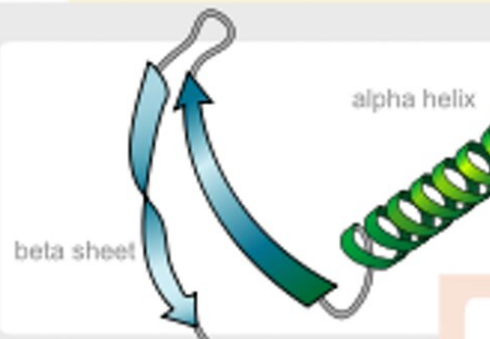
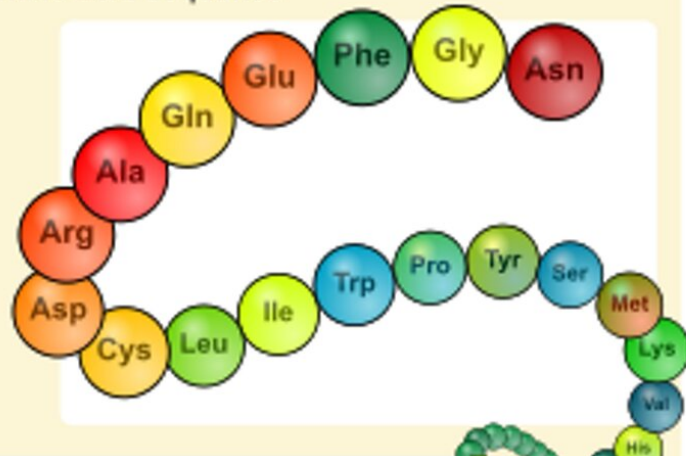


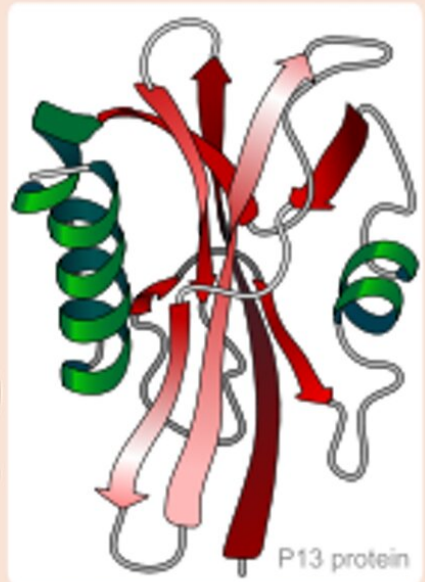
When researchers don't have the proteins they need, they can get AI to 'hallucinate' new structures

January 6 2022, by Ivan Anishchenko

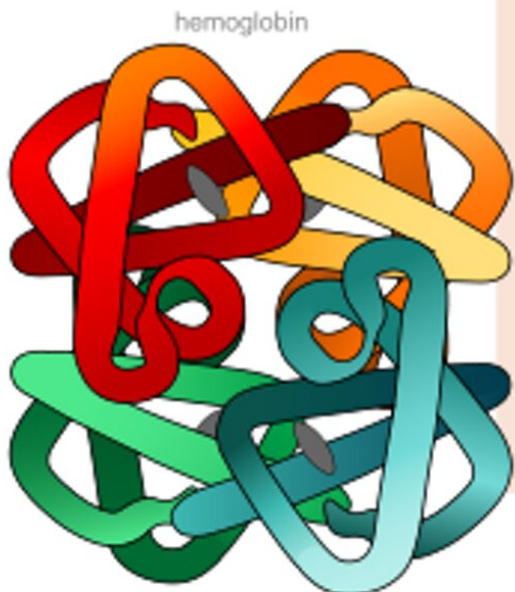
Primary structure
amino acid sequence



Secondary structure
regular sub-structures



Tertiary structure
three-dimensional structure



Quaternary structure
complex of protein molecules

Amino acid chains fold into proteins. Credit: LadyofHats/Wikimedia Commons

All living organisms use proteins, which encompass a vast number of complex molecules. They perform a wide array of functions, from allowing plants to [use solar energy for oxygen production](#) to helping your [immune system](#) fight against pathogens to letting your [muscles](#) perform physical work. [Many drugs](#) are also based on proteins.

