Critical Assessment of Methods for Predicting the 3D Structure of Proteins and Protein Complexes.

Shoshana J Wodak, Sandor Vajda, Marc Lensink, Dima Kozakov, Paul A Bates

Annual Review of Biophysics

Critical Assessment of Methods for Predicting the 3D Structure of Proteins and Protein Complexes

Shoshana J. Wodak,[1] Sandor Vajda,[2,3] Marc F. Lensink,[4] Dima Kozakov,[5,6] and Paul A. Bates[7]

[1]VIB-VUB Center for Structural Biology, Vrije Universiteit Brussel, Brussels, Belgium; email: Shoshana.wodak@gmail.com

[2]Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA; email: vajda@bu.edu

[3]Department of Chemistry, Boston University, Boston, Massachusetts, USA

[4]Univ. Lille, CNRS, UMR 8576-UGSF-Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France; email: marc.lensink@univ-lille.fr

[5]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York, USA; email: midas@laufercenter.org

[6]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York, USA

[7]Biomolecular Modelling Laboratory, The Francis Crick Institute, London, United Kingdom; email: paul.bates@crick.ac.uk

Annu. Rev. Biophys. 2023. 52:183–206

First published as a Review in Advance on January 10, 2023

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-102622-084607

Keywords

protein structure prediction, critical assessment of structure predictions, CASP, critical assessment of predicted interactions, CAPRI, artificial intelligence, protein interactions

Abstract

Advances in a scientific discipline are often measured by small, incremental steps. In this review, we report on two intertwined disciplines in the protein structure prediction field, modeling of single chains and modeling of complexes, that have over decades emulated this pattern, as monitored by the community-wide blind prediction experiments CASP and CAPRI. However, over the past few years, dramatic advances were observed for the accurate prediction of single protein chains, driven by a surge of deep learning methodologies entering the prediction field. We review the main

scientific developments that enabled these recent breakthroughs and feature the important role of blind prediction experiments in building up and nurturing the structure prediction field. We discuss how the new wave of artificial intelligence–based methods is impacting the fields of computational and experimental structural biology and highlight areas in which deep learning methods are likely to lead to future developments, provided that major challenges are overcome.

Contents

INTRODUCTION

The problem of predicting the native 3D structure of a protein from its amino acid sequence has occupied a prominent position in protein modeling research for over five decades owing to its inherent scientific interest and to the many potential applications that robust structure prediction algorithms would offer in areas such as the prediction of function from genome sequence and designing new drugs to treat disease (106). In comparison, although the important functional role of protein–protein interactions and complexes was recognized in the 1960s, methods for predicting the structure of complexes has become a booming research area only since the turn of the century (141), fueled by the realization of the ubiquitous involvement of protein complexes in nearly all cellular processes.

The past decade has seen major advances in both types of prediction methodologies, due to a variety of factors. Notable has been the application of artificial intelligence (AI) methods, culminating with the recent phenomenal success of the AI-based algorithm AlphaFold2 by DeepMind in predicting the structures of single protein chains to accuracy levels rivaling those of experimental methods (56). Important in nurturing and catalyzing these developments have been the blind prediction experiments of CASP (Critical Assessment of Structure Predictions) and CAPRI (Critical Assessment of PRedicted Interactions), which focus, respectively, on the critical assessment of methods for predicting the structures of proteins and those of protein complexes.

In this review, we outline the progress made in the methods developed for these two prediction tasks. We describe how the performance of prediction methods is evaluated by CASP and CAPRI and how progress is assessed. We highlight the role of blind predictions in building up the communities of method developers and shaping the field. We end by offering our view on the impact that the new wave of AI-based methods is having on the field of computational and experimental structural biology and where the remaining challenges lie.

PREDICTION OF PROTEIN 3D STRUCTURE FROM SEQUENCE

Computational analysis of protein structures was initiated in the 1960s by Shneior Lifson and his group, who extended the molecular mechanics approach developed for modeling small organic molecules to large molecular systems (42, 43). They introduced the Consistent Force Field (CFF) energy function, which led to the development of some of the most important all-atom potentials used today in protein modeling, including CHARMM (11), Amber (138), and ECEPP (50). All three potentials include covalent, noncovalent, and electrostatic energy terms, as in the original CFF, with some additional terms specific to each force field. These classical potentials have served well whenever various intrinsic properties of the protein needed to be investigated in a vacuum; however, they were proven to be inadequate for a thermodynamic description of stable compact protein folds in solution and unable to discriminate between native proteins and incorrectly folded models (99). The main reason for this was the failure to account for solvation effects, an important determinant of protein stability. These effects were usually incorporated by using these potentials in molecular dynamics (MD) simulations of the protein immersed in a box of explicit solvent molecules, an exercise that remained prohibitive for protein structure prediction due to its computational burden, leading to problems of convergence and inadequate conformational sampling.

The next step forward was the addition of implicit solvation terms to the classical potentials. An early approach was based on surface area–dependent empirical transfer free energy models used in conjunction with atomic solvation parameters (30). This was superseded by continuum electrostatic models evaluating the electrostatic contribution to the solvation free energy. The latter were formulated using the finite difference Poisson–Boltzmann (PB) method (38) and various approximations to the Generalized Born (GB) treatment (101). Augmented with a surface area–based term to represent the nonpolar contribution to solvation and integrated into the classical potential functions, the resulting force fields could identify the native states of peptides or proteins, albeit with limited accuracy (45, 153).

The shortcomings of models based on molecular mechanics and continuum electrostatics led to interest in extracting effective potentials from experimentally determined protein structures. A frequently used approach to derive such potentials consists of computing frequencies of structural features (structural frequencies) and converting these frequencies into free energy contributions (127). Following this approach, many statistical (or knowledge-based) potentials were proposed (52, 120). Most of these potentials used simplified residue-based representations of the protein, reminiscent of the coarse-grained potentials used decades earlier in protein folding calculations (86). These relatively simple, computationally efficient potentials helped score and rank predicted protein models. When combined with various energy optimization methods, they were also able to model the structures of very small proteins from their amino acid sequence in the so-called ab initio protein modeling approach. However, sampling the vast conformational space of average-size proteins remained a problem. Data on protein sequences and known protein structures have been increasingly relied upon to address this problem.

Evolutionarily related proteins adopt similar 3D structures (22); with the increasing number of experimentally determined protein structures, this property gave rise to the method of homology

modeling, also known as comparative or template-based modeling (29). The atomic-resolution structure of the target protein is modeled from its amino acid sequence and an experimental 3D structure of an evolutionarily related protein (the template). Aligned regions of the template backbone are simply copied into the target, whereas special prediction methods are used for adding loops in the nonaligned regions (32) and also for placing the side chains of nonconserved residues (32), and the resulting models are refined using molecular mechanics or MD methods.

A notable development in conformational sampling, which in some ways bridges template-based and ab initio methods, has been the use of fragment-based assembly approaches, whereby models are built from short contiguous backbone fragments (typically 3–15 residues in length) taken from proteins of known structure and assembled into full-length models using Monte Carlo simulated annealing or equivalent techniques (68, 123).

The next major advance in protein modeling was the effective use of coevolutionary information, enabled by the growing number of related sequences (39). The underlying hypothesis was that, if mutations occurring at two positions in the aligned sequences are correlated, then these positions are likely to form a contact in 3D space (105). Finding true evolutionary covariation between residues is difficult because one must minimize the effect of transitive correlations, i.e., indirect correlations that are observed, for example, when two residues contact the same third residue but do not actually contact each other. Transitive correlations can be removed by global statistical approaches involving direct coupling analysis (92), pseudolikelihood optimization (58), or machine learning (139). The approach was first used to identify residue pairs that are in contact and further extended to derive residue distance and dihedral angle distributions, all used as restraints in ab initio modeling (105, 149). The more recent neural network–based learning methods further extend the use of multiple sequence alignment to end-to-end protein structure prediction, achieving previously unimaginable accuracy for a significant fraction of proteins (56), as is discussed in the final section of this review.

PREDICTION OF THE 3D STRUCTURE OF PROTEIN COMPLEXES

Efforts to model protein–protein interactions began in the 1970s, driven by the desire to explain aberrant protein–protein interactions caused by a single point mutant in sickle-cell hemoglobin (Hb-S) (85). The first protein docking algorithm, formulated as the task of modeling the atomic structure of a native protein complex from the structures of its components, was developed a few years later. This early incarnation of docking treated the interacting proteins as rigid bodies, used a coarse-grained representation of the protein developed for protein folding calculations (86), and searched for large surface patches with complementary shapes. Shape complementarity was evaluated using the interface area formed by the contacting proteins (21), a geometric quantity representing the loss of solvent-accessible surface area upon binding, itself related to the hydrophobic contribution to the binding free energy (20).

Ab Initio Docking Methods

Over the following two decades, a variety of docking procedures were proposed (141, 142), including most notably the fast Fourier transform (FFT)-based methods (59) that currently dominate the field of protein docking. FFT-based methods enable speedy coarse-grained rigid-body searches capable of detecting shape complementarity, as well as the evaluation of different properties of protein interfaces, such as hydrophobicity (132) or electrostatic and van der Waals interactions (9). Following these advances, strategies were proposed to speed up high-resolution searches, required for accurately defining the molecular positions and orientations. These include the use of spherical polar Fourier expansion coefficients, shown to significantly accelerate the search for solutions that optimize properties of generated interfaces (114).

