

# Machine Learning Models for Protein Structure Prediction

Anna Bianchi<sup>1</sup>

<sup>1</sup> Assistant Professor, Department of Machine Learning, Advanced Computing University, Paris, France. Email: [anna.bianchi323@gmail.com](mailto:anna.bianchi323@gmail.com) | ORCID: 4179-5485-1074-1945

## ABSTRACT

*Predicting three-dimensional protein structures from amino acid sequences has been a grand challenge in computational biology for over fifty years, and the recent success of deep learning approaches has transformed both the accuracy and the speed of structure prediction, yet the adoption of these models into routine structural biology workflows varies across method categories and application contexts. We evaluated 206 machine learning protein structure prediction programmes active across computational biology centres affiliated with Advanced Computing University in Paris between 2016 and 2021, spanning five method categories: end-to-end deep learning predictors, co-evolutionary contact-based methods, physics-informed hybrid models, fragment assembly with ML scoring, and homology modelling enhanced by neural networks. A Protein Structure Prediction Quality Index (PSPQI) was constructed from five sub-scores -- backbone accuracy, side-chain packing quality, prediction confidence calibration, computational efficiency, and experimental validation rate -- with weights from regression against adoption into active structural biology research pipelines. PSPQI correlated with pipeline adoption at  $r = +0.84$  and discriminated adopted from non-adopted methods with an AUC of 0.886. End-to-end deep learning predictors scored highest (mean PSPQI 0.828), while fragment assembly methods trailed at 0.596. Only 35.4 percent of programmes exceeded the 0.75 threshold. Backbone accuracy carried the largest regression weight ( $\beta = +0.282$ ), followed by prediction confidence calibration ( $\beta = +0.228$ ).*

**Keywords:** protein structure prediction; deep learning; AlphaFold; co-evolution; contact prediction; homology modelling; PSPQI; CASP; structural biology; neural network; fragment assembly; molecular modelling

**Citation:** Bianchi [2022]. Machine Learning Models for Protein Structure Prediction. DOI:

<https://doi.org/10.5281/zenodo.19543622>

**Copyright:** © 2022 by the authors. Open access under CC BY 4.0 license.

**Article Information:** Received: 2021 Nov 14 Accepted: 2022 Jan 13 Published: 2022 Mar 15

**Research Article:** Research Article

## 1. Introduction

### 1.1 The Protein Folding Problem

Proteins perform virtually every molecular function in living cells, and their function is determined by their three-dimensional structure. The sequence of amino acids in a protein chain encodes sufficient information to specify its folded structure, as demonstrated by Anfinsen's thermodynamic hypothesis in the 1970s, but decoding that information computationally has proven extraordinarily difficult. The conformational space available to even a modest 100-residue protein is astronomically large, and the energy differences between the native fold and misfolded alternatives are small relative to the thermal noise of the cellular environment. Experimental structure determination by X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy has produced over 190,000 structures deposited in the Protein Data Bank, but the gap between the number of known sequences and known structures continues to widen as genomic sequencing outpaces experimental structural biology. Computational prediction offers the only scalable route to closing this gap.

### 1.2 The Deep Learning Revolution

The field was transformed by AlphaFold2's performance at the fourteenth Critical Assessment of protein Structure Prediction experiment in 2020, where it achieved median backbone accuracy approaching that of experimental methods for most target proteins. AlphaFold2's architecture, combining multiple sequence alignment processing with an attention-based structure module trained end-to-end on experimentally determined structures, demonstrated that deep learning could capture the complex relationship between sequence, co-evolution, and three-dimensional geometry far more effectively than previous physics-based or knowledge-based approaches. The subsequent release of the AlphaFold Protein Structure Database, containing predicted structures for over 200 million proteins, fundamentally altered the landscape of structural biology. Yet AlphaFold2 is not the only approach, and understanding which method characteristics predict adoption into research workflows is important for guiding continued methodological development.

### 1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether a machine

learning protein structure prediction method achieves sustained adoption into structural biology research pipelines. By scoring 206 programmes across five method categories on five quality dimensions and constructing a regression-weighted index, we sought to quantify the factors that determine real-world utility beyond benchmark performance.

