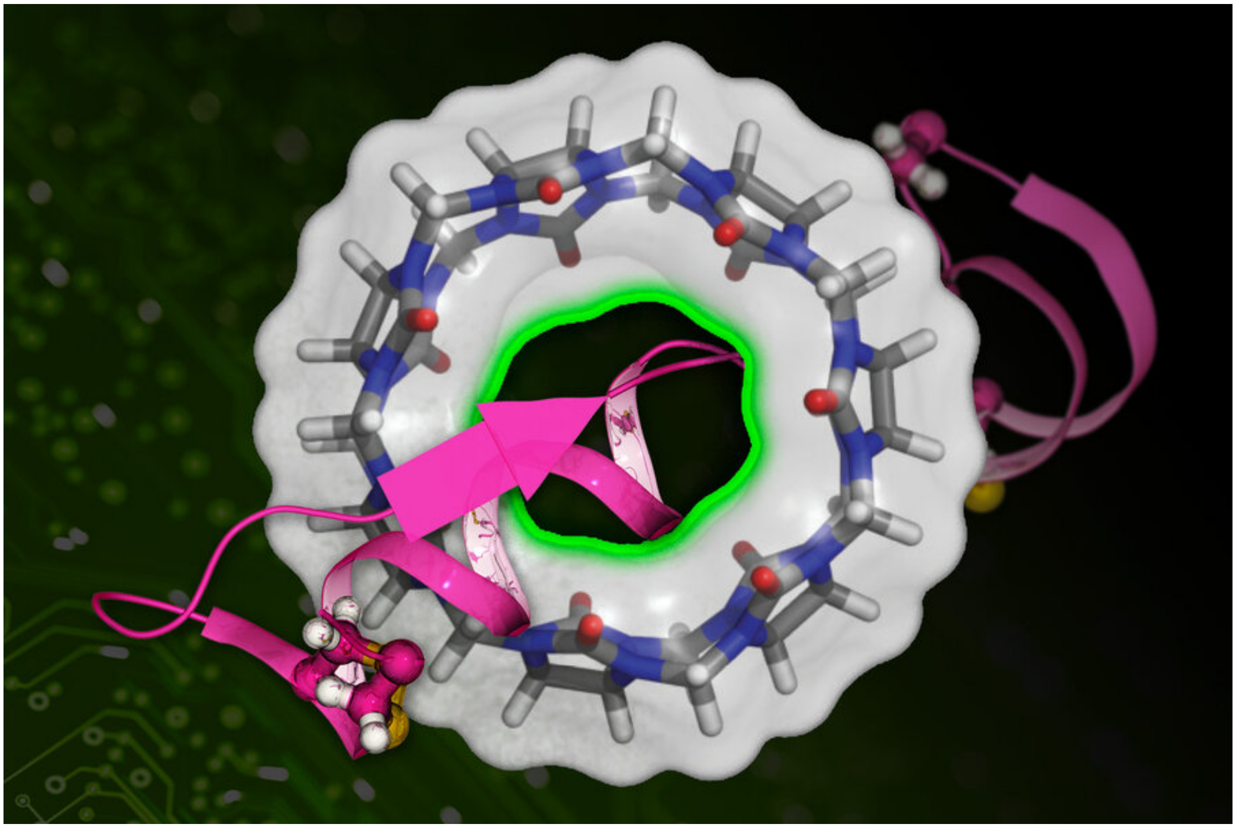
Faster drug discovery through machine learning

March 15 2021, by Daniel Ackerman



MIT researchers have developed a machine learning-based technique to more quickly calculate the binding affinity of a drug molecule (represented in pink) with a target protein (the circular structure). Credit: MIT News, and circular structure courtesy of the researchers

Drugs can only work if they stick to their target proteins in the body.

Assessing that stickiness is a key hurdle in the drug discovery and screening process. New research combining chemistry and machine learning could lower that hurdle.

The new technique, dubbed DeepBAR, quickly calculates the binding affinities between drug candidates and their targets. The approach yields precise calculations in a fraction of the time compared to previous state-of-the-art methods. The researchers say DeepBAR could one day quicken the pace of drug discovery and protein engineering.

