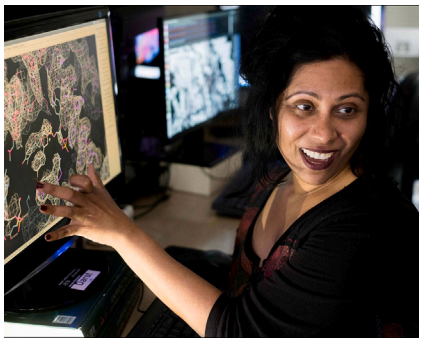
CellPress

Molecular Cell

Voices

AlphaFold: Research accelerator and hypothesis generator

To celebrate the 50th anniversary of Cell Press and the *Cell* focus issue on structural biology, we discussed with scientists working across diverse fields how AlphaFold has changed their research and brought structural biology to the masses.



Elizabeth A. Campbell
The Rockefeller University, USA

A greater appreciation for AI

AlphaFold2 has revolutionized biological research by providing researchers with high-confidence structural models for proteins, offering valuable insights into their functions. Trained on experimentally derived structures, this artificial intelligence (AI) program extends its impact beyond structural biology, fostering a greater appreciation for AI within the scientific community.

As a structural microbiologist focused on therapeutic applications, my research centers on pathogenic bacterial and coronaviral transcription complexes. Given that transcription is a validated therapeutic target, our efforts span both fundamental processes and pathogen-specific regulation, aiming to uncover new biology and identify potential inhibition targets. AlphaFold2 plays a crucial role in generating preliminary models for new transcription factors, guiding low-resolution cryoelectron microscopy (cryo-EM) modeling within larger transcription complexes, including the RNA polymerase (RNAP). Additionally, it aids in designing folded and active fusion proteins, facilitating the assessment of individual domain functions *in vivo* and *in vitro*. However, AlphaFold2's current limitations include its inability to model larger hetero-macromolecules, relying on previous structures for docking regulatory factors.

Despite its impact, AlphaFold2 faces challenges in predicting dynamic choreography and conformational changes, particularly in processes like the unwinding of DNA by RNAP during transcription. Understanding intricate transcription cycle steps necessitates context, rules, and observations beyond AlphaFold2's current scope.

The advancing methodologies of RoseTTA2Fold All Atom and AlphaFold's latest in modeling protein-DNA/RNA/small-molecule interactions drive *in silico* and experimental structure-based drug discovery. Recognized for its power in *in silico* docking, models in the AlphaFold Database are increasingly utilized by researchers, including Jiankun (JK) Lyu, who is assessing its potential and accuracy to template structure-based ligand discovery. This application is poised to thrive as these deep-learning-based structure-prediction methods further refine their capabilities in modeling protein-ligand structures.

A current limitation in AlphaFold2 in this arena lies in its dependence on a restricted number of ligand-bound structures for training, necessitating additional experimentally derived data. The forthcoming deep-learning-based programs may struggle to capture ligand-induced conformational changes without prior experimental derivation. For instance, our recent cryo-EM work on the coronaviral RNAP revealed diverse nucleotide binding poses in the NiRAN domain. Such predictions will challenge AlphaFold's capabilities for binding configurations, water interactions, and accompanying conformational changes. This highlights the crucial role of complementary experimental methods in drug development and underscores AlphaFold's need for expanded structural datasets.

Concerns about potential errors propagating also arise, as more scientists use AlphaFold2 to predict structures. Incorporating a code within the folding algorithm to exclude AI-derived models to address this concern may be challenging given the growing use of AlphaFold2 to build hybrid structures that are then deposited in the PDB database of experimentally derived structures.

In conclusion, despite its impressive accuracy, AlphaFold remains an evolving predictive, albeit powerful, program requiring validation. Currently emerging are integrative methods combining AlphaFold's prediction with genetics, biochemistry, molecular dynamics, and structure/function experiments. These experimental methodologies promise to add scientific rigor to AlphaFold's models.

