École Polytechnique Fédérale de Lausanne (EPFL)

Solubilizing integral membrane proteins to facilitate binder screens

or how to turn hard-to-study cell-membrane proteins into soluble mimetics

Author: Aminata Touré

Program: Laidlaw Scholars Leadership and Research

Laboratory: Laboratory of Protein Design and Immunoengineering

Supervisor(s): Petra Balbi, Dr. Bruno Correia

September 30, 2025

Abstract

Integral membrane proteins are challenging targets for structural and functional studies due to their hydrophobic nature. This study aimed to redesign Somatostatin Receptor 2 into water-soluble mimetics and validate them experimentally, using a computational pipeline combining RFdiffusion and solProteinMPNN followed by bacterial expression, purification, and binding analysis. Out of eight candidates, three were successfully expressed, purified, and demonstrated significant binding to somatostatin derivatives, showing that the pipeline can produce functional mimetics and providing a framework for future design of soluble variants of other membrane proteins.

ACKNOWLEDGMENTS

I am deeply grateful to the members of the Laboratory of Protein Design and Immunoengineering, for their warm welcome and constant willingness to teach. Most of all, I sincerely thank Petra Balbi and Prof. Bruno Correia for their constant patience, guidance and support over the course of the research internship. They have inspired in me a lasting commitment to scientific research. Moreover, I am thankful to the Laidlaw Foundation for offering me such an educational and inspirational opportunity, and to Dan-Thi Nguyen for having the kindness to award me the privilege of being a Laidlaw Scholar.

Contents

ACKNOWLEDGMENTS i

1 INTRODUCTION 1

2 METHODOLOGY 3

3 RESULTS 7

4 DISCUSSION 10

5 CONCLUSION 11

List of Figures

2.1 *Computational pipeline for redesigning SSTR2 into soluble mimetics.* . . 4

2.2 *Circular map of the pPB42 plasmid (5835 bp) represented on benchling.com* 5

3.1 *Size-exclusion chromatography (SEC) profile of purified protein sample pPB42. The UV absorbance (blue trace) shows three main peaks eluting at approximately 44.9 mL, 55.7 mL, and 63.8 mL, corresponding to different oligomeric or aggregated states of the protein.* 7

3.2 *Size-exclusion chromatography (SEC) profile of purified protein sample pPB39. The UV absorbance trace (blue) shows several very small elution peaks, with main signals at 8.5 mL, 11.2 mL, and 26.2 mL. pPB39 did not express significantly* 8

3.3 *Response Unit (RU) versus time (s) for sample pPB41. It shows that pPB41 failed to significantly bind to Somatostatin derivative Octreotide.* 8

3.4 *Response Unit (RU) versus time (s) for sample pPB41. It shows that pPB45 significantly binds to Somatostatin derivative Octreotide.* 9

Chapter 1

INTRODUCTION

Background on Integral Membrane Proteins

Proteins are a highly diverse class of macromolecules that carry out an extensive array of functions and are indispensable to nearly all biological processes. Each protein's unique amino acid sequence folds into a specific three-dimensional shape, which in turn dictates its activity and function. In the organism, some proteins exist that are permanently embedded within the cell membrane, and who are referred to as integral membrane proteins. They have one or more stretches of hydrophobic amino acids that interact with the lipid bilayer hydrophobic core. Transmembrane proteins are a subclass of those, and have the particularity to span the entire lipid bilayer from one side to another, giving them hydrophobic interior portions that interact with the membrane's fatty acid tails, and hydrophilic exterior portions that project into the aqueous environment inside and outside the cell. They include receptors that pass on external signals, ion channels that control ion flow, transporters and pumps that move specific molecules, and adhesion proteins that link cells to each other and to their surroundings. Together, these proteins help keep cells stable and coordinate how they respond to their environment.

Studying transmembrane proteins is essential because they mediate fundamental processes such as signal transduction, molecular transport, and cell adhesion, and they represent major drug targets. Understanding their structure and function provides critical knowledge to apply to the development of new therapeutic strategies.

