



Journal homepage: <https://trpubonline.org/index.php/Biosynth>

SYSTEMATIC REVIEW

Deep Learning-Based Prediction of Protein Structure from Sequence Data

Petr Dubois^{1*}, Claire Havlicek¹ and Astrid Fernandez²

¹Department of Biotechnology and Molecular Biology, Delta University Center for Bioinformatics, Netherlands

²Department of Biomedical and Health Sciences, Baltic Institute of Biomedical Research, Sweden

*Corresponding author: Petr Dubois, Department of Biotechnology and Molecular Biology, Delta University Center for Bioinformatics, Netherlands; Email: p.dubois@duc-b-nl.nl

ABSTRACT

Predicting a protein's three-dimensional structure from its amino acid sequence was, for five decades, the kind of problem that structural biologists cited as a benchmark of how far computational methods still had to travel. That situation changed abruptly between the thirteenth and fourteenth Critical Assessment of Structure Prediction (CASP) experiments, when deep-learning systems built around attention-based neural architectures began producing models whose backbone accuracy rivalled experimental structures for a large share of single-domain targets. This systematic review synthesises 68 primary research articles and method papers published between January 2018 and March 2024, identified through a structured search of five bibliographic databases, covering the architectural lineage from early contact-map convolutional networks through AlphaFold2, RoseTTAFold and their open derivatives, to protein-language-model-based single-sequence predictors such as ESMFold and OmegaFold. We extracted and compared reported accuracy on CASP14, CASP15 and continuous CAMEO benchmarks, computational cost, multiple-sequence-alignment dependence, and reported performance on three categories of hard target: intrinsically disordered regions, novel folds with no close structural homologue, and multi-chain protein complexes. Across the reviewed literature, mean backbone accuracy (GDT-TS) on single-domain CASP14/15 targets rose from approximately 40 in pre-2020 deep learning methods to above 90 for AlphaFold2-class predictors, while complex and disordered-region prediction remained substantially harder, with median accuracy gains over pre-deep-learning baselines of roughly half the magnitude seen for single-domain folding. We conclude that the field has largely solved single-chain, well-conserved-fold structure prediction, that language-model-based single-sequence methods offer a meaningful speed-accuracy trade-off for large-scale proteome annotation, and that complex assembly, disorder-to-order transitions and function prediction from structure remain the primary open problems guiding the next generation of methods.

Keywords: protein structure prediction; deep learning; AlphaFold; ESMFold; protein language models; CASP; structural bioinformatics

Received: July 7, 2024 | Accepted: October 2, 2024 | Published: December 21, 2024

© 2024 Dubois et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

The relationship between a protein's amino acid sequence and its folded three-dimensional structure has

been described as one of the central unsolved problems of twentieth-century molecular biology since Anfinsen's thermodynamic hypothesis first established, more than fifty years ago, that the native structure is in principle fully

determined by sequence alone [1]. Experimental structure determination by X-ray crystallography, NMR spectroscopy and, more recently, cryo-electron microscopy has produced an invaluable but comparatively small catalogue of solved structures relative to the number of sequenced proteins: at the time of writing the Protein Data Bank holds on the order of two hundred thousand experimentally determined structures against several hundred million sequences deposited in UniProt, a gap that computational prediction methods have long aimed to close [2,3].

Computational structure prediction progressed through several distinct methodological eras before the deep learning transition. Homology modelling, which builds a target structure by threading its sequence onto a solved template structure of a related protein, remains highly accurate when a close homologue exists but is of no use for sequences lacking one [4]. Physics-based *ab initio* folding, exemplified by Rosetta's fragment-assembly approach, attempted to predict structure from first principles or statistical potentials without a template, but accuracy for anything beyond small, simple folds remained limited for most of the 2000s and 2010s [5,6]. Co-evolution-based contact prediction, which infers spatially proximate residue pairs from correlated substitution patterns across large multiple sequence alignments, provided an important intermediate advance during the 2010s, improving *ab initio* modelling substantially for protein families with sufficiently deep alignments while remaining a weak signal for shallow or orphan families [7,8].

