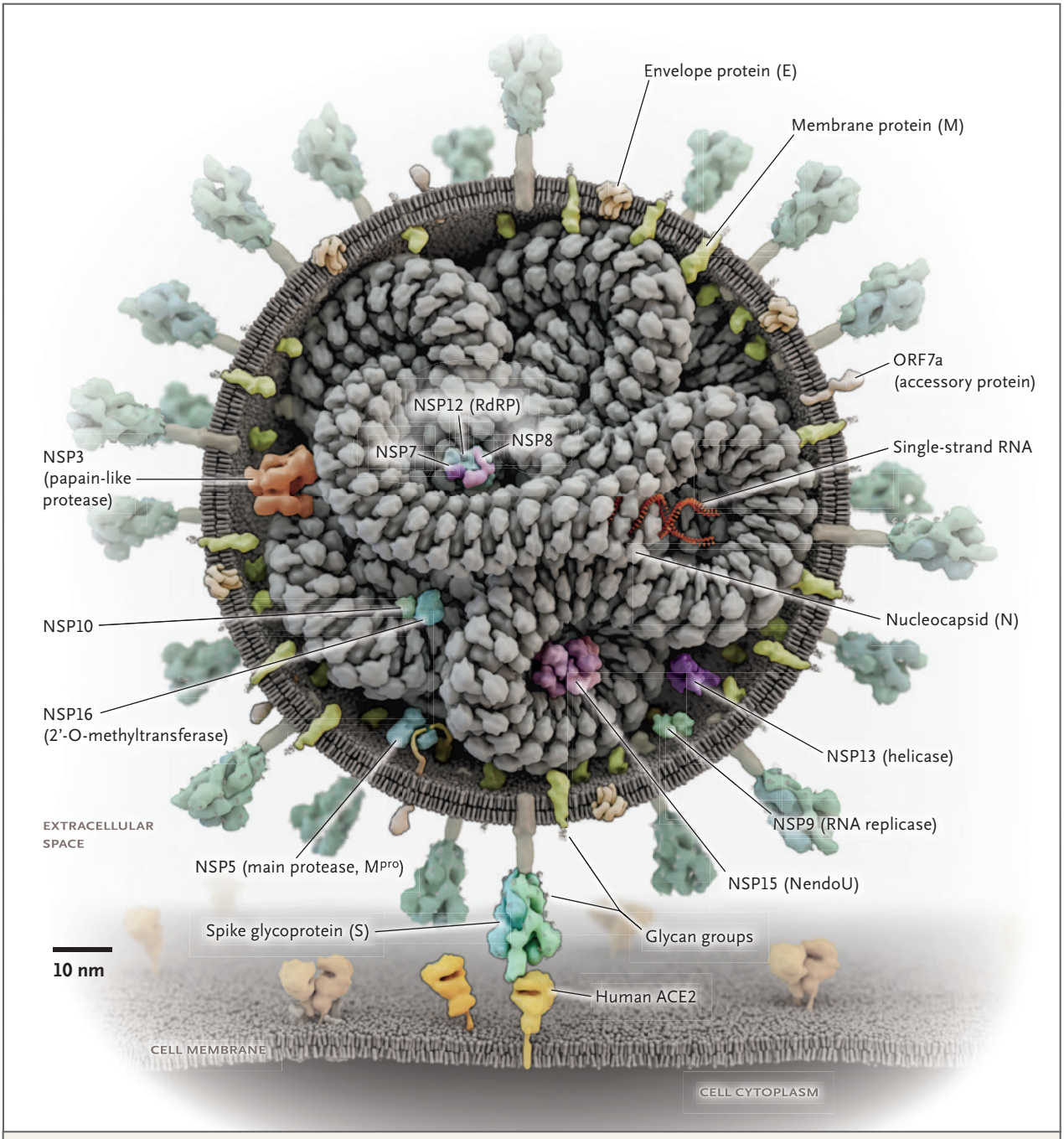

Figure 1. The SARS-CoV-2 Virion and Its Proteins.

Although all the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins are potential drug targets, some are likely to be more easy to find a drug against, in part because they play principal roles in the viral lifecycle and also lack human protein homologues. Examples include the spike glycoprotein, the papain-like protease, the chymotrypsin-like main protease, and the RNA-dependent RNA polymerase. A list of the Worldwide Protein Data Bank identifiers of the structures shown is provided in the Supplementary Appendix, available with the full text of this article at NEJM.org. ACE2 denotes angiotensin-converting enzyme 2, NSP nonstructural protein, ORF open reading frame, and RdRP RNA-dependent RNA polymerase.

