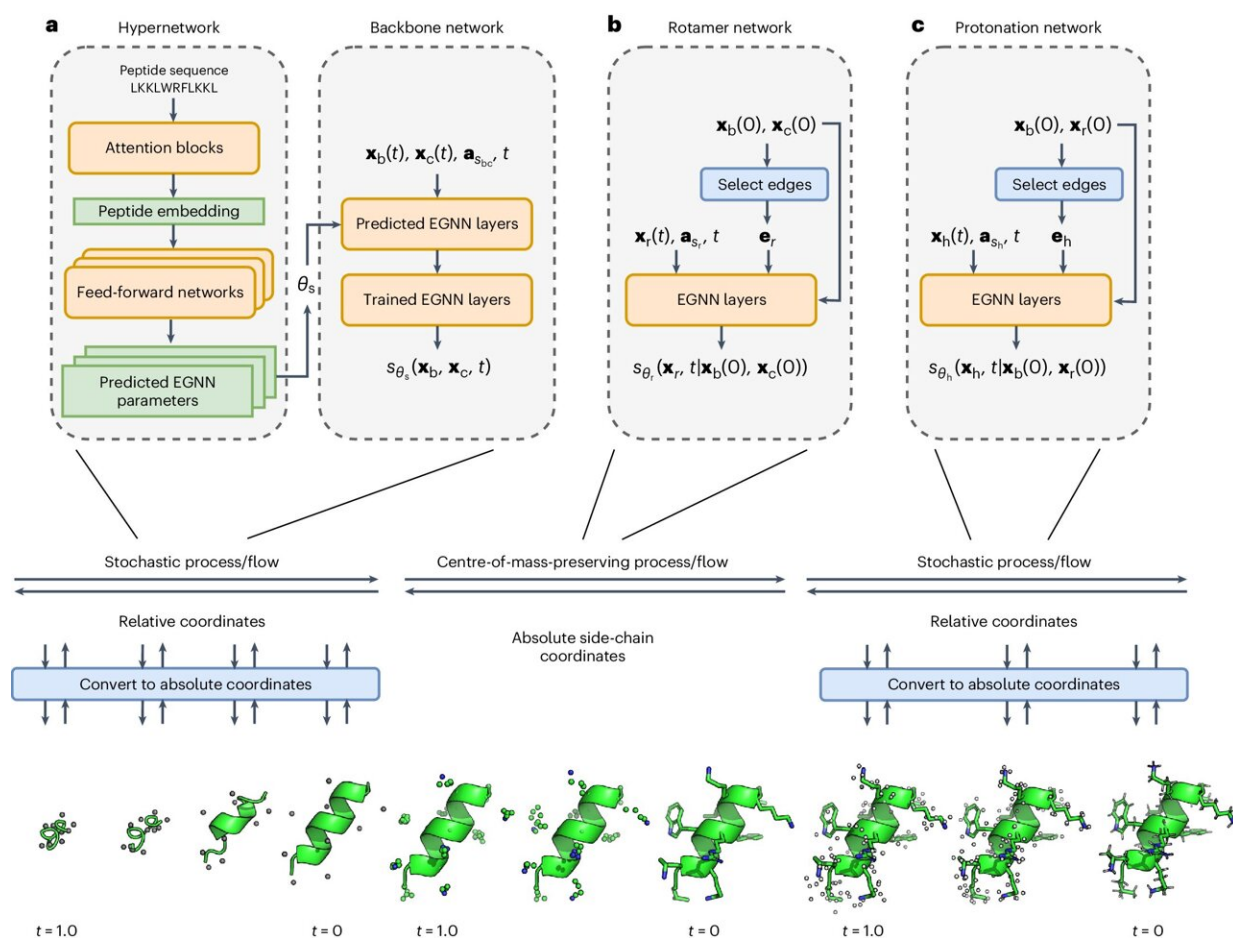# Researchers develop deep-learning model that outperforms Google AI system to predict peptide structures

June 27 2024

Schematic of the PepFlow architecture. Credit: *Nature Machine Intelligence* (2024). DOI: 10.1038/s42256-024-00860-4

Researchers at the University of Toronto have developed a deep-learning model, called PepFlow, that can predict all possible shapes of peptides—chains of amino acids that are shorter than proteins, but perform similar biological functions.

PepFlow combines machine learning and physics to model the range of folding patterns that a peptide can assume based on its energy landscape. Peptides, unlike proteins, are very dynamic molecules that can take on a range of conformations.

"We haven't been able to model the full range of conformations for peptides until now," said Osama Abdin, first author on the study and recent Ph.D. graduate of molecular genetics at U of T's Donnelly Centre for Cellular and Biomolecular Research. "PepFlow leverages deep-learning to capture the precise and accurate conformations of a peptide within minutes. There's potential with this model to inform drug development through the design of peptides that act as binders."

The study was published today in the journal *Nature Machine Intelligence*.

A peptide's role in the human body is directly linked to how it folds, as its 3D structure determines the way it binds and interacts with other molecules. Peptides are known to be highly flexible, taking on a wide range of folding patterns, and are thus involved in many biological processes of interest to researchers in the development of therapeutics.

"Peptides were the focus of the PepFlow model because they are very important biological molecules and they are naturally very dynamic, so we need to model their different conformations to understand their function," said Philip M. Kim, principal investigator on the study and a professor at the Donnelly Centre. "They're also important as therapeutics, as can be seen by the GLP1 analogs, like Ozempic, used to

treat diabetes and obesity."

Peptides are also cheaper to produce than their larger protein counterparts, said Kim, who is also a professor of computer science at U of T's Faculty of Arts & Science.

The new model expands on the capabilities of the leading Google Deepmind AI system for predicting protein structure, AlphaFold. PepFlow can outperform AlphaFold2 by generating a range of conformations for a given peptide, which AlphaFold2 was not designed to do.

What sets PepFlow apart is the technological innovations that power it. For instance, it is a generalized model that takes inspiration from Boltzmann generators, which are highly advanced physics-based machine learning models.

PepFlow can also model peptide structures that take on unusual formations, such as the ring-like structure that results from a process called macrocyclization. Peptide macrocycles are currently a highly promising venue for drug development.

While PepFlow improves upon AlphaFold2, it has limitations of its own, being the first version of a model. The study authors noted a number of ways in which PepFlow could be improved, including training the model with explicit data for solvent atoms, which would dissolve the peptides to form a solution, and for constraints on the distance between atoms in ring-like structures.

PepFlow was built to be easily expanded to account for additional considerations and new information and potential uses. Even as a first version, PepFlow is a comprehensive and efficient model with potential for furthering the development of treatments that depend on peptide

binding to activate or inhibit biological processes.

"Modeling with PepFlow offers insight into the real energy landscape of peptides," said Abdin. "It took two-and-a-half years to develop PepFlow and one month to train it, but it was worthwhile to move to the next frontier, beyond models that only predict one structure of a peptide."

More information: Osama Abdin et al, Direct conformational sampling from peptide energy landscapes through hypernetwork-conditioned diffusion, *Nature Machine Intelligence* (2024). DOI: 10.1038/s42256-024-00860-4

Provided by University of Toronto