Helen Walden
University of Glasgow, UK

Generating structurally literate biologists

When AlphaFold was released, we could see immediately how transformative it would be for our work as structural biologists. I view AlphaFold as "the luckiest postdoc in the lab"—we determine structures as a means to understanding function, generating new hypotheses, or explaining the mechanism of a particular protein or complex, not just to see what a protein looks like. The immediate relief is that we no longer have to spend 10 years growing crystals in order to get some physical insight into how a system might work. The huge difference with AlphaFold is the accuracy of the models; we've always used secondary-structure prediction in the design of constructs and to make educated guesses, but the level of accuracy is truly game changing. It does have limitations, particularly in dealing with intrinsically disordered regions. However, the limitations are mitigated by the confidence measures and ease of replicates, provided people pay attention to them. Of course, the power of AlphaFold comes from the huge efforts of the structural biology community, and that dependency is not yet weaned; I often joke that when we find something AlphaFold doesn't predict, it's only a matter of time and that we are simply feeding the machine, for the training sets will only improve.

We use it as a quick way to generate testable hypotheses; before AlphaFold and ColabFold, if we had a "hunch" that two proteins might interact based on them being in the same pathway, we would have to express, purify, and test the hunch, which would optimistically take a few months. Now we can run as many potential possibilities as we can think of and focus the labor-intensive efforts on those that appear plausible.

Many colleagues comment that structural biologists are now "out of a job." Personally, I think they are very mistaken. If anything, AlphaFold is as revolutionary as the recombinant DNA technologies of the 1980s and likely to breed a whole new generation of structurally literate biologists. Gone are the days where a protein could be chopped into small pieces based on linear sequence in order to determine "functional" sites. There is no longer any rationale for a structurally agnostic approach in mechanistic studies, and all biologists now need to know how to interpret structural (chemical) entities.

Johannes C. Walter
Harvard Medical School, USA

Toward a global "predictome" of PPIs

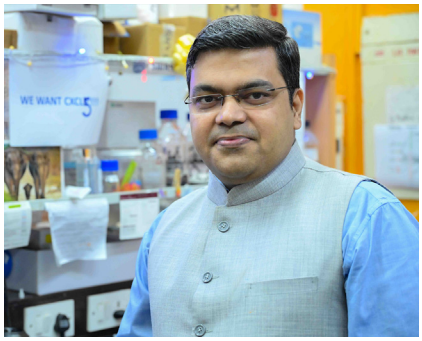
Protein-protein interactions (PPIs) are integral to virtually all biological processes. Stable PPIs form multi-subunit assemblies; transient interactions underpin complex and dynamic processes such as DNA replication and signaling. Therefore, as a first step to understand how a protein works, we usually ask "what other proteins does it bind to?" Until recently, no reliable methods were available to rapidly and comprehensively detect endogenous PPIs.

Into this void stepped the structure prediction algorithm AlphaFold-Multimer (AF-M), a variation of AlphaFold designed to predict the structure of PPIs. My laboratory recently used AF-M to uncover the function of an enigmatic protein called DONSON. We knew that DONSON was somehow required for assembly of the replicative CMG helicase, but how DONSON functions was unclear. Having failed to find DONSON interactors by conventional means, we used AF-M to screen scores of DNA replication proteins for potential DONSON partners. The top four virtual hits pointed to DONSON functioning as a chaperone that delivers an essential helicase co-factor to its binding site on CMG. Site-directed mutagenesis of DONSON and complementary mutations in its predicted partners confirmed the AF-M model.

Seeing how AF-M accelerates mechanistic discovery was amazing, inviting its widespread use. However, it is crucial that every structure prediction be rigorously tested. We recommend that for every predicted PPI, interfacial mutations be made in both binding partners to verify that they prevent complex formation and disrupt function. Needless to say, experimental structure determination is still essential to obtain precise residue, backbone, and domain orientations, which are critical for a complete understanding of biological mechanisms.

I am most excited about AI's potential to systematically identify novel PPIs on a large scale. We have initiated such an effort by "folding" the core human genome maintenance proteins with each other using AF-M, yielding tens of thousands of structure predictions (https://predictomes.org/). No doubt, all 20,000 human proteins will eventually be folded with each other. However, at this scale (200 million unique pairs), even a small percentage of false-positive predictions would obscure meaningful interactions. Therefore, a major challenge is to identify biologically relevant structure predictions, which will probably involve a combination of experimental (e.g., cross-linking mass spectrometry) and computational approaches.

