Floppy and flexible: How Machine Learning helps to build new proteins

A team of researchers from the Heidelberg Institute for Theoretical Studies (HITS) and the Max Planck Institute for Polymer Research (MPIP) have developed a model that learns how to generate proteins whose structures are highly flexible, even with patterns that are uncommon in natural proteins. Their work, presented at the International Conference on Machine Learning (ICML), marks a step towards the goal of designing new proteins for applications in biotechnology, therapeutics and environmental research.

The natural protein universe is vast, and yet, going beyond and designing new proteins not observed in nature can yield new functions and can solve problems in medicine or materials science. The past few years have marked the golden age of de novo protein design: Machine learning methods have led to an unprecedented level of modeling accuracy. This progress enables researchers to design protein structures with specific functional properties never observed before. This is of particular interest for biotechnological applications, therapeutics development and sustainability problems, such as plastic degradation.

One of the key features of functional proteins – large biomolecules with complex structures – is their inherent structural flexibility: They wiggle, jiggle and change shape. But current designs largely lack this important feature. For a team of researchers from the Heidelberg Institute for Theoretical Studies (HITS) and the Max Planck Institute for Polymer Research (MPIP) this was the starting point to deliberate about whether one could design proteins with a custom flexibility from scratch. They presented the results of their work at the International Conference on Machine Learning (ICML) in Vancouver, Canada.

Matching the Flow: A model for de novo proteins

"We wanted to build a model that learns how to generate proteins such that their structures are flexible to a given extent at a given position", says first author Vsevolod Viliuga (MPIP). To that end, the team introduced a framework for generating flexible protein structures. This framework is based both on a neural network trained to predict flexibilities of protein backbones and a generative model for protein structure. "Natural proteins are so excellent in fulfilling their tasks because they are flexible wherever needed", says co-author Leif Seute (HITS). "We now can design novel proteins that mimic this key property." HITS group leader Jan Stühmer adds: "It is an extension of the Geometric Algebra Flow Matching model, in short: GAFL, that we developed last year." GAFL is three times faster than comparable models and not only achieves high designability, but also resembles the natural proteins better in various aspects.

In the end, the team showed that the model can generate proteins with the desired flexibility patterns, even for patterns that are uncommon in natural proteins. Frauke Gräter (MPIP), one of the team leaders, resumes: "This work is a step forward to design new proteins for applications where flexibility is required, such as enzyme catalysts."

Publication:
Vsevolod Viliuga, Leif Seute, Nicolas Wolf, Simon Wagner, Arne Elofsson, Jan Stühmer, Frauke Gräter: Flexibility-conditioned protein structure design with flow matching.
https://icml.cc/virtual/2025/poster/46289

This study received funding from the Klaus Tschira Stiftung gGmbH (HITS Lab).

Press release

23-Jul-2025
Source: HITS gGmbH Heidelberg Institute for Theoretical Studies

Further information

- Heidelberg Institute for Theoretical Studies (HITS)