Production, Characterization and Engineering of a Variable Lymphocyte Receptor that Targets the Extracellular Matrix of the Brain

By

Elizabeth A. Appelt

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemical and Biological Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 08/28/2024

The dissertation is approved by the following members of the Final Oral Committee:
Eric V. Shusta, Professor, Chemical and Biological Engineering (Advisor)
Quentin M. Dudley, Assistant Professor, Chemical and Biological Engineering
Hazel M. Holden, Professor, Biochemistry
Regina M. Murphy, Professor, Chemical and Biological Engineering
Ivan Rayment, Professor, Biochemistry

Production, Characterization and Engineering of a Variable Lymphocyte Receptor that Targets the Extracellular Matrix of the Brain

Elizabeth A. Appelt

Under the supervision of Professor Eric V. Shusta

at the University of Wisconsin-Madison

ABSTRACT

Variable lymphocyte receptors (VLRs) are a type of antigen receptor derived from the adaptive immune system of jawless vertebrates such as lamprey (*Petromyzon marinus*). Such vertebrates diverged from mammals more than 500 million years ago. Consequently, lamprey VLRs can target mammalian antigens that traditional immunoglobulins fail to recognize as antigenic due to self-tolerance. In previous investigations the Shusta laboratory identified a VLR, namely P1C10, capable of targeting the brain extra-cellular matrix (ECM). P1C10 was identified from a library of VLRs derived from a lamprey immunized with murine brain microvessel plasma membranes. The library of VLRs subsequently was enriched for brain ECM binding, and VLR P1C10 was identified as a candidate VLR capable of binding the brain ECM. As such, P1C10 has been studied as a possible drug delivery vehicle for glioblastoma (GB) therapies. Utilizing murine GB models, subsequent *in vivo* studies in the Shusta laboratory demonstrated that systemically administered

P1C10 showed selective accumulation in the brain at the site of the tumor. When P1C10 was conjugated to doxorubicin-loaded liposomes and administered to GB-bearing mice, there was a significant increase in survival of mice dosed with the P1C10 therapeutic conjugate compared to the no treatment cohort. These exciting results prompted further characterization of the structure and function of VLR P1C10.

To further characterize P1C10, it was critical to develop a robust VLR expression platform. VLRs are disulfide bonded proteins that are often challenging to produce with expression involving genetic modifications, fusion partners, non-scalable host cell lines, inclusion body formation, and poor yield. Previously, P1C10 had been produced as a fusion protein in mammalian expression systems with low yield and poor stability or as a cell-bound construct in a yeast surface display format. To investigate a potential high-yield expression platform for soluble, monomeric P1C10, the SHuffle *E. coli* strain was tested. The SHuffle strain has been genetically altered to allow cytoplasmic disulfide bond formation by mutations to thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) to create an oxidative cytoplasm. Furthermore, the SHuffle strain expresses disulfide bond isomerase DsbC in the cytoplasm to promote correct disulfide bond pairing. Therefore, we hypothesized the SHuffle strain could be used as an expression platform for VLRs. Here, we demonstrate that the Shuffle strain produced multiple unique, monomeric VLRs of high yield, purity and expected activity. Specifically, P1C10 was produced at yields of 20 to 38 mg of protein per liter of culture and concentrated to 18 mg/ml compared to the previous production method in mammalian cells that resulted in poor stability of the VLR that could only be concentrated to 3 mg/ml.

With the high yield and stable monomeric construct of P1C10 produced in SHuffle strain, we were able to pursue the downstream characterization of the VLR. We initiated an X-ray crystallographic analysis of P1C10 and were able to determine the molecular model of P1C10 at 1.3 Å resolution. We show P1C10 bound to a buffer molecule and compare its molecular architecture and binding residues to previously reported VLRs. We use this analysis to speculate on the possible ECM-binding residues of P1C10 and demonstrate that the sequence variability of P1C10 relative to other VLRs may be responsible for the unique binding properties of P1C10.

Expanding on the structural analysis of P1C10, binding properties of P1C10 were further investigated with the putative binding partners of the VLR. The Shusta laboratory previously performed a mass spectrometry pulldown experiment to identify putative ECM binding partners and identified Carboxypeptidase E as a potential binding partner to P1C10. Thus, we investigated the binding of P1C10 to Carboxypeptidase E and other carboxypeptidase family isoforms including Carboxypeptidases A, A6 and B. Molecular modeling with AlphaFold Multimer was used to model P1C10 in complex with the carboxypeptidases and compare binding residues from the molecular models of the complex with our solved P1C10 crystal structure. Alanine mutants were produced and used in binding assays with carboxypeptidases to investigate whether the molecular modeling predictions translated to the biophysical interaction between P1C10 and the carboxypeptidases. We determine that mutations of the binding residues identified through molecular modeling did not impact carboxypeptidase binding. Moreover, we show that binding to the carboxypeptidases was limited as measured by ELISA, isothermal titration calorimetry and crystallization screens. Thus, we are unable to confirm that carboxypeptidases are the *in vivo* target of P1C10.

Given that P1C10 is of lamprey origin, it could elicit immunogenic reactions if used as a human therapy. Thus, we also began work to engineer a “humanized” VLR scaffold that could be used for further development of VLRs as a therapy. Since the general binding motif of VLRs is found along the leucine-rich (LRR) repeat concave binding surface, we hypothesize that the LRR binding motif can be leveraged to develop humanized, VLR-like binding proteins. Here we present the development of a human scaffold based on the human LRR protein SLIT2 Domain 2 (SLIT2D2) that has substantial structural similarity to a VLR. We present our engineered library that introduced variability into the concave surface of SLIT2D2 at amino acid locations known to be involved in ligand binding in both VLRs and SLIT2D2. The resultant SLIT2D2-based library eventually could be used to identify humanized VLRs that bind to any target of interest. For example, a SLIT2D2 variant from the library which has P1C10-like binding could be identified, while also minimizing any potential immunogenicity that the native P1C10 VLR might have. Taken together, the production, characterization, and engineering of P1C10 may enable future development of VLRs and their derivatives as human therapies.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude for all the mentors, collaborators, and friends who have supported me during my Ph.D. I am especially grateful for my thesis advisor, Professor Eric Shusta. Professor Shusta provided the funding and initial framework for my research, and he mentored me as I developed my thesis. I am grateful to have had his support in pursuing my own ideas throughout my thesis research. I will always be grateful for the opportunities he provided me and how he shaped my career by accepting me as one of his Ph.D. students. I was especially fortunate to have an advisor who was also a Wisconsin alum and shared the same pride and spirit for our university as my family. I look forward to his continued advice and mentorship as I navigate my career in biotechnology. Go Badgers!

