Computation-aided protein engineering for targeted therapeutic delivery

Thesis by
Xiaozhe Ding

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended Feb 28, 2023

Xiaozhe Ding
ORCID: 0000-0002-0267-0791

ACKNOWLEDGEMENTS

The adventure as a grad student at Caltech will undoubtedly be a memorable piece of my life journey. By the time I write this paragraph, the Caltech campus has already become the neighborhood where I have spent the most years of my life. This school is a place that I feel I belong to — my nonstop nerdy thoughts and robot-like life habits can easily find their kind on this campus. I am grateful that there is still a protected space for people like me in this messy world, and I feel fortunate that I could spend my 20s in this lovely place.

I am grateful to my advisor, Dr. Viviana Gradinaru, who allowed me to grow with radical freedom, trust, and support. Under her protection, I had the luxury of enjoying the fun of tackling scientific challenges without worrying about anything else for years. Moreover, I had the precious experience of choosing exciting scientific problems on my own and getting comprehensive training, from proposing ideas to writing a manuscript, which gave me complete confidence to run an independent research program.

I wish to thank my rotation lab advisors and committee members, Dr. Mikhail Shapiro, Dr. James Heath, Dr. Frances Arnold, Dr. Dave Tirrell, Dr. Pamela Björkman, and Dr. Rob Phillips. They allowed me to learn about the breadth of multi-disciplinary research on this amazing campus and gave me utterly-needed feedback and technical help whenever needed.

I thank all my failed experiments. They taught me to keep moving when things do not go as well as expected. One example I wanted to cite here is a very challenging project on AAV engineering that I took on since I joined the lab. I devoted several restless years to tackling that problem and tested thousands of molecular constructs. Once for nearly a year, I thought we had made a breakthrough and got so excited about the positive results. However, only after seeing more and more puzzling control experiment results did I realize that the phenomenon I observed was partially an artifact caused by some self-propagating contamination in the lab. Although a failure from one perspective, this dramatic roller-coaster experience that spanned most of my Ph.D. career taught me more than anything I had accomplished successfully. It became an unforgettable lesson to be ruthlessly critical of my own biases and wishful thinking.

I am also thankful for researchers advancing artificial intelligence technology. The

tools they created not only have been a critical powerhouse for my doctoral research (AlphaFold by DeepMind) but have also helped me to improve the language and clarity of this thesis (rephrasing tools like Writing Assistant by Stork and Grammarly), particularly as a non-native speaker.

I want to thank my friends who have accompanied me through this journey. One reason I love Caltech is that I have met so many people with the right chemistry — we share the same joy of doing science and the same intense curiosity about everything around us. We will cherish the precious and pure friendships we made here.

My family, of course. Neither of my parents was in a STEM major, but they allowed me and provided me with the environment to explore anything I was interested in since I was a kid. They have always given me unconditional support for all my choices, even when they barely understand my research and career options. No one else in my extended family is a scientist, but they all protected and nurtured my curiosity and passion for science.

I wish to dedicate this thesis to my grandpa Yangzheng Ding (1945-2023), a farmer from a rural village in northern China, a self-taught electrician, carpenter, welder, veterinary, and beekeeper, a proud father of two college graduates, a kind person, and the ultimate protector in my childhood.

ABSTRACT

My Ph.D. projects centered on using computational structural biology tools to develop protein engineering methods for targeted therapeutic delivery, emphasizing delivering molecules to the brain. In this thesis, I focus on three main projects. First, utilizing computational structural biology techniques, I investigate the molecular mechanism that enables engineered adeno-associated viral (AAV) capsids to cross the blood-brain barrier (BBB). I develop a pipeline to model the vast and dynamic complex between engineered AAV capsids and their BBB receptors. I also apply a tool, recently developed by myself and discussed in Chapter 3, to distinguish capsids that bind to different receptors. The findings of this study can lead to novel approaches for developing chemicals and biologicals that can penetrate the human brain (Chapter 2). Second, I describe the development of Automated Pairwise Peptide-Receptor AnalysIs for Screening Engineered proteins (APPRAISE). This computational pipeline predicts the receptor binding propensity of engineered proteins based on competitive modeling and physics-grounded analysis. I show that APPRAISE is capable of distinguishing between receptor-dependent and receptor-independent adeno-associated viral vectors and ranking various engineered proteins, such as miniproteins binding to the SARS-CoV-2 spike and nanobodies binding to a G-protein-coupled receptor. A top performer in an *in silico* screening using APPRAISE was validated experimentally (Chapter 3). Third, I show an example to engineer a genetically encoded transmitter indicator (GETI), which may eventually be a cargo delivered to the brain. The GETI has a novel scaffold based on bacterial repressors, a class of transcriptional regulators that are critical for bacteria to respond to environmental chemicals. I repurposed an antibiotic-sensing repressor protein to bind a neurotransmitter, melatonin, using machine-learning-guided directed evolution. A melatonin indicator was then created by integrating the repurposed receptor with a fluorescent protein. This engineering platform may be adapted to create bio-orthogonal GETIs for various neurotransmitters (Chapter 4).

PUBLISHED CONTENT AND CONTRIBUTIONS

Ding, Xiaozhe et al. (2023). "Fast, accurate ranking of engineered proteins by receptor binding propensity using structural modeling". In: *bioRxiv*. DOI: 10.1101/2023.01.11.523680.

X.D. conceived the project, developed the software, and performed the in silico experiments. (**First and co-corresponding author.**)

Shay, Timothy F. et al. (2023). "Primate-conserved carbonic anhydrase IV and murine-restricted LY6C1 enable blood-brain barrier crossing by engineered viral vectors". In: *Science Advances* 9.16, eadg6618. DOI: 10.1126/sciadv.adg6618.

X.D. developed computational structural modeling strategies, performed APPRAISE-AAV, and participated in some experimental designs (**Co-first author**).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

The efficacy and safety of therapeutics rely not only on their capability to correct the disease-causing mechanism but also on how they're delivered (Langer, 1998). To produce a drug's intended efficacy, the drug molecules must reach their target site of action. Similarly, drug safety also depends on how specifically the molecules are distributed. This is because potent drugs are often double-edged swords, and much of the toxicity and adverse side effects are caused by healthy tissues being exposed to a drug. Thanks to its ability to promote efficacy and safety, targeted therapeutic delivery is a fundamental pillar for modern precision medicine (Manzari et al., 2021).

1.1 Targeted therapeutic delivery

Targeted delivery aims to guide drugs to their intended action site, enhance the amount of drug that reaches the target area, and decrease the amount of drug wasted on non-specific targets. To attain these goals, there are two categories of approaches: physical targeting and chemical targeting.

Physical targeting approaches

Controlling the injection route. Opting for the proper injection site and the route is the most mature physical targeting method. For instance, for slow and sustained delivery of drugs to a target tissue adjacent to the skin or muscle, intramuscular (IM) injection (administering drugs into a muscle) or subcutaneous (SC) injection (administering drugs below the skin) can be employed (Jin et al., 2015). Another example is specific delivery to the central nervous system (CNS), such as intrathecal (IT) injection or intracerebroventricular (ICV) injection. Both techniques involve injecting directly into the cerebrospinal fluid (CSF) surrounding the spinal cord and the brain (Atkinson Jr, 2017).

Injection-based targeted drug delivery provides accurate, regulated administration of drugs to specific sites, improving treatment effectiveness and limiting contact with healthy tissues. However, drawbacks include: (1) Limited target sites: injection-based delivery is restricted by target site accessibility, such as veins, muscles, or spinal fluid; (2) Invasiveness: some injection routes, especially targeting the CNS,

can be invasive, raising the risk of complications (e.g., infection, bleeding, nerve damage); (3) Low repeatability: if the target area is deep in the body or the drug get easily metabolized, repeated injections may not be possible or results inconsistent.

Magnetic or ultrasound-mediated drug targeting. Non-invasive physical modalities penetrating biological tissues, like magnetism or ultrasound, have been used to manipulate therapeutics (Owen, Pankhurst, and Stride, 2012) spatially. Both modalities enable accurate, dynamic control of drugs in deep tissues.

The core idea of magnetic delivery is to conjugate magnetic materials, usually magnetic nanoparticles, to small molecules, proteins (McBain, Yiu, and Dobson, 2008), or even probiotic bacteria (Buss et al., 2021), and to use a magnetic field to guide the particles to target tissues. On the other hand, ultrasound-mediated delivery typically uses ultrasound's penetrant mechanical energy to reversibly disrupt the local tissue structures, such as the blood-brain barrier, permitting the entrance of therapeutic molecules (Szablowski, Bar-Zion, and Shapiro, 2019).

Despite being emerging drug delivery technologies, magnetic and ultrasound-mediated drug targeting faces difficulties. One main difficulty is the technical complexity: both technologies necessitate specialized equipment and expertise, which may impede the technologies' extensive application in clinical settings. Moreover, creating magnetic particles or ultrasound contrast agents for drug delivery is a complex process still being improved. Because of these reasons, most studies involving magnetic or ultrasound-mediated drug targeting have been limited to animal models (McBain, Yiu, and Dobson, 2008).

Chemical targeting approaches

Modifying the physicochemical properties of the drug. Therapeutic molecules exhibit various physicochemical properties that influence their biodistribution profiles. Modifying these properties, including lipid solubility, surface activity, charge, and molecular weight, may can lead to improved passive diffusion to desired tissues. One notorious example is the creation of heroin, an internationally controlled opioid drug derived from morphine. Two hydroxyl groups of morphine are acetylated to increase the molecule's lipid solubility, yielding a two orders-of-magnitude increase in BBB-crossing capability (Oldendorf et al., 1972). However, such improvement in delivery efficiency is rare, as similar modifications often result in physiologically unstable compounds (Dong, 2018; Pardridge, 2012). Besides the stability issue, this strategy do not provide high specificity and can only apply to small-molecule drugs.

Engineering therapeutics for specific receptor binding. Receptor-mediated targeted drug delivery takes advantage of the specific interaction between a drug molecule and a membrane receptor protein on specific target cells or tissues. This strategy often involves conjugating drugs to a receptor-binding targeting module. This rational approach improves the specificity of drug distribution, as the molecules are preferentially delivered to cell populations that express the targeted receptor. An important advantage of this strategy is its versatility — the strategy is compatible with various therapeutic modalities, including small molecules, peptides, and biologics. Moreover, because of the clear mechanisms involved, this strategy makes the translatability of the biodistribution from animal models to humans theoretically predictable (Figure 1.1).

Receptor-mediated targeted therapeutic delivery faces several practical challenges, including the limited availability of known receptors on target tissues and promiscuous ligands binding to multiple receptors. Identifying specific binders to a receptor is currently a technically challenging and costly process that necessitates specialized knowledge and equipment.

In summary, targeted therapeutic delivery can dramatically alter how diseases are treated, enhancing both the efficacy and safety of existing treatments and reducing the dosage needed. Despite the promises, both physical and chemical approaches for targeted delivery have challenges and limitations. The following thesis will focus on our endeavor to enhance brain delivery through the receptor-mediated approach.

1.2 Targeting therapeutics to the brain through receptors

The blood-brain barrier (BBB)

The Blood-Brain Barrier (BBB) is a highly selective membrane that isolates the CNS from the bloodstream. It is formed by the tight junctions between the endothelial cells that line the blood vessels in the brain, which restrict the free diffusion of many substances from the bloodstream into the CNS. Will the BBB preserves the equilibrium of the CNS and shields the brain from hazardous substances in the bloodstream (Sweeney et al., 2019; Profaci et al., 2020), the protective barrier also keeps most therapeutic medications from penetrating the brain (Banks, 2016; Sweeney et al., 2019; Zhao and Zlokovic, 2020; Terstappen et al., 2021a).

The BBB is not completely impenetrable, and specific membrane proteins on the BBB can facilitate the passage of specific substances into the CNS. These so-

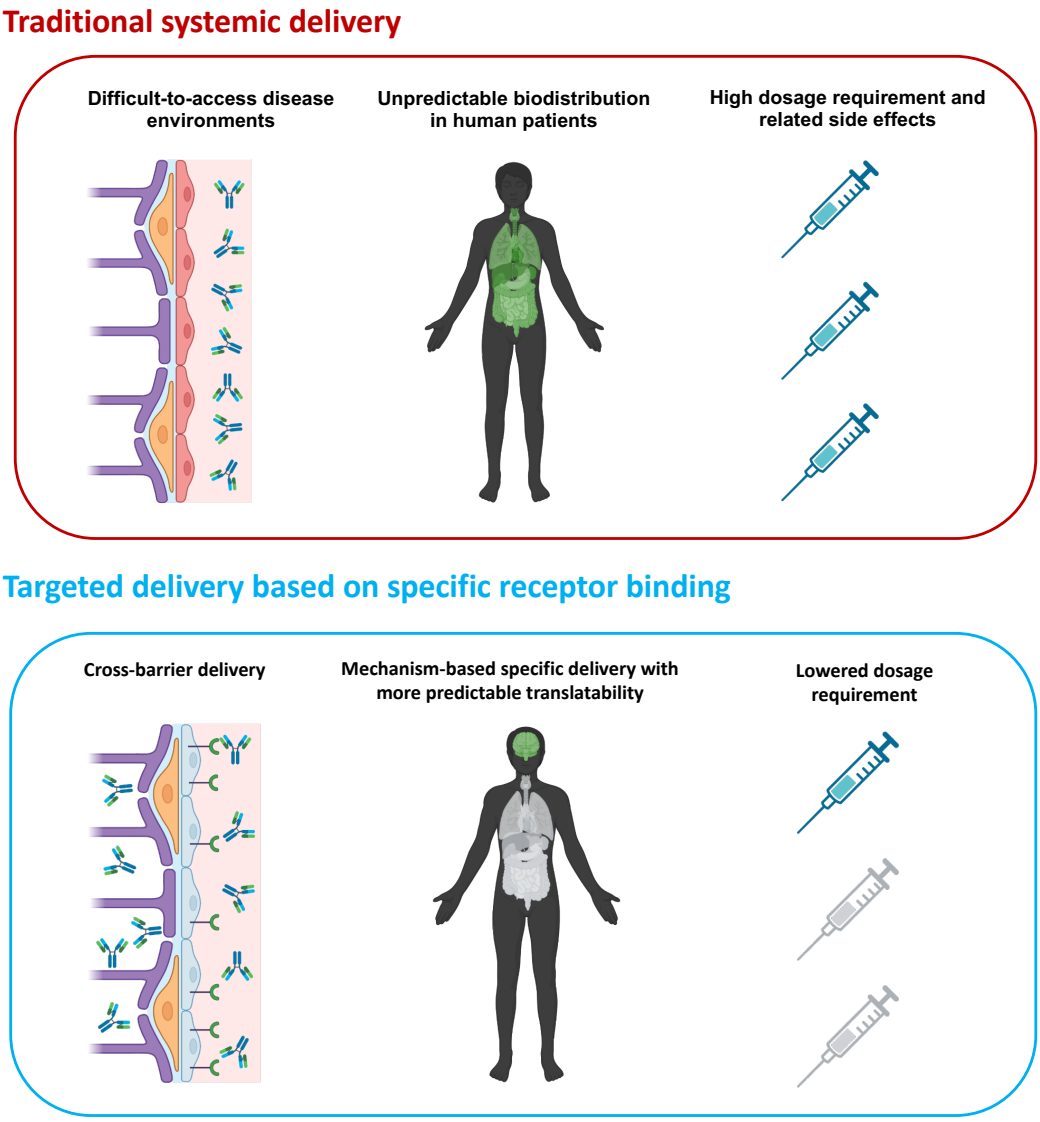


Figure 1.1: **Benefits of targeted delivery based on specific receptor binding** Systemic drug delivery faces several major challenges, including inaccessibility to the disease environment, inconsistent biodistribution in human patients, and the need for high and risky doses. Receptor-mediated targeted delivery is a promising approach to address these challenges.

called BBB receptors are paramount in allowing selective drug uptake into the CNS (Jones and Shusta, 2007; Watts and Dennis, 2013; Pulgar, 2019; Terstappen et al., 2021b). The most well-characterized BBB receptors include transferrin receptors, insulin receptors, low-density lipoprotein receptors, and solute carrier family proteins(Fishman et al., 1987; Zuchero et al., 2016; Zhang et al., 2020; Terstappen et al., 2021b).

Receptor-mediated transcytosis (RMT)

Binding between a ligand and a BBB receptor may trigger transcytosis across the barrier, and such receptor-mediated transcytosis (RMT) is a popular approach to transport drugs into the CNS. In this process, drugs are often conjugated to specific receptor-binding ligands, and the complex thus formed is then taken up by the BBB and transported beyond the barrier (Jones and Shusta, 2007; Terstappen et al., 2021a). The success of RMT-based delivery strategies depends heavily on the availability and specificity of BBB receptor proteins.

Transferrin receptor (TfR). TfR has been extensively studied and is the most commonly used BBB receptor (Jefferies et al., 1984; Fishman et al., 1987; Friden et al., 1991; Roberts, Fine, and Sandra, 1993; Yu et al., 2011; Couch et al., 2013; Yu et al., 2014; Johnsen et al., 2019; Logan et al., 2021). TfR is a widely expressed on the BBB with a natural function to transport transferrin, an iron-binding protein, into the CNS. Studies utilizing TfR as a model system to achieve RMT have yielded numerous insights, for example: 1) finetuning TfR binding affinity can improve brain uptake and peripheral exposure of antibodies (Yu et al., 2011; Couch et al., 2013), 2) monovalent binding is more effective than bivalent binding in inducing RMT of an antibody (Niewoehner et al., 2014), and 3) binding to a specific epitope on the receptor (e.g., the apical lobe of TfR) is essential to initiate RMT (Niewoehner et al., 2014; Terstappen et al., 2021b). Some of these insights may generally apply to other BBB receptors.

CD98 heavy chain (CD98hc). CD98 heavy chain (CD98hc), a solute carrier family transporter that exchanges amino acids between the cytoplasm and the extracellular fluid, was recently identified as a BBB receptor (Zuchero et al., 2016). Too prove CD98hc's BBB receptor function, bispecific antibodies targeting both $A\beta$ and CD98hc were developed, and these antibodies indeed show a significantly boosted brain accumulation and a robust pharmacodynamic response following intravenous administration (Zuchero et al., 2016).

Interestingly, latest studies revealed a striking difference in the pharmacokinetics of antibodies targeting CD98hc v.s. those targeting TfR receptors (Chew et al. and Lucas et al., presentations at Drug Delivery to the Brain Conference 2023). First, CD98hc-binding antibodies have a steady and prolonged presence in the brain, while TfR-binding antibodies have a fast yet short brain distribution. Second, in contrast to TfR binders, CD98hc-binding antibodies do not cause down-regulate the receptor's membrane level, which explains the antibodies' more steady pharmacokinetic

profiles. Third, unlike TfR, which is broadly expressed in the peripheral organs, the off-target distribution for CD98hc-targeting antibodies is mainly restricted to the spleen. These differences between antibodies targeting the two receptors suggest that there is no single solution for all in the quest for receptor-mediated CNS delivery, and more BBB receptor options will serve patients with diverse biological backgrounds.

Discovery of novel receptors

The demand for diverse BBB receptors has incentivized research campaigns to discover novel receptors. These campaigns sped up in recent years thanks to the advancement in molecular biology, cell biology, and bioinformatic techniques. These techniques are used in combination following two different strategies.

Forward engineering strategy. Forward engineering strategy studies brain endothelial cells to identify potential BBB receptors. A typical forward engineering campaign was demonstrated in the aforementioned CD98hc's case (Zuchero et al., 2016). Such campaign begins with extensive transcriptomic and proteomic analysis of brain endothelial cells and control tissues. Bioinformatics is then employed to analyze large and noisy multi-omics datasets and look for promising candidate receptors with high expression in brain endothelial cells, low expression in peripheral tissues, and steady expression on the apical side (but not the basolateral side) of endothelial cells. Finally, ligands against the putative receptors are created to validate these receptors' function in transporting cargo across the BBB.

Reverse engineering strategy. BBB receptors can also be identified through a "reverse engineering" strategy – hypothetical receptors were first postulated based on the membrane proteins' interactions with a known BBB-crossing cargo, such as endogenous transferrin and insulin (Jefferies et al., 1984; Pardridge et al., 1995). Such "reverse engineering" has become more feasible in recent years thanks to the availability of many more engineered BBB-crossing molecules. For example, directed evolution techniques have enabled developing a variety of brain-targeting AAV capsids (Choudhury et al., 2016; Deverman et al., 2016; Ojala et al., 2017; Hudry et al., 2018; Hanlon et al., 2019; Davidsson et al., 2019; Ravindra Kumar et al., 2020; Weinmann et al., 2020; Nonnenmacher et al., 2021; Challis et al., 2022), which are enriched in the CNS after systemic dosing. To identify the BBB receptors underlying the brain transduction of these AAV capsids, *in vitro*, high-throughput screening can be conducted using either cultured cells or organoids, and the BBB-crossing

protein's cellular uptake level is measured in these assays. To confirm whether the cellular uptake is dependent on a candidate receptor, genetic perturbation techniques such as over-expression, RNA interference, or CRISPR/Cas9-mediated gene editing can be used to alter the expression level of the candidate. Recently, some commercial services emerged to facilitate such high-throughput screenings. For instance, Retrogenix provides a chip-based platform that can detect the binding between a protein and thousands of putative membrane receptors expressed in cultured cells.

1.3 Structure-guided protein engineering techniques for receptor targeting

As a protein's structure dictates its function (A.Petsko and Ringe, 2003), structural knowledge is valuable in guiding the engineering of therapeutic proteins like IgG antibodies, antibody-like proteins, or viral vectors for receptor-mediated targeted delivery.

Common scaffolds to engineer for receptor targeting

IgG antibodies. Traditional engineering of IgG antibody (Diskin et al., 2011; Barnes et al., 2020) has focused on the complementarity-determining regions (CDRs), particularly the CDR3, as this region is essential for the precise antigen recognition of IgG molecules. This CDR engineering alone, however, is not enough to create IgG antibodies with receptor-mediated targeting. The latter requires the molecule to bind to multiple different proteins, including the therapeutic target and the tropism-determining receptor, with specificity (Watts and Dennis, 2013). Innovative designs, such as 'knobs-into-holes' engineering, generate this bi-specificity by allowing the controlled heterodimerization between two variable fragments (Ridgway, Presta, and Carter, 1996; Merchant et al., 1998). Other designs focus on adding targeting capability to the Fc domain, such as appending an additional single-chain variable fragment (scFv) or single-domain antibodies (nanobodies or VHHs) to the C-terminal end of the Fc domain (Brinkmann and Kontermann, 2017). Another creative way to engineer an Fc domain is to substitute particular surface exposed loops in the domain (Kariolis et al., 2020) with receptor-binding sequences. One example is the so-called Fc-based transport vehicle (TV), which binds to a BBB receptor and induces RMT to transport IgG or other cargo across the BBB (Kariolis et al., 2020).

Adeno-associated viral (AAV) capsids. AAV is a non-pathogenic virus widely used as a gene therapy delivery vector. Its safety, effectiveness, and broad and engineerable tropism have made recombinant AAV a promising agent for treating

various genetic disorders (Bedbrook, Deverman, and Gradinaru, 2018; Wang, Tai, and Gao, 2019; Challis et al., 2022). The surface of an AAV capsid contains several variable regions (VRs) that are not conserved between serotypes, making them ideal targets for engineering efforts(Agbandje-McKenna and Kleinschmidt, 2011). Due to their structural flexibility and proximity to receptor binding sites, VR VIII and VR IV are most frequently used VRs for peptide display libraries (Choudhury et al., 2016; Deverman et al., 2016; Ojala et al., 2017; Hudry et al., 2018; Hanlon et al., 2019; Davidsson et al., 2019; Ravindra Kumar et al., 2020; Weinmann et al., 2020; Nonnenmacher et al., 2021; Bedbrook, Deverman, and Gradinaru, 2018; Wang, Tai, and Gao, 2019; Challis et al., 2022).

Directed evolution for protein engineering

Directed evolution combines genetic and biochemical methods to create mutations in a targeted protein and then chooses the variants with the desired characteristics via selection or screening procedures (Romero and Arnold, 2009). Through directed evolution, protein engineers have been able to modify proteins, such as AAV capsids and antibodies, to enhance their desired characteristics, like receptor binding.

In vivo selection methods have been highly successful in targeting AAV vectors to desired tissues, particularly the brain, as demonstrated by several studies (Choudhury et al., 2016; Deverman et al., 2016; Ojala et al., 2017; Hudry et al., 2018; Hanlon et al., 2019; Davidsson et al., 2019; Ravindra Kumar et al., 2020; Weinmann et al., 2020; Nonnenmacher et al., 2021; Challis et al., 2022). Additionally, *in vitro* receptor-based selection methods have recently been developed (Huang et al., 2022) to find BBB receptor binders and the resulting capsids were shown to cross the BBB *in vivo*.

Directed evolution has been extensively utilized to engineer antibodies and antibody-like proteins for specific targeting. The adaptive immune system is sophisticated for performing "*in vivo* directed evolution" to generate therapeutic agents against invading pathogens. In addition to such *in vivo* directed evolution, *in vitro* directed evolution has been employed to select specific receptor binders, for instance, Fc fragments targeting TfR (Kariolis et al., 2020).

Despite its successes, directed evolution is a time-consuming and costly, especially for in vivo selections. It should be complemented with rational design and insights from mechanism studies, such as structural biology studies and computational modeling, to make the process more efficient and cost-effective.

Computational tools for protein structure prediction

Computational structure prediction, popularly known as the protein folding problem, predicts the three-dimensional protein structures based on the primary sequences, i.e., finds out the P(structure|sequence) distribution. This concept has a long past, beginning in the early days of molecular biology and computer science. In recent years, computational protein structure prediction accuracy has been significantly enhanced due to the introduction of deep learning-based methods and the accumulation of high-quality experimental structural data over the decades. As a result, such computational methods have become critical supplementary tools for protein engineers in addition to experimental structural biological methods. This section will briefly overview the most influential tools made available.

Rosetta. One of the first widely adopted protein structure-prediction tools was the Rosetta software suite (Simons et al., 1999; Leaver-Fay et al., 2011). The key idea behind Rosetta was to search the structural space using fragment recombination and evaluates each candidate pose using an "energy function" based on physical principles and statistical knowledge. Due to its unique physics-based features, to date Rosetta is still a popular tool used for many purposes, such as analyzing interface energy or relaxing local structure.

AlphaFold1 and trRosetta. Protein structure prediction came to a new era by adopting deep learning techniques and coevolutionary information from multiple sequence alignments (MSAs). The key idea behind using MSAs was that protein contact is evolutionarily conserved, and the contacting residues tend to co-evolve together (Yanofsky, Horn, and Thorpe, 1964). Therefore, a "contact map" of a target protein can be inferred from MSAs by measuring coevolution. Such contact maps can then be used to interpret inter-residue interactions and resulting tertiary structures. One successful example was AlphaFold1 from DeepMind, a neural network-based structure-prediction method that uses MSAs as a major input Senior et al., 2020, which topped in the blinded competition CASP13 (2018) (Kryshtafovych et al., 2019). A similar approach, trRosetta, further improved the prediction precision (Yang et al., 2020).

AlphaFold2 and the latest MSA-based methods. The release of AlphaFold2 is a revolutionary landmark in the field of protein structure prediction. AlphaFold2 uses a novel architecture that integrates both coevolutionary information from MSAs and geometrical/physical information from solved templates (Jumper et al., 2021). The model won the CASP14 (2020) competition with dominant scores by predicting

single-chain protein structures with near-experimental accuracy (Kryshtafovych et al., 2021). At present, models with AlphaFold2-like architectures (Baek et al., 2021; Evans et al., 2021; Mirdita et al., 2022; Aderinwale, Christoffer, and Kihara, 2022; Motmaen et al., 2022; Ruffolo et al., 2022; Li et al., 2022; Wu et al., 2022; Wang et al., 2022a; Cheng et al., 2022; Liu et al., 2022) still the most accurate structure-prediction tools available.

ESMFold and protein language model-based methods. In addition to AlphaFold2-like architectures, another emerging class of tools first "embed" protein sequences as vectors using pre-trained protein language models and subsequently use the lower-dimensional vectors to predict the structure. Examples in this class include ESMFold (Lin et al., 2022), RGN2(Chowdhury et al., 2022) or trRosettaX-Single (Wang, Peng, and Yang, 2022), which have achieved comparable accuracy to that of AlphaFold2 in proteins with small sizes, with superior performance in "orphan" proteins with few homologous sequences. In addition to the improved performance in orphan proteins, another main advantage of these language model-based prediction tools is their fast inference speed, which is often orders of magnitude higher than AlphaFold2-like, MSA-based methods.

Other neural networks for specialized prediction tasks. Some recent structure-prediction tools have been tailored for specific targets such as antibody monomers (Ruffolo et al., 2022) or MHC-peptide complexes (Motmaen et al., 2022). Other tools were built to enhance the functionality of AlphaFold2. For example, AlphaFill can transfer small-molecule ligands and ions from experimentally-solved structures to AlphaFold2-predicted apoprotein structures (Hekkelman et al., 2022).

In summary, computational structural modeling has seen significant progress over the years, and modern tools such as AlphaFold and ESMfold provide novel insights into protein structures that may accelerate the development of new drugs and therapies. However, the current capabilities of the structure-prediction tools are limited, particularly when dealing with large complexes, dynamic complexes, protein-ligand complexes, post-translational modifications, and lipids surrounding membrane proteins (Phillips et al., 2009), etc. These missing yet critical functions call for further method development.

Computational tools for protein design

Protein design finds coding sequences for proteins with the desired structure or function. To date, it is still challenging to directly design proteins with specific

functions, although promises start to emerge with the progress in large protein language models (Madani et al., 2023). In comparison, designing protein structures is a much more achievable goal thanks to the breakthroughs in its inverse problem, i.e., protein structure prediction, which I just discussed above. The overarching goal of protein structure design is to find primary amino acid sequences that fold to a given structure, i.e., to figure out the P(sequence|structure) distribution of proteins.

Rosetta. In addition to the aforementioned structure prediction functions, the Rosetta software suite (Simons et al., 1999; Leaver-Fay et al., 2011) also features protein design functions. The software generates protein sequences that fold into a desired structure through a Monte Carlo search that explores the sequence space using a physics-and-statistics-based energy function. Numerous landmark *de novo* proteins were designed using Rosetta, with examples including TOP7(Kuhlman et al., 2003), the first protein with an artificially designed protein fold, and I53-50(Bale et al., 2016), an ultrastable self-assembling protein nanocage used for antigen display.

Hallucination and inpainting. RoseTTAFold Design, or RFdesign, takes advantage of the well-established architecture of protein prediction neural networks like tr-Rosetta (Yang et al., 2020) or RoseTTAFold (Baek et al., 2021) and inverse it for the protein design problem (Wang et al., 2022b). Importantly, RFdesign provides two methods, "hallucination" and "inpainting", that are especially useful when a design task demands the preservation of functional sites required for protein-protein interaction or enzymatic activity. In protein design, "hallucination" refers to generating novel protein backbones using protein structure prediction networks by iteratively introducing mutations and accepting mutations that sharpen the contact map (Anishchenko et al., 2021). Built upon this basic idea, the RFdesign-Hallucination method was equipped with customized loss functions that ensure protein backbones include a desired functional motif. The second method, RFDesign-Inpainting, uses a retrained structure prediction network that can "restore" both the masked sequence and the entire protein structure when only the active site of a protein is known (Wang et al., 2022b).

Diffusion. RosettaFold Diffusion (RFdiffusion) (Watson et al., 2022) is a general protein design framework inspired by denoising diffusion probabilistic models (DDPMs). DDPMs have recently succeeded in generative image models (e.g., DALL-E 2) that can create realistic pictures by "denoising". Like image-generating DDPMs, RFdiffusion employs the RoseTTAFold structure prediction network to conduct a "denoising" task on protein structure, yielding highly "realistic" protein

backbones suitable for downstream sequence design. According to Watson et al., 2022, RFdiffusion outperformed hallucination and inpainting in all protein design tasks examined and successfully created *de novo* protein binders, new scaffolds with functional motifs, and symmetrical proteins.

ProteinMPNN. Whereas hallucination, inpainting, and diffusion focus on the issue of protein backbone generation, ProteinMPNN uses a novel graph neural network-based architecture to address the problem of designing the optimal amino acid sequence based on a fixed backbone (Dauparas et al., 2022). When generating sequences for backbones from nature proteins, ProteinMPNN has a much higher sequence recovery rate of 52.4 % compared to 32.9 % for Rosetta. ProteinMPNN can also improve the solubility and expression level of previously failed designs created with Rosetta or Alphafold (Dauparas et al., 2022). ProteinMPNN has frequently been used downstream of other design methods that generate initial backbones to improve overall design success rates.

1.4 Utilizing the computational protein engineering methods for targeted delivery

Although many powerful methods for computational protein engineering have become available in recent years, these methods have been under-utilized in research efforts for targeted delivery due to a lack of customized tools. This thesis aims to bridge this gap by developing and applying tools and pipelines based on state-of-the-art computational tools to facilitate the development of targeted therapeutic delivery systems like receptor-targeting Adeno-associated viral (AAV) vectors.

Chapter 2

UNDERSTANDING MOLECULAR MECHANISM UNDERLYING BBB-CROSSING AAV CAPSIDS

2.1 Abstract

The blood-brain barrier (BBB) presents a major challenge to delivering large molecules to study and treat the central nervous system (CNS). This is due in part to the scarcity of effective targets for BBB crossing, the identification of which is the crucial first step of drug development. Here, we leveraged a panel of adeno-associated viruses (AAVs) previously identified through directed evolution for improved BBB transport to reverse engineer protein targets for enhanced BBB crossing. We identify both murine-restricted LY6C1 and primate-conserved carbonic anhydrase IV (CA-IV; CA4) as receptors for crossing the BBB. We demonstrate how these receptors can unlock experimental and computational target-focused engineering strategies by creating the enhanced LY6C1-binding vector AAV-PHP.eC and by applying AlphaFold-enabled in silico methods to rank capsids against identified receptors and generate capsid-receptor binding models. The identification of CA-IV and structural insights from computational modeling enable receptor-targeted paths toward efficient human brain-penetrant chemicals (drugs) and biologicals (including gene delivery).