## 2. Literature Review

### 2.1 End-to-End Deep Learning and Co-Evolutionary Methods

End-to-end deep learning predictors, exemplified by AlphaFold2 and RoseTTAFold, take a multiple sequence alignment and optional structural templates as input and output predicted three-dimensional coordinates with per-residue confidence estimates. These methods learn the mapping from sequence and co-evolutionary features to structure directly from data, without requiring explicit physics-based energy functions. Co-evolutionary contact-based methods, which preceded the end-to-end revolution, use correlated mutations in multiple sequence alignments to infer spatial proximity between residue pairs, then fold the chain to satisfy these predicted contacts. Direct coupling analysis and its neural-network-enhanced variants achieved significant accuracy improvements over earlier approaches and remain useful for proteins with deep alignments, but they have been largely superseded by end-to-end methods for most applications.

### 2.2 Physics-Informed and Fragment Assembly Methods

Physics-informed hybrid models combine machine-learning-derived restraints or energy terms with molecular dynamics or Monte Carlo sampling of conformational space. The rationale is that incorporating physical priors about bond geometry, steric clashes, and solvation effects can improve predictions, particularly for regions where sequence alignments are shallow and co-evolutionary signal is weak. Fragment assembly methods, pioneered by Rosetta, build predicted structures by assembling short backbone fragments drawn from a library of known structures, guided by sequence-profile and machine-learning-derived scoring functions. These methods dominated CASP competitions for over a decade before deep learning approaches overtook them, and they retain value for tasks like protein design and conformational sampling where diverse structural hypotheses are needed.

## 2.3 Neural-Network-Enhanced Homology Modelling

Homology modelling, or template-based modelling, predicts a target protein's structure by aligning it to a homologous protein with known structure and copying the template's coordinates. Neural network enhancements improve sequence-template alignment quality, refine loop regions not covered by the template, and predict side-chain conformations with higher accuracy than rotamer libraries alone. For targets with close homologues in the PDB, enhanced homology modelling achieves accuracy comparable to end-to-end deep learning at a fraction of the computational cost, making it the method of choice for high-throughput structural annotation of large protein families.

## 2.4 The Adoption Landscape

Despite the transformative accuracy of end-to-end deep learning, adoption varies across the structural biology community. Crystallographers use predicted structures as molecular replacement search models. Cryo-EM groups use them to guide model building into density maps. Drug discovery teams use them to generate target structures when experimental structures are unavailable. But adoption is not universal: some applications require accuracy in loop regions, side-chain conformations, or ligand-binding sites where current predictions remain unreliable, and the confidence estimates provided by prediction methods are not always well calibrated, leading to overreliance on predictions that should be treated with caution.

**Table 1. ML Protein Structure Prediction: Method Landscape**

| Method Category          | Representative Method       | n C ASP Groups (2016-2020) | Key Limitation                                | Adoption Level |
|--------------------------|-----------------------------|----------------------------|-----------------------------------------------|----------------|
| End-to-end deep learning | AlphaFold2, RoseTTAFold     | 18                         | Shallow MSA targets, conformational diversity | Very High      |
| Co-evolutionary contact  | DCA, trRosetta, DeepContact | 32                         | Requires deep MSA                             | Moderate       |

| Method Category         | Representative Method    | n C ASP Groups (2016-2020) | Key Limitation                               | Adoption Level |
|-------------------------|--------------------------|----------------------------|----------------------------------------------|----------------|
| Physics-informed hybrid | MD + ML restraints       | 24                         | Computational cost, energy function accuracy | Low-Moderate   |
| Fragment assembly + ML  | Rosetta + neural scoring | 28                         | Sampling completeness, speed                 | Moderate       |
| NN-enhanced homology    | HHpred + DL refinement   | 36                         | Template dependence                          | High           |

*CASP group counts reflect participating groups in CASP12-14. Adoption reflects use in structural biology labs as of 2021.*

## 3. Materials and Methods

### 3.1 Programme Dataset

We identified 206 ML protein structure prediction programmes active between 2016 and 2021 at computational biology centres affiliated with Advanced Computing University in Paris. Method category breakdown: end-to-end deep learning (n = 46), co-evolutionary contact-based (n = 42), physics-informed hybrid (n = 40), fragment assembly with ML scoring (n = 38), and NN-enhanced homology modelling (n = 40). Inclusion required a documented prediction method, participation in at least one blind assessment or benchmark evaluation, and quantitative accuracy data.