I am incredibly grateful to Professors Hazel Holden and Ivan Rayment, members of my thesis committee and collaborators. Professor Holden provided invaluable expertise in x-ray crystallography and structure analysis. Professor Rayment made himself available to assist in troubleshooting all the protein production complications (and there were many!), and he provided me more than 20 jars of homemade jam throughout my Ph.D. I am grateful to them for being my mentors and moral supporters. Thank you for always encouraging me to be curious.

I also must thank, Dr. Jim Thoden in Professor Holden's lab. Dr. Thoden was in the lab with me every day. He was a constant supporter and collaborator of all the research in this thesis. He helped maintain lab supplies and instrumentation. He also was an invaluable resource for research suggestions and bouncing ideas around.

I am very grateful for my mentor, collaborator, and friend, Professor Moriah Katt. Professor Katt helped me adjust to academic research after having been in industry prior to my Ph.D. She also provided invaluable advice on cell culture and trained me on many early protocols that I learned in the Shusta lab. As a protein engineer, I was very fortunate to learn cell culture techniques from one of the leading experts in stem cell research.

I would like to thank Professors Regina Murphy and Brian Pfleger, as well. Both Professors Murphy and Pfleger served on my thesis committee at different points of my Ph.D., and they both provided influential feedback and advice from my preliminary exam through my fifth year.

Thank you to my friend, mentor, and former boss Dr. Shawn Russell. Dr. Russell was the first person to get me interested in protein engineering while I was completing my M.S. Throughout my Ph.D. he provided countless feedback and advice on challenging proteins, purification methods, and analytical tools. I am grateful for Dr. Russell's continued career support and friendship.

Many thanks to my two advisees—Hannah Bachmeier and Seth Gehrke. Hannah helped produce P1C10 before we knew how to produce it in SHuffle cells; she maintained countless volumes of culture and refolding experiments. Seth helped develop the SLIT2 library; he was instrumental in making the library a reality in the lab.

Finally, thank you to all my friends and family who supported me throughout this Ph.D. It was quite a journey, but we made it.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – Variable Lymphocyte Receptors for Biomedical Applications

1.1 Abstract

Variable lymphocyte receptors are an emerging class of antigen receptors derived from the alternative adaptive immune system of lamprey. Lamprey diverged from mammals > 500 million years ago on the evolutionary scale. As a result, lamprey can be immunized with human-derived antigens to produce variable lymphocyte receptors capable of targeting human tissue. The unique targeting potential of variable lymphocyte receptors can be leveraged for biomedical applications. In this review, we discuss the origin, structure, and function of variable lymphocyte receptors as well as the ability to engineer variable lymphocyte receptors for diagnostics and therapeutics. Notably, we introduce variable lymphocyte receptor P1C10—a variable lymphocyte receptor capable of targeting the extracellular matrix of the brain.

1.2 Introduction

Lampreys are a jawless vertebrate of the phylum Chordata in the Cyclostomes clade [1, 2]. Within the phylum Chordata, Cyclostomes belong to the Craniates [1, 2]. The Craniates are considered more evolved than the other two taxa within phylum Chordata—the Cephalochordates and the Urochordates [1]. Craniates are more evolved due to their "true" brain, skull, paired external sensory organs, pharyngeal skeleton, and neural crest cells with pluripotency [1]. However, of the Craniates, Cyclostomes are the closest to Urochordates, making Cyclostomes the least evolved of the Craniates [1]. An exclusive group of two, the Cyclostome clade includes lamprey and hagfish[2]. Unfortunately, hagfish research has been hindered by the inability to fertilize eggs *in vitro* [1]. Thus, understanding this evolutionary branching point of this unique group is contingent on the lamprey. This position on the evolutionary timeline makes lamprey a crucial species for research. While the lamprey evolved enough to have many organ systems similar to humans, lamprey diverged from homo sapiens more than 560 million years ago [3]. Thus, lamprey reside at a unique branching point of evolution where archaic biological systems might be leveraged in modern biotechnology.

1.3 Lamprey Immune System

Biomedical research has taken special interest in the lamprey immune system. Unlike jawed vertebrates, jawless vertebrates do not possess a traditional adaptive immune system (Figure 1-1). Lamprey do not have a thymus, T Cell Receptors (TCRs), immunoglobulins (Igs), or Major Histocompatibility Complex (MHC) molecules [4]. However, early histological studies of lamprey revealed lympho-hematopoietic structures suggesting the presence of an alternative anticipatory immune system [4, 5]. Pathology of lamprey tissue sections identified lymphocyte-like cells in lamprey [4-6]. The lymphoid cells were investigated with an immune challenge in the lamprey [4]. Lamprey larvae were challenged with an antigen cocktail and two boosters [4]. Three days after the second booster, the fraction of large lymphocyte-like cells in the test group increased 13-fold compared to the control cohort, suggesting the lamprey immune response was activated [4]. Indicative of an immune response, the large lymphocyte-like cells were approximately double the size of the small lymphocyte-like cells and had azurophilic cytoplasm and prominent nucleoli as seen in activated lymphocytes of mammals [4]. The activated lymphocyte-like cells were sorted, and the mRNA was used to generate a cDNA library of the activated lymphocyte-like cells [4]. The cDNA libraries of the activated lymphocyte-like cells compared to the small lymphocyte-like cells showed the most abundant difference with the number of leucine-rich repeat (LRR) proteins [4]. The stimulated cDNA library returned more than 300 clones of LRR proteins with variable sequences and numbers of LRRs [4]. For this reason, the proteins were termed Variable Lymphocyte Receptors (VLRs) [4].

VLRs have a largely conserved structure with variable regions present in the LRRs. VLRs have a signal peptide, N-terminal LRR (LRRNT), an LRR1, 1-9 variable LRRs (LRRVs), an end LRRV (LRRVE), an LRR connecting peptide (CP), and a C-terminal LRR (LRRCT) composing the ectodomain (Figure 1-2) [4, 7]. The VLRs are tethered to the immune cells via the threonine/proline-rich stalk, a glycosyl-phosphatidyl-inositol (GPI)-anchor, and a hydrophobic tail [4]. Five different types of VLRs have been identified: VLRA, VLRB, VLRC, VLRD, and VLRE [4, 8-11]. VLRA, VLRC, VLRD, and VLRE are thought to be derived from T Cell-like cells and remain tethered to the cell via the stalk, anchor, and tail. In contrast, VLRBs are derived from B Cell-like cells and are secreted as multivalent, soluble proteins[8].

