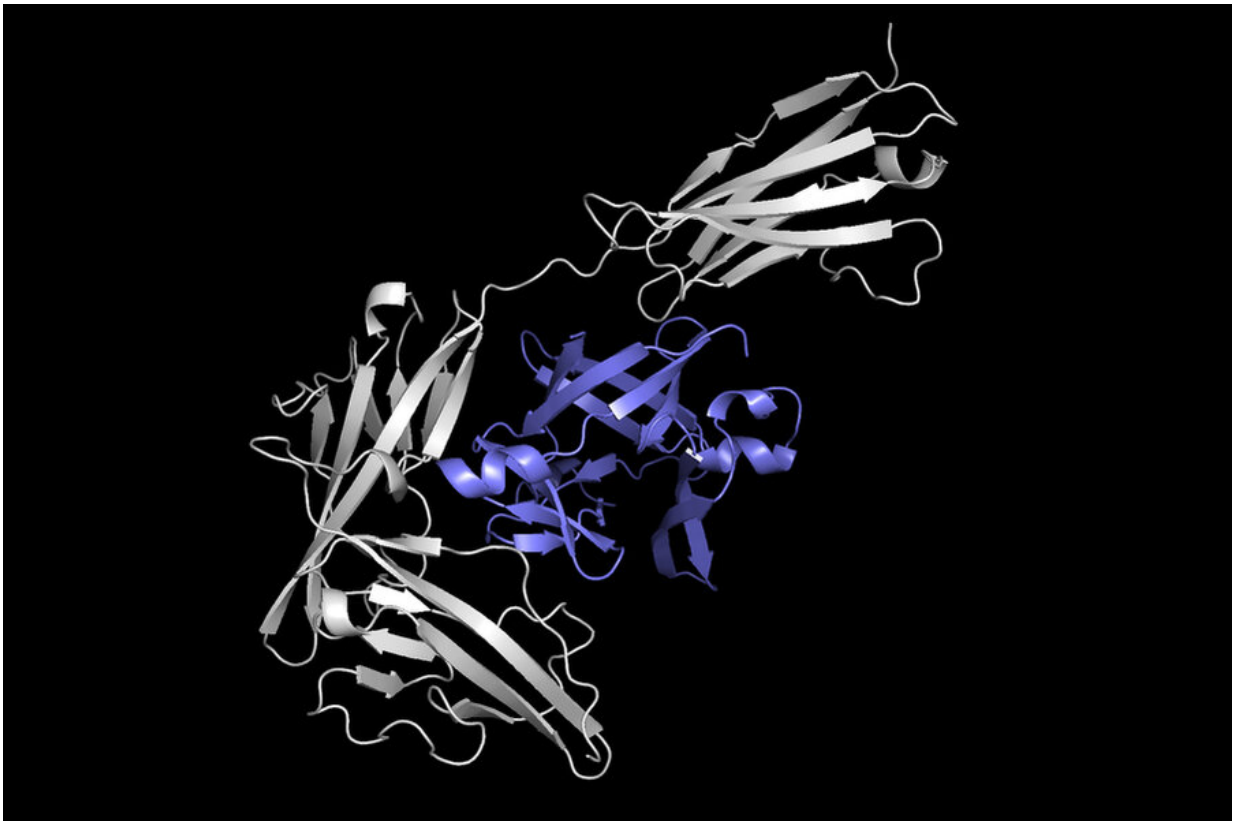


# Artificial intelligence system rapidly predicts how two proteins will attach

February 1 2022, by Adam Zewe

---



One protein (in gray) docking with another protein (in purple) to form a protein complex. Equidock, the machine learning system the researchers developed, can directly predict a protein complex like this in a matter of seconds. Credit: Massachusetts Institute of Technology

Antibodies, small proteins produced by the immune system, can attach

to specific parts of a virus to neutralize it. As scientists continue to battle SARS-CoV-2, the virus that causes COVID-19, one possible weapon is a synthetic antibody that binds with the virus' spike proteins to prevent the virus from entering a human cell.

To develop a successful synthetic antibody, researchers must understand exactly how that attachment will happen. Proteins, with lumpy 3D structures containing many folds, can stick together in millions of combinations, so finding the right [protein](#) complex among almost countless candidates is extremely time-consuming.