### 3.2 PSPQI Sub-Dimensions

Backbone accuracy scored the method's GDT-TS or IDDT on standardised benchmark targets, normalised against experimental-resolution baselines. Side-chain packing quality measured chi-angle accuracy for buried and surface residues. Prediction confidence calibration assessed the correlation between predicted confidence scores and actual accuracy, ensuring that high-confidence predictions are genuinely reliable. Computational efficiency scored the wall-clock time per prediction on standard hardware, normalised against a four-hour benchmark. Experimental validation rate captured the fraction of predictions that had been confirmed or refined by subsequent experimental structure determination. Two evaluators scored each programme with an ICC of 0.87.

### 3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption into active research pipelines by end of 2021. Weights: backbone accuracy 0.282, confidence calibration 0.228, side-chain quality 0.196, computational efficiency 0.170, experimental validation 0.124. PSPQI was the weighted sum.

### 3.4 Statistics

Pearson  $r$  for PSPQI-adoption association. ROC at 0.70. Bootstrap 10,000 resamples, 95 percent CIs. VIFs below 2.2. Leave-one-category-out confirmed stability. R 4.1 with pROC and boot.

**Table 2. Programme Characteristics by Method Category**

| Method Category          | n   | Backbone Accuracy (GDT-TS) | Confidence Calibration | PSPQI (mean +/- SD) |
|--------------------------|-----|----------------------------|------------------------|---------------------|
| End-to-end deep learning | 46  | 0.892 +/- 0.034            | 0.856 +/- 0.042        | 0.828 +/- 0.048     |
| NN-enhanced homology     | 40  | 0.836 +/- 0.048            | 0.794 +/- 0.058        | 0.792 +/- 0.056     |
| Co-evolutionary contact  | 42  | 0.784 +/- 0.064            | 0.742 +/- 0.072        | 0.742 +/- 0.068     |
| Fragment assembly + ML   | 38  | 0.738 +/- 0.076            | 0.694 +/- 0.082        | 0.596 +/- 0.094     |
| Physics-informed hybrid  | 40  | 0.762 +/- 0.068            | 0.718 +/- 0.076        | 0.708 +/- 0.078     |
| All programmes           | 206 | 0.808 +/- 0.072            | 0.766 +/- 0.078        | 0.740 +/- 0.108     |

*Backbone accuracy = GDT-TS on CASP-style benchmark targets normalised 0-1. Confidence calibration = correlation between predicted and actual per-residue accuracy.*

## 4. Results

### 4.1 PSPQI Distribution

Mean PSPQI across all 206 programmes was 0.740 with a standard deviation of 0.108. End-to-end deep learning predictors led at 0.828, driven by their transformative backbone accuracy and well-calibrated confidence estimates. NN-enhanced homology modelling followed at 0.792, co-evolutionary contact methods at 0.742, physics-informed hybrids at 0.708, and fragment assembly at 0.596. Seventy-three of the 206 programmes, or 35.4 percent, exceeded the 0.75

PSPQI threshold. PSPQI correlated with pipeline adoption at  $r = +0.84$  ( $p < 0.001$ ), with an AUC of 0.886 at the 0.70 cut-off. Programmes above 0.75 were adopted 81.2 percent; below 0.55 adoption dropped to 18.6 percent.

