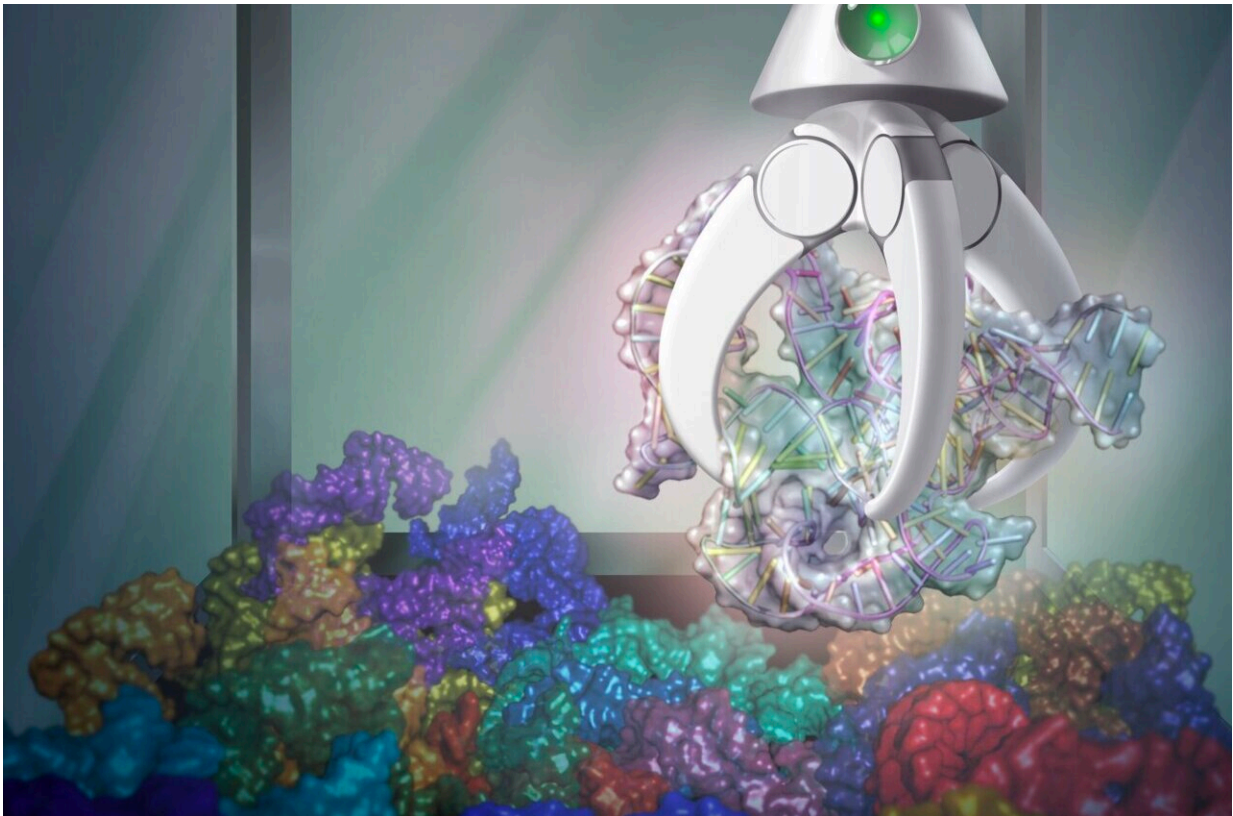# AI algorithm solves structural biology challenges

August 26 2021

A new artificial intelligence algorithm can pick out an RNA molecule's 3D shape from incorrect shapes. Computational prediction of the structures into which RNAs fold is particularly important – and particularly difficult – because so few structures are known. Credit: Camille L.L. Townshend

Determining the 3-D shapes of biological molecules is one of the hardest

problems in modern biology and medical discovery. Companies and research institutions often spend millions of dollars to determine a molecular structure—and even such massive efforts are frequently unsuccessful.

Using clever, new machine learning techniques, Stanford University Ph.D. students Stephan Eismann and Raphael Townshend, under the guidance of Ron Dror, associate professor of computer science, have developed an approach that overcomes this problem by predicting accurate structures computationally.

Most notably, their approach succeeds even when learning from only a few known structures, making it applicable to the types of molecules whose structures are most difficult to determine experimentally.

Their work is demonstrated in two papers detailing applications for RNA molecules and multi-protein complexes, published in *Science* on Aug. 27, 2021, and in *Proteins* in December 2020, respectively. The paper in *Science* is a collaboration with the Stanford laboratory of Rhiju Das, associate professor of biochemistry.

"Structural biology, which is the study of the shapes of molecules, has this mantra that structure determines function," said Townshend.

The algorithm designed by the researchers predicts accurate molecular structures and, in doing so, can allow scientists to explain how different molecules work, with applications ranging from fundamental biological research to informed drug design practices.

"Proteins are molecular machines that perform all sorts of functions. To execute their functions, proteins often bind to other proteins," said Eismann. "If you know that a pair of proteins is implicated in a disease and you know how they interact in 3-D, you can try to target this

interaction very specifically with a drug."

Eismann and Townshend are co-lead authors of the *Science* paper with Stanford postdoctoral scholar Andrew Watkins of the Das lab, and also co-lead authors of the *Proteins* paper with former Stanford Ph.D. student Nathaniel Thomas.

Designing the algorithm

Instead of specifying what makes a structural prediction more or less accurate, the researchers let the algorithm discover these molecular features for itself. They did this because they found that the conventional technique of providing such knowledge can sway an algorithm in favor of certain features, thus preventing it from finding other informative features.

"The problem with these hand-crafted features in an algorithm is that the algorithm becomes biased towards what the person who picks these features thinks is important, and you might miss some information that you would need to do better," said Eismann.

"The network learned to find fundamental concepts that are key to molecular structure formation, but without explicitly being told to," said Townshend. "The exciting aspect is that the algorithm has clearly recovered things that we knew were important, but it has also recovered characteristics that we didn't know about before."

Having shown success with proteins, the researchers next applied their algorithm to another class of important biological molecules, RNAs. They tested their algorithm in a series of "RNA Puzzles" from a long-standing competition in their field, and in every case, the tool outperformed all the other puzzle participants and did so without being designed specifically for RNA structures.

Broader applications

The researchers are excited to see where else their approach can be applied, having already had success with [protein](#) complexes and RNA molecules.

"Most of the dramatic recent advances in machine learning have required a tremendous amount of data for training. The fact that this method succeeds given very little training data suggests that related methods could address unsolved problems in many fields where data is scarce," said Dror, who is senior author of the *Proteins* paper and, with Das, co-senior author of the *Science* paper.

Specifically for [structural biology](#), the team says that they're only just scratching the surface in terms of scientific progress to be made.

"Once you have this fundamental technology, then you're increasing your level of understanding another step and can start asking the next set of questions," said Townshend. "For example, you can start designing new molecules and medicines with this kind of information, which is an area that people are very excited about."

Other co-authors of the Science paper include Stanford Ph.D. students Ramya Rangan and Maria Karelina. Other co-authors of the Proteins paper include former Stanford students Milind Jagota and Bowen Jing.

Provided by Stanford University

Citation: AI algorithm solves structural biology challenges (2021, August 26) retrieved 10 April 2024 from https://phys.org/news/2021-08-ai-algorithm-biology.html