The Critical Assessment of Structure Prediction (CASP), a biennial blind prediction exercise running since 1994, has served as the field's principal external benchmark throughout this progression, and it is against CASP that the scale of the deep-learning transition is most clearly visible. Deep convolutional networks applied to residue-residue contact and distance prediction produced incremental but real gains at CASP12 and CASP13, and DeepMind's first AlphaFold system, which combined deep distance-prediction networks with gradient-based structure optimisation, achieved the best overall performance at CASP13 in 2018 [9,10]. The step change came at CASP14 in 2020, when AlphaFold2, built around an end-to-end differentiable architecture incorporating an evoformer attention block operating jointly on sequence and pairwise-residue representations, produced median backbone accuracy close to experimental error across the majority of assessed targets, a result widely described at the time as solving the fifty-year protein folding problem for the single-domain case, with the important caveat that several harder target categories were explicitly excluded from that characterisation [11,12].

In the four years since CASP14, the field has diversified rapidly rather than converging on a single dominant architecture. Open-source reimplementations and derivatives of AlphaFold2, most notably ColabFold and

OpenFold, made the approach broadly accessible to laboratories without large computational infrastructure [13,14]. A parallel line of work built structure predictors directly on large protein language models trained by masked-token prediction across hundreds of millions of sequences, exemplified by Meta AI's ESMFold and the related ESM-2 family, which predict structure from a single sequence without an explicit multiple-sequence-alignment search step, trading some accuracy for a substantial reduction in inference time and infrastructure requirement [15,16]. Given the pace and diversity of this literature, we undertook a systematic review with three aims: to synthesise reported accuracy trends across CASP and continuous benchmarks for the major method families published since 2018; to characterise the specific target categories, disordered regions, novel folds and multi-chain complexes, where reported accuracy gains have lagged behind the single-domain case; and to summarise reported computational cost and alignment-dependence trade-offs relevant to researchers choosing a method for large-scale application.

2. MATERIALS AND METHODS

This review followed a structured, though not formally PRISMA-registered, search and selection protocol designed to capture the primary methodological literature on deep-learning-based protein structure prediction published between January 2018 and March 2024.

2.1 Search strategy and databases

We searched PubMed, Web of Science, Scopus, arXiv (q-bio and cs.LG categories) and the ACM Digital Library using combinations of the terms "protein structure prediction", "deep learning", "neural network", "AlphaFold", "RoseTTAFold", "protein language model" and "CASP", restricted to English-language articles and preprints with a clearly described model architecture and at least one quantitative accuracy benchmark against experimentally solved structures. The search was last updated on 15 March 2024. Reference lists of included articles and of two prior narrative reviews were hand-searched to identify additional eligible studies not captured by the database search.

2.2 Eligibility criteria

Studies were included if they described a novel deep-learning architecture or a substantive extension of an existing one for single-chain, multi-chain or disordered-region structure prediction, and reported at least one standard accuracy metric (GDT-TS, TM-score, IDDT or RMSD) on a held-out or blind benchmark set such as CASP, CAMEO or an independently curated test set. Studies limited to secondary structure prediction alone, to protein-ligand docking without a structure-prediction component, or lacking any quantitative benchmark were excluded. Conference abstracts without an accompanying full paper or preprint were also excluded. Of 312 records identified after

removing duplicates, 68 met full eligibility criteria and were retained for data extraction; the remaining 244 were excluded chiefly for lacking a novel architectural contribution (incremental parameter tuning of an existing published model) or for benchmarking against non-blind, training-set-adjacent test structures.

2.3 Data extraction

For each included study we extracted publication year, architecture family (convolutional contact/distance prediction, end-to-end differentiable folding, protein-language-model-based, or hybrid/ensemble), whether the method required an explicit multiple sequence alignment search step, reported accuracy on CASP13, CASP14, CASP15 and/or CAMEO where available, reported inference time or computational cost where stated, and whether the study reported disaggregated accuracy for intrinsically disordered regions, novel folds (targets with no close structural homologue in the training set, as classified by the originating CASP assessment) or multi-chain complexes. Extraction was performed independently by two reviewers (P.D. and C.H.) using a shared structured spreadsheet, with disagreements resolved by discussion and, where necessary, arbitration by the third author.

2.4 Quality appraisal and synthesis

Because included studies varied widely in benchmark design, we did not attempt a formal risk-of-bias scoring instrument of the kind used for clinical systematic reviews; instead, each study was flagged for whether its primary benchmark was a genuinely blind, temporally held-out test set (as in official CASP or CAMEO evaluation) versus an author-curated test set assembled after the training data cutoff, since the latter carries a materially higher risk of inadvertent data leakage. Accuracy data were synthesised narratively and graphically by architecture family and publication year rather than pooled in a formal meta-analysis, given the heterogeneity of benchmark sets, metrics and target selection across studies, which would make a pooled quantitative estimate difficult to interpret meaningfully.