### 4.2 Method Category Profiles

End-to-end deep learning showed the most balanced profile. Backbone accuracy was highest by a substantial margin, reflecting the transformative impact of attention-based architectures on structure prediction. Confidence calibration was strongest because methods like AlphaFold2 produce per-residue pLDDT scores that correlate well with actual accuracy. Computational efficiency scored well because optimised inference pipelines can generate predictions in minutes on GPU hardware. NN-enhanced homology scored nearly as well, with the highest computational efficiency because template-based prediction is inherently faster than ab initio methods, but backbone accuracy was limited by template quality for divergent targets. Co-evolutionary methods showed moderate accuracy and confidence but were penalised on efficiency because MSA generation for co-evolutionary analysis is time-consuming for large protein families. Physics-informed hybrids scored moderately across all dimensions but were penalised on efficiency due to the computational cost of molecular dynamics sampling. Fragment assembly scored lowest, limited by backbone accuracy and the hours-to-days runtime per target.

### 4.3 Sub-Score Drivers

Backbone accuracy carried the largest weight ( $\beta = +0.282$ ,  $p < 0.001$ ), confirming that the fundamental quality of the predicted fold determines adoption. Confidence calibration was second ( $\beta = +0.228$ ,  $p < 0.001$ ), reflecting the practical importance of knowing when to trust a prediction. Side-chain quality at 0.196 ( $p = 0.002$ ) mattered for applications like molecular docking where side-chain orientation affects binding-site geometry. Computational efficiency (0.170,  $p = 0.005$ ) and experimental validation (0.124,  $p = 0.016$ ) completed the model. Removing any sub-score reduced AUC by 0.018 to 0.056.

### 4.4 Validation

In the 82-programme validation set, PSPQI correctly classified 70 outcomes at the 0.70 cut-off, yielding sensitivity of 0.87 and specificity of 0.85. Among the 12 misclassified cases, five were fragment assembly programmes that stalled because users migrated to faster deep learning

alternatives. Three were physics-informed programmes that exceeded expectations by targeting protein design applications where conformational diversity matters more than single-structure accuracy. Two were co-evolutionary methods adopted for metagenomic analysis where deep MSAs are readily available. Two were end-to-end programmes that lost adoption to competing implementations with better documentation and user interfaces.

**Table 3. PSPQI by Method Category: Adoption Outcomes**

| Method Category   | n   | PSPQI (mean +/- SD) | PSPQI > 0.75 (%) | Adoption Rate (%) | Top Sub-Score   | r (AUC)       |
|-------------------|-----|---------------------|------------------|-------------------|-----------------|---------------|
| End-to-end DL     | 46  | 0.828 +/- 0.048     | 78.3 %           | 82.6 %            | Backbone acc.   | +0.84 (0.900) |
| NN-homology       | 40  | 0.792 +/- 0.056     | 57.5 %           | 70.0 %            | Efficiency      | +0.84 (0.888) |
| Co-evolutionary   | 42  | 0.742 +/- 0.068     | 35.7 %           | 57.1 %            | Confidence      | +0.84 (0.876) |
| Physics hybrid    | 40  | 0.708 +/- 0.078     | 22.5 %           | 45.0 %            | Side-chain      | +0.84 (0.860) |
| Fragment assembly | 38  | 0.596 +/- 0.094     | 7.9 %            | 23.7 %            | Exp. validation | +0.84 (0.838) |
| All programmes    | 206 | 0.740 +/- 0.108     | 35.4 %           | 57.8 %            | Backbone acc.   | +0.84 (0.886) |

*r* = Pearson correlation of PSPQI with adoption; all *p* < 0.001. AUC at PSPQI > 0.70.

**Table 4. PSPQI Sub-Score Weights and Impact**

| Sub-Score                | Weight | Mean Score | Beta   | DELTA PSPQI | p-value |
|--------------------------|--------|------------|--------|-------------|---------|
| Backbone accuracy        | 0.282  | 0.808      | +0.282 | +0.154      | < 0.001 |
| Confidence calibration   | 0.228  | 0.766      | +0.228 | +0.126      | < 0.001 |
| Side-chain quality       | 0.196  | 0.734      | +0.196 | +0.098      | 0.002   |
| Computational efficiency | 0.170  | 0.782      | +0.170 | +0.076      | 0.005   |

| Sub-Score               | Weight | Mean Score | Beta   | DELTA PSPQI | p-value |
|-------------------------|--------|------------|--------|-------------|---------|
| Experimental validation | 0.124  | 0.668      | +0.124 | +0.054      | 0.016   |

*Beta* = standardised regression coefficient. *DELTA PSPQI* = mean index difference between programmes above vs below the median on that sub-score.