Motivation of the study

Because of their central role in cellular communication and their potential as drug targets, transmembrane proteins are prime candidates for efforts to develop protein-based binders through the screening of a pre-assembled synthetic nanobody-yeast library. Nanobodies, with their convex paratopes and long CDR3 loops, can access integral membrane proteins cryptic and concave sites effectively, making them ideal candidates for targeting integral membrane proteins [1], [2]. However, membrane proteins are espe-

cially laborious to work with experimentally, as their hydrophobic nature often leads to instability and aggregation during expression and purification. Purification of membrane proteins is usually labor intensive and often involves using detergent, that forms membrane-mimicking micelles around the protein, fusion to a highly soluble protein to increase solubility and/or introducing mutations to render the surface more hydrophilic and the core more stable. Therefore, if using an integral membrane protein as a target, one of the main limiting steps in any protein-based binder discovery pipeline is production of sufficient target protein. To overcome this problem, the LPDI has implemented a simple idea : computationally recreate those membrane proteins into hydrophilic proteins, that preserve the initial binding capacities of the protein of interest, thus bypassing the solubility problem. If confirmed as plausible, the technique could make a wide range of integral membrane proteins accessible to binder design campaigns which were previously not accessible, due to their inherent characteristics that make them challenging to work with experimentally

Aim of the study

This study aims to validate a newly implemented computational pipeline designed to redesign integral membrane proteins into water-soluble mimetics. Specifically, it seeks to experimentally express and purify the engineered mimetics, to assess their folding and structural integrity, and to determine whether they retain the capacity to bind the same ligand as the original transmembrane protein of interest. By doing so, the study will provide a test of the pipeline's potential to facilitate structural and functional studies of otherwise challenging membrane proteins.

Chapter 2

METHODOLOGY

Experimental Target: Somatostatin Receptor 2

To carry out our experiment, we devised to work with the transmembrane protein Somatostatin Receptor 2 (SSTR2). Somatostatin receptor 2 (SSTR2) is one of five somatostatin receptors and belongs to the large family of G-protein–coupled receptors (GPCRs). It has seven segments that cross the cell membrane, with an outer part that binds the hormone somatostatin and an inner part that activates signaling inside the cell. When Somatostatin-14 or a derivative binds, SSTR2 reduces the production of cAMP and changes the activity of ion channels, which lowers hormone secretion (such as insulin or growth hormone), neurotransmitter release, and cell growth. SSTR2 is found in the brain, pancreas, and digestive system, and is often present at high levels in neuroendocrine tumors [3]. Developing new binders for SSTR2 offers the possibility for finer control of hormone secretion, yet this is made difficult by the receptor's hydrophobic character, which explains why it was selected as a target in this study.

Computational pipeline

The computational pipeline relies on the principle that, for transmembrane proteins such as SSTR2, the part responsible for the binding to the ligand, and therefore the important part is the extra-cellular domain (ECD). Importantly, the ECD is hydrophilic since it is positioned outside of the lipidic bilayer and therefore does not pose an issue during expressing and purifying. As such, it is fixed, and only the hydrophobic backbone is redesigned using RFdiffusion, a diffusion-based generative model for protein design developed by the Baker Lab. It creates new 3-D protein structures in the Protein Database (PDB) format, by learning how to reverse a gradual noise-adding process[4]. From this first step, we generated 6 streamlines, which are reverse-diffusion trajectory from random noise to a final structure, that each produced 150 backbones, giving us a total of 900 backbones. Those backbones were then given as input to soluble Protein Message Passing Neural Network (solProteinMPNN), a deep-learning model from the Baker Lab that

takes a fixed 3-D protein backbone and designs amino acid sequences likely to fold into that backbone. It is particular in the sense that it has been trained on soluble proteins beforehand, which ensures it favors amino acid distributions appropriate for aqueous environments (more polar residues on the surface, hydrophobic residues inside). At this point, the pipeline yielded 3000 designs per streamline, for a total of 18 000 possible amino acids sequences. To score and filter out the candidates with the higher chance of success, we submitted the sequences to AlphaFold 2 (AF2) . It estimated for each amino acid sequence the Root-Mean-Square Deviation (RMSD), the square root of the mean squared distance in the superimposition of the candidate against the original SSTR2, and the Predicted Local Distance Difference Test score (pLDDT), a per-residue confidence score (0–100) indicating how reliable AF2 thinks the local structure is. Ultimately, the 8 designs that showed the most promise, with a plDDT superior to 80 and an RMSD lower than 3.5 Ca, were selected to go through expression and purification.

