

NVIDIA, Evozyne Create Generative AI Model for Proteins

Scientists use NVIDIA BioNeMo for large language models that generate high-quality proteins that can speed drug design and help create a more sustainable environment.

Author: Rick Merritt

Using a pretrained AI model from NVIDIA, startup Evozyne created two proteins with significant potential in healthcare and clean energy.

A joint paper released today describes the process and the biological building blocks it produced. One aims to cure a congenital disease, another is designed to consume carbon dioxide to reduce global warming.

Initial results show a new way to accelerate drug discovery and more.

"It's been really encouraging that even in this first round the AI model has produced synthetic proteins as good as naturally occurring ones," said Andrew Ferguson, Evozyne's co-founder and a co-author of the paper. "That tells us it's learned nature's design rules correctly."

Evozyne used NVIDIA's implementation of ProtT5, a transformer model that's part of NVIDIA BioNeMo, a software framework and service for creating AI models for healthcare.

"BioNeMo really gave us everything we needed to support model training and then run jobs with the model very inexpensively — we could generate millions of sequences in just a few seconds," said Ferguson, a molecular engineer working at the intersection of chemistry and machine learning.

The model lies at the heart of Evozyne's process called ProT-VAE. It's a workflow that combines BioNeMo with a variational autoencoder that acts as a filter.

"Using large language models combined with variational autoencoders to design proteins was not on anybody's radar just a few years ago," he said.

Like a student reading a book, NVIDIA's transformer model reads sequences of amino acids in millions of proteins. Using the same techniques neural networks employ to understand text, it learned how nature assembles these powerful building blocks of biology.

The model then predicted how to assemble new proteins suited for functions Evozyne wants to address.

"The technology is enabling us to do things that were pipe dreams 10 years ago," he said.

Machine learning helps navigate the astronomical number of possible protein sequences, then efficiently identifies the most useful ones.

The traditional method of engineering proteins, called directed evolution, uses a slow, hit-or-miss approach. It typically only changes a few amino acids in sequence at a time.

By contrast, Evozyne's approach can alter half or more of the amino acids in a protein in a single round. That's the equivalent of making hundreds of mutations.

"We're taking huge jumps which allows us to explore proteins never seen before that have new and useful functions," he said.

Using the new process, Evozyne plans to build a range of proteins to fight diseases and climate change.

“NVIDIA’s been an incredible partner on this work,” he said.

“They scaled jobs to multiple GPUs to speed up training,” said Joshua Moller, a data scientist at Evozyne. “We were getting through entire datasets every minute.”

That reduced the time to train large AI models from months to a week. “It allowed us to train models — some with billions of trainable parameters — that just would not be possible otherwise,” Ferguson said.

The horizon for AI-accelerated protein engineering is wide.

“The field is moving incredibly quickly, and I’m really excited to see what comes next,” he said, noting the recent rise of diffusion models.

“Who knows where we will be in five years’ time.”

Sign up for early access to the NVIDIA BioNeMo to see how it can accelerate your applications.

Original URL: <https://blogs.nvidia.com/blog/2023/01/12/generative-ai-proteins-evozyne/>