A series of rigid docking algorithms with variations on these fundamental principles (19, 35, 62, 114) underpin most of the docking procedures used today. One in particular forms the basis of a well-frequented automatic docking server, ClusPro (64), which enables the reduction of the search space from 6° to 5° of freedom by employing a Fourier transform in polar coordinate space, resulting in a 10-fold increase in speed over classical FFT approaches without compromising accuracy (102). A few alternative sampling algorithms, such as enhanced sampling Monte Carlo procedures (148), and algorithms incorporating heuristic methods based on particle swarm optimization (91) have also been quite effective.

Also worth mentioning are the so-called data-driven docking methods, which involve the incorporation of distance restraints obtained from biophysical or biochemical data into the modeling protocol, thereby reducing the search space for the location of the native complex. One of the first approaches using these principles, now operating as a publicly available server (24), is the program HADDOCK (28). Similar restraints-based methodologies have been added to other protein docking methods (104, 137, 145). More recently, intermolecular contacts derived from data on residue coevolution were also used as restraints in docking calculations but with only modest success (110).

Scoring docking poses. To further prioritize the large number of solutions (often numbering hundreds of thousands) produced by the docking calculations, these solutions are reranked using more sophisticated scoring functions. An important requirement for such functions has been that they be able to reliably percolate the most native-like binding modes to the top of the list. Awareness of this challenge has led to a major focus on the development of scoring schemes over the past two decades. There are a wide variety of such schemes, and they are often combined with model optimization. Use is being made of atom or residue pair potentials, sometimes in combination with classical potential energy terms but increasingly implementing different flavors of knowledge-based potentials. The latter are adapted from potentials developed for the structure prediction of single protein chains (46, 48, 151, 152). Among the most effective are methods combining knowledge-based potentials with the evaluation of interatomic contact areas using Voronoi tessellation (100), methods that enrich knowledge-based potentials with evolutionary relationships (98), and methods augmented with deep learning models (88) or replaced with such models (113). A notable example is the Rosetta all-atom multicomponent energy function (2), which has been used broadly in various molecular modeling applications including the evaluation of docking models. For many scoring schemes, the rank of native-like solutions can be bolstered by clustering the top-ranking docking poses based on the similarity of their interfaces and using cluster size and stability to perturbation (63) to rank models alone or as part of more complex ranking procedures (64).

Handling protein flexibility. Given that rigid-body search algorithms make up the core component of most docking procedures, it is not surprising that these procedures do poorly when the interacting proteins exhibit moderate to high levels of flexibility (26). Nevertheless, with modifications to some rigid-body docking algorithms for so-called soft docking—allowing for some atomic clashes to be alleviated subsequently by standard MD—this problem can be mitigated to some extent. Another strategy coined ensemble docking involves generating ensembles of conformers for individual components of a complex by MD (125) or normal mode analysis (27) and systematically docking conformers from both ensembles to one another; however, this strategy has had mixed results (72).

The inability to adequately address flexibility led to the design of algorithms that introduced protein backbone movements and sidechain repacking of putative interface residues during the sampling process, with some success for small to medium conformational changes upon complex formation (4, 10). However, to date, it remains debatable whether modeling protein flexibility using available methods improves the quality of docked models sufficiently to justify the typically

higher computational cost entailed (109). Nevertheless, new methodologies are continuing to address the problems associated with significant conformational change upon complex formation; some of these methodologies are part of a new wave of machine learning approaches (44).

Template-Based Docking

Template-based docking arose due to the increasing success of the homology modeling techniques for single protein chains described above. With the growing number of available experimentally determined protein structures, it was soon realized that homology- or template-based modeling may be extended to pairs of homologous complexes if at least some of their component parts show a degree of sequence similarity (3). This then led to the idea that the 3D structure of a complex can be modeled directly from the experimentally determined structures of other complexes. The accuracy of the method hinges on sensitive sequence searches and alignment to the appropriate complexed proteins (124). Interestingly, bearing in mind the constraints on accuracy described above, enough experimentally determined protein complex templates are available to model most native protein–protein interactions for any organism (71)—as underlined by the successful employment of the methodology in several recent CAPRI blind trials (75). Rapid searching for homologous complexes is now supported by annotated databases of such predicted relationships (70).

The processes of scoring and ranking models derived from template-based docking conform to the same principles as for classical docking but with the potential advantage of having to score and rank fewer models. However, there is an obvious caveat: If the modeled complexes do not comply with the principle of conservation of homologous interfaces, then the native complex will not be sampled. This contrasts with the ab initio methods, in which there is always a chance, provided that flexibility does not dominate, of at least having a near-native model in the complete list of models generated. Moreover, there are clearly certain categories of interactions that are not conducive to this form of docking; the classic example is antibody–antigen complexes, where evolutionary relationships between the binding partners are not expected to be prevalent, thereby limiting the options to ab initio docking methods (40).

Ab initio and template-based docking methods are clearly not mutually exclusive, and procedures are actively being developed to employ both to model the widest range of interactions possible, and to the highest level of accuracy, should the appropriate levels of sequence homology prevail (146); such pipelines have already been encoded into some automatic docking servers (111, 147). With the recent advances in deep learning approaches (see the final section) for modeling of both components and full complexes, these two principal methodologies are likely to become seamlessly merged.

THE BLIND PREDICTION CHALLENGES

CASP

CASP is a community-wide double-blind experiment for testing and comparing protein structure predictions (96). Every two years, sequences of soon-to-be experimentally determined protein structures are collected and passed on to registered predictors. Predictors fall into two categories: teams of participants, who usually have a period of three weeks to complete their work, and automatic servers, which must return a model within 72 hours, in principle without human intervention. Predictions are evaluated by independent assessors using well-developed criteria. CASP provides research groups with an opportunity to test their protein structure prediction methods and delivers an independent assessment of the state of the art in protein structure modeling to the research community and software users. The results show what progress has been made during the previous two years and expose where future approaches should focus to improve the methodology.

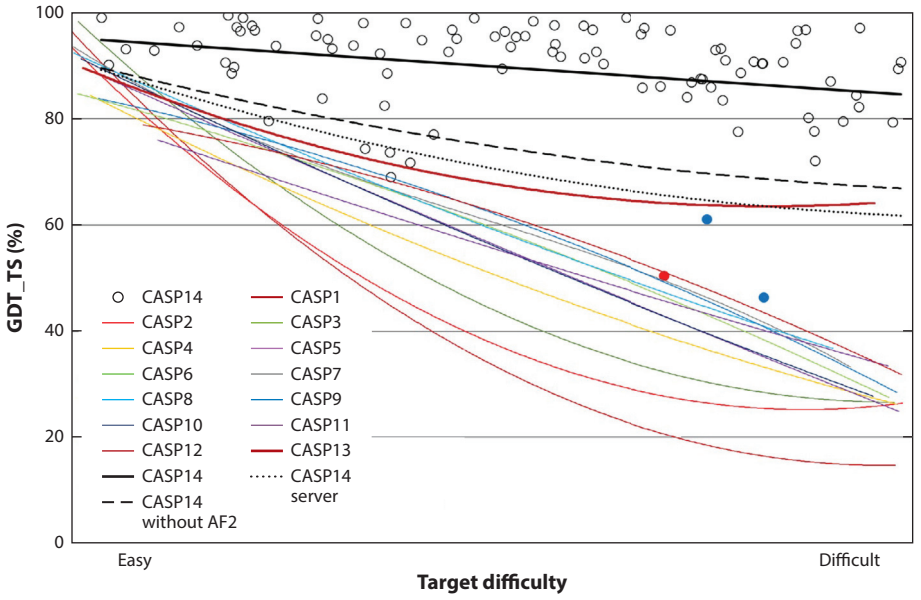


Figure 1

Backbone accuracy of the best models in each of the 14 CASP rounds. Individual target points are shown for CASP14. The two targets with the lowest agreement with experiment are NMR structures, colored blue; the red point represents the model of a subunit of a cryo-EM-derived large heteromeric structure. The agreement metric, GDT_TS, is a multiscale indicator of the closeness of the Cα atoms in a model to those in the corresponding experimental structure and is reported as a percentage. Because of experimental errors and artifacts, models with GDT_TS > 90 are considered compatible with experiment in backbone accuracy (67). Target difficulty is based on sequence and structure similarity to other proteins with known experimental structures. Figure reproduced from Reference 67, with permission from Wiley and Sons. Abbreviations: cryo-EM, cryo-electron microscopy; GDT_TS, Global Distance Test Total Score; NMR, nuclear magnetic resonance.

