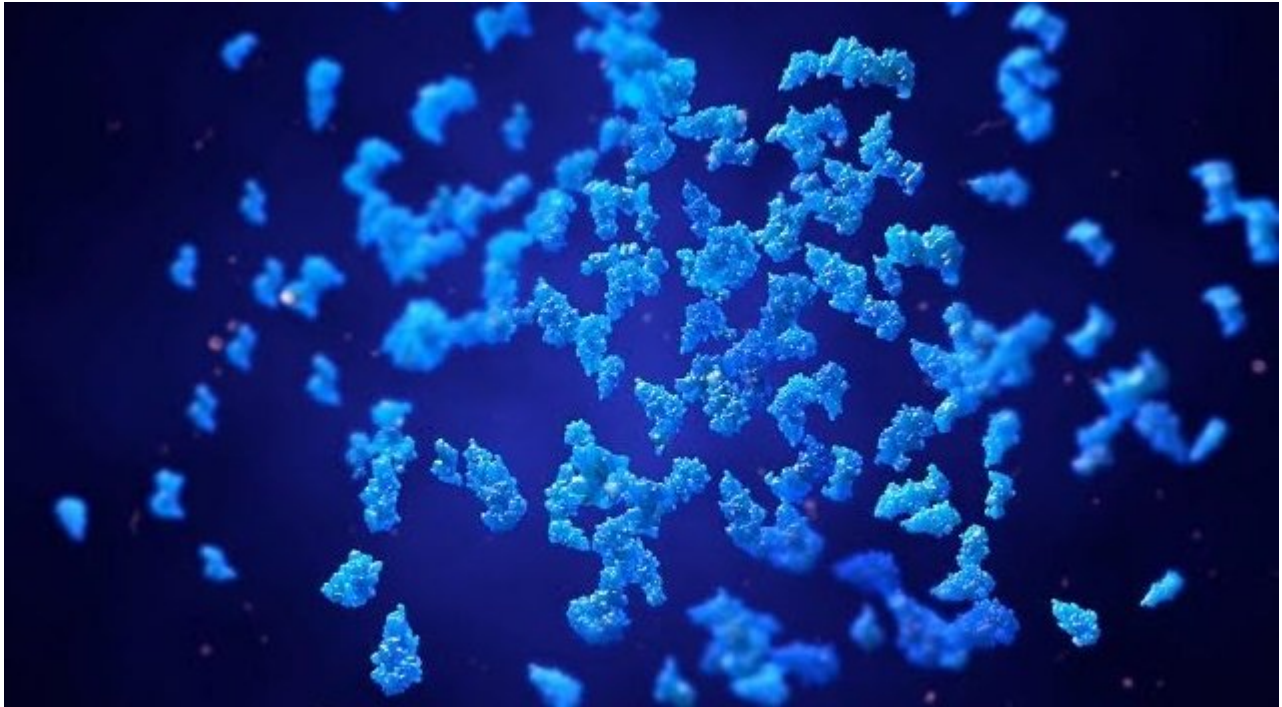


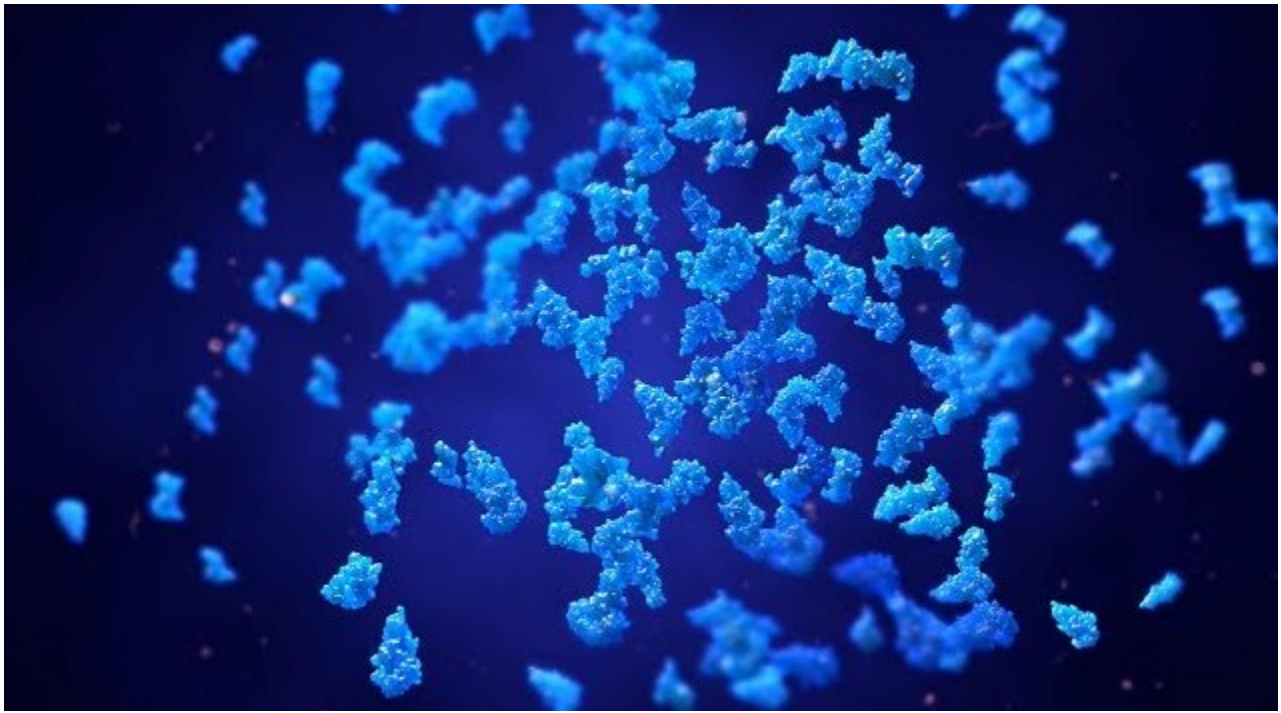
Scientists use AI to Learn the Intricate Language of Biomolecules

Category: Business

written by News Mall | November 4, 2025



Scientists at the University of Glasgow have harnessed a powerful supercomputer, normally used by astronomers and physicists to study the universe, to develop a new machine learning model which can help translate the language of proteins.



University of Glasgow Researchers Develop AI to Decode Protein Structures and Interactions

In a new study, published in [Nature Communications](#), the cross-disciplinary team developed a large language model (LLM), called PLM-Interact, to better understand protein interactions, and even predict which mutations will impact how these crucial molecules ‘talk’ to one another.

Early tests of PLM-interact, a protein language model (PLM), show that it outperforms competing models in understanding and predicting how proteins interact with one another. The team’s research demonstrates PLM-interact could help us better understand key areas of medical science, including the development of diseases such as cancer and virus infection.

The research team – led by Dr Ke Yuan from the University’s School of Cancer Sciences and the Cancer Research UK Scotland Institute, Prof Craig Macdonald from the School of Computing

Science and Prof David L Robertson, from the MRC-University of Glasgow Centre for Virus Research (CVR) – are developing these types of AI model to add much-needed detail on how diseases arise.

PLM-interact could also provide new insight into how viruses interact with their host species. In the future, it is possible this approach could even be used to predict a virus's pandemic potential and identify new drug targets.

Proteins are the main structural components of all cells and viruses and play a key role in biological processes by interacting with other proteins. Disruption of these protein-to-protein interactions (PPIs) is often linked with disease formation, including cancers and genetic diseases. Additionally, protein-to-protein interactions play an important role in viral infections, with viruses relying on the proteins in our cells to help them replicate and continue the infection process.

A better understanding of protein interactions would offer scientists vital new insights into disease and infections, potentially paving the way for the development of new therapies or vaccines. However, currently identifying protein-to-protein interactions experimentally can be both costly and time-consuming, and new ways to speed up the learning process are required.

PLM-interact was first '**trained**' on more than 421,000 human protein pairs and their interactions with data-processing support from the UK's DiRAC High Performance Super Computer

facility. Specifically, Tursa, originally developed to help theoretical physicists simulate aspects of the workings of the universe, provided the team with access to a highly optimised GPU cluster that helped them more quickly build and fine-tune the model, which involves more than 650 million individual parameters.