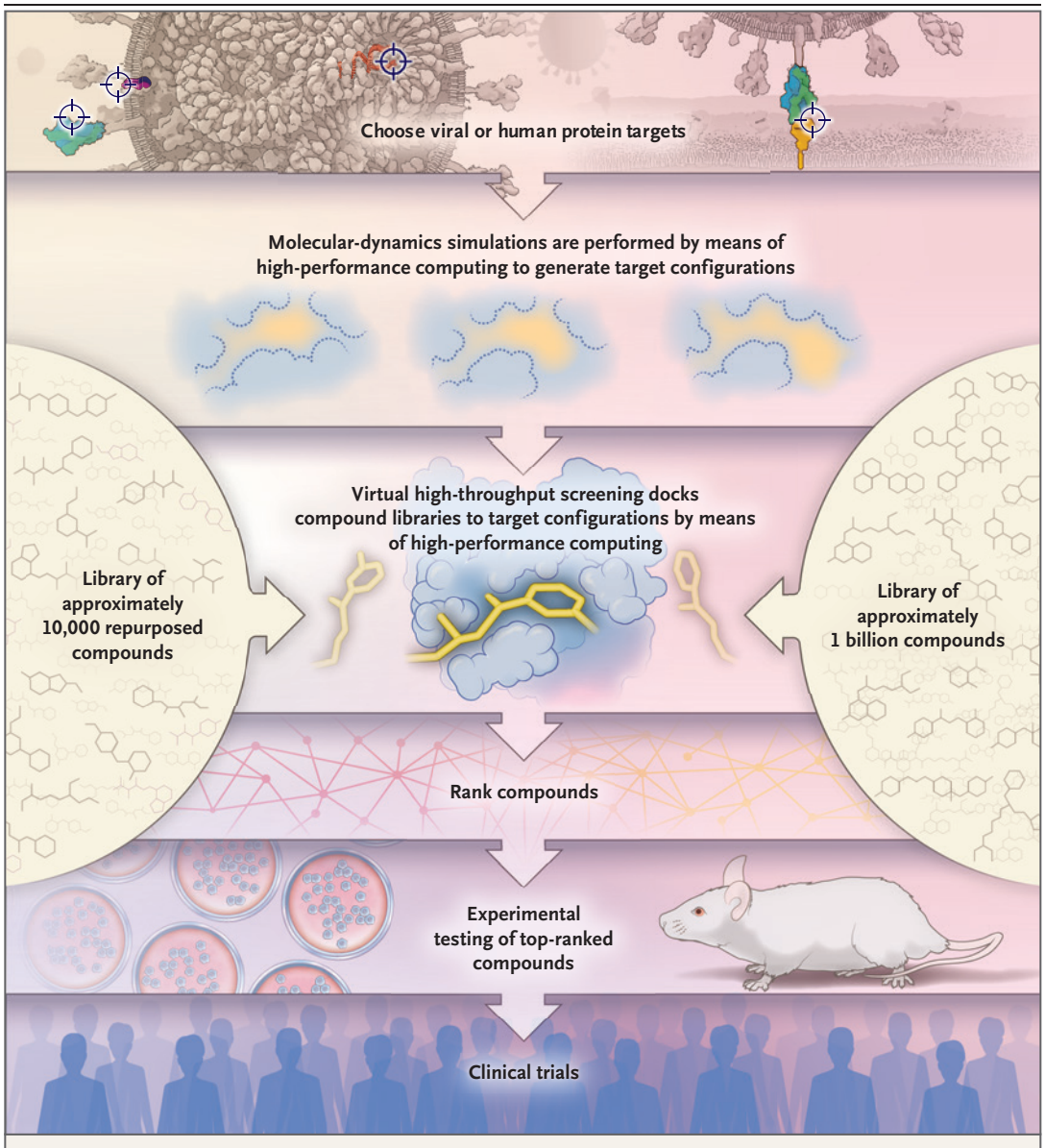

Figure 2. Computational High-Throughput Ensemble Docking in Drug Discovery.
Docking a repurposing library can lead to the discovery of rapidly deployable compounds. Larger libraries, which can contain billions of compounds, are useful for quickly discovering new compounds as yet untested in humans.

parallel, each exploring a slightly different conformational space. Thus, a comprehensive simulation model of a SARS-CoV-2 protein drug target can be obtained with the use of Summit in a day, whereas it would take months with the use of a typical computer cluster. Supercomputers are also used in rapid parallel docking of large databases of compounds. The structure-based drug-discovery field is thus primed for quick results.

So, what is happening now? The laborious, decade-long, classic pathway for the discovery and approval of new drugs could hardly be less well suited to the present pandemic. Repurposing existing drugs offers a potentially rapid mechanism to deployment, since the safety pro-

files are known. Therefore, a preliminary report of a supercomputer-driven ensemble docking study of a repurposing compound database to the viral S protein was published on a preprint server in mid-February, with 8000 compounds ranked according to the calculated binding affinity to the receptor-binding domain of the S protein.[3] Top-ranked compounds from the original S-protein virtual screen are being tested for activity against the live virus. The results will inform future calculations in a speedy, iterative process.

However, in the surreal, accelerated world of Covid-19 research, advances are quickly out of date. Many new experimental three-dimensional structures of the S protein and other viral targets are being reported in quick succession, a process that requires the simulations and docking to be refined and repeated. Artificial intelligence is being used to predict drug binding. Different types of experimental laboratory screening programs have been set up all over the world and are ramping up. Meanwhile, for several SARS-CoV-2 proteins, the virtual high-throughput screening and ensemble docking pipeline is in full production mode, both on supercomputers and with the use of vast cloud-computing resources. None of this guarantees success within any given time frame, but a combination of rationality, scientific insight, and ingenuity with the most powerful tools available will give us our best shot.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Thomas Splettstoesser, Ph.D., for compiling the list of the Worldwide Protein Data Bank identifiers of the structures shown in Figure 1 and for contributing to the conceptual design of the figure.

From the University of Tennessee–Oak Ridge National Laboratory Center for Molecular Biophysics, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge (J.M.P., J.C.S.), and the Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville (J.C.S.).

This article was published on May 20, 2020, at NEJM.org.

1. Gorgulla C, Boeszoermenyi A, Wang ZF, et al. An open-source drug discovery platform enables ultra-large virtual screens. Nature 2020 March 9 (Epub ahead of print).
2. Amaro RE, Baudry J, Chodera J, et al. Ensemble docking in drug discovery. Biophys J 2018;114:2271-8.
3. Smith MD, Smith JC. Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. ChemRxiv. March 11, 2020 (https://chemrxiv.org/articles/Repurposing_Therapeutics_for_the_Wuhan_Coronavirus_nCov-2019_Supercomputer-Based_Docking_to_the_Viral_S_Protein_and_Human_ACE2_Interface/11871402).

DOI: 10.1056/NEJMcibr2007042