

The Nobel Prize in chemistry has been awarded to David Baker for succeeding in the seemingly impossible feat of building new proteins and to Demis Hassabis and John M. Jumper for solving the half-a-century old problem of predicting proteins' complex structures using an artificial intelligence (AI) model called AlphaFold that they developed. It used to take years of painstaking experiments to determine the 3D structure of a protein, which in many cases have only been partially determined. Determining structure is the first and most important step in determining protein function. Proteins are long ribbons in which the 20 different amino acid building blocks can be sequentially arranged to form innumerable combinations. Even if researchers know the sequence of amino acids in a ribbon, the ribbon can twist and fold in an astronomical number of possible shapes for each sequence, thereby making protein structure determination extremely challenging. For instance, if a protein consists of only 100 amino acids, the protein can assume at least 1,047 different 3D structures. Till a few years ago, of the 20,000-odd proteins found in humans, the structure of only a third was partially determined even experimentally. Hassabis and Jumper made protein structure prediction child's play. AlphaFold has now predicted the structure of almost all 200 million proteins from nearly a million species. Baker developed computerised methods to create proteins that did not previously exist and which, "in many cases, have entirely new functions".

In 2018, Hassabis and Jumper achieved an accuracy of 60% in predicting protein structure. In 2020, the Al model's performance was comparable with X-ray crystallography. Though still not perfect, the Al model estimates how correct the structure that has been produced is, allowing researchers to know the reliability of the prediction. The code for the Alpha Fold model has been publicly available since 2021, and the Al tool has been used by more than two million people from 190 countries. Baker used his computer software Rosetta to generate new proteins that never existed naturally. Instead of predicting the protein structure based on amino acid sequences, he created new protein structures and used Rosetta to determine the amino acid sequence by searching a database of all known protein structure. Rosetta then optimised these fragments and proposed an amino acid sequence. Like Hassabis and Jumper, Baker too made the code for Rosetta freely available so that researchers can develop the software and find new areas of application.



GS