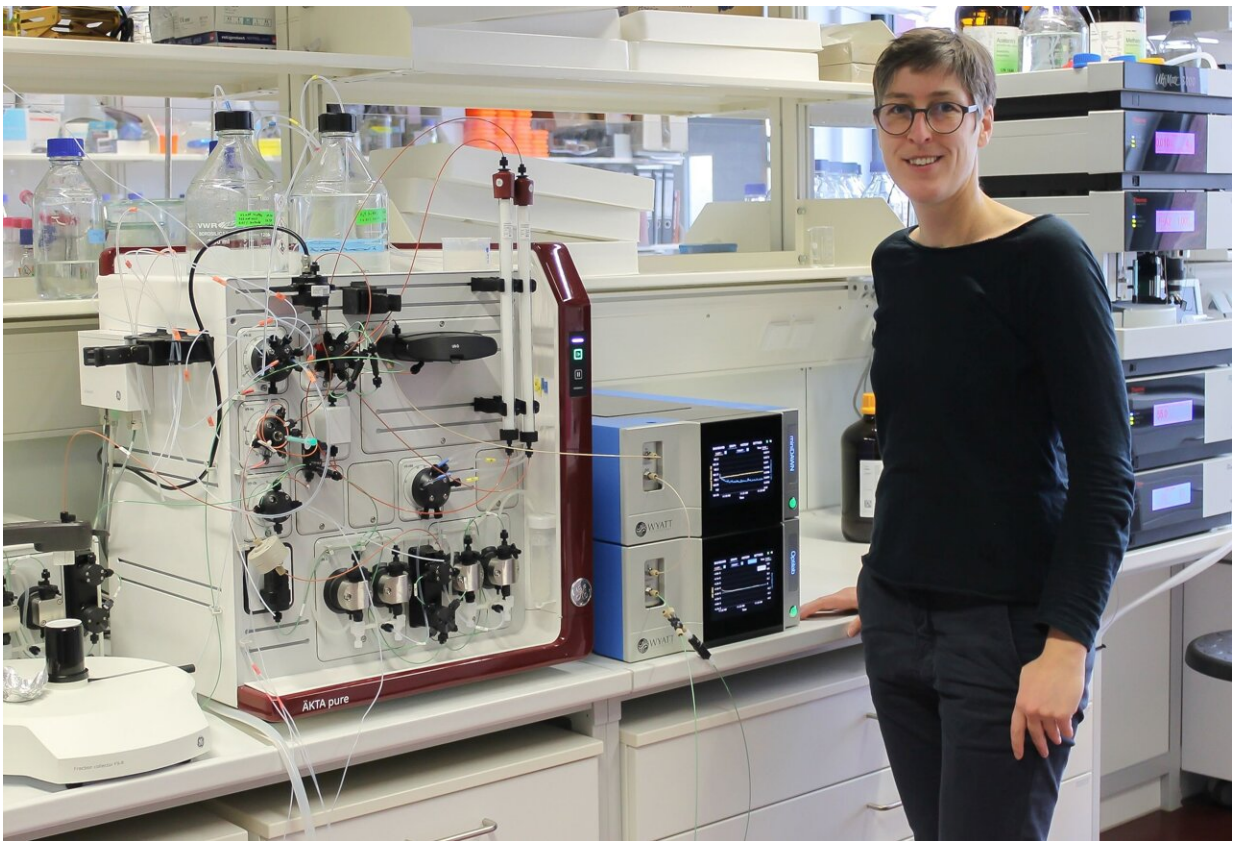


# Proteins and natural language: Artificial intelligence enables the design of novel proteins

August 4 2022, by Christian Wißler

---



Prof. Dr. Birte Höcker in a biochemistry laboratory on the Bayreuth campus.  
Credit: UBT / Christian Wißler

Artificial intelligence (AI) has created new possibilities for designing

tailor-made proteins to solve everything from medical to ecological problems. A research team at the University of Bayreuth led by Prof. Dr. Birte Höcker has now successfully applied a computer-based natural language processing model to protein research. Completely independently, the ProtGPT2 model designs new proteins that are capable of stable folding and could take over defined functions in larger molecular contexts. The model and its potential are detailed scientifically in *Nature Communications*.

Natural languages and proteins are actually similar in structure. Amino acids arrange themselves in a multitude of combinations to form structures that have specific functions in the living organism—similar to the way words form sentences in different combinations that express certain facts. In recent years, numerous approaches have therefore been developed to use principles and processes that control the computer-assisted processing of [natural language](#) in [protein research](#). "Natural language processing has made extraordinary progress thanks to new AI technologies. Today, models of language processing enable machines not only to understand meaningful sentences but also to generate them themselves. Such a [model](#) was the starting point of our research. With detailed information concerning about 50 million sequences of natural proteins, my colleague Noelia Ferruz trained the model and enabled it to generate [protein sequences](#) on its own. It now understands the language of proteins and can use it creatively. We have found that these creative designs follow the basic principles of natural proteins," says Prof. Dr. Birte Höcker, Head of the Protein Design Group at the University of Bayreuth.

The language processing model transferred to protein evolution is called ProtGPT2. It can now be used to design proteins that adopt stable structures through folding and are permanently functional in this state. In addition, the Bayreuth biochemists have found out, through complex investigations, that the model can even create proteins that do not occur

in nature and have possibly never existed in the history of evolution. These findings shed light on the immeasurable world of possible proteins and open a door to designing them in novel and unexplored ways. There is a further advantage: Most proteins that have been designed de novo so far have idealized structures. Before such structures can have a potential application, they usually must pass through an elaborate functionalization process—for example by inserting extensions and cavities—so that they can interact with their environment and take on precisely defined functions in larger system contexts. ProtGPT2, on the other hand, generates proteins that have such differentiated structures innately, and are thus already operational in their respective environments.

"Our new model is another impressive demonstration of the systemic affinity of [protein design](#) and [natural language processing](#). Artificial intelligence opens up highly interesting and promising possibilities to use methods of language processing for the production of customized proteins. At the University of Bayreuth, we hope to contribute in this way to developing innovative solutions for biomedical, pharmaceutical, and ecological problems," says Prof. Dr. Birte Höcker.

**More information:** Noelia Ferruz et al, ProtGPT2 is a deep unsupervised language model for protein design, *Nature Communications* (2022). [DOI: 10.1038/s41467-022-32007-7](https://doi.org/10.1038/s41467-022-32007-7)

Noelia Ferruz et al, Controllable protein design with language models, *Nature Machine Intelligence* (2022). [DOI: 10.1038/s42256-022-00499-z](https://doi.org/10.1038/s42256-022-00499-z)

Provided by Bayreuth University

Citation: Proteins and natural language: Artificial intelligence enables the design of novel proteins (2022, August 4) retrieved 26 April 2024 from <https://phys.org/news/2022-08-proteins->

[natural-language-artificial-intelligence.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.