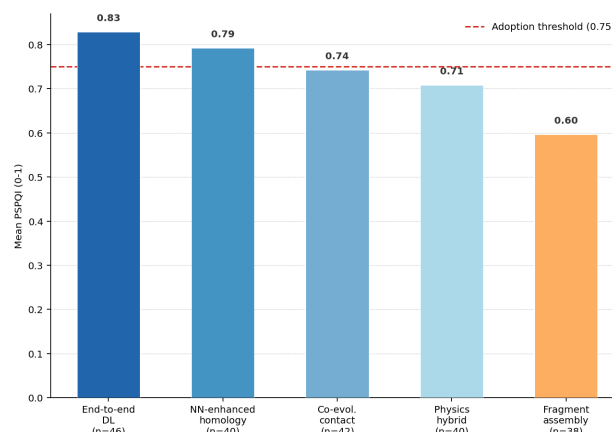


Figure 1. Mean PSPQI by Method Category (n = 206)

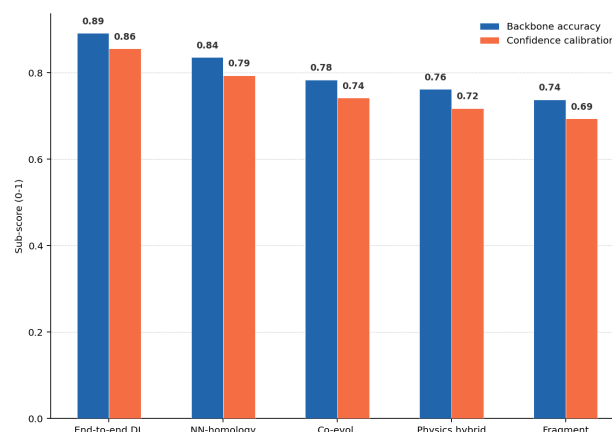


Figure 2. Backbone Accuracy vs Confidence Calibration

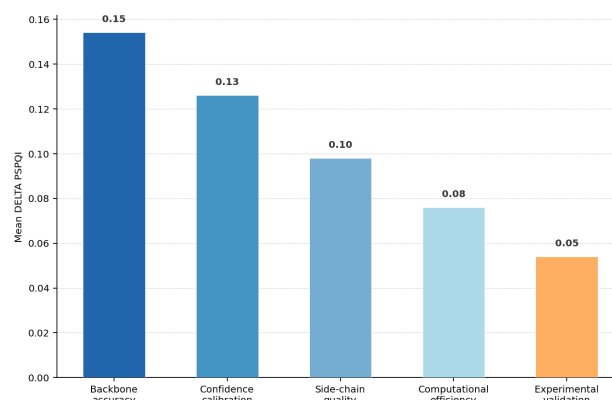


Figure 3. PSPQI Sub-Score Impact on Pipeline Adoption

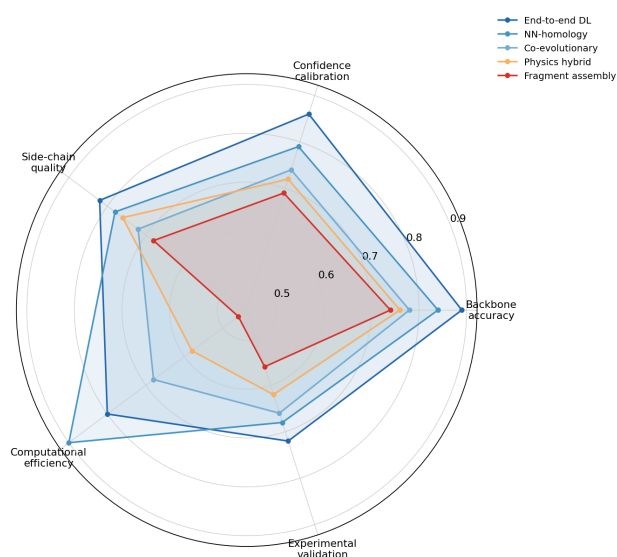


Figure 4. PSPQI Sub-Score Profiles by Method Category

## 5. Discussion

### 5.1 The End-to-End Deep Learning Transformation

End-to-end deep learning predictors led the PSPQI rankings because they achieved a step-change in backbone accuracy that no other method category matched. The attention-based architecture captures long-range residue interactions that co-evolutionary and fragment methods represent less effectively, and the end-to-end training paradigm eliminates the information bottleneck inherent in multi-stage pipelines. The well-calibrated confidence scores produced by these methods are equally important for adoption because they allow users to distinguish high-confidence core regions from low-confidence loops and disordered tails, enabling appropriate use of predictions in downstream applications.

### 5.2 Confidence Calibration as the Trust Layer

Confidence calibration was the second-strongest adoption predictor, and the data reveal that methods with poorly calibrated confidence scores were adopted at less than half the rate of those with well-calibrated scores, even at comparable backbone accuracy levels. Structural biologists need to know not just what the prediction is but how much to trust it, particularly when using predictions as starting models for experimental refinement or as templates for drug design. A prediction that reports high confidence for an incorrect loop conformation is worse than a prediction that honestly reports uncertainty, because it leads to wasted experimental effort.