2.2 Introduction

The blood-brain barrier (BBB) presents a fundamental bottleneck to the development of effective research tools and therapeutics for the central nervous system (CNS) (Banks, 2016; Sweeney et al., 2019; Zhao and Zlokovic, 2020). This complex structure, comprised mainly of brain endothelial cells (Sweeney et al., 2019; Profaci et al., 2020), requires large molecules to be delivered via invasive intracranial injections, technically challenging focused ultrasound (Szablowski, Bar-Zion, and Shapiro, 2019), or receptor-mediated transcytosis (Watts and Dennis, 2013; Pulgar, 2019). Rational design of BBB-crossing large molecules has long been hampered by our imperfect understanding of the mechanisms involved in transcytosis, with only a handful of targets, such as transferrin receptor (Friden et al., 1991; Yu et al., 2011; Johnsen et al., 2019; Logan et al., 2021), validated for research and therapies (Pulgar, 2019; Terstappen et al., 2021b; Cho et al., 2017; Bergmann et al., 2018).

Directed evolution is a powerful method for generating biomolecules with enhanced fitness for desired properties despite an incomplete understanding of the underlying biological systems (Romero and Arnold, 2009). Importantly, the outcomes of directed evolution libraries could in turn be used to unlock previously unknown biology by probing the mechanism of action for molecules with evolved properties. We have decided to apply this paradigm of reverse-engineering directed evolution hits to the accumulating wealth of data resulting from selective pressure on adeno-associated virus (AAV) libraries for CNS enrichment after systemic administration (Choudhury et al., 2016; Deverman et al., 2016; Ojala et al., 2017; Hudry et al., 2018; Hanlon et al., 2019; Davidsson et al., 2019; Ravindra Kumar et al., 2020; Weinmann et al., 2020; Nonnenmacher et al., 2021; Challis et al., 2022).

One such improved rodent BBB-crossing AAV capsid is PHP.eB (Chan et al., 2017) which we previously identified by Cre-recombination-based AAV targeted evolution (CREATE) method (Deverman et al., 2016) on the parent capsid AAV9 (Gao et al., 2004; Foust et al., 2009). Following systemic injection in genetically divergent mouse strains mice, capsids such as PHP.eB can either show potent CNS tropism (as in C57BL/6J, FVB/NCrl, and DBA/2J) or akin to AAV9 (as in BALB/cJ) (Hordeaux et al., 2018; Matsuzaki et al., 2019; Challis et al., 2019). This is explained by the receptor for PHP.eB, the GPI-anchored protein Ly6a, being strongly expressed in C57BL/6J (the directed evolution selection host), FVB/NCrl and other mouse strains but non-functional in the BALB/cJ strain and other mammals (**Hordeaux**; Batista et al., 2020; Huang et al., 2019). Consistently, in non-human primates (NHPs), PHP.eB and others display AAV9-like CNS infectivity (Hordeaux et al., 2018; Liguore et al., 2019; Matsuzaki et al., 2018; Goertsen et al., 2022a). As AAVs have become the vector of choice for human gene therapies, including for therapies of the CNS (Samulski and Muzyczka, 2014; Kuzmin et al., 2021; Burdett and Nuseibeh, 2022), complications from directly applying mouse-evolved AAVs in diverse genetic backgrounds contributed to a shift toward performing AAV directed evolution in NHPs. In so doing, researchers hope to increase the likelihood of identifying AAVs whose enhanced tropism will translate to humans. As both the pre-clinical validation and, increasingly, the generation of engineered capsids occurs in NHPs however (Tabebordbar et al., 2021; Goertsen et al., 2022a; Chen et al., 2022a), the animals' scarcity (Tian, 2021; Subbaraman, 2021) and costs slow the identification of engineered capsids while the risk of NHP-specific AAVs entering clinical trials remains. Nevertheless, examples are beginning to accumulate of capsids that can cross both rodent and NHP BBB (Goertsen et al., 2022a) but, of utter importance,

also of at least one capsid reported to cross the macaque but not the rodent BBB (Chen et al., 2022a). Collectively, this diverse set of capsids engineered over the past decade by many groups represents, through their yet unexplored mechanisms of action, an unprecedented opportunity to start unraveling new targets for crossing the BBB across strains and species.

As patients treated with any AAV are likely to develop neutralizing antibodies toward most future AAVs and preclude them from future AAV treatments (Samulski and Muzyczka, 2014; Kuzmin et al., 2021; Burdett and Nuseibeh, 2022), methods to efficiently identify targets of brain-enhanced AAVs are critically necessary. Thus, NHP-optimized and validated AAV capsids that might not cross the human BBB are a concerning possibility. Here, we demonstrate a path forward by identifying protein targets or BBB-crossing mechanisms that may directly translate to human drug development. Using brain-enhanced engineered AAVs previously identified by directed evolution selections in mice (Hanlon et al., 2019; Ravindra Kumar et al., 2020; Nonnenmacher et al., 2021; Goertsen et al., 2022a), we validate a pipeline to reverse engineer targets for potent BBB crossing. Focusing on engineered AAVs whose enhanced CNS infectivity upon systemic injection in mice is conserved across both C57BL/6J and BALB/cJ strains (Hanlon et al., 2019; Ravindra Kumar et al., 2020; Nonnenmacher et al., 2021), we identified using an in vitro screening model with in vivo validation two novel receptors for enhanced BBB crossing by engineered AAVs: murine-restricted LY6C1 and primate-conserved carbonic anhydrase IV (Car4; CA4).

To demonstrate how a known target can unlock new engineering strategies, (1) we created an enhanced LY6C1-binding AAV variant, AAV-PHP.eC, (2) we utilized a new method for Automated Pairwise Peptide-Receptor Analysis for Screening Engineered AAVs (APPRAISE-AAV for short) that uses AlphaFold-Multimer(Jumper et al., 2021) to screen peptides against potential receptors in silico and (3) we generated AAV-receptor interaction models, including the first high-resolution binding model of PHP.eB with LY6A. Our experimental and computational pipeline for learning how evolved AAVs enact their enhanced BBB crossing tropisms demonstrates target-driven capsid engineering with LY6C1 and establishes in vivo that primate-conserved carbonic anhydrase IV is a receptor target for enhanced CNS transduction by AAVs. In this thesis, I will report and discuss my contributions to this work, particularly points (2) and (3).

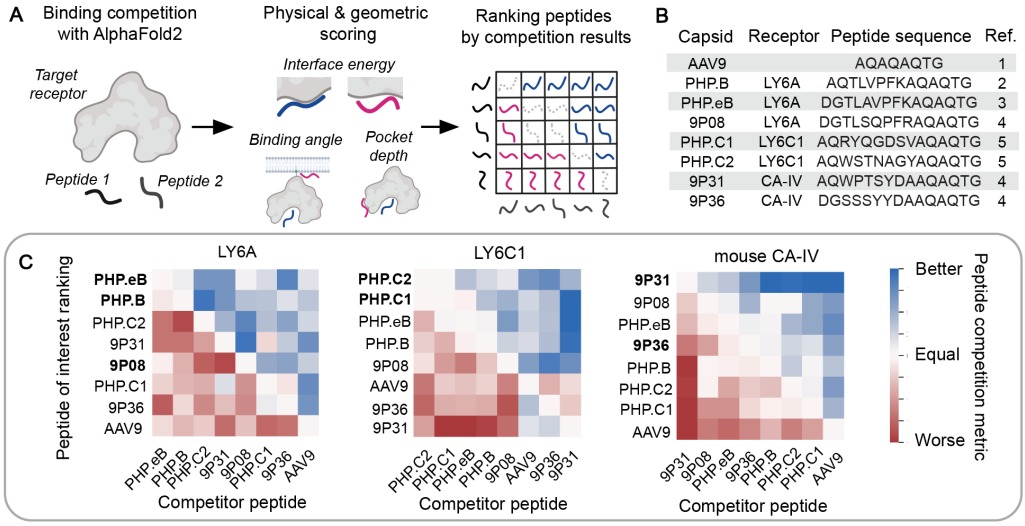


Figure 2.1: ***In silico*** **ranking of AAV variants by their receptor-binding propensities.** **A,** Overview of AlphaFold-based in silico Automated Pairwise Peptide-Receptor Analysis for Screening Engineered AAVs (APPRAISE-AAV for short) (Ding et al., 2023) Surface peptides from AAV variants are put in pairwise binding competition using AlphaFold-Multimer. A peptide competition metric is calculated according to each peptide's interface energy, binding angle, and pocket depth (see Materials and Methods section for details) before being assembled into broader ranked matrices of interaction likelihood. Competition results reflect the relative peptide binding probability encoded in the AlphaFold neural network. **B,** Table of engineered AAV capsids, their confirmed receptor, and the capsid peptide sequence used in APPRAISE-AAV. References: 1(Gao et al., 2004), 2(Deverman et al., 2016), 3(Chan et al., 2017), 4(Nonnenmacher et al., 2021), 5(Ravindra Kumar et al., 2020) **C,** Matrices ranking AAV peptides by their average competition metric over ten replicate conditions for LY6A, LY6C1, and mouse CA-IV. AAV peptide labels in bold indicate those experimentally identified to interact with the corresponding receptor. Metric values out of range (-100 - 100) were capped to range limits.

2.3 Results

Identifying receptor-binding capsids using structure modeling

Having identified a panel of receptor and AAV capsid pairings using a reverse engineering strategy (see the full version of the receptor identification story in Shay et al., 2023), we aimed to see if we could capitalize on rapid advances in protein structure prediction to generate binding poses for engineered AAVs and their newly identified receptors.

We began by applying an AlphaFold2-based computational method (Evans et al., 2021; Jumper et al., 2021) for Automated Pairwise Peptide Receptor Analysis for Screening Engineered AAV (APPRAISE-AAV for short) (Ding et al., 2023).

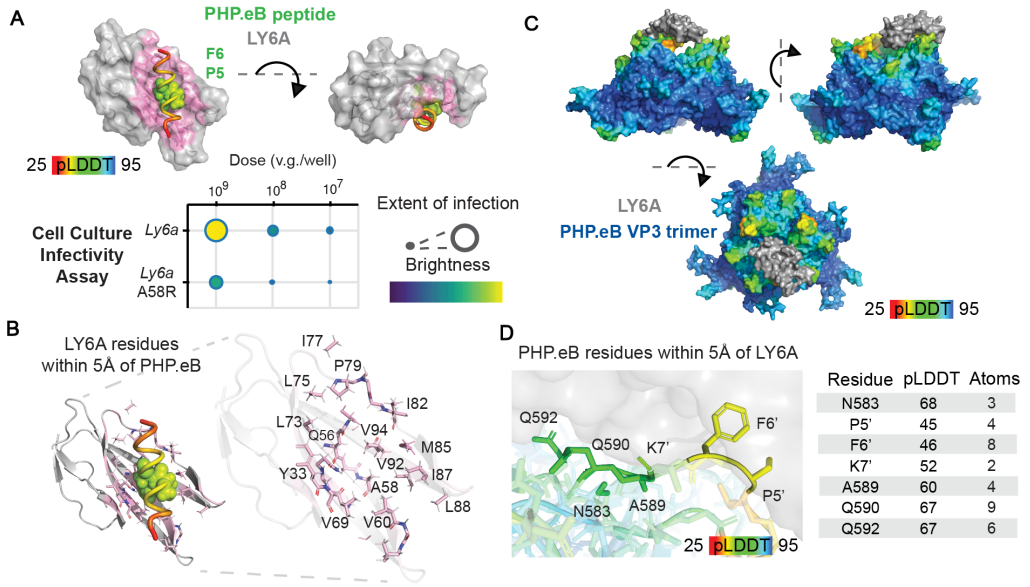


Figure 2.2: **Engineered AAV interactions with LY6A. A,** AlphaFold-Multimer-predicted LY6A-PHP.eB peptide complex structure. PHP.eB peptide is colored by pLDDT (predicted Local Distance Difference Test) score, a per-residue estimate of the model confidence. The highest confidence side chains, P5' and F6', are shown as spheres. LY6A A58R mutation, chosen to disrupt the predicted peptide interaction, resulted in reduced potency in the cell culture infectivity assay. Extent of infection (Max: 0.29, Min: 0.03), Total brightness per signal area (Max: 0.61, Min: 0.16). **B,** LY6A residues with at least 2 atoms within 5 angstroms (Å) of the modeled PHP.eB peptide. **C,** Complete model of the PHP.eB trimer and LY6A complex. The AlphaFold-Multimer structural prediction from **A** was combined with a capsid monomer-receptor structural prediction and optimized using Rosetta Remodel within the context of the AAV trimer (Figure 2.3A). **D,** Zoom-in view of the PHP.eB-LY6A binding interface in modeled PHP.eB-LY6A complex and PHP.eB residues with at least 2 atoms within 5 Å of LY6A.

Inspired by recent work (Chang and Perez, 2022a), this method uses AlphaFold-Multimer to place surface-exposed peptides spanning mutagenic insertions (AA587 – 594) from two distinct AAV variants in competition to interact with a potential receptor (Figure 2.1A). This comprises the minimal peptide encompassing the solvent-exposed residues of capsid variable region VIII. A combination of physical and geometric scoring parameters that include interface energy, binding angle, and binding pocket depth calculations are used to generate a peptide competition metric. Results from these individual pairwise competitions can be assembled into larger matrices that rank sets of AAV capsid insertion peptides according to their receptor-binding probability encoded in the AlphaFold2 neural network. When applied to LY6A and our newly identified receptors, we found that the experimentally-verified

LY6A, LY6C1, and CA-IV insertion peptides rise to the top of their respective rankings (Figure 2.1B-C). Some false negatives were also observed, however, as in 9P08 with LY6A or 9P36 with CA-IV.

AlphaFold2-based methods to understand the structural basis underlying receptor binding

In addition to predictions of whether a peptide binds to a receptor, we can also computationally interrogate the structural details of the binding interaction. We generated binding poses by pairing the top AAV insertion peptide with its receptor and validated the binding pose for each pairing by repeating our cell culture screen with receptors containing point mutations hypothesized to disrupt the high-confidence region of the binding interface (as determined by the per-residue estimated model confidence pLDDT score and consistency between replicate models) in these predicted poses. The PHP.eB peptide is predicted to nestle in a groove in LY6A, forming strong interactions at Pro5' and Phe6' (Figure 2.2A) with several LY6A residues (Figure 2.2B). We therefore introduced a point mutation in this groove, LY6A Ala58Arg, and found that it disrupts PHP.eB's enhanced infectivity with the wild-type receptor. This experimental result further bolsters confidence in in silico APPRAISE-AAV rankings.

To gain a full picture of the AAV-receptor interaction, we next modeled the PHP.eB insertion peptide and LY6A receptor complex within the context of the AAV capsid three-fold symmetry spike. This structure is challenging for standard modeling tools because of the large size of an AAV capsid (200kDa per trimer) as well as the often weak and dynamic binding interactions between engineered capsids and receptors (μM affinities possible without avidity (Xu et al., 2022)). AlphaFold-Multimer failed to capture direct contact between full-length PHP.eB capsid and LY6A in either a monomer-receptor or trimer-receptor configuration (data not shown). To address this challenge, we developed an integrative structure modeling pipeline. In this pipeline, an initial model of a AAV capsid trimer predicted using AlphaFold-Multimer (Evans et al., 2021) is structurally aligned with an AlphaFold2-predicted peptide-receptor complex model through the high-confidence Pro5' and Phe6' residues of the peptide insertion and RosettaRemodel (Huang et al., 2011) optimization of the linking peptide residues within the context of the AAV capsid three-fold symmetry spike (Figure 2.3A). This complete binding model (Figure 2.2C) provides a snapshot of a dynamic interaction that has thus far proven resistant to high-resolution structural characterization (Xu et al., 2022).

The PHP.eB-LY6A model is consistent with available experimental results. RMSD (root mean square deviation) between our monomeric PHP.eB model and a cryo-EM-based model (PDB ID: 7WQO) is 0.36 angstrom. RMSD increases in PHP.eB's engineered loop to 1.36 angstroms. The only high-confidence deviation from cryo-EM structures of un-complexed PHP.eB is the side chain of Phe6', which shows no substantial electron density, indicating flexibility, but forms a stable interaction with LY6A in our model (Figure 2.3B, right). The high confidence prediction of Pro5' and Phe6' aligns with recent evidence showing that PFK 3-mer insertion alone is sufficient to gain LY6A binding (Jang et al., 2022). While LY6A can bind any insertion loop of a trimer, additional interactions induce steric clashes supporting a ratio of one LY6A per capsid trimer. Interestingly, a PHP.eB-LY6A complex ensemble image forced to contain 60 bound copies of LY6A resembles a recently reported CryoEM map, whose analysis pipeline would average over all 60 singly-occupied binding sites to form a composite map (Figure 2.3C) (Xu et al., 2022). Our model shows that a single copy of both LY6A and AAVR PKD2 domain may bind to the same three-fold spike simultaneously without clashing (Figure 2.3D), in agreement with saturation binding experiments (Xu et al., 2022). Consistent with previous work showing the LY6A SNP D63G does not affect PHP.eB binding (Huang et al., 2019), the residue is greater than 10 angstroms from the PHP.eB peptide atoms in our models. The PHP.eB-LY6A complex model includes several interactions involving AAV insertion-adjacent residues, which is consistent with a previous report (Figure 2.2D) (Martino et al., 2021).

Application of the structure modeling methods to new receptors

We next applied these structural modeling methods to our newly identified receptors. Interestingly, unlike for LY6A and CA-IV, the predicted binding pose for PHP.C2 peptide with LY6C1 was found to vary with the version of AlphaFold-Multimer used, with v1 predictions closely matching mutational data from our cell culture infectivity assay (Figure 2.5A). Such complementarity between versions has been reported previously (Johansson-Åkhe and Wallner, 2022). In mouse CA-IV, 9P31 peptide invades the catalytic pocket of the enzyme (Figure 2.4A). The 9P31 tyrosine residue shared with 9P36 approaches the enzyme active site and 9P31's divergent tryptophan finds purchase in an ancillary pocket (Figure 2.4B). This predicted binding pose is competitive with the binding site of brinzolamide (PDB ID 3NZC) (Stams et al., 1998), a broad carbonic anhydrase inhibitor that is prescribed for glaucoma60. In our cell culture infectivity assay, brinzolamide shows a dose-dependent inhibition

of 9P31 and 9P36 potency while PHP.eB is unaffected (Figure 2.4B). The smaller brinzolamide binds deep in the catalytic core of CA-IV where side chains are largely conserved between species (Figure 2.4C). 9P31 peptide however extends to the surface of the enzyme where there is considerable sequence divergence. Similarly, while brinzolamide binds to both mouse and human CA-IV(Stams et al., 1998; Supuran, Scozzafava, and Casini, 2003), 9P31 and 9P36 are selective for mouse CA-IV (Figure 2.4D). Chimeric receptors that swap a highly divergent loop of the 9P31 binding site show that this region is necessary but not sufficient to control 9P31 and 9P36 potency. A second potential CA-IV binding site is also suggested by lower-ranked poses with 9P31 and 9P36, but mutagenesis experiments show inconsistent effects (Figure 2.5B). Engineering a human CA4-binding AAV with optimal BBB crossing properties is both critically important and not trivial without ready in vivo model systems for validation, as illustrated by the extensive, multi-year efforts realizing the transferrin receptor's potential.

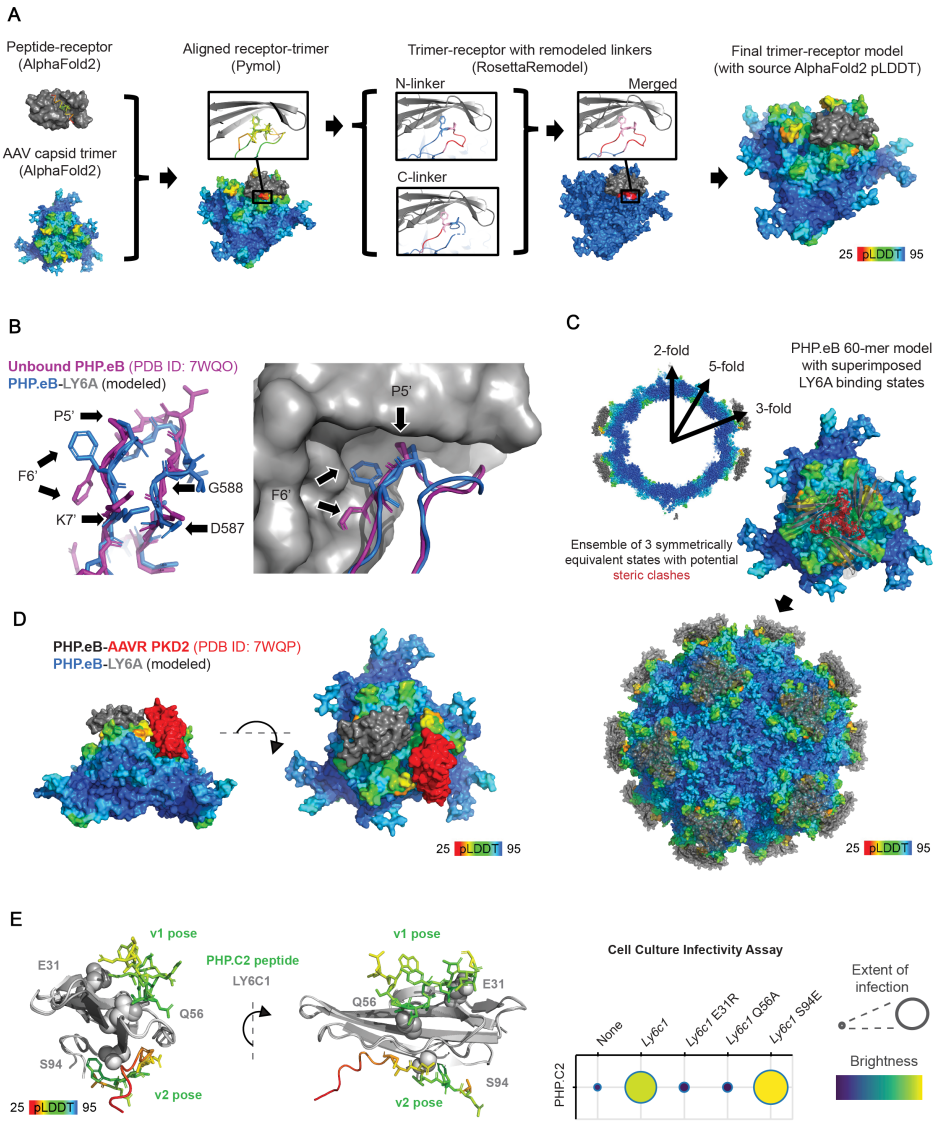


Figure 2.3: **An integrative structure modeling method yields a snapshot of PHP.eB-LY6A interaction. A,** A workflow for modeling engineered AAV-receptor complex structures (PHP.eB: colored by pLDDT score, LY6A: grey). **B,** Comparison between computationally modeled PHP.eB-LY6A (PHP.eB: blue, LY6A: grey) and CryoEM structure of unbound PHP.eB (purple, PDB ID: 7WQO). All high-confidence residues (pLDDT > 45, see left panel arrow) within the inserted peptide showed consistent conformations between the two models except for F6', which did not have clear side chain density in the CryoEM map. The side chain of F6' predicted in the unbound PHP.eB model would cause a steric clash in the predicted complex model. **C,** Assembled PHP.eB capsid-LY6A model representing an ensemble of all LY6A binding states, despite steric clashes, as would occur in cryo-EM particle reconstruction (PHP.eB: rainbow, LY6A:grey). Top: cross section of the assembled capsid model. Middle: zoom-in of a 3-fold spike, highlighting steric clashes (red) between three different binding states. **D,** Overlaid structures of computationally modeled PHP.eB-LY6A (PHP.eB: rainbow, LY6A: grey) and CryoEM model of PHP.eB-AAVR PKD2 (AAVR PKD2: red, PDB ID: 7WQP).

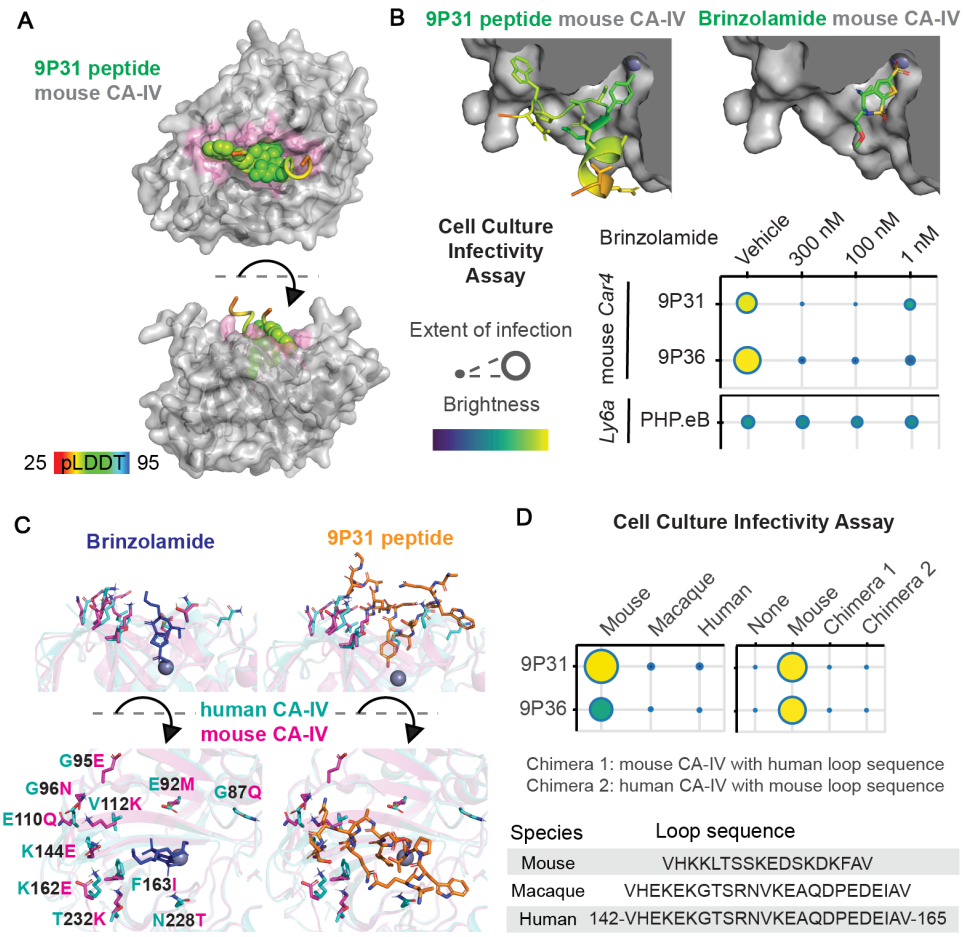


Figure 2.4: **Engineered AAV interactions with carbonic anhydrase IV. A,** AlphaFold-predicted mouse CAR4-9P31 peptide complex structure. 9P31 peptide is colored by pLDDT score at each residue with the highest confidence side chains shown as spheres. **B,** Cut-away view of mouse CAR4 catalytic pocket with modeled 9P31 peptide binding pose (top left) and crystallographic brinzolamide binding pose (PDB ID: 3ZNC, top right). Cell culture infectivity assay of brinzolamide's effects on engineered AAVs (bottom). Extent of infection (Max: 0.63, Min: 0.04), Total brightness per signal area (Max: 0.75, Min: 0.18) **C,** Views of amino acid side chains that differ between mouse (PDB ID: 3ZNC) and human (PDB ID: 1ZNC) carbonic anhydrase IV in relation to brinzolamide and 9P31 peptide binding poses. **D,** Potency in cell culture infectivity assay of 9P31 and 9P36 in HEK293T cells transfected with mouse, rhesus macaque, or human carbonic anhydrase IV receptors, as well as two chimeric receptors of mouse and human carbonic anhydrase IV that exchange the loop sequences depicted. Extent of infection (left, Max: 0.52, Min: 0.05, right, Max: 0.65, Min: 0.03), Total brightness per signal area (left, Max: 0.78, Min: 0.46, right, Max: 0.75, Min: 0.13)

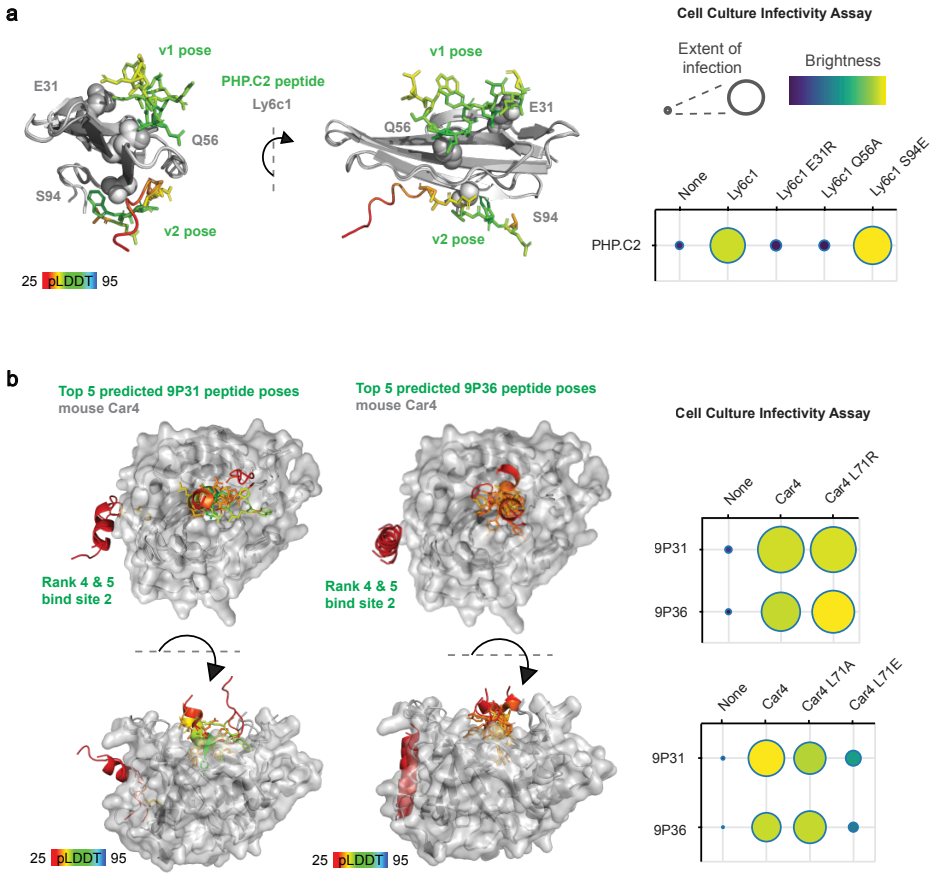


Figure 2.5: **Modeling peptide complexes with LY6C1 and mouse CA-IV a left**, Comparison between the computationally modeled PHP.C2-LY6C1 binding poses predicted by AlphaFold-Multimer v1 and v2. PHP.C2 peptides are colored by residue according to pLDDT score. LY6C1 residues chosen for mutation are shown as spheres. **right,** Potency of PHP.C2 for HEK293T cells transfected with wild type or mutant LY6C1 receptors. Extent of infection (Max: 0.67, Min: 0.13), Total brightness per signal area (Max: 0.72, Min: 0.31). **b left,** The top five AlphaFold2-predicted mouse CA-IV-9P31 and CA-IV-9P36 peptide complex structures. 9P31 and 9P36 peptides are colored by pLDDT score at each residue. **right,** Potency of 9P31 and 9P36 for HEK293T cells transfected with wild type or mutant CA-IV receptors designed to disrupt putative site 2 binding. Extent of infection (top, Max: 0.99, Min: 0.12, bottom, Max: 0.67, Min: 0.13), Total brightness per signal area (top, Max: 0.89, Min: 0.33, bottom, Max: 0.72, Min: 0.31).

2.4 Discussion

The blood-brain barrier restricts access to the CNS by large molecule research tools and therapeutics, limiting our ability to study and treat the brain (Banks, 2016; Sweeney et al., 2019; Zhao and Zlokovic, 2020; Profaci et al., 2020). Here we sought to expand the roster of protein targets through which biologicals and chemicals may access the CNS by de-orphanizing engineered AAVs selected through directed evolution for enhanced brain potency. While directed evolution methods have identified several engineered AAVs with enhanced tissue potency after systemic injection (Challis et al., 2022), the mechanisms by which engineered AAVs gain their enhancements are, with a few recent notable exceptions (Tabebordbar et al., 2021; Havlik et al., 2021), largely unknown. This is particularly true for engineered AAVs with enhanced potency in the CNS, where PHP.eB, which was found to use LY6A in many mouse strains (**Hordeaux**; Batista et al., 2020; Huang et al., 2019), stands alone in being de-orphanized. The strain dependence and murine restriction of PHP.eB's LY6A interaction accelerated a push toward NHPs for engineered capsid identification and validation for translational vectors. However, human gene therapy's increasing embrace of engineered AAV capsids in human clinical trials (Kuzmin et al., 2021; Burdett and Nuseibeh, 2022) coupled with the scarcity and costs of NHP(Tian, 2021; Subbaraman, 2021), highlight the need for higher throughput methods to validate engineered AAVs with diverse, and conserved, mechanisms for crossing the BBB. By screening a curated pool of 40 candidate receptors selected for the intersection of their CNS expression level and endothelial-cell specificity, we were able to identify LY6C1 and carbonic anhydrase IV as molecular receptors for enhanced blood-brain barrier crossing of ten LY6A-independent engineered AAVs (as well as LY6A-dependent PHP.N) (see the full version of Shay et al., 2023). These findings allow for more efficient allocation of NHPs, inform future directed evolution library designs, and enable receptor-guided engineering directly for human protein interaction.

Interestingly, neither LY6C1 nor CA-IV had been identified as among the most enriched proteins in CNS endothelial cells compared to peripheral endothelial cells (Munji et al., 2019). Given the distinct capsid sites for peptide insertion and galactose (Penzes et al., 2021) or AAVR interaction (Zhang et al., 2019; Meyer et al., 2019) and our model predicting simultaneous AAVR PKD2 and LY6A interaction without steric clashes, it is likely that the receptors identified here work in concert with AAV9's endogenous interaction partners to shape each AAV's tropism.

Using our integrative modeling pipeline, we generated a complete, experimentally-validated receptor complex model for PHP.eB with LY6A, which has otherwise resisted high-resolution structural characterization (Xu et al., 2022; Jang et al., 2022). This model illustrates the complementarity of PHP.eB to LY6A and predicts additional interactions outside of the insertion peptide. This insight provides opportunities for improved capsid engineering by both rational design (via in vitro selection for Ly6 family members with desirable expression patterns or conservation across species) and directed evolution (via negative selection pre-screens against purified Ly6 family proteins to encourage other BBB-crossing solutions). Our new APPRAISE-AAV in silico method is well suited to such screens. This method is also readily applied to any existing engineered capsid library dataset to mine for capsid variants likely to interact with a chosen target receptor, including CA-IV. Our modeling pipeline also provides high-confidence binding models for AAV receptor complexes that have proven difficult to structurally resolve. We note that the APPRAISE methodology is not limited to AAVs, and the pipeline for generating full AAV trimer complex structures may readily be employed to guide the translation of engineered peptide insertions identified through directed evolution in AAVs to other protein modalities.

