Inverse folding improves structure prediction for confidence-aware peptide deorphanization

A.Dougha[1], S. Murail[1], J. Leprince[2,3], P. Tufféry[1]

[1]BFA, INSERM ERL 1133, CNRS UMR 8251, Université Paris Cité, F-75013 Paris, France
[2]Univ Rouen Normandie, Inserm, NorDiC UMR 1239, Rouen, France
[3]Univ Rouen Normandie, Inserm, CNRS, HeRacLeS US 51 UAR 2026, PRIMACEN, Rouen, France
Contact: pierre.tuffery@u-paris.fr, alexis.dougha@u-paris.fr

https://doi.org/10.17952/38EPS.2026.P269

1. Background and objectives

- Peptide hormones are involved in the regulation of a wide range of physiological processes, including but not limited to homeostasis, immunity and growth. They bind to their cognate receptors to trigger downstream signaling pathways (see **Figure 1**), but the identity of the receptors is not always known. In some cases, experimental evidence suggests that the peptide of interest binds to a G protein-coupled receptor (GPCR) which remains to be identified.

- Structure prediction-based confidence metrics like the interface predicted Template Modeling (ipTM) score have been applied to solve this issue[1]. While promising, the protocol suffers from the imperfection of the ipTM, which makes its use limited under a realistic peptide deorphanization scenario where only a small number of candidates can be tested.

- We aim to identify the best protocol to deorphanize peptides, benchmarking structure prediction (AlphaFold-Multimer[2], AlphaFold 3[3], Boltz-2[4]) and scoring functions including popular ones (ipTM, ipSAE[5], LIS[6], pDockQ[7], pDockQ2[8], peptide pLDDT) and others based on inverse folding used in this context for the first time (peptide perplexity computed with ESM-IF1[9] and ProteinMPNN[10]).

- Specifically, we address the deorphanization process by considering the value of the proposed protocol in a real-world scenario where, beyond ranking candidate pairs, our protocol enables evaluation of the probability of success when testing only a small number of top predictions (i.e., confidence-aware peptide deorphanization).

- First, all the combinations of structure prediction models and scoring functions are benchmarked on a rat dataset (17 peptides, 359 GPCRs, 29 interacting pairs). Then, the best protocol was applied to a larger human dataset (122 peptides, 393 GPCRs, 244 interacting pairs) assembled from IUPHAR[11].

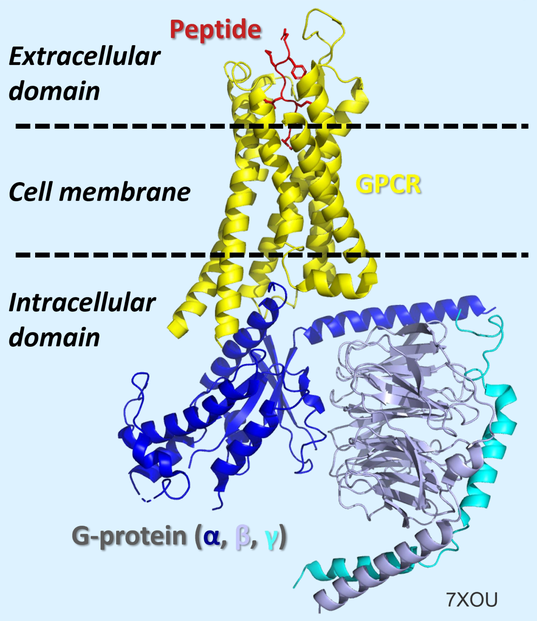


Figure 1. A peptide–GPCR–G protein complex resolved by cryo-EM. Cholecystokinin (CCK) binds to the CCK receptor type A, promoting G protein recruitment. CCK is known to regulate digestion.

2. Computational workflow

- For each peptide (**Figure 2**), we predict its structure in interaction with all the GPCRs to be tested *in silico*. We generated 25 structures per peptide-GPCR pair for each model evaluated.

- All the scores are evaluated with all the structure prediction models on the rat dataset, so that we keep only the best combinations for the large-scale prediction of the human dataset (2 * 25 models * 122 peptides * 393 GPCRs = 2,397,300 predicted structures). In **Figure 2**, only the best scoring function associated with AFmm (ProteinMPNN perplexity) and AF3 (ipSAE) is shown.

- We employed the ColabFold[12] implementation of AFmm (MMSeqs2 MSA and dropout enabled), and AF3 with default parameters (data pipeline using JackHMMER MSA).