The CASP experiment had a fairly modest start in 1994 with 35 participating research groups (97). Targets were provided in three prediction categories: comparative modeling, fold recognition or threading, and ab initio folding (97). The results of CASP1 demonstrated a sobering failure of prediction methods using physics-based potentials, shocking the protein folding community. The only meaningful predictions were obtained using comparative modeling, and only for easy targets with closely related known structures. Such negative results made it difficult for protein scientists to continue publishing theoretical papers without participating in CASP. The preeminence of template-based approaches was further emphasized by CASP2 in 1996. By that time, CASP had become more widely recognized as a much needed blind experiment that had the potential to introduce a new era of reproducibility and openness in protein structure prediction. The number of participating groups grew to 70, and there was some improvement in the predictions of more difficult targets (**Figure 1**). Improved sequence alignment tools, the use of multiple templates, and fragment assembly approaches further improved results at CASP3 and CASP4, but after that, improvements remained essentially moderate until CASP13 in 2018. Although contact maps based on coevolutionary information were already present in several methods at CASP10, CASP11, and CASP12, their effective use with deep learning led to a jump in prediction quality, particularly for difficult targets (**Figure 1**), only at CASP13 (with approximately 100 participating groups). This improvement was already very significant, and CASP14 in 2020 led to a further revolution

by AlphaFold2, a neural network–based end-to-end prediction method that is discussed in more detail below. Predictions by other predictor groups and servers have also become much better (**Figure 1**).

The CASP predictions are evaluated using a variety of quality measures (65) that are listed on the Protein Structure Prediction Center website (**https://predictioncenter.org/casp12/doc/help.html**). The most important measure is the Global Distance Test Total Score (GDT_TS), shown in **Figure 1**. Another important measure is the Local Distance Difference Test, a superposition-free score that evaluates local distance differences of all atoms in a model, including validation of stereochemical plausibility. However, the ranking of CASP predictions is generally based on GDT_TS, computed over the alpha carbon atoms and reported as a percentage, with higher values indicating a closer fit of a model to a given reference structure (see **Figure 1** for details).

At CASP14, the targets were assigned to one of four classes of modeling difficulty, based on sequence and structure similarity to already experimentally determined structures: TBM-Easy for straightforward template modeling targets, TBM-Hard for more difficult homology modeling targets, FM/TBM for those with only remote structural homologies, and FM for the most difficult set with no detectable homology to known structures (TBM indicates template-based modeling, and FM indicates free modeling) (67). However, with the significant improvement in prediction quality, for the ongoing CASP15, the distinction between template-based and template-free modeling is eliminated. As shown in **Figure 1**, until CASP13, the predictions became substantially less accurate as the level of difficulty increased. However, this changed in 2018 with the introduction of deep learning methods at CASP13 that were able to yield predictions with GDT_TS over 60 for the most difficult targets. This trend was further strengthened at CASP14, where prediction quality started at a GDT_TS of approximately 95 and rarely went below 80. Although the outstanding performance at CASP14 is dominated by AlphaFold2, **Figure 1** shows that other groups also made substantial advances.

As shown in **Figure 1**, at CASP14, automated servers had similar performance to human groups (with the exception of the results for Alphafold2). Lastly, we note that the quality of many protein structure prediction servers is continuously evaluated by CAMEO (Continuous Automated Model EvaluatiOn), a fully automated assessment platform, which is a complement to the biannual CASP experiment (41, 115).

CAPRI

CAPRI is a community-wide experiment inspired by CASP. It was established in 2001 (51) to offer computational biologists the opportunity to test their algorithms in blind predictions of experimentally determined 3D structures of protein complexes, the targets, provided to CAPRI prior to publication. Experiments focusing on this prediction task were attempted only twice before, including once in 1996 by CASP (51), attracting only limited interest. Since its inception, CAPRI prediction rounds have been managed in collaboration with the Protein Databank Europe at the European Bioinformatics Institute (EBI) (**https://www.ebi.ac.uk/pdbe/complex-pred/capri/**).

Since the rate at which structures of protein complexes are being determined and offered as targets for prediction is slower than for single chain proteins, CAPRI runs prediction rounds on a rolling basis as targets become available (51). As in CASP, participants include automatic servers, which must return models within 72 hours, and human predictors who are given 6–8 weeks to complete their predictions. Initially limited to homo and hetero protein–protein complexes, the panel of targets diversified to include protein–peptide interactions and complexes of proteins with

RNA, DNA, and oligosaccharides (79). With time, target size and complexity also increased, especially with the availability of large multimeric complexes solved to high resolution by cryo-electron microscopy (cryo-EM). Recognizing the essential role that scoring functions play in identifying native-like association modes, CAPRI also offers a scoring challenge upon the completion of each prediction round. In this challenge, a larger set of anonymized models predicted by different groups and comprising both correct and incorrect binding modes is made available to all participants to test new scoring functions independently from docking calculations (79, 83). The data comprising the consolidated ensemble of predicted complexes made available in the CAPRI scoring experiments have been compiled in a freely available benchmark data set [the ScoreSet (84)] that was recently extended to include 170,310 models of diverse complexes predicted by different methods for all CAPRI targets whose structure has been published (**https://scoreset.org/**).

The way in which the prediction problem is formulated for a given target has likewise evolved over the years to ensure the best use of the available data on the known structures of single protein chains and complexes. In early CAPRI rounds, predictor groups were offered the coordinates of the unbound structures of the components of the complex to predict. Occasionally, the bound conformation of one of the components was provided as input in random orientation with sidechain coordinates stripped away (51). As the repertoire of 3D structures of single protein chains increased, thanks, in particular, to structural genomics initiatives (18), participants were invited to predict the structure of the target assembly starting from sequence information alone. This requires the integration of homology-based modeling of individual subunits with docking calculations or relying entirely on template-based modeling when adequate templates for the entire complex are available (see the section titled Template-Based Docking). Identifying adequate templates is not trivial and has been a task that members of the CAPRI community, who specialize in ab initio docking calculations, have had to learn to master using available resources such as HHPRED (36), PPI3D (23), or GalaxyWEB (61). The increased reliance on homology- and template-based modeling was further catalyzed starting in 2014, when CASP included the prediction of protein assemblies in their biannual prediction season in collaboration with CAPRI (81), a collaboration that has been continuing since.

Model Quality Measures

Objective criteria for independently evaluating the quality of predicted models by comparing them against the target structure are key components of blind prediction challenges such as CAPRI and CASP. When dealing with models representing protein complexes comprising two or more interacting partners, the evaluation criteria need to account for both local and global parameters of the molecular assembly. In CAPRI, the standard evaluation protocol involves the evaluation of three parameters for a given pair of interacting subunits (77): F_{nat}, the fraction of residue–residue contacts in the native interface that are recalled in the interface of the predicted complex; *i_rms*, the root mean square deviation of the backbone atoms of interface residues in the model versus the target; and *L_rms*, the root mean square deviation of backbone atoms of the smaller chain (ligand) of the model versus the target after superposition of the larger chain (receptor), quantifying the relative rigid-body displacement of the binding partners in the model versus the target. Based on ranges in the values of these three parameters, defined by expert evaluation and validated by the community, models are assigned to four discrete categories: high, medium, and acceptable quality and incorrect (77). For targets representing higher-order assemblies with multiple distinct interfaces, submitted models are evaluated by comparing each pair of interacting subunits in the model to each of the relevant pairs of interacting subunits in the target (81). The quality categories of individual interfaces are then used to derive a global quality score for the full assembly. Several

different formulations have been tested over time; the latest is a weighted average of the number of individual interfaces of the assembly predicted at acceptable, medium, or high quality (75).

This interface-centric evaluation reflects the essential role of the binding interface in defining the 3D structure of the complex. Even models that reproduce the structure of the native interface to lower accuracy provide useful information that can be further exploited. For example, a sizable fraction of acceptable-quality and borderline incorrect models were shown to correctly predict the residues that contribute to the binding interface (75), information that can be exploited in mutagenesis experiments. In contrast, incorrect models that entirely miss the binding interface are generally of little utility. The performance of individual groups is therefore ranked based only on the best model of acceptable, medium, or high quality that each group produces for a given target. This results in a coarse-grained ranking, which has its advantages (80) but may also be deemed cumbersome in comparison to continuous model quality metrics, which can be used to evaluate performance at different graininess levels, are more amenable to statistical analyses, and may serve as a target function to train machine learning models. The continuous DockQ metric, which combines the three CAPRI model quality parameters (F_{nat}, *i_rms*, and *L_rms*) into a single score with values ranging from 0–1 (7; see also **Figure 2*a***), is an attractive alternative that CAPRI and other studies (13) have already been using to evaluate prediction results for protein–protein complexes. Similar continuous metrics will need to be independently parameterized for complexes with smaller binding partners, such as short peptides, which involve fewer interactions that need to be modeled with high accuracy (82). Likewise, still missing are continuous evaluation metrics that seamlessly integrate quality measures of the interface region (such as DockQ) with those of the remainder of the protein structure.