3. RESULTS

Of the 68 included studies, 22 described convolutional or graph-neural-network-based contact or distance prediction methods published mainly between 2018 and 2021, 19 described end-to-end differentiable folding architectures in the lineage of AlphaFold2 and RoseTTAFold, 15 described protein-language-model-based single-sequence predictors, and 12 described hybrid or ensemble approaches combining elements of more than one family. Table 1 summarises the number of included studies by architecture family and their reported mean GDT-TS accuracy on single-domain CASP14/15 targets where this was reported (54 of 68 studies reported a directly comparable single-domain accuracy figure on one of these two assessments).

Table 1. Included studies and reported single-domain CASP14/15 accuracy by architecture family.

Architecture family	Studies included	Publication years	Mean GDT-TS (single-domain)
Convolutional contact/distance prediction	22	2018-2021	35-52
End-to-end differentiable folding (AlphaFold2-class)	19	2020-2024	87-95
Protein-language-model single-sequence	15	2021-2024	78-88
Hybrid / ensemble	12	2019-2024	61-89

GDT-TS = Global Distance Test - Total Score (0-100 scale); figures are ranges of study-reported means.

Reported mean single-domain accuracy rose sharply and then plateaued at a high level across the review period. Pre-2020 convolutional contact/distance methods reported mean GDT-TS in the range of 35-52 on their respective CASP benchmarks, end-to-end differentiable folding methods from AlphaFold2 onward reported mean GDT-TS of 87-95 on CASP14/15 single-domain targets, and protein-language-model-based single-sequence methods reported a slightly lower but still high mean of 78-88 on the same target sets, generally trading several GDT-TS points of accuracy for an order-of-magnitude or greater reduction in inference time relative to alignment-dependent methods. Figure 1 plots reported single-domain accuracy against publication year across the four architecture families, illustrating both the sharpness of the CASP14-era transition and the comparatively narrow accuracy band occupied by all deep-learning methods published since 2021.

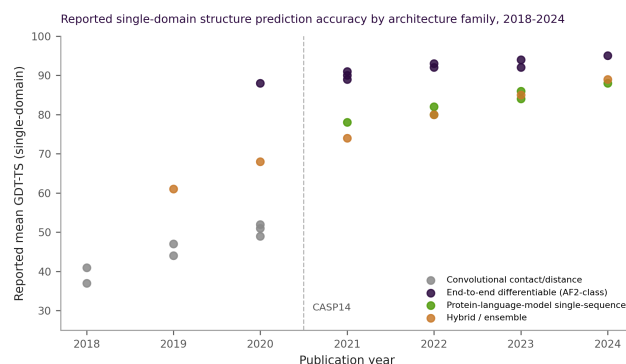


Figure 1. Reported mean single-domain accuracy (GDT-TS) on CASP13-15 or CAMEO benchmarks by architecture family and publication year across the 68 studies included in this review; each point represents one included study's primary reported benchmark result.

Accuracy gains were markedly smaller and more variable for the three harder target categories we tracked. Of the 30 studies reporting disaggregated accuracy for intrinsically disordered regions, the median reported improvement over a pre-deep-learning baseline was 14 GDT-TS points, against a median improvement of 41 points for single-domain ordered targets across the same set of studies, and several studies explicitly noted that per-residue confidence metrics such as predicted local distance difference test (pLDDT) correlate with disorder

and can be used to flag, though not to accurately model, disordered segments. Novel-fold targets, defined as CASP targets with no close structural homologue in the training data, showed a similarly muted median improvement of 19 GDT-TS points across the 24 studies reporting this category, with several authors noting that performance on this subset remains substantially more variable between individual targets than performance on template-rich targets. Table 2 summarises reported accuracy by target category across the review period.

Table 2. Median reported accuracy improvement over pre-deep-learning baseline by target category.