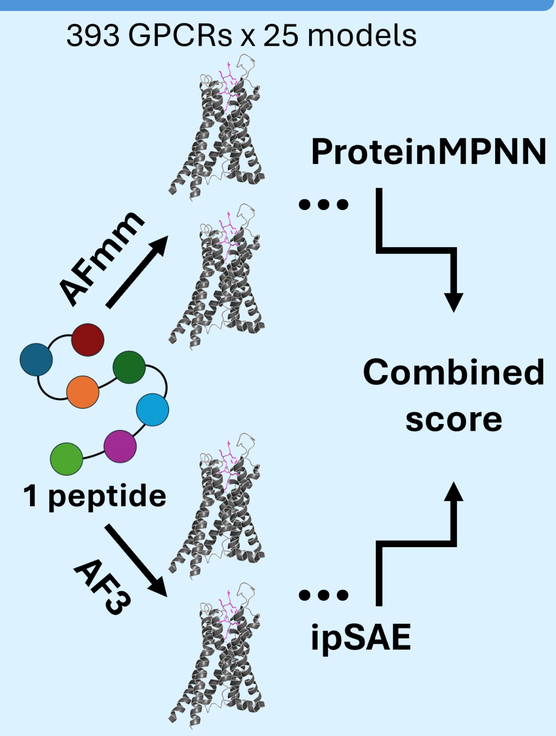


Figure 2. Computational workflow applied for each case study. All the structures are scored with a range of scoring functions benchmarked. Boltz-2 was also employed for the rat dataset.

3. Benchmark with rat interactions

- Interestingly, ipSAE is the best score (in term of average precision AP) when employed with AF3, while for AFmm and Boltz-2, rescoring with the peptide perplexity computed with ProteinMPNN is beneficial (**Figure 3**).

- Although Boltz-2 was trained on a larger dataset, AF3 and AFmm achieve better performance.

- Even if AFmm achieves a higher AP than AF3, we speculate that the combination of both protocols could improve the results since AF3 performs better than AFmm in the high-recall region of the curve.

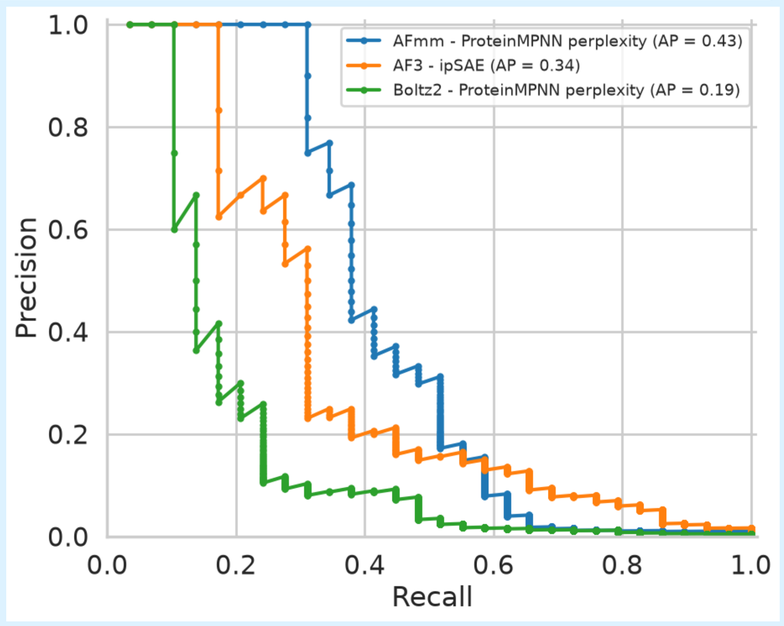


Figure 3. Precision-recall curves showing only the best scoring function associated with each structure prediction model tested (rat dataset). The average precision (AP) values summarize the overall performance of each method. While AFmm has the highest AP, AF3 performs better in the high-recall region (above 0.6). Boltz-2 performs poorly compared with the two other protocols, so it was not employed for the human dataset.

4. Validation on human interactions

- The final protocol considers both AFmm and AF3-derived scores in a new combined one. For each score threshold, both precisions were calculated as the proportion of correct pairs among all GPCR-peptide pairs with a score exceeding the corresponding threshold. The combined score was then obtained by averaging the precision values of the two scores.

- The protocols based on AFmm + ProteinMPNN and AF3 + ipSAE achieve similar performance when considering only the top *N* predictions where $N \leq 7$, which is the most important region of the curve (**Figure 4**) for the peptide deorphanization problem (we want to experimentally test a small number of GPCRs for a given peptide of interest).

- The combined score outperforms the two separate ones (based on AFmm only or AF3 only) to identify a larger number of peptide-GPCR pairs when only considering a small number of potential GPCRs for each peptide.