To streamline the process, MIT researchers created a [machine-learning model](#) that can directly predict the complex that will form when two proteins bind together. Their technique is between 80 and 500 times faster than state-of-the-art software methods, and often predicts protein structures that are closer to actual structures that have been observed experimentally.

This technique could help scientists better understand some biological processes that involve protein interactions, like DNA replication and repair; it could also speed up the process of developing new medicines.

"Deep learning is very good at capturing interactions between different proteins that are otherwise difficult for chemists or biologists to write experimentally. Some of these interactions are very complicated, and people haven't found good ways to express them. This [deep-learning](#) model can learn these types of interactions from data," says Octavian-Eugen Ganea, a postdoc in the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) and co-lead author of the paper.

Ganea's co-lead author is Xinyuan Huang, a graduate student at ETH Zurich. MIT co-authors include Regina Barzilay, the School of Engineering Distinguished Professor for AI and Health in CSAIL, and

Tommi Jaakkola, the Thomas Siebel Professor of Electrical Engineering in CSAIL and a member of the Institute for Data, Systems, and Society. The research will be presented at the International Conference on Learning Representations.

## **Protein attachment**

The model the researchers developed, called Equidock, focuses on rigid body docking—which occurs when two proteins attach by rotating or translating in 3D space, but their shapes don't squeeze or bend.

The model takes the 3D structures of two proteins and converts those structures into 3D graphs that can be processed by the neural network. Proteins are formed from chains of amino acids, and each of those amino acids is represented by a node in the graph.

The researchers incorporated geometric knowledge into the model, so it understands how objects can change if they are rotated or translated in 3D space. The model also has mathematical knowledge built in that ensures the proteins always attach in the same way, no matter where they exist in 3D space. This is how proteins dock in the human body.

Using this information, the machine-learning system identifies atoms of the two proteins that are most likely to interact and form chemical reactions, known as binding-pocket points. Then it uses these points to place the two proteins together into a complex.

"If we can understand from the proteins which individual parts are likely to be these binding pocket points, then that will capture all the information we need to place the two proteins together. Assuming we can find these two sets of points, then we can just find out how to rotate and translate the proteins so one set matches the other set," Ganea explains.

One of the biggest challenges of building this model was overcoming the lack of training data. Because so little experimental 3D data for proteins exist, it was especially important to incorporate geometric knowledge into Equidock, Ganea says. Without those geometric constraints, the model might pick up false correlations in the dataset.

## **Seconds vs. hours**

Once the model was trained, the researchers compared it to four software methods. Equidock is able to predict the final protein complex after only one to five seconds. All the baselines took much longer, from between 10 minutes to an hour or more.

In quality measures, which calculate how closely the predicted protein complex matches the actual protein complex, Equidock was often comparable with the baselines, but it sometimes underperformed them.

"We are still lagging behind one of the baselines. Our method can still be improved, and it can still be useful. It could be used in a very large virtual screening where we want to understand how thousands of proteins can interact and form complexes. Our method could be used to generate an initial set of candidates very fast, and then these could be fine-tuned with some of the more accurate, but slower, traditional methods," he says.

In addition to using this method with traditional models, the team wants to incorporate specific atomic interactions into Equidock so it can make more accurate predictions. For instance, sometimes atoms in proteins will attach through hydrophobic interactions, which involve water molecules.

Their technique could also be applied to the development of small, drug-like molecules, Ganea says. These molecules bind with protein surfaces

in specific ways, so rapidly determining how that attachment occurs could shorten the drug development timeline.

In the future, they plan to enhance Equidock so it can make predictions for flexible protein docking. The biggest hurdle there is a lack of data for training, so Ganea and his colleagues are working to generate synthetic data they could use to improve the model.

**More information:** Independent SE(3)-Equivariant Models for End-to-End Rigid Protein Docking. [openreview.net/forum?id=GQjaI9mLet](https://openreview.net/forum?id=GQjaI9mLet)

*This story is republished courtesy of MIT News ([web.mit.edu/newsoffice/](https://web.mit.edu/newsoffice/)), a popular site that covers news about MIT research, innovation and teaching.*

Provided by Massachusetts Institute of Technology

Citation: Artificial intelligence system rapidly predicts how two proteins will attach (2022, February 1) retrieved 26 April 2024 from <https://phys.org/news/2022-02-artificial-intelligence-rapidly-proteins.html>

<p>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.</p>
--