"Our method is orders of magnitude faster than before, meaning we can have drug discovery that is both efficient and reliable," says Bin Zhang, the Pfizer-Laubach Career Development Professor in Chemistry at MIT, an associate member of the Broad Institute of MIT and Harvard, and a co-author of a new paper describing the technique.

The research appears today in the *Journal of Physical Chemistry Letters*. The study's lead author is Xinqiang Ding, a postdoc in MIT's Department of Chemistry.

The affinity between a drug molecule and a target protein is measured by a quantity called the binding free energy—the smaller the number, the stickier the bind. "A lower binding free energy means the drug can better compete against other molecules," says Zhang, "meaning it can more effectively disrupt the protein's normal function." Calculating the binding free energy of a drug candidate provides an indicator of a drug's potential effectiveness. But it's a difficult quantity to nail down.

Methods for computing binding free energy fall into two broad categories, each with its own drawbacks. One category calculates the quantity exactly, eating up significant time and computer resources. The second category is less computationally expensive, but it yields only an approximation of the binding free energy. Zhang and Ding devised an approach to get the best of both worlds.

Exact and efficient

DeepBAR computes binding free energy exactly, but it requires just a fraction of the calculations demanded by previous methods. The new technique combines traditional chemistry calculations with recent advances in machine learning.

The "BAR" in DeepBAR stands for "Bennett acceptance ratio," a decades-old algorithm used in exact calculations of binding free energy. Using the Bennet acceptance ratio typically requires a knowledge of two "endpoint" states (e.g., a drug molecule bound to a protein and a drug molecule completely dissociated from a protein), plus knowledge of many intermediate states (e.g., varying levels of partial binding), all of which bog down calculation speed.

DeepBAR slashes those in-between states by deploying the Bennett acceptance ratio in machine-learning frameworks called deep generative models. "These models create a reference state for each endpoint, the bound state and the unbound state," says Zhang. These two reference states are similar enough that the Bennett acceptance ratio can be used directly, without all the costly intermediate steps.

In using deep generative models, the researchers were borrowing from the field of computer vision. "It's basically the same model that people use to do computer image synthesis," says Zhang. "We're sort of treating each molecular structure as an image, which the model can learn. So, this project is building on the effort of the machine learning community."

While adapting a computer vision approach to chemistry was DeepBAR's key innovation, the crossover also raised some challenges.

"These models were originally developed for 2D images," says Ding. "But here we have proteins and molecules—it's really a 3D structure. So, adapting those methods in our case was the biggest technical challenge we had to overcome."

A faster future for drug screening

In tests using small protein-like molecules, DeepBAR calculated binding free energy nearly 50 times faster than previous methods. Zhang says that efficiency means "we can really start to think about using this to do drug screening, in particular in the context of COVID. DeepBAR has the exact same accuracy as the gold standard, but it's much faster." The researchers add that, in addition to drug screening, DeepBAR could aid protein design and engineering, since the method could be used to model interactions between multiple proteins.

DeepBAR is "a really nice computational work" with a few hurdles to clear before it can be used in real-world drug discovery, says Michael Gilson, a professor of pharmaceutical sciences at the University of California at San Diego, who was not involved in the research. He says DeepBAR would need to be validated against complex experimental data. "That will certainly pose added challenges, and it may require adding in further approximations."

In the future, the researchers plan to improve DeepBAR's ability to run calculations for large proteins, a task made feasible by recent advances in computer science. "This research is an example of combining traditional computational chemistry methods, developed over decades, with the latest developments in machine learning," says Ding. "So, we achieved something that would have been impossible before now."

More information: Xinqiang Ding et al. DeepBAR: A Fast and Exact Method for Binding Free Energy Computation, *The Journal of Physical*

Chemistry Letters (2021). DOI: 10.1021/acs.jpclett.1c00189

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

Provided by Massachusetts Institute of Technology

Citation: Faster drug discovery through machine learning (2021, March 15) retrieved 29 April 2024 from https://phys.org/news/2021-03-faster-drug-discovery-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.