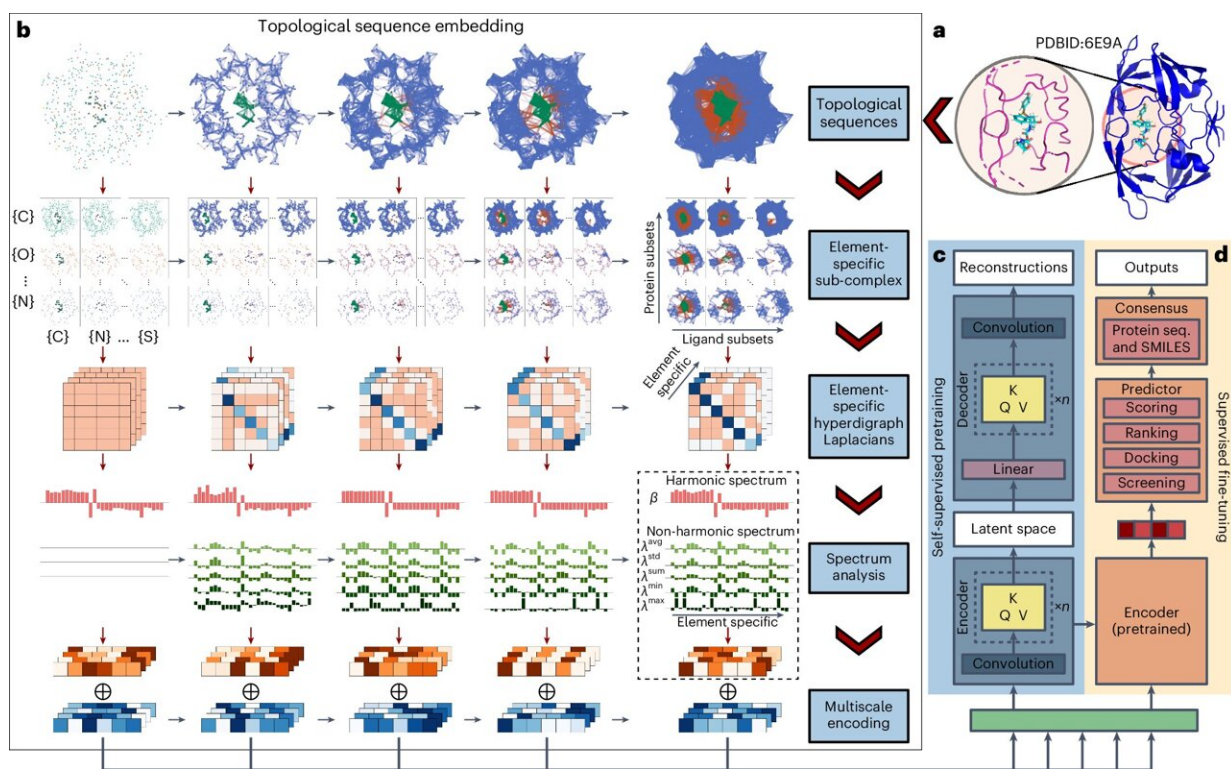


Transforming drug discovery with AI: New program transforms 3D information into data that typical models can use

June 21 2024, by Jude Coleman



Schematic illustration of the overall TopoFormer model. Credit: *Nature Machine Intelligence* (2024). DOI: 10.1038/s42256-024-00855-1

A new AI-powered program will allow researchers to level up their drug discovery efforts.

The program, called TopoFormer, was developed by an interdisciplinary team led by Guowei Wei, a Michigan State University Research Foundation Professor in the Department of Mathematics. TopoFormer translates three-dimensional information about molecules into data that typical AI-based drug-interaction models can use, expanding those models' abilities to predict how effective a drug might be.

"With AI, you could make drug discovery faster, more efficient and cheaper," said Wei, who also holds appointments in the Department of Biochemistry and Molecular Biology and the Department of Electrical and Computer Engineering.

Wei and his team published a paper about their work in the journal [Nature Machine Intelligence](#).

Instructions for structure

In the United States, developing a single drug is roughly a decade-long process that costs around \$2 billion, Wei said. Testing the drug with trials eats up roughly half of that time, he added, but the other half goes into discovering a new therapeutic candidate to test.

TopoFormer has the potential to shrink development time. In doing so, it can reduce development costs, which could lower the price of the drug for consumers downstream. That could be particularly useful for [rare diseases](#), because the limited number of patients means drug companies need to charge more to recoup costs.

Although researchers currently use computer models to aid in [drug discovery](#), there are limitations, stemming from the myriad variables of the problem.

"In our body we have over 20,000 proteins," Wei said. "When a disease

comes up, some or one of those is targeted."

The first step, then, is learning which [protein](#) or proteins a disease affects. Those proteins also become the targets for researchers, who want to find molecules that can prevent, minimize or counteract the effects of the disease.

"When I have a target, I try to find a lot of potential drugs for that particular target," Wei said.

Once scientists know which proteins to target with a drug, they can input molecular sequences from the [protein](#) and potential drugs into conventional computer models. The models predict how the drugs and target will interact, guiding decisions on which drugs to develop and test in clinical trials.

While these models can predict some interactions based on the drug and protein's chemical makeup alone, they also miss vital interactions that come from molecular shape and three-dimensional, or 3D, structure.

Ibuprofen, discovered by chemists in the 1960s, is one example of this. There are two different ibuprofen molecules that share the exact same chemical sequence but have slightly different 3D structures. Only one arrangement is shaped in a way that can bind to pain-related proteins and erase a headache.

"Current deep learning models can't account for the shape of drugs or proteins when predicting how they'll work together," Wei said.

That's where TopoFormer comes in. It's a transformer model, the same type of artificial intelligence used by Open AI's chatbot, ChatGPT (the GPT stands for "generative pre-trained transformer").

That means that TopoFormer is trained to read information in one form and turn it into another form. In this case, it takes three-dimensional information about how proteins and drugs interact based on their shapes and recreates it as one-dimensional information that [current models](#) can understand.

In fact, "Topo" stands for "topological Laplacian," which refers to mathematical tools Wei and his team invented to convert 3D structures into 1D sequences.

The new model is trained on tens of thousands of protein-drug interactions, where each interaction between two molecules is recorded as a piece of code, or a "word." The words are strung together to create a description of the drug-protein complex, creating a record of its shape.

"In such a way, you have many, many words knitted together like a sentence," Wei said.

Those sentences can then be read by other models that predict new drug interactions, and give them more context. If a new drug is a book, TopoFormer can take a rough story idea and turn it into a fully-fledged plotline, ready to be written.

More information: Dong Chen et al, Multiscale topology-enabled structure-to-sequence transformer for protein–ligand interaction predictions, *Nature Machine Intelligence* (2024). [DOI: 10.1038/s42256-024-00855-1](https://doi.org/10.1038/s42256-024-00855-1)

Provided by Michigan State University

Citation: Transforming drug discovery with AI: New program transforms 3D information into

data that typical models can use (2024, June 21) retrieved 26 June 2024 from
<https://phys.org/news/2024-06-drug-discovery-ai-3d-typical.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.