Like the Igs produced by T Cells and B Cells in jawed vertebrates, the VLRs produced by lamprey have been identified to be antigen receptor proteins that compose an alternative adaptive immune system. Early research demonstrated the ability to immunize lamprey with cross-species blood cells to generate an anti-blood cell response. Lamprey immunized with sheep red blood cells and human Type O blood cells produced unique immune responses suggestive of an adaptive immune system able to upregulate antigen receptors specific to the antigen challenge [12-15]. These antigen receptors were later identified to be the VLRs produced by the lymphocyte-like cells of the lamprey. Lamprey immunized with anthrax spores produced VLRs capable of recognizing anthrax spores with progressive affinity over the booster period [16]. Once the VLRs were identified as the antigen receptors of the lamprey adaptive immune system, the assembly and diversification potential were investigated.

VLRs are assembled according to a combinatorial process. Rather than encoding complete VLRs, the lamprey germline encodes LRR cassettes [4, 16, 17]. As many as 454 VLRB and 393 VLRA unique LRR cassettes have been identified in the lamprey germline [11]. The cassettes were identified to have 6-50 nucleotides of overlap at each terminus to promote a combinatorial assembly [11]. While vertebrates assemble Ig fragments using the RAG-dependent system, the lamprey immune system uses a AID-APOBEC family cytosine deaminase to combine LRR cassettes [11]. Through V(D)J recombination and somatic diversity, homo sapiens have been estimated to be able to assemble 10^{12} different Igs [18] whereas lamprey are estimated to have the potential to assemble, at minimum, 10^{14} to 10^{17} unique VLRs with the combinatorial process hinting at even greater potential diversity of VLRs in lamprey with each lymphocyte producing a unique VLR [4, 11, 17]. The potential to produce countless, unique VLRs against specific antigen targets has fostered an interest VLR development for biomedical applications.

1.4 Variable Lymphocyte Receptor Applications

With the discovery of VLRs in the lamprey adaptive immune system, lamprey immunization with biomedically relevant antigens has been investigated. Because lamprey diverged from humans > 500 million years ago, the lamprey immune system recognizes human targets as antigenic [19]. Thus, lamprey immunized with human-derived antigens generate a robust adaptive immune response with VLRs capable of binding human targets. VLRs have been demonstrated to have high specificity for antigenic targets with dissociation constants as strong as picomolar affinity [19-27]. VLRs used in biotechnology applications commonly are derived from VLRBs or engineered to mimic the soluble structure of VLRBs eliminating the stalk, tail, and anchor of the wildtype VLRs such that the VLRs exist as soluble monomers.

One of the first VLRs generated from a targeted lamprey immunization was VLR RBC36 [27]. RBC36 was discovered from a library of VLRs generated from lamprey immunized with human blood group O erythrocytes. Consequently, RBC36 binds H-trisaccharide derived from the H-antigen of human blood group O erythrocytes [27]. The structure of RBC36 in complex with H-trisaccharide was solved at 1.67 Å (PDB ID: 5UFD) to shed light on the binding interface of VLRs. Like other LRR proteins, RBC36 bound its antigen along the β sheet of the concave face of the VLR composed of its LRRs strands suggesting the binding specificity is imparted by the variable residues of the LRRs.

The ability of RBC36 to bind a human sugar led to the investigation of VLR-binding to other biomedically relevant glycans. A library of 10^8 VLRs was collected from lamprey in the wild without any known immune challenge specific to human glycans [26]. Hong and colleagues

screened the library for binding to tumor-associated carbohydrate antigens (Tn and TFα), Lewis antigens (LeA and LeX), N-glycolylneuraminic acid, HIV antibody targets (poly-Man9 and the HIV gp120), and glycoproteins asialo-ovine submaxillary mucin (aOSM) and asialo-human glycophorin A (aGPA). VLR binders for each of the sugars were found in the library [26]. The VLRs bound each of the sugars with nanomolar affinity suggesting that VLRs could be used in diagnostics or drug delivery.

Given the high affinity binders found in wildtype lamprey, the ability to immunize lamprey with specific antigens for strong binders was investigated using β-galactosidase (β-gal) and hen egg lysozyme (HEL) [28]. VLR libraries from the immunized lamprey were screened for binding to β-gal and HEL. VLR binders to both antigens were identified and binding was assessed to be on the single digit picomolar scale [28] comparable to very tight binding Igs. The ability to generate a library of VLRs with picomolar scale binding motivates the use of antigen-specific challenged lamprey as opposed to relying on wildtype lamprey VLR libraries.

The ability to detect a specific clinical indication with an immune-challenged VLR library was confirmed using human plasma cells [21]. Lamprey larvae were immunized with bone marrow from a multiple myeloma patient to generate a library of VLRs against antigens specific to plasma cells [21]. A VLR clone was identified from the library capable of binding CD38 ectoenzyme, a cell marker unique to plasma cells [21]. Thus, the VLR could be used a diagnostic tool for comparing the number of plasma cells in a healthy individual to someone with a higher count of plasma cells such is the case for multiple myeloma.

Non-human targets with biomedical applications have been explored for VLR binding, as well. In one such study, lampreys were immunized with the exosporium of B. anthracis spores [22]. The subsequent VLR library was screened for binding to the major protein of the exosporium, BclA [22]. A VLRB, VLR5, was identified to bind BclA. The residues in VLR5 responsible for binding were mutated to demonstrated knockdown binding affinity in response to changes along the LRRs in the concave β sheet [22]. This study demonstrated the ability to target a pathogen with VLRs, and it highlighted the ability to mutate surface residues along the LRR binding interface of VLRs, suggesting the potential to engineer VLRs with specific binding features.

1.5 Variable Lymphocyte Receptor Engineering

Given the ability for lamprey to generate VLRs with unique binding specificity, high affinity, and strong avidity for biomedically relevant target, the next opportunity was to further engineer the VLRs for improved production and use. Protein engineering strategies have been utilized in VLR development to improve the solubility of VLRs, impart binding specificity, and improve the half-life of VLRs *in vivo*. Such efforts have made VLRs more clinically relevant and capable of rivaling traditional Ig diagnostics and therapies.

The repebody was one of the first examples of an engineered VLR [7]. In the repebody design, the modularity of VLRs is exploited to improve the manufacturability of VLRs. VLRs exist with consensus repeat units of LRRs. Much of the LRRs are conserved sequences except for the variable regions responsible for binding. In nature, VLRs exist with 1 to 9 variable LRRs each with 20-29 conserved residues and 4 variables residues that are generally responsible for binding specificity [7]. Thus, it was conceived that a binding scaffold could be engineered with repeat modules of LRRs and termed the "repebody" for the repeat units [7]. The repebody was engineered with a consensus LRRNT from internalin-B. An alternative to lamprey LRRNT was used to increase the solubility of the scaffold protein. Wildtype lamprey VLRs are poorly soluble in bacterial expression system. Previous work demonstrated that LRR proteins fold via an N-terminal transition state as that of internalin-B [29]. The internalin-B LRRNT can act a folding chaperone for other LRR protein. Therefore, Internalin-b was chosen for the LRRNT of the repebody to promote solubility of the scaffold in bacteria and because homology modeling indicated structural similarities between the internalin-b LRRNT and the conserved LRRNT of

lamprey VLR. Following the LRRNT, LRRVs were inserted using a consensus LRRV sequence from VLRs found in nature. 3 to 6 LRRVs were used in the repebody scaffold to mimic the most abundant VLRs found in nature. Solubility of the repebody was inversely related to the number of LRRVs inserted in the scaffold. Consensus CP and LRRCT were chosen from the most abundant sequences found in lamprey VLRs. The scaffold was able to be produced in *E. coli* with yields ranging from 50-80 mg of protein per liter of culture indicating that the modular assembly of the repebody was successful [7].