Dr Ke Yuan, one of the paper's corresponding authors, said, *"It's great to think that DiRAC, which was developed to help scientists understand the laws of nature from the smallest subatomic particles to the largest scales in the Universe, has helped us build this new model to explore the inner space of protein interactions instead."*

"Colleagues from our School of Computing Science provided support with the language modelling aspects of creating PLM-interact, but in order to train the model itself, we needed access to vast amounts of computing power. Working with DiRAC to tap into their GPU computing resources, as well as their training, technical support and software engineering resources, helped us do that much more quickly and effectively."

PLM-interact can predict protein interactions with between 16% and 28% more accuracy than other state-of-the art AI protein models. In addition, PLM-interact was able to accurately predict five key protein interactions that govern essential biological functions including RNA polymerisation and protein transportation. Notably other protein AI tools, including the Google DeepMind-powered AlphaFold3, were only able to predict one of the five protein-to-protein interactions.

Researchers were also able to show that PLM-interact could accurately identify the impact of mutations on protein interactions, both for mutations that cause negative consequences (including genetic diseases) and for mutations that inhibit essential protein-protein interactions, causing diseases such as cancers.

The research team also trained PLM-interact with a further 22,383 protein-to-protein interactions, this time from 5,882 human and 996 virus proteins. Once again PLM-interact outperformed existing protein models in its ability to predict how human and virus proteins interacted, demonstrating the model's power as an accurate virus prediction tool.

Prof David L Robertson, head of CVR Bioinformatics at the University of Glasgow, is the paper's other corresponding author. He said: *"The urgency to understand virus-host interactions during COVID-19 pandemic is a good illustration of why a tool like PLM-interact could be invaluable in the future. Being able to quickly and accurately gain insight into how viruses interact with our proteins could help us better understand virus emergence and disease risks, which in turn can help speed up the development of new treatments and therapies."*

"Our results are a very promising contribution to developing a system capable of predicting protein interactions at an unprecedented scale and level of accuracy. This is Dan Liu's, the paper's first author, PhD work and is a remarkably strong platform to build on for the future. We're already looking at expanding our team to help us explore the full potential of PLM-interact for a wide range of applications in the future."

The study, '**PLM-interact: extending protein language models to predict protein-protein interactions**' is published in Nature Communications. The work was funded by European Union's Horizon 2020 research and innovation 562 program, the Medical Research Council with support from Cancer Research UK, Prostate Cancer UK and the Biotechnology and Biological Sciences Research Council.

Notes

This work used the DiRAC Extreme Scaling service (Tursa, <https://www.epcc.ed.ac.uk/hpc-services/dirac-tursa-gpu>) at the University of Edinburgh, managed by the Edinburgh Parallel Computing Centre on behalf of the STFC DiRAC HPC Facility. Tursa is composed of 178 GPU nodes, each of which has 4xA100 GPUs for a total of 712 GPUs. A 200 Gb/s HDR Infiniband non-blocking interconnect supports running jobs very efficiently across multiple nodes, with 4x 200 Gb/s interfaces being available on each node. As it is GPU-based, Tursa is extremely energy efficient, delivering approximately 5 times more scientific throughput using only half the power when compared to its predecessor. This energy efficiency is increased further by the direct liquid cooling of all compute nodes. Tursa is supported by storage including a 5PB lustre fast file system and a large tape backup and archive service. The DiRAC service at Edinburgh was funded by BEIS, UKRI and STFC capital funding and STFC operations grants. DiRAC is part of the UKRI Digital Research Infrastructure.

About the University of Glasgow

The University of Glasgow has been changing the world since 1451. It is placed 76th in the QS World University Rankings

2024 and is a member of the prestigious Russell Group of leading UK research universities.

Its people have always been at the forefront of innovation, including eight Nobel Laureates, two UK Prime Ministers, three First Ministers of Scotland, 10 Fellows of the Royal Society and 11 Fellows of the British Academy.

It is committed to tackling the most pressing global challenges facing humanity and the natural world. It is delivering against the United Nation's 17 Sustainable Development Goals (SDGs) and has pledged to be carbon neutral by 2030. It is the first European university to declare it would divest from fossil fuels by 2024 and the first in Scotland to declare a climate emergency.

For more details, please visit: <https://www.gla.ac.uk>