Evaluating Progress

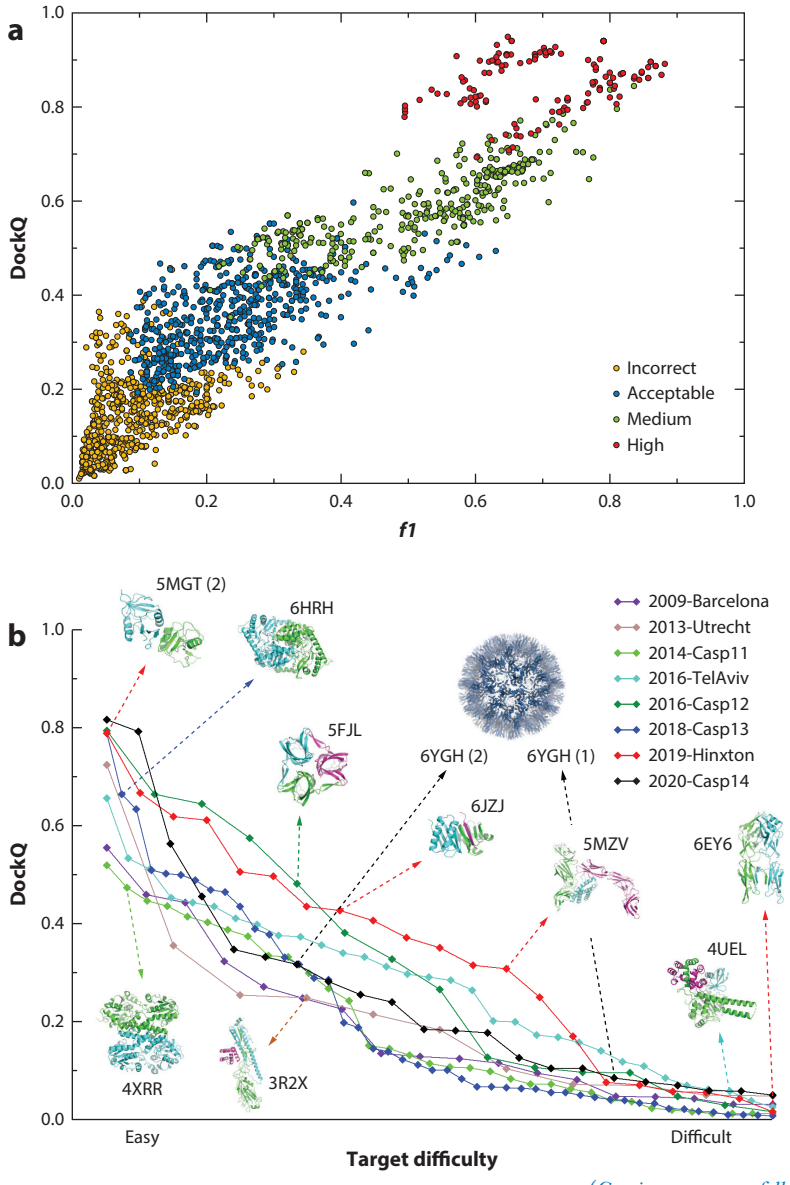
CAPRI has been evaluating progress in intervals of approximately three years, with each evaluation performed on results for 10–30 targets achieved during the prediction rounds of the intervening period. In addition, the CAPRI team, which has been evaluating CAPRI prediction results since its inception, independently evaluated the results obtained by participants in the assembly prediction challenges of the CASP11–CASP14 prediction seasons. All of the evaluations were performed on results achieved by human predictor groups (approximately 40 on average), by automatic servers (increasing from approximately 3 to as many as 12 over the years), and by participants in the scoring challenges (15–20 groups on average). Detailed evaluations of these results and assessments of the progress achieved by the community have been reported in the associated publications, amply cited in this review. In this section, we provide an overview of the main trends.

Figure 2*b* plots the quality of predicted binding modes measured by the average DockQ score across human predictor groups as a function of the level of modeling difficulty of the corresponding targets. These binding modes were evaluated in the successive dated periods since 2009, setting the chronological order of the results. The plots clearly illustrate the substantial variability in target difficulty levels during individual evaluation periods that persists over time, highlighting the difficulty of evaluating differences in performance across time periods and challenges. Nonetheless, one observes that model quality improves with time for easy- and medium-difficulty targets but remains low for difficult targets. Examples of targets in different modeling categories and the typical characteristics of these targets, detailed in **Figure 2*b***, clearly indicate that the modeling challenge differs substantially depending on the system at hand. Some large multicomponent assemblies solved to high resolution by cryo-EM (not shown) are particularly challenging to model when they combine several characteristics of difficult-to-model complexes.

An important contribution of the CAPRI community has been the development of automatic servers, the performance of which has steadily improved and diversified to the point of often

rivaling that of human predictors (76). Several of the best-performing servers, such as ClusPro (64), GalaxyPPDock (74), MDockPP (47), integrate docking procedures with template-based modeling and offer a panoply of handy tools for various modeling tasks, thereby gaining popularity with the wider scientific community.

Another area where CAPRI has helped break new ground is the prediction of protein–peptide complexes. This is an important category of complexes for which interest is rapidly growing, given the important role that recognition of short peptide motifs by protein domains plays in many regulatory processes (133). Recent methods, also implemented in several automatic servers, rose to the task of mastering the problem of modeling this challenging category of transient complexes, often



(*Caption appears on following page*)

Figure 2 (*Figure appears on preceding page*)

Accuracy levels of the best models of protein–protein complexes in CAPRI and CASP–CAPRI prediction rounds and the relation of the DockQ score to the CAPRI model quality categories. (*a*) Scatter plots of DockQ values for models submitted by predictors for individual targets evaluated in Reference 79 (vertical axis) as a function of $f1$ [$f1 = 2F_{nat}\ (1 - F_{non\_nat})/(F_{nat} + (1 - F_{non\_nat})$], where F_{nat} is the fraction of native contacts recalled in the model, and $(1 - F_{non\_nat})$ is the fraction of predicted contacts that are native (80). Individual points are color-coded according to the CAPRI model quality category: incorrect (*yellow*), acceptable (*blue*), medium (*green*), and high (*red*), illustrating that DockQ essentially reproduces the CAPRI model quality categories. (*b*) DockQ values for the best models as a function of target difficulty. Individual color-coded plots refer to best models evaluated in individual CAPRI assessment periods and CASP–CAPRI prediction rounds between 2009 and 2020, as indicated in the legend. Examples of targets of different difficulty levels are shown together with their Protein Data Bank (PDB) codes. Arrows indicate the rounds in which they were offered; numbers shown in parentheses following the PDB codes refer to the models of individual interfaces of the targets in question. Target difficulty is based on sequence and structure similarity to other proteins with known experimental structures.

to medium and high accuracy, despite their small binding interfaces and the significant degree of flexibility of the bound peptide ligands (79).

For the other modeling problems that CAPRI occasionally confronts, such as the prediction of protein–nucleic acid complexes (82), interface side-chain conformations, positions of interfacial water molecules (78), and protein oligosaccharide complexes (79), encouraging results were obtained. However, the number of targets for these complexes was too low to draw any conclusions.

Considering the state of protein science before these blind prediction experiments, it is difficult to imagine that the current level of prediction technology could have been reached without CASP leading the way. CAPRI followed suit a few years later, and both experiments created a higher level of transparency not only in protein structure prediction but also in computational biology in general by requiring the source code for most publications. They also built competitive yet collaborative communities, promoting the exchange of ideas and thus speeding up method development.

THE BREAKTHROUGH OF DEEP LEARNING–BASED PREDICTION METHODS: THE CURRENT STATE OF PLAY

The past few years have witnessed a breakthrough in modeling the 3D structure of proteins. This breakthrough can be attributed to two primary factors. The first is the extraordinary growth in protein sequence databases (131) coupled with a less prolific, yet notable, growth in the database of experimentally determined structures (136); both types of databases are freely available in public depositories. The second is the progressive introduction of cutting-edge methods in deep learning to a maturing protein modeling field (5, 126, 129). A key role was also played by the community-wide initiatives that enabled the critical evaluation of the recent breakthrough methods for predicting the structure of single protein chains, recorded in the CASP13 and CASP14 prediction seasons (16). Without these three components, the extraordinary achievement of the company Google DeepMind (122) would not have been possible. In this section, we examine the impact of this remarkable achievement on charting the structural landscape of native proteins and their complexes using computations and experiments.

Predicting the Structure of Single Chains

For the reasons discussed above, deep learning, a subfield of machine learning that utilizes multilayered artificial neural networks to extract patterns within large data sets without the need to explicitly define features of the data (73, 117), was uniquely primed to make substantial

contributions to the protein structure prediction field. Taking advantage of this, the DeepMind scientists designed and employed their prediction engine, AlphaFold2 (56), in the 2020 CASP season (55). This engine was trained on approximately 170,000 experimentally determined protein structures in the Protein Data Bank (PDB) (8) and a massive data set on multiply aligned protein sequences of related proteins, many of unknown structure (131), to produce models that rival in accuracy experimentally determined protein structures, surpassing the results of their first program, AlphFold1 (119), which was shown to perform well 2 years earlier (118). The power of AlphaFold2 lies in its novel multicomponent architecture that jointly embeds features from multiple sequence alignments and a residue pair representation, encoding spatial relationships between residues, and integrates graph-based components with attention learning (60). The pipeline also involves iterative refinements of predicted residue interactions based on their predicted interactions with other residues, enabling it to encapsulate structural features to a higher level of accuracy in a fully differentiable end-to-end deep learning method (56).