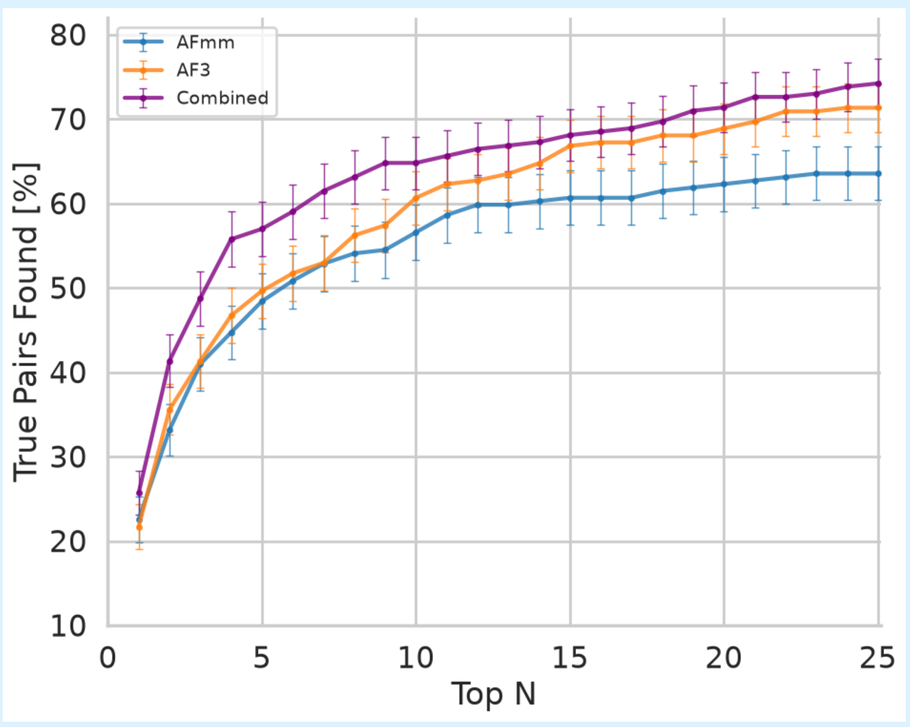


Figure 4. Enrichment curves showing the percentage of true pairs recovered within the top *N* predictions (human dataset). For each value of *N*, the percentage of true pairs recovered was calculated across 1,000 bootstrap resamples. Points represent the mean across bootstrap replicates, with error bars indicating the corresponding standard deviation.

5. Conclusion

1. Inverse folding can be used in combination with structure prediction to prioritize potential GPCRs that can interact with a peptide of interest.
2. Combining "orthogonal" protocols (AFmm/ProteinMPNN + AF3/ipSAE) improves the performance.
3. The combined protocol was employed to deorphanize four peptides of known activity but unknown target. For each one, at least one GPCR was predicted with a confidence score over 53%.
4. 133 potential secondary peptide–GPCR interactions were identified with high confidence in humans (defined as not annotated as interacting in IUPHAR, predicted rank = 1 or confidence score over 50%).
5. *In vitro* validation is ongoing to validate the predictions of points 3 and 4.

References

[1] Teufel, F. et al, *J. Chem. Inf. Model.* **2023**, *63(9)*, 2651-2655.
[2] Evans, R. et al, *biorxiv* **2021**.
[3] Abramson, J. et al, *Nature* **2024**, *630(8016)*, 493-500.
[4] Passaro, S. et al, *biorxiv* **2025**.
[5] Dunbrack Jr, R. L., *biorxiv* **2025**.
[6] Kim, A. R., *biorxiv* **2024**.
[7] Bryant, P. et al, *Nat. Commun.* **2022**, *13(1)*, 1265.
[8] Zhu, W. et al, *Bioinformatics* **2023**, *39(7)*, btad424.
[9] Hsu, C. et al, *ICML* **2022**.
[10] Dauparas, J. et al, *Science* **2022**, *378(6615)*, 49-56.
[11] Harding, S. D. et al, *Nucleic Acids Res.* **2026**, *54(D1)*, D1446-D1456.
[12] Mirdita, M. et al, *Nat. Methods* **2022**, *19(6)*, 679-682.
[13] Alland, C. et al, *Nucleic Acids Res.* **2026**, *33(suppl_2)*, W44-W49.

Acknowledgements

This work was supported by Université Paris Cité. Computational resources were partly provided by RPBS[13]. This project was provided with computing HPC and storage resources by GENCI at IDRIS thanks to the grant 2026-AD010717472 on the supercomputer Jean Zay's CSL, V100, A100, H100 partitions.