If this challenge can be met, I expect a new era will be upon us in which a well-curated cellular "predictome" suggests first drafts of molecular mechanisms and drives a systems-level understanding of the interactions underlying cell and organismal physiology.



Arun K. Shukla
Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, India

Potential for virtual drug screening

We work with G-protein-coupled receptors (GPCRs), one of the most versatile and incredibly dynamic membrane proteins, with tremendous therapeutic implications in several human disorders. We have used models predicted by AlphaFold to guide our efforts to build GPCR structures in cryo-EM maps and compared the experimental structures with AlphaFold predictions. The ability of AlphaFold to predict the structure of the receptor core region has been robust, and accessibility has been impressive. On the other hand, accurate prediction of highly flexible loop regions and intrinsically disordered segments, such as the carboxyl terminus, still requires further improvement. This is perhaps true for other membrane proteins, including ion channels and transporters, as well. Structural dynamics of GPCRs is an important driver of the range and diversity of signaling through these receptors, including biased agonism and context-dependent interactions. Visualizing distinct conformational states of GPCRs with confidence still remains a challenge for AlphaFold, although it may change positively as the structural coverage expands further. Along the same lines, accurate prediction of the interaction interface of GPCRs with their signal transducers, especially those governed by post-translational modifications, is still in its infancy, although recent developments appear to have made significant strides in predicting biomolecular interaction interfaces in general.

It is interesting that some people are starting to use the term "AlphaFold structure" even in scientific talks and even appear to carry an impression that AlphaFold is on the way to substitute experimental structure determination processes carried out using crystallography or cryo-EM. While the possibilities that AlphaFold have unfurled are enormous, it is also important to understand the challenges that still need to be surmounted through further optimization and development. For example, structure-guided ligand screening on GPCRs has emerged as a major direction in the field as the structural coverage is expanding at a staggering pace. An obvious direction that could be immensely impactful if it works is the use of structural templates of GPCRs predicted by AlphaFold for virtual screening of drug-like molecules. Although some promising evidence has started to surface recently to suggest the feasibility, at least for some targets, future studies are still needed to broadly assess the robustness and generality of this approach. Despite some limitations, AlphaFold has been a groundbreaking development that should be celebrated widely and put to good use, including widespread and collective efforts to make it more effective for a multitude of targets and applications.

Martin Beck
Max Planck Institute of Biophysics, Germany

Helping solve the protein puzzle with confidence

For many years, understanding the architecture of the human nuclear pore complex has been our flagship project. We approached this problem, which may be conceptualized as solving a big 3D puzzle with 1,000 protein pieces, by integrative modeling. We used cryoelectron tomographic maps of entire human nuclear pores as a framework that we annotated step by step with high-resolution structures of the individual nucleoporins. Thereby, one caveat always was that we had to project structural knowledge obtained from various eukaryotic species into the human system. Nucleoporins are neither very well conserved nor very accessible to experimental structure determination. The nucleoporin structures solved by many laboratories in the field thus scatter over various species. As a consequence, homology modeling of human nucleoporins was our bread-and-butter work but, at the same time, somewhat of a dark art. An objective score that would assess the quality of our homology models was missing. AI-based structure prediction has changed the way we operate. It allowed us to obtain structural models of basically any human nucleoporin with an attached confidence score. And not only that. It would also allow us to assess candidate protein pairs for possible interfaces, thus enabling an *in silico* complex walking approach. The resulting structural models often explained complementary experimental data that we had already available and did not understand beforehand. They were so accurate that further pieces of the puzzle would seamlessly fall into place. As of today, AI-based structure prediction has become part of our routine workflows. We use it to build very large structures almost comprehensively, to predict protein interfaces and short linear motifs, or to pick interactors from a list of candidates. AI-based structure prediction is an amazing research accelerator and hypothesis generator. We very much look forward to new implementations that will incorporate post-translational modifications, small molecules, and protein-nucleic acid interactions!