The ability to change the binding specificity of the repebody was tested. Myeloid differentiation protein-2 (MD2) and HEL were chosen as the two initial antigen targets for repebody mutants [7]. MD2 and HEL were chosen as initial antigens for the repebody because they had both been previously solved in complex with a structurally similar LRR protein. To rationally design repebodies with specificity for each antigen, a homology model of the repebody scaffold was aligned with the solved LRR proteins in complex with MD2 (PDB ID: 2Z65) and HEL (PDB ID: 3G3A). The surface residues in the variable regions of the repebody LRRs were compared to the known binding residues of the LRR proteins in the solved crystal structures. The positions in the repebody scaffold that aligned with the known binding positions of the LRR proteins in the crystal structures were mutated to be the same residues as that of the known binders. The resulting repebodies were produced in *E. coli* and demonstrated binding on the same scale as that of the LRR proteins from which the repebodies were modeled [7]. These results demonstrated the feasibility of rationally designed VLRs.

In addition to rational design of the engineered repebody scaffold, the ability to create a library of repebodies was also explored. For the library, the repebody scaffold was mutated with NNK codons at three surface residues in the binding motif of the first two LRRVs for a total of six hypervariable positions in the scaffold amino acid sequence. The NNK library was estimated to be 10^8 in size and screened against a biomedically relevant target, interleukin-6 (IL-6). Following four rounds of enrichment, three candidate clones were identified from the repebody library with nanomolar scale binding for IL-6 [7]. Taken together, the repebody scaffold engineering shows the ability to rationally design VLRs as well as use directed evolution of engineered VLRs to alter binding specificity.

Following the success of the repebodies engineered against MD2, HEL, and IL-6, additional targeting scaffolds were engineered. In subsequent iterations of the repebody scaffold, a binder to ectodomain of human epidermal growth factor receptor EGFR (hEGFR) was identified [30, 31]. hEGFR was used as a target to demonstrate the ability of repebodies to target disease markers on malignant tumors for diagnostics and drug delivery [31]. A repebody able to bind hEGFR was discovered from initial rational design of the scaffold and a series of phage library enrichment with hEGFR. The resulting repebody was able to be labeled with fluorescent dye or radioisotope for *in vivo* applications [31, 32].

In recent work, the repebody was further engineered for improved half-life stability [33]. Wildtype VLRs are smaller than most Igs with a molecular weight ranging from < 20 kDa to < 50 kDa depending on the number of repeat modules present [27, 33]. The small size and lack of Fc allow the VLRs to evade FcRn re-uptake and make them easily cleared by the renal system [33].

Fast clearance makes VLRs less efficacious as therapeutics as they require more frequent dosing which can be costly and time prohibitive. Therefore, methods to increase the half-life of VLRs in circulation were explored. A repebody engineered with affinity for human serum albumin (HAS) was designed to address the issues of fast clearance. Previous studies indicated that HSA-binding proteins extended the circulatory time of drugs in vivo [34]. Therefore, an HSA-binding repebody could extend the half-life of other repebodies. As a proof of concept, the HSA-binding repebody was fused to GLP1, an existing therapeutic for endocrine disorders. The repebody-GLP1 fusion had a 5-fold higher half-life than that of GLP1 on its own [34]. These results suggest that HSA-binding repebodies could act as fusion partners for other repebodies with bivalent targeting to increase half-life *in vivo*.

Like the repebody, other engineered VLRs have been described for biomedical applications including the “lambody.” The lambody is named for its design as a lamprey-derived antibody fusion [26]. In the lambody design, the ectodomain of VLRBs are expressed as fusion to the Fc fragment of Igs. The Fc fusion increases the solubility of VLRBs in mammalian expression systems, and it promotes increased half-life via the FcRn reuptake mechanism. Furthermore, the Fc dimerizes to form a dimeric VLRB potentially increasing the avidity of the VLRB fusions.

1.6 Variable Lymphocyte Receptors that Target the Central Nervous System

The ability to engineer VLRs using common protein engineering techniques enabled the development of VLRs for more complex disease indications. Specifically, VLRs capable of targeting the Central Nervous System (CNS) have recently been reported [35]. Diseases of the CNS evade many existing therapeutics due to the Blood-Brain Barrier (BBB). The BBB is responsible for maintaining the essential homeostasis of the CNS by enabling the selective uptake of recognized nutrients to the parenchyma while inhibiting unrecognized molecules such as toxins from entering the CNS. The BBB is formed by the brain endothelial cells that line the vasculature of the CNS with paracellular tight junction proteins linking the endothelial cells of the BBB and inhibiting paracellular transport of any non-regulated substances from the lumen of the blood vessels [36]. Endothelial cells compose the luminal-facing layer of the BBB while pericytes and astrocytic endfeet surround the perivascular space of the endothelial cell basement membranes on the basal side of the BBB [36]. Pericytes support endothelial maturation and permeability via growth factor secretion and contribute to selective uptake of regulated nutrients [36]. Astrocytes biochemically support endothelial cells and deliver regulated nutrients from the basal compartment of the BBB to neurons in the parenchyma [36]. As a result, minimal paracellular and transcellular diffusion occurs across the BBB. Instead, most regulated substances cross the BBB via receptor-mediated transcytosis (RMT), cell-mediated transcytosis (CMT), transporter-mediated transcytosis (TMT), and adsorptive mediated transcytosis (AMT) [37]. The transcytosis pathways for regulated substances to cross the BBB involve highly specific binding interactions between the transcytosis binding moiety on the BBB and the molecule being transported across the BBB. For example, glucose alone has twelve unique transporters on the

BBB that respond to specific metabolic states of the CNS to regulate the uptake of glucose in the CNS [38].