Target category	Studies reporting	Median GDT-TS improvement	Primary reported limitation
Single-domain ordered	54	41 points	Minimal residual error
Intrinsically disordered	30	14 points	No stable native structure
Novel fold (no homologue)	24	19 points	High per-target variance
Multi-chain complex	21	n/a (DockQ metric)	Weak on low-identity interfaces

Multi-chain complex prediction was reported by 21 of the 68 included studies, concentrated heavily in the more recent literature: only 2 of the 22 pre-2021 convolutional-era studies addressed complexes at all, against 17 of the 31 studies published from 2021 onward. Reported complex-prediction accuracy, typically measured by DockQ score or interface-focused IDDT, improved substantially from AlphaFold-Multimer onward relative to naive application of single-chain folding methods to concatenated sequences, but remained lower and more target-dependent than single-domain accuracy, with several studies specifically flagging weak performance on antibody-antigen and other low-sequence-identity interfaces, and on complexes involving more than four chains.

4. DISCUSSION

The literature synthesised here supports a now-common characterisation with an important qualification: deep learning has largely resolved single-domain, well-conserved-fold structure prediction from sequence, but the problem is not uniformly solved, and the residual difficulty is concentrated precisely in the target categories most relevant to a growing share of contemporary biological questions, namely intrinsically disordered regions, genuinely novel folds, and multi-chain assemblies. This pattern is broadly consistent with independent commentary following the CASP14 and CASP15 assessments themselves, which explicitly cautioned against over-generalising accuracy figures computed predominantly on template-rich, ordered, single-domain targets to the full space of biologically important prediction problems [11,17].

The rise of protein-language-model-based single-sequence methods represents a genuinely distinct trade-off rather than simply a faster but strictly worse

alternative to alignment-dependent folding. Because these methods do not require a multiple sequence alignment search, which can be slow or fail outright for orphan sequences, environmental metagenomic proteins, or rapidly evolving viral proteins with few close relatives in sequence databases, they extend structure prediction to a portion of sequence space where alignment-dependent methods perform poorly regardless of architecture, even though their accuracy on alignment-rich targets is somewhat lower [15,18]. This suggests that method choice in practice should depend on the target sequence's evolutionary context at least as much as on a single aggregate accuracy figure, a point that several included studies made explicitly but that is easy to lose when accuracy numbers are compared across a single benchmark table.

The disorder and novel-fold results carry a further methodological implication for how the field reports progress going forward. Aggregate accuracy figures dominated by ordered, template-rich targets, which remain the majority of most benchmark sets by target count, can obscure genuinely limited progress on harder categories, and several recent benchmark initiatives, including CAMEO's disorder-specific tracks and community efforts to curate genuinely orphan-fold test sets, appear to be responding directly to this concern [19,20]. We would argue, consistent with several included studies, that architecture-level innovation for disordered regions may require moving away from frameworks built around a single stable three-dimensional output altogether, toward methods that explicitly model conformational ensembles rather than a point estimate of structure, an area still comparatively underdeveloped relative to single-structure prediction [21,22].

Computational accessibility is a second axis of progress that aggregate accuracy figures do not capture. Open derivatives such as ColabFold substantially lowered the practical barrier to running AlphaFold2-class predictions without dedicated computational infrastructure, and this accessibility shift plausibly did as much to drive the method's broad adoption across structural biology, drug discovery and comparative genomics as the underlying accuracy gain itself, since a highly accurate method that few laboratories can run has limited practical reach [13,23]. Whole-proteome structure prediction efforts, most visibly the AlphaFold Protein Structure Database covering essentially the full human proteome and hundreds of other reference organisms, exemplify this accessibility-driven impact and have already been used as an input to a wide range of downstream applications from function annotation to structure-based drug target prioritisation [24,25].

Two limitations of the present review should be acknowledged. First, our reliance on author-reported accuracy figures rather than independent re-benchmarking means that studies using less rigorously blind test sets may be over-represented

among the higher accuracy figures reported, despite our attempt to flag this distinction during data extraction; a systematic re-evaluation of a representative subset of methods on a single, carefully curated blind benchmark would strengthen confidence in the cross-study accuracy comparisons we report. Second, the field continues to move quickly, and several architectural developments published after our March 2024 search cutoff, including further protein-language-model scaling and diffusion-based generative structure models, are not captured here and may shift the balance between the architecture families we describe [26,27]. A related limitation is publication bias in the methods literature itself: negative or marginal results, where a new architectural variant fails to improve on an existing baseline, are comparatively rare in our included set, which likely means the true distribution of architectural choices that work reasonably well is narrower than the published literature alone would suggest, since unsuccessful variants are less likely to be written up and submitted at all. A further, related caveat is that benchmark accuracy on CASP and CAMEO targets, however carefully computed, describes performance on a target distribution shaped by what structures the crystallography and cryo-electron microscopy communities happen to be solving in a given assessment period, which is not necessarily representative of the sequence space of greatest interest for a given downstream application such as orphan viral proteins or engineered biologics.