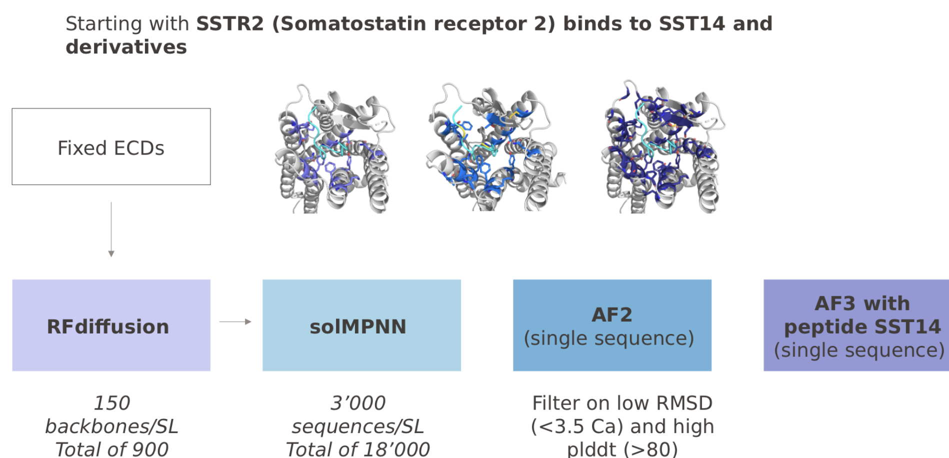

Figure 2.1: *Computational pipeline for redesigning SSTR2 into soluble mimetics.*

Experimental process

The experimental process consists of two major steps : protein expression and the subsequent purification of the obtained samples. We will from now on refer to the selected proteins with the names pPB39 up until pPB46. We began by ordering DNA inserts coding for the amino acid sequences selected, and inserting them into bacterial circular DNA molecules called plasmids through the Golden Gate Cloning process. To do so, we relied on BsaI HFv2, a restriction enzyme who cut both the plasmids and inserts at specific sites, leaving sticky ends, allowing the plasmid and insert to be recombined into a singular new plasmid that encodes our engineered protein. The obtained plasmid is then transformed into HB101 E.Coli competent cells, to amplify the quantity of recombined DNA. The HB101 strain is optimized for cloning, and after it replicates, it produces millions of copies of the assembled plasmid. Once the bacteria has multiplied, we extract

the plasmid DNA from the culture using a miniprep kit. This yields clean plasmid DNA free from most cellular contaminants that can then be sent for sequencing, to confirm that the plasmid contains the correct insert, orientation, and sequence. Investing the effort to verify the correctness of the recombined plasmid at this step ensures that no time is wrongfully invested afterwards in downstream experiments like expression, purification or binding analysis of a protein from an incorrect plasmid.

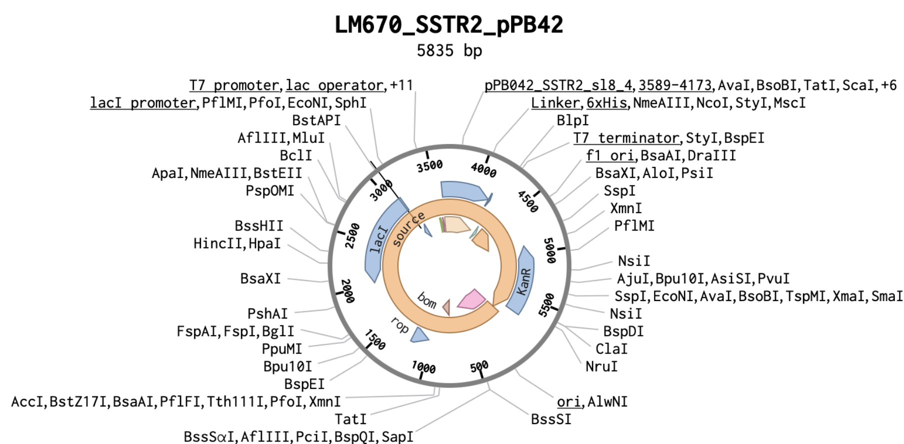

Figure 2.2: *Circular map of the pPB42 plasmid (5835 bp) represented on benchling.com*

After verification of the purified plasmid, it is transformed again, into BL21 T7 express E.Coli competent cells. The BL21 T7 strain is optimized for high-level recombinant protein expression using T7 RNA polymerase, which ensures maximum protein expression, and should provide a sufficient quantity of protein for each mimetic. The E.Coli cells carrying the recombinant plasmid are then left to incubate until Optical Density (OD), is in the range of 0.6 to 0.8, which indicates sufficient cell concentration in solution for the addition of IPTG, a stable chemical inducer that activates the transcription of genes controlled by lac/T7 promoters to produce large amounts of recombinant protein. Following an overnight growth period, the cells are finally harvested, and lysed to extract the protein mimetics. The cells undergo chemical lysis in lysis buffer (composed of lysozyme, PMSF, and DNase) as well as mechanical lysis, in the sonicator, to completely break apart the cell and retrieve the proteins of interest. The process of protein expression is therefore complete, however, at this stage, the soluble mimetics are still mixed with cell debris and various other proteins, hence the reason they must still be purified.