While the BBB safeguards the CNS, it also prevents pharmaceuticals from entering the brain parenchyma. In the event a pharmaceutical passes the BBB, efflux transporters exist on the basal membrane to pump the invader back to the lumen of the blood vessel [35]. Additionally, the CNS has several immune cells specific to the vasculature of the brain that act to eliminate any foreign bodies that evade the previously mentioned barriers of the BBB [36, 39]. Perivascular macrophages (PVM) exist in the perivascular space of the BBB, and vessel-associated microglia (VAM) exist in the basal compartment just beyond the astrocytic endfeet of the BBB [39]. Both PVMs and VAMs phagocytose foreign bodies that cross the BBB and contribute to immunogenic responses in the CNS [39].

The complexity of the BBB causes therapeutics that are efficacious for diseases outside the CNS to fail to demonstrate any efficacy in CNS diseases. The inability of FDA-approved drugs to cross the BBB is especially evident in cases of metastatic disease. Such an example exists with FDA-approved monoclonal antibodies (mAbs) trastuzumab, rituximab and bevacizumab which are efficacious against primary breast cancer. However, in clinical trials involving metastatic disease, these therapeutics are ineffective against breast cancer metastases to the brain even in the presence of the breast cancer tumor markers HER2, CD20 and VEGF in the CNS [40, 41]. Thus, much of the problem in treating CNS diseases does not rely on novel therapeutics, but rather developing drug delivery mechanisms to shuttle existing therapies across the BBB. Receptor-mediated transcytosis has been explored as a transport system to conjugate FDA-approved drug

payloads, such as the previously mentioned mAbs, to natural binding partners of the BBB. In this delivery scheme, the BBB-targeting molecule binds to a receptor on the luminal side of the endothelial cells and triggers a vesicular endocytosis of the conjugated therapeutic which undergoes transcytosis to the basement membrane and release into the perivascular space of the basal compartment of the BBB for uptake into the parenchyma. Several receptors on the BBB are targets for this drug delivery method including the low-density lipoprotein-related receptors [42], transferrin receptor [43], insulin receptor [44], basigin [45] and CD98hc [46]. Antibody targeting of receptors has demonstrated as much as 50 nM accumulation of the antibody in the brain as demonstrated with anti-transferrin [44]. Thus, development of BBB-targeting drug delivery systems using antibodies for transcytosis presents a possible solution to the problem of delivering drugs to the CNS.

Current antibody-based drug delivery methods for CNS therapies have pitfalls that prevent clinical efficacy. The current receptor targets such as transferrin and insulin that are used for BBB transcytosis are expressed in tissues outside the CNS which causes systemic uptake of the drug. Non-specific drug uptake leads to only 0.5-4% of each dose reaching the brain [43, 44]. Loss of >95% of the drug payload is not financially viable for manufacturing and dosing nor is it sustainable biologically. Systemic uptake of drug payloads such as potent chemotherapies in healthy tissue can lead to serious adverse events in patients [47]. The systemic uptake of antibody delivery systems to the brain presents an urgent need for novel antibody drug delivery systems with higher specificity for the BBB than current methods.

To overcome the challenges of dosing therapeutics across the BBB, VLRs have been explored as potential drug delivery vehicles to the brain. Lamprey immunized with mouse brain microvessel plasma membranes produce a robust immune response generating a library of VLRs with BBB targets [35]. BBB-binding VLRs potentially can bind and transcytose the BBB via the mechanisms utilized by CNS nutrients such as RMT. As such, the resulting VLR library was screened against the plasma membranes for positive binders and successive rounds of enrichment were performed using mouse brain endothelial cells to sort for extracellular binders. A negative selection was done using decellularized mouse brain endothelial cells, leaving behind the extra-cellular matrix (ECM), to remove ECM-binding VLRs from the library to focus the screen on cellular epitopes. Positive binders were selected from the screen and expressed as VLR-Fc fusions according to the previously described method of the lambody VLRs. The VLRs showed positive internalization in mouse brain endothelial cells with signature punctate indicative of endocytosis. Specifically, two of the VLRs were identified to internalize in the endothelial cells via caveolin-mediated endocytosis while one of the VLRs internalized via clathrin-mediated endocytosis [35]. The three candidate VLRs were found to bind mouse brain endothelial cells with nanomolar affinity. In vivo studies with mice demonstrated positive and specific uptake of mouse brain microvessels with the VLRs [35]. One of the candidate VLRs demonstrated both positive endocytosis in the mouse brain endothelial cells and complete transcytosis across the cytoplasm to the brain parenchyma [35]. The results suggest the ability to engineer VLRs with CNS specificity for drug delivery to the brain.

1.7 Variable Lymphocyte Receptor P1C10

The potential to use VLRs for drug delivery was further investigated by identifying brain ECM-binding VLRs. Previously, the library of VLRs enriched for mouse brain microvessel plasma membrane binders was negatively sorted for ECM-binders [35]. The negative sort from the previous library was used to generate an enriched library for VLRs capable of binding the brain ECM [48]. In many diseases of the CNS, the BBB is pathologically disrupted at the site of disease. Pathology that disrupts the BBB includes common diseases of the CNS such as HIV-1 encephalitis, Alzheimer's disease, ischemia, tumors, multiple sclerosis, and Parkinson's disease [36]. Thus, one approach to targeted drug delivery to the site of disease is to target the ECM of the brain that is exposed at the site of BBB disruption (Figure 1-2). The ECM off the brain exists on the basal side of the endothelial cells that compose the BBB which is normally unexposed to the bloodstream that runs along the luminal side of the BBB. However, in the case of pathologically-disrupted BBB, the ECM of the endothelial cells is exposed at the disease site, making the ECM a ubiquitous and accessible targeting moiety. Therefore, a library of VLRs capable of binding brain ECM could be leveraged as drug delivery vehicles in diseases of the CNS that disrupt the BBB.

VLR P1C10 was identified as a candidate VLR in the brain ECM-binding VLR library [48]. P1C10 was identified from the brain ECM-binding VLR library following two rounds of enrichment using decellularized mouse brain endothelial cells to enrich for ECM binders. The library was negatively sorted for binding to fibroblast ECM. The library was evaluated for binding via ECM ELISAs and labeling of murine brain sections with P1C10 being a top performing clone. Further characterization of P1C10 demonstrated that the VLR positively labeled mouse brain section as

well as freshly resected, snap-frozen, histologically normal human brain sections and samples from human glioblastoma tumors [48]. The binding was quantified using an ECM ELISA and P1C10 which revealed the dissociated constant of monomeric P1C10 to be nanomolar scale [48]. P1C10 was evaluated in mouse models. Healthy mice were dosed with P1C10 and demonstrated no brain retention. However, when the BBB was disrupted with mannitol, P1C10 uptake in the brain was observed confirming brain ECM-binding of P1C10 [48].

