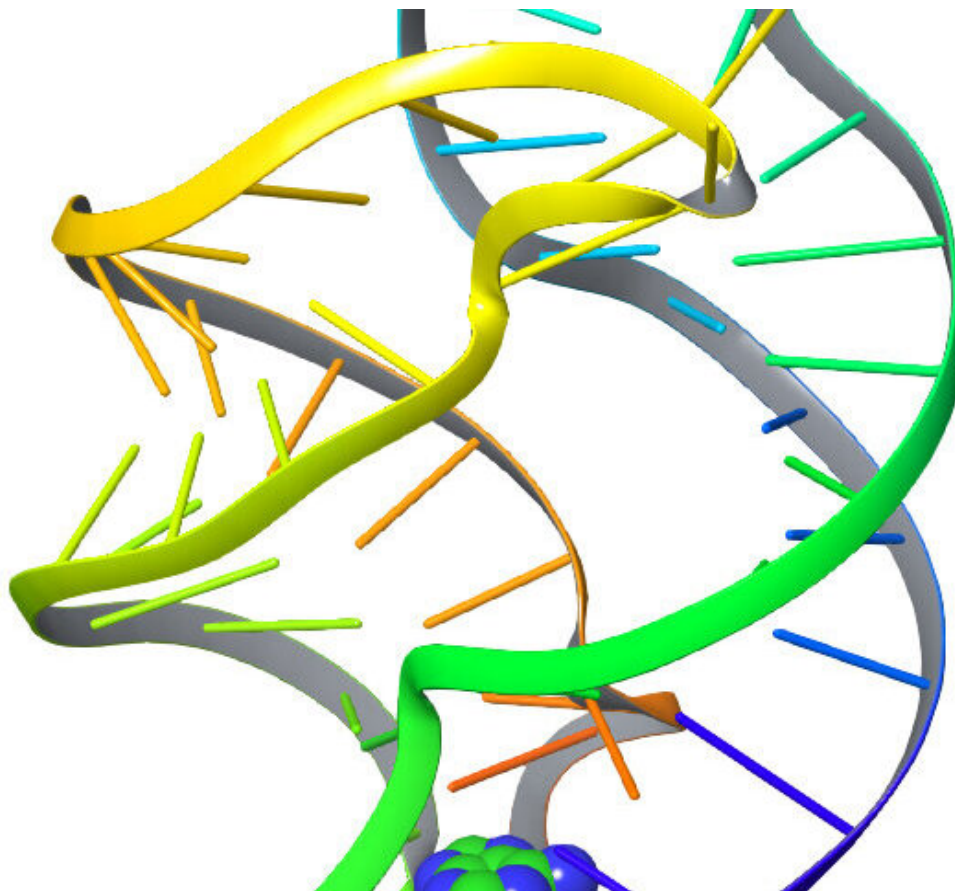


# Researchers use artificial intelligence language tools to decode molecular movements

October 9 2020

---



Scientists from the University of Maryland applied a language processing system to the movements of a riboswitch molecule pictured here, to understand how and when the molecule takes different forms. Credit: Zachary Smith/UMD

By applying natural language processing tools to the movements of protein molecules, University of Maryland scientists created an abstract language that describes the multiple shapes a protein molecule can take and how and when it transitions from one shape to another.

A [protein molecule](#)'s function is often determined by its shape and structure, so understanding the dynamics that control shape and structure can open a door to understanding everything from how a protein works to the causes of disease and the best way to design targeted drug therapies. This is the first time a [machine learning algorithm](#) has been applied to biomolecular dynamics in this way, and the method's success provides insights that can also help advance artificial intelligence (AI). A [research paper](#) on this work was published on October 9, 2020, in the journal *Nature Communications*.

"Here we show the same AI architectures used to complete sentences when writing emails can be used to uncover a [language](#) spoken by the [molecules](#) of life," said the paper's senior author, Pratyush Tiwary, an assistant professor in UMD's Department of Chemistry and Biochemistry and Institute for Physical Science and Technology. "We show that the [movement](#) of these molecules can be mapped into an [abstract language](#), and that AI techniques can be used to generate biologically truthful stories out of the resulting abstract words."

Biological molecules are constantly in motion, jiggling around in their environment. Their shape is determined by how they are folded and twisted. They may remain in a given shape for seconds or days before suddenly springing open and refolding into a different shape or structure. The transition from one shape to another occurs much like the stretching of a tangled coil that opens in stages. As different parts of the coil release and unfold, the molecule assumes different intermediary conformations.

But the transition from one form to another occurs in picoseconds (trillionths of a second) or faster, which makes it difficult for experimental methods such as high-powered microscopes and spectroscopy to capture exactly how the unfolding happens, what parameters affect the unfolding and what different shapes are possible. The answers to those questions form the biological story that Tiwary's new method can reveal.

Tiwary and his team applied Newton's laws of motion—which can predict the movement of atoms within a molecule—with powerful supercomputers, including UMD's Deepthought2, to develop statistical physics models that simulate the shape, movement and trajectory of individual molecules.

Then they fed those models into a machine learning algorithm, like the one Gmail uses to automatically complete sentences as you type. The algorithm approached the simulations as a language in which each molecular movement forms a letter that can be strung together with other movements to make words and sentences. By learning the rules of syntax and grammar that determine which shapes and movements follow one another and which don't, the algorithm predicts how the protein untangles as it changes shape and the variety of forms it takes along the way.

To demonstrate that their method works, the team applied it to a small biomolecule called riboswitch, which had been previously analyzed using spectroscopy. The results, which revealed the various forms the riboswitch could take as it was stretched, matched the results of the spectroscopy studies.

"One of the most important uses of this, I hope, is to develop drugs that are very targeted," Tiwary said. "You want to have potent drugs that bind very strongly, but only to the thing that you want them to bind to. We

can achieve that if we can understand the different forms that a given biomolecule of interest can take, because we can make drugs that bind only to one of those specific forms at the appropriate time and only for as long as we want."

An equally important part of this research is the knowledge gained about the language processing system Tiwary and his team used, which is generally called a [recurrent neural network](#), and in this specific instance a long short-term memory network. The researchers analyzed the mathematics underpinning the network as it learned the language of molecular motion. They found that the network used a kind of logic that was similar to an important concept from statistical physics called path entropy. Understanding this opens opportunities for improving recurrent neural networks in the future.

"It is natural to ask if there are governing physical principles making AI tools successful," Tiwary said. "Here we discover that, indeed, it is because the AI is learning path entropy. Now that we know this, it opens up more knobs and gears we can tune to do better AI for biology and perhaps, ambitiously, even improve AI itself. Anytime you understand a complex system such as AI, it becomes less of a black-box and gives you new tools for using it more effectively and reliably."

**More information:** "Learning molecular dynamics with simple language model built upon long short-term memory neural network," *Nature Communications* (2020). [DOI: 10.1038/s41467-020-18959-8](https://doi.org/10.1038/s41467-020-18959-8)

Provided by University of Maryland

Citation: Researchers use artificial intelligence language tools to decode molecular movements (2020, October 9) retrieved 2 May 2024 from <https://phys.org/news/2020-10-artificial->

[intelligence-language-tools-decode.html](https://www.phys.org/intelligence-language-tools-decode.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.