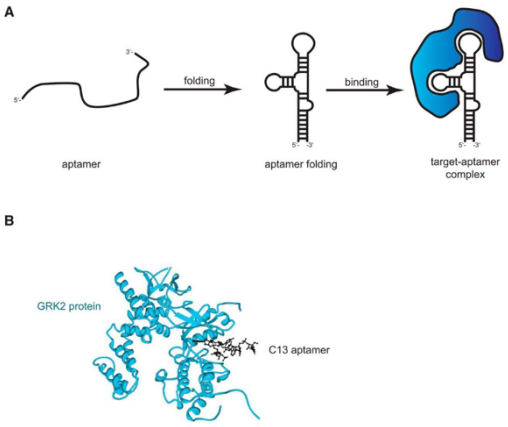
AI-Driven Evaluation of Aptamer–Protein Interaction Prediction

This project was completed in collaboration with Imose Iduozee, Matias Ylitolva, Joonatan Lindfors, Sofia Hänninen, Jaakko Pulkkinen and Linda Jalonen.

Introduction



Aptamers are short DNA or RNA molecules capable of binding proteins with high affinity and specificity, making them attractive alternatives to antibodies in diagnostics and therapeutics. However, identifying effective aptamers—and especially compatible aptamer sandwich pairs—typically requires extensive experimental screening.

Recent advances in AI-driven structural biology offer the possibility of accelerating this process by predicting nucleic acid folding and biomolecular interactions in silico. Despite their promise, the reliability and limitations of these tools for modeling flexible aptamer–protein systems remain insufficiently understood.

This project evaluates whether modern AI-based and docking-based computational tools can plausibly predict aptamer–protein interactions, and how consistent their predictions are under realistic biological and computational constraints.

Objective

The goal of this project was not to develop new machine learning models, but to systematically evaluate and compare existing computational tools for predicting aptamer–protein interactions.

The project aimed to:

- Assess how different AI-driven and physics-based tools model the same aptamer–protein systems
- Compare confidence metrics and binding scores across methods
- Identify agreement, disagreement, and failure modes between tools
- Explore whether these predictions could support early-stage aptamer screening for diagnostic applications

Skills demonstrated

- Comparative evaluation of AI-based and physics-based computational models
- Interpretation of model confidence metrics and uncertainty

- Domain-aware reasoning in structural bioinformatics
- Analysis of heterogeneous outputs (structures, docking scores, confidence maps)
- Critical assessment of AI applicability in real biological settings
- Technical communication of complex modeling results

Data

Target proteins

- SARS-CoV-2 spike protein
- Thrombin (used as a benchmark due to well-characterized aptamer binding)

Ligands

- DNA aptamers reported in prior literature

Interaction types

- Single aptamer–protein complexes
- Dual-aptamer "sandwich" complexes

Environment

- Physiological salt conditions (Na^+ ions)

Thrombin–aptamer interactions were used as a reference point to contextualize model behavior, while SARS-CoV-2 interactions served as the primary application case.

Methods & Tools

Multiple complementary computational approaches were evaluated:

AlphaFold 3

- Deep learning model for predicting biomolecular complex structures
- Used to generate 3D structures of protein–aptamer and dual-aptamer complexes
- Key confidence metrics analyzed:
    - ipTM (interface confidence)
    - pTM (overall structure confidence)
    - pLDDT (per-residue confidence)
- Notably allowed modeling of multi-component complexes, unlike most docking tools

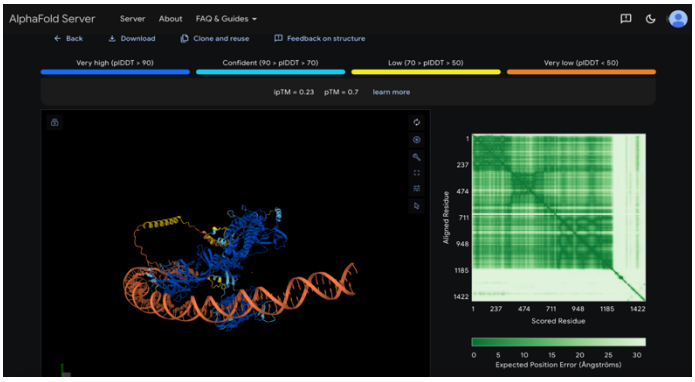


Figure 2. Results from AlphaFold 3

AptaTrans

- Deep learning pipeline for predicting aptamer–protein interaction (API) scores
- Used to rank candidate aptamers based on predicted interaction strength
- Results were compared against known benchmark interactions

PyMOL

- Structural visualization and geometric analysis
- Used to inspect predicted binding interfaces, measure residue proximity, and analyze inter-aptamer distances

HADDOCK

- Physics-based docking server
- Provided binding scores, cluster statistics, RMSD, and energetic components
- Used as an independent comparison to AI-based predictions

HADDOCK score	170.3 +/- 10.4
Cluster size	5
RMSD from the overall lowest-energy structure	14.1 +/- 0.5
Van der Waals energy	-98.8 +/- 6.4
Electrostatic energy	-286.6 +/- 42.9
Desolvation energy	40.2 +/- 4.5
Restraints violation energy	2863.0 +/- 87.9
Buried Surface Area	3046.3 +/- 199.8
Z-Score	-1.3

Figure x. Results from Haddock

RoseTTAFold2NA was also explored but could not be fully deployed due to extensive hardware and installation requirements, highlighting practical constraints of some state-of-the-art tools.