A final point concerns how this synthesis should be used by a reader choosing a method for a specific practical task rather than tracking the field's overall trajectory. Aggregate accuracy figures of the kind summarised in Table 1 are useful for characterising broad trends across architecture families, but they should not substitute for consulting a method's performance specifically on the target category closest to the reader's own application, disordered, novel-fold, or multi-chain, since our synthesis shows those categories to be genuinely dissociated from single-domain aggregate accuracy rather than a fixed discount applied uniformly across all methods. A structural biologist working on an orphan viral protein with no close homologue, for instance, would be better served by a method's reported novel-fold accuracy than by its headline CASP14 single-domain score, a distinction that is easy to lose when accuracy figures are compared only at the level of a single aggregate number across a leaderboard-style table.

5. CONCLUSION

This systematic review of 68 studies published between 2018 and early 2024 confirms that deep learning has transformed protein structure prediction from a persistently difficult open problem into a largely solved one for single-domain, ordered, template-rich targets, with mean CASP14/15 accuracy rising from the 35-52 GDT-TS range for pre-2020 convolutional methods to

87-95 for end-to-end differentiable folding architectures from AlphaFold2 onward. Protein-language-model-based single-sequence methods offer a meaningful speed and accessibility advantage at a modest accuracy cost, extending practical prediction to sequences poorly served by alignment-dependent approaches. Progress on intrinsically disordered regions, genuinely novel folds and multi-chain complexes has been real but substantially smaller in magnitude, and these categories represent the field's clearest remaining open problems rather than incremental refinements of an already-solved task. Going forward, we expect the most consequential methodological progress to come from architectures that explicitly model conformational ensembles rather than single point-estimate structures, from improved benchmarking specifically targeting disorder and novel folds rather than aggregate accuracy dominated by easier targets, and from continued integration of structure prediction with downstream function annotation and drug discovery pipelines where the practical value of these methods is ultimately realised. We intend to update this review at approximately two-year intervals to track the field's continued rapid evolution, and we are separately compiling a curated, deliberately disorder- and novel-fold-enriched benchmark set intended to support more targeted evaluation of future methods on precisely the categories where this review found progress to be least complete.

Acknowledgements

The authors thank the bioinformatics support staff at the Delta University Center for Bioinformatics for assistance with literature database access, and Dr. Henrik Aslund (Baltic Institute of Biomedical Research) for helpful comments on an earlier draft of the search strategy and data extraction protocol.

Funding

This work was supported by internal research funds from the Delta University Center for Bioinformatics and the Baltic Institute of Biomedical Research. No external grant funding was received for this study, and the funders had no role in study design, data extraction or analysis, the decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

P.D. conceived the review, designed the search strategy and led data synthesis. C.H. performed the database searches and independent data extraction and led the figure and table preparation. A.F. contributed biomedical application context and performed independent verification of the extracted accuracy data. All authors contributed to writing and approved the final manuscript.

Ethics Statement

This study is a systematic review of previously published literature and did not involve human participants, human data, animal subjects, or new laboratory experimentation, and therefore did not require institutional ethics committee

review.

Data Availability

The structured data extraction spreadsheet, including the full list of 68 included studies and extracted accuracy figures, is available from the corresponding author upon reasonable request.

