

# Model analyzes how viruses escape the immune system

January 14 2021

---



A colored scanning electron micrograph of the SARS-CoV-2 virus. Credit: NIAID

One reason it's so difficult to produce effective vaccines against some viruses, including influenza and HIV, is that these viruses mutate very rapidly. This allows them to evade the antibodies generated by a

particular vaccine, through a process known as "viral escape."

MIT researchers have now devised a new way to computationally model viral escape, based on models that were originally developed to analyze language. The model can predict which sections of viral surface proteins are more likely to mutate in a way that enables viral escape, and it can also identify sections that are less likely to mutate, making them good targets for new vaccines.

"Viral escape is a big problem," says Bonnie Berger, the Simons Professor of Mathematics and head of the Computation and Biology group in MIT's Computer Science and Artificial Intelligence Laboratory. "Viral escape of the surface [protein](#) of influenza and the envelope surface protein of HIV are both highly responsible for the fact that we don't have a universal flu vaccine, nor do we have a vaccine for HIV, both of which cause hundreds of thousands of deaths a year."

In a study appearing today in *Science*, Berger and her colleagues identified possible targets for vaccines against influenza, HIV, and SARS-CoV-2. Since that paper was accepted for publication, the researchers have also applied their model to the new variants of SARS-CoV-2 that recently emerged in the United Kingdom and South Africa. That analysis, which has not yet been peer-reviewed, flagged viral genetic sequences that should be further investigated for their potential to escape the existing vaccines, the researchers say.

Berger and Bryan Bryson, an assistant professor of biological engineering at MIT and a member of the Ragon Institute of MGH, MIT, and Harvard, are the senior authors of the paper, and the lead author is MIT graduate student Brian Hie.

## **The language of proteins**

Different types of viruses acquire [genetic mutations](#) at different rates, and HIV and influenza are among those that mutate the fastest. For these mutations to promote viral escape, they must help the virus change the shape of its surface proteins so that antibodies can no longer bind to them. However, the protein can't change in a way that makes it nonfunctional.

The MIT team decided to model these criteria using a type of computational model known as a language model, from the field of natural language processing (NLP). These models were originally designed to analyze patterns in language, specifically, the frequency which with certain words occur together. The models can then make predictions of which words could be used to complete a sentence such as "Sally ate eggs for ..." The chosen word must be both grammatically correct and have the right meaning. In this example, an NLP model might predict "breakfast," or "lunch."

The researchers' key insight was that this kind of model could also be applied to biological information such as genetic sequences. In that case, grammar is analogous to the rules that determine whether the protein encoded by a particular sequence is functional or not, and semantic meaning is analogous to whether the protein can take on a new shape that helps it evade antibodies. Therefore, a mutation that enables viral escape must maintain the grammaticality of the sequence but change the protein's structure in a useful way.

"If a virus wants to escape the human immune system, it doesn't want to mutate itself so that it dies or can't replicate," Hie says. "It wants to preserve fitness but disguise itself enough so that it's undetectable by the human immune system."

To model this process, the researchers trained an NLP model to analyze patterns found in genetic sequences, which allows it to predict new

sequences that have new functions but still follow the biological rules of protein structure. One significant advantage of this kind of modeling is that it requires only sequence information, which is much easier to obtain than protein structures. The model can be trained on a relatively small amount of information—in this study, the researchers used 60,000 HIV sequences, 45,000 influenza sequences, and 4,000 coronavirus sequences.

"Language models are very powerful because they can learn this complex distributional structure and gain some insight into function just from sequence variation," Hie says. "We have this big corpus of viral sequence data for each amino acid position, and the model learns these properties of amino acid co-occurrence and co-variation across the training data."

### **Blocking escape**

Once the model was trained, the researchers used it to predict sequences of the coronavirus spike protein, HIV envelope protein, and influenza hemagglutinin (HA) protein that would be more or less likely to generate escape mutations.

For influenza, the model revealed that the sequences least likely to mutate and produce viral escape were in the stalk of the HA protein. This is consistent with recent studies showing that antibodies that target the HA stalk (which most people infected with the flu or vaccinated against it do not develop) can offer near-universal protection against any flu strain.

The model's analysis of coronaviruses suggested that a part of the spike protein called the S2 subunit is least likely to generate escape mutations. The question still remains as to how rapidly the SARS-CoV-2 virus mutates, so it is unknown how long the vaccines now being deployed to

combat the COVID-19 pandemic will remain effective. Initial evidence suggests that the virus does not mutate as rapidly as influenza or HIV. However, the researchers recently identified new mutations that have appeared in Singapore, South Africa, and Malaysia, that they believe should be investigated for potential viral escape (these new data are not yet peer-reviewed).

In their studies of HIV, the researchers found that the V1-V2 hypervariable region of the protein has many possible escape mutations, which is consistent with previous findings, and they also found sequences that would have a lower probability of escape.

The researchers are now working with others to use their [model](#) to identify possible targets for cancer vaccines that stimulate the body's own immune system to destroy tumors. They say it could also be used to design small-molecule drugs that might be less likely to provoke resistance, for diseases such as tuberculosis.

"There are so many opportunities, and the beautiful thing is all we need is sequence data, which is easy to produce," Bryson says.

**More information:** B. Hie et al., "Learning the language of viral evolution and escape," *Science* (2020). [science.sciencemag.org/cgi/doi ... 1126/science.abd7331](https://science.sciencemag.org/cgi/doi/10.1126/science.abd7331)

Y.-A. Kim et al., "The language of a virus," *Science* (2020). [science.sciencemag.org/cgi/doi ... 1126/science.abf6894](https://science.sciencemag.org/cgi/doi/10.1126/science.abf6894)

Provided by Massachusetts Institute of Technology

Citation: Model analyzes how viruses escape the immune system (2021, January 14) retrieved 25

April 2024 from <https://phys.org/news/2021-01-viruses-immune.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.