Protein purification is the set of techniques used to isolate one specific protein from a complex mixture of debris, other proteins and nucleic acids. In our experiment, pPB39 to pPB46 were engineered with a 6-Histidine residue called a His-tag, which enabled us to perform Immobilized Metal Affinity Chromatography (IMAC). Histidines bind strongly

to metal ions such as nickel or cobalt, which are attached to a chromatography resin. When a cell extract is passed over this resin, the His-tagged protein sticks while most other proteins flow through. The target protein can then be released by adding imidazole or changing the pH, giving a purified preparation of the protein.

The purified solution of protein is then to be subjected to a final purification, through a technique called Size Exclusion Chromatography. It separates proteins and other molecules by size using a column filled with porous beads. Large molecules cannot enter the pores and come out first, while small molecules enter the pores and come out later. This gentle method is often used as a final purification step to check a protein's size and purity, and separate the actual folded target protein from aggregate. It reveals if a protein is expressed in a significant amount, so that its binding affinity towards Somatostatin and derivatives might be qualified.

In summary, the experimental process combined a validated cloning strategy with optimized bacterial expression and a two-step purification workflow to produce samples of the engineered SSTR2 mimetics (pPB39–pPB46). This approach ensured that each construct was correctly assembled, efficiently expressed, and purified under non-denaturing conditions, resulting in protein samples suitable for downstream structural and functional analysis, including ligand-binding studies.

Downstream functional analysis

Ultimately, we obtained purified proteins that are soluble in aqueous solution; however, one key question remains : whether the newly designed proteins function as true SSTR2 mimetics, that is, whether they bind Somatostatin and its derivatives, and with what affinity. To determine whether or not the obtained proteins are functional mimetics, we use a technique called Surface Plasmon Resonance (SPR). It is a method used to measure how biomolecules interact. A ligand is fixed on a thin gold surface, and an analyte flows over it. When binding occurs, the local refractive index changes, which shifts the angle of reflected light. By tracking this shift over time, SPR provides both the rate constants of association and dissociation and the overall binding affinity (KD), allowing detailed analysis of interaction strength and specificity. In our experiment, we immobilized SST14 and Octreotide (a Somatostatin derivative) and flowed our candidates over it. The final SSTR2 mimetics were the proteins amongst which we observed a change in Response Unit (RU) that indicated binding and affinity to Somatostatin and derivatives.

Chapter 3

RESULTS

Presentation of the experiment's results

Following the computational redesign of the SSTR2 transmembrane protein, eight candidates were selected for experimental evaluation. Size-exclusion chromatography revealed that pPB39 and pPB40 were expressed at negligible levels, leaving six proteins with significant expression.

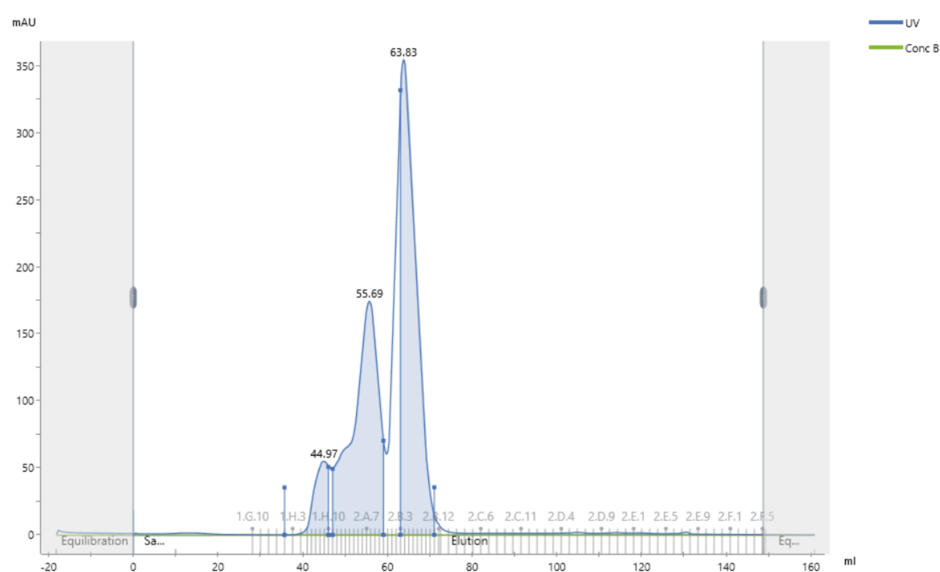

Figure 3.1: *Size-exclusion chromatography (SEC) profile of purified protein sample pPB42. The UV absorbance (blue trace) shows three main peaks eluting at approximately 44.9 mL, 55.7 mL, and 63.8 mL, corresponding to different oligomeric or aggregated states of the protein.*