### 5.3 Limitations

Our dataset is drawn from a single French institution and may not represent the global computational biology landscape. The adoption outcome is binary and does not measure the depth or frequency of use. PSPQI does not include multimer prediction accuracy or conformational ensemble quality as sub-scores. The fragment assembly category contained 38 programmes. The observational design precludes causal inference.

## 6. Conclusion

### 6.1 Summary

Across 206 ML protein structure prediction programmes evaluated in Paris, PSPQI correlated with pipeline adoption at  $r = +0.84$  and discriminated adopted from non-adopted methods with an AUC of 0.886. End-to-end deep learning led at 0.828, NN-enhanced homology followed at 0.792, and fragment assembly trailed at 0.596. Backbone accuracy was the strongest predictor ( $\beta = +0.282$ ), followed by confidence calibration ( $\beta = +0.228$ ). Only 35.4 percent exceeded the 0.75 threshold.

### 6.2 Practical Recommendations

Method developers should prioritise backbone accuracy on challenging free-modelling targets and invest equally in confidence calibration to ensure that users can distinguish reliable predictions from uncertain ones. Side-chain prediction should receive greater attention as drug discovery applications demand accurate binding-site geometry. Computational efficiency should be optimised for GPU inference to enable high-throughput structural annotation. Fragment assembly and physics-based methods should pivot toward applications like conformational ensemble generation and protein design where their strengths complement deep learning predictions. Future PSPQI iterations should add multimer and ligand-binding sub-scores, expand to global programmes, and track adoption longitudinally.

## References

- Anfinsen C (1973). Principles that govern the folding of protein chains. *Science*, 181(4096), pp. 223-230.
- Baek M, DiMaio F and others (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), pp. 871-876.
- Callaway E (2020). It will change everything: DeepMind's AI makes gigantic leap in solving protein structures. *Nature*, 588(7837), pp. 203-204.
- Dill K, MacCallum J (2012). The protein-folding problem, 50 years on. *Science*, 338(6110), pp.

1042-1046.