For many areas of biomedical research and drug development, however, there are no natural proteins that can serve as suitable starting points to build new proteins. Researchers designing new drugs to [prevent COVID-19 infection](#), or developing proteins that can [turn genes on or off](#) or [turn cells into computers](#), had to create new proteins from scratch.

This process of [de novo protein design](#) can be difficult to get right. [Protein engineers like me](#) have been trying to figure out ways to more efficiently and accurately design new proteins with the properties we need.

Luckily, a form of artificial intelligence called [deep learning](#) may provide an elegant way to create proteins that did not exist previously—[hallucination](#).

Designing proteins from scratch

Proteins are made up of hundreds to thousands of smaller building blocks called [amino acids](#). These amino acids are connected to one another in long chains that fold up to form a [protein](#). The order in which these amino acids are connected to one another determines each [protein](#)

's unique structure and function.

The biggest challenge protein engineers face when designing new proteins is coming up with a [protein structure](#) that will perform a desired function. To get around this problem, researchers typically create design templates based on naturally occurring proteins with a similar function. These templates have instructions on how to create the unique folds of each particular protein. However, because a template must be created for each individual fold, this strategy is time-consuming, labor-intensive and limited by what proteins are available in nature.

Over the past few years, various research groups, [including the lab I work in](#), have developed a number of dedicated [deep neural networks](#)—computer programs that use multiple processing layers to "learn" from input data to make predictions about a desired output.

When the desired output is a new protein, millions of parameters describing different facets of a protein are put into the [network](#). What's predicted is a randomly chosen sequence of [amino acids](#) mapped onto the most probable 3D structure that sequence would take.

Network predictions for a random amino [acid](#) sequence are blurry, meaning the final structure of the protein is not very clear-cut, while both naturally occurring proteins and proteins built from scratch produce much more well-defined protein structures.

Hallucinating new proteins

These observations hint at one way that new proteins can be generated from scratch—by tweaking random inputs to the network until predictions yield a well-defined structure.

The [protein generation method my colleagues](#) and I developed is

conceptually similar to [computer vision](#) methods such as [Google's DeepDream](#), which finds and enhances patterns in images.

These methods work by taking networks trained to recognize human faces or other patterns in images, like the shape of an animal or an object, and inverting them so that they learn to recognize these patterns where they don't exist. In DeepDream, for example, the network is given arbitrary input images that are adjusted until the network can recognize a face or some other shape in the image. While the final image doesn't look much like a face to a person looking at it, it would to the neural network.

The products of this technique are often referred to as [hallucinations](#), and this is what we call our designed proteins, too.

[Our method](#) starts by passing a random amino acid sequence through a deep neural network. The resulting predictions are initially blurry, with unclear structures, as expected for random sequences. Next, we introduce a mutation that changes one amino acid in the chain into a different one and pass this new sequence through the network again. If this change gives the protein a more defined structure, then we keep the amino acid and we introduce another mutation into the sequence.

With each repetition of this process, the proteins get closer and closer to the real shape they would take if they were produced in nature. Thousands of repetitions are required to create a brand-new protein.

Using this process, we generated 2,000 new protein sequences predicted to fold into well-defined structures. Of these, we selected over 100 that were the most distinct in shape to physically recreate in the lab. Finally, we chose three of the top candidates for detailed analysis and confirmed that they were close matches to the shapes predicted by our hallucinated models.

Why hallucinate new proteins?

Our hallucination approach greatly simplifies the protein design pipeline. By eliminating the need for templates, researchers can directly focus on creating a protein based on desired functions and let the network take care of figuring out the [structure](#) for them.

Our work opens up multiple avenues for researchers to explore. Our lab is [currently investigating](#) how to best use this hallucination approach to generate even more specificity in the function of designed proteins. Our approach can also be readily extended to design new proteins using [other recently developed](#) deep neural networks.

The potential applications of de novo proteins are vast. With deep neural networks, researchers will be able to create even more proteins that can [break down plastics](#) to reduce environmental pollution, [identify and respond](#) to unhealthy cells and [improve vaccines](#) against existing and new pathogens—just to name a few.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: When researchers don't have the proteins they need, they can get AI to 'hallucinate' new structures (2022, January 6) retrieved 20 June 2024 from <https://phys.org/news/2022-01-dont-proteins-ai-hallucinate.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.