To enable the community to benefit from their monumental advance, DeepMind made their source code for the trained model of AlphaFold2 freely available to anyone wishing to make new predictions (**https://github.com/deepmind/alphafold**). Following suit, the group of David Baker released their deep learning protocols for protein prediction and design, RoseTTAFold, an algorithm exploring similar ideas to those of AlphaFold (6). In a further move to accelerate scientific research, DeepMind recently partnered with the EBI to create AlphaFold-DB (135), which provides access to predicted structures of single protein chains for the human proteome and other key organisms, as well as to the majority of the manually curated Uniport entries (SwissProt), further extending the coverage to over 200 million catalogued proteins.

This new treasure trove of structural data and the associated software tools have had a watershed effect on the field of computational and experimental structural biology, generating a flurry of studies (16). Examples include the optimization of multiple sequence alignments fed into AlphaFold2 (12) and a community-wide study evaluating various aspects of the structural information that AlphaFold produced and the applications that it enables (1). Evaluation of these applications has been greatly aided by two confidence measures that AlphaFold2 assigns to its predicted structures: a per-residue measure of confidence assigned to the local backbone structure and another measuring the confidence associated with residue-pairwise distances (56). The first is usually high for structured domains but low for linker regions, which may be flexible, intrinsically disordered, or structured only in the context of a larger complex. The second is useful for assessing more global features, such as domain packing.

Results obtained from these first analyses suggest that AlphaFold2 can be used to substantially extend the structural information for model proteomes beyond what is enabled by homology modeling, provided that its confidence metrics are critically interpreted. For example, while the atomic coordinates of regions modeled with low confidence may not be trustworthy, they can nevertheless be used to predict disordered regions more accurately than state-of-the-art methods (1). Similar analyses currently underway to characterize the astronomical number of structures in AlphaFold-DB should shed further light on the information that may be safely extracted from these predicted structures. Clearly missing, however, is information on the dynamic properties of proteins, many of which adopt multiple conformational states that are essential for their function (i.e., binding other proteins, nucleic acids, and small molecule ligands or switching between functionally active and inactive states) (95, 143). This is currently a serious limitation of deep learning approaches such as AlphaFold2, as has recently been discussed (33). Tackling this limitation is an important goal that is receiving increased attention, as further noted below.

Notwithstanding these limitations, ready access to the predicted protein structures in AlphaFold-DB and to the freely available code and resources, such as ColabFold (90), that can be

used to predict new structures with AlphaFold2 or RoseTTAFold efficiently with more modest computational resources is having a resounding impact on experimental determination of protein structures. Rather than making experimental structural biology obsolete, it is offering new opportunities like never before. Combining these opportunities with the recent spectacular advances in cryo-EM techniques is propelling the field to new levels. In several instances, hard-to-solve X-ray or cryo-EM structures have been elucidated by using AlphaFold models in molecular replacement protocols (66, 89). AlphaFold and RoseTTAFold models have been used successfully to fit residual electron density of cryo-EM maps, most notably in a recent assembly of the human nuclear pore complex (93). This is clearly an area that should soon see major advances from closer integration of deep learning–based and other structure modeling approaches with emerging deep learning–based and experimental cryo-EM techniques (130, 150).

Prediction of Protein Complexes and Assemblies

An obvious next frontier for deep learning–based protein structure prediction methods is the accurate prediction of complexes and larger protein assemblies; a flurry of recent studies have reported forays toward this goal. Several benchmarking studies suggest that extensions of deep learning–based methods to the prediction of protein complexes will provide a major advance over traditional docking methods. In these studies, AlphaFold2 was tricked into successfully modeling the structure of a set of protein complexes of known stoichiometry, albeit not consistently to high accuracy, by feeding it the concatenated sequences of the interacting component proteins. (13). Better performance was reported for AlphaFold-Multimer, the inference engine of AlphaFold, directly trained on protein complexes from the PDB (31). For the same benchmark data set of heteromeric interfaces, AlphaFold-Multimer produced acceptable predictions (DockQ $\geq$ 0.23) for approximately 67% of the interfaces but high-accuracy predictions (DockQ $\geq$ 0.80) for only 23%, an improvement of 25% and 11% respectively, over the modified AlphaFold2 version. More modest improvement was achieved for homomeric interfaces generally associated with higher binding affinity, for which larger fractions could be predicted to acceptable (69%) and high (34%) accuracy. At the same time, methods have been proposed to integrate AlphaFold2 predictions of complexes with classical docking calculations and use the predicted complexes as templates for AlphaFold2 to significantly improve performance over either method used independently (37).

Interestingly, in some instances, AlphaFold2 accurately predicts the bound conformation of individual subunits of protein assemblies in the absence of any information on stoichiometry. Two examples of such instances are illustrated in **Figure 3**. In one, the AlphaFold2-predicted structures of the L-ring and P-ring proteins of the flagella LP ring from *Salmonella* (**Figure 3*a***) reproduce quite accurately the experimentally determined structures of these proteins (**Figure 3*b***) when they are part of the assembled flagella ring (54) (PDB code 7BGL) (**Figure 3*c***). The second example concerns the TnpA transposase from *Bacillus thuringiensis*, a protein displaying a high degree of structural plasticity. The recent cryo-EM structures of this protein were determined to high resolution in the apo state and in complex with transposon (DNA), where TnpA forms dimers adopting significantly different conformations (121) (**Figure 3*e,f***). AlpfaFold2 predicts a single structure for the TnpA chain that accurately reproduces the structure of the TnpA protomer in the complex with transposon (**Figure 3*d***). In both examples, no structural templates could be identified, and although the predicted highly nonglobular protein conformations are unlikely to represent stable states, they should still help to model the assembly and provide useful information on the functional state of the protein.

Several studies have also suggested that AlphaFold and RoseTTAfold can be used to extend the structural coverage of model interactomes beyond what is enabled by homology modeling

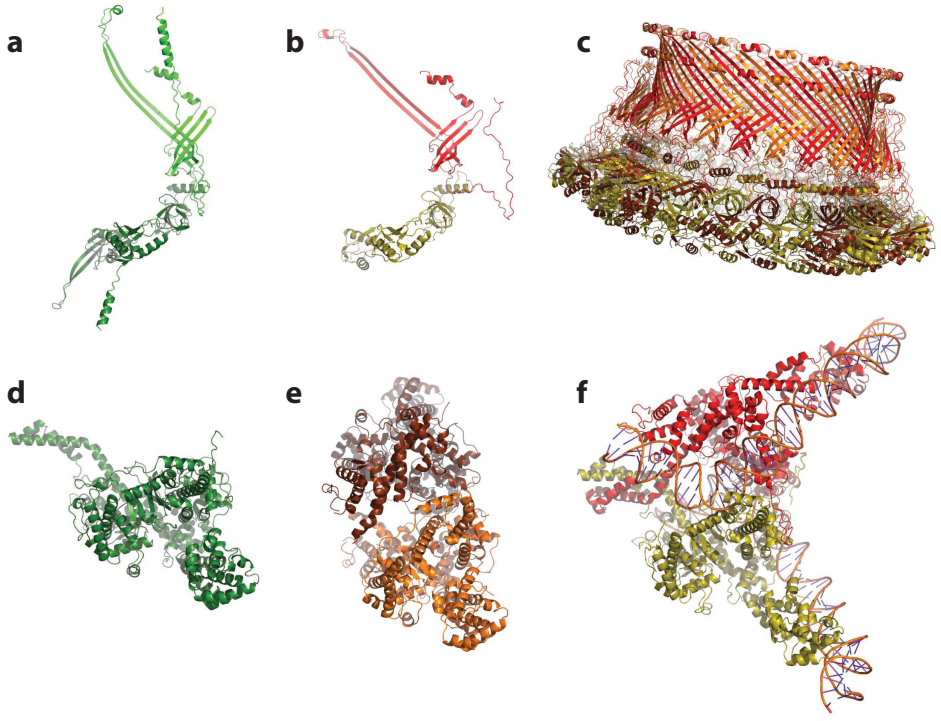


Figure 3

Examples of AlphaFold2 accurately predicting the bound conformation of individual subunits of protein assemblies. (*a*) The structures predicted by AlphaFold2 for the L-ring and P-ring proteins of the flagella LP ring from *Salmonella*. (*b*) The experimentally determined structures of these proteins when they are part of the assembled flagella ring. (*c*) Structure of the LP flagella ring (Protein Data Bank code 7BGL). (*d*) The structure of the TnpA protomer predicted by AlphaFold2, displaying a high degree of similarity (1.82 Å backbone rms) to the experimental structure of the TnpA protomer in the dimeric complex with transposon (DNA) (121), shown in panel *f*. (*e*) Structure of the apo TnpA dimer, adopting a globular structure and exhibiting significant conformational changes relative to the transposon complex (8.87 Å backbone rms of individual protomers). (*f*) Experimental structure of the TnpA dimer in complex with transposon (121).

(of complexes or of single chains followed by docking) (94). For example, for the human interactome, AlphaFold predicted approximately 1,400 high-confidence models of complexes displaying no homology to a known structure (15). Both AlphaFold and RoseTTAFold were used to identify interacting proteins and model their complexes in baker's yeast (49) and the human mitochondrion (107), deriving in each case new structural information for functionally important complexes.

