

The coronavirus might have weak spots—and machine learning could help find them

May 18 2020, by Roberto Molar Candanosa



Amino acids are critical to the structure of proteins, which are often visualized as 3D ribbon structures. Northeastern biochemists are studying the chemistry of amino acids within SARS-CoV-2 to predict the reactions they enable. Credit: Ruby Wallau/Northeastern University

Chemically speaking, proteins might be the most sophisticated molecules

out there. Millions of different kinds of them live within our cells and work together as a fine-tuned orchestra catalyzing the biochemical reactions that keep us alive.

Few things in the world would function without proteins—not the cells within your body, and certainly not SARS-CoV-2, the coronavirus responsible for COVID-19.

The proteins in the coronavirus facilitate its remarkable ability to infect human cells without resulting in visible symptoms of COVID-19 for long periods of time. That's why researchers around the world have been investigating the roles of each of the 29 proteins packed inside SARS-CoV-2.

By learning more about each of those proteins at the [molecular level](#), researchers want to pin down the exact parts of the SARS-CoV-2 proteins that enable it to bind itself to other proteins on the surface of human cells and enable the virus to replicate. The idea is to inhibit those [chemical reactions](#) right from the start, and render the coronavirus ineffective.

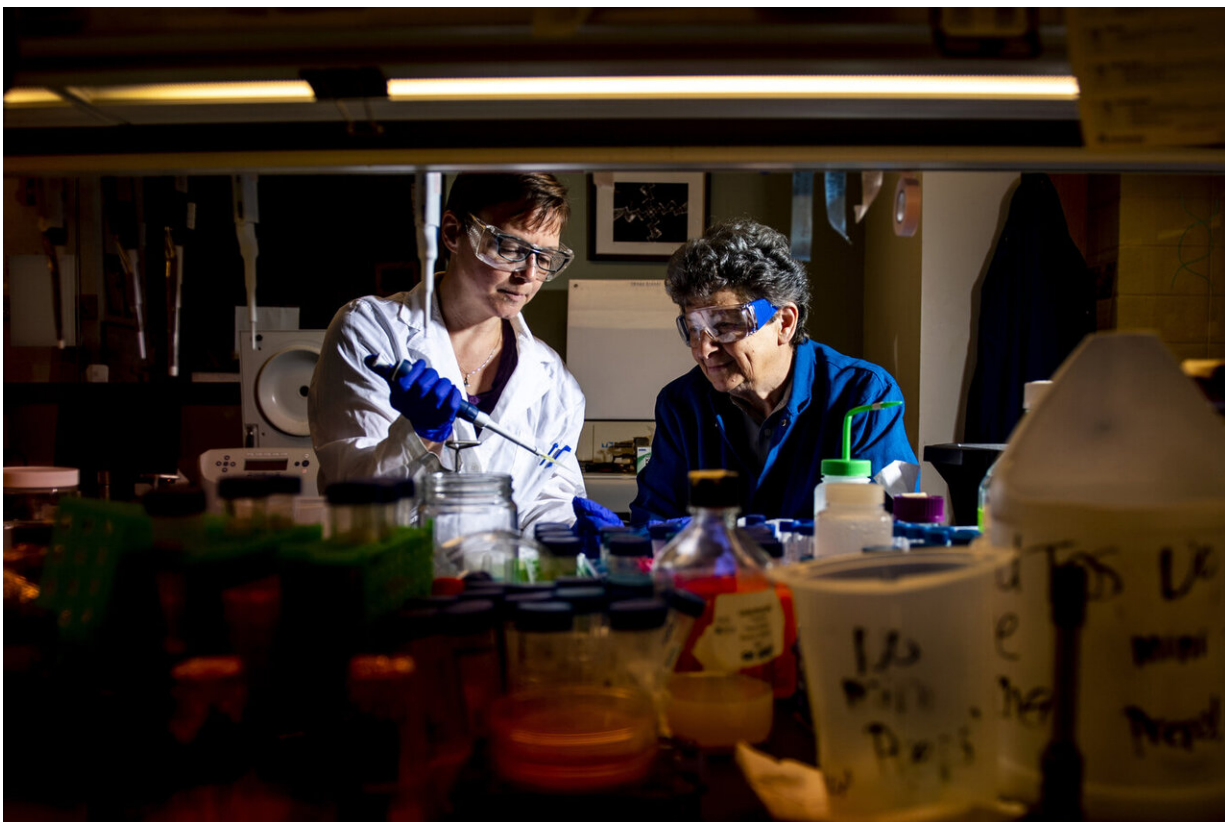
To analyze those [protein](#) interactions, Northeastern researchers are bringing another set of tools to study the coronavirus' proteins down to their amino acids, the building blocks of all proteins.

