Check for updates

Method of the Year 2021: Protein structure prediction

Deep Learning based approaches for protein structure prediction have sent shock waves through the structural biology community. We anticipate far-reaching and long-lasting impact.

The potential to predict protein three-dimensional (3D) structures given a linear sequence of amino acids has captivated computational biologists for decades. While considerable progress had been made in the field, no approach had been able to reliably produce models that approached, let alone matched, the quality of experimentally determined structures. In the past year, the deep-learning-based methods AlphaFold2 and RoseTTAfold have managed to achieve this feat over a range of targets, forever altering the course of the structural biology field. More impressively, a collaboration between the European Molecular Biology Laboratory and DeepMind has predicted structures for over 350,000 proteins for 21 model organisms and made them freely available at the AlphaFold Protein Structure Database — with plans for expanding predictions to millions of structures in 2022. For these remarkable achievements, we have chosen protein structure prediction as the Method of the Year 2021.

The 3D shape of a protein dictates its biological function and provides vital information for potentially altering it to provide useful biotechnology tools or modulating its function. Solving structures experimentally is a slow and laborious process, and despite recent method advances, especially in cryo-electron microscopy (cryo-EM), it remains challenging. Computational researchers have always believed that a theoretical approach to solving the 'protein folding problem' would be feasible given a sequence of amino acids — the building blocks of proteins — and sufficient understanding of their biochemical and biophysical behavior. Numerous approaches have been explored over the past several decades, but historically, progress has occurred in short bursts with long phases of stagnation. The biennial Critical Assessment of Structure Prediction (CASP) protein-folding challenge, a blind competition held since 1994, has monitored and facilitated this progress. Challenge participants predict structures of particularly difficult proteins whose structures have been experimentally resolved but not yet released to the public.

A year ago, at the CASP14 meeting, AlphaFold2 from DeepMind outperformed all other approaches, and by a wide margin. On average, the fraction of a protein structure that AlphaFold2 correctly predicted crossed the 90% mark. A leap in performance of this magnitude was frankly not anticipated for another decade or so. It was therefore not a surprise that many deemed the protein folding problem essentially solved.

AlphaFold's success can be attributed to its neural network architecture and the training procedure that takes into account the available 3D structures of experimentally resolved proteins. In a Comment, AlphaFold developers John Jumper and Demis Hassabis describe the inner workings of the algorithm and its anticipated impact on the broader structural biology field.

Inspired by AlphaFold's approach, while the paper and related code were not yet released, an academic team led by David Baker developed RoseTTAFold, which performs nearly as well. Minkyung Baek and Baker discuss these new approaches in a Comment.

Admittedly, none of this would have been feasible without the availability of a large volume of experimental structural data serving as a training data resource for deep learning. Over the past 50 years, structural biologists have arduously solved the structures of over 170,000 proteins and openly shared these in a central macromolecular data archive, the Protein Data Bank (PDB). Fortuitously, this decision to openly share data at a time when data repositories were hardly the norm turned out to be one of the best investments for the field.

A new computational race has started. Since publication, both AlphaFold and RoseTTAFold have been further optimized to predict multi-protein complexes. Several other preprints are available that extend the AlphaFold method or apply it to more specific problems, such as predicting protein dynamics and ligand binding. Deep learning is also making an impact on the RNA structure prediction field. A Comment from David T. Jones and Janet M. Thornton examines AlphaFold, its ongoing impact on structural biology, and the caveats of predicted structures.

The burning question, however, is, now that it is possible to predict accurate structures for the large majority of proteins, what lies in the future for experimental structural biology?

In our opinion, having a potential structure already in hand gives structural biologists a massive head start in tackling more complex and interesting biological questions, but experiments will continue to remain important for testing hypotheses based on these predicted structures. In a Comment, Sriram Subramaniam and Gerard J. Kleywegt discuss how the future of structural biology will involve a stronger partnership between structure prediction and the experimental techniques of cryo-EM and cryo-electron tomography — in particular, to capture protein conformational dynamics and in situ structural complexity.

An inadequacy in our understanding of protein structure and function pertains to intrinsically disordered regions, which adopt specific secondary structures only when interacting with a binding partner. Around 30% of regions in the human proteome are estimated to be intrinsically disordered. More generally, discerning the structures that proteins adopt exclusively in a functional context is not feasible from static structure predictions. Abbas Ourmazd and colleagues argue for a pivot to predicting protein function directly from amino acid sequence in their Comment. We expect this to become an important focus for the field.

Our Technology Feature presents personal perspectives from the scientific community on the leap that AlphaFold delivers. The excitement is palpable. These methods have provided a true paradigm shift, and we look forward to seeing many exciting new methods spurred by this advance.

We hope you enjoy reading this special issue. We also highlight technologies we expect and hope to make a splash in the near future in our Methods to Watch section.

Here's to a happy 2022! ❐