While these early results are very promising, they also indicate that significant room remains for improvement. Further optimizing these methods to tackle complexes spanning a wide range of binding partners, binding affinities, and functional states has the potential to lead to significant breakthroughs for these prediction problems. However, fulfilling this potential will not be effortless. The structural coverage of the protein complexes that form in living cells—the body of data that AI methods need to learn from—is orders of magnitude smaller than the current structural coverage of single protein chains. Furthermore, the formation of many of the more transient complexes featuring lower binding affinities, such as those associated with signal transaction processes,

is highly context dependent. Currently, however, the ability of experimental structural biology to adequately sample the physiologically relevant contexts of complexes is limited. Third, modeling the dynamic properties of the component proteins, which govern the conformational changes associated with binding, will be an important problem to overcome.

WHAT NEXT?

Although this new wave of applied deep learning methods in the protein structure prediction field have rocked the biological sciences, with numerous potential applications (57), much work will be needed to mitigate important current limitations. Top of the list is the problem of accounting for dynamic flexibility within single chains, the association process, and the complex (108). This is important for understanding and modeling the functional states of proteins, including those of intrinsically disordered proteins, of which there is a natural abundance (69). Next, and related to the above, deep learning–based methods are still incapable of interpreting the effects of single point mutations; backbone movements are simply not replicated when one amino acid is substituted for another, as benchmarked in several recent studies (14, 103).

In this case, too, deep learning methodologies are offering a way forward by first allowing understanding of the conformational states that a protein samples (i.e., in known structures) and the likely transition paths between them (112). This understanding is then used to further extend the sampled conformational space by generating experimentally unobserved but native-like protein conformations, as recently described with an Autoencoder method (25). These descriptions of multiple conformations of a given protein may have to be integrated with the data on multiple sequence alignments to model structures corresponding to specific functional states. Crucial for training and testing this type of method will be the development of benchmark data sets of physiologically pertinent structures of single chain proteins and complexes that incorporate information on the sampled conformational landscape.

Addressing these challenges will impact all areas of protein structure predictions, including prediction of protein assemblies. In this case, progress will also depend on the ability of deep learning algorithms to restrain the sampling of the vast number of potential binding modes closer to the basin of native-like solutions, even in cases where coevolution signals derived from the multiple sequence alignments are weak (such as for complexes with antibodies or host pathogen interactions), a problem that AlphaFold seems to struggle with (37). Improving the ability to recognize native-like binding modes—the model ranking (and scoring) problem—will be of crucial importance. Promising results toward this goal have recently been reported by standalone deep learning models implementing convolutional neural networks and other methods (17, 113, 140). Ultimately, rankings of models of protein complexes need to display some correlation with binding affinities, a goal that scoring methods have pursued with modest success (34, 134).

As end-to-end machine learning methodologies are improved and mastered by the wider structural biology community, it will become routine to model a significant fraction of proteins and the complexes that they form just from their amino acid sequences, ultimately negating the need for intermediate steps, such as searching for and utilizing close structural templates. Likewise, one expects these new methodologies to be extended to modeling nucleic acids, particularly RNA but also DNA, and the complexes that they form with proteins within the cell. In this case, too, a major challenge will be to collect enough experimental data to train and validate machine learning methods. The way forward would be a closer integration of computational and experimental approaches. This would involve combining emerging methods for extracting information on structural heterogeneity in macromolecular complexes from the cryo-EM data obtained from endogenous material with AI-based structure prediction algorithms and molecular simulation techniques. All of these

are important developing areas where blind prediction initiatives should continue to play a major role.

POSTSCRIPT

As this review was being completed, the first results of the CASP15 blind prediction experiment, held in May through August 2022, became available, shedding further light on where the field is headed.

These results indicate a consolidation of the successful performance of the prediction of single protein chains heralded by AlphaFold2 from DeepMind in 2020. This consolidation was achieved by different groups (DeepMind did not participate in the 2022 experiment) employing various implementations and different parameters of AlphaFold2, as well as bolstering and tweaking the contents of the multiple sequence alignments. Most notable, however, is the substantial progress recorded in modeling protein–protein interactions and complexes, evaluated in collaboration with CAPRI. The primary driver of this success has been a similar creative employment of AlphaFold2 and AlphaFold-multimer inference engines, often modifying them to sample a much larger number of models (137a) and/or using multiple sequence alignments augmented with sequences from various sources. The performance of both approaches seems to greatly benefit from the use of AlphaFold confidence metrics to score and rank sampled models, a practice backed by recent findings that these metrics closely estimate the true quality of the candidate structures, outperforming other state-of-the-art model accuracy estimates (116).

Nevertheless, while high-quality models (DockQ > 75) were produced for nearly half of the target complexes (40%, a notable increase over the 8% achieved in CASP14), average model accuracy is still below that for single chain proteins, and even the current level of success is not uniform across all interaction classes. Antigen–antibody complexes are poorly predicted by AlphaFold-based tools, since the lack of evolutionary relationships between the binding partners precludes the construction of deep multiple sequence alignment capable of guiding prediction. An interesting observation is that there are signs that these more protracted problems may soon be solvable, as several reports have emerged on artificial intelligence–based transformer protein language models (87, 144) that enable end-to-end atomic-resolution structure prediction directly from a single sequence, thereby negating the need for multiple sequence alignments at the inference stage. Such approaches, if shown to withstand scrutiny in terms of the accuracy of the predicted structures and their general applicability, should enhance our ability to model the structures of proteins (and protein complexes) of the immune system, orphan proteins, and de novo designed proteins.

In summary, these are undoubtedly exciting times, as artificial intelligence–based methods continue to push back the barriers of protein structure prediction. Nevertheless, a cautionary note must be sounded concerning the model accuracy levels expected from these methods in the immediate future. A comprehensive study comparing AlphaFold models with experimentally determined electron density maps of recent crystal structures (128) indicates that, while some AlphaFold models match experimental electron density maps closely, most do not. Differences involve whole domain orientations, as well as local backbone and sidechain conformations, and occur even in parts of AlphaFold models that are predicted with high confidence. These findings challenge the confident and routine use of predicted structures by artificial intelligence–based methods for the precision engineering of rationally designed drugs to modulate protein function, an issue deserving further scrutiny.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

S.V. and D.K. acknowledge support from grants R35GM118078, R01 GM140098, and RM1135136 from the National Institute of General Medical Sciences and grant DMS 2054251 from the National Science Foundation. P.A.B. acknowledges support from the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001003), the UK Medical Research Council (FC001003), and the Wellcome Trust (FC001003). We thank Rouslan Efremov, author of the cryo-EM structures of TnpA, for disclosing these structures prior to publication and offering the TnpA-transposon complex as a target to CAPRI.

LITERATURE CITED

1. Akdel M, Pires DEV, Porta Pardo E, Jänes J, Zalevsky AO, et al. 2021. A structural biology community assessment of AlphaFold 2 applications. bioRxiv 2021.09.26.461876. **https://doi.org/10.1101/2021.09.26.461876**
2. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, et al. 2017. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* 13:3031–48
3. Aloy P, Ceulemans H, Stark A, Russell RB. 2003. The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.* 332:989–98
4. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. 2008. Principles of flexible protein-protein docking. *Proteins* 73:271–89
5. Baek M, Baker D. 2022. Deep learning and protein structure modeling. *Nat. Methods* 19:13–14
6. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, et al. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373:871–76
7. Basu S, Wallner B. 2016. DockQ: a quality measure for protein-protein docking models. *PLOS ONE* 11:e0161879
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–42
9. Blom NS, Sygusch J. 1997. High resolution fast quantitative docking using Fourier domain correlation techniques. *Proteins* 27:493–506
10. Bonvin AM. 2006. Flexible protein-protein docking. *Curr. Opin. Struct. Biol.* 16:194–200
11. Brooks BR, Brooks CL III, Mackerell AD Jr., Nilsson L, Petrella RJ, et al. 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30:1545–614
12. Bryant AS, Goddard CA, Huguenard JR, Knudsen EI. 2015. Cholinergic control of gamma power in the midbrain spatial attention network. *J. Neurosci.* 35:761–75
13. Bryant P, Pozzati G, Elofsson A. 2022. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* 13:1265
14. Buel GR, Walters KJ. 2022. Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* 29:1–2
15. Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, et al. 2021. Towards a structurally resolved human protein interaction network. bioRxiv 2021.11.08.467664. **https://doi.org/10.1101/2021.11.08.467664**
16. Callaway E. 2020. "It will change everything": DeepMind's AI makes gigantic leap in solving protein structures. *Nature* 588:203–4
17. Cao Y, Shen Y. 2020. Energy-based graph convolutional networks for scoring protein docking models. *Proteins* 88:1091–99
18. Chandonia JM, Brenner SE. 2006. The impact of structural genomics: expectations and outcomes. *Science* 311:347–51
19. Chen R, Li L, Weng Z. 2003. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 52:80–87
20. Chothia C. 1974. Hydrophobic bonding and accessible surface area in proteins. *Nature* 248:338–39
21. Chothia C, Janin J. 1975. Principles of protein-protein recognition. *Nature* 256:705–8
22. Chothia C, Lesk AM. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5:823–26