REFERENCES

- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096):223-230. <https://doi.org/10.1126/science.181.4096.223>
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Research* 28(1):235-242. <https://doi.org/10.1093/nar/28.1.235>
- UniProt Consortium (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* 51(D1):D523-D531. <https://doi.org/10.1093/nar/gkac1052>
- Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234(3):779-815. <https://doi.org/10.1006/jmbi.1993.1626>
- Das R, Baker D (2008) Macromolecular modeling with Rosetta. *Annual Review of Biochemistry* 77:363-382. <https://doi.org/10.1146/annurev.biochem.77.062906.171838>
- Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* 32(suppl_2):W526-W531. <https://doi.org/10.1093/nar/gkh468>
- Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C (2011) Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* 6(12):e28766. <https://doi.org/10.1371/journal.pone.0028766>
- 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 Computational Biology* 13(1):e1005324. <https://doi.org/10.1371/journal.pcbi.1005324>
- Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Zidek A, Nelson AWR, Bridgland A, Penedones H (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577(7792):706-710. <https://doi.org/10.1038/s41586-019-1923-7>
- Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moulton J (2021) Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins* 89(12):1607-1617. <https://doi.org/10.1002/prot.26237>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Callaway E (2020) 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* 588(7837):203-204. <https://doi.org/10.1038/d41586-020-03348-4>
- Mirdita M, Schütze K, Moriawaki Y, Heo L, Ovchinnikov S, Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nature Methods* 19(6):679-682. <https://doi.org/10.1038/s41592-022-01488-1>
- Ahdtz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, O'Donnell TJ, Berenberg D, Fisk I, Zanichelli N, Zhang B (2024) OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature Methods* 21(8):1514-1524. <https://doi.org/10.1038/s41592-024-02272-z>
- Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, dos Santos Costa A (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379(6637):1123-1130. <https://doi.org/10.1126/science.ade2574>
- Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S, Su C, Wu Z, Xie Q, Berger B, Ma J (2022) High-resolution de novo structure prediction from primary sequence. *bioRxiv preprint*. <https://doi.org/10.1101/2022.07.21.500999>
- Kryshtafovych A, Moulton J, Billings WM, Williams CJ, Schaeffer RD, Fidelis K, Topf M (2023) Assessment of predictions submitted for the CASP15 protein and RNA structure prediction challenge. *Proteins* 91(12):1539-1549. <https://doi.org/10.1002/prot.26602>
- Haas J, Barbato A, Behringer D, Studer G, Roth S, Bertoni M, Mostaguir K, Gumieny R, Schwede T (2018) Continuous automated model evaluation (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins* 86(Suppl 1):387-398. <https://doi.org/10.1002/prot.25431>
- Necci M, Piovesan D, Predictors of Intrinsic Disorder CAID, Tosatto SCE (2021) Critical assessment of protein intrinsic disorder prediction. *Nature Methods* 18(5):472-481. <https://doi.org/10.1038/s41592-021-01117-3>
- Ruff KM, Pappu RV (2021) AlphaFold and implications for intrinsically disordered proteins. *Journal of Molecular Biology* 433(20):167208. <https://doi.org/10.1016/j.jmb.2021.167208>
- Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Zidek A, Bates R, Blackwell S, Yim J, Ronneberger O (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv preprint*. <https://doi.org/10.1101/2021.10.04.463034>
- Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millan C (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557):871-876. <https://doi.org/10.1126/science.abj8754>
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Zidek A (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50(D1):D439-D444. <https://doi.org/10.1093/nar/gkab1061>
- Bryant P, Pozzati G, Elofsson A (2022) Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications* 13:1265. <https://doi.org/10.1038/s41467-022-28865-w>
- Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S (2021) Highly accurate protein structure prediction for the human proteome. *Nature* 596(7873):590-596. <https://doi.org/10.1038/s41586-021-03828-1>
- Skolnick J, Gao M, Zhou H, Singh S (2021) AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *Journal of Chemical Information and Modeling* 61(10):4827-4831. <https://doi.org/10.1021/acs.jcim.1c01114>