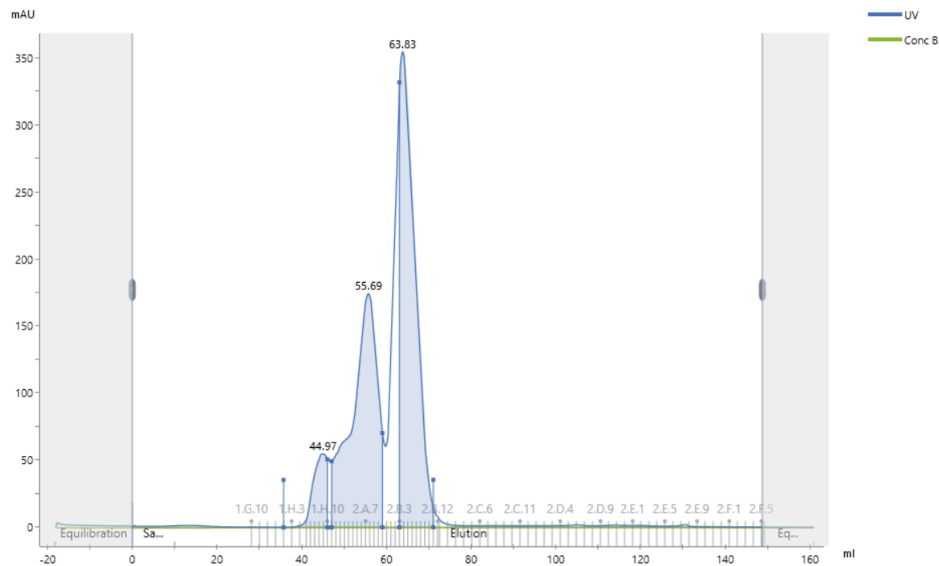

Figure 3.2: *Size-exclusion chromatography (SEC) profile of purified protein sample pPB39. The UV absorbance trace (blue) shows several very small elution peaks, with main signals at 8.5 mL, 11.2 mL, and 26.2 mL. pPB39 did not express significantly*

Among these six, only pPB42, pPB45, and pPB46 displayed substantial binding affinity toward SST14 or octreotide, identifying them as the functional mimetics produced by the computational pipeline and subsequent experimental process. (SPR pictures)

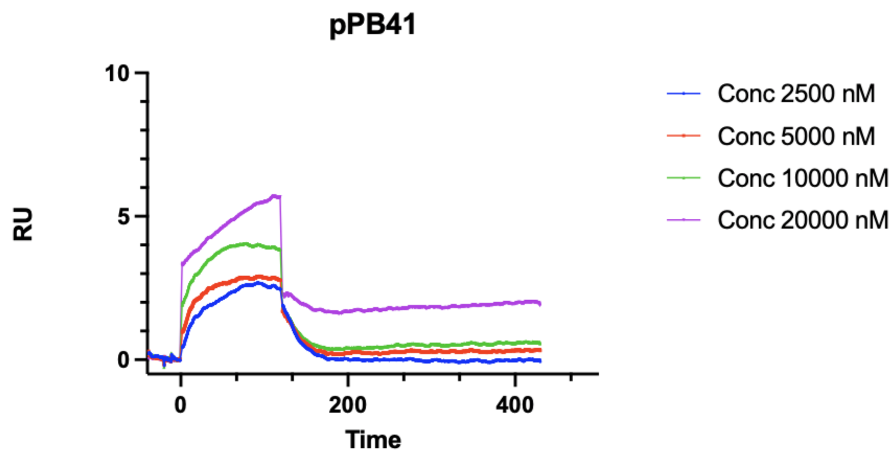

Figure 3.3: *Response Unit (RU) versus time (s) for sample pPB41. It shows that pPB41 failed to significantly bind to Somatostatin derivative Octreotide.*