23. Dapkunas J, Timinskas A, Olechnovic K, Margelevicius M, Diciunas R, Venclovas C. 2017. The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinformatics* 33:935–37
24. de Vries SJ, van Dijk M, Bonvin AM. 2010. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 5:883–97
25. Degiacomi MT. 2019. Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure* 27:1034–40.e3
26. Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. 2020. Performance and its limits in rigid body protein-protein docking. *Structure* 28:1071–81.e3
27. Dobbins SE, Lesk VI, Sternberg MJ. 2008. Insights into protein flexibility: the relationship between normal modes and conformational change upon protein-protein docking. *PNAS* 105:10390–95
28. Dominguez C, Boelens R, Bonvin AM. 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125:1731–37
29. Dunbrack RL Jr., Gerloff DL, Bower M, Chen X, Lichtarge O, Cohen FE. 1997. Meeting review: the Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2), Asilomar, California, December 13–16, 1996. *Fold Des.* 2:R27–42
30. Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature* 319:199–203
31. Evans R, O'Neill M, Pritzel A, Antropova N, Senior AW, et al. 2021. Protein complex prediction with AlphaFold-Multimer. bioRxiv 2021.10.04.463034. **https://doi.org/10.1101/2021.10.04.463034**
32. Fiser A, Sali A. 2003. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* 19:2500–1
33. Fleishman SJ, Horovitz A. 2021. Extending the new generation of structure predictors to account for dynamics and allostery. *J. Mol. Biol.* 433:167007
34. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, et al. 2011. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J. Mol. Biol.* 414:289–302
35. Gabb HA, Jackson RM, Sternberg MJ. 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 272:106–20
36. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, et al. 2020. Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr. Protoc. Bioinform.* 72:e108
37. Ghani U, Desta I, Jindal A, Khan O, Jones G, et al. 2021. Improved docking of protein models by a combination of Alphafold2 and ClusPro. bioRxiv 2021.09.07.459290. **https://doi.org/10.1101/2021.09.07.459290**
38. Gilson MK, Sharp KA, Honig BH. 1988. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.* 9:337–35
39. Gobel U, Sander C, Schneider R, Valencia A. 1994. Correlated mutations and residue contacts in proteins. *Proteins* 18:309–17
40. Guest JD, Vreven T, Zhou J, Moal I, Jeliazkov JR, et al. 2021. An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. *Structure* 29:606–21.e5
41. Haas J, Roth S, Arnold K, Kiefer F, Schmidt T, et al. 2013. The Protein Model Portal—a comprehensive resource for protein structure and model information. *Database* 2013:bat031
42. Hagler AT, Huler E, Lifson S. 1974. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* 96:5319–27
43. Hagler AT, Lifson S. 1974. Energy functions for peptides and proteins. II. The amide hydrogen bond and calculation of amide crystal properties. *J. Am. Chem. Soc.* 96:5327–35
44. Harmalkar A, Gray JJ. 2021. Advances to tackle backbone flexibility in protein docking. *Curr. Opin. Struct. Biol.* 67:178–86
45. Ho BK, Dill KA. 2006. Folding very short peptides using molecular dynamics. *PLOS Comput. Biol.* 2:e27
46. Huang SY, Zou X. 2008. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* 72:557–79
47. Huang SY, Zou X. 2010. MDockPP: a hierarchical approach for protein-protein docking and its application to CAPRI rounds 15–19. *Proteins* 78:3096–103

48. Huang SY, Zou X. 2011. Statistical mechanics-based method to extract atomic distance-dependent potentials from protein structures. *Proteins* 79:2648–61
49. Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, et al. 2021. Computed structures of core eukaryotic protein complexes. *Science* 374:eabm4805
50. Isogai Y, Nemethy G, Scheraga HA. 1977. Enkephalin: conformational analysis by means of empirical energy calculations. *PNAS* 74:414–18
51. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, et al. 2003. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* 52:2–9
52. Jernigan RL, Bahar I. 1996. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.* 6:195–209
53. Deleted in proof.
54. Johnson S, Furlong EJ, Deme JC, Nord AL, Caesar JJE, et al. 2021. Molecular structure of the intact bacterial flagellar basal body. *Nat. Microbiol.* 6:712–21
55. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Applying and improving AlphaFold at CASP14. *Proteins* 89:1711–21
56. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–89
57. Jumper J, Hassabis D. 2022. Protein structure predictions to atomic accuracy with AlphaFold. *Nat. Methods* 19:11–12
58. Kamisetty H, Ovchinnikov S, Baker D. 2013. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *PNAS* 110:15674–79
59. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *PNAS* 89:2195–99
60. Knyazev B, Taylor GW, Amer M. 2019. Understanding attention and generalization in graph neural networks. In *NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pp. 4202–12. New York: ACM
61. Ko J, Park H, Heo L, Seok C. 2012. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* 40:W294–97
62. Kozakov D, Brenke R, Comeau SR, Vajda S. 2006. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65:392–406
63. Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, et al. 2010. Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. *Proteins* 78:3124–30
64. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, et al. 2017. The ClusPro web server for protein-protein docking. *Nat. Protoc.* 12:255–78
65. Kryshtafovych A, Monastyrskyy B, Fidelis K. 2016. CASP11 statistics and the prediction center evaluation system. *Proteins* 84(Suppl. 1):15–19
66. Kryshtafovych A, Moult J, Albrecht R, Chang GA, Chao K, et al. 2021. Computational models in the service of X-ray and cryo-electron microscopy structure determination. *Proteins* 89:1633–46
67. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. 2021. Critical assessment of methods of protein structure prediction (CASP): Round XIV. *Proteins* 89:1607–17
68. Kuhlman B, Bradley P. 2019. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* 20:681–97
69. Kulkarni P, Uversky VN. 2018. Intrinsically disordered proteins: the dark horse of the dark proteome. *Proteomics* 18:e1800061
70. Kundrotas PJ, Anishchenko I, Dauzhenka T, Kotthoff I, Mnevets D, et al. 2018. Dockground: a comprehensive data resource for modeling of protein complexes. *Protein Sci.* 27:172–81
71. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. 2012. Templates are available to model nearly all complexes of structurally characterized proteins. *PNAS* 109:9438–41
72. Kuroda D, Gray JJ. 2016. Pushing the backbone in protein-protein docking. *Structure* 24:1821–29
73. LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature* 521:436–44

74. Lee H, Seok C. 2017. Template-based prediction of protein-peptide interactions by using GalaxyPepDock. *Methods Mol. Biol.* 1561:37–47
75. Lensink MF, Brysbaert G, Mauri T, Nadzirin N, Velankar S, et al. 2021. Prediction of protein assemblies, the next frontier: the CASP14-CAPRI experiment. *Proteins* 89:1800–23
76. Lensink MF, Brysbaert G, Nadzirin N, Velankar S, Chaleil RAG, et al. 2019. Blind prediction of homo- and hetero-protein complexes: the CASP13-CAPRI experiment. *Proteins* 87:1200–21
77. Lensink MF, Mendez R, Wodak SJ. 2007. Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins* 69:704–18
78. Lensink MF, Moal IH, Bates PA, Kastritis PL, Melquiond AS, et al. 2014. Blind prediction of interfacial water positions in CAPRI. *Proteins* 82:620–32
79. Lensink MF, Nadzirin N, Velankar S, Wodak SJ. 2020. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins* 88:916–38
80. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ. 2018. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins* 86(Suppl. 1):257–73
81. Lensink MF, Velankar S, Kryshtafovych A, Huang SY, Schneidman-Duhovny D, et al. 2016. Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment. *Proteins* 84(Suppl. 1):323–48
82. Lensink MF, Velankar S, Wodak SJ. 2017. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* 85:359–77
83. Lensink MF, Wodak SJ. 2013. Docking, scoring, and affinity prediction in CAPRI. *Proteins* 81:2082–95
84. Lensink MF, Wodak SJ. 2014. Score_set: a CAPRI benchmark for scoring protein complexes. *Proteins* 82:3163–69
85. Levinthal C, Wodak SJ, Kahn P, Dadivanian AK. 1975. Hemoglobin interaction in sickle cell fibers. I. Theoretical approaches to the molecular contacts. *PNAS* 72:1330–34
86. Levitt M. 1976. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* 104:59–107
87. Lin Z, Akin H, Rao R, Hie B, Zhu Z, et al. 2022. Evolutionary-scale prediction of atomic level protein structure with a language model. bioRxiv 2022.07.20.500902. **https://doi.org/10.1101/2022.07.20.500902**
88. Liu J, Wu T, Guo Z, Hou J, Cheng J. 2022. Improving protein tertiary structure prediction by deep learning and distance prediction in CASP14. *Proteins* 90:58–72
89. McCoy AJ, Sammito MD, Read RJ. 2022. Implications of AlphaFold2 for crystallographic phasing by molecular replacement. *Acta Crystallogr. D* 78:1–13
90. Mirdita M, Ovchinnikov S, Steinegger M. 2021. ColabFold: making protein folding accessible to all. *Nat. Methods* 19:679–82
91. Moal IH, Chaleil RAG, Bates PA. 2018. Flexible protein-protein docking with SwarmDock. *Methods Mol. Biol.* 1764:413–28
92. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, et al. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *PNAS* 108:E1293–301
93. Mosalaganti S, Obarska-Kosinska A, Siggel M, Taniguchi R, Turonova B, et al. 2022. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* 376:eabm9506
94. Mosca R, Ceol A, Aloy P. 2013. Interactome3D: adding structural details to protein networks. *Nat. Methods* 10:47–53
95. Motlagh HN, Wrabl JO, Li J, Hilser VJ. 2014. The ensemble nature of allostery. *Nature* 508:331–39
96. Moult J. 2005. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* 15:285–89
97. Moult J, Pedersen JT, Judson R, Fidelis K. 1995. A large-scale experiment to assess protein structure prediction methods. *Proteins* 23:ii–v
98. Nadaradjane AA, Guerois R, Andreani J. 2018. Protein-protein docking using evolutionary information. *Methods Mol. Biol.* 1764:429–47
99. Novotny J, Bruccoleri R, Karplus M. 1984. An analysis of incorrectly folded protein models: implications for structure predictions. *J. Mol. Biol.* 177:787–818