Mary Jo Ondrechen, a professor of chemistry and [chemical biology](#), wants to identify all of the amino acids responsible for the abilities of the coronavirus to infect and thrive at the expense of human cells. Together with Penny Beuning, a professor of chemistry and chemical biology, Ondrechen recently received a grant from the National Science Foundation to use machine learning algorithms and experimental lab work to do just that.

Proteins are long chains of molecules that function through cascading interactions with amino acids from other proteins. But those interactions don't always occur in the same place within the structure of a protein where the protein carries out its chemical reaction. Often, although the interactions happen outside of that site, they still control the reaction. A specific site within a protein can also control the action of different proteins, helping or hindering a specific chemical reaction.

Changes in protein behavior resulting from these networks of interactions, or from preventing interactions, are known as allosteric regulation. Ondrechen's algorithm predicts many of these and other types of interactions based on the specific molecular structures of proteins.

Research led by her and Beuning could help researchers gain a better understanding of the biochemistry of SARS-CoV-2, and serve as the basis for developing new drugs to inhibit its infectious abilities.



Mary Jo Ondrechen and Penny Beuning, professors of chemistry and chemical biology at Northeastern, are using machine learning to study the proteins within the coronavirus that enable it to infect human cells and replicate. Credit: Ruby Wallau/Northeastern University

Researchers around the world have been rushing to develop new chemicals that show promise as compounds that could hinder the coronavirus by interacting with its main active proteins.

Still, scientists are just beginning to understand many of the coronavirus' proteins. And, Ondrechen says, there might be sites within those poorly understood proteins that researchers might be failing to notice.

The program, which Ondrechen's lab invented in 2009, analyzes the

chemical properties of each of the individual amino acids within a protein. It could predict the roles of important but subtle interactions in SARS-CoV-2 involving [amino acids](#) that aren't directly linked to the main reaction sites, and which would be too difficult to analyze with conventional bioinformatic research.

"In the main protease, everybody knows where the catalytic site is, in the RNA transferase, everybody knows where the catalytic site is," Ondrechen says. "Our technology is special because we could predict exo-sites, allosteric sites, and other binding sites or interaction sites that can control."

The program will run those predictions against databases that include tens of thousands of compounds with anti-viral properties and compounds found in food, all in a major attempt to find proteins that might hit the predicted sites of protein interaction.

Once the program runs the computational analysis to find candidate proteins to inhibit SARS-CoV-2, it will guide Beuning's experimental tests in her lab.

"We'll be looking at the protein level: Do the compounds actually bind those proteins, and do they modulate the activity of the protein?" Beuning says. "Ideally, they would inhibit the activity of the protein, and then impair the virus."

For the past 10 years, Ondrechen and Beuning have been combining their computational and experimental power to understand such questions as how proteins control the production of our DNA, and how proteins enable our bodies to carry out some of the most important metabolic functions.

Now, they are planning to move as fast as possible to identify important

protein interactions in SARS-CoV-2, test them in the lab, and move on with further tests in live organisms.

"Our plans are to finish in six months," Ondrechen says. "If we come up with interesting compounds in vitro, hopefully we can find a collaborator that could do in vivo testing."

Provided by Northeastern University

Citation: The coronavirus might have weak spots—and machine learning could help find them (2020, May 18) retrieved 11 July 2024 from <https://phys.org/news/2020-05-coronavirus-weak-spotsand-machine.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.