Analytical Approach

Rather than optimizing predictions from a single model, the analysis focused on **comparative evaluation**:

- Benchmarked tools using the known RE31–thrombin aptamer–protein pair
- Compared predicted binding quality across AlphaFold 3, AptaTrans, and HADDOCK
- Assessed how confidence metrics aligned—or failed to align—between methods
- Evaluated the effect of input ordering on AlphaFold 3 predictions
- Used structural inspection to contextualize numerical scores and confidence values

The emphasis was on **relative consistency and interpretability**, not absolute prediction accuracy.

Results

Substantial variability across tools

- Predictions for the same aptamer–protein pair often differed markedly between methods

- In some cases, strong binding was predicted by one tool but not supported by others

Benchmark inconsistencies

- While AlphaFold 3 and HADDOCK produced plausible predictions for the RE31–thrombin benchmark, AptaTrans yielded unexpectedly low interaction scores
- This highlighted potential limitations related to model training and generalization

Sensitivity to modeling assumptions

- AlphaFold 3 predictions were sensitive to the order in which molecules were provided as input
- Global structure confidence (pTM) was often high even when interface confidence (ipTM) remained low

Structural insights from visualization

- Residue-level inspection revealed frequent involvement of aromatic residues, particularly tyrosine, at predicted binding interfaces
- Dual-aptamer sandwich predictions for SARS-CoV-2 exhibited larger inter-aptamer distances than the thrombin benchmark, providing a plausible structural explanation for weaker predicted dual binding

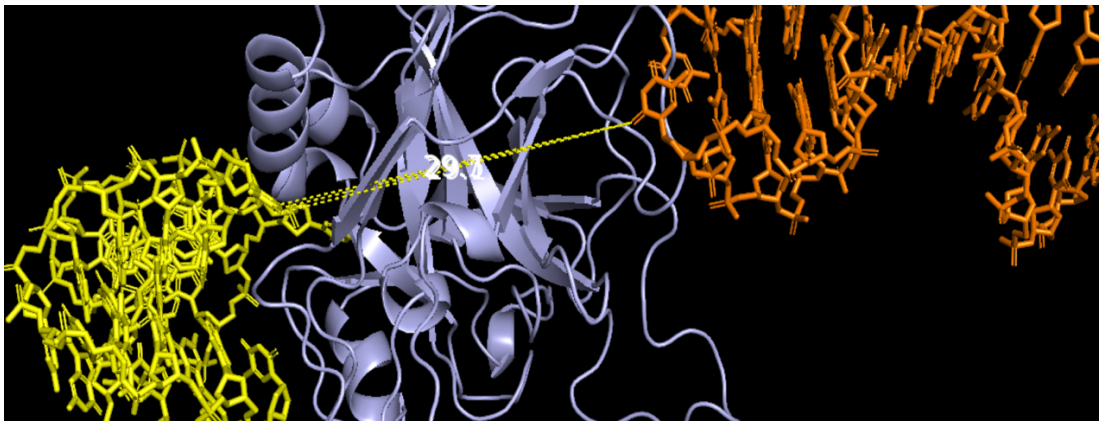


Figure 4. The closest distance between aptamers RE31 and HD22 (Thrombin).

While tools such as AlphaFold 3, AptaTrans, and HADDOCK can accelerate hypothesis generation, their predictions were often sensitive to modeling assumptions, training limitations, and molecular flexibility. Confidence metrics did not always align across methods, and high global confidence did not necessarily imply reliable interface prediction. These findings emphasize that AI-based interaction models must be interpreted comparatively and cautiously, with close attention to uncertainty, biological plausibility, and computational feasibility, rather than treated as definitive predictors.

Key takeaway - This project demonstrates the ability to critically evaluate modern AI tools in a realistic scientific context, integrating confidence metrics, structural inspection, and

biological reasoning to assess when computational predictions are informative—and when they are not. Rather than treating AI outputs as definitive answers, the work emphasizes comparative analysis, uncertainty awareness, and methodological limitations, skills that are essential in applied data science and bioinformatics research.
This project was completed in collaboration with Imose Iduozee, Matias Ylitolva, Joonatan Lindfors, Sofia Hänninen, Jaakko Pulkkinen and Linda Jalonen.