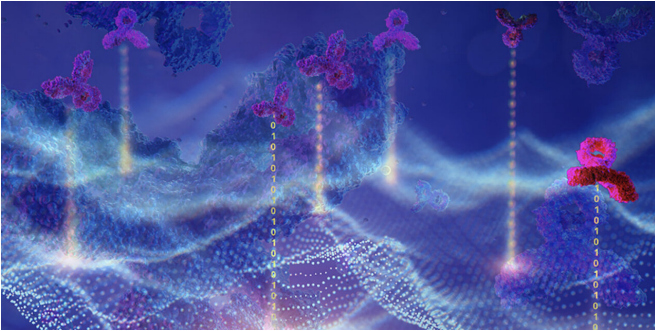# Designing better antibody drugs with artificial intelligence

15 April 2021

Machine learning helps develop optimal antibody drugs. Credit: ETH Zurich

Antibodies are not only produced by our immune cells to fight viruses and other pathogens in the body. For a few decades now, medicine has also been using antibodies produced by biotechnology as drugs. This is because antibodies are extremely good at binding specifically to molecular structures according to the lock-and-key principle. Their use ranges from oncology to the treatment of autoimmune diseases and neurodegenerative conditions.

However, developing such antibody drugs is anything but simple. The basic requirement is for an antibody to bind to its target molecule in an optimal way. At the same time, an antibody drug must fulfill a host of additional criteria. For example, it should not trigger an immune response in the body, it should be efficient to produce using biotechnology, and it should remain stable over a long period of time.

Once scientists have found an antibody that binds to the desired molecular target structure, the development process is far from over. Rather, this marks the start of a phase in which researchers use bioengineering to try to improve the antibody's properties. Scientists led by Sai Reddy, a professor

at the Department of Biosystems Science and Engineering at ETH Zurich in Basel, have now developed a machine learning method that supports this optimisation phase, helping to develop more effective antibody drugs.

Robots can't manage more than a few thousand

When researchers optimize an entire antibody molecule in its therapeutic form (i.e. not just a fragment of an antibody), it used to start with an antibody lead candidate that binds reasonably well to the desired target structure. Then researchers randomly mutate the gene that carries the blueprint for the antibody in order to produce a few thousand related antibody candidates in the lab. The next step is to search among them to find the ones that bind best to the target structure. "With automated processes, you can test a few thousand therapeutic candidates in a lab. But it is not really feasible to screen any more than that," Reddy says. Typically, the best dozen antibodies from this screening move on to the next step and are tested for how well they meet additional criteria. "Ultimately, this approach lets you identify the best antibody from a group of a few thousand," he says.

Candidate pool massively increased by machine learning

Reddy and his colleagues are now using machine learning to increase the initial set of antibodies to be tested to several million. "The more candidates there are to choose from, the greater the chance of finding one that really meets all the criteria needed for drug development," Reddy says.

The ETH researchers provided the proof of concept for their new method using Roche's antibody cancer drug Herceptin, which has been on the market for 20 years. "But we weren't looking to make suggestions for how to improve it—you can't just retroactively change an approved drug," Reddy explains. "Our reason for choosing this antibody is

because it is well known in the scientific community and because its structure is published in open-access databases."

Computer predictions

Starting out from the DNA sequence of the Herceptin antibody, the ETH researchers created about 40,000 related antibodies using a CRISPR mutation method they developed a few years ago. Experiments showed that 10,000 of them bound well to the target protein in question, a specific cell surface protein. The scientists used the DNA sequences of these 40,000 antibodies to train a machine learning algorithm.

They then applied the trained algorithm to search a database of 70 million potential antibody DNA sequences. For these 70 million candidates, the algorithm predicted how well the corresponding antibodies would bind to the target protein, resulting in a list of millions of sequences expected to bind.

Using further computer models, the scientists predicted how well these millions of sequences would meet the additional criteria for drug development (tolerance, production, physical properties). This reduced the number of candidate sequences to 8,000.

Improved antibodies found

From the list of optimized candidate sequences on their computer, the scientists selected 55 sequences from which to produce antibodies in the lab and characterize their properties. Subsequent experiments showed that several of them bound even better to the target protein than Herceptin itself, as well as being easier to produce and more stable than Herceptin. "One new variant may even be better tolerated in the body than Herceptin," says Reddy. "It is known that Herceptin triggers a weak immune response, but this is typically not a problem in this case." However, it is a problem for many other antibodies and is necessary to prevent for drug development.

The ETH scientists are now applying their artificial intelligence method to optimize antibody drugs that are in clinical development. To this end, they recently founded the ETH spin-off deepCDR Biologics, which partners with both early stage and established biotech and pharmaceutical companies for antibody drug development.

The study is published in *Nature Biomedical Engineering*.

Provided by ETH Zurich