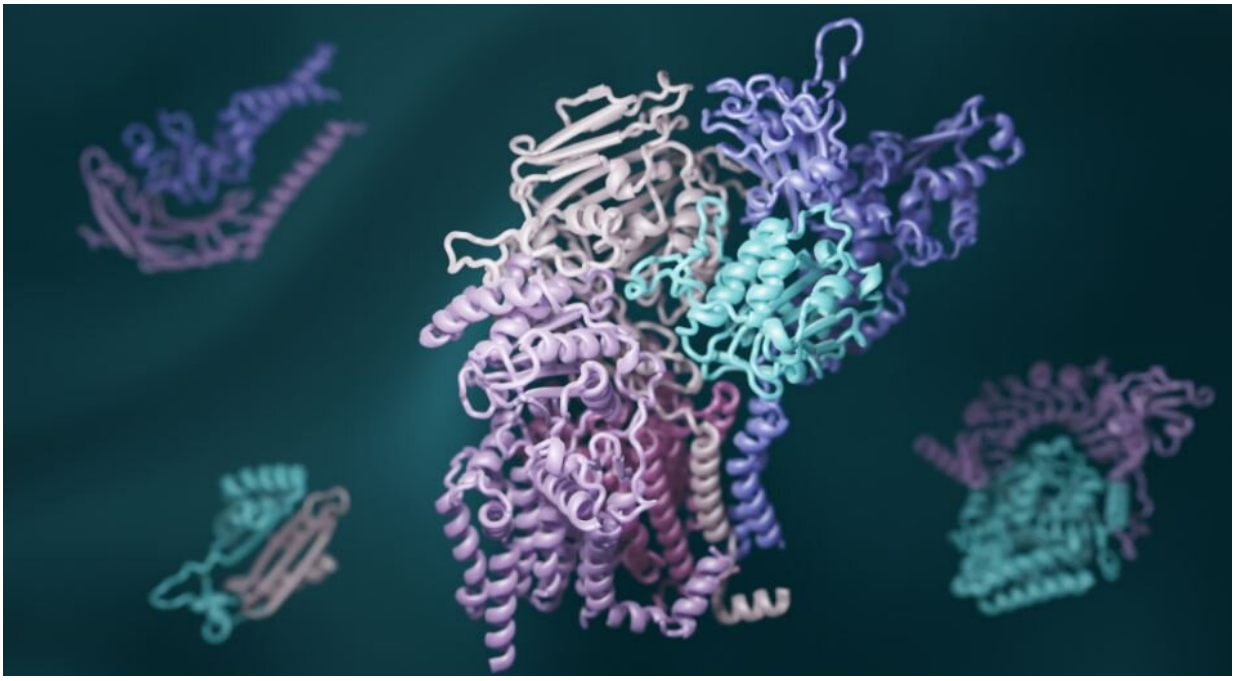
Deep learning reveals how proteins interact

November 23 2021, by Ian Haydon



3-dimensional computed structural modeling of protein interactions is made possible by deep learning and evolutionary analysis. Credit: Ian Haydon/UW Medicine Institute for Protein Design

Scientist are now combining recent advances in evolutionary analysis and deep learning to build three-dimensional models of how most proteins in eukaryotes interact. (Eukaryotes are organisms whose cells have a membrane-bound nucleus to hold genetic materials.)

The research effort has implications for understanding the biochemical

processes that are common to all animals, plants, and fungi. The open-access work appears Nov. 11 in *Science*.

As part of a multi-institutional collaboration, the lab of David Baker at the UW Medicine Institute for Protein Design helped guide this new development.

"To really understand the cellular conditions that give rise to health and disease, it's essential to know how different proteins in a cell work together," Baker said. "In this paper, we provide detailed information on protein interactions for nearly every core process in eukaryotic cells. This includes over a hundred interactions that have never been seen before."

Proteins, the workhorses of all cells, rarely act alone. Different proteins often must fit together to form precise complexes that carry out specific tasks. These can include reading genes, digesting nutrients, and responding to signals from neighboring cells and the outside world. When protein complexes malfunction, disease can result.

"This work shows that deep learning can now generate real insights into decades-old questions in biology—not just what a particular protein looks like, but also which proteins come together to interact," said senior author Qian Cong, assistant professor of biophysics at the University of Texas Southwestern Medical Center.

To exhaustively map the interactions that give rise to protein complexes, a team of structural biologists from UW Medicine, University of Texas Southwestern Medical Center, Harvard University, and several other institutions examined all known gene sequences in yeast. Using advanced statistical analyses, they identified pairs of genes that naturally acquire mutations in a linked fashion. They reasoned that such shared mutations are a sign that the proteins the genes encode must physically interact.

The researchers also used new deep-learning software to model the three-dimensional shapes of these interacting proteins. RoseTTAFold, invented at UW Medicine, and AlphaFold, invented by the Alphabet subsidiary DeepMind, were both tapped to generate hundreds of detailed pictures of protein complexes.

"As computer methods become more powerful, it is easier than ever to generate large amounts of scientific data, but to make sense of it still requires scientific experts," said Baker, who is a University of Washington School of Medicine professor of biochemistry. "So we recruited a village of expert biologists to interpret our 3-D protein models. This is community science at its best."

The hundreds of newly identified protein complexes provide rich insights into how cells function. For example, one complex contains the protein RAD51, which is known to play a key role in DNA repair and cancer progression in humans. Another includes the poorly understood enzyme glycosylphosphatidylinositol transamidase, which has been implicated in neurodevelopmental disorders and cancer in humans. Understanding how these and other proteins interact may open the door to the development of new medications for a wide range of health disorders.

The protein structures generated in this work are available to download from the ModelArchive. The researchers thank and remember the late John Westbrook at the Protein Data Bank for his support in establishing formats and software code to allow efficient deposition of the models into the archive. The Science paper reporting the results was in preparation at the time of Westbrook's death.

The project on computed structures of core eukaryotic proteins complexes was led by Ian Humphrey, Aditya Krishnakumar, and Minkyung Baek, all at UW Medicine, as well as Jimin Pei at the

University of Texas Southwestern Medical Center. Collaborating institutions include UW Medicine, UT Southwestern, Harvard University, Wayne State University, Cornell University, MRC Laboratory of Molecular Biology, Memorial Sloan Kettering Cancer Center, Gerstner Sloan Kettering Graduate School of Biomedical Sciences, Fred Hutchinson Cancer Research Center, Columbia University, University of Würzburg, St Jude Children's Research Hospital, FIRC Institute of Molecular Oncology, and Istituto di Genetica Molecolare, Consiglio Nazionale delle Ricerche.

More information: Ian R. Humphreys et al, Computed structures of core eukaryotic protein complexes, *Science* (2021). DOI: 10.1126/science.abm4805

Provided by University of Washington School of Medicine