Glioblastoma mouse models were used to assess the efficacy of P1C10 as a drug delivery vehicle to the tumor site. Glioblastoma was chosen as this is a currently incurable disease with very poor therapeutic options as a result of the pseudopalisading pathology that makes complete surgical resection impossible [49]. Furthermore, the disease targeting of existing chemotherapies to the tumor site is poor, with most of the therapeutics being systemically cleared [50]. In the glioblastoma mouse models, IV-injected P1C10 demonstrated selective uptake at the tumor site suggesting that P1C10 was able to specifically identify and bind its ECM target *in vivo* when systemically administered [48].

Further evaluation of the in vivo efficacy of P1C10 was done using doxorubicin-loaded liposomes. Survival studies of glioblastoma-bearing mouse models were conducted compare the survival time of P1C10 doses mice with control cohorts. Indeed, the mice dosed with doxorubicin-loaded liposomes conjugated to P1C10 had a significantly increased survival time when compared with the non-P1C10 cohorts [48]. Taken together, these exciting results suggest that P1C10 could be used as a drug delivery vehicle for existing cancer therapies to reach the tumor site in glioblastoma.

1.8 Summary

The lamprey immune system has provided protein engineers more >10^{14} antigen receptors, termed VLRs, to discover and design for biomedical applications. Because lamprey diverged more than 500 million years ago from humans, lamprey immunized with human-derived antigens produce a robust immune response of VLRs targeting human antigens. These VLRs can be engineered to enhance manufacturability and alter the binding specificity, avidity, and affinity to broaden their potential biomedical applications. Notably, VLRs can be engineered for drug delivery to the CNS. VLR P1C10 has shown exciting drug delivery properties in mouse glioblastoma models. However, the specific binding target of P1C10 remains unknown. The structure and function of P1C10 can be further evaluated to leverage the VLR's unique drug delivery properties in the CNS.

1.9 Figures

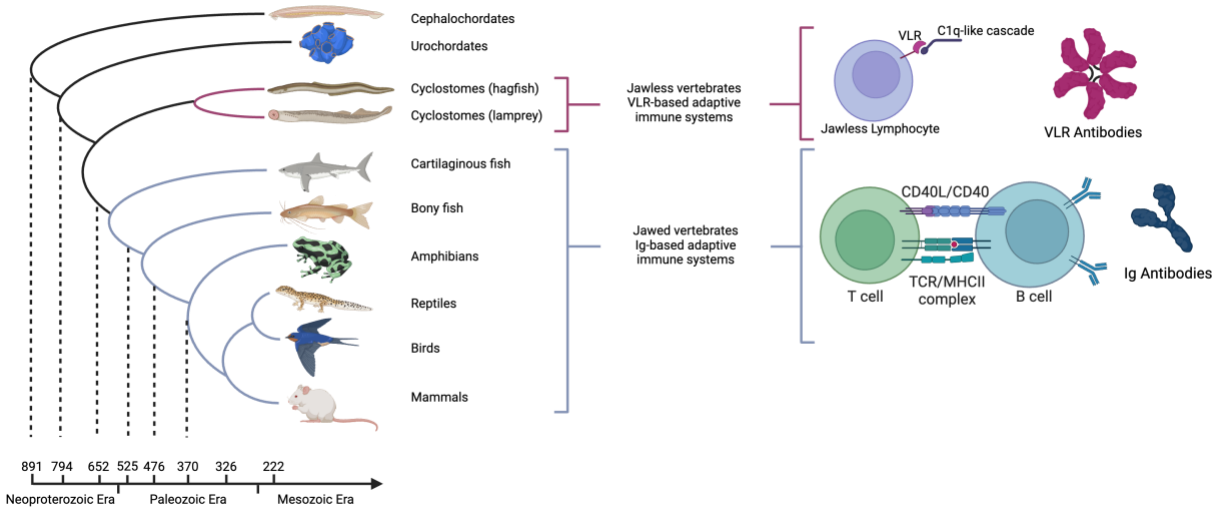


Figure 1-1 Adaptive immune system of jawless and jawed vertebrates. Jawless vertebrates have lymphocyte-like cells that are stimulated by a C1q-like cascade to produce VLR antibody-like proteins. The soluble VLRs form multimers of dimers via their proline/threonine-rich stalk. Jawed vertebrates have T Cell and B Cell lymphocyte that use a complex signaling process that includes CD40L/CD40 and TCR/MHCII signaling cascades to produce Ig antibodies.