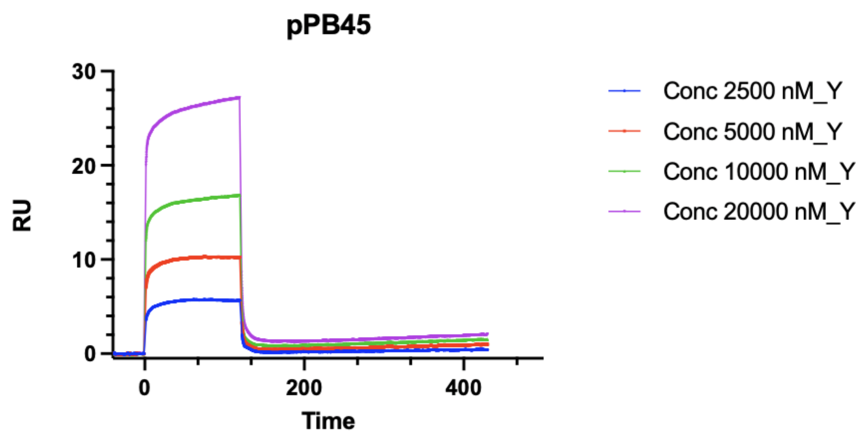

Figure 3.4: *Response Unit (RU) versus time (s) for sample pPB41. It shows that pPB45 significantly binds to Somatostatin derivative Octreotide.*

Chapter 4

DISCUSSION

Interpretation of the experiment's results

The experiment yielded three water-soluble mimetics of an integral membrane protein out of eight computationally selected candidates. This corresponds to a 3/8 success rate, which is generally consistent with typical success rates reported in comparable protein engineering studies, given the complexity of protein folding and expression. Furthermore, the performance of the computational pipeline is influenced by downstream experimental conditions and manipulations, which may also explain why only three of the eight candidates displayed the desired characteristics.

Future endeavors

In the continuity of the achieved work, two directions must be considered : firstly, optimisation of the computational pipeline, by possibly introducing better and more diverse scoring functions or additional constraints for folding. Secondly, the soluble mimetics obtained must be actually tested to conduct the screen of the selected nanobody yeast library, through a positive selection followed afterward by a negative selection to filter out nanobodies that may recognize elements other than Somatostatin. We hope to find a suitable binder to the mimetic, so that it may bind SSTR2 itself. The results of these solubilisation efforts will therefore provide us with information about the limits of solubilizing integral membrane proteins. Development of solubilisation strategies as described in the previous sections is of major interest to the pharmaceutical industry and extends our laboratory's previously reported success in solubilisation of integral membrane proteins. The technology can make a wide range of integral membrane proteins accessible to binder design campaigns which were previously not accessible, due to their inherent characteristics that make them challenging to work with experimentally.

Chapter 5

CONCLUSION

This study set out to validate a computational pipeline for redesigning the transmembrane receptor SSTR2 into water-soluble mimetics through experimental expression, purification, and binding analysis. From eight computationally selected candidates, three mimetics were successfully expressed, purified, and demonstrated measurable binding to somatostatin derivatives. These findings demonstrate that the pipeline can produce functional mimetics of an integral membrane protein and suggest that the scoring metrics correlate with experimental success. However, the limited success rate and variation in binding affinities highlight the need for further optimization of both the computational design steps and experimental conditions. Future work will focus on refining sequence prediction models, incorporating feedback from experimental testing, and extending the approach to other transmembrane targets to accelerate binder discovery

Bibliography

[1] B. Thomas, K. Chockalingam, and Z. Chen, "Methods for engineering binders to multi-pass membrane proteins," *Bioengineering*, vol. 10, no. 12, Dec. 2023. DOI: 10.3390/BIOENGINEERING10121351.

[2] K. A. Henry and C. R. MacKenzie, "Antigen recognition by single-domain antibodies: Structural latitudes and constraints," *mAbs*, vol. 10, no. 6, p. 815, Aug. 2018. DOI: 10.1080/19420862.2018.1489633.

[3] U. Kumar, "Somatostatin and somatostatin receptors in tumour biology," *International Journal of Molecular Sciences*, vol. 25, no. 1, p. 436, Dec. 2023. DOI: 10.3390/ijms25010436.

[4] J. L. Watson et al., "De novo design of protein structure and function with rf diffusion," *Nature*, vol. 620, no. 7976, pp. 1089–1100, Jul. 2023. DOI: 10.1038/s41586-023-06415-8.