Carbonic anhydrase IV is broadly conserved across vertebrates and has similar CNS expression profiles in humans (Tolvanen et al., 2013; Le Roy et al., 2014), with a recent single cell analysis of human brain vasculature confirming CA4's expression in the human BBB (Yang et al., 2022). Thus, CA4-interacting AAVs are attractive candidates for translation across diverse model organisms and potentially in human gene therapies. Both 9P31 and 9P36 display enhanced potency with mouse CA-IV but not rhesus macaque or human CA4. While neither virus would be expected to translate from mice to these species, we have identified a therapeutic target and mechanism for BBB crossing that may. The potential for specific engineered AAV binding epitopes to experience genetic drift between even closely related species confronts all products of directed evolution whose intended final use differs from their selection conditions. This potential takes on increasing importance when considering the potential for failed trials to preclude patients from future AAV treatments by eliciting cross-reactive neutralizing antibodies (Tenenbaum, Lehtonen, and Monahan, 2003; Ronzitti, Gross, and Mingozzi, 2020; Rapti and Grimm, 2021). Future rational engineering of new AAVs against species-appropriate CA4, aided by our APPRAISE-AAV method, is a promising new avenue for the generation of non-invasive vectors with enhanced CNS potency. Targeting CA4 may also find

application across diverse protein and chemical modalities.

2.5 Methods

Automated Pairwise Peptide Receptor Analysis for Screening Engineered AAVs (APPRAISE-AAV) FASTA-format files containing a target receptor amino acid sequence (mature protein part only) as well as peptide sequences corresponding to amino acids 587 through 594 (wild-type AAV9 VP1 indices) from two AAV capsids of interest were used for structural prediction using a batch version of ColabFold (Mirdita et al., 2022) (alphafold-colabfold 2.1.14), a cloud-based implementation of multiple sequence alignment (Mirdita et al., 2017; Mirdita, Steinegger, and Söding, 2019; Mitchell et al., 2019; Steinegger et al., 2019) 85–88, and AlphaFold2 Multimer (Evans et al., 2021). The ColabFold Jupyter notebook was run on a Google Colaboratory session using a GPU (NVIDIA Tesla V100 SXM2 16GB; we found that the same model of the GPU yielded the most consistent results). We chose alphafold-multimer-v2 as the default AlphaFold version unless otherwise specified. Each model was recycled three times, and ten models were generated from each competition. Models were quantified with PyMol (version 2.3.3) using a custom script to count the total number of atoms in the interface ($N_{contact}^{POI}$, defined by a distance cutoff of 5 angstroms), the total number of atoms in the peptide that are clashing with the receptor (N_{clash}^{POI}, defined by a distance cutoff of 1 Å), the binding angle of the peptide (θ, defined as the angle between the vector from receptor gravity center to receptor anchor and the vector from receptor gravity center to peptide gravity center), and the binding depths (d, defined as the difference of the distance between the closest point on the peptide to the receptor center and the minor radius of the ellipsoid hull of the receptor normalized by the minor radius) of the peptide in each putative peptide-receptor complex model. The minor radii of the ellipsoid hulls of receptors were measured using HullRad 8.1(Fleming and Fleming, 2018) (LY6A: 13.4 Å, LY6C1: 12.7 Å, mouse CA-IV: 23.0 Å). Finally, the metric $\Delta B^{(POI,competitor)}$ for ranking the propensity of receptor binding was calculated by subtracting the total binding score of the competing peptide from the counterpart score of the peptide of interest. The mean number of this metric across replicates was used to form a matrix and plot a heatmap. Peptides in the heatmap were ranked by the total number of competitions each peptide won minus the total number of competitions it lost (competitions with $\Delta B^{(POI,competitor)}$ scores that have p-values greater than 0.05 in the one-sample Student's t-test were excluded). Details of the APPRAISE method can be found in Chapter 3 of this thesis.

Computational structure modeling of receptor-AAV complexes Peptide-receptor structures were modeled using a similar procedure as described in the APPRAISE-AAV section but with only one single peptide of interest in the input file to achieve higher accuracy. AAV trimer-receptor complex models were produced using an integrative structure modeling method (Figure **??**A). Trimers at the AAV three-fold symmetry interface were chosen as the minimal complete binding interface with a putative receptor that might recapitulate the entire viral particle while optimizing computational efficiency. First, a peptide-receptor model was generated by modeling the 15mer peptide sequence between the residues 587 and 594 (both in wild-type AAV9 VP1 indices) from the AAV variant of interest in complex with the target receptor as described above. Then, a trimer model of the AAV variant of interest was modeled using AlphaFold2 Multimer. The two residues with the highest confidence score (pLDDT score) in the 15mer peptide of the peptide-receptor model, Pro5' and Phe6', were structurally aligned to the corresponding residues on the first chain of the trimer model. A coarse combined model was then generated by combining the receptor and the two high-confidence AAV residues from the peptide receptor model with the remaining AAV residues from the trimer model. The two loops between Pro5' and Phe6' and the high-confidence AAV9 backbone in the coarse combined model (corresponding to residues 588-(588+)4' and residues (588+)7'-590, respectively) were then individually remodeled using RosettaRemodel (Huang et al., 2011) from the Rosetta software bundle (release 2018.48.60516). Finally, these remodeled loops were merged to generate a final model. The pLDDT scores for each residue from the original AlphaFold2 outputs were used to color images of the final model.

2.6 Acknowledgment

We thank Catherine Oikonomou for help with manuscript editing, Michael Altermatt for assisting in filtering scRNAseq datasets for membrane proteins, Min Jee Jang for designing RNA sequencing variant barcodes, Josette Medicielo for plasmid purification, and Julie Miwa (Lehigh) for sharing a lynx1 expression plasmid. We thank the Caltech Protein Expression Center supported by the Beckman Institute for SPR studies, protein expression and purification. We thank Andrew D. Steele (CPP) for his contributions to the Caltech - CPP ASPIRE Program and the entire Gradinaru lab and the Beckman Institute CLOVER center staff for helpful discussion. Figures were created using images from BioRender.com.

2.7 Contributions

T.F.S. and V.G. conceived the project. T.F.S. and V.G. wrote the manuscript and prepared figures with input from all authors. X.D. and A.W.L. optimized LY6A-Fc expression protocol and A.W.L. produced LY6A-Fc protein. T.F.S., E.E.S., and D.A.W. produced AAVs. T.F.S. and J.V. performed SPR experiments. T.F.S. and D.B. analyzed the scRNAseq dataset. T.F.S. and E.E.S. developed and E.E.S. performed the cell culture infectivity assay. D.G. developed and implemented the in vitro transduction quantification and plotting pipeline, performed data analysis, and, with T.F.S, prepared in vitro transduction quantification plots. E.E.S. and M.B. performed immunofluorescence experiments. E.E.S., X.C., and M.B. performed CAR4-KO experiments, X.D. performed APPRAISE-AAV and developed computational structural modeling strategies. T.F.S., S.R.K., E.E.S, and X.C. performed and analyzed AAV-PHP.eC selections. E.E.S, X.C., A.F.C. and D.A.W. tested AAVs in wild type mice. T.F.S. and V.G. supervised and V.G. funded the project.

2.8 Declaration of Interests

The California Institute of Technology has filed a provisional patent for this work with T.F.M., X.D., and V.G. listed as inventors. V.G. is a co-founder and board member of Capsida Biotherapeutics, a fully integrated AAV engineering and gene therapy company. The remaining authors declare no competing interests

2.9 Related publication

Shay, Timothy F. et al. (2023). "Primate-conserved carbonic anhydrase IV and murine-restricted LY6C1 enable blood-brain barrier crossing by engineered viral vectors". In: *Science Advances* 9.16, eadg6618. DOI: 10.1126/sciadv.adg6618.

Chapter 3

APPRAISE: FAST, ACCURATE RANKING OF ENGINEERED PROTEINS BY RECEPTOR BINDING PROPENSITY USING STRUCTURE MODELING

3.1 Abstract

Deep learning-based methods for protein structure prediction have achieved unprecedented accuracy. However, the power of these tools to guide the engineering of protein-based therapeutics remains limited due to a gap between the ability to predict the structures of candidate proteins and the ability to assess which of those proteins are most likely to bind to a target receptor. Here we bridge this gap by introducing Automated Pairwise Peptide-Receptor AnalysIs for Screening Engineered proteins (APPRAISE), a method for predicting the receptor binding propensity of engineered proteins. After generating models of engineered proteins competing for binding to a target using an established structure-prediction tool such as AlphaFold-Multimer or ESMFold, APPRAISE performs a rapid (under 1 CPU second per model) scoring analysis that takes into account biophysical and geometrical constraints. As a proof-of-concept, we demonstrate that APPRAISE can accurately classify receptor-dependent vs. receptor-independent adeno-associated viral vectors and diverse classes of engineered proteins such as miniproteins targeting the SARS-CoV-2 spike, nanobodies targeting a G-protein-coupled receptor, and peptides that specifically bind to transferrin receptor or PD-L1. APPRAISE can be accessed through a web-based notebook interface using Google Colaboratory (https://tiny.cc/APPRAISE). With its accuracy, interpretability, and generalizability, APPRAISE promises to expand the utility of protein structure prediction and accelerate protein engineering for biomedical applications.

3.2 Introduction

Many protein-based biologics rely on precise targeting. As a result, protein engineers have devoted considerable efforts to create specific binders, using methods such as directed evolution (Deverman et al., 2016; Ravindra Kumar et al., 2020; Ring et al., 2013; Lee et al., 2001) and rational design (Cao et al., 2020; Chevalier et al., 2017; Yin et al., 2021). Currently, the costly experimental evaluation of candidate binders using *in vitro* and *in vivo* assays presents a bottleneck, which can be eased using

computational prioritization. (Sliwoski et al., 2014).

Two strategies are employed to predict protein functions: end-to-end sequence-function and two-step sequence-structure/structure-function. End-to-end, sequence-function models can predict complex functions such as enzyme activities or ion channel conductivity (Yang, Wu, and Arnold, 2019; Bedbrook et al., 2017), which are challenging to calculate using physical principles (Bolon, Voigt, and Mayo, 2002). However, such specialized models require domain-specific, high-quality training datasets for accurate prediction. In comparison, the two-step sequence-structure/structure-function strategy offers a more generalizable solution, particularly for functions with well-understood biophysical mechanisms such as protein-protein binding.

The rapid development of deep learning-based methods has brought unprecedented accuracy to the first step of the sequence-structure/structure-function strategy. Since AlphaFold2 (AF2)'s outstanding performance in CASP14 in 2020 (Jumper et al., 2021), several new deep learning-based structure-prediction tools have been released (Baek et al., 2021; Evans et al., 2021; Mirdita et al., 2022; Aderinwale, Christoffer, and Kihara, 2022; Motmaen et al., 2022; Ruffolo et al., 2022; Lin et al., 2022; Li et al., 2022; Wu et al., 2022; Wang et al., 2022a; Cheng et al., 2022; Liu et al., 2022), providing a diverse tool set for generating protein models with atomic-level precision. While the original AlphaFold2 can predict peptide-protein complexes (Tsaban et al., 2022), there are enhanced versions such as AlphaFold-Multimer that can model multi-chain complexes with greater accuracy (Evans et al., 2021; Motmaen et al., 2022). Importantly, these structure-prediction tools allow the generation of models in less than one GPU hour each, a level of throughput that experimental methods cannot match.

The second step, ranking target binding propensities based on structure predictions, has been less attended than the first. Structure-prediction tools generate confidence scores for predicted multimer models, such as pLDDT and pTM scores (used by AF2)(Jumper et al., 2021), interface pTM scores (used by AF-Multimer)(Evans et al., 2021), which have been used "off-label" as metrics to evaluate the probability of binding (Bennett et al., 2022; Motmaen et al., 2022). However, previous reports (Chang and Perez, 2022b) and our experience revealed that these scores alone are, in some cases, not reflective of binding propensities, particularly when the interaction is weak or transient. Extracting additional information stored in the 3D coordinates using biophysical principles may help improve the accuracy of binder ranking.

Ranking the binding probability of engineered proteins through modeled structures presents unique challenges. For example, the high sequence similarity between candidate molecules imposes a frequent challenge. Engineered protein variants are often constructed by modifying a short variable region to a common scaffold. Due to this similarity, the energy difference between the candidate binders can be very small, sometimes buried in the error of the energy function used for candidate ranking (Baker, 2019; Alford et al., 2017). This problem is compounded by structure-prediction methods that rely heavily on co-evolutionary information or homology, causing them to generate similar binding poses for the candidate proteins. Another major challenge is assessing a large number of predicted structure models efficiently. Direct quantification of protein-protein interface energy using interpretable, physics-based methods trades off between accuracy and speed (Gonzalez et al., 2020). For instance, molecular dynamics simulation methods can cost more than 10^3 CPU hours per model. Faster, less rigorous methods with the better-than-random ability to predict the impact of interface mutations still require 1 CPU minute to 1 CPU hour per non-antibody-antigen model (Gonzalez et al., 2020). In the post-AlphaFold era, an interpretable and efficient method of predicting the target binding of a large number of models would greatly accelerate protein engineering efforts.

Recently, Chang and Perez utilized competitive modeling with AF-Multimer to demonstrate a correlation between competition results and peptide binding affinities (Chang and Perez, 2022b). Nevertheless, the study's method to assess the competition results necessitates a comparison of the modeled structures to an experimentally solved "native" structure, which is not available for many engineered proteins.

To bridge the remaining gap between structure prediction and protein engineering, here we present Automated Pairwise Peptide-Receptor AnalysIs for Screening Engineered proteins (APPRAISE), a readily interpretable and generalizable method for ranking the receptor binding propensity of engineered proteins based on competitive structure modeling and fast physics-informed structure analysis.

3.3 Results

The workflow of APPRAISE (Figure 3.1) comprises four main components. In the first step, pairs of peptides from N candidate protein molecules (N^2 pairs total) are modeled in complex with a target receptor using a state-of-the-art structure method such as AF-Multimer (Evans et al., 2021). In the second stage, a simplified energetic binding score is calculated for each peptide (i.e., the peptide of interest (POI) and its

competitor). In the third optional step, geometrical constraints for effective binding are applied to these scores. Finally, the result of each competition is decided using the score difference between the POI and the competitor, and the peptides are ranked based on the matrix of competition results.

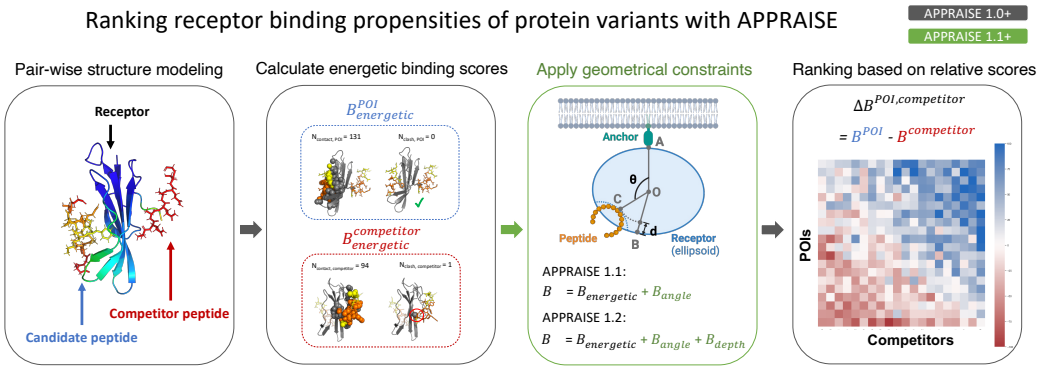


Figure 3.1: **Workflow of Automated Pairwise Peptide-Receptor AnalysIs for Screening Engineered proteins (APPRAISE).** First, peptides from the engineered protein candidates' receptor-binding region are modeled in competing pairs with the target receptor using tools like AF-Multimer or ESMFold. Second, a non-negative energetic binding score based on atom counting is calculated for each peptide. Third, in APPRAISE 1.1+, additional geometrical constraints critical for peptide binding, including the binding angle and pocket depth, are considered. Finally, a relative score for each match is calculated by taking the difference between the scores for the two peptides. The averaged relative scores form a matrix that determines the final ranking.

APPRAISE can accurately classify receptor-mediated brain transduction of viral vectors

We first developed APPRAISE to predict the binding propensities of engineered Adeno-Associated Viral (AAV) capsids for brain receptors. Recombinant AAVs are widely used as delivery vectors for gene therapy due to their relative safety as well as their broad and engineerable tropism. *In vivo* selections from libraries of randomized peptide-displaying AAV variants have yielded capsids that can transduce the animal brain (Deverman et al., 2016; Chan et al., 2017; Ravindra Kumar et al., 2020; Nonnenmacher et al., 2021; Goertsen et al., 2022b; Chuapoco et al., 2022; Chen et al., 2022b), an organ tightly protected by the blood-brain barrier (BBB). Widely-known examples among these capsids are AAV-PHP.B (Deverman et al., 2016) and AAV-PHP.eB (Chan et al., 2017), two AAV9-based (Gao et al., 2004) variants displaying short (7aa-9aa) surface peptides. The two variants can efficiently deliver

genetic cargo to the brains of a subset of rodent strains. Genetic and biophysical studies have revealed that the BBB receptor for PHP.B/PHP.eB in these trains is LY6A, a GPI-anchored membrane protein (Huang et al., 2019; Hordeaux et al., 2019; Batista et al., 2019). A dataset comprising peptide-displaying AAV capsids that were engineered in a similar way as PHP.B/eB was collected in order to train the APPRAISE method(Figure 3.5). Although binding between the AAV and the LY6A receptor is dynamic (Xu et al., 2022; Jang et al., 2022) and therefore challenging to be quantitatively measured, we could infer the binary LY6A-binding profiles of AAV capsids from their differential brain transduction profiles in mice strains with and without the receptor, producing a training set of peptide-displaying AAV capsids(Figure 3.5).

One challenge for modeling AAV capsids is that they are huge complexes made of 30,000+ amino acids (aa). In order to reduce computational costs for structure modeling and avoid complications arising from non-specific interactions, we modeled each AAV capsid variant using a single peptide spanning the engineered region (Figure 3.2a). This peptide (residues 587-594 in the VP1 sequence) includes 7 inserted residues and 8 contextual residues flanking the insertion. All of these residues are surface-exposed and may make direct contact with the receptor in the assembled capsid. Modeling this surface peptide (15 aa) is far less computationally intensive than modeling the entire capsid or even an asymmetric capsid subunit (500+ aa). In addition, compared to the latter, it may improve accuracy by eliminating competing interactions of residues normally buried in inter-subunit interfaces.

To discriminate relatively small differences in receptor binding propensities of candidate peptides, we modeled the peptides pairwise in competition for the target receptor (Morrone et al., 2017; Chang and Perez, 2022b). To evaluate the competition results efficiently, we designed a score based on simple atom counting as a rough estimate of the interface free energy between the peptide of interest (POI) and the receptor in a structure model (Figure 3.2b). This score, which we term the energetic binding score ($B^{POI}_{energetic}$, simplified as B^{POI}_{0}), is a non-negative value calculated from the numbers of contacting and clashing atoms at the interface (Eq. 3.1). We describe the detailed rationale behind this score in Methods.

$$B^{POI}_{0} = B^{POI}_{energetic} = \max(N^{POI}_{contact} - 10^{3} \cdot N^{POI}_{clash}, 0) \tag{3.1}$$

To take full advantage of the information encoded in the competitive models, we

further derived a "relative binding score", inspired by the "specificity strategy" for protein-protein interface design (Bolon et al., 2005). The relative score takes the difference between the absolute scores for the POI and competitor peptide (Eq. 3.2), rewarding POIs destabilizing competing peptides' binding.

$$\Delta B_0^{POI,competitor} = B_0^{POI} - B_0^{competitor} \tag{3.2}$$

An engineered protein must meet certain geometrical constraints to effectively bind to a membrane receptor (Figure 3.2c). To utilize this geometrical information, which is likely unused by structure-prediction tools, we incorporated two essential constraints for effective binding: the binding angle and the binding depth (Figure 3.2c-e).

The first constraint comes from the angle a binding protein can make (Figure 3.2c,d). In modeling a peptide-receptor complex using the extracellular domain of the membrane receptor (e.g., LY6A), most structure predictors (e.g., AF-Multimer) would consider the whole surface of the domain to be accessible by the peptide. However, in biological conditions, the membrane-facing side of the receptor is inaccessible to the engineered peptide. This polarity of accessibility is a general property of any receptor that is closely anchored to a larger complex. To account for the potentially huge energy cost of an engineered peptide binding these inaccessible locations, we used a steep polynomial term to penalize peptides that bind to the anchor-facing part of the receptor (Figure 3.2d, defined in the Methods by Eq. 3.1). B_0^{POI} is adjusted by this geometrical constraint term, rectified to be non-negative, and $\Delta B_0^{POI,competitor}$ is also re-calculated accordingly, yielding new scores B_1^{POI} and $\Delta B_1^{POI,competitor}$ (Eq. 3.3).

$$\begin{aligned} \Delta B_1^{POI,competitor} &= B_1^{POI} - B_1^{competitor} \\ &= \max(B_{energetic}^{POI} + B_{angle}^{POI}, 0) \\ &\quad - \max(B_{energetic}^{competitor} + B_{angle}^{competitor}, 0) \end{aligned} \tag{3.3}$$

The second constraint concerns the binding pocket depth (Figure 3.2c,e). We hypothesized that peptides binding to a deeper pocket on the receptor surface might

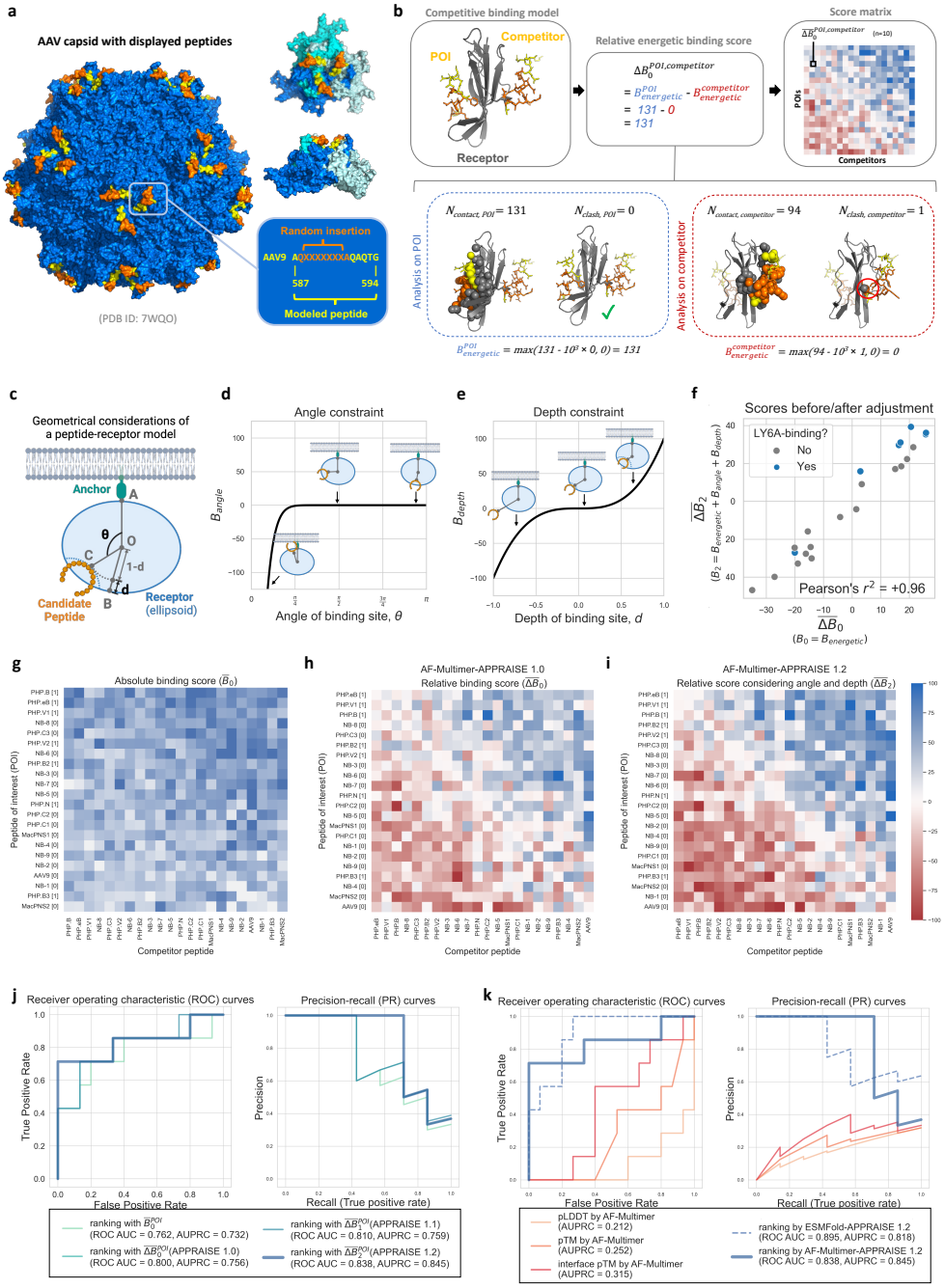


Figure 3.2: **Binary classification of receptor-binding AAV capsids using physical and geometrical principles.** See the next page for the caption.

benefit from longer receptor residence time, which is vital for the efficacy of many therapeutics (Copeland, 2016). Based on this hypothesis, we included a pocket depth consideration in APPRAISE's scoring function. We used a relative pocket depth measurement instead of an absolute peptide-receptor distance measurement to avoid possible bias caused by the sizes of different receptors. We then used an

Figure 3.2: **Binary classification of receptor-binding AAV capsids using physical and geometrical principles. a**, a structure model of AAV-PHP.eB, highlighting the site for inserting the displayed peptide (orange) and the peptide used for APPRAISE modeling (yellow or orange). The left image shows the AAV capsid of 60 structurally identical subunits. The two images on the top right show a top view and a side view around the 3-fold axis, respectively. The three subunits that make the trimer are colored blue, cyan, and white. The sequence corresponding to the peptides is shown in the bottom right. **b**, An example showing the calculation process of a relative energetic binding score. The number of contacting atoms (< 5 Å) and the number of clashing atoms (< 1 Å) for each peptide in the competition are counted, and an absolute energetic binding score is calculated based on the counts according to Eq. 3.1. A difference between the two numbers, or the relative energetic binding score, is then calculated. The competition result between two peptides is determined using the average of relative binding scores across 10 replicates. The matrix of the mean scores is then used to rank the peptides of interest (POIs). **c**, A simplified geometrical representation of a peptide-receptor model, where the hull of the receptor is represented by an ellipsoid (blue). Point O: the center of mass of the receptor. Point A: the receptor's terminus attached to an anchor. Segment OB: the minor axis of the ellipsoid receptor hull. Point C: the deepest point on the candidate peptide (orange). θ: the binding angle of the peptide. d: the binding pocket depth of the peptide. **d**, The angle constraint function. Three representative scenarios with different binding angles are highlighted. **e**, The depth constraint function. Three representative scenarios with different binding depths are highlighted. **f**, Comparison of the averaged relative binding energy scores before geometry-based adjustments vs. after adjustments. **g-i** Heatmaps representing the matrix of mean scores 22 AAV9-based capsid variants, including **g** mean absolute binding scores, **h** mean relative binding scores, and **i** mean relative binding scores that have considered both angle and depth constraints. All heatmap matrices were sorted by point-based round-robin tournaments (Methods). Bracketed numbers in the row labels are LY6A-binding profiles of the capsids inferred from experimental evidence (Figure 3.5). Each block in the heatmap represents the mean score measured from 10 independent models. **j-k**, comparison of different ranking methods used as binary classifiers to predict the LY6A-binding profile of 22 AAV9-based capsid variants.**j**, comparison between rankings given by different versions of APPRAISE scores using AF-Multimer as the structure prediction tool. **k**, comparison between rankings given by confidence scores of AF-Multimer versus rankings given by APPRAISE 1.2 using either AF-Multimer or ESMFold as prediction engines. The sequence and shape parameters of LY6A used for the modeling and analyses are included in Table 3.1.

odd polynomial term to reward peptides that insert into deep pockets on the receptor while penalizing peptides that attach to surface humps (Figure 3.2e, defined in the Methods by Eq. 3.2). The addition of the depth term gives us an adjusted score $\Delta B_2^{POI,competitor}$ (Eq. 3.4).

$$\begin{aligned}\Delta B_2^{POI,competitor} &= B_2^{POI} - B_2^{competitor} \\ &= \max(B_{energetic}^{POI} + B_{angle}^{POI} + B_{depth}^{POI}, 0) \\ &\quad - \max(B_{energetic}^{competitor} + B_{angle}^{competitor} + \\ &\quad B_{depth}^{competitor}, 0)\end{aligned} \tag{3.4}$$

We compared different versions of scoring methods based on competitive modeling results using AF-Multimer modeling (Figure 3.2g-i). Individual matching scores with statistical significance were used to determine wins and losses, and the total matching points in a tournament were used to rank all candidate proteins (Methods). We found that simple atom-counting-based B_0^{POI} can already differentiate LY6A-binding peptides from non-binders(Figure 3.2g, j). Compared to B_0^{POI} alone, the relative score $\Delta B_0^{POI,competitor}$ showed improved prediction power, a ROC-AUC of 0.800 and an AUPRC of 0.756 for the training dataset (Figure 3.2h, j, k). Adding both geometrical terms, B_{angle} and B_{depth}, into consideration indeed improved the prediction accuracy of the binding score (Figure 3.2i-k), yielding a ROC-AUC of 0.838 and an AUPRC of 0.845 (Figure 3.2j, k). Importantly, the improvement in ROC-AUC mainly came from the low-false-positive-rate segment of the ROC curve, which is crucial for *in silico* screening of engineered proteins. For clarity, we name the version that considers only the angle constraint (through score ΔB_1) APPRAISE 1.1 (Figure 3.6a) and the version that considers both angle and depth constraints (through score ΔB_2) APPRAISE 1.2 (Figure 3.2i).

We then compared AF-Multimer-based APPRAISE 1.2 with other structure-based peptide affinity ranking methods on the AAV dataset (Figure 3.2k). With this particular dataset, the model confidence scores pLDDT, pTM, and interface pTM fail to differentiate whether an AAV variant is an LY6A binder, producing worse-than-random prediction (ROC AUC < 0.5). This is possibly due to the dynamic nature of the interaction between LY6A-binding AAV variants and the receptor (Xu et al., 2022; Jang et al., 2022), which causes the confidence scores of the complex models to be generally low. APPRAISE 1.2 utilizing ESMFold as the structure prediction engine, however, performed at a comparable level to AF-Multimer-APPRAISE 1.2 (Figure 3.6b), with a ROC AUC of 0.895 and AUPRC 0f 0.818 (Figure 3.2k).

AF-Multimer-APPRAISE 1.2 ranking outperformed all other ranking methods at the low-false-positive-rate end of the ROC curve, with a true positive rate of 0.714 and no false positive predictions. The performance with stringent cut-off values is particularly relevant for protein engineering applications, where the goal is typically to identify a few positive binders from a large number of negative, non-binding candidates. The superiority of AF-Multimer-APPRAISE 1.2 in dealing with this kind of imbalanced library is also evidenced by its highest AUPRC. Because of this, we chose to characterize AF-Multimer-APPRAISE 1.2 further. In the following text, 'APPRAISE' will be used to refer to AF-Multimer-APPRAISE 1.2 unless otherwise specified.

APPRAISE is generally applicable to diverse classes of engineered proteins

To determine the applicability of APPRAISE to different classes of engineered proteins, we applied the method to four classes of engineered protein binders targeting four representative receptors for therapeutics.

We first applied APPRAISE 1.2 to other short peptide binders (Figure 3.3a-d). In the first trial, the method successfully ranked a peptide selected by phage display to bind human transferrin receptor (Lee et al., 2001), a well-characterized BBB receptor, over non-binding counterparts from the same selection (Lee et al., 2001) (Figure 3.3a). In the second trial, evaluating two 47aa-long, rationally-designed PD-L1 binding peptides (Yin et al., 2021)) against the scaffold and length-matched AAV variable region fragments, both designed PD-L1 binding peptides were clear winners, with the higher-affinity MOPD-1 peptide topping the list despite the sequence similarity (Figure 3.3c, Figure 3.7a).