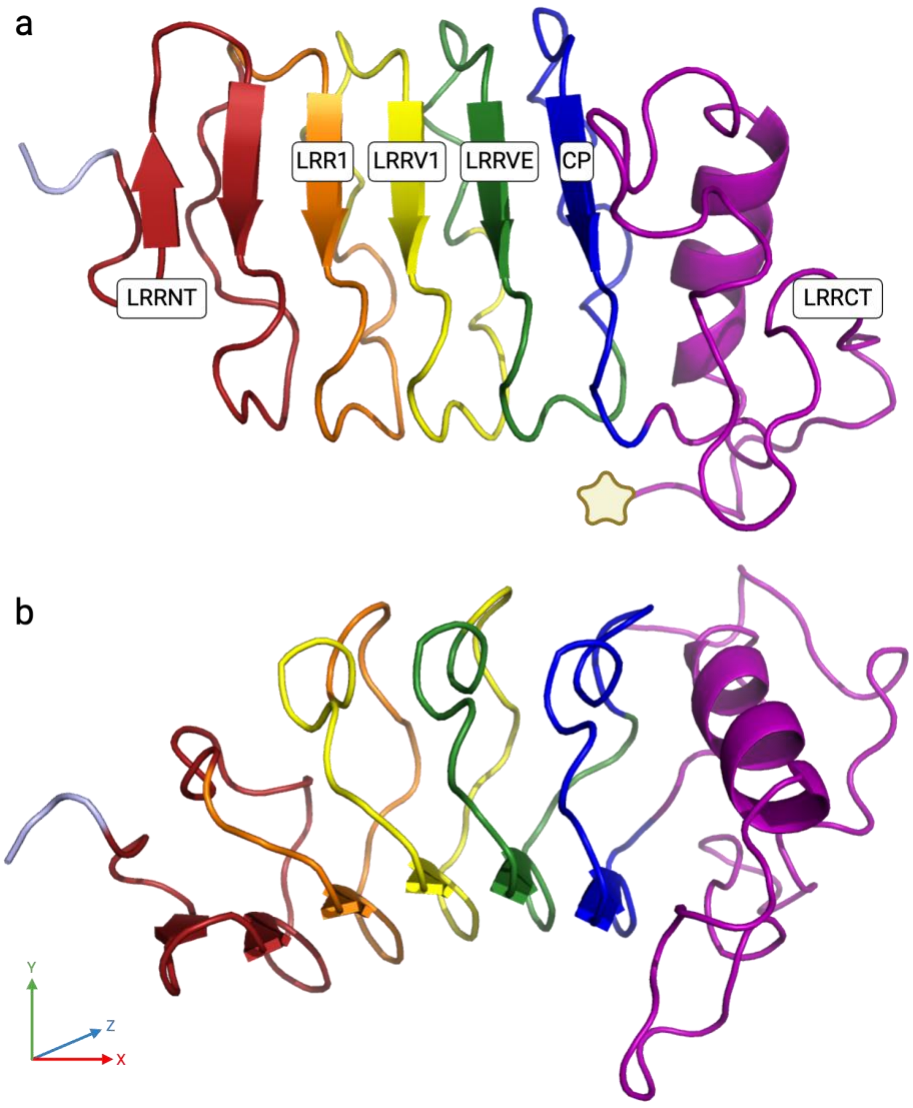


Figure 1-2 Structure of a 1.4 Variable Lymphocyte Receptor. (a) The β sheet is composed of β strands with each β strand representing one of the LRRs. Each LRR is color-coded: LRRNT (red), LRR1 (orange), LRRV1 (yellow), LRRVE (green), CP (blue), LRRCT (purple). In naturally occurring VLRs, the end of the LRRCT (marked with the star) connects to the threonine/proline-rich stalk that feeds into the GPI anchor and hydrophobic tail which tether the VLR to the lymphocyte-like cell surface. (b) The VLR is rotated 90 degrees along the z-axis to depict the solenoid shape of the concave β sheet.

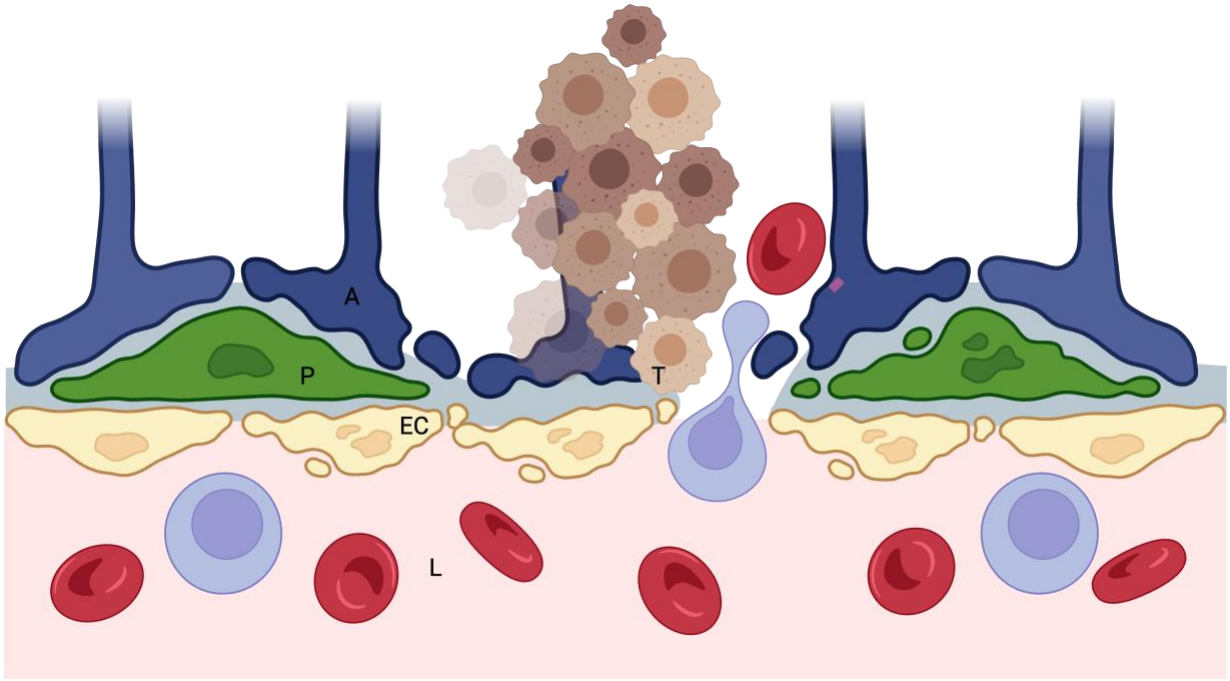


Figure 1-3 Pathologically-disrupted Blood-Brain Barrier. At the site of tumor (T), the astrocytic endfeet (A) and pericytes (P) lining the endothelial cells (EC) are compromised by the invading tumor cells, and the tight junctions of the endothelial cells are separated leading to fenestrations in the BBB at the site of the tumor. This enables material from the lumen (L) of the blood vessels to pass through the BBB to the parenchyma in the tumor space.

CHAPTER 2 – A Method for Facile Production of Variable Lymphocyte Receptors Using SHuffle *E. coli*

This chapter was adapted from the submitted manuscript "A Method for Facile Production of Variable Lymphocyte Receptors Using SHuffle *E. coli.*" This chapter was completed with tremendous support from Dr. Jim Thoden who helped with many of the daily lab tasks, provided input on construct and plasmid design, and supported many of the large scale-up productions of protein and purification. Thank you, Jim! This chapter was all completed with the support of Prof. Moriah Katt and Candis Dancy who provided support with the bEnd.3 ECM assay; Hannah Bachmeier who support many of the initial refolding experiments; Seth Gehrke who provided support with characterization of the VLRs; Prof. Ivan Rayment who provided support on the construct design and purification; Prof. Hazel Holden who provided lab resources, and Prof. Eric Shusta who supported experimental design and manuscript revisions.

2.1 Abstract

Variable lymphocyte receptors (VLRs) are the antigen receptors of jawless vertebrates such as lamprey. VLRs are of growing biotechnological interest for their ability to bind antigenic targets with higher affinity than traditional immunoglobulins. However, VLRs are disulfide bonded proteins that are often challenging to produce with expression involving genetic modifications, fusion partners, non-scalable host cell lines, inclusion body formation, and poor yield. As a potential expression platform option, the SHuffle *E. coli* strain has been genetically altered to allow cytoplasmic disulfide bond formation by mutations to thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) to create an oxidative cytoplasm that could be used as an expression platform for VLRs. Furthermore, the SHuffle strain expresses disulfide bond isomerase DsbC in the cytoplasm to promote correct disulfide bond pairing. Here, we demonstrate that the Shuffle strain can produce high yield, monomeric VLRs with titers ranging from 3 to 38 milligrams of VLR per liter of SHuffle culture. Three VLRs (P1C10, RBC36, VLRA.R2.1) were expressed in SHuffle *E. coli* with direct comparison to the Rosetta *E. coli* strain. All VLRs were validated for correct sequence, purity, and activity. For all VLRs, SHuffle *E. coli* produced 2 to 9 times more soluble VLRs than Rosetta *E. coli*. Furthermore, the fraction of soluble material relative to the insoluble material was 2 to 6 times greater in SHuffle *E. coli* than Rosetta *E. coli* for all VLRs. Overall, these results suggest that the *E. coli* SHuffle strain is a convenient and effective expression system for producing large amounts of native, monomeric VLR.

