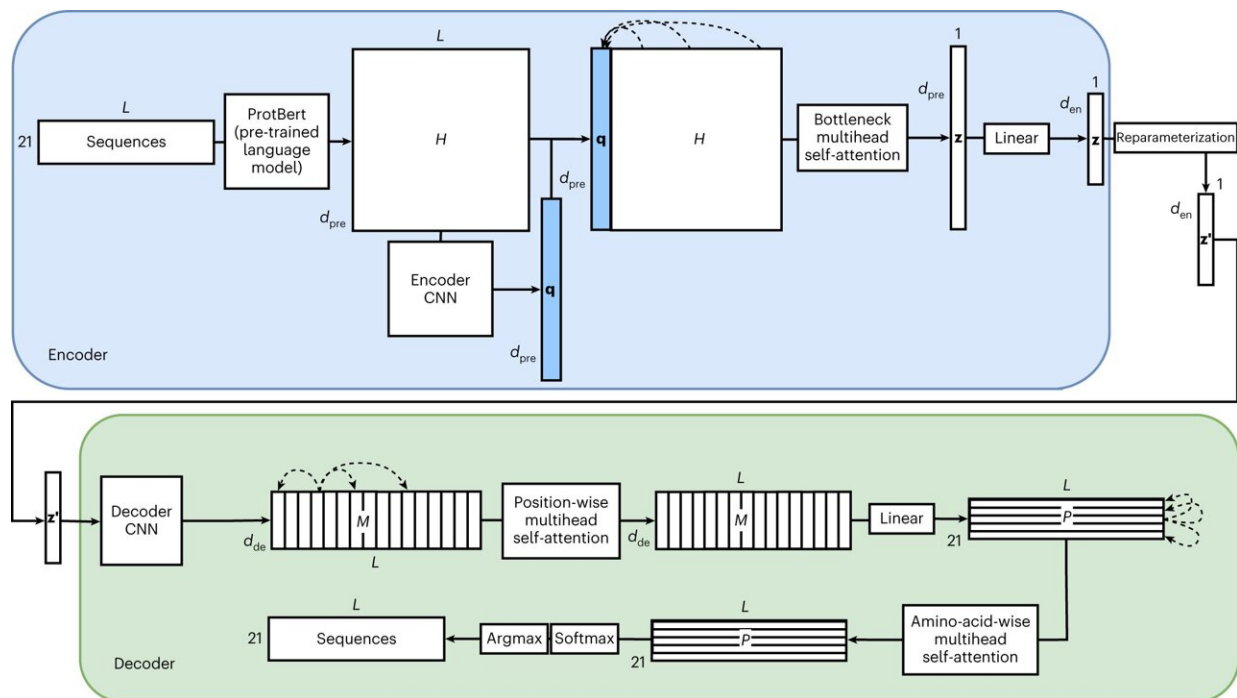


New AI model designs proteins to deliver gene therapy

January 29 2024, by Qin Dai



ProteinVAE architecture. Encoder: pre-trained language model (for example, ProtBert) converts the input one-hot-encoded sequences to amino-acid-level representation H . Encoder CNN extracts crude sequence-level representation q , which is used as the fixed query vector in the bottleneck attention mechanism to produce the refined global representation z . Dimensionality of z is adjusted with a linear layer. Reparameterization samples z' from the distribution defined by z . Decoder: deconvolutional networks are used for upsampling the latent vector z' . Self-attention along the length dimension is used to capture long-distance dependency. A linear layer converts hidden representation M to logits P , and another self-attention along the channel dimension is used to adjust for amino acid usage difference. The final logits are passed through a softmax layer to

obtain probabilities, with the highest probability amino acid being selected as the prediction at each position. d_{en} , latent dimension; d_{pre} , hidden dimension of pre-trained language model; CNN, convolutional neural network; L, sequence length. Credit: *Nature Machine Intelligence* (2024). DOI: 10.1038/s42256-023-00787-2

Researchers at the University of Toronto have used an artificial intelligence framework to redesign a crucial protein involved in the delivery of gene therapy.

The study, [published](#) in *Nature Machine Intelligence*, describes new work optimizing proteins to mitigate immune responses, thereby improving the efficacy of gene therapy and reducing side effects.

"Gene therapy holds immense promise, but the body's pre-existing [immune response](#) to viral vectors greatly hampers its success. Our research zeroes in on hexons, a fundamental protein in adenovirus vectors, which—but for the immune problem—hold huge potential for gene therapy," says Michael Garton, an assistant professor at the Institute of Biomedical Engineering in the Faculty of Applied Science & Engineering.

"Immune responses triggered by serotype-specific antibodies pose a significant obstacle in getting these vehicles to the right target; this can result in reduced efficacy and severe adverse effects."

To address the issue, Garton's lab used AI to custom-design variants of hexons that are distinct from natural sequences.

"We want to design something that is distant from all human variants and is, by extension, unrecognizable by the immune system," says Ph.D. candidate Suyue Lyu, who is lead author of the study.

Traditional methods of designing new proteins often involve extensive trial and error as well as mounting costs. By using an AI-based approach for [protein design](#), researchers can achieve a higher degree of variation, reduce costs and quickly generate simulation scenarios before homing in on a specific subset of targets for experimental testing.

While numerous protein-designing frameworks exist, it can be challenging for researchers to properly design new variants because of the lack of natural sequences available and hexons' relatively large size—consisting, on average, of 983 amino acids.

With this in mind, Lyu and Garton developed a different AI framework. Dubbed ProteinVAE, the model can be trained to learn the characteristics of a long protein using limited data. Despite its compact design, ProteinVAE exhibits a generative capability comparable to larger available models.

"Our model takes advantage of pre-trained protein language models for efficient learning on small datasets. We also incorporated many tailored engineering approaches to make the model suitable for generating long proteins," says Lyu, adding that ProteinVAE was intentionally designed to be lightweight.

Lyu adds, "Unlike other, considerably larger models that demand high computational resources to design a long protein, ProteinVAE supports fast training and inference on any standard GPUs. This feature could make the model more friendly for other academic labs. Our AI model, validated through molecular simulation, demonstrates the ability to change a significant percentage of the protein's surface, potentially evading immune responses."

The next step is experimental testing in a wet lab, Lyu explains.

Garton believes the AI-model can be utilized beyond [gene therapy](#) protein design and could likely be expanded to support protein design in other disease cases as well.

"This work indicates that we are potentially able to design new subspecies and even species of biological entities using generative AI," he says, "and these entities have therapeutic value that can be used in novel medical treatments."

More information: Suyue Lyu et al, Variational autoencoder for design of synthetic viral vector serotypes, *Nature Machine Intelligence* (2024). [DOI: 10.1038/s42256-023-00787-2](https://doi.org/10.1038/s42256-023-00787-2)

Provided by University of Toronto

Citation: New AI model designs proteins to deliver gene therapy (2024, January 29) retrieved 28 April 2024 from <https://phys.org/news/2024-01-ai-proteins-gene-therapy.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.