Analyzing The Potential of Alphafold in Drug Discovery

2022-09-17
Study finds computer models that predict molecular interactions need improvement before they can help identify drug mechanisms of action.

Over the past few decades, very few new antibiotics have been developed, largely because current methods for screening potential drugs are prohibitively expensive and time-consuming. One promising new strategy is to use computational models, which offer a potentially faster and cheaper way to identify new drugs.

A new study from MIT reveals the potential and limitations of one such computational approach. Using protein structures generated by an artificial intelligence program called AlphaFold, the researchers explored whether existing models could accurately predict the interactions between bacterial proteins and antibacterial compounds. If so, then researchers could begin to use this type of modeling to do large-scale screens for new compounds that target previously untargeted proteins. This would enable the development of antibiotics with unprecedented mechanisms of action, a task essential to addressing the antibiotic resistance crisis.

However, the researchers, led by James Collins, the Termeer Professor of Medical Engineering and Science in MIT's Institute for Medical Engineering and Science (IMES) and Department of Biological Engineering, found that these existing models did not perform well for this purpose. In fact, their predictions performed little better than chance.

"Breakthroughs such as AlphaFold are expanding the possibilities for in silico drug discovery efforts, but these developments need to be coupled with additional advances in other aspects of modeling that are part of drug discovery efforts," Collins says. "Our study speaks to both the current abilities and the current limitations of computational platforms for drug discovery."

In their new study, the researchers were able to improve the performance of these types of models, known as molecular docking simulations, by applying machine-learning techniques to refine the results. However, more improvement will be necessary to fully take advantage of the protein structures provided by AlphaFold, the researchers say.

Collins is the senior author of the study, which appears today in the journal [Molecular Systems Biology](). MIT postdocs Felix Wong and Aarti Krishnan are the lead authors of the paper.

Molecular interactions

The new study is part of an effort recently launched by Collins' lab called the Antibiotics-AI Project, which has the goal of using artificial intelligence to discover and design new antibiotics.

AlphaFold, an AI software developed by DeepMind and Google, has accurately predicted protein structures from their amino acid sequences. This technology has generated excitement among researchers looking for new antibiotics, who hope that they could use the AlphaFold structures to find drugs that bind to specific bacterial proteins.

To test the feasibility of this strategy, Collins and his students decided to study the interactions of 296 essential proteins from E. coli with 218 antibacterial compounds, including antibiotics such as tetracyclines.

The researchers analyzed how these compounds interact with E. coli proteins using molecular docking simulations, which predict how strongly two molecules will bind together based on their shapes and physical properties.

This kind of simulation has been successfully used in studies that screen large numbers of compounds against a single protein target, to identify compounds that bind the best. But in this case, where the researchers were trying to screen many compounds against many

potential targets, the predictions turned out to be much less accurate.

By comparing the predictions produced by the model with actual interactions for 12 essential proteins, obtained from lab experiments, the researchers found that the model had false positive rates similar to true positive rates. That suggests that the model was unable to consistently identify true interactions between existing drugs and their targets.

Using a measurement often used to evaluate computational models, known as auROC, the researchers also found poor performance. "Utilizing these standard molecular docking simulations, we obtained an auROC value of roughly 0.5, which basically says you're doing no better than if you were randomly guessing," Collins says.

The researchers found similar results when they used this modeling approach with protein structures that have been experimentally determined, instead of the structures predicted by AlphaFold.

"AlphaFold appears to do roughly as well as experimentally determined structures, but we need to do a better job with molecular docking models if we're going to utilize AlphaFold effectively and extensively in drug discovery," Collins says.

Better predictions

One possible reason for the model's poor performance is that the protein structures fed into the model are static, while in biological systems, proteins are flexible and often shift their configurations.

To try to improve the success rate of their modeling approach, the researchers ran the predictions through four additional machine-learning models. These models are trained on data that describe how proteins and other molecules interact with each other, allowing them to incorporate more information into the predictions.

"The machine-learning models learn not just the shapes, but also chemical and physical properties of the known interactions, and then use that information to reassess the docking predictions," Wong says. "We found that if you were to filter the interactions using those additional models, you can get a higher ratio of true positives to false positives."

However, additional improvement is still needed before this type of modeling could be used to successfully identify new drugs, the researchers say. One way to do this would be to train the models on more data, including the biophysical and biochemical properties of proteins and their different conformations, and how those features influence their binding with potential drug compounds.

This study both lets us understand just how far we are from realizing full machine-learning-based paradigms for drug development, and provides fantastic experimental and computational benchmarks to stimulate and direct and guide progress towards this future vision," says Roy Kishony, a professor of biology and computer science at Technion (the Israel Institute of Technology), who was not involved in the study.

With further advances, scientists may be able to harness the power of AI-generated protein structures to discover not only new antibiotics but also drugs to treat a variety of diseases, including cancer, Collins says. "We're optimistic that with improvements to the modeling approaches and expansion of computing power, these techniques will become increasingly important in drug discovery," he says. "However, we have a long way to go to achieve the full potential of in silico drug discovery."

Read the [original article](#) on Massachusetts Institute of Technology (MIT).