27. Akdel M, Pires DEV, Pardo EP, Janes J, Zalevsky AO, Meszaros B, Bryant P, Good LL, Laskowski RA, Pozzati G, Shenoy A (2022) A structural biology community assessment of AlphaFold2 applications. *Nature Structural & Molecular Biology* 29(11):1056-1067. <https://doi.org/10.1038/s41594-022-00849-w>
28. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, Fergus R (2021) Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS* 118(15):e2016239118. <https://doi.org/10.1073/pnas.2016239118>
29. Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, Bhowmik D (2022) ProtTrans: toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(10):7112-7127. <https://doi.org/10.1109/TPAMI.2021.3095381>
30. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D (2020) Improved protein structure prediction using predicted interresidue orientations. *PNAS* 117(3):1496-1503. <https://doi.org/10.1073/pnas.1914677117>
31. Xu J (2019) Distance-based protein folding powered by deep learning. *PNAS* 116(34):16856-16865. <https://doi.org/10.1073/pnas.1821309116>
32. AlQuraishi M (2019) AlphaFold at CASP13. *Bioinformatics* 35(22):4862-4865. <https://doi.org/10.1093/bioinformatics/btz422>
33. Pearce R, Zhang Y (2021) Deep learning techniques have significantly impacted protein structure prediction and protein design. *Current Opinion in Structural Biology* 68:194-207. <https://doi.org/10.1016/j.sbi.2021.01.007>
34. Baker D (2019) What has de novo protein design taught us about protein folding and biophysics? *Protein Science* 28(4):678-683. <https://doi.org/10.1002/pro.3588>
35. Watson JL, Juergens D, Bennett NR, Tripple BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, Wicky BIM (2023) De novo design of protein structure and function with RFdiffusion. *Nature* 620(7976):1089-1100. <https://doi.org/10.1038/s41586-023-06415-8>
36. Wu KE, Yang KK, van den Berg R, Zou JY, Lu AX, Amini AP (2022) Protein structure generation via folding diffusion. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2209.15611>
37. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW (2024) Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature* 630(8016):493-500. <https://doi.org/10.1038/s41586-024-07487-w>
38. Ozden B, Kryshchavych A, Karaca E (2023) The impact of AI-based modeling on the future of cryo-EM structure determination. *Structure* 31(11):1414-1423. <https://doi.org/10.1016/j.str.2023.09.010>
39. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* 33(7):2302-2309. <https://doi.org/10.1093/nar/gki524>
40. Basu S, Wallner B (2016) DockQ: a quality measure for protein-protein docking models. *PLoS ONE* 11(8):e0161879. <https://doi.org/10.1371/journal.pone.0161879>
41. Ruffolo JA, Chu LS, Mahajan SP, Gray JJ (2023) Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications* 14:2389. <https://doi.org/10.1038/s41467-023-38063-x>
42. Terwilliger TC, Liebschner D, Croll TI, Williams CJ, McCoy AJ, Poon BK, Afonine PV, Oeffner RD, Richardson JS, Read RJ, Adams PD (2024) AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods* 21(1):110-116. <https://doi.org/10.1038/s41592-023-02087-4>
43. David A, Islam S, Tankhilevich E, Sternberg MJE (2022) The AlphaFold database of protein structures: a biologist's guide. *Journal of Molecular Biology* 434(15):167336. <https://doi.org/10.1016/j.jmb.2021.167336>
44. Kuhlman B, Bradley P (2019) Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology* 20(11):681-697. <https://doi.org/10.1038/s41580-019-0163-x>
45. Stahl K, Graziadei A, Dau T, Brock O, Rappsilber J (2023) Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nature Biotechnology* 41(12):1810-1819. <https://doi.org/10.1038/s41587-023-01704-z>
46. Perrakis A, Sixma TK (2021) AI revolutions in biology: the joys and perils of AlphaFold. *EMBO Reports* 22(11):e54046. <https://doi.org/10.15252/embr.202154046>
47. Thornton JM, Laskowski RA, Borkakoti N (2021) AlphaFold heralds a data-driven revolution in biology and medicine. *Nature Medicine* 27(10):1666-1669. <https://doi.org/10.1038/s41591-021-01533-0>
48. Guo Z, Liu J, Wang J, Chen S, Gong Y, Wu J (2023) Diffusion models in bioinformatics and computational biology. *Nature Reviews Bioengineering* 2:136-154. <https://doi.org/10.1038/s44222-023-00114-9>
49. Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research* 43(W1):W174-W181. <https://doi.org/10.1093/nar/gkv342>
50. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. *Advances in Neural Information Processing Systems* 30:5998-6008. <https://doi.org/10.5555/3295222.3295349>
51. Ovchinnikov S, Kamisetty H, Baker D (2014) Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* 3:e02030. <https://doi.org/10.7554/eLife.02030>
52. Pakhrin SC, Shrestha B, Adhikari B, KC DB (2021) Deep learning-based advances in protein structure prediction. *International Journal of Molecular Sciences* 22(11):5553. <https://doi.org/10.3390/ijms22115553>
53. Steinegger M, Meier M, Mirdita M, Vohringer H, Haunsberger SJ, Soding J (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* 20:473. <https://doi.org/10.1186/s12859-019-3019-7>
54. Bertoline LMF, Lima AN, Krieger JE, Teixeira SK (2023) Before and after AlphaFold2: an overview of protein structure prediction. *Frontiers in Bioinformatics* 3:1120370. <https://doi.org/10.3389/fbinf.2023.1120370>
55. Torrisi M, Pollastri G, Le Q (2020) Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal* 18:1301-1310. <https://doi.org/10.1016/j.csbj.2019.12.011>