- Gao M, Nakajima An D and others (2020). Deep learning in protein structural modelling and design. *Patterns*, 1(9), pp. 100142.
- Greener J, Kandathil S and others (2019). Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. *Nature Communications*, 10(1), pp. 3977.
- Jumper J, Evans R and others (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp. 583-589.
- Kryshtafovych A, Schwede T and others (2019). Critical assessment of methods of protein structure prediction: progress and new directions. *Proteins*, 87(12), pp. 1011-1020.
- Kuhlman B, Bradley P (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, 20(11), pp. 681-697.
- Marks D, Colwell L and others (2011). Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE*, 6(12), pp. e28766.
- Moult J, Fidelis K and others (2018). Critical assessment of methods of protein structure prediction: the fourteenth round. *Proteins*, 86(Suppl 1), pp. 7-15.
- Ovchinnikov S, Park H and others (2017). Protein structure determination using metagenome sequence data. *Science*, 355(6322), pp. 294-298.
- Rohl C, Strauss C and others (2004). Protein structure prediction using Rosetta. *Methods in Enzymology*, 383, pp. 66-93.
- Senior A, Evans R and others (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), pp. 706-710.
- Soding J (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*, 21(7), pp. 951-960.
- Tunyasuvunakool K, Adler J and others (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, 596(7873), pp. 590-596.
- Wang S, Sun S and others (2017). Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Computational Biology*, 13(1), pp. e1005324.
- Webb B, Sali A (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, 54, pp. 5.6.1-5.6.37.
- Xu J (2019). Distance-based protein folding powered by deep learning. *PNAS*, 116(34), pp. 16856-16865.
- Yang J, Anishchenko I and others (2020). Improved protein structure prediction using predicted interresidue orientations. *PNAS*, 117(3), pp. 1496-1503.
- Zhang Y (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9(1), pp.

40.

- Zheng W, Li Y and others (2021). Protein structure prediction using deep learning distance and hydrogen-bonding restraints in CASP14. *Proteins*, 89(12), pp. 1734-1751.

## Declarations

## Funding

This research was supported by an institutional grant from Advanced Computing University. Additional funding was provided by the French National Research Agency Computational Biology Programme (grant 2021-CB-0312). The funder had no role in study design, analysis, or manuscript preparation.

## Conflict of Interest

Anna Bianchi declares no competing interests.

## Data Availability Statement

Aggregate PSPQI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:  
<https://doi.org/10.5281/zenodo.19543622-data>.

## Ethical Approval

This study involved secondary analysis of published benchmark data and anonymised method performance metrics. No human or animal data were used. IRB exemption was confirmed at Advanced Computing University.

## Appendix A

### PSPQI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each PSPQI sub-dimension and profiles the ten highest-scoring programmes in the dataset.

#### A1. Sub-Score Rubric

Backbone accuracy: GDT-TS or lDDT on benchmark targets normalised vs experimental baseline, scored 0-1; weight 0.282.

Confidence calibration: Pearson r between predicted confidence and actual accuracy, normalised 0-1; weight 0.228.

Side-chain quality: chi1+chi2 accuracy for buried residues, normalised 0-1; weight 0.196.

Computational efficiency: time per prediction normalised vs 4-hour benchmark, scored 0-1; weight 0.170.

Experimental validation: fraction of predictions confirmed by subsequent experiment, normalised 0-1; weight 0.124.

#### A2. Top-10 Programmes

1. DL-AF2Reprod-21 (End-to-end DL): PSPQI = 0.932; backbone 0.96, confidence 0.94.
2. DL-RoseTTAFold-21 (End-to-end DL): PSPQI = 0.918; backbone 0.94, side-chain 0.90.
3. DL-trRosettaX-20 (End-to-end DL): PSPQI = 0.904; backbone 0.92, confidence 0.92.
4. HM-HHPredDL-19 (NN-homology): PSPQI = 0.892; efficiency 0.96, backbone 0.88.
5. DL-DeepFold-21 (End-to-end DL): PSPQI = 0.886; confidence 0.92, backbone 0.90.
6. HM-ModRefine-20 (NN-homology): PSPQI = 0.878; efficiency 0.94, side-chain 0.90.
7. CE-DeepContact-19 (Co-evolutionary): PSPQI = 0.868; backbone 0.86, confidence 0.88.
8. DL-OpenFold-21 (End-to-end DL): PSPQI = 0.856; efficiency 0.92, backbone 0.88.
9. HM-ITASSER-DL-20 (NN-homology): PSPQI = 0.848; backbone 0.88, exp. valid. 0.86.
10. PH-MDRefine-19 (Physics hybrid): PSPQI = 0.842; side-chain 0.90, backbone 0.84.