STRUCTURAL BIOLOGY

Deep learning 3D structures

Recent developments in deep-learning-based methods improve protein structure prediction.

Protein structure prediction has been an active area of research for several decades, and theoretical methods have given insight into the structures of experimentally intractable proteins. In parallel, as experimental methods to determine protein structures have improved, the availability of larger quantities of higher-quality structural data has resulted in improvements in training-data quality, and consequently in the accuracy of these predictive algorithms. The ultimate goal would be to accurately predict the 3D structure of a protein from only its sequence; this is of course easier in cases where the structure of a close homolog is available.

For proteins lacking a close homolog, accurate structure prediction remains a challenge. Evolutionary covariance data have been used to enhance structure prediction. Multiple-sequence alignments (MSA) of sequences related to the target sequence are used to identify amino acids that show correlated changes through the course of evolution, the rationale being that these coevolving residues will lie in close proximity or contact in the 3D structure of the protein. These contact maps have been incorporated with some success in several popular approaches.

Deep-learning-based methods demonstrated high accuracy in the recent 13th Critical Assessment of Protein Structure Prediction (CASP13) structure prediction challenge and were among the top performers in the free modeling (FM) category (in which there are no available homologs). Rookie entrant AlphaFold, from Google's DeepMind, won the competition (Senior et al.). It predicted the greatest number of correct structures in the FM category — 24 out of 43 proteins — and performed better than or comparably to other methods in the template-based category (although AlphaFold did not use a template).

The accuracy of the method comes from the high accuracy of distance predictions. AlphaFold employs a convolutional neural network trained on protein structures from the Protein Data Bank. Given an input sequence and its MSA, it predicts pairwise distances and torsion angles between the residues. These distances are optimized

using gradient descent minimization to obtain well-packed protein structures. The advantage of using distances over contacts is that they provide more specific information about the structure. In addition, the neural network also provides information about the variances of their distance predictions, which indicates the level of confidence that should be associated with each prediction, explains Andrew Senior, from DeepMind, London. They took on the problem of protein structure prediction as deep learning challenge, but the DeepMind team intends to keep working on the problem to further improve the algorithm's predictive capabilities.

Building on DeepMind's advances, David Baker's research group at the University of Washington, Seattle, and collaborators have developed transform-restrained Rosetta (trRosetta). "trRosetta uses both residue-residue distances and orientations, which gives richer information on the structure compared to distances only," explains Baker. The web tool is available at <https://yanglab.nankai.edu.cn/trRosetta/>. Their publication (Yang et al.) highlights how this approach works with a Rosetta-based optimization scheme and combines the predicted information with additional components of the Rosetta energy function to generate protein models. Blind reanalysis of the CASP13 targets gave slightly better results than AlphaFold's performance at the competition, although the researchers recognize that the extremely challenging problem of protein folding requires building on each other's success. The Baker lab is looking to expand the method to protein-protein interaction modeling and protein design.

Arunima Singh

Published online: 4 March 2020
<https://doi.org/10.1038/s41592-020-0779-y>

Research papers

Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

Yang, J. et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl Acad. Sci. USA* **117**, 1496–1503 (2020).

**Working together
we open the potential
of scientific innovation**

Our enzymes have been discovering facts about life for decades.

That's why Worthington, the primary enzyme producer, has more citations in respected journals like *Nature Methods* and rank higher in search engines like *Bioz*®.

COLLAGENASES → **99% Bioz Rating**
 DNASE I → **99% Bioz Rating**
 PAPAIN → **99% Bioz Rating**

Unlock the power of Worthington with our new **Introduction to Enzymes** guide. Go to:
Worthington-Biochem.com