100. Olechnovic K, Venclovas C. 2017. VoroMQA: assessment of protein structure quality using interatomic contact areas. *Proteins* 85:1131–45
101. Onufriev A, Case DA, Bashford D. 2002. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* 23:1297–304
102. Padhorny D, Kazennov A, Zerbe BS, Porter KA, Xia B, et al. 2016. Protein-protein docking by fast generalized Fourier transforms on 5D rotational manifolds. *PNAS* 113:E4286–93
103. Pak MA, Markhieva KA, Novikova MS, Petrov DS, Vorobyev IS, et al. 2021. Using AlphaFold to predict the impact of single mutations on protein stability and function. bioRxiv 2021.09.19.460937. **https://doi.org/10.1101/2021.09.19.460937**
104. Pallara C, Jimenez-Garcia B, Romero M, Moal IH, Fernandez-Recio J. 2017. pyDock scoring for the new modeling challenges in docking: protein-peptide, homo-multimers, and domain-domain interactions. *Proteins* 85:487–96
105. Pearce R, Zhang Y. 2021. Deep learning techniques have significantly impacted protein structure prediction and protein design. *Curr. Opin. Struct. Biol.* 68:194–207
106. Pearce R, Zhang Y. 2021. Toward the solution of the protein structure prediction problem. *J. Biol. Chem.* 297:100870
107. Pei J, Zhang J, Cong Q. 2021. Human mitochondrial protein complexes revealed by large-scale coevolution analysis and deep learning-based structure modeling. *Bioinformatics* 38:4301–11
108. Perrakis A, Sixma TK. 2021. AI revolutions in biology: the joys and perils of AlphaFold. *EMBO Rep.* 22:e54046
109. Porter KA, Desta I, Kozakov D, Vajda S. 2019. What method to use for protein-protein docking? *Curr. Opin. Struct. Biol.* 55:1–7
110. Porter KA, Padhorny D, Desta I, Ignatov M, Beglov D, et al. 2019. Template-based modeling by ClusPro in CASP13 and the potential for using co-evolutionary information in docking. *Proteins* 87:1241–48
111. Quignot C, Rey J, Yu J, Tuffery P, Guerois R, Andreani J. 2018. InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. *Nucleic Acids Res.* 46:W408–16
112. Ramaswamy VK, Musson SC, Willcocks CG, Degiacomi MT. 2021. Deep learning protein conformational space with convolutions and latent interpolations. *Phys. Rev. X* 11:011052
113. Renaud N, Geng C, Georgievska S, Ambrosetti F, Ridder L, et al. 2021. DeepRank: a deep learning framework for data mining 3D protein-protein interfaces. *Nat. Commun.* 12:7068
114. Ritchie DW, Kemp GJ. 2000. Protein docking using spherical polar Fourier correlations. *Proteins* 39:178–94
115. Robin X, Haas J, Gumienny R, Smolinski A, Tauriello G, Schwede T. 2021. Continuous Automated Model EvaluatiOn (CAMEO): perspectives on the future of fully automated evaluation of structure prediction methods. *Proteins* 89:1977–86
116. Roney JP, Ovchinnikov S. 2022. State-of-the-art estimation of protein model accuracy using AlphaFold. bioRxiv 2022.03.11.484043. **https://doi.org/10.1101/2022.03.11.484043**
117. Schmidhuber J. 2015. Deep learning in neural networks: an overview. *Neural Netw.* 61:85–117
118. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, et al. 2019. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins* 87:1141–48
119. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, et al. 2020. Improved protein structure prediction using potentials from deep learning. *Nature* 577:706–10
120. Shen MY, Sali A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15:2507–24
121. Shkumatov AV, Aryanpour N, Oger CA, Goossens G, Hallet BF, Efremov RG. 2022. Metamorphism of catalytic domain controls transposition in Tn3 family transposases. bioRxiv 2022.02.23.481423. **https://doi.org/10.1101/2022.02.23.481423**
122. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529:484–89

123. Simons KT, Kooperberg C, Huang E, Baker D. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* 268:209–25
124. Singh A, Dauzhenka T, Kundrotas PJ, Sternberg MJE, Vakser IA. 2020. Application of docking methodologies to modeled proteins. *Proteins* 88:1180–88
125. Smith GR, Sternberg MJ, Bates PA. 2005. The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *J. Mol. Biol.* 347:1077–101
126. Suh D, Lee JW, Choi S, Lee Y. 2021. Recent applications of deep learning methods on evolution- and contact-based protein structure prediction. *Int. J. Mol. Sci.* 22:6032
127. Tanaka S, Scheraga HA. 1976. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* 9:945–50
128. Terwilliger TC, Liebschner D, Croll TI, Williams CJ, McCoy AJ, et al. 2022. AlphaFold predictions: great hypotheses but no match for experiment. bioRxiv 2022.11.21.517405. **https://doi.org/10.1101/2022.11.21.517405**
129. Torrisi M, Pollastri G, Le Q. 2020. Deep learning methods in protein structure prediction. *Comput. Struct. Biotechnol. J.* 18:1301–10
130. Tuting C, Kyrilis FL, Muller J, Sorokina M, Skalidis I, et al. 2021. Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. *Nat. Commun.* 12:6933
131. UniProt Consort. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49:D480–89
132. Vakser IA, Aflalo C. 1994. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins* 20:320–29
133. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, et al. 2014. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* 114:6733–78
134. Vangone A, Bonvin AM. 2015. Contacts-based prediction of binding affinity in protein-protein complexes. *eLife* 4:e07454
135. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, et al. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50:D439–44
136. Velankar S, Burley SK, Kurisu G, Hoch JC, Markley JL. 2021. The Protein Data Bank Archive. *Methods Mol. Biol.* 2305:3–21
137. Vreven T, Schweppe DK, Chavez JD, Weisbrod CR, Shibata S, et al. 2018. Integrating cross-linking experiments with ab initio protein-protein docking. *J. Mol. Biol.* 430:1814–28
137a. Wallner B. 2022. AFsample: improving multimer prediction with AlphaFold using aggressive sampling. bioRxiv 2022.12.20.521205. **https://doi.org/10.1101/2022.12.20.521205**
138. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004. Development and testing of a general amber force field. *J. Comput. Chem.* 25:1157–74
139. Wang S, Sun S, Li Z, Zhang R, Xu J. 2017. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLOS Comput. Biol.* 13:e1005324
140. Wang X, Flannery ST, Kihara D. 2021. Protein docking model evaluation by graph neural networks. *Front. Mol. Biosci.* 8:647915
141. Wodak SJ, Janin J. 2002. Structural basis of macromolecular recognition. *Adv. Protein Chem.* 61:9–73
142. Wodak SJ, Mendez R. 2004. Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* 14:242–49
143. Wodak SJ, Paci E, Dokholyan NV, Berezovsky IN, Horovitz A, et al. 2019. Allostery in its many disguises: from theory to applications. *Structure* 27:566–78
144. Wu R, Ding F, Wang R, Shen R, Zhang X, et al. 2022. High-resolution *de novo* structure prediction from primary sequence. bioRxiv 2022.07.21.500909. **https://doi.org/10.1101/2022.07.21.500999**
145. Xia B, Vajda S, Kozakov D. 2016. Accounting for pairwise distance restraints in FFT-based protein-protein docking. *Bioinformatics* 32:3342–44
146. Yan Y, Wen Z, Wang X, Huang SY. 2017. Addressing recent docking challenges: a hybrid strategy to integrate template-based and free protein-protein docking. *Proteins* 85:497–512

147. Yan Y, Zhang D, Zhou P, Li B, Huang SY. 2017. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 45:W365–73
148. Zhang Z, Schindler CE, Lange OF, Zacharias M. 2015. Application of enhanced sampling Monte Carlo methods for high-resolution protein-protein docking in Rosetta. *PLOS ONE* 10:e0125941
149. Zheng W, Li Y, Zhang C, Zhou X, Pearce R, et al. 2021. Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14. *Proteins* 89:1734–51
150. Zhong ED, Bepler T, Berger B, Davis JH. 2021. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* 18:176–85
151. Zhou H, Skolnick J. 2011. GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys. J.* 101:2043–52
152. Zhou H, Zhou Y. 2002. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* 11:2714–26
153. Zhou R. 2003. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* 53:148–61

Annual Review of Biophysics
Volume 52, 2023

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Biophysics* articles may be found at http://www.annualreviews.org/errata/biophys