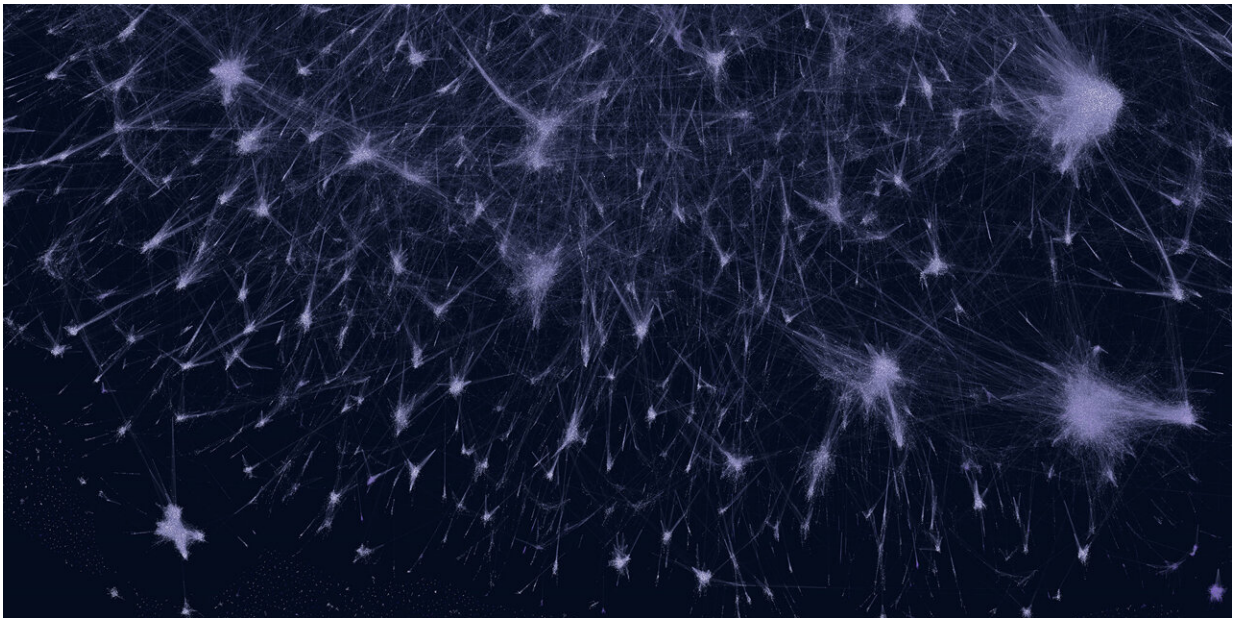# Artificial intelligence tools shed light on millions of proteins

September 20 2023

A snapshot of the interactive network "Protein Universe Atlas". Credit: University of Basel, Biozentrum

A research team at the University of Basel and the SIB Swiss Institute of Bioinformatics uncovered a treasure trove of uncharacterized proteins. Embracing the recent deep learning revolution, they discovered hundreds of new protein families and even a novel predicted protein fold. The study has now been published in *Nature*.

In the past years, AlphaFold has revolutionized [protein](#) science. This [artificial intelligence](#) (AI) tool was trained on protein data collected by [life scientists](#) for over 50 years, and is able to predict the 3D shape of proteins with high accuracy. Its success prompted the modeling of an astounding 215 million proteins last year, providing insights into the shapes of almost any protein. This is particularly interesting for proteins that have not been studied experimentally, a complex and time-consuming process.

"There are now many sources of protein information, enclosing [valuable insights](#) into how proteins evolve and work," says Joana Pereira, the leader of the study. Nevertheless, research has long been faced with a data jungle. The research team led by Professor Torsten Schwede, group leader at the Biozentrum, University of Basel, and the Swiss Institute of Bioinformatics (SIB), has now succeeded in decrypting some of the concealed information.

A bird's eye view reveals new protein families and folds

The researchers constructed an interactive network of 53 million proteins with high quality AlphaFold structures. "This network serves as a valuable source for theoretically predicting unknown protein families and their functions on a large scale," says Dr. Janani Durairaj, the first author. The team was able to identify 290 new protein families and one new protein fold that resembles the shape of a flower.

Building on the expertise of the Schwede group in developing and maintaining the leading software SWISS-MODEL, they made the network available as an interactive web resource, termed the "Protein Universe Atlas."

AI as a valuable tool in research

The team has employed deep learning-based tools for finding novelties in this network, paving the way to innovations in life sciences, from basic to applied research. "Understanding the structure and function of proteins is typically one of the first steps to develop a new drug, or modify their functions by protein engineering, for example," says Pereira.

The work was supported by a "kickstarter" grant from SIB to encourage the adoption of AI in life science resources. It underscores the transformative potential of deep learning and intelligent algorithms in research.

With the Protein Universe Atlas, scientists can now learn more about proteins relevant to their research. "We hope this resource will help not only researchers and biocurators but also students and teachers by providing a new platform for learning about protein diversity, from structure, to function, to evolution," says Janani Durairaj.

More information: Janani Durairaj et al, Uncovering new families and folds in the natural protein universe, *Nature* (2023). DOI: 10.1038/s41586-023-06622-3

Protein Universe Atlas: uniprot3d.org/atlas/AFDB90v4

Provided by University of Basel