Lori A. Passmore
MRC Laboratory of Molecular Biology, UK

Fast-forwarding to testable hypotheses

In 2020, the emergence of new methods that predict protein structure opened the field to everyone—it is now possible to model the structures and interactions of most proteins. These can be used to generate testable hypotheses about protein function and interactions. Many more predictions, including the prediction of ligand and nucleic acid interactions and conformational states, are becoming possible as the field advances. And *de novo* protein design is now becoming a reality.

My lab studies large multi-protein complexes that regulate mRNAs and DNA repair. We want to understand how these complexes are dynamically regulated to control their activities and ensure fidelity. However, our experimental investigation of many of the critical interactions has been confounded by protein flexibility, instability, and low-affinity interactions. This is where AlphaFold has started to revolutionize our work. In particular, we use AlphaFold to predict the molecular details of interaction within the complexes we study. So instead of spending a lot of time identifying protein-protein interactions experimentally, we are now fast-forwarding to test whether disruption of putative interactions affects complex assembly, *in vitro* biochemical activities, and cellular functions. In addition, AlphaFold "screens" are now widely used to predict entirely new protein interactors, for example, to discover new components of a pathway.

So where does experimental structure determination fit in? Some proteins are still recalcitrant to structure prediction. For example, interactions with intrinsically disordered regions and coiled coils seem particularly problematic, generating many false positives and false negatives in AlphaFold. But more generally, we are not able to figure out all the details simply by using structure prediction methods—at least for now. For example, how do different structures, conformational states, ligands, and post-translational modifications fit together in time to regulate a cellular process? What are the molecular details of an enzymatic reaction, and how is this regulated? Experimental work is still necessary to understand mechanisms.

Moving forward, I am particularly excited about the combination of structure prediction and advanced cryoelectron tomography methods to view the molecular details of cellular processes *in situ*, something that is not routinely possible with current methods. Personally, I am excited to embrace new AI-based methods to advance our mechanistic understanding of life.

H. Eric Xu
Shanghai Institute of Materia Medica, Chinese Academy of Sciences, China

Diligence in evaluating confidence metrics

AlphaFold continues to astound with its rapid advances in predicting biomolecular structures. The latest iteration detailed in the Google DeepMind report showcases dramatic improvements in modeling complexes. AlphaFold can now jointly predict bound structures including proteins, ligands, nucleic acids, and modifications—a technical feat.

We've utilized AlphaFold models in cryo-EM structure determination, with most backbone structure traces reliably matching in density. However, improvements are still needed for side-chain conformations. The new ability to directly incorporate experimental data will be a game changer, driving accuracy beyond our imaginations.

By co-modeling interfaces, this new AlphaFold captures intricate protein-ligand and antigen-antibody complexes. It even outperforms specialist predictors of protein-DNA and protein-RNA complex structures. The promise for antibody engineering and nucleic acid design is tremendous. However, diligence in evaluating confidence metrics is still essential, as lower-confidence regions require experimental validation. We must remember that these are models, not experimental structures.

The future looks bright for leveraging AlphaFold in drug discovery too. It demonstrates remarkable accuracy in predicting protein-ligand complexes, sometimes matching or exceeding traditional computational docking. As algorithmic small-molecule modeling improves, AlphaFold could revolutionize structure-based design. However, the community as a whole is eagerly waiting for accessibility to the new version of AlphaFold to test its utilities in drug design.

Overall, AlphaFold has enriched our research programs in myriad ways. By providing a first glimpse at previously unseeable structures, it accelerates discovery across biomedical frontiers. Of course, experiments remain essential, but combining AlphaFold's models with data validates and refines the structures, producing a whole greater than the sum of its parts. I am excited to continue exploring the myriad ways this technology can enhance our understanding of biological mechanisms and enable rational drug design.

DECLARATION OF INTERESTS

J.C.W. is a cofounder of MOMA Therapeutics, in which he has a financial interest. L.A.P., A.K.S., and H.E.X. are members of the *Molecular Cell* advisory board.