We next tested whether APPRAISE 1.2 can be used to evaluate larger proteins, for example, computationally designed miniproteins (50-90 aa) that bind to the receptor-binding domain (RBD) of SARS-CoV-2 spike protein (Cao et al., 2020) (Figure 3.3e-g). Among the designed miniproteins, 5 can neutralize live SARS-CoV-2 virus *in vitro* with IC50 from 20 pM to 40 nM (Cao et al., 2020). The APPRAISE 1.2 rankings of the 5 neutralizing miniproteins matched well with their IC50 rankings (Spearman's $\rho = 0.90$, $p = 0.037$, Figure 3.3g). The predictive accuracy of APPRAISE decreased when non-neutralizing miniproteins (Cao et al., 2020) and control AAV fragments were included (Spearman's $\rho = 0.88$, $p < 0.001$, Figure 3.3g); nevertheless, the top 4 binders still remained on the top. In contrast, the ranking given by the iPTM score of AF-Multimer only achieved a Spearman's

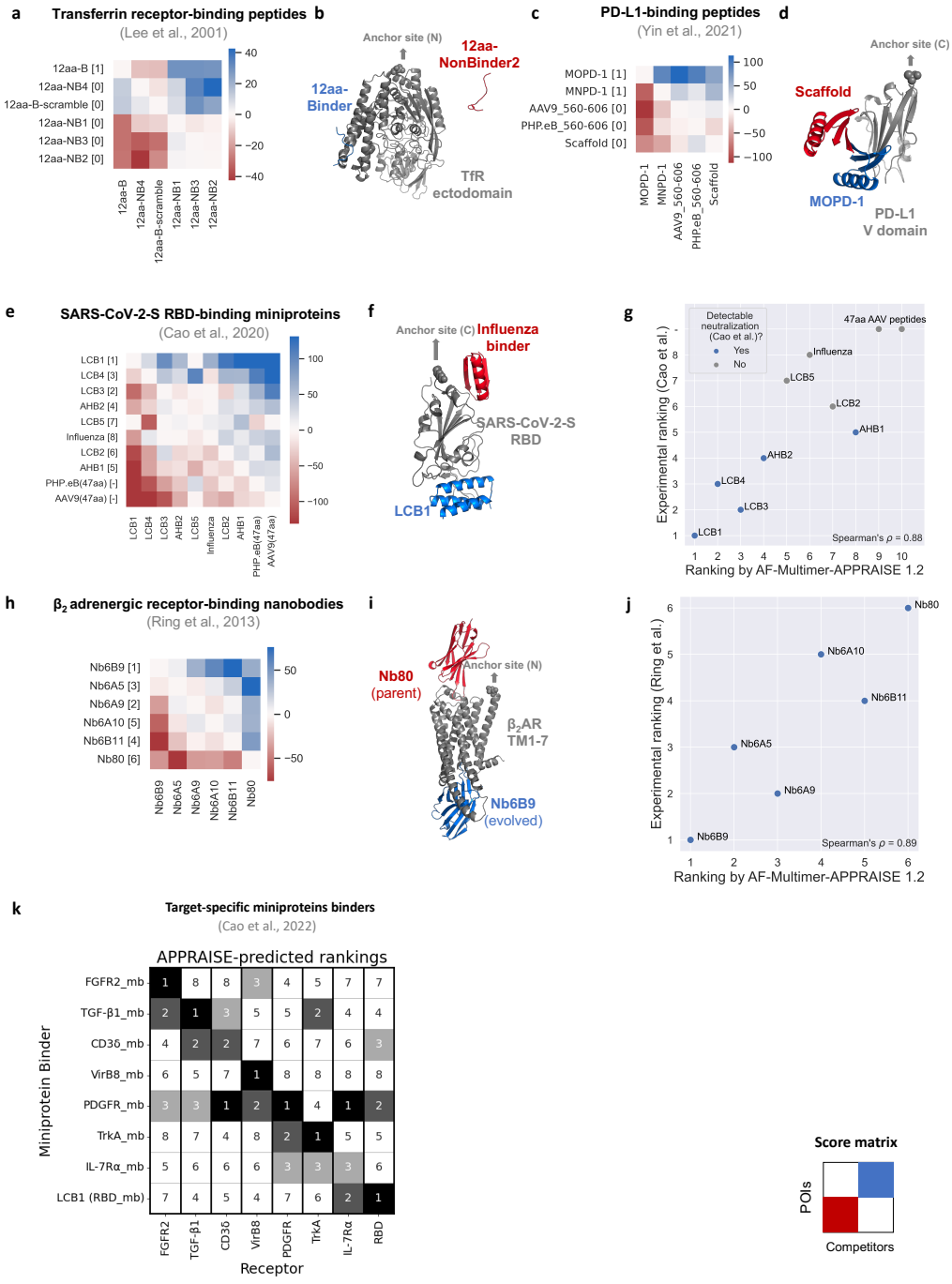


Figure 3.3: **AF-Multimer-APPRAISE 1.2 accurately ranks binding propensities of different classes of engineered proteins.** See the next page for the caption

ρ of 0.67 (p = 0.035).

We also used APPRAISE to rank 6 nanobodies (120 aa) that were evolved experimentally (Ring et al., 2013) with highly similar scaffolds (Figure 3.7b) to bind to an activated conformation of β_2 adrenergic receptor (β_2AR), a G-protein-coupled

Figure 3.3: **AF-Multimer-APPRAISE 1.2 accurately ranks binding propensities of different classes of engineered proteins. a-b,** APPRAISE 1.2 ranking of transferrin receptor-binding peptides and non-binding control peptides (Lee et al., 2001). **a,** Pairwise score matrix and ranking of a panel of 12-aa peptides given by APPRAISE 1.2. Bracketed numbers in the row labels are experimentally determined transferrin receptor-binding profiles of each peptide (Lee et al., 2001). **b,** A representative AF-Multimer model result of a binding peptide (blue) competing against a non-binding peptide (red) for binding to transferrin receptor. **c-d,** APPRAISE 1.2 ranking of PD-L1-binding peptides and non-binding control peptides (Yin et al., 2021). **c,** Pairwise score matrix and ranking of a panel of 47-aa peptides given by APPRAISE 1.2. Bracketed numbers in the row labels show the PD-L1-binding profile of each peptide determined either experimentally (for MOPD-1, MNPD-1, and scaffold protein) or by expectation (for AAV9 and PHP.eB) (Yin et al., 2021). **d,** A representative AF-Multimer model result of MOPD-1 (blue), a designed binding peptide, competing against a non-binding scaffold peptide (red) for binding to PD-L1. **e-g,** APPRAISE 1.2 ranking of SARS-CoV-2-S RBD-binding miniproteins(Cao et al., 2020). **e,** Pairwise score matrix and ranking given by APPRAISE 1.2. Bracketed rankings in the row labels are determined based on experimentally measured IC50 of each miniprotein to neutralize live SARS-CoV-2 (Cao et al., 2020). **f,** A representative AF-Multimer model result of LCB1 (blue), a SARS-CoV-2-S RBD-binding miniprotein, competing against an influenza virus-binding miniprotein (Chevalier et al., 2017) (red). **g,** A scatter plot showing the correlation between APPRAISE-predicted ranking and experimentally-measured IC50 ranking of all miniproteins tested. Blue points highlight binders that showed the capability of complete neutralization of the SARS-CoV-2 virus in the tested range of concentration *in vitro*. **h-j,** APPRAISE 1.2 ranking of β_2 adrenergic receptor-binding nanobodies (Ring et al., 2013). **h,** Pairwise score matrix and ranking given by APPRAISE 1.2. Bracketed numbers in the row labels are rankings of experimentally measured binding of each nanobody (Ring et al., 2013).**i,** A representative AF-Multimer model result of Nb6B9 (blue), the strongest evolved binder to active β_2AR, competing against Nb80 (red), the nanobody used for the evolution. **j,** A scatter plot showing the correlation between APPRAISE-predicted ranking and experimentally-measured ranking by β_2AR binding of all nanobodies tested. Each block in the heatmap represents the mean score measured from 10 independent models. For comparison, rankings given by AF-Multimer-APPRAISE 1.0, ESMFold-APPRAISE 1.2, and interface pTM of SARS-Cov2-S RBD binding miniproteins and β_2 adrenergic receptor-binding nanobodies are shown in Figure 3.8. **k,** A summary of APPRAISE rankings of eight miniproteins (Cao et al., 2022) designed to bind to eight different target receptors. Figure 3.9 displays the score matrices utilized for rankings with individual receptors. Tables 3.1 and 3.2 include sequences and shape parameters of all receptors.

receptor (GPCR) (Figure 3.3h-j). APPRAISE 1.2 correctly found the strongest

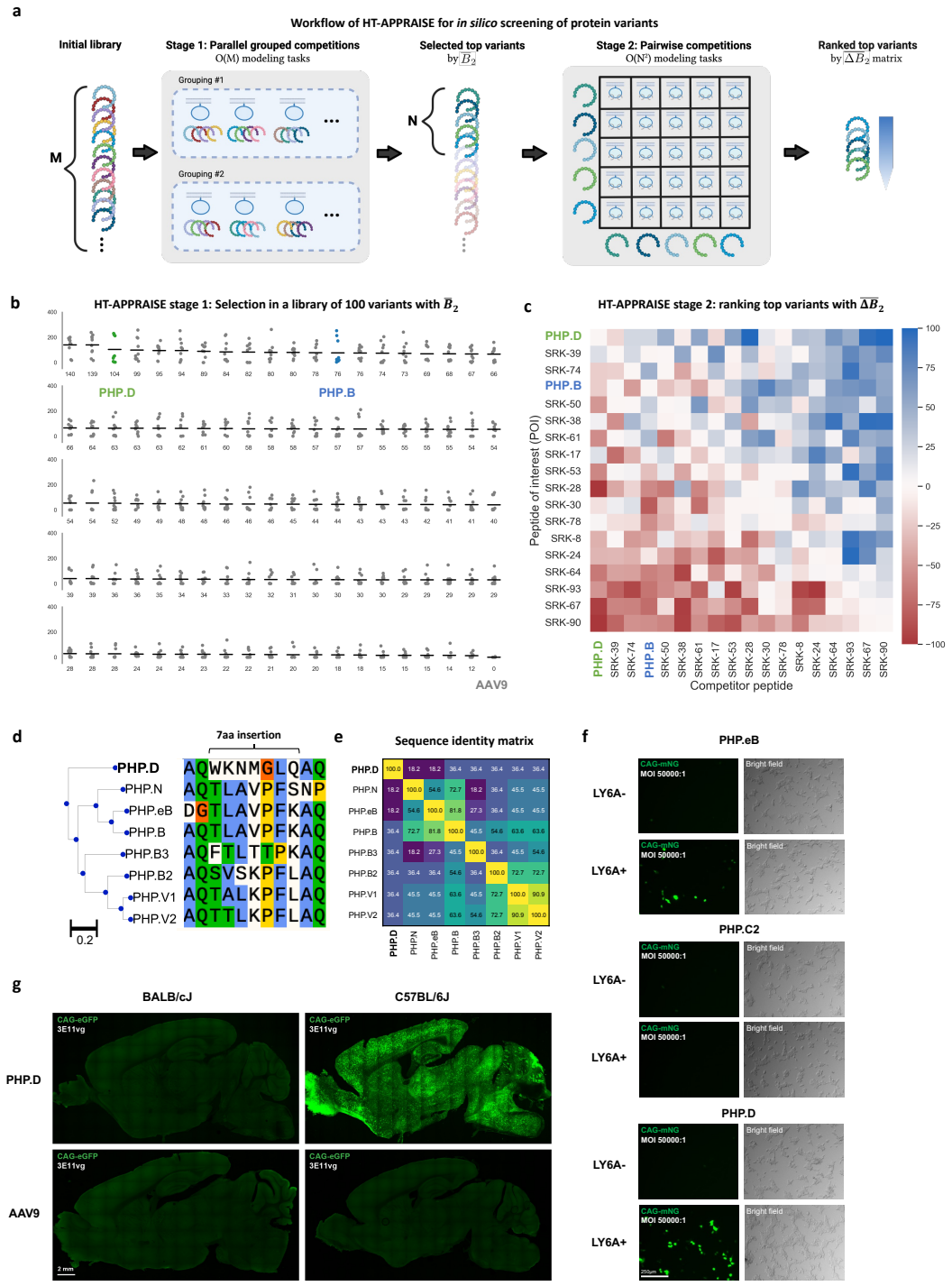


Figure 3.4: **An *in silico* HT-APPRAISE screening of a medium-sized AAV library identifies a LY6A-dependent variant with a distinct sequence.** See the next page for the caption.

evolved binder and placed the parent (the weakest binder among all candidates) at the bottom (Figure 3.3h). The overall predicted ranking correlated well with

Figure 3.4: **An *in silico* HT-APPRAISE screening of a medium-sized AAV library identifies a LY6A-dependent variant with a distinct sequence. a,** A schematic showing the two-stage strategy for *in silico* screening of a variant library. In the first stage, M variants of interest are randomly pooled into groups of 4 and compete for receptor binding. At least two parallel groupings are used to reduce bias. Each peptide's mean absolute binding score in the pool competitions is used for selecting the top N variants. In the second stage, the top N variants compete pairwise using standard APPRAISE for a more accurate ranking. **b-c,** Results from a proof-principle screening with 100 AAV9-based variants, including the wild-type control and variants with 7aa insertions. Using a standard random algorithm, a total of 97 variants were picked from a list of 9000 variants (Ravindra Kumar et al., 2020) that demonstrated higher brain enrichment than the wild-type AAV9 after one round of screening in C57BL/6J mice ("Round 2 library" in Ravindra Kumar et al., 2020). PHP.B and PHP.D, two known brain-transducing capsids, and wild-type AAV9, are spiked into the library. Table 3.4 shows the peptide sequences used in the screening. **b,** Stage 1 result. Dots indicate absolute binding scores measured from individual structure models. Horizontal bars indicate the mean scores of each variant. Scores of PHP.D (ranked 3^{rd}), PHP.B (ranked 13^{th}), and AAV9 (ranked 100^{th}) are highlighted. **c,** Stage 2 result. Rows corresponding to scores of PHP.D (ranked 1^{st}), PHP.B (ranked 4^{th}) are highlighted. Each block in the heatmap represents the mean score measured from 10 independent models. **d-g,** Characterization of PHP.D, a variant that tops the *in silico* screening. **d,** Sequence alignment and phylogenetic tree of known LY6A-dependent brain transducing variants. The sequence of PHP.D is very different from all other variants. The alignment and sequence distances were generated with Clustal Omega (Sievers et al., 2011). The colored alignment is plotted with Snapgene software. Blue: conserved hydrophobic residues; green: conserved hydrophilic residues; orange or yellow: conserved unique residues (glycine or proline). **e,** Sequence identity matrix of the LY6A-dependent variants. **f,** *In vitro* infectivity assay in HEK293T cells. PHP.D and PHP.eB showed LY6A-enhanced transduction, while the negative control PHP.C2 did not show LY6A-enhanced transduction. AAV capsids carrying a fluorescent protein expression cassette were applied to HEK293T cells either transfected with LY6A or not at 5×10^8 vg per well in a 96-well plate. Images were taken 24hr after transduction. n=3 per condition. Scale bar, 250 µm **g,** *In vivo* brain transduction of PHP.D vs. AAV9 in two mice strains. PHP.D showed transduction only in the LY6A+ strain, C57BL/6J. AAVs carrying CAG-mNeonGreen transgene were injected intravenously at 3×10^{11} vg per animal, and the tissues were harvested and imaged 3 weeks after injection. n=3 per condition. Scale bar, 2 mm.

the ranking from experimentally determined binding readouts (Ring et al., 2013) (Spearman's $\rho = 0.89$, $p = 0.02$, Figure 3.3j), surpassing the prediction given by the iPTM score of AF-Multimer (Spearman's $\rho = 0.49$, $p = 0.329$, Figure 3.8f). The ability of APPRAISE to rank the binding affinity of nanobodies, a widely-used

therapeutic modality, has the potential to expand its value in drug design and development. However, predicting the structure of larger adaptive immune complexes like IgG-antigen complexes is still considered a difficult task in general, and further improvements in the underlying structure prediction methods that APPRAISE relies on are required to generalize the ranking capability to these targets.

To evaluate the cross-receptor capabilities of APPRIASE, we used the method to rank eight recently developed miniproteins binders targeting eight different therapeutically significant receptors (Cao et al., 2022). This ranking included all receptors with a ligand binding domain of less than 250 amino acids in Cao and coworkers' study. APPRAISE is capable of accurately identifying the correct binder within the top 3 in every instance, and 6 out of 8 times, the correct binder was ranked as the top 1 (Figure 3.3k, Figure 3.9).

Our analysis involved comparing the performance of AF-Multimer-APPRAISE 1.2 to alternative methods on both the miniprotein dataset and the nanobody datasets. AF-Multimer-APPRAISE1.2 again yielded the most accurate predictions when compared to AF-Multimer-APPRAISE 1.0, ESMFold-APPRAISE 1.2, or interface pTM scores given by AF-Multimer (Figure 3.8), reflected by higher Spearman's correlation to experimental rankings. ESMFold-APPRAISE 1.2 failed completely with the miniprotein dataset (Figure 3.8b). Upon further inspection, we found that the unfolded SARS-Cov-2-S RBD structure in ESMFold-generated complex models can explain the failed ranking prediction.

Without any fine-tuning, AF-Multimer-APPRAISE 1.2 demonstrated consistent prediction ability for ranking all four classes of proteins, including experimentally-selected and rationally-designed peptides, computationally-designed miniproteins, and nanobodies. Realizing the potential general applicability of the APPRAISE method, we have created a web-based notebook interface to make it readily accessible to the protein engineering community (Figure 3.10).

HT-APPRAISE screening can identify novel receptor-dependent capsid variants

We next adapted APPRAISE 1.2 for *in silico* screening. The computational cost in the pairwise competition mode grows quadratically with the number of input variants, which is unsuitable for high-throughput screening. To address this scalability issue, we designed a two-stage screening strategy named high-throughput (HT)-APPRAISE (Figure 3.4a). The first stage aims to shrink the size of the variant

library using a less accurate yet more scalable strategy. Variants are randomly pooled into groups and compete for receptor binding. The variants are then ranked by their absolute score $\overline{B}_2^{POI}$. The number of pooled competitions grows linearly with the number of variants in the starting library, making the first stage of HT-APPRAISE suitable for larger libraries. In the second stage, the top variants selected from the first stage compete pairwise, yielding a matrix of $\overline{\Delta B}_2^{POI,competitor}$ and a more accurate ranking.

We used our HT-APPRAISE *in silico* screening to find LY6A binders in a library of 100 capsid variants (Figure 3.4b-c). This library is composed of 97 capsid variants randomly chosen from a list of 9,000 variants showing superior brain enrichment in C57BL/6J mice compared to the wild-type AAV9 capsid (Ravindra Kumar et al., 2020) as well as three spiked-in capsids: the variants PHP.B and PHP.D, and the wild-type AAV9 (Table 3.4). PHP.D is a brain-transducing capsid identified in a recent directed evolution campaign in our lab, the relevant receptor for which was unknown.

The HT-APPRAISE screening took 24 hours using 3 parallel Google Colaboratory GPU sessions and a laptop computer. In both stages of the HT-APPRAISE, the most time-consuming step was the structural prediction, which took approximately 0.1-1 GPU minute per peptide-LY6A model (a complex made of 114 aa total). In comparison, the time cost for structural analysis was negligible, taking less than 1 second per model on a CPU.

After the first stage of screening, both PHP.D and PHP.B appeared in the top 15% of the library (ranked by $\overline{B}_2^{POI}$) (Figure 3.4b).). In the second stage, the top 18 capsids were ranked using pairwise APPRAISE 1.2 (Figure 3.4c). PHP.D and PHP.B were the 1^{st} and the 4^{th} in the final ranking.

The most intriguing aspect of PHP.D's result is that its variable region bears little sequence similarity to any of the LY6A-dependent variants used to develop the APPRAISE method (Figure 3.4d-e). To confirm this prediction result, we experimentally tested PHP.D's LY6A dependency. An *in vitro* viral infection assay showed that PHP.D indeed exhibits LY6A-enhanced transduction of HEK293T cells (Figure 3.4f). In addition, *in vivo* systemic delivery of PHP.D packaging a ubiquitously-expressed fluorescent protein revealed that the brain transduction capability of this capsid variant is restricted to LY6A-expressing mouse strains (Figure 3.4g). The ability of HT-APPRAISE to identify binders with distinct sequences highlights the generalizability of the physics-informed, sequence-structure/structure-function

strategy.

3.4 Discussion

Here we describe APPRAISE, a structure-based, physics-informed method that accurately ranks receptor binding propensities of engineered proteins. APPRAISE uses a competitive, pairwise modeling strategy to capture affinity differences between even proteins with similar sequences and takes into account biophysical and geometrical principles.

The competition-based structure modeling strategy addresses the challenge of assessing small differences in binding affinity with high accuracy. This challenge was highlighted by a recent benchmarking study using AF2 models for molecular docking of small-molecule antibiotic candidates. The authors reported that the prediction power of the direct physics-based scoring is no better than a random model (Wong et al., 2022). By contrast, a competition-based modeling strategy might have helped cancel the shared noise and amplify the signal arising from the small affinity differences. The competition setup, in many cases, forces the structure-prediction neural network to put only the more probable binder close to the receptor (Figure 3.3 d, i), converting a small probabilistic difference into a binary output.

The generalizability of AF-Multimer-APPRAISE is shown by its accurate ranking of five different classes of engineered proteins and twelve different receptors. This generalizability may be grounded in the physical principles learned by AlphaFold (**PhysRevLett.129.238101**; Chang and Perez, 2022b). For example, consistent with the report of Chang and Perez, 2022b, we also observed a recurring trend in our tests with AF-Multimer where binders ranked at the top are frequently predicted with secondary structures, which may indicate a stable binding interface.

A key feature of APPRAISE is that its analysis module uses only information stored in the 3D coordinates, making the modular pipeline compatible with other computational tools for protein engineering. For example, the structure-prediction tool used in APPRAISE can be replaced by any current or future structure-prediction tool. Moreover, APPRAISE can be an orthogonal validation tool for structure-based protein design methods (Ovchinnikov and Huang, 2021; Anand et al., 2022; Dauparas et al., 2022), particularly those that rely on optimization of predicted confidence scores (Wang et al., 2022b).

The scalable, two-stage HT-APPRAISE strategy we designed allows *in silico* screen-

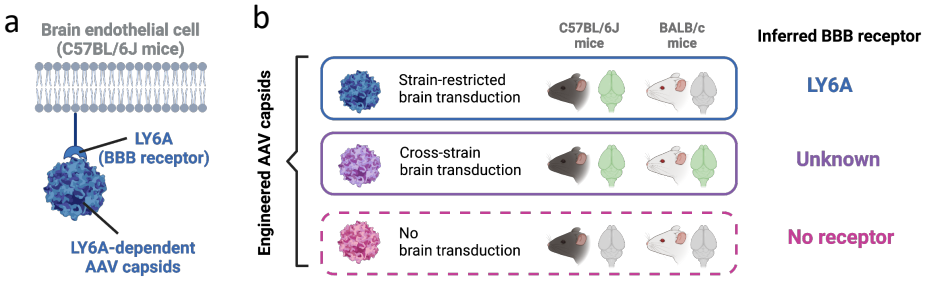
ing of protein candidates for receptor binding (Figure 3.4). Such screening can help prioritize leading candidates during drug discovery, reducing the huge time, financial, and environmental costs of experimental validation. For example, the computational tasks needed to screen 100 AAV variants that we presented here (Figure 3.4) could be completed within 24hr with research-grade computational resources. *In vivo* characterization of the capsids at a comparable scale would have taken several months.

As a competition-based ranking method, APPRAISE faces several intrinsic limitations. One such limitation is that APPRAISE only outputs the relative, not the absolute, probability of binding. Therefore, unless there are positive controls with known binding to compare against, a variant's position at the top of the ranking does not indicate that the variant has an experimentally detectable binding affinity. Another limitation lies in APPRAISE's assumption that the binding of competing proteins is mutually exclusive. Counterexamples arise if the competing proteins exhibit cooperative binding or attach to epitopes situated at a considerable distance. Furthermore, certain candidate proteins may exhibit a tendency to interact with one another rather than with the designated receptor. Additionally, the geometrical scores utilized in APPRAISE 1.1+ were computed assuming the receptor has a predominantly convex structure. Thus, these scores are most effective when applied to single protein domains with convex shapes.

Other limitations of APPRAISE may arise from the protein structure prediction engine that it relies on. For example, ESMFold-APPRAISE fails when the language-model-based structure prediction tool cannot properly fold the protein in a complex (Figure 3.8b). At the same time, AF-Multimer-APPRAISE results can be biased by the specific selection of multiple-sequence alignments due to the dependence on co-evolutionary information by AF-Multimer. So far, the APPRAISE pipeline's ability to accurately rank IgG antibodies is limited due to the complexity of predicting antibody-antigen complex structures. Moreover, the accuracy and speed of APPRAISE may be compromised when the modeled proteins contain long disordered regions or large domains that are unnecessary for binding. As a result, pre-screening of several truncated protein constructs for minimal folding domains with the particular structure prediction tool (analogous to the common practice in structural biology) is always helpful. Additionally, the APPRAISE method is ineffective in ranking weak binders in a pool (e.g., Figure 3.3g), perhaps because the predicted structures do not offer many opportunities for meaningful interaction, resulting in

near-zero competition scores. Fortunately, this should not be a practical concern for most protein engineering applications since the most valuable candidates usually bind with higher affinities. Considering these limitations, it is essential to conduct spot checks on model results to confirm their physical soundness.

While APPRAISE has succeeded in ranking the binding propensities of different protein variants, its accuracy and speed could be further improved. For example, the parameters of APPRAISE 1.2's scoring function have only been minimally tuned to proper orders of magnitudes to decrease the risk of over-fitting (Methods). With further fine-tuning of parameters and the ever-growing power of protein structure prediction, the APPRAISE method promises to streamline the process of engineering protein-based therapeutics.



c

Variant name	Ref.	Modeled peptide sequence	C57BL/6J brain?	BALB/c brain?	Inferred LY6A dependency
PHP.B	Deverman et al.	AQTLAVPFKAQAQTG	✓	×	True*
PHP.eB	Chan et al.	DGTLAVPFKAQAQTG	✓	×	True**
PHP.B2	Deverman et al.	AQSVSKPFLAQAQTG	✓	×	True*
PHP.B3	Deverman et al.	AQFTLTTPKAQAQTG	✓	×	True*
PHP.N	Kumar et al.	AQTLAVPFSNPAQTG	✓	×	True
PHP.V1	Kumar et al.	AQTALKPFLAQAQTG	✓	×	True
PHP.V2	Kumar et al.	AQTTLKPFLAQAQTG	✓	×	True
PHP.C1	Kumar et al.	AQRYQGDSVAQAQTG	✓	✓	False
PHP.C2	Kumar et al.	AQWSTNAGYAQAQTG	✓	✓	False
PHP.C3	Kumar et al.	AQERVGFAQAQAQTG	✓	✓	False
AAV9	Gao et al.	AQ-------AQAQTG	×	×	False**
MacPNS1	Chen et al.	AQPHEGSSRAQAQTG	×	-	False
MacPNS2	Chen et al.	AQPNASVNSAQAQTG	×	-	False
NB-1	Kumar et al.	AQSLDTTKGAQAQTG	×	-	False
NB-2	Kumar et al.	AQESRGLSLAQAQTG	×	-	False
NB-3	Kumar et al.	AQMIRGLDTAQAQTG	×	-	False
NB-4	Kumar et al.	AQNGDLTRSAQAQTG	×	-	False
NB-5	Kumar et al.	AQTKREIDYAQAQTG	×	-	False
NB-6	Kumar et al.	AQMHHGMEIAQAQTG	×	-	False
NB-7	Kumar et al.	AQTPRMDTFAQAQTG	×	-	False
NB-8	Kumar et al.	AQVHRDLNFAQAQTG	×	-	False
NB-9	Kumar et al.	AQTARKEEFAQAQTG	×	-	False

* LY6A binding had been confirmed by cell binding assay
** LY6A dependency had been biochemical and genetic assays

Figure 3.5: **Prior experimental studies revealing the receptor dependency of some brain-transducing AAV variants. a,** A schematic showing an AAV capsid binding to a blood-brain barrier receptor (BBB) receptor that is only expressed at a high level on the endothelial cells of certain mouse strains (Huang et al., 2019; Hordeaux et al., 2019; Batista et al., 2019). **b,** A schematic showing how we can infer whether an AAV capsid can use LY6A, a mouse BBB receptor, by characterizing its brain transduction across different strains. A capsid with strain-restricted brain transduction in C57BL/6J mice is likely LY6A-dependent, while a capsid with cross-strain brain transduction or does not transduce the brain is not LY6A-dependent. **c,** A table summarizing all 22 capsids used in Figure 3.2 with their source literature, sequence, brain transduction profile, and inferred LY6A dependency.

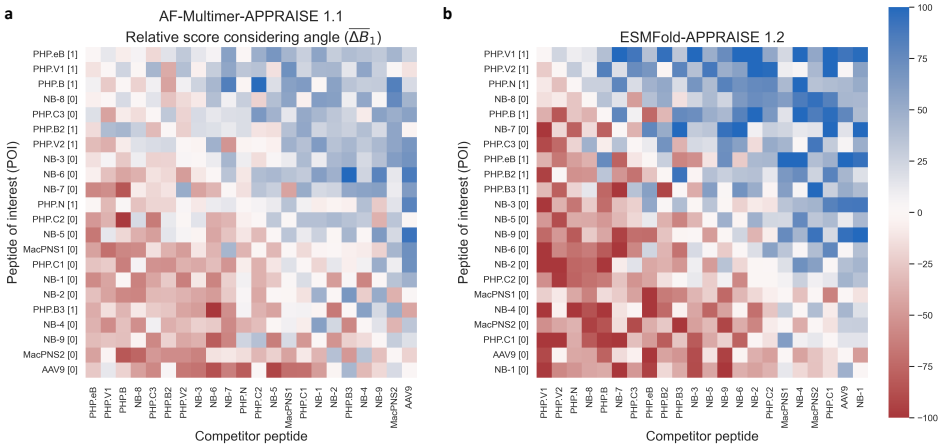


Figure 3.6: **Heatmaps representing score matrices of AF-Multimer-APPRAISE 1.1 and ESMFold-APPRAISE 1.2.**

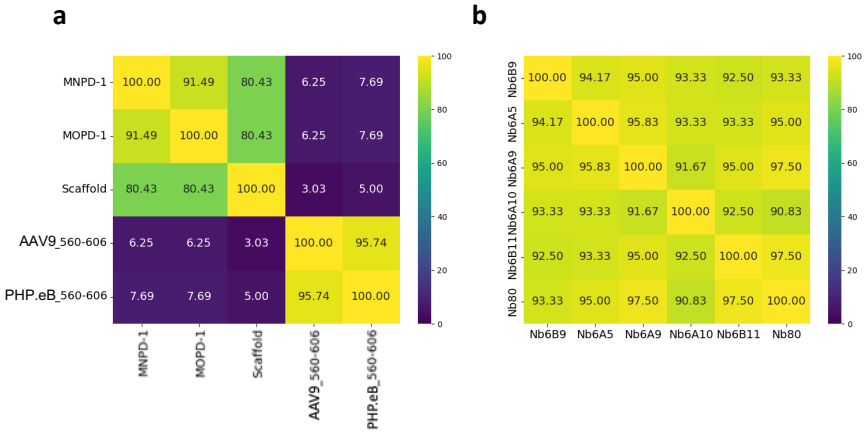


Figure 3.7: **Sequence identity between some engineered proteins.** Sequence identity matrix generated with Clustal Omega (Sievers et al., 2011) for **a)** PD-L1-binding peptides and **b)** β_2 adrenergic receptor-binding nanobodies, demonstrating that these proteins share similar sequences.

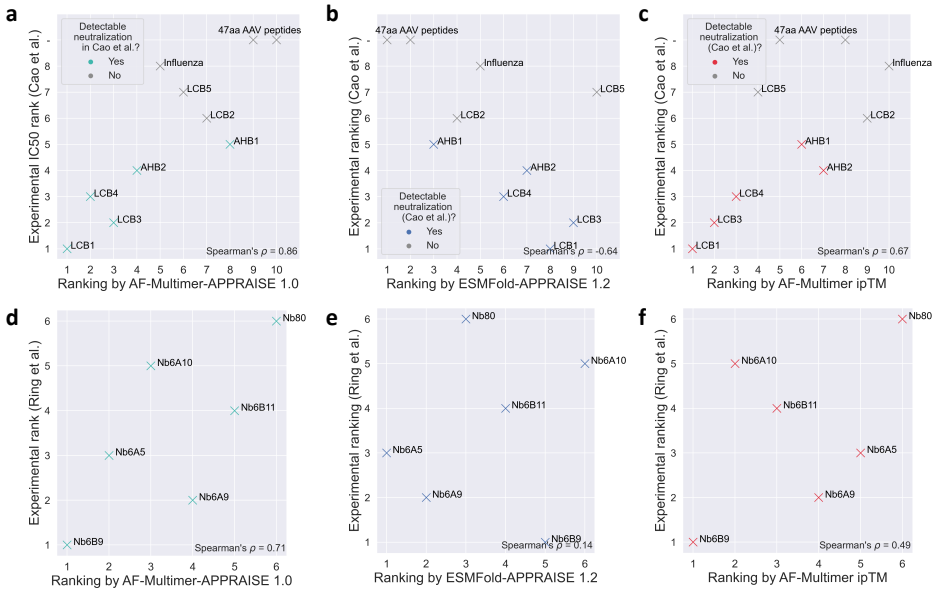


Figure 3.8: **Ranking protein binders using alternative methods.** Rankings of two groups of peptides analyzed in Figure 3.3g, j based on **a, d)** AF-Multimer-APPRAISE 1.0, **b, e)** ESMFold-APPRAISE 1.2, and **c, f)** interface pTM given by AlphaFold-Multimer.

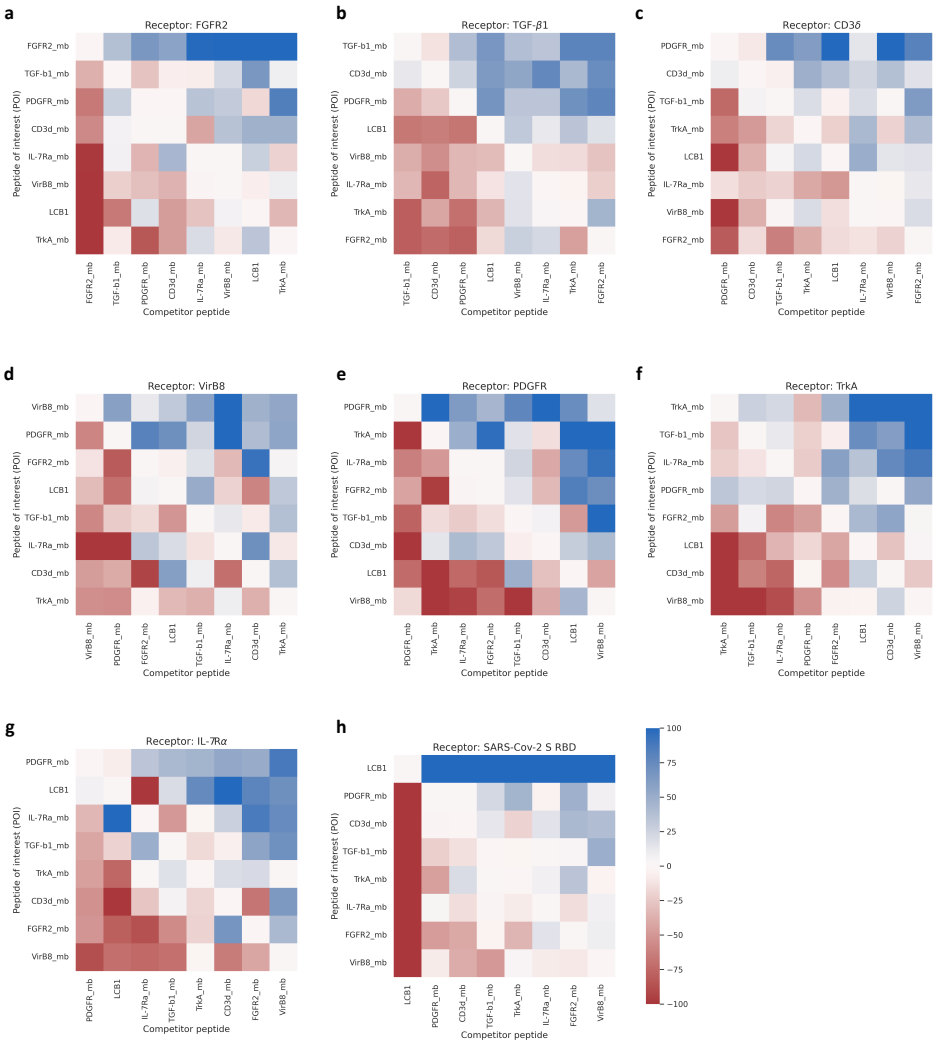


Figure 3.9: **Score matrices for APPRAISE rankings of miniprotein binders with individual receptors.** Detailed rankings of miniproteins with individual receptors summarized in Figure 3.3k using AF-Multimer-APPRAISE 1.2. The miniprotein sequences were from Cao et al., 2022

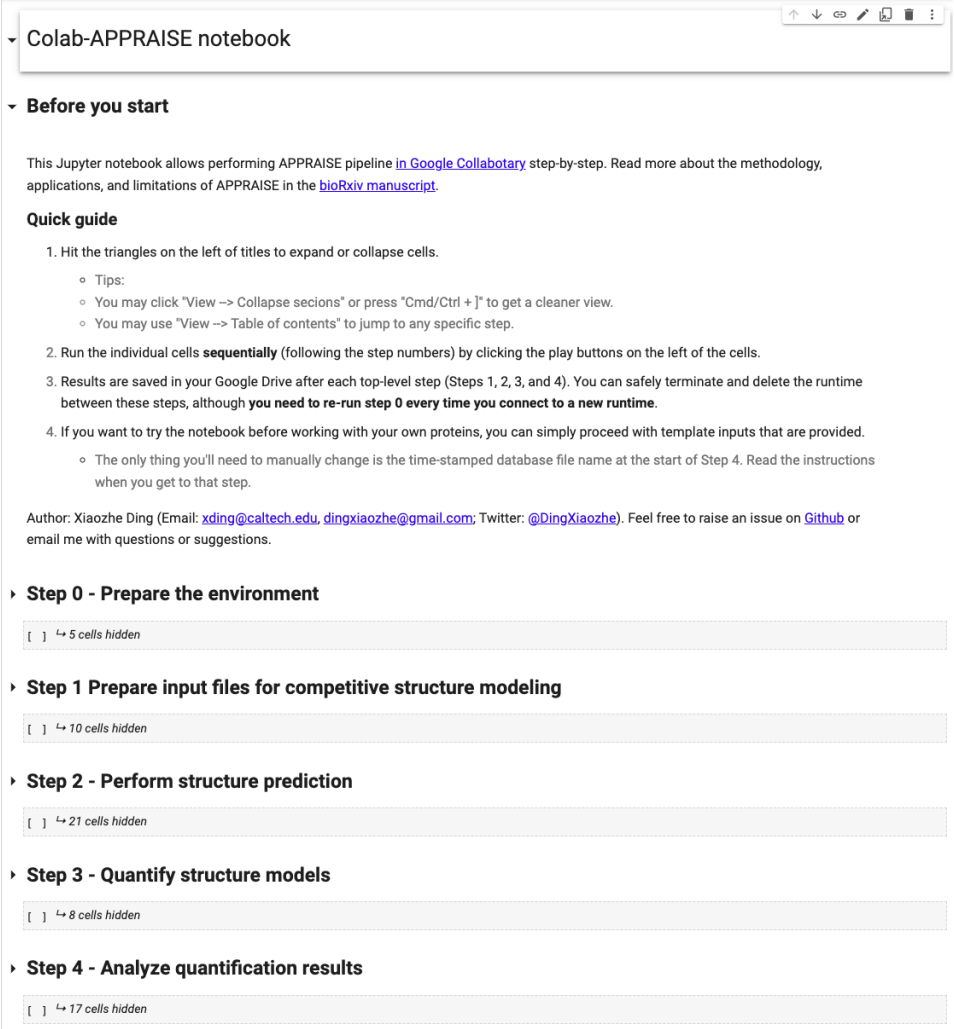


Figure 3.10: **A screenshot of the interface of Colab-APPRAISE.** APPRAISE can be easily accessed by running a web-based notebook on Google Colaboratory (`https://tiny.cc/APPRAISE`).

Table 3.1: Receptor sequences and parameters used for APPRAISE analysis (part 1 of 2)

Receptor name	Organism	Uniprot accession ID	Domain used for modeling	Residue indices	D_{max} (Å)	Axial ratio	R_{minor} (Å)	Anchor site	Sequence used for modeling
LY6A	Mus musculus (Mouse)	P05533	Mature protein	27-110	46.68	1.74	13.4	C-term	LECYQCYGVPFETSCPSITCPY PDGVCVTQEAAVIVDSQTRKVK NNLCLPICPPNIESMEILGTKV NVKTSCCQEDLCNVAVP
PD-L1	Homo sapiens (Human)	Q9NZQ7	V domain	18-132	46.50	1.51	15.4	C-term	AFTVTVPKDLYVVEYGSNMTIE CKFPVEKQLDLAALIVYWEMED KNIIQFVHGEEDLKVQHSSYRQ RARLLKDQLSLGNAALQITDVK LQDAGVYRCMISYGGADYKRIT VKVNA
Beta-2 adrenergic receptor	Homo sapiens (Human)	P07550	TM1-TM7	29-342	93.8	2.23	21.0	N-term	DEVWVVGMGIVMSLIVLAIVFG NVLVITAIAKFERLQTVTNYFI TSLACADLVMGLAVVPFGAAHI LMKMWTFGNFWCEFWTSIDVLC VTASIETLCVIAVDRYFAITSP FKYQSLLTKNKARVIILMVWIV SGLTSFLPIQMHWYRATHQEAI NCYANETCCDFFTNQAYAIASS IVSFYVPLVIMVFVYSRVFQEA KRQLQKIDKSEGRFHVQNLSQV EQDGRTGHGLRRSSKFCLKEHK ALKTLGIIMGTFTLCWLPFFIV NIVHVIQDNLIRKEVYILLNWI GYVNSGFNPLIYCRSPDFRIAF QELLCL
Transferrin receptor 1	Homo sapiens (Human)	P02786	Ectodomain	122-760	86.50	1.47	29.4	N-term	LYWDDLKRKLSEKLDSTDFTGT IKLLNENSYVPREAGSQKDENL ALYVENQFREFKLSKVWRDQHF VKIQVKDSAQNSVIIVDKNGRL VYLVENPGGYVAYSKAATVTGK LVHANFGTKKDFEDLYTPVNGS IVIVRAGKITFAEKVANAESLN AIGVLIYMDQTKFPIVNAELSF FGHAHLGTGDPYTPGFPSFNHT QFPPSRSSGLPNIPVQTISRAA AEKLFGNMEGDCPSDWKTDSTC RMVTSESKNVKLTVSNVLKEIK ILNIFGVIKGFVEPDHYVVVGA QRDAWGPGAAKSGVGTALLLKL AQMFSDMVLKDGFQPSRSIIFA SWSAGDFGSVGATEWLEGYLSS LHLKAFTYINLDKAVLGTSNFK VSASPLLYTLIEKTMQNVKHPV TGQFLYQDSNWASKVEKLTLDN AAFPFLAYSGIPAVSFCFCEDT DYPYLGTTMDTYKELIERIPEL NKVARAAAEVAGQFVIKLTHDV ELNLDYERYNSQLLSFVRDLNQ YRADIKEMGLSLQWLYSARGDF FRATSRLTTDFGNAEKTDRFVM KKLNDRVMRVEYHFLSPYVSPK ESPFRHVFWGSGSHTLPALLEN LKLRKQNNGAFNETLFRNQLAL ATWTIQGAANALSGDVWDIDNE F
Spike	SARS-CoV-2	P0DTC2	RBD	331-529	67.93	1.87	18.2	C-term	NITNLCPFGEVFNATRFASVYA WNRKRISNCVADYSVLYNSASF STFKCYGVSPTKLNDLCFTNVY ADSFVIRGDEVRQIAPGQTGKI ADYNYKLPDDFTGCVIAWNSNN LDSKVGGNYNYLYRLFRKSNLK PFERDISTEIYQAGSTPCNGVE GFNCYFPLQSYGFQPTNGVGYQ PYRVVVLSFELLHAPATVCGPK K

Table 3.2: Receptor sequences and parameters used for APPRAISE analysis (part 2 of 2)

Receptor name	PDB ID	Chain ID	Regions used for modeling	Construct length	D_{max} (Å)	Axial ratio	R_{minor} (Å)	Anchor site	Sequence used for modeling
FGFR2	1DJS	A	Structured region	216	105.2	3.06	17.19	C-term	TLEPEGAPYWTNTEKMEKRLHA VPAANTVKFRCPAGGNPMPTMR WLKNGKEFKQEHRIGGYKVRNQ HWSLIMESVVPSDKGNYTCVVE NEYGSINHTYHLDVVERSPHRP ILQAGLPANASTVVGGDVEFVC KVYSDAQPHIQWIKHVEKNGSK YGPDGLPYLKVLKAAGVNTTDK EIEVLYIRNVTFEDAGEYTCLA GNSIGISFHSAWLTVLPA
TGF-β1	3KFD	A	Structured region	112	65.88	2.4	13.73	C-term	ALDTNYCFSSTEKNCCVRQLYI DFRKDLGWKWIHEPKGYHANFC LGPCPYIWSLDTQYSKVLALYN QHNPGASAAPCCVPQALEPLPI VYYVGRKPKVEQLSNMIVRSCK CS
CD3δ	1XIW	B	Structured region	74	48.7	2.06	11.82	C-term	MKIPIEELEDRVFVNCNTSITW VEGTVGTLLSDITRLDLGKRIL DPRGIYRCNGTDIYKDKESTVQ VHYRMCQS
VirB8	4O3V	A	Cytoplasmic domain	143	56.27	1.62	17.37	C-term	ANPYISVANIMLQNYVKQREKY NYDTLKEQFTFIKNASTSIVYM QFANFMNIDNSLSPVIRYQKLY RRSINIISINNINNNEATVTFE SLAQNNTGEILENMLWEAKIGF IMDSISTSTLHNMPFHFIVTSY KLKLLRNKNQQ
PDGFR	3MJG	C	Ig-like C2-type domains 2&3	202	86.04	2.33	18.46	C-term	DERKRLYIFVPDPTVGFLPNDA EELFIFLTEITEITIPCRVTDP QLVVTLHEKKGDVALPVPYDHQ RGFSGIFEDRSYICKTTIGDRE VDSDAYYVYRLQVSSINVSVNA VQTVVRQGENITLMCIVIGNEV VNFEWTYPRKESGRLVEPVTDF LLDMPYHIRSILHIPSAELEDS GTYTCNVTESVNDHQDEKAINI TVVE
TrkA	2IFG	A	Ig-like C2-type domain 2	100	50.52	1.8	14.03	C-term	SFPASVQLHTAVEMHHWCIPFS VDGQPAPSLRWLFNGSVLNETS FIFTEFLEPAANETVRHGCLRL NQPTHVNNGNYTLLAANPFGQA SASIMAAFMDNP
IL-7Rα	3DI3	B	Structured region	193	70.65	2.07	17.07	C-term	DYSFSCYSQLEVNGSQHSLTCA FEDPDVNTTNLEFEICGALVEV KCLNFRKLQEIYFIETKKFLLI GKSNICVKVGEKSLTCKKIDLT TIVKPEAPFDLSVVYREGANDF VVTFNTSHLQKKYVKVLMHDVA YRQEKDENKWTHVNLSSTKLTL LQRKLQPAAMYEIKVRSIPDHY FKGFWSEWSPSYYFRTP

Table 3.3: Sequences of engineered proteins used in APPRAISE tests in Figure 3.3

Protein Name	Description	Source	Protein sequence
12aa-B	Peptide from a phage display selection for Transferrin binding (positive)	Lee et al. 2001	THRPPMWSPVWP
12aa-B-scramble	Scrambled sequence of 12aa-B	Lee et al. 2001	PWRPSHPVWMPT
12aa-NB1	Peptide from a phage display selection for Transferrin binding (negative)	Lee et al. 2001	SSHMENTPDTLR
12aa-NB2	Peptide from a phage display selection for Transferrin binding (negative)	Lee et al. 2001	SNIRLSNSPMNT
12aa-NB3	Peptide from a phage display selection for Transferrin binding (negative)	Lee et al. 2001	YSYTPHATSMYS
12aa-NB4	Peptide from a phage display selection for Transferrin binding (negative)	Lee et al. 2001	SDMYPSTTLPVI
MNPD-1	Peptide rationally designed to bind to PD-L1	Yin et al. 2021	AQIREYKRCGQDEERVRRECKERGERQNCVYNIYKE GNCYVCGIICL
MOPD-1	Peptide rationally designed to bind to PD-L1	Yin et al. 2021	AQIREYKRCGQDEERVRRECKERGERQNCVYNIYKE GNCYVCGIICL
Scaffold	Scaffold of the designed peptides	Yin et al. 2021	GSEERRYKRCGQDEERVRRECKERGERQNCQYQIRK EGNCYVCEIRC
AAV9_560-606	Negative control sequence for PD-L1 binding (from AAV capsid)	negative control	ITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGW VQNQGILPGMV
PHP.eB_560-606	Negative control sequence for PD-L1 binding (from AAV capsid)	negative control	ITNEEEIKTTNPVATESYGQVATNHQSDGTLAVPFK AQAQTGWVQNQGILPGMV
AHB1	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	DEDLEELERLYRKAEEVAKEAKDASRRGDDERAKEQ MERAMRLFDQVFELAQELQEKQTDGNRQKATHLDKA VKEAADELYQRVR
AHB2	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	ELEEQVMHVLDQVSELAHELLHKLTGEELERAAYFN WWATEMMLELIKSDDEREIREIEEEARRILEHLEEL ARK
Influenza	Miniproteins designed to bind to Influenza virus	Chevalier et al., 2017	CIEQSFTTLFACQTAAEIWRAFGYTVKIMVDNGNCR LHVC
LCB1	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	DKEWILQKIYEIMRLLDELGHAEASMRVSDLIYEFM KKGDERLLEEAERLLEEVER
LCB2	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	SDDEDSVRYLLYMAELRYEQGNPEKAKKILEMAEFI AKRNNNEELERLVREVKKRL
LCB3	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	NDDELHMLMTDLVYEALHFAKDEEIKKRVFQLFELA DKAYKNNDRQKLEKVVEELKELLERLLS
LCB4	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	QREKRLKQLEMLLEYAIERNDPYLMFDVAVEMLRLA EENNDERIIERAKRILEEYE
LCB5	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	SLEELKEQVKELKKELSPEMRRLIEEALRFLEEGNP AMAMMVLSDLVYQLGDPRVIDLYMLVTKT
AAV9_560-606	Negative control sequence for SARS-Cov-2-S binding (from AAV capsid)	Cao et al. 2020	ITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGW VQNQGILPGMV
PHP.eB_560-606	Negative control sequence for SARS-Cov-2-S binding (from AAV capsid)	Cao et al. 2020	ITNEEEIKTTNPVATESYGQVATNHQSDGTLAVPFK AQAQTGWVQNQGILPGMV
Nb80	Nanobody that binds to beta2 adrenergic receptor (parent)	Ring et al. 2013	QVQLQESGGGLVQAGGSLRLSCAASGSIFSINTMGW YRQAPGKQRELVAAIHSGGSTNYANSVKGRFTISRD NAANTVYLQMNSLKPEDTAVYYCNVKDYGAVLYEYD YWGQGTQVTVSS
Nb6A10	Nanobody selected to bind to beta2 adrenergic receptor using yeast display	Ring et al. 2013	QVQLQESGGGLVQAGGSLRLSCAASGGIFGFNTMGW YRQAPGKQRELVAAILSGGTTYYANSVKGRFTISRD NAANTVYLQMNSLKPEDTAVYYCNVKDHGSIIYDYD YWGQGTQVTVSS
Nb6B11	Nanobody selected to bind to beta2 adrenergic receptor using yeast display	Ring et al. 2013	QVQLQESGGGLVQAGGSLRLSCAASGTIFSINTMGW YRQAPGKQRELVAAIHSGGSTYYANSVKGRFTISRD NAANTVYLQMNSLKPEDTAVYYCNVKDYGAVLYDYD YWGQGTQVTVSS
Nb6A5	Nanobody selected to bind to beta2 adrenergic receptor using yeast display	Ring et al. 2013	QVQLQESGGGLVQAGGSLRLSCAASGSIFTFNTMGW YRQAPGKQRELVAAIHSGGNTDYANSVKGRFTISRD NAANTVYLQMNSLKPEDTAVYYCNVKDYGAIIYEYD YWGQGTQVTVSS
Nb6A9	Nanobody selected to bind to beta2 adrenergic receptor using yeast display	Ring et al. 2013	QVQLQESGGGLVQAGGSLRLSCAASGSIFAINTMGW YRQAPGKQRELVAAIHSGGSTNYANSVKGRFTISRD NAANTVYLQMNSLKPEDTAVYYCNVKDYGAIVYEYD YWGQGTQVTVSS
Nb6B9	Nanobody selected to bind to beta2 adrenergic receptor using yeast display	Ring et al. 2013	QVQLQESGGGLVQAGGSLRLSCAASGSIFALNIMGW YRQAPGKQRELVAAIHSGGTTNYANSVKGRFTISRD NAANTVYLQMNSLKPEDTAVYYCNVKDFGAIIYDYD YWGQGTQVTVSS
FGFR2_mb	Miniprotein designed to bind to FGFR2	Cao et al. 2022	DRRKEMDKVYRTAFKRITSTPDKEKRKEVVKEATEQ LRRIAKDEEEKKKAAYMILFLKTLG
TGF-β1_mb	Miniprotein designed to bind to TGF-β1	Cao et al. 2022	HCTIEVVGVDPEKVEAIAAAYGAEVCEKDGKFEIHL DDPHSAESAAVAISVLTNRPVRLQC
CD3δ_mb	Miniprotein designed to bind to CD3δ	Cao et al. 2022	NHIACEIHNPEAAKEIAKVANVRRVYFIKQPGNRYF VLLKNADPEGVKKVRSKYNVRCVIRE
VirB8_mb	Miniprotein designed to bind to VirB8	Cao et al. 2022	NAEEITEKATLVGIEAWLLAKDEEQKKKVRTLNRQV KKLLQQNDLDQAKRVLDQLKSVLEDLKS
PDGFR_mb	Miniprotein designed to bind to PDGFR	Cao et al. 2022	DDERLATLAFRALIKRAGVKNLDVKVTNGKVRVTIT GRDQASFKALQLVFALARLGLQVQIDTR
TrkA_mb	Miniprotein designed to bind to TrkA	Cao et al. 2022	RDEIKERIFKAVVRAIVTGNPEQLKEAKKLLEKLKK LGRLDQDAKKFEKAIRQVEKRLRS
IL-7Rα_mb	Miniprotein designed to bind to IL-7Rα	Cao et al. 2022	SVIEKLRKLEKQARKQGDEVLVMLARMVLEYLEKGW VSEEDADESADRIEEVLKK
LCB1	Miniproteins designed to bind to SARS-Cov-2-S	Cao et al. 2020	DKEWILQKIYEIMRLLDELGHAEASMRVSDLIYEFM KKGDERLLEEAERLLEEVER

Table 3.4: Peptides used for *in silico* screening

Peptide name	Peptide sequence (residues 587-594 in VP1)	Original source
AAV9	AQ---AQAQTG	Gao et al., 2002
PHP.B	AQTLAVPFKAQAQTG	Deverman et al. 2016
PHP.D	AQWKNMGLQAQAQTG	Unpublished in vivo selection
SRK-1	AQLYHGGSTAQAQTG	Kumar et al. 2020
SRK-2	AQNNSVRQLAQAQTG	Kumar et al. 2020
SRK-3	AQVNSTRNVAQAQTG	Kumar et al. 2020
SRK-4	AQGNMTKFTAQAQTG	Kumar et al. 2020
SRK-5	AQTAIQPPKAQAQTG	Kumar et al. 2020
SRK-6	AQITTDQPFAQAQTG	Kumar et al. 2020
SRK-7	AQDTANTARAQAQTG	Kumar et al. 2020
SRK-8	AQTHDAQAWAQAQTG	Kumar et al. 2020
SRK-9	AQQPLAEEAAQAQTG	Kumar et al. 2020
SRK-10	AQTALANQKAQAQTG	Kumar et al. 2020
SRK-11	AQTGTERLSAQAQTG	Kumar et al. 2020
SRK-12	AQNGVTQSKAQAQTG	Kumar et al. 2020
SRK-13	AQWTEQRLVAQAQTG	Kumar et al. 2020
SRK-14	AQDTGLNNRAQAQTG	Kumar et al. 2020
SRK-15	AQPLPPTSIAQAQTG	Kumar et al. 2020
SRK-16	AQSDPGKFMAQAQTG	Kumar et al. 2020
SRK-17	AQTTMGTMLAQAQTG	Kumar et al. 2020
SRK-18	AQKQTQDSSAQAQTG	Kumar et al. 2020
SRK-19	AQLAHNSALAQAQTG	Kumar et al. 2020
SRK-20	AQVVPSTYRAQAQTG	Kumar et al. 2020
SRK-21	AQFRHLTGAAQAQTG	Kumar et al. 2020
SRK-22	AQSANLLSSAQAQTG	Kumar et al. 2020
SRK-23	AQFSNTHALAQAQTG	Kumar et al. 2020
SRK-24	AQFNSKLQLAQAQTG	Kumar et al. 2020
SRK-25	AQFKTNISAAQAQTG	Kumar et al. 2020
SRK-26	AQYPVPLKQAQAQTG	Kumar et al. 2020
SRK-27	AQHVNHMAPAQAQTG	Kumar et al. 2020
SRK-28	AQIVSNQMSAQAQTG	Kumar et al. 2020
SRK-29	AQPRPERMYAQAQTG	Kumar et al. 2020
SRK-30	AQNMKIQHVAQAQTG	Kumar et al. 2020
SRK-31	AQNTNVPAMAQAQTG	Kumar et al. 2020
SRK-32	AQSAQLRSSAQAQTG	Kumar et al. 2020
SRK-33	AQSHHEQVSAQAQTG	Kumar et al. 2020
SRK-34	AQGATGHLTAQAQTG	Kumar et al. 2020
SRK-35	AQHNLRDSIAQAQTG	Kumar et al. 2020
SRK-36	AQGPGTSFKAQAQTG	Kumar et al. 2020
SRK-37	AQSPPVQGLAQAQTG	Kumar et al. 2020
SRK-38	AQTLYNAIHAQAQTG	Kumar et al. 2020
SRK-39	AQLGDITGFAQAQTG	Kumar et al. 2020
SRK-40	AQGFNSMKPAQAQTG	Kumar et al. 2020
SRK-41	AQSNGLNGLAQAQTG	Kumar et al. 2020
SRK-42	AQVRIPGALAQAQTG	Kumar et al. 2020
SRK-43	AQDMGTDNLAQAQTG	Kumar et al. 2020
SRK-44	AQNYATKSQAQAQTG	Kumar et al. 2020
SRK-45	AQSVTTSHVAQAQTG	Kumar et al. 2020
SRK-46	AQTSGTDGIAQAQTG	Kumar et al. 2020
SRK-47	AQARTAHGYAQAQTG	Kumar et al. 2020
SRK-48	AQHSANMSKAQAQTG	Kumar et al. 2020
SRK-49	AQHDERANMAQAQTG	Kumar et al. 2020
SRK-50	AQNNFNASLAQAQTG	Kumar et al. 2020
SRK-51	AQSASLVSHAQAQTG	Kumar et al. 2020
SRK-52	AQAPRIDNAAQAQTG	Kumar et al. 2020
SRK-53	AQLTSSNALAQAQTG	Kumar et al. 2020
SRK-54	AQTLNSIRAAQAQTG	Kumar et al. 2020
SRK-55	AQSGTGRQQAQAQTG	Kumar et al. 2020
SRK-56	AQKTTLASGAQAQTG	Kumar et al. 2020
SRK-57	AQMRVNTEEAQAQTG	Kumar et al. 2020
SRK-58	AQFETLHKTAQAQTG	Kumar et al. 2020
SRK-59	AQTQHRFEMAQAQTG	Kumar et al. 2020
SRK-60	AQHTAEKAPAQAQTG	Kumar et al. 2020
SRK-61	AQNHMVRELAQAQTG	Kumar et al. 2020
SRK-62	AQRFQPSSAAQAQTG	Kumar et al. 2020
SRK-63	AQRSVANVPAQAQTG	Kumar et al. 2020
SRK-64	AQVFQATRTAQAQTG	Kumar et al. 2020
SRK-65	AQEQRTPSPAQAQTG	Kumar et al. 2020
SRK-66	AQGSSTASLAQAQTG	Kumar et al. 2020
SRK-67	AQQVPHLHSAQAQTG	Kumar et al. 2020
SRK-68	AQPSQPYTKAQAQTG	Kumar et al. 2020
SRK-69	AQTHTRDQGAQAQTG	Kumar et al. 2020
SRK-70	AQINPGITLAQAQTG	Kumar et al. 2020
SRK-71	AQLQPTKSSAQAQTG	Kumar et al. 2020
SRK-72	AQQDAKVTTAQAQTG	Kumar et al. 2020
SRK-73	AQGASTHNAAQAQTG	Kumar et al. 2020
SRK-74	AQIPVSIQAAQAQTG	Kumar et al. 2020
SRK-75	AQVTSAHPVAQAQTG	Kumar et al. 2020
SRK-76	AQTASLIASAQAQTG	Kumar et al. 2020
SRK-77	AQDRGTRTVAQAQTG	Kumar et al. 2020
SRK-78	AQTAYLEVKAQAQTG	Kumar et al. 2020
SRK-79	AQATTQMSSAQAQTG	Kumar et al. 2020
SRK-80	AQKYDASQSAQAQTG	Kumar et al. 2020
SRK-81	AQTGTSHLHAQAQTG	Kumar et al. 2020
SRK-82	AQTMTPSGIAQAQTG	Kumar et al. 2020
SRK-83	AQTPSSSGNAQAQTG	Kumar et al. 2020
SRK-84	AQKDVVNSNAQAQTG	Kumar et al. 2020
SRK-85	AQRSPATMLAQAQTG	Kumar et al. 2020
SRK-86	AQYDQKSLAAQAQTG	Kumar et al. 2020
SRK-87	AQMGARNLPAQAQTG	Kumar et al. 2020
SRK-88	AQLPISATEAQAQTG	Kumar et al. 2020
SRK-89	AQTRHTSLTAQAQTG	Kumar et al. 2020
SRK-90	AQNKLTANGAQAQTG	Kumar et al. 2020
SRK-91	AQNGDSHSHAQAQTG	Kumar et al. 2020
SRK-92	AQVRTDMDMAQAQTG	Kumar et al. 2020
SRK-93	AQSVSTPRGAQAQTG	Kumar et al. 2020
SRK-94	AQVSRQFEPAQAQTG	Kumar et al. 2020
SRK-95	AQSANNVRGAQAQTG	Kumar et al. 2020
SRK-96	AQIGTKSTNAQAQTG	Kumar et al. 2020
SRK-97	AQGSELRTGAQAQTG	Kumar et al. 2020

3.5 Methods and Materials

Structure Modeling

Modeling of peptide-receptor complexes using AF2-multimer Peptide-receptor models are modeled using Colabfold (Python package index: alphafold-colabfold 2.1.14), an implementation of integrated multiple-sequence alignment generation with MMseqs2 and structure modeling with AF-Multimer-v2 (Mirdita et al., 2022; Steinegger et al., 2019; Mitchell et al., 2019; Mirdita, Steinegger, and Söding, 2019; Mirdita et al., 2017; Evans et al., 2021). First, batches of *.fasta files containing combined receptor sequences (Tables 3.1 and 3.2) and peptide sequences for the pairwise competition or pooled competition, where the protein chains are separated by the ":" symbol, are prepared using the function (appraise.input_fasta_prep.get_complex_fasta()). Second, the *.fasta files are used as input files for the "batch" Jupyter notebook in Colabfold package, and the notebook is run on Google Colaboratory using a V100 SXM2 16GB GPU or an A100 SXM4 40GB GPU. The settings used for the modeling are listed below:

```
msa_mode = "MMseqs2 (UniRef+Environmental)"
num_models = 5
num_recycles = 3
stop_at_score = 100
use_custom_msa = False
use_amber = False
use_templates = True
model_type = "auto" #or "AF2-multimer-v2"
```

Modeling of peptide-receptor complexes using ESMFold To model the peptide-receptor complexes using ESMFold, a process analogous to the one employed for AF2-multimer modeling is implemented. First, batches of *.fasta files containing combined receptor sequences (Table 3.1) and peptide sequences for the pairwise competition or pooled competition, where the protein chains are separated by a poly-glycine linker (30 glycine residues), are prepared using the same Python function(get_complex_fasta()) mentioned above. Second, the *.fasta files are used as input files for a custom Jupyter notebook with codes adapted from the Colabfold package for batch modeling using ESMFold, and the notebook is run on Google Colaboratory using an A100 SXM4 40GB GPU. The custom Colab notebook is included in the APPRAISE package:

`appraise/misc_utilities/ColabFold_ESMFold_batch_run.ipynb`.

Physics-informed analysis of individual structure models

The output folder containing *.pdb files generated by alphafold-colabfold is downloaded to a local computer for processing. Key parameters in a predicted structure model are measured, and the measurements are used to generate binding scores for each peptide in a model.

Automated quantification of the peptide-receptor models. The structure models are analyzed using PyMOL 2.3.3 using a custom PyMOL script. Briefly, the script loads all *.pdb models in a directory, extracts metadata from the file names, and measures the relevant contact atom numbers, angles, and distances. The measurements are saved as a *.csv file. The custom PyMOL script is included in the APPRAISE package:

`appraise/pymol_quantify_peptide_binding.py`.

Measurement of the R_{minor} of the receptor hull. The receptor shape parameter R_{minor}, which is necessary for APPRAISE 1.2, is obtained by measuring the shape parameters of an AlphaFold-modeled receptor structure. Briefly, the monomeric receptor (Table 3.1) is modeled using Colabfold (Python package index: alphafold-colabfold 2.1.14). The top model is then analyzed using HullRad v8.1(Fleming and Fleming, 2018) to obtain its major axis diameter D_{max} and aspect ratio P. R_{minor} is then calculated using the formula $R_{minor} = D_{max}/P/2$. Before the analysis, R_{minor} measurement is manually added as a column to the pandas dataframe storing PyMOL measurements with the column "R_minor".

Construction and calculation of $B_{energetic}$. We defined a contact atom as a non-hydrogen atom of either the receptor or the peptide within 5Å of the binding partner in the peptide-receptor model since atoms within this distance cutoff are responsible for most protein-protein interactions (Salamanca Viloria et al., 2017). We defined a clashing term as the number of non-hydrogen atoms in the peptide that are within 1Å of the receptor since this distance is smaller than the typical diameter of an atom and can cause a huge Van der Waals strain. To find the suitable weight for the clashing term, we estimated the relative energy scales using Lennard-Jones' potential. We concluded that an order of magnitude of 10^3 should be proper (Eq. 3.1). Since most interfaces between the engineered peptide and the receptor have up to a few hundred non-hydrogen atoms (tens of residues) in the interface, this heavy weight for the clashing atom practically sets the B_0^{POI} of any peptide with steric clashing

against the receptor to 0. Thus, Eq. 3.1 is practically equivalent to:

$$B_{energetic}^{peptide} = \begin{cases} N_{contact}^{peptide}, & \text{if } N_{clash}^{peptide} = 0 \\ 0, & \text{if } N_{clash}^{peptide} >= 1 \end{cases}$$

Construction and calculation of geometrical scores. The binding angle θ is defined as the angle between the vector from receptor center of mass to receptor anchor $\overrightarrow{OA}$ and the vector from receptor center of mass to peptide center of mass $\overrightarrow{OC'}$ (Figure **??**f. Note that the peptide center of mass C' is usually very close to the deepest point C, and therefore point C and point C' are undifferentiated in this schematic). A steep function is used to penalize inaccessible binding angles that are close to the anchor point:

$$B_{angle}^{peptide} = \begin{cases} -10^3 \cdot (1 - \frac{\theta}{\frac{\pi}{2}})^{10}, & \text{if } \theta < \frac{\pi}{2} \\ 0, & \text{if } \frac{\pi}{2} <= \theta <= \pi \end{cases} \tag{3.1}$$

The definition of binding depth d is a simplification of previously defined travel depth (Coleman and Sharp, 2006): we first calculate the hydrodynamic radius of the hull of the receptor at the minor axis (R_{minor}) using HullRad (Fleming and Fleming, 2018), and then take the difference of the distance between the "closest point on the peptide" to the receptor center and R_{minor}. The ratio between the difference and R_{minor} is defined as the depth. In other words, binding depth $d = \frac{\|OB\| - \|OC\|}{\|OB\|}$ where $\|OB\|$ is the minor axis radius (in Å) of the receptor hull when considering it as an ellipsoid (Figure **??**f), and $\|OC\|$ is the distance (in Å) between the center of mass of the receptor and the closest point on the peptide (Figure **??**f). An odd polynomial function is used to construct the score to reflect both the positive effect of a deep binding pocket and the negative effect of a convex binding site:

$$B_{depth}^{peptide} = 10^2 \cdot d^3 \tag{3.2}$$

Calculation of scores for each peptide in a model. The total binding scores for each peptide in a model are calculated using Eqs. 3.1-3.4 in the main text.

Generation of the score matrix and a ranking. The total binding scores of a POI vs. a competitor across 10 replicate models are averaged to get $\overline{\Delta B}^{POI,competitor}$ ($\overline{\Delta B_0}^{POI,competitor}$

for APPRAISE 1.0, $\overline{\Delta B}_1^{POI,competitor}$ for APPRAISE 1.1, or $\overline{\Delta B}_2^{POI,competitor}$ for APPRAISE 1.2). These averaged competition scores are then used to create a matrix and are plotted as a heatmap.

In the final score matrix, the POIs are ranked using a point-based round-robin tournament system (McCarthy and Benjamin, 1996) to avoid the bias caused by individual competitions with unusually high scores. Briefly, each $\overline{\Delta B}^{POI,competitor}$ in the matrix is considered as the match result between a POI and a competitor. A POI gains 1 point for winning over each match and loses 1 point for losing each match. (In the cases when $|\overline{\Delta B}^{POI,competitor}|$ does not reach the threshold of $p < 0.05$ using a one-sample, two-sided, Student's t test (degree of freedom=9), the match is called a tie, and the POI gets 0 points from the match.)

Code Availability

All codes are available in a GitHub repository:

`github.com/GradinaruLab/APPRAISE`

In addition, APPRAISE is made accessible through a web-based notebook interface using Google Colaboratory. The notebook can be found in the GitHub repository above or be directly accessed through the link:

`tiny.cc/APPRAISE`

The Colab-APPRAISE notebook includes pre-filled templates that can be used to demonstrate the workflow of APPRAISE. More demos can be found under `demo` folder in the GitHub repository.

Experimental Validations

***In vitro* infectivity assay.** HEK293T (ATCC, CRL-3216) cells were seeded in 6-well plates at 80% confluency and maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% fetal bovine serum (FBS), 1% non-essential amino acids (NEAA), and 100 U/mL penicillin-streptomycin at 37°C in 5% CO2. Cells were transiently transfected with 2.53 µg plasmid DNA encoding an expression cassette for the LY6A receptor. The following day, receptor-expressing cells were transferred to black, clear bottom 96-well plates at 20% confluency and maintained in FluoroBrite™ DMEM supplemented with 0.5% FBS, 1% NEAA, 100 U/mL penicillin-streptomycin, 1x GlutaMAX, and 15 µM HEPES at 37°C in 5% CO2. Engineered AAV variants packaging a CAG-mNeonGreen transgene were dosed in triplicate at 5E8 vg per well once the cells were attached. Plates were imaged 24

hours after AAV was introduced to cells with a Keyence BZ-X700 using a 4x objective and NucBlue™ Live ReadyProbes™ Reagent (Hoechst 33342) to autofocus each well.

***In vivo* mouse experiment.** For all the experiments performed in this study, the animals were randomly assigned, and the experimenters were not blinded while performing the experiments unless mentioned otherwise. All animal procedures in mice were approved by the California Institute of Technology Institutional Animal Care and Use Committee (IACUC), Caltech Office of Laboratory Animal Resources (OLAR) and were carried out in accordance with guidelines and regulations.

For the profiling of the novel AAVs in C57BL/6J mice (The Jackson Laboratory, 000664) and BALB/cJ mice (The Jackson Laboratory, 000651), the AAV vectors were injected intravenously via the retro-orbital route to 6-8 week-old adult mice at a dose of 3×10^{11} vg per mouse. Retro-orbital injections were performed as described previously (Challis et al., 2019). To harvest the tissues of interest after 3 weeks of expression, the mice were anesthetized with Euthasol (pentobarbital sodium and phenytoin sodium solution, Virbac AH) and transcardially perfused using 50 mL of 0.1 M phosphate-buffered saline (PBS) (pH 7.4), followed by 50 mL of 4% paraformaldehyde (PFA) in 0.1 M PBS. The organs were collected and post-fixed 24 h in 4% PFA at 4°C. Following this, the tissues were washed with 0.1 M PBS and stored in fresh PBS-azide (0.1 M PBS containing 0.05% sodium azide) at 4°C. Before imaging, the 100 µm tissue slices were cut using a Leica VT1000S. Brain images were acquired with a Zeiss LSM 880 confocal microscope using a Plan-Apochromat 10× 0.45 M27 (working distance 2.0 mm) objective. Zen Black 2.3 SP1 was used to process the images.

Ethical approval. All animal procedures in mice were approved by the California Institute of Technology Institutional Animal Care and Use Committee (IACUC), Caltech Office of Laboratory Animal Resources (OLAR) and were carried out in accordance with guidelines and regulations.

Materials Availability

The plasmid expressing the AAV-PHP.D capsid reported in this manuscript is deposited to Addgene (ID: 197055).

3.6 Acknowledgment

We thank Elisha Mackey, Zhe Qu, and Pat Anguiano for administrative assistance, Catherine Oikonomou for help with manuscript editing, and Sripriya R. Kumar, Seongmin Jang, Jimin Park, and Changfan Lin for helpful discussions. Schematics in this manuscript were created with BioRender.com. The study was funded by an NIH Director's Pioneer Award DP1OD025535 (to V.G.).

3.7 Related publication

Ding, Xiaozhe et al. (2023). "Fast, accurate ranking of engineered proteins by receptor binding propensity using structural modeling". In: *bioRxiv*. DOI: 10.1101/2023.01.11.523680.

Chapter 4

ENGINEERING A PROKARYOTIC REPRESSOR AS A SCAFFOLD FOR AN ORTHOGONAL GENETICALLY ENCODED TRANSMITTER INDICATOR

4.1 Abstract

The blood-brain barrier (BBB) presents a major challenge for delivering large molecules to study and treat the central nervous system. This is due in part to the scarcity of targets known to mediate BBB-crossing. To identify novel targets, we leverage a panel of adeno-associated viruses (AAVs) previously identified through mechanism-agnostic directed evolution for improved BBB transcytosis. Screening potential cognate receptors for enhanced BBB crossing, we identify two targets: murine-restricted LY6C1 and widely-conserved carbonic anhydrase IV (CA-IV; CA4). We apply AlphaFold-based in silico methods to generate capsid-receptor binding models to predict the affinity of AAVs for these identified receptors. Demonstrating how these new tools can unlock target-focused engineering strategies, we create an enhanced LY6C1-binding vector, AAV-PHP.eC, that, unlike our prior PHP.eB, also works in Ly6a-deficient mouse strains such as BALB/cJ . Combined with structural insights from computational modeling, the identification of primate-conserved CA-IV opens a path for designing more specific and potent human brain-penetrant chemicals and biologicals, including for gene delivery.

4.2 Introduction

Chemical transmission between neurons is crucial to the formation of neural circuitry and computation in the nervous system. Genetically encoded transmitter indicators (GETIs) that enable fluorescence imaging of such chemical transmissions by neurotransmitters, neuromodulators, and neurohormones have become powerful tools for neurobiologists (Andreoni, Davis, and Tian, 2019; Lin and Schnitzer, 2016; Patriarchi et al., 2019; Robinson et al., 2019; Wang, Jing, and Li, 2018). Most current GETIs are based on native membrane receptors, such as G-protein coupled receptors (GPCRs). In such GPCR-based GETIs, the native-like ligand-binding events at the GPCRs are coupled to either the transcription of a reporter protein (Lee et al., 2017) or direct conformational change of fluorescent proteins (Feng et al., 2019; Jing et al., 2018; Patriarchi et al., 2018; Sun et al., 2018). However, there are

inevitable concerns about the unpredictable influence on normal signaling pathways and cell physiology by the overexpression of such receptors.

One strategy to avoid such concerns is to use scaffolds that are "orthogonal" to the native signaling pathway. GETIs based on bacterial periplasmic binding proteins (PBPs) have been successfully developed for glutamate (Marvin et al., 2013; Marvin et al., 2018) and GABA (-aminobutyric acid) (Marvin et al., 2019). However, the PBP-based approach requires finding a natural bacterial PBP that natively binds to the specific transmitter molecule of interest in bacteria, and the lack of availability of such natural periplasmic binding proteins becomes a rate-limiting factor for the development (Marvin et al., 2019).

We hypothesized that prokaryotic repressors, a class of allosteric transcriptional regulators, can be engineered to be a new family of orthogonal sensing domains for GETIs because they have highly diverse and convertible ligand specificities (Dietrich, McKee, and Keasling, 2010; Tang and Cirino, 2011), a well-characterized allosteric sensing mechanism (Hinrichs et al., 1994; Orth et al., 2000; Reichheld, Yu, and Davidson, 2009), and a screenable functional readout (Ellefson, Ledbetter, and Ellington, 2018; Kimura et al., 2020).

As an example, the TetR family of repressors, a family of one-component transcriptional regulators with over 200,000 putative sequences, have been reported to sense a large variety of bioactive small molecules such as antibiotics, metabolites, quorum signaling molecules (Cuthbertson and Nodwell, 2013) with a variety of ligand binding affinity and specificity (data not shown). A typical TetR family repressor binds to a specific DNA sequence with its DNA-binding domain (DBD) in the absence of a specific "inducer" molecule. When an inducer molecule binds to the regulatory domain of the repressor, it triggers an allosteric conformational change in DBD which promotes dissociation of the repressor-DNA complex, allowing the transcription of the set of genes being regulated (Figure 4.1A). The TetR family members share the same protein fold, and their allosteric sensing mechanisms upon ligand binding events have been characterized in atomic details by over 200 protein structures in both bound and unbound states (Hinrichs et al., 1994; Orth et al., 2000) as well as circular dichroism (Reichheld, Yu, and Davidson, 2009) (Figure 4.1B).

In this study, we provide a proof-of-principle example of developing a GETI based on a TetR family bacterial repressor, TtgR, to sense a monoamine neurohormone, melatonin (N-acetyl-5-methoxy tryptamine). Melatonin, a molecule produced mainly by the pineal gland, carries out a diversity of roles including phasing circadian rhythms

and antagonizing inflammation and excitotoxicity by acting via GPCRs, nuclear receptors, or other proteins. The molecule has been related to not only the control of sleep, but also multiple diseases including metabolic syndromes, obesity, and type 2 diabetes (Hardeland et al., 2011; Reiter, Tan, and Fuentes-Broto, 2010). Despite its intriguing functions in neurophysiology, to date, there hasn't been a GETI reported detecting the melatonin molecule.

4.3 Results

Our development pipeline includes three stages (Figure 4.1C). 1.) Parent selection: identification of a natural parent repressor that shows promiscuous activity to the target molecule. 2.) Directed evolution in bacteria: mutate the repressor sequence and screen for variants with improved sensitivity and specificity for the target transmitter molecule with a fluorescence-based assay; in each round of evolution, new knowledge gained from structural analysis and machine learning models is incorporated into the diversification process to improve the efficiency of the evolution. 3.) Backbone swap and test in mammalian cells: the evolved repressor is plugged into a GETI backbone with a transcriptional reporter readout or an immediate fluorescence intensity readout; the GETI is then tested for response to the target transmitter in mammalian cells.

A critical component in the development pipeline is the rapid, high-throughput screening of repressor variants. We developed a fluorescent assay-based directed evolution method to improve the protein's sensitivity and specificity to melatonin. Random libraries of repressor variants are transformed into E.coli. Single colonies are picked into multi-well plates semi-automatically, and the grid colonies are replicated in multiple conditions to grow. In the backbone of the library plasmid (Rogers et al., 2015), the repressor variant also regulates the expression of a fluorescent reporter. Fluorescence intensities of each colony under different conditions are recorded. In each round of evolution, only variants with improvement in both specificity and induced fold change are chosen for the next steps (Figure 4.1D).

To identify a parent repressor for sensing melatonin, we performed a small-scale prescreening on 17 different repressors (Stanton et al., 2014a) against melatonin in a version of fluorescence-based screening in E.coli, and validated the promising hits from the prescreening by repeating the experiment with quadruplicates (Figure S1). One repressor protein, TtgR, stands out by showing a promiscuous, concentration-dependent response to melatonin (Figure 4.2A). We found that the wild-type TtgR

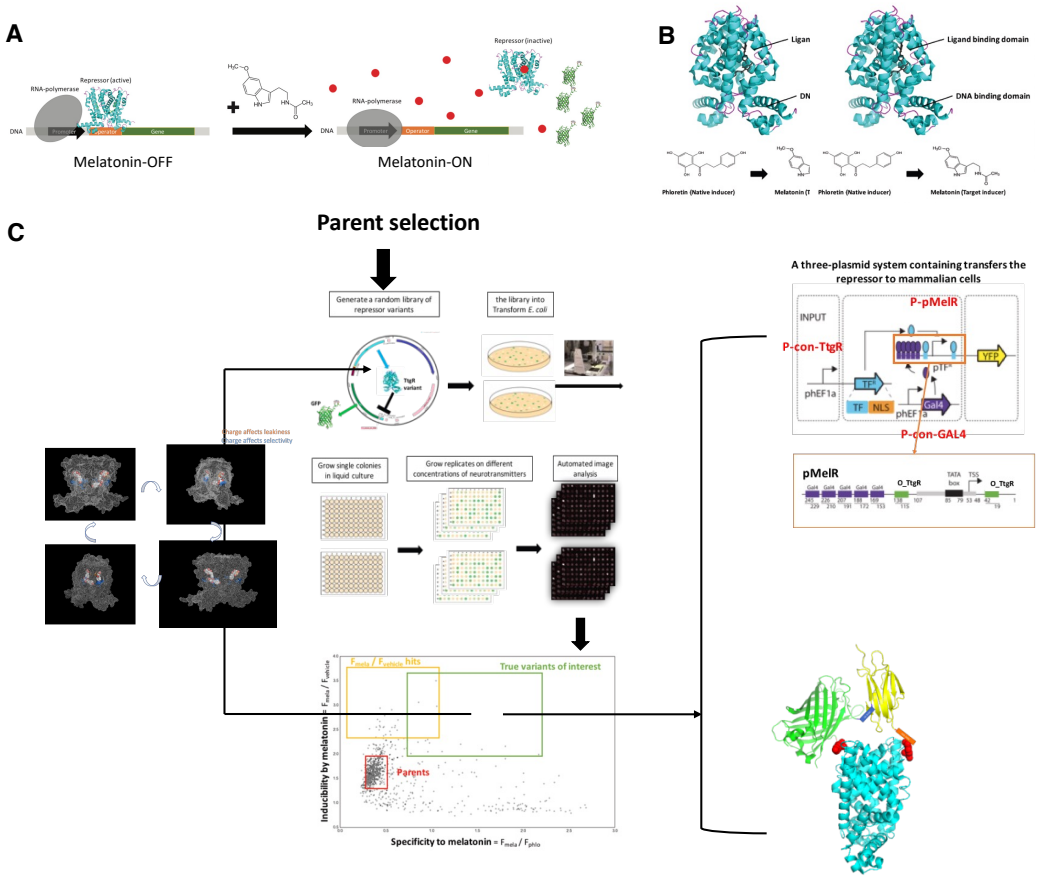


Figure 4.1: **Overview of the sensor development pipeline. A,** schematic showing the mechanism of genetic regulation of a melatonin-inducible repressor in bacteria. Melatonin binding promotes dissociation of the repressor-operator complex, thus promoting a transcriptional readout of melatonin presence. **B,** allosteric sensing mechanism of TetR family repressor. **C,** an overview of the development pipeline.

may cross-react, to a lower degree, with L-Dopa and arachidonic acid, but does not respond to a panel of other molecules including DOPAC, dopamine, GABA, glycine, glutamic acid, or serotonin (Figure S1). To improve the melatonin sensitivity of wild-type TtgR to a more physiologically relevant concentration. We evolved TtgR with the above-mentioned directed evolution approach with error-prone PCR to generate random mutations across the whole coding gene of TtgR. The first three rounds of evolution increased the specificity and induced fold change to melatonin. The fluorescent response to 250uM melatonin increased from 0 in wild-type TtgR to 1-fold in the hit of the third round of evolution (Figure 4.2B). Further evolution kept improving the induced fold change by melatonin (Data not shown).

To better understand which residues might contribute to the ligand specificity of

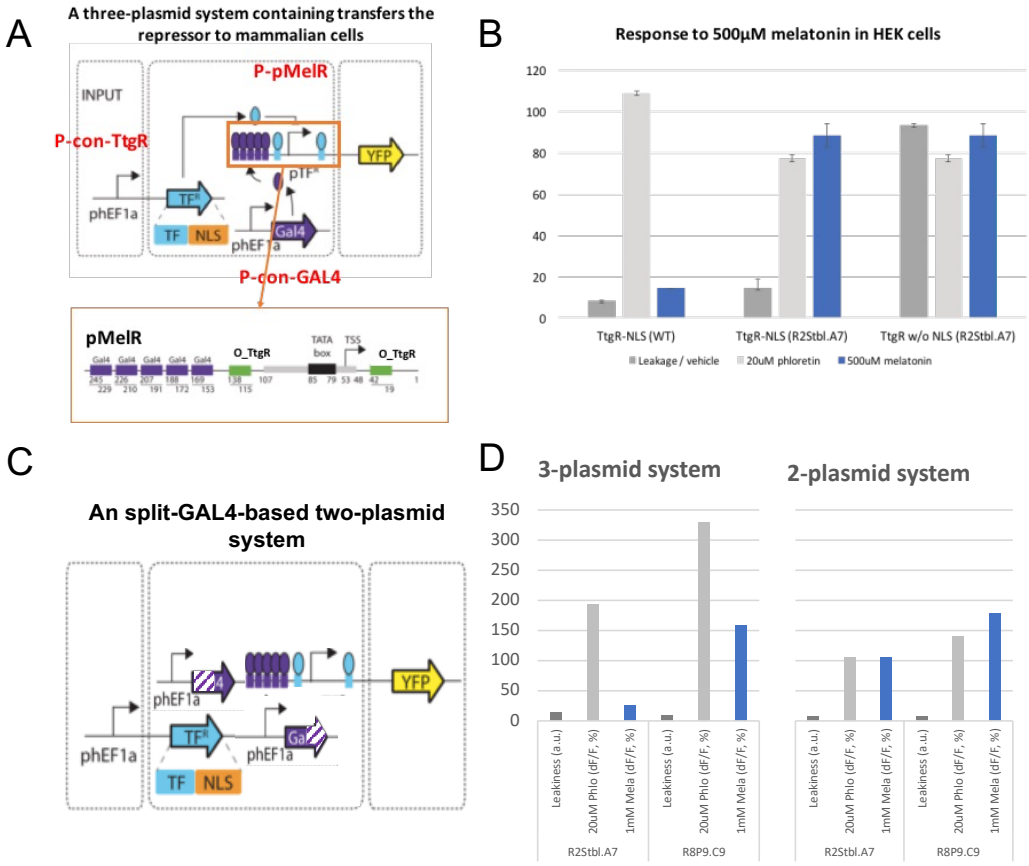


Figure 4.2: **Evolution of a prokaryotic repressor for sensing melatonin. A,** in an LB agar plate-based fluorescent assay, wild type TetR family repressor TtgR showed a concentration-dependent response to melatonin added to the culture medium. Each data point is one E.coli colony. **B,** fluorescence response, specificity, and leakage of the chosen repressors from the first three rounds of evolution, measured in the same LB agar plate-based assay as described in panel C. **C,** Computational structural model gives hints on the mechanism of the improvement from blindly-introduced mutations. The model was obtained by docking melatonin to the crystal structure of TtgR (PDB:2UIX) with RosettaLigand protocol. **D,** Linear SVC models revealed key amino acid features that determine repressors' melatonin-specificity and leakiness. Binding pocket amino acid sequences predict selectivity and leakiness, but not responsiveness of repressor variants.

TtgR, we performed computational docking of melatonin into the binding pocket of TtgR (PDB: 2UIX) with RosettaLigand protocol (Figure 4.2C). We hypothesized that the residues in close contact with the melatonin molecule and phloretin molecule (PDB:2UIX) within the ligand-binding domain might be critical for the repressor's specificity. These residues include H67, N110, G140, and V175. Based on the structural analysis, we hypothesize that combinatorial site saturation of these

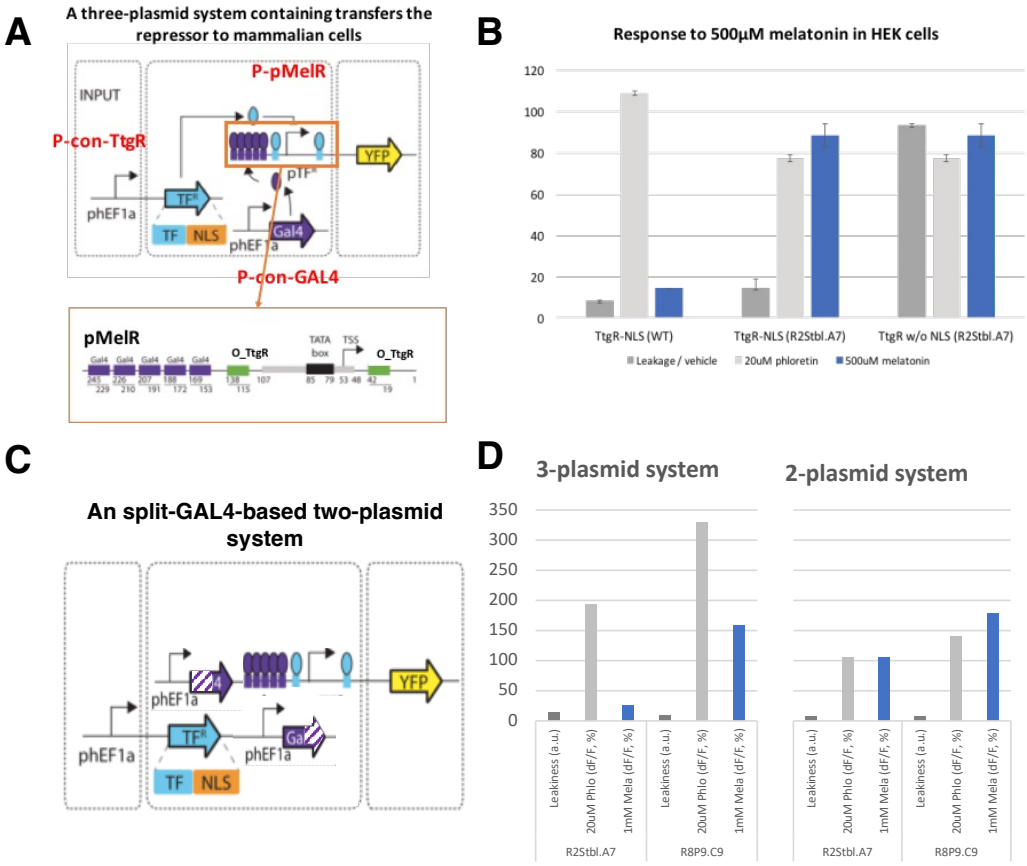


Figure 4.3: **The engineered repressor can be transferred to mammalian cells as a transcriptional sensor. A,** schematic of a 3-plasmid system containing a reporter plasmid (P-pMelR-EYFP), a repressor plasmid (P-con-TtgR) and an activator plasmid (P-con-GAL4-VP16) which transfers the repressor to mammalian cells, adapted from Stanton et al., 2014b. **B,** Average fluorescence (a.u.) of HEK293T cells transiently transfected with the 3-plasmids system measured by flow cytometry. Cells were treated in medium containing specified drugs for 24 hours before the flow cytometry. **C,** schematic of a 2-plasmid system based on split-GAL4 is designed to reduce the number of plasmids while maintaining the flexibility of tuning the repressor. **D,** average fluorescence (a.u.) of HEK293T cells transiently transfected with the 2-plasmids system measured by flow cytometry. Cells were treated in medium containing specified drugs for 24 hours before the flow cytometry.

residues will alter the volume and chemical properties of the binding pocket and in doing so, alter both its relative affinity for phloretin, its native substrate, and melatonin, the neurotransmitter of interest. In the screening of the library, we identified four mutations that are capable of dramatically changing the interaction and chemical properties of the binding pocket. The construct with these four mutations in initial screening showed a 35-fold improvement in δ and a 7-fold improvement

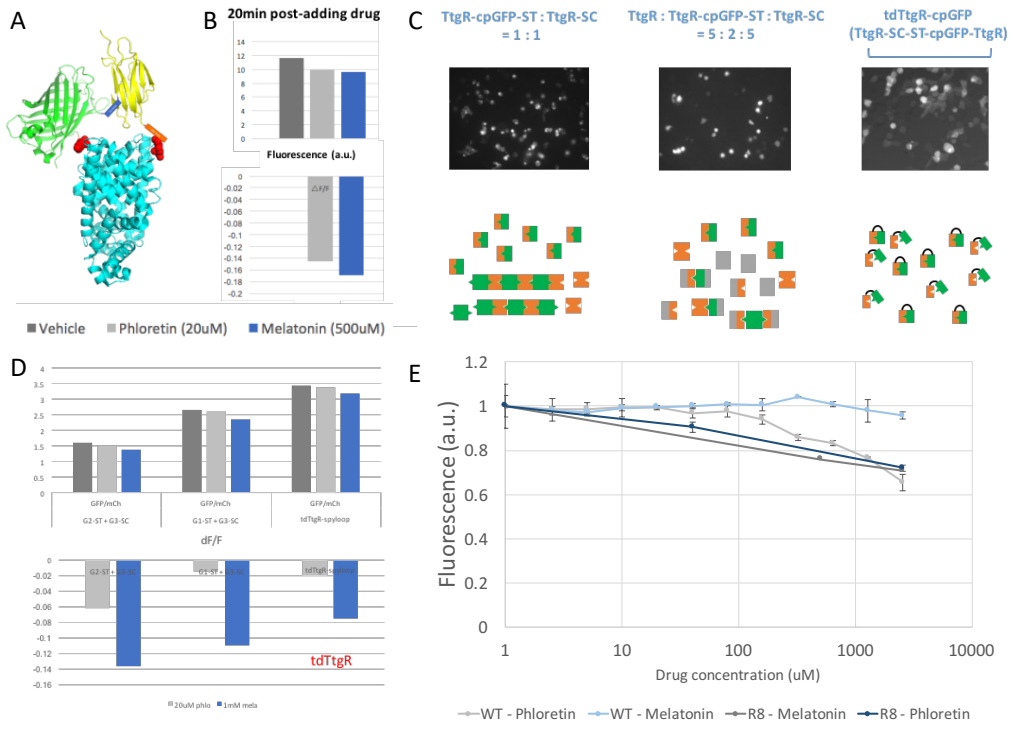


Figure 4.4: **TdTtgR, a genetic fusion of the repressor, a circularly permutated GFP and a SpyLoop, is a fluorescent indicator for melatonin. A,** schematic showing the hypothetical structure of dimerized TtgR-cpGFP-SpyTag and TtgR-SpyCatcher. **B,** average fluorescence (a.u.) of HEK293T cells transiently transfected with TtgR-cpGFP-SpyTag and TtgR-SpyCatcher by flow cytometry. Cells were treated in medium containing specified drugs for 20 minutes before the flow cytometry.**C,** representative images showing the fluorescent aggregations of cells expressing TtgR-cpGFP-SpyTag and TtgR-SpyCatcher (left), reduced aggregation with non-reactive TtgR blockers (middle), and minimal aggregation in tandom dimer TtgR (tdTtgR). **D,** average fluorescence (a.u.) of HEK293T cells transiently transfected with two versions of TtgR-cpGFP-SpyTag and TtgR-SpyCatcher pairs as well as tdTtgR. **E,** fluorescence of bacterial lysate expressing tdTtgR in response to different concentrations of drugs. Fluorescence intensity is normalized to that of untreated lysate. Each data point is an average of at least 3 measurements.

in specificity for melatonin over phloretin. This was the first construct identified which showed a preference for melatonin over phloretin.

During the screening of the site-saturation mutagenesis library, we also generated a data set that couples the residue identities of the binding pocket amino acids and correspondent functional readouts. With interpretable machine learning models, this data set would allow us to understand the key features in the sequence level that determine the repressor functions(CB17; Yang, 2018). This is essential because

our combinatorial site-saturation library has too many constructs to screen it its entirety experimentally so being able to explore larger portions in silico would greatly expand our abilities to engineer the repressors. With linear SVC models, we identified three key amino acid features of the binding pocket residues (specify:) that account for most of the repressors' melatonin-specificity and leakiness (Figure 4.2D, left two boxes). (The mutations in the binding pocket predominantly changed the electrochemical properties of the pocket but there was no appreciable change in either the molecular weight of the amino acids present or their polarity.) However, these binding pocket amino acid sequences do not but not predict the induced fold changes of the repressor variants (Figure 4.2D, right box). This implies that the repressor protein has evolved to function modularly, and mutations in the binding pocket only affect inducer recognition, but not signal transduction.

One way to report neurotransmitter concentration with the engineered transmitter-responsive repressor is by a transcriptional readout. To build such a transcriptional sensor for melatonin in mammalian cells, we adapted a previously reported three-plasmid system (Stanton et al., 2014b) which contains a repressor plasmid (P-con-TtgR) that expresses the evolved melatonin-responsive TtgR, an activator plasmid (P-con-GAL4-VP16) which expresses GAL4-VP16, and a reporter plasmid (P-pMelR-EYFP) which expresses a fluorescent reporter under the regulation of both the repressor and the activator. (Figure 4.3A)

After optimization of the plasmid ratio for repressor: activator: reporter, HEK293T cells transfected with this three-plasmid system show inducible expression of the reporter fluorescent protein at the presence of melatonin. Engineered TtgR showed significantly improved fluorescence fold change by melatonin compared to wildtype TtgR. (Figure 4.3B) It is worth noting that the engineered TtgRs still show cross-response to phloretin, TtgR's natural inducer. However, to our best knowledge, phloretin is not expressed in animals, and thus the cross-response should not be a concern for its application.

Although having three separate plasmids in the system allows easy tuning of the ratio of components in the system, it also imposes a challenge on the co-delivery of the system. We thus designed a two-plasmid system based on split-GAL4 to reduce the number of plasmids while maintaining the flexibility of tuning the repressor-activator ratio. The gene expression cassettes encoding two halves of intein-based split-GAL4 (Wang et al., 2018) were inserted to the reporter plasmid (P-pMelR-EYFP) and the repressor plasmid (P-con-TtgR) respectively. (Figure 4.2C) After the optimization

of the plasmid ratio and usage of promoters, the two-plasmid system produced comparable inducibility to melatonin compared to the three-plasmid system.

Another way of reporting melatonin concentration is through direct coupling of melatonin binding to the conformational change of fluorescence proteins. One design to utilize the repressor as a ligand-binding domain is by expressing two fusion proteins based on TtgR: TtgR-cpGFP-SpyTag and TtgR-SpyCatcher. The two protein products would hypothetically react and dimerize to form Spyloop-cpGFP-dTtgR (d stands for "dimer"), which couples ligand-induced conformational change of TtgR (WT) directly to fluorescence change. Indeed, the Spyloop-cpGFP-TtgR design shows drug-induced fluorescence change as measured by flow cytometry. (Figure 4.4B)

However, the separately expressed TtgR-cpGFP-SpyTag and TtgR-SpyCatcher suffer from severe aggregations in the cells. This phenomenon is hypothetically caused by a chain reaction that oligomerizes the TtgR proteins. (Figure 4.4C left) Consistent with the hypothesis, overexpressing non-reactive TtgR helped to reduce aggregation by blocking chain reaction. (Figure 4.4C middle) A tandem-dimer design, tdTtgR appears to show minimal aggregation while maintaining its negative response to inducers in HEK293T cells. (Figure 4.4C right, Figure 4.4D) Furthermore, bacterial lysate expressing tdTtgR also showed a concentration-dependent response to phloretin and melatonin, validating that the repressor-based fluorescent sensor responds to its native inducer in vitro. The tdTtgR molecule with evolved TtgR (R8) showed a significantly larger response to melatonin compared to wildtype TtgR (WT). (Figure 4.4E)

4.4 Discussion

Here we presented a generalizable method to neurotransmitter-sensing repressor proteins, which can be used as an orthogonal ligand-binding module for building GETIs. Our data showed that TtgR, a natural repressor, can be evolved to be sensitive and specific to a chemical transmitter, melatonin. The melatonin-inducible transcriptional regulation system can be transferred to mammalian cells by a 3-plasmid or 2-plasmid system, allowing targeted expression of reporters or neural modulators. The evolved repressor can be further used as a melatonin-sensing module in a novel fluorescent GETI, tdTtgR.

To our knowledge, this is the first report of a GETI for melatonin. The current sensitivity of the evolved repressor to melatonin (100uM) is close to the physiolog-

ical concentrations of melatonin (10uM-200uM) in pineal glands in typical model organisms.

The directed evolution method, the neurotransmitter-inducible genetic regulation system and the fluorescent protein-based scaffold of a GETI presented here provide a promising platform for the development of GETIs and targeted expression of other chemical transmitters in the nervous system.

4.5 Methods

Plasmids

All plasmids were generated with standard molecular biology techniques: restriction enzyme digestion (New England Biolabs), Q5 polymerase PCR (New England Biolabs), and InFusion assembly (Takara). DNA was transformed into Stellar competent E. coli cells (Takara) with carbenicillin (100 µg/ml) selection. Plasmid DNA was isolated using plasmid miniprep or maxiprep kits (Qiagen). Sequences of the inserted region are verified by Sanger sequencing (performed by Laragen).

1. The expression vector used for screening in E.coli., pJKR-H-ttgR, was a gift from Dr. George Church (Addgene plasmid 62565). In round 8, a slightly modified version of pJKR, pJ3 was generated by adding a terminator (BBa_B0062 (Shetty, Endy, and Knight, 2008)) downstream of sfGFP coding region and adding an expression cassette for an mCherry reporter. 2. The three-plasmid expression system in mammalian cells are constructed based on the system described in previous literature(Stanton et al., 2014b). The constitutive repressor plasmid, P-con-TtgR, was synthesized (VectorBuilder) and assembled into a backbone with hEF1a promoter. The reporter plasmid, P-pMelR-EYFP, was cloned based on The constitutive GAL4 plasmid, P-con-GAL4 was a gift of Dr. Chris Voigt. 3. The two-plasmid expression system was cloned based on the constitutive repressor plasmid and the reporter plasmid, with two halves of split-intein inserted into each of the plasmid. 4. Inserts for the cpGFP or splitGFP fusion plasmids are synthesized (IDT or Qinglan Biotech) and cloned into pMV promoter.

Maps for all plasmids are included as a separate file. Plate-Based Binding Assay for small-scale screening for natural TetR repressors We designed a plate-based assay for comparing the induction levels of the parental constructs with the different neurotransmitters. BL21(DE3) bacteria (NEB cat. C2527H) were transformed with each construct. The vector has a promoter for the repressor gene which is repressed by a Lac repressor protein which is itself induced by IPTG and a binding site for

that repressor just upstream of the gene that transcribes YFP (Figure S1). In brief, individual colonies were selected and inoculated in 200 L cultures for 16 hours to generate homogeneous, saturated cultures of each construct. 1.5 L of this culture was then plated onto treated M9 minimal media plates (recipe for 1L: 200 mL of M9 salts, 20 mL of a 20% glucose solution, 2 mL of magnesium sulfate, 0.1 mL of Calcium chloride, and 780 mL of water) for minimal background auto-fluorescence of the media.

Each plate was treated prior to culture plating and dried at 37 °C for 1 hr for better absorption in later steps. Then a solution of neurotransmitters at different concentrations (and depending on the condition the corresponding amount of IPTG) was spread onto the surface of the plate and allowed to dry and diffuse into the plate. For each treatment condition (Table 2.2) the treatment was done for one IPTG+ M9 plate and one IPTG- M9 plate. The E. coli containing the appropriate construct were then aliquoted in 1.5 L droplets onto the plate spread about 2 cm apart. After treatment and seeding of bacteria, the plates were left in the 37 °C incubator overnight. The next day colonies were assayed for YFP fluorescence level by imaging all the plates on a blue light gel imager at fixed exposure and focus conditions. A python script was used for image processing to calculate the ratio of the fluorescence of IPTG+ plates to IPTG- plates. The difference between IPTG+ vs IPTG- plates for each condition was compared.

Sequence diversification by Error-prone PCR

Error-prone PCR was performed with 200 µM MnCl2 in addition to 1x Taq Polymerase Mastermix (NEB). Primers (Forward: 5'- CTCTACAAATAATTTTGTTTAACTTTGAAATAAGGAGGTAATACAAATG -3', Reverse: 5'- CATGTTGGTTTCCTACATTCAATTTTTTAGTCGCTTATTA -3', synthesized by IDT) were designed to have a predicted Tm of 57 ºC and 15 bp overhangs homologous to vector backbone. PCR reactions were run on a C1000 Touch Thermal Cycler (Biorad) for 25 cycles with an annealing temperature of 52. PCR products were purified with gel electrophoresis. Sequence diversification by Site-saturation mutagenesis Site-saturation libraries were made using fragment PCR. Primers (synthesized by IDT) are designed for each site to be saturatedly mutated. At each site, a pair of forward and reverse primers was designed to contain NNK at the codon to be mutated. Individual PCR were performed by the standard NEB Q5 protocol using the forward primer from the upstream site and the reverse primer from the nearest downstream site.

Fragment 1. forward: 5'- CTCTACAAATAATTTTGTTTAACTTTGAAATAAG-GAGGTAATACAAATG -3', reverse: 5'- GCGTTTCMNNCAGAGAATCCAGCAG -3'

Fragment 2. forward: 5'-CTGGATTCTCTGNNKGAAACGCATGATC-3', reverse: 5'- GATTTCMNNAATGCGACGGGTACG -3'

Fragment 3. forward: 5'- CGTACCCGTCGCATTNNKGAAATCC-3', reverse: 5'- CCAGGGTGATMNNTTTATGACAATCC -3'

Fragment 4. forward: 5'-GATTGTCATAAANNKATCACCCTGGC-3', reverse: 5'- CMNNCAGGCCATCCACATAGG-3'

Fragment 5. forward: 5'- CTATGTGGATGGCCTGNNKGGTC-3', reverse: 5'- CATGTTGGTTTCCTACATTCAATTTTTTAGTCGCTTATTA -3'

The fragments were then gel purified and amplified with the standard Q5 protocol (NEB) using the forward and reverse primers for ttgR (forward: 5'- CTC-TACAAATAATTTTGTTTAACTTTGAAATAAGGAGGTAATACAAATG -3', reverse: 5'- CATGTTGGTTTCCTACATTCAATTTTTTAGTCGCTTATTA -3'). This full ttgR gene was then subcloned into the pJ3 vector.

Preparation of linearized vectors

Linearized vector was prepared in a 100 µL PCR reaction from pJ3 vector backbone with standard manufacturer's PCR protocol and 2X Hot-Start Q5 Master Mix (NEB). Primers (forward: 5'-TAGGAAACCAACATGTTCACACAGGAAAC-3', reverse: 5'-AAATTATTTGTAGAGGGAAACCGTTGTGGTC-3') were synthesized by IDT. PCR products were purified with gel electrophoresis.

Preparation of a variant library in E.coli. Step 1 - Measurement of mutation rate with a test library: The insert was then recombined with the vector backbone using Takara's In-Fusion Cloning Kit with 40 ng of insert and 100 ng of linearized vector following standard protocol from Takara. Reaction product was transformed into Stellar chemically competent E. coli (Takara) using the manufacturer's transformation protocol. Mutation rates of this test library are measured by sequencing >10 colonies from the test library.

Step 2 - Preparation of a variant library for screening: The same method used to produce test library is scaled up to generate enough number of colonies for screening. In parallel to transformation of the variant library, a glycerol stock of the parent of the library was streaked onto an LB agar plates with 100 µg/ml carbenicillin so that

the individual colonies could be isolated. Both the transformed library and streaked parent were incubated overnight at 37 ºC before being screened with the 96-well format fluorescent assay.

Screening with a 96-well format fluorescent assay

1. Preparation of the library cultured in 96-well plates. Each well in a 96-Well Deep Well plate (VWR) was filled with 400 µL of LB media with 100mg/ml ampicillin. Individual colonies from the variant library were then picked into this plate with autoclaved pipette tips. The borders were left as sterile technique controls. 57 colonies were picked into the plate. In wells E6, F6, G6, individual colonies of the parent of the library were picked from the streaked glycerol stocks. Cells were then cultured in deep-well plate for 16 hours in a 37-degree shaker at 250 rpm.

2. Preparation of LB agar plates containing ligands. LB agar plates containing the ligands of interest at appropriate concentrations and with 100mg/ml ampicillin. This was done by preparing a 1000X stock solution in ethanol (200 Proof, KOPTEC) for both phloretin (Sigma-Aldrich) and melatonin (Sigma-Aldrich). 40 µL of stock solution was added to 40 mL of liquid LB Agar at about 50 ºC. Solutions were mixed thoroughly. These solutions were then poured into OmniTrays (ThermoFischer) and were allowed to cool with flame on.

3. Replication of 96-well-formatted library to LB agar plates containing ligands. 900 µL of fresh LB media with 100 mg/ml ampicillin was then added to each well of the culture plate. Cultures were then replicated with a 96 standard pin replicator (Scinomix) onto the LB Agar plates containing relevant ligands made in (6). The LB agar plates were grown in a 37 ºC incubator for 16 hours.

4. Fluorescent imaging of the plates. Using a ChemiDoc MP fluorescence imaging system (Biorad), a standard imaging protocol at three different channels (epi white light: no filter, auto exposure for intense bands; blue epi-illumination, 530/28 filter, 3 ms exposure; red epi-illumination, 695/55 filter, 200 ms, exposure). Image segmentation and measurements were performed using CellProfiler software and a customized pipeline. Fold changes for individual colonies were calculated using a customized python script. The variants shown with an increased response, increased specificity, and decreased leakiness in the fluorescent assay were selected as hits and streaked out onto LB agar plates with 100 µg/ml carbenicillin.

5. Validation of hits. Ten colonies were picked from each of the hit plates and cultures were inoculated in 96-well deep well plates. The deep-well plate was

then cultured overnight for 16 hours, replicated onto LB Agar plates containing the neurotransmitters of interest, and grown for 16 hours before being imaged as described in (4). Variants that exhibit the desired traits across replicates and trials can then be used as the parents in future rounds of evolution.

Flow cytometry

At day 0, HEK293T cells were split into 24-well plates at 50% confluency, with 500ul DMEM media (with 5% FBS, wo antibiotics) in each well. 24hr later, each well was transfected with 1 µg of total plasmids at the specified ratio with FuGENE6 (Promega).

48hr after transfection, each well was resuspended with 8 µL TryPLE EXPRESS (Thermo Fisher), and then neutralized with 500 µL of resuspension buffer (HBSS with 2.5% BSA, 0.5% DNAseI, and 1 mM MgCl2). Cells were spun down at 300 x g, 3min before being resuspended in 500 µL of resuspension buffer. The resuspended samples were homogenized with 40um cell strainers before being transferred to 96-well U-shaped plates. The samples were then analyzed with a MACSQuant VYB Flow Cytometer (Miltenyi Biotech).

4.6 Acknowledgement

Nicholas Hutchins, Claire N. Bedbrook, and Viviana Gradinaru are contributing authors of this unpublished work. The authors thank Dr. Pamela Bjorkman, Dr. Mikhail Shapiro, Dr. Rob Philips, Dr. David Tirrell, Dr. Sarkis Mazmanian, Sripriya Kumar, Elisha Mackey, Ronghui Zhu, Di Wu, Dr. Dan Piraner, Dr. Han Wang, and Andrew Hill for inspiring discussions. The sokut GAL4 plasmids are generous gifts of Dr.Han Wang and Dr.Paul Sternberg. The automatic colony picking was performed with the help of Dr. Andy Cameron, Ann Cutting, and Kari Koppitch in the Center for computational regulatory genomics at Caltech. Fluorescence imaging of the bacterial culture plates was performed in Dr. Mikhail Shapiro's lab. Flow cytometry was performed at Caltech Flow Cytometry Facility, with instructions from Rochelle Diamond and Ronghui Zhu (flow cytometry).

Chapter 5

FUTURE WORK

5.1 Computational design of BBB receptor binders.

Currently, there is a plethora of protein design tools available. Besides the established Rosetta framework that combines physics and knowledge (Simons et al., 1999; Leaver-Fay et al., 2011), there are now a number of deep learning-based protein design tools that have been experimentally validated, e.g. ProteinMPNN which utilizes graph neural networks (Dauparas et al., 2022), and architectures inspired by AlphaFold2 that incorporate deep learning concepts from computer vision including hallucination, inpainting(Wang et al., 2022b), and diffusion (Watson et al., 2022). These tools open the chance to computationally design peptide binders for target BBB receptors. To ensure the optimal performance of protein design tools, which are better suited for small proteins (< 200 aa), a pipeline will be developed to integrate the binding modules and the desired scaffolds (e.g., AAV capsids).

5.2 High-throughput *In silico* screening of receptor-binding molecules.

The HT-APPRAISE pipeline (Figure 3.4) was able to process a library of 100 variants, however, trials with larger-sized libraries using the current version of AlphaFold-Multimer-based HT-APPRAISE have been unsuccessful due to the long computation hours and the increasing number of false-positive variants that impede efficient pairwise APPRAISE. To extend the screening capability of HT-APPRAISE, novel deep-learning-based structure prediction tools and techniques such as preclustering will be implemented.

5.3 Engineering small molecule shuttles for brain targeting.

Carbonic anhydrase IV (CA-IV, CA4) is a primate-conserved BBB receptor that we have recently identified and is known to bind to small molecule inhibitors, including the FDA-approved, generic small molecule drug brinzolamide. The binding pose of brinzolamide to CA-IV is remarkably similar to the binding pose between BBB-crossing AAV capsid 9P31 and CA-IV (Figure 2.4). It is possible that the binding of brinzolamide, or other CA-IV inhibitors, could be capable of inducing transcytosis across the BBB in a manner similar to that of the CA-IV-dependent, brain-transducing AAV capsids 9P31 and 9P36, which were reported in (Nonnen-

macher et al., 2021; Shay et al., 2023). We, therefore, hypothesize that we can use these CA-IV inhibitors as molecular shuttles to transport therapeutic molecules across the BBB via CA-IV binding (Figure 5.1). To test this hypothesis, conjugates of therapeutic molecules with varying sizes and CA-IV inhibitors with different affinities will be evaluated for their ability to be uptaken by CA-IV-expressing cells *in vitro* and cross CA-IV-expressing animals *in vivo*.

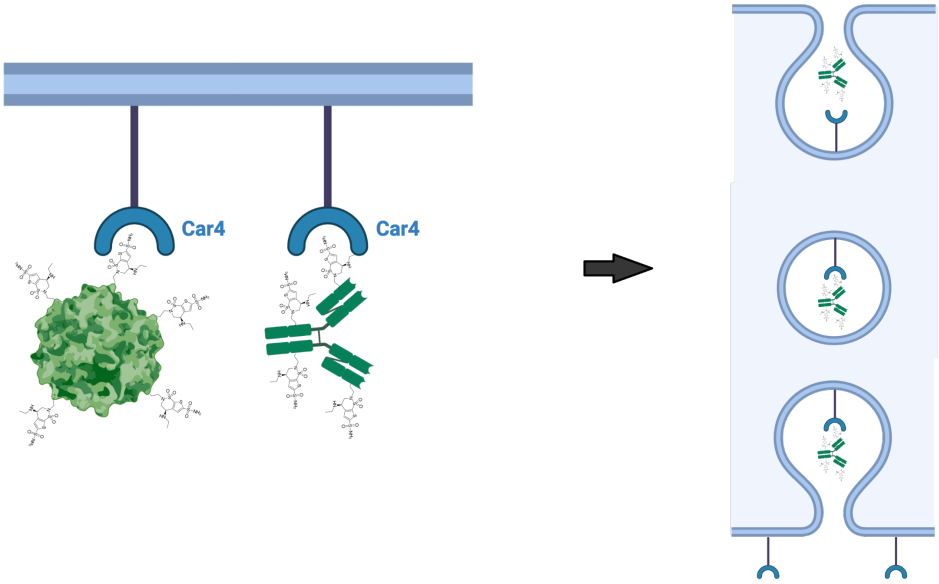


Figure 5.1: **The concept of small molecule shuttle.** A therapeutic cargo, conjugated to a small CA-IV inhibitor such as brinzolamide, may be able to induce transcytosis across the BBB when bound to the catalytic site of carbonic anhydrase IV, as demonstrated by the CA-IV-dependent AAV capsids 9P31 or 9P36 (Nonnenmacher et al., 2021; Shay et al., 2023).

BIBLIOGRAPHY

Aderinwale, Tunde, Charles Christoffer, and Daisuke Kihara (2022). *RL-MLZerD: Multimeric protein docking using reinforcement learning*. URL: https://www.frontiersin.org/articles/10.3389/fmolb.2022.969394.

Agbandje-McKenna, Mavis and Jürgen Kleinschmidt (2011). "AAV capsid structure and cell interactions". eng. In: *Methods in Molecular Biology (Clifton, N.J.)* 807, pp. 47–92. ISSN: 1940-6029. DOI: 10.1007/978-1-61779-370-7_3.

Alford, Rebecca F et al. (2017). "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design". In: *Journal of Chemical Theory and Computation* 13.6, pp. 3031–3048. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.7b00125. URL: https://doi.org/10.1021/acs.jctc.7b00125.

Anand, Namrata et al. (2022). "Protein sequence design with a learned potential". In: *Nature Communications* 13.1, p. 746. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28313-9. URL: https://doi.org/10.1038/s41467-022-28313-9.

Andreoni, Alessio, Carolyn M. O. Davis, and Lin Tian (Dec. 2019). "Measuring brain chemistry using genetically encoded fluorescent sensors". en. In: *Current Opinion in Biomedical Engineering*. Molecular & CellularEngineering: single molecule technology Neural Engineering: High Resolution Cell Imaging 12, pp. 59–67. ISSN: 2468-4511. DOI: 10.1016/j.cobme.2019.09.008. URL: http://www.sciencedirect.com/science/article/pii/S2468451119300431 (visited on 04/06/2020).

Anishchenko, Ivan et al. (2021). "De novo protein design by deep network hallucination". In: *Nature* 600.7889, pp. 547–552.

A.Petsko, Gregory and Dagmar Ringe (2003). "Protein structure and function". In.

Atkinson Jr, Arthur J (2017). "Intracerebroventricular drug administration". In: *Translational and Clinical Pharmacology* 25.3, pp. 117–124.

Baek, Minkyung et al. (Aug. 2021). "Accurate prediction of protein structures and interactions using a three-track neural network". In: *Science* 373.6557. Publisher: American Association for the Advancement of Science, pp. 871–876. DOI: 10.1126/science.abj8754. URL: https://www.science.org/doi/10.1126/science.abj8754 (visited on 09/02/2022).

Baker, David (2019). "What has de novo protein design taught us about protein folding and biophysics?" In: *Protein Science* 28.4, pp. 678–683. ISSN: 0961-8368. DOI: https://doi.org/10.1002/pro.3588. URL: https://doi.org/10.1002/pro.3588.

Bale, Jacob B. et al. (July 2016). "Accurate design of megadalton-scale two-component icosahedral protein complexes". In: *Science* 353.6297, pp. 389–394. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.aaf8818. URL: http://www.ncbi.nlm.nih.gov/pubmed/27463675.

Banks, William A. (Apr. 2016). "From blood–brain barrier to blood–brain interface: new opportunities for CNS drug delivery". en. In: *Nature Reviews Drug Discovery* 15.4, pp. 275–292. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/nrd.2015.21. URL: https://www.nature.com/articles/nrd.2015.21 (visited on 06/23/2022).

Barnes, Christopher O et al. (2020). "SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies". In: *Nature* 588.7839, pp. 682–687.

Batista, Ana Rita et al. (2019). "Ly6a Differential Expression in Blood–Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B". In: *Human Gene Therapy* 31.1-2, pp. 90–102. ISSN: 1043-0342. DOI: 10.1089/hum.2019.186. URL: https://doi.org/10.1089/hum.2019.186.

Batista, Ana Rita et al. (Jan. 2020). "*Ly6a* Differential Expression in Blood–Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B". en. In: *Human Gene Therapy* 31.1-2, pp. 90–102. ISSN: 1043-0342, 1557-7422. DOI: 10.1089/hum.2019.186. URL: https://www.liebertpub.com/doi/10.1089/hum.2019.186 (visited on 09/28/2021).

Bedbrook, Claire N., Benjamin E. Deverman, and Viviana Gradinaru (July 2018). "Viral Strategies for Targeting the Central and Peripheral Nervous Systems". en. In: *Annual Review of Neuroscience* 41.1, pp. 323–348. ISSN: 0147-006X, 1545-4126. DOI: 10.1146/annurev-neuro-080317-062048. URL: https://www.annualreviews.org/doi/10.1146/annurev-neuro-080317-062048 (visited on 09/29/2021).

Bedbrook, Claire N et al. (2017). "Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization". In: *PLOS Computational Biology*, pp. 1–21. ISSN: 1111111111. DOI: 10.1371/journal.pcbi.1005786. URL: http://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1005786&type=printable.

Bennett, Nathaniel et al. (2022). "Improving de novo Protein Binder Design with Deep Learning". In: *bioRxiv*, p. 2022.06.15.495993. DOI: 10.1101/2022.06.15.495993. URL: http://biorxiv.org/content/early/2022/06/17/2022.06.15.495993.abstract.

Bergmann, Sonja et al. (Dec. 2018). "Blood–brain-barrier organoids for investigating the permeability of CNS therapeutics". en. In: *Nature Protocols* 13.12, pp. 2827–2843. ISSN: 1754-2189, 1750-2799. DOI: 10.1038/s41596-018-0066-x. URL:

http://www.nature.com/articles/s41596-018-0066-x (visited on 02/23/2022).

Bolon, Daniel N, Christopher A Voigt, and Stephen L Mayo (Apr. 2002). "De novo design of biocatalysts". In: *Current Opinion in Chemical Biology* 6.2, pp. 125–129. ISSN: 1367-5931. DOI: 10.1016/S1367-5931(02)00303-4. URL: https://www.sciencedirect.com/science/article/pii/S1367593102003034.

Bolon, Daniel N et al. (2005). "Specificity versus stability in computational protein design". In: *Proceedings of the National Academy of Sciences* 102.36, pp. 12724–12729. DOI: 10.1073/pnas.0506124102. URL: https://doi.org/10.1073/pnas.0506124102.

Brinkmann, Ulrich and Roland E Kontermann (2017). "The making of bispecific antibodies". In: *MAbs*. Vol. 9. 2. Taylor & Francis, pp. 182–212.

Burdett, Tamara and Samir Nuseibeh (Sept. 2022). "Changing trends in the development of AAV-based gene therapies: a meta-analysis of past and present therapies". en. In: *Gene Therapy*. ISSN: 0969-7128, 1476-5462. DOI: 10.1038/s41434-022-00363-0. URL: https://www.nature.com/articles/s41434-022-00363-0 (visited on 09/29/2022).

Buss, Marjorie T et al. (2021). "Spatial control of probiotic bacteria in the gastrointestinal tract assisted by magnetic particles". In: *Advanced Materials* 33.17, p. 2007473.

Cao, Longxing et al. (Oct. 2020). "De novo design of picomolar SARS-CoV-2 miniprotein inhibitors". In: *Science* 370.6515. Publisher: American Association for the Advancement of Science, pp. 426–431. DOI: 10.1126/science.abd9909. URL: https://www.science.org/doi/10.1126/science.abd9909 (visited on 09/01/2022).

Cao, Longxing et al. (May 2022). "Design of protein-binding proteins from the target structure alone". en. In: *Nature* 605.7910, pp. 551–560. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-022-04654-9. URL: https://www.nature.com/articles/s41586-022-04654-9 (visited on 10/12/2022).

Challis, Rosemary C. et al. (Feb. 2019). "Systemic AAV vectors for widespread and targeted gene delivery in rodents". en. In: *Nature Protocols* 14.2, pp. 379–414. ISSN: 1754-2189, 1750-2799. DOI: 10.1038/s41596-018-0097-3. URL: http://www.nature.com/articles/s41596-018-0097-3 (visited on 09/28/2021).

Challis, Rosemary C. et al. (July 2022). "Adeno-Associated Virus Toolkit to Target Diverse Brain Cells". en. In: *Annual Review of Neuroscience* 45.1, annurev–neuro–111020–100834. ISSN: 0147-006X, 1545-4126. DOI: 10.1146/annurev-neuro-111020-100834. URL: https://www.annualreviews.org/doi/10.1146/annurev-neuro-111020-100834 (visited on 05/27/2022).

Chan, Ken Y et al. (Aug. 2017). "Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems". en. In: *Nature Neuroscience* 20.8, pp. 1172–1179. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/nn.4593. URL: http://www.nature.com/articles/nn.4593 (visited on 09/28/2021).

Chang, Liwei and Alberto Perez (Mar. 2022a). *AlphaFold encodes the principles to identify high affinity peptide binders*. en. preprint. Biophysics. DOI: 10.1101/2022.03.18.484931. URL: http://biorxiv.org/lookup/doi/10.1101/2022.03.18.484931 (visited on 06/22/2022).

– (2022b). "Ranking Peptide Binders by Affinity with AlphaFold". In: *Angewandte Chemie*. ISSN: 1521-3757. DOI: 10.1002/ANGE.202213362. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/ange.202213362https://onlinelibrary.wiley.com/doi/abs/10.1002/ange.202213362https://onlinelibrary.wiley.com/doi/10.1002/ange.202213362.

Chen, Xinhong et al. (May 2022a). "Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems". en. In: *Neuron*, S0896627322004111. ISSN: 08966273. DOI: 10.1016/j.neuron.2022.05.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627322004111 (visited on 06/21/2022).

Chen, Xinhong et al. (2022b). "Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems". In: *Neuron* 110.14, 2242–2257.e6. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2022.05.003. URL: https://www.sciencedirect.com/science/article/pii/S0896627322004111.

Cheng, Shenggan et al. (2022). *FastFold: Reducing AlphaFold Training Time from 11 Days to 67 Hours*. DOI: 10.48550/ARXIV.2203.00854. URL: https://arxiv.org/abs/2203.00854.

Chevalier, Aaron et al. (Oct. 2017). "Massively parallel de novo protein design for targeted therapeutics". en. In: *Nature* 550.7674. Number: 7674 Publisher: Nature Publishing Group, pp. 74–79. ISSN: 1476-4687. DOI: 10.1038/nature23912. URL: https://www.nature.com/articles/nature23912 (visited on 09/02/2022).

Cho, Choi-Fong et al. (Aug. 2017). "Blood-brain-barrier spheroids as an in vitro screening platform for brain-penetrating agents". en. In: *Nature Communications* 8.1, p. 15623. ISSN: 2041-1723. DOI: 10.1038/ncomms15623. URL: http://www.nature.com/articles/ncomms15623 (visited on 02/23/2022).

Choudhury, Sourav R et al. (July 2016). "In Vivo Selection Yields AAV-B1 Capsid for Central Nervous System and Muscle Gene Therapy". en. In: *Molecular Therapy* 24.7, pp. 1247–1257. ISSN: 15250016. DOI: 10.1038/mt.2016.84. URL: https://linkinghub.elsevier.com/retrieve/pii/S1525001616303574 (visited on 10/05/2022).

Chowdhury, Ratul et al. (2022). "Single-sequence protein structure prediction using a language model and deep learning". In: *Nature Biotechnology* 40.11, pp. 1617–1623.

Chuapoco, Miguel R et al. (2022). "Intravenous gene transfer throughout the brain of infant Old World primates using AAV". In: *bioRxiv*, p. 2022.01.08.475342. DOI: 10.1101/2022.01.08.475342. URL: http://biorxiv.org/content/early/2022/01/09/2022.01.08.475342.abstract.

Coleman, Ryan G and Kim A Sharp (2006). "Travel Depth, a New Shape Descriptor for Macromolecules: Application to Ligand Binding". In: *Journal of Molecular Biology* 362.3, pp. 441–458. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2006.07.022. URL: https://www.sciencedirect.com/science/article/pii/S0022283606008746.

Copeland, Robert A (2016). "The drug–target residence time model: a 10-year retrospective". In: *Nature Reviews Drug Discovery* 15.2, pp. 87–95. ISSN: 1474-1784. DOI: 10.1038/nrd.2015.18. URL: https://doi.org/10.1038/nrd.2015.18.

Couch, Jessica A. et al. (May 2013). "Addressing Safety Liabilities of TfR Bispecific Antibodies That Cross the Blood-Brain Barrier". en. In: *Science Translational Medicine* 5.183. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.3005338. URL: https://www.science.org/doi/10.1126/scitranslmed.3005338 (visited on 11/07/2022).

Cuthbertson, Leslie and Justin R. Nodwell (Sept. 2013). "The TetR family of regulators". eng. In: *Microbiology and molecular biology reviews: MMBR* 77.3, pp. 440–475. ISSN: 1098-5557. DOI: 10.1128/MMBR.00018-13.

Dauparas, J et al. (2022). "Robust deep learning based protein sequence design using ProteinMPNN". In: *bioRxiv*, p. 2022.06.03.494563. DOI: 10.1101/2022.06.03.494563. URL: http://biorxiv.org/content/early/2022/06/04/2022.06.03.494563.abstract.

Davidsson, Marcus et al. (Dec. 2019). "A systematic capsid evolution approach performed in vivo for the design of AAV vectors with tailored properties and tropism". en. In: *Proceedings of the National Academy of Sciences* 116.52, pp. 27053–27062. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1910061116. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1910061116 (visited on 09/28/2021).

Deverman, Benjamin E et al. (Feb. 2016). "Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain". en. In: *Nature Biotechnology* 34.2, pp. 204–209. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.3440. URL: http://www.nature.com/articles/nbt.3440 (visited on 09/28/2021).

Dietrich, Jeffrey A., Adrienne E. McKee, and Jay D. Keasling (2010). *High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection*. Vol. 79. ISBN: 0-626-08095-9. DOI: 10.1146/annurev-biochem-

062608-095938. URL: http://www.annualreviews.org/doi/10.1146/annurev-biochem-062608-095938.

Ding, Xiaozhe et al. (2023). "Fast, accurate ranking of engineered proteins by receptor binding propensity using structural modeling". In: *bioRxiv*. DOI: 10.1101/2023.01.11.523680.

Diskin, Ron et al. (2011). "Increasing the potency and breadth of an HIV antibody by using structure-based rational design". In: *Science* 334.6060, pp. 1289–1293.

Dong, Xiaowei (2018). "Current strategies for brain drug delivery". In: *Theranostics* 8.6, p. 1481.

Ellefson, Jared W., Michael P. Ledbetter, and Andrew D. Ellington (Apr. 2018). "Directed evolution of a synthetic phylogeny of programmable Trp repressors". en. In: *Nature Chemical Biology* 14.4. Number: 4 Publisher: Nature Publishing Group, pp. 361–367. ISSN: 1552-4469. DOI: 10.1038/s41589-018-0006-7. URL: https://www.nature.com/articles/s41589-018-0006-7 (visited on 04/12/2020).

Evans, Richard et al. (Oct. 2021). *Protein complex prediction with AlphaFold-Multimer*. en. preprint. Bioinformatics. DOI: 10.1101/2021.10.04.463034. URL: http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034 (visited on 06/21/2022).

Feng, Jiesi et al. (May 2019). "A Genetically Encoded Fluorescent Sensor for Rapid and Specific In Vivo Detection of Norepinephrine". en. In: *Neuron* 102.4, 745–761.e8. ISSN: 08966273. DOI: 10.1016/j.neuron.2019.02.037. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627319301722 (visited on 04/07/2020).

Fishman, JB et al. (1987). "Receptor-mediated transcytosis of transferrin across the blood-brain barrier". In: *Journal of neuroscience research* 18.2, pp. 299–304.

Fleming, Patrick J. and Karen G. Fleming (Feb. 2018). "HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties". en. In: *Biophysical Journal* 114.4, pp. 856–869. ISSN: 00063495. DOI: 10.1016/j.bpj.2018.01.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0006349518300651 (visited on 09/29/2022).

Foust, Kevin D et al. (Jan. 2009). "Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes". en. In: *Nature Biotechnology* 27.1, pp. 59–65. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.1515. URL: http://www.nature.com/articles/nbt.1515 (visited on 09/28/2021).

Friden, P M et al. (June 1991). "Anti-transferrin receptor antibody and antibody-drug conjugates cross the blood-brain barrier." en. In: *Proceedings of the National Academy of Sciences* 88.11, pp. 4771–4775. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.88.11.4771. URL: https://pnas.org/doi/full/10.1073/pnas.88.11.4771 (visited on 06/23/2022).

Gao, Guangping et al. (June 2004). "Clades of Adeno-Associated Viruses Are Widely Disseminated in Human Tissues". en. In: *Journal of Virology* 78.12, pp. 6381–6388. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.78.12.6381-6388.2004. URL: https://journals.asm.org/doi/10.1128/JVI.78.12.6381-6388.2004 (visited on 06/22/2022).

Goertsen, David et al. (Jan. 2022a). "AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset". en. In: *Nature Neuroscience* 25.1, pp. 106–115. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/s41593-021-00969-4. URL: https://www.nature.com/articles/s41593-021-00969-4 (visited on 01/14/2022).

Goertsen, David et al. (2022b). "AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset". In: *Nature Neuroscience* 25.1, pp. 106–115. ISSN: 1546-1726. DOI: 10.1038/s41593-021-00969-4. URL: https://doi.org/10.1038/s41593-021-00969-4.

Gonzalez, Tawny R et al. (2020). "Assessment of software methods for estimating protein-protein relative binding affinities". In: *PLOS ONE* 15.12, e0240573. URL: https://doi.org/10.1371/journal.pone.0240573.

Hanlon, Killian S. et al. (Dec. 2019). "Selection of an Efficient AAV Vector for Robust CNS Transgene Expression". en. In: *Molecular Therapy - Methods & Clinical Development* 15, pp. 320–332. ISSN: 23290501. DOI: 10.1016/j.omtm.2019.10.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S2329050119301159 (visited on 09/28/2021).

Hardeland, Ruediger et al. (Mar. 2011). "Melatonin-A pleiotropic, orchestrating regulator molecule". English. In: *Progress in Neurobiology* 93.3. Place: Oxford Publisher: Pergamon-Elsevier Science Ltd WOS:000289339300003, pp. 350–384. ISSN: 0301-0082. DOI: 10.1016/j.pneurobio.2010.12.004.

Havlik, L. Patrick et al. (Sept. 2021). "Receptor Switching in Newly Evolved Adeno-associated Viruses". en. In: *Journal of Virology* 95.19. Ed. by Colin R. Parrish, e00587–21. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00587-21. URL: https://journals.asm.org/doi/10.1128/JVI.00587-21 (visited on 02/23/2022).

Hekkelman, Maarten L et al. (2022). "AlphaFill: enriching AlphaFold models with ligands and cofactors". In: *Nature Methods*, pp. 1–9.

Hinrichs, W. et al. (Apr. 1994). "Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance". en. In: *Science* 264.5157. Publisher: American Association for the Advancement of Science Section: Reports, pp. 418–420. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.8153629. URL: https://science.sciencemag.org/content/264/5157/418 (visited on 04/12/2020).

Hordeaux, Juliette et al. (Mar. 2018). "The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice". en. In: *Molecular Therapy* 26.3, pp. 664–668. ISSN: 15250016. DOI: 10.1016/j.ymthe.2018.01.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S152500161830025X (visited on 09/28/2021).

Hordeaux, Juliette et al. (2019). "The GPI-Linked Protein LY6A Drives AAV-PHP.B Transport across the Blood-Brain Barrier". In: *Molecular Therapy* 27.5, pp. 912–921. ISSN: 1525-0016. DOI: 10.1016/j.ymthe.2019.02.013. URL: https://doi.org/10.1016/j.ymthe.2019.02.013.

Huang, Po-Ssu et al. (Aug. 2011). "RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design". en. In: *PLoS ONE* 6.8. Ed. by Vladimir N. Uversky, e24109. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0024109. URL: https://dx.plos.org/10.1371/journal.pone.0024109 (visited on 06/21/2022).

Huang, Qin et al. (2019). "Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids". In: *PLOS ONE* 14.11, e0225206. URL: https://doi.org/10.1371/journal.pone.0225206.

Huang, Qin et al. (2022). "Targeting AAV vectors to the CNS via de novo engineered capsid-receptor interactions". In: *bioRxiv*. DOI: 10.1101/2022.10.31.514553. eprint: https://www.biorxiv.org/content/early/2022/11/01/2022.10.31.514553.full.pdf. URL: https://www.biorxiv.org/content/early/2022/11/01/2022.10.31.514553.

Hudry, Eloise et al. (Sept. 2018). "Efficient Gene Transfer to the Central Nervous System by Single-Stranded Anc80L65". en. In: *Molecular Therapy - Methods & Clinical Development* 10, pp. 197–209. ISSN: 23290501. DOI: 10.1016/j.omtm.2018.07.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S232905011830069X (visited on 10/05/2022).

Jang, Seongmin et al. (2022). "Structural basis of receptor usage by the engineered capsid AAV-PHP. eB". In: *Molecular Therapy-Methods & Clinical Development* 26, pp. 343–354. DOI: 10.1016/j.omtm.2022.07.011.

Jefferies, Wilfred A et al. (1984). "Transferrin receptor on endothelium of brain capillaries". In: *Nature* 312, pp. 162–163.

Jin, Jing-fen et al. (2015). "The optimal choice of medication administration route regarding intravenous, intramuscular, and subcutaneous injection". In: *Patient preference and adherence*, pp. 923–942.

Jing, Miao et al. (Sept. 2018). "A genetically encoded fluorescent acetylcholine indicator for in vitro and in vivo studies". en. In: *Nature Biotechnology* 36.8. Number: 8 Publisher: Nature Publishing Group, pp. 726–737. ISSN: 1546-1696. DOI: 10.1038/nbt.4184. URL: https://www.nature.com/articles/nbt.4184 (visited on 04/07/2020).

Johansson-Åkhe, Isak and Björn Wallner (Sept. 2022). "Improving peptide-protein docking with AlphaFold-Multimer using forced sampling". In: *Frontiers in Bioinformatics* 2, p. 959160. ISSN: 2673-7647. DOI: 10.3389/fbinf.2022.959160. URL: https://www.frontiersin.org/articles/10.3389/fbinf.2022.959160/full (visited on 09/29/2022).

Johnsen, Kasper Bendix et al. (Oct. 2019). "Targeting the transferrin receptor for brain drug delivery". en. In: *Progress in Neurobiology* 181, p. 101665. ISSN: 03010082. DOI: 10.1016/j.pneurobio.2019.101665. URL: https://linkinghub.elsevier.com/retrieve/pii/S0301008219301340 (visited on 06/23/2022).

Jones, Angela R and Eric V Shusta (2007). "Blood–brain barrier transport of therapeutics via receptor-mediation". In: *Pharmaceutical research* 24, pp. 1759–1771.

Jumper, John et al. (Aug. 2021). "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature* 596.7873, pp. 583–589. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03819-2. URL: https://www.nature.com/articles/s41586-021-03819-2 (visited on 06/21/2022).

Kariolis, Mihalis S. et al. (May 2020). "Brain delivery of therapeutic proteins using an Fc fragment blood-brain barrier transport vehicle in mice and monkeys". en. In: *Science Translational Medicine* 12.545, eaay1359. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.aay1359. URL: https://www.science.org/doi/10.1126/scitranslmed.aay1359 (visited on 06/21/2022).

Kimura, Yuki et al. (Mar. 2020). "Directed Evolution of the Stringency of the LuxR Vibrio fischeri Quorum Sensor without OFF-State Selection". In: *ACS Synthetic Biology* 9.3. Publisher: American Chemical Society, pp. 567–575. DOI: 10.1021/acssynbio.9b00444. URL: https://doi.org/10.1021/acssynbio.9b00444 (visited on 04/12/2020).

Kryshtafovych, Andriy et al. (2019). "Critical assessment of methods of protein structure prediction (CASP)—Round XIII". In: *Proteins: Structure, Function, and Bioinformatics* 87.12, pp. 1011–1020.

– (2021). "Critical assessment of methods of protein structure prediction (CASP)—Round XIV". In: *Proteins: Structure, Function, and Bioinformatics* 89.12, pp. 1607–1617.

Kuhlman, Brian et al. (2003). "Design of a novel globular protein fold with atomic-level accuracy". In: *science* 302.5649, pp. 1364–1368.

Kuzmin, Dmitry A. et al. (Mar. 2021). "The clinical landscape for AAV gene therapies". en. In: *Nature Reviews Drug Discovery* 20.3, pp. 173–174. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/d41573-021-00017-7. URL: http://www.nature.com/articles/d41573-021-00017-7 (visited on 09/29/2021).

Langer, Robert (1998). "Drug delivery and targeting." In: *Nature* 392.6679 Suppl, pp. 5–10.

Le Roy, Nathalie et al. (Dec. 2014). "The evolution of metazoan -carbonic anhydrases and their roles in calcium carbonate biomineralization". en. In: *Frontiers in Zoology* 11.1, p. 75. ISSN: 1742-9994. DOI: 10.1186/s12983-014-0075-8. URL: https://frontiersinzoology.biomedcentral.com/articles/10.1186/s12983-014-0075-8 (visited on 10/07/2021).

Leaver-Fay, Andrew et al. (2011). "ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules". In: *Methods in enzymology*. Vol. 487. Elsevier, pp. 545–574.

Lee, Dongmin et al. (2017). "A calcium- and light-gated switch to induce gene expression in activated neurons". In: *Nature Biotechnology* May. DOI: 10.1038/nbt.3902. URL: http://www.nature.com/doifinder/10.1038/nbt.3902.

Lee, J. H. et al. (Apr. 2001). "Receptor mediated uptake of peptides that bind the human transferrin receptor". eng. In: *European Journal of Biochemistry* 268.7, pp. 2004–2012. ISSN: 0014-2956. DOI: 10.1046/j.1432-1327.2001.02073.x.

Li, Ziyao et al. (2022). "Uni-Fold: An Open-Source Platform for Developing Protein Folding Models beyond AlphaFold". In: *bioRxiv*, p. 2022.08.04.502811. DOI: 10.1101/2022.08.04.502811. URL: http://biorxiv.org/content/early/2022/08/30/2022.08.04.502811.abstract.

Liguore, William A. et al. (Nov. 2019). "AAV-PHP.B Administration Results in a Differential Pattern of CNS Biodistribution in Non-human Primates Compared with Mice". en. In: *Molecular Therapy* 27.11, pp. 2018–2037. ISSN: 15250016. DOI: 10.1016/j.ymthe.2019.07.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S1525001619303569 (visited on 09/28/2021).

Lin, Michael Z and Mark J Schnitzer (Aug. 2016). "Genetically encoded indicators of neuronal activity". In: *Nature neuroscience* 19.9, pp. 1142–1153. ISSN: 1097-6256. DOI: 10.1038/nn.4359. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5557009/ (visited on 04/06/2020).

Lin, Zeming et al. (2022). "Language models of protein sequences at the scale of evolution enable accurate structure prediction". In: *bioRxiv*, p. 2022.07.20.500902. DOI: 10.1101/2022.07.20.500902. URL: http://biorxiv.org/content/early/2022/07/21/2022.07.20.500902.abstract.

Liu, Sirui et al. (2022). *PSP: Million-level Protein Sequence Dataset for Protein Structure Prediction*. DOI: 10.48550/ARXIV.2206.12240. URL: https://arxiv.org/abs/2206.12240.

Logan, Todd et al. (Sept. 2021). "Rescue of a lysosomal storage disorder caused by Grn loss of function with a brain penetrant progranulin biologic". en. In: *Cell* 184.18, 4651–4668.e25. ISSN: 00928674. DOI: 10.1016/j.cell.2021.08.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867421009442 (visited on 06/23/2022).

Madani, Ali et al. (2023). "Large language models generate functional protein sequences across diverse families". In: *Nature Biotechnology*, pp. 1–8.

Manzari, Mandana T et al. (2021). "Targeted drug delivery strategies for precision medicines". In: *Nature Reviews Materials* 6.4, pp. 351–370.

Martino, R. Alexander et al. (Sept. 2021). "Context-Specific Function of the Engineered Peptide Domain of PHP.B". en. In: *Journal of Virology* 95.20. Ed. by Rozanne M. Sandri-Goldin. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.01164-21. URL: https://journals.asm.org/doi/10.1128/JVI.01164-21 (visited on 09/28/2021).

Marvin, Jonathan S. et al. (2013). "An optimized fluorescent probe for visualizing glutamate neurotransmission". In: *Nature Methods* 10.2, pp. 162–170. ISSN: 1548-7105 (Electronic)\r1548-7091 (Linking). DOI: 10.1038/nmeth.2333. URL: http://www.nature.com/doifinder/10.1038/nmeth.2333.

Marvin, Jonathan S. et al. (Nov. 2018). "Stability, affinity, and chromatic variants of the glutamate sensor iGluSnFR". en. In: *Nature Methods* 15.11. Number: 11 Publisher: Nature Publishing Group, pp. 936–939. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0171-3. URL: https://www.nature.com/articles/s41592-018-0171-3 (visited on 04/07/2020).

Marvin, Jonathan S. et al. (Aug. 2019). "A genetically encoded fluorescent sensor for in vivo imaging of GABA". en. In: *Nature Methods* 16.8. Number: 8 Publisher: Nature Publishing Group, pp. 763–770. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0471-2. URL: https://www.nature.com/articles/s41592-019-0471-2 (visited on 04/06/2020).

Matsuzaki, Yasunori et al. (Feb. 2018). "Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain". en. In: *Neuroscience Letters* 665, pp. 182–188. ISSN: 03043940. DOI: 10.1016/j.neulet.2017.11.049. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304394017309588 (visited on 01/14/2022).

Matsuzaki, Yasunori et al. (Apr. 2019). "Neurotropic Properties of AAV-PHP.B Are Shared among Diverse Inbred Strains of Mice". en. In: *Molecular Therapy* 27.4, pp. 700–704. ISSN: 15250016. DOI: 10.1016/j.ymthe.2019.02.016. URL: https://linkinghub.elsevier.com/retrieve/pii/S1525001619300577 (visited on 09/28/2021).

McBain, Stuart C, Humphrey HP Yiu, and Jon Dobson (2008). "Magnetic nanoparticles for gene and drug delivery". In: *International journal of nanomedicine* 3.2, pp. 169–180.

McCarthy, Clifford A and Arthur T Benjamin (1996). "Determinants of the Tournaments". In: *Mathematics Magazine* 69.2, pp. 133–135. ISSN: 0025-570X. DOI: 10.1080/0025570X.1996.11996410. URL: https://doi.org/10.1080/0025570X.1996.11996410.

Merchant, A Margaret et al. (1998). "An efficient route to human bispecific IgG". In: *Nature biotechnology* 16.7, pp. 677–681.

Meyer, Nancy L et al. (May 2019). "Structure of the gene therapy vector, adeno-associated virus with its cell receptor, AAVR". en. In: *eLife* 8, e44707. ISSN: 2050-084X. DOI: 10.7554/eLife.44707. URL: https://elifesciences.org/articles/44707 (visited on 05/27/2022).

Mirdita, Milot, Martin Steinegger, and Johannes Söding (Aug. 2019). "MMseqs2 desktop and local web server app for fast, interactive sequence searches". en. In: *Bioinformatics* 35.16. Ed. by John Hancock, pp. 2856–2858. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/bty1057. URL: https://academic.oup.com/bioinformatics/article/35/16/2856/5280135 (visited on 06/21/2022).

Mirdita, Milot et al. (Jan. 2017). "Uniclust databases of clustered and deeply annotated protein sequences and alignments". en. In: *Nucleic Acids Research* 45.D1, pp. D170–D176. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkw1081. URL: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1081 (visited on 06/21/2022).

Mirdita, Milot et al. (June 2022). "ColabFold: making protein folding accessible to all". en. In: *Nature Methods* 19.6, pp. 679–682. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-022-01488-1. URL: https://www.nature.com/articles/s41592-022-01488-1 (visited on 06/21/2022).

Mitchell, Alex L et al. (Nov. 2019). "MGnify: the microbiome analysis resource in 2020". en. In: *Nucleic Acids Research*, gkz1035. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz1035. URL: https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz1035/5614179 (visited on 06/21/2022).

Morrone, Joseph A et al. (2017). "Computed Binding of Peptides to Proteins with MELD-Accelerated Molecular Dynamics". In: *Journal of Chemical Theory and Computation* 13.2, pp. 870–876. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6b00977. URL: https://doi.org/10.1021/acs.jctc.6b00977.

Motmaen, Amir et al. (July 2022). *Peptide binding specificity prediction using fine-tuned protein structure prediction networks*. en. Pages: 2022.07.12.499365 Section: New Results. DOI: 10.1101/2022.07.12.499365. URL: https://www.biorxiv.org/content/10.1101/2022.07.12.499365v1 (visited on 09/02/2022).

Munji, Roeben Nocon et al. (Nov. 2019). "Profiling the mouse brain endothelial transcriptome in health and disease models reveals a core blood–brain barrier dysfunction module". en. In: *Nature Neuroscience* 22.11, pp. 1892–1902. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/s41593-019-0497-x. URL: http://www.nature.com/articles/s41593-019-0497-x (visited on 02/23/2022).

Niewoehner, Jens et al. (2014). "Increased brain penetration and potency of a therapeutic antibody using a monovalent molecular shuttle". In: *Neuron* 81.1, pp. 49–60.

Nonnenmacher, Mathieu et al. (Mar. 2021). "Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning". en. In: *Molecular Therapy - Methods & Clinical Development* 20, pp. 366–378. ISSN: 23290501. DOI: 10.1016/j.omtm.2020.12.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S2329050120302540 (visited on 09/28/2021).

Ojala, David S. et al. (2017). "In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the Subventricular Zone". In: *Molecular Therapy* 26.1. DOI: 10.1016/j.ymthe.2017.09.006.

Oldendorf, WH et al. (1972). "Blood-brain barrier: penetration of morphine, codeine, heroin, and methadone after carotid injection". In: *Science* 178.4064, pp. 984–986.

Orth, Peter et al. (Mar. 2000). "Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system". en. In: *Nature Structural Biology* 7.3. Number: 3 Publisher: Nature Publishing Group, pp. 215–219. ISSN: 1545-9985. DOI: 10.1038/73324. URL: https://www.nature.com/articles/nsb0300_215 (visited on 04/12/2020).

Ovchinnikov, Sergey and Po-Ssu Huang (Dec. 2021). "Structure-based protein design with deep learning". In: *Mechanistic Biology * Machine Learning in Chemical Biology* 65, pp. 136–144. ISSN: 1367-5931. DOI: 10.1016/j.cbpa.2021.08.004. URL: https://www.sciencedirect.com/science/article/pii/S1367593121001125.

Owen, Joshua, Quentin Pankhurst, and Eleanor Stride (2012). "Magnetic targeting and ultrasound mediated drug delivery: Benefits, limitations and combination". In: *International Journal of Hyperthermia* 28.4, pp. 362–373.

Pardridge, William M. (2012). "Drug transport across the blood-brain barrier". In: *Journal of Cerebral Blood Flow and Metabolism* 32.11, pp. 1959–1972. ISSN: 1559-7016 (Electronic)\r0271-678X (Linking). DOI: 10.1038/jcbfm.2012.126. URL: http://dx.doi.org/10.1038/jcbfm.2012.126.

Pardridge, William M et al. (1995). "Human insulin receptor monoclonal antibody undergoes high affinity binding to human brain capillaries in vitro and rapid transcytosis through the blood–brain barrier in vivo in the primate". In: *Pharmaceutical research* 12, pp. 807–816.

Patriarchi, Tommaso et al. (2018). "Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors". eng. In: *Science (New York, N.Y.)* 360.6396. ISSN: 1095-9203. DOI: 10.1126/science.aat4422.

Patriarchi, Tommaso et al. (Dec. 2019). "Imaging neuromodulators with high spatiotemporal resolution using genetically encoded indicators". en. In: *Nature Protocols* 14.12. Number: 12 Publisher: Nature Publishing Group, pp. 3471–3505. ISSN: 1750-2799. DOI: 10.1038/s41596-019-0239-2. URL: https://www.nature.com/articles/s41596-019-0239-2 (visited on 04/06/2020).

Penzes, Judit J. et al. (Sept. 2021). "Adeno-associated Virus 9 Structural Rearrangements Induced by Endosomal Trafficking pH and Glycan Attachment". en. In: *Journal of Virology* 95.19. Ed. by Rozanne M. Sandri-Goldin, e00843–21. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00843-21. URL: https://journals.asm.org/doi/10.1128/JVI.00843-21 (visited on 05/27/2022).

Phillips, Rob et al. (2009). "Emerging roles for lipids in shaping membrane-protein function". In: *Nature* 459.7245, pp. 379–385.

Profaci, Caterina P. et al. (Apr. 2020). "The blood–brain barrier in health and disease: Important unanswered questions". en. In: *Journal of Experimental Medicine* 217.4, e20190062. ISSN: 0022-1007, 1540-9538. DOI: 10.1084/jem.20190062. URL: https://rupress.org/jem/article/doi/10.1084/jem.20190062/151582/The-bloodbrain-barrier-in-health-and-disease (visited on 02/23/2022).

Pulgar, Victor M. (Jan. 2019). "Transcytosis to Cross the Blood Brain Barrier, New Advancements and Challenges". In: *Frontiers in Neuroscience* 12, p. 1019. ISSN: 1662-453X. DOI: 10.3389/fnins.2018.01019. URL: https://www.frontiersin.org/article/10.3389/fnins.2018.01019/full (visited on 01/20/2022).

Rapti, Kleopatra and Dirk Grimm (Oct. 2021). "Adeno-Associated Viruses (AAV) and Host Immunity – A Race Between the Hare and the Hedgehog". In: *Frontiers in Immunology* 12, p. 753467. ISSN: 1664-3224. DOI: 10.3389/fimmu.2021.753467. URL: https://www.frontiersin.org/articles/10.3389/fimmu.2021.753467/full (visited on 09/29/2022).

Ravindra Kumar, Sripriya et al. (May 2020). "Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types". en. In: *Nature Methods* 17.5, pp. 541–550. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-020-0799-7. URL: http://www.nature.com/articles/s41592-020-0799-7 (visited on 09/28/2021).

Reichheld, Sean E., Zhou Yu, and Alan R. Davidson (Dec. 2009). "The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.52, pp. 22263–22268. ISSN: 1091-6490. DOI: 10.1073/pnas.0911566106.

Reiter, Russel J., Dun-Xian Tan, and Lorena Fuentes-Broto (2010). "Melatonin: a multitasking molecule". English. In: *Neuroendocrinology: The Normal Neuroendocrine System*. Ed. by L. Martini et al. Vol. 181. ISSN: 0079-6123 Journal Abbreviation: Prog. Brain Res. WOS:000287859100009. Amsterdam: Elsevier Science Bv, pp. 127–151. ISBN: 978-0-444-53617-4. DOI: 10.1016/S0079-6123(08)81008-4.

Ridgway, John B.B., Leonard G. Presta, and Paul Carter (1996). "'Knobs-into-holes' engineering of antibody C $_{H}$ 3 domains for heavy chain heterodimerization". en. In: *"Protein Engineering, Design and Selection"* 9.7, pp. 617–621. ISSN: 1741-0126, 1741-0134. DOI: 10.1093/protein/9.7.617. URL: https://academic.oup.com/peds/article-lookup/doi/10.1093/protein/9.7.617 (visited on 10/12/2022).

Ring, Aaron M. et al. (Oct. 2013). "Adrenaline-activated structure of the 2-adrenoceptor stabilized by an engineered nanobody". In: *Nature* 502.7472, pp. 575–579. ISSN: 0028-0836. DOI: 10.1038/nature12572. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3822040/ (visited on 09/01/2022).

Roberts, Richard L, Richard E Fine, and Alexander Sandra (1993). "Receptor-mediated endocytosis of transferrin at the blood-brain barrier". In: *Journal of cell science* 104.2, pp. 521–532.

Robinson, J Elliott et al. (Sept. 2019). "Optical dopamine monitoring with dLight1 reveals mesolimbic phenotypes in a mouse model of neurofibromatosis type 1". In: *eLife* 8. Ed. by Inna Slutsky, Eve Marder, and Camilla Bellone. Publisher: eLife Sciences Publications, Ltd, e48983. ISSN: 2050-084X. DOI: 10.7554/eLife.48983. URL: https://doi.org/10.7554/eLife.48983 (visited on 04/06/2020).

Rogers, Jameson K. et al. (Sept. 2015). "Synthetic biosensors for precise gene control and real-time monitoring of metabolites". In: *Nucleic Acids Research* 43.15, pp. 7648–7660. ISSN: 0305-1048. DOI: 10.1093/nar/gkv616. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4551912/ (visited on 04/17/2020).

Romero, Philip A. and Frances H. Arnold (Dec. 2009). "Exploring protein fitness landscapes by directed evolution". en. In: *Nature Reviews Molecular Cell Biology* 10.12, pp. 866–876. ISSN: 1471-0072, 1471-0080. DOI: 10.1038/nrm2805. URL: http://www.nature.com/articles/nrm2805 (visited on 10/05/2022).

Ronzitti, Giuseppe, David-Alexandre Gross, and Federico Mingozzi (Apr. 2020). "Human Immune Responses to Adeno-Associated Virus (AAV) Vectors". In: *Frontiers in Immunology* 11, p. 670. ISSN: 1664-3224. DOI: 10.3389/fimmu.2020.00670. URL: https://www.frontiersin.org/article/10.3389/fimmu.2020.00670/full (visited on 09/29/2022).

Ruffolo, Jeffrey A. et al. (Apr. 2022). *Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies*. en. Pages: 2022.04.20.488972 Section: New Results. DOI: 10.1101/2022.04.20.488972. URL: https://www.biorxiv.org/content/10.1101/2022.04.20.488972v1 (visited on 09/02/2022).

Salamanca Viloria, Juan et al. (2017). "An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass". In: *Scientific*

Reports 7.1, p. 2838. ISSN: 2045-2322. DOI: 10.1038/s41598-017-01498-6. URL: https://doi.org/10.1038/s41598-017-01498-6.

Samulski, R. Jude and Nicholas Muzyczka (Nov. 2014). "AAV-Mediated Gene Therapy for Research and Therapeutic Purposes". en. In: *Annual Review of Virology* 1.1, pp. 427–451. ISSN: 2327-056X, 2327-0578. DOI: 10.1146/annurev-virology-031413-085355. URL: http://www.annualreviews.org/doi/10.1146/annurev-virology-031413-085355 (visited on 09/29/2021).

Senior, Andrew W et al. (2020). "Improved protein structure prediction using potentials from deep learning". In: *Nature* 577.7792, pp. 706–710.

Shay, Timothy F. et al. (2023). "Primate-conserved carbonic anhydrase IV and murine-restricted LY6C1 enable blood-brain barrier crossing by engineered viral vectors". In: *Science Advances* 9.16, eadg6618. DOI: 10.1126/sciadv.adg6618.

Shetty, Reshma P., Drew Endy, and Thomas F. Knight (Apr. 2008). "Engineering BioBrick vectors from BioBrick parts". en. In: *Journal of Biological Engineering* 2.1, p. 5. ISSN: 1754-1611. DOI: 10.1186/1754-1611-2-5. URL: https://doi.org/10.1186/1754-1611-2-5 (visited on 04/19/2020).

Sievers, Fabian et al. (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega". In: *Molecular Systems Biology* 7.1, p. 539. ISSN: 1744-4292. DOI: https://doi.org/10.1038/msb.2011.75. URL: https://doi.org/10.1038/msb.2011.75.

Simons, Kim T et al. (1999). "Ab initio protein structure prediction of CASP III targets using ROSETTA". In: *Proteins: Structure, Function, and Bioinformatics* 37.S3, pp. 171–176.

Sliwoski, Gregory et al. (Jan. 2014). "Computational Methods in Drug Discovery". en. In: *Pharmacological Reviews* 66.1. Ed. by Eric L. Barker. Publisher: American Society for Pharmacology and Experimental Therapeutics Section: Review Article, pp. 334–395. ISSN: 0031-6997, 1521-0081. DOI: 10.1124/pr.112.007336. URL: https://pharmrev.aspetjournals.org/content/66/1/334 (visited on 09/20/2022).

Stams, Travis et al. (Mar. 1998). "Structures of murine carbonic anhydrase IV and human carbonic anhydrase II complexed with brinzolamide: Molecular basis of isozyme-drug discrimination". en. In: *Protein Science* 7.3, pp. 556–563. ISSN: 0961-8368, 1469-896X. DOI: 10.1002/pro.5560070303. URL: https://onlinelibrary.wiley.com/doi/10.1002/pro.5560070303 (visited on 09/29/2021).

Stanton, Brynne C. et al. (Feb. 2014a). "Genomic mining of prokaryotic repressors for orthogonal logic gates". en. In: *Nature Chemical Biology* 10.2. Number: 2 Publisher: Nature Publishing Group, pp. 99–105. ISSN: 1552-4469. DOI: 10.1038/nchembio.1411. URL: https://www.nature.com/articles/nchembio.1411 (visited on 04/17/2020).

Stanton, Brynne C. et al. (Dec. 2014b). "Systematic Transfer of Prokaryotic Sensors and Circuits to Mammalian Cells". In: *ACS Synthetic Biology* 3.12. Publisher: American Chemical Society, pp. 880–891. DOI: 10.1021/sb5002856. URL: https://doi.org/10.1021/sb5002856 (visited on 04/13/2020).

Steinegger, Martin et al. (Dec. 2019). "HH-suite3 for fast remote homology detection and deep protein annotation". en. In: *BMC Bioinformatics* 20.1, p. 473. ISSN: 1471-2105. DOI: 10.1186/s12859-019-3019-7. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3019-7 (visited on 06/21/2022).

Subbaraman, Nidhi (July 2021). "The US is boosting funding for research monkeys in the wake of COVID". en. In: *Nature* 595.7869, pp. 633–634. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/d41586-021-01894-z. URL: http://www.nature.com/articles/d41586-021-01894-z (visited on 09/29/2022).

Sun, Zhoutong et al. (2018). "[ASAP] Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability". In: *Chemical Reviews*. DOI: 10.1021/acs.chemrev.8b00290. URL: http://dx.doi.org/10.1021/acs.chemrev.8b00290.

Supuran, Claudiu T., Andrea Scozzafava, and Angela Casini (Mar. 2003). "Carbonic anhydrase inhibitors". en. In: *Medicinal Research Reviews* 23.2, pp. 146–189. ISSN: 0198-6325, 1098-1128. DOI: 10.1002/med.10025. URL: https://onlinelibrary.wiley.com/doi/10.1002/med.10025 (visited on 09/29/2022).

Sweeney, Melanie D. et al. (Jan. 2019). "Blood-Brain Barrier: From Physiology to Disease and Back". en. In: *Physiological Reviews* 99.1, pp. 21–78. ISSN: 0031-9333, 1522-1210. DOI: 10.1152/physrev.00050.2017. URL: https://www.physiology.org/doi/10.1152/physrev.00050.2017 (visited on 06/23/2022).

Szablowski, Jerzy O., Avinoam Bar-Zion, and Mikhail G. Shapiro (Sept. 2019). "Achieving Spatial and Molecular Specificity with Ultrasound-Targeted Biomolecular Nanotherapeutics". en. In: *Accounts of Chemical Research* 52.9, pp. 2427–2434. ISSN: 0001-4842, 1520-4898. DOI: 10.1021/acs.accounts.9b00277. URL: https://pubs.acs.org/doi/10.1021/acs.accounts.9b00277 (visited on 06/23/2022).

Tabebordbar, Mohammadsharif et al. (Sept. 2021). "Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species". en. In: *Cell* 184.19, 4919–4938.e22. ISSN: 00928674. DOI: 10.1016/j.cell.2021.08.028. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867421010023 (visited on 02/23/2022).

Tang, Shuang Yan and Patrick C. Cirino (2011). "Design and application of a mevalonate-responsive regulatory protein". In: *Angewandte Chemie - Interna-*

tional Edition 50.5, pp. 1084–1086. ISSN: 1521-3773 (Electronic) 1433-7851 (Linking). DOI: 10.1002/anie.201006083.

Tenenbaum, L., E. Lehtonen, and P. Monahan (Dec. 2003). "Evaluation of Risks Related to the Use of Adeno-Associated Virus-Based Vectors". en. In: *Current Gene Therapy* 3.6, pp. 545–565. ISSN: 15665232. DOI: 10.2174/1566523034578131. URL: http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1566-5232&volume=3&issue=6&spage=545 (visited on 09/29/2022).

Terstappen, Georg C et al. (2021a). "Strategies for delivering therapeutics across the blood–brain barrier". In: *Nature Reviews Drug Discovery* 20.5, pp. 362–383.

Terstappen, Georg C. et al. (May 2021b). "Strategies for delivering therapeutics across the blood–brain barrier". en. In: *Nature Reviews Drug Discovery* 20.5, pp. 362–383. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/s41573-021-00139-y. URL: http://www.nature.com/articles/s41573-021-00139-y (visited on 01/18/2022).

Tian, Chao-Yang (Aug. 2021). "China is facing serious experimental monkey shortage during the COVID-19 lockdown". en. In: *Journal of Medical Primatology* 50.4, pp. 225–227. ISSN: 0047-2565, 1600-0684. DOI: 10.1111/jmp.12528. URL: https://onlinelibrary.wiley.com/doi/10.1111/jmp.12528 (visited on 09/29/2022).

Tolvanen, Martti E.E. et al. (Mar. 2013). "Analysis of evolution of carbonic anhydrases IV and XV reveals a rich history of gene duplications and a new group of isozymes". en. In: *Bioorganic & Medicinal Chemistry* 21.6, pp. 1503–1510. ISSN: 09680896. DOI: 10.1016/j.bmc.2012.08.060. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968089612006979 (visited on 09/28/2021).

Tsaban, Tomer et al. (Jan. 2022). "Harnessing protein folding neural networks for peptide–protein docking". en. In: *Nature Communications* 13.1, p. 176. ISSN: 2041-1723. DOI: 10.1038/s41467-021-27838-9. URL: https://www.nature.com/articles/s41467-021-27838-9 (visited on 11/07/2022).

Wang, Dan, Phillip W. L. Tai, and Guangping Gao (May 2019). "Adeno-associated virus vector as a platform for gene therapy delivery". en. In: *Nature Reviews Drug Discovery* 18.5, pp. 358–378. ISSN: 1474-1784. DOI: 10.1038/s41573-019-0012-9. URL: https://www.nature.com/articles/s41573-019-0012-9 (visited on 02/01/2020).

Wang, Guoxia et al. (2022a). *HelixFold: An Efficient Implementation of AlphaFold2 using PaddlePaddle*. DOI: 10.48550/ARXIV.2207.05477. URL: https://arxiv.org/abs/2207.05477.

Wang, Han et al. (2018). "Split cGAL, an intersectional strategy using a split intein for refined spatiotemporal transgene control in Caenorhabditis elegans". eng. In:

Proceedings of the National Academy of Sciences of the United States of America 115.15, pp. 3900–3905. ISSN: 1091-6490. DOI: 10.1073/pnas.1720063115.

Wang, Huan, Miao Jing, and Yulong Li (June 2018). "Lighting up the brain: genetically encoded fluorescent sensors for imaging neurotransmitters and neuromodulators". en. In: *Current Opinion in Neurobiology*. Neurotechnologies 50, pp. 171–178. ISSN: 0959-4388. DOI: 10.1016/j.conb.2018.03.010. URL: http://www.sciencedirect.com/science/article/pii/S0959438817301915 (visited on 04/06/2020).

Wang, Jue et al. (2022b). "Scaffolding protein functional sites using deep learning". In: *Science* 377.6604, pp. 387–394. DOI: 10.1126/science.abn2100. URL: https://doi.org/10.1126/science.abn2100.

Wang, Wenkai, Zhenling Peng, and Jianyi Yang (2022). "Single-sequence protein structure prediction using supervised transformer protein language models". In: *Nature Computational Science* 2.12, pp. 804–814.

Watson, Joseph L et al. (2022). "Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models". In: *bioRxiv*, pp. 2022–12.

Watts, Ryan J and Mark S Dennis (June 2013). "Bispecific antibodies for delivery into the brain". en. In: *Current Opinion in Chemical Biology* 17.3, pp. 393–399. ISSN: 13675931. DOI: 10.1016/j.cbpa.2013.03.023. URL: https://linkinghub.elsevier.com/retrieve/pii/S1367593113000537 (visited on 06/23/2022).

Weinmann, Jonas et al. (Dec. 2020). "Identification of a myotropic AAV by massively parallel in vivo evaluation of barcoded capsid variants". en. In: *Nature Communications* 11.1, p. 5432. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19230-w. URL: http://www.nature.com/articles/s41467-020-19230-w (visited on 10/05/2022).

Wong, Felix et al. (2022). "Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery". In: *Molecular Systems Biology* 18.9, e11081. ISSN: 1744-4292. DOI: https://doi.org/10.15252/msb.202211081. URL: https://doi.org/10.15252/msb.202211081.

Wu, Ruidong et al. (2022). "High-resolution de novo structure prediction from primary sequence". In: *bioRxiv*, p. 2022.07.21.500999. DOI: 10.1101/2022.07.21.500999. URL: http://biorxiv.org/content/early/2022/07/22/2022.07.21.500999.abstract.

Xu, Guangxue et al. (Sept. 2022). "Structural basis for the neurotropic AAV9 and the engineered AAVPHP.eB recognition with cellular receptors". en. In: *Molecular Therapy - Methods & Clinical Development* 26, pp. 52–60. ISSN: 23290501. DOI: 10.1016/j.omtm.2022.05.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S2329050122000742 (visited on 06/22/2022).

Yang, Andrew C. et al. (Feb. 2022). "A human brain vascular atlas reveals diverse mediators of Alzheimer's risk". en. In: *Nature*. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-04369-3. URL: https://www.nature.com/articles/s41586-021-04369-3 (visited on 02/23/2022).

Yang, Jianyi et al. (2020). "Improved protein structure prediction using predicted interresidue orientations". In: *Proceedings of the National Academy of Sciences* 117.3, pp. 1496–1503.

Yang, Kevin K., Zachary Wu, and Frances H. Arnold (July 2019). "Machine-learning-guided directed evolution for protein engineering". En. In: *Nature Methods*, p. 1. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0496-6. URL: https://www.nature.com/articles/s41592-019-0496-6 (visited on 07/23/2019).

Yanofsky, Charles, Virginia Horn, and Deanna Thorpe (1964). "Protein structure relationships revealed by mutational analysis". In: *Science* 146.3651, pp. 1593–1594.

Yin, Huawu et al. (Nov. 2021). "Rational Design of Potent Peptide Inhibitors of the PD-1:PD-L1 Interaction for Cancer Immunotherapy". In: *Journal of the American Chemical Society* 143.44. Publisher: American Chemical Society, pp. 18536–18547. ISSN: 0002-7863. DOI: 10.1021/jacs.1c08132. URL: https://doi.org/10.1021/jacs.1c08132 (visited on 09/02/2022).

Yu, Y. Joy et al. (May 2011). "Boosting Brain Uptake of a Therapeutic Antibody by Reducing Its Affinity for a Transcytosis Target". en. In: *Science Translational Medicine* 3.84. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.3002230. URL: https://www.science.org/doi/10.1126/scitranslmed.3002230 (visited on 06/23/2022).

Yu, Y. Joy et al. (Nov. 2014). "Therapeutic bispecific antibodies cross the blood-brain barrier in nonhuman primates". en. In: *Science Translational Medicine* 6.261. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.3009835. URL: https://www.science.org/doi/10.1126/scitranslmed.3009835 (visited on 10/12/2022).

Zhang, Ran et al. (Apr. 2019). "Adeno-associated virus 2 bound to its cellular receptor AAVR". en. In: *Nature Microbiology* 4.4, pp. 675–682. ISSN: 2058-5276. DOI: 10.1038/s41564-018-0356-7. URL: http://www.nature.com/articles/s41564-018-0356-7 (visited on 05/27/2022).

Zhang, Wandong et al. (2020). "Differential expression of receptors mediating receptor-mediated transcytosis (RMT) in brain microvessels, brain parenchyma and peripheral tissues of the mouse and the human". In: *Fluids and Barriers of the CNS* 17.1, pp. 1–17.

Zhao, Zhen and Berislav V. Zlokovic (July 2020). "Therapeutic TVs for Crossing Barriers in the Brain". en. In: *Cell* 182.2, pp. 267–269. ISSN: 00928674. DOI:

10.1016/j.cell.2020.06.041. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867420308187 (visited on 06/21/2022).

Zuchero, Y. Joy Yu et al. (2016). "Discovery of Novel Blood-Brain Barrier Targets to Enhance Brain Uptake of Therapeutic Antibodies". In: *Neuron* 89.1, pp. 70–82. ISSN: 0896-6273. DOI: https://doi.org/10.1016/j.neuron.2015.11.024. URL: https://www.sciencedirect.com/science/article/pii/S0896627315010302.