2.2 Introduction

Variable lymphocyte receptors (VLRs) are a class of leucine rich repeat (LRR) proteins that are of interest in protein engineering research. As antigen receptors derived from the adaptive immune systems of jawless vertebrates like hagfish and lamprey [22, 51, 52], VLRs share similar antigen-targeting capabilities as that of immunoglobulin (Ig) antibodies [26, 27]. Unlike mammalian host species of traditional Ig-antibodies, lamprey diverged from mammals over 500 million years ago, giving VLRs an advantage of recognizing antigenic targets that mammalian antibodies struggle to recognize [53]. Notably, VLRs that are capable of targeting the brain vasculature [53], carbohydrates [27, 54], cell markers [55], and many other conserved mammalian proteins [19, 56, 57] have been identified.

As a result of growing interest in their antigen-targeting capabilities, recombinant VLRs have been produced in a variety of expression systems including *Saccharomyces cerevisiae* [26, 48, 53, 58, 59], HEK293 mammalian cells [22, 48, 53, 55, 57], *Escherichia coli* [7, 58-61], and Hi5 insect cells [54, 56]. While these expression systems have proven capable of producing VLRs, existing production methods present several challenges. In their typical recombinant form, VLRs exist as monomeric, solenoid-shaped, LRR proteins with four disulfide bonds. To improve production of recombinant VLRs, various strategies have been utilized to express properly folded, monomeric VLRs. Yeast Surface Display (YSD) technology has been employed for successful high-throughput screening of VLRs tethered to the surface of yeast and for soluble expression in *S. cerevisiae*, but secreted protein yields are limited [48, 53]. VLRs also have been

engineered with folding partners, including Fc domains, for improved expression in HEK293 cells, but these approaches require further downstream processing to release the unaltered monomeric VLRs [48, 53, 55]. Furthermore, modifications to the amino-terminus of VLRs have been used to improve expression in *E. coli* [7, 60]. When larger amounts of monomeric material are desired without genetic changes to the protein, VLRs can be produced by stable transfection in high-density bioreactors [22], baculovirus transfection systems [62], and inclusion body isolation and refolding [58, 61]. However, many of these strategies are notoriously labor intensive, produce low yields of properly folded protein, do not work for all VLRs, and do not address the need for genetically unmodified, monomeric VLR production at larger scales [63-65].

While *E. coli* is generally a preferred host for large-scale recombinant protein expression due to its accessibility, ease of use, affordability, and ability to overexpress large quantities of recombinant protein [66], standard *E. coli* production strains like the Rosetta strain are not always ideal for proteins like VLRs that possess disulfide bonds. *E. coli* is maladapted for disulfide bond formation due to the naturally reducing environment of the cytoplasm [67]. Recombinant proteins are often manufactured in the cytoplasm of *E. coli,* but disulfide bonds are only able to form in the oxidative environment of the *E. coli* periplasm [67]. Thus, any recombinant proteins that require disulfide bond formation must be shuttled to the periplasm via chaperones and leader sequences such as pelB [68]. Unfortunately, since proteins are translated in the cytoplasm, many proteins begin to fold before reaching the periplasm [68-70]. When proteins prematurely fold in the cytoplasm, insoluble inclusion bodies can be formed [68-

70]. For proteins that remain unfolded until they are shuttled to the periplasm, oftentimes only a small percentage will adopt the correct disulfide bond formation due to the large flux of overexpressed proteins and the requirement for oxidative protein folding [58]. Specifically, with respect to VLRs, inclusion bodies are often formed in standard *E. coli* strains in the absence of genetic modification or folding partners as described above [58, 61].

To address these issues, researchers have developed engineered *E. coli* strains adapted for disulfide bond formation. Origami and SHuffle strains are two unique *E. coli* strains capable of cytoplasmic production of recombinant proteins with disulfide bonds [67]. Both Origami, a BL21(DE3) *E. coli* derivative, and SHuffle, a K12 *E. coli* derivative, have a modified disulfide bond formation pathway that diverts disulfide bond formation to the cytoplasm [67]. In brief, thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) are suppressed in Origami and SHuffle *E. coli* to create an oxidative cytoplasmic state to enable the formation of disulfide bonds, as opposed to the naturally occurring reductive environment of wild type *E. coli* cytoplasm [67]. However, unique to the SHuffle strain, disulfide bond isomerase DsbC is expressed in the cytoplasm to aid disulfide bonds isomerization and form the native disulfide linkages in proteins that have multiple non-consecutive disulfide bonds, such as VLRs [67]. In this work, we describe a new approach for VLR expression using the SHuffle *E. coli* strain. We demonstrate large-scale production, with titers ranging from 6 to 38 milligrams of VLR per liter of culture, for producing correctly folded, soluble, monomeric VLRs. Furthermore, the proteins bound their known binding partners, confirming the correct fold of the VLRs.

2.3 Materials & Methods

2.3.1 Construction of Variable Lymphocyte Receptor genes in pET vectors

VLR genes were constructed in pET vectors for expression in *E. coli* (Table S1). Each VLR was engineered as two separate constructs. The first construct had an amino-terminus six-histidine (NT-6xHis) epitope tag with a recombinant Tobacco Etch Virus (TEV) cleavage site. The second construct had a carboxyl-terminus six-histidine (CT-6xHis) epitope tag (Figure 1). RBC36 NT-6xHis, RBC36 CT-6xHis, VLRA.R2.1 NT-6xHis and VLRA.R2.1 CT-6xHis were purchased in pET-30a(+) from Genscript. P1C10 NT-6xHis was purchased in pET-28a(+) from Genscript. P1C10 CT-6xHis was purchased in pET-21b(+) from Genscript. Plasmids were expanded in DH5α chemically competent *E. coli* by the heat shock transformation method [71]. pET30 and pET28 plasmids were plated on LB agar containing 50 μg/ml kanamycin antibiotic as a selection marker, and pET21 plasmid was plated on LB agar containing 100 μg/ml ampicillin antibiotic as a selection marker for overnight incubation at 37 °C. Three colonies were picked from each plate and each colony suspended in 3 ml of LB media containing the appropriate antibiotic selection in disposable inoculation tubes (Thermo Scientific #150268). pET30 and pET28 plasmids were suspended in LB media containing 50 μg/ml kanamycin antibiotic, and pET21 plasmid was suspended in LB media containing 100 μg/ml ampicillin. Cultures were grown overnight at 37 °C with 250 RPM shaking. Plasmids were purified using the ZymoPURE Plasmid Miniprep Kit (#D4208