Designing Proteins Targeting PDGFR β With RFdiffusion

Elisa BILLARD & Céline HIRSCH
EPFL Life Sciences Engineering Section

***Abstract*—This project focuses on designing protein binders targeting the Platelet-Derived Growth Factor Receptor β (PDGFR-β) using advanced computational tools, including RFdiffusion, ProteinMPNN, and AlphaFold. PDGFR plays a critical role in angiogenesis and cancer progression, making it a significant therapeutic target. We evaluate various protein design methods and ultimately employ RFdiffusion for generating high-affinity binders with optimal structural properties.**

I. INTRODUCTION

PDGFR-β is a receptor tyrosine kinase involved in cellular processes such as proliferation, migration, and angiogenesis. It plays a pivotal role in tissue repair and development by mediating cellular responses to extracellular growth factor signals. Aberrant activation of PDGFR contributes to tumor growth and metastasis, making it an attractive target for protein engineering.

Protein design is a promising tool for generating new binders to inhibit PDGFR activity. Protein binders can act as inhibitors, blocking ligand-receptor interactions, and downstream signaling. Compared to small molecules, protein-based binders often exhibit higher specificity and reduced off-target effects, making them ideal candidates for targeted therapies.

This project explores multiple computational approaches to design high-affinity binders. In the end, we selected RFdiffusion because of its ability to design binders with relatively low computational cost and good results.

II. DESIGN PIPELINE

The pipeline we used combines multiple methods: RFdiffusion for structural design, ProteinMPNN for inverse folding, and AlphaFold for in silico validation.

RFdiffusion [4] leverages generative diffusion models to design proteins by iteratively sampling structural features. This method creates new scaffolds with high binding specificity and stability, making it particularly suited for designing binders against complex targets like PDGFR. It also has a relatively low computational complexity, which makes it well suited for this project and its use in GoogleColab.

For the structural design part using RFdiffusion, multiple parameters had to be adjusted:

- Binder length: This is the length (number of amino acids) of the desired binder. We tested lengths ranging from 30 aa to 70 aa.
- Number of iterations: This refers to the number of refinement steps that the model performs to optimize the generated protein structure toward the desired design objectives. For computational reasons, we could only try two values for this parameter: 25 and 50 iterations.
- Hotspots: This parameter allows to specify the key residue positions on the target protein that the binder protein should interact with, guiding the model to prioritize these regions during structure generation. It is not mandatory to specify this parameter. After visually inspecting the interaction between PDGFR and its ligand PDGF (platelet-derived growth factor), we identified multiple residues that seemed important for the interaction: X133, X243-245, and X296. We tested all of these, as well as specifying none.

The final RFdiffusion parameters that we used to generate our 10 best designs can be seen in Table I.

Binder	Length	Iterations	Hotspots
1	50	25	X133
2	50	25	-
3	50	25	-
4	50	25	-
5	50	25	-
6	50	25	-
7	50	25	-
8	50	25	-
9	50	25	X296
10	50	25	-

TABLE I: RFdiffusion parameters for top 10 designs.

Inverse folding was performed using ProteinMPNN, which generates sequence variants based on a given structural template. For this, 3 parameters had to be specified:

- Number of sequences to be generated for each protein structure. We chose to generate 4 to 8 sequences for each design.
- Sampling temperature, which controls the diversity of the generated sequences. Higher values promote more variability and lower values favor sequences closer to the optimal prediction. We chose to use the default temperature of 0.1.
- Amino acids to remove. For most of our designs, we chose to exclude cysteines, as they can form disulfide bonds under oxidizing conditions, potentially disrupting the protein's structure and stability.

III. Evaluation

The top 10 designed protein binders generated with the RFdiffusion pipeline were evaluated using AlphaFold2-based in silico validation, leveraging two orthogonal metrics to ensure structural and functional validity:

- **pAE (predicted Aligned Error)**: This metric provides insight into the confidence of the structural predictions, with lower pAE values indicating higher reliability in the predicted binder-target complex.
- **pLDDT (predicted Local Distance Difference Test)**: This metric assesses the per-residue confidence of the predicted structures, with higher pLDDT scores reflecting greater reliability and structural accuracy of the designed binders.

Binder	pLDDT [%]	pAE [Å]
1	0.90	6.74
2	0.88	7.46
3	0.84	8.46
4	0.86	8.53
5	0.85	8.57
6	0.86	8.74
7	0.83	9.65
8	0.84	10.13
9	0.86	10.21
10	0.81	10.59

TABLE II: Comparison with AlphaFold for validation.

Table II shows the metrics obtained for our top 10 designs. Additionally, visual inspection of binding interfaces using PyMOL ensured that the designs were physically plausible.

IV. Other Methods Tested

Before choosing the RFdiffusion pipeline, we tested other computational methods for protein design, but without success.

A. BindCraft

BindCraft [3] employs AlphaFold-based backpropagation to design proteins with improved binding interfaces. By iteratively optimizing the interaction energy between the target and the binder, BindCraft generates sequences tailored for binding specificity. While powerful, its computational complexity limited its application in this project. The method did not converge within our computational limit, even with very small target proteins.

B. ProteinMPNN and ColabFold

ProteinMPNN [1] generates sequence variants based on a given structural template using inverse folding techniques. ColabFold [2] predicts structural accuracy and binding interfaces for the generated sequences. This combination provides a robust framework for sequence generation and validation but not for binder design.

V. Discussion

This project highlights the potential of RFdiffusion for designing therapeutic protein binders. The combination of RFdiffusion's generative capabilities and robust validation metrics ensures the generation of high-quality designs. Despite promising results, reliance on computational tools has limitations, including computational complexity, the potential for overfitting to in silico metrics and the lack of experimental feedback. Future work will involve wet-lab assays to validate binding affinity and specificity and to refine the computational pipeline.

VI. Conclusion

This study successfully applied RFdiffusion to generate novel protein binders against PDGFR-β, validated using pAE scores and binding energy calculations. These results emphasize the efficacy of machine-learning-based design pipelines in producing therapeutic candidates with strong potential for clinical applications.

References

[1] J Dauparas et al. "Scalable, robust, and flexible generative design for protein sequences with ProteinMPNN". In: *Science* 378.6615 (2022), pp. 49–56.

[2] M Mirdita et al. "ColabFold: Making protein folding accessible to all". In: *Nature Methods* 19.6 (2022), pp. 679–682.

[3] M Pacesa et al. "BindCraft: Protein-ligand binding optimization through deep learning and backpropagation". In: *bioRxiv* (2023).

[4] JL Watson et al. "RFDiffusion: Diffusion probabilistic modeling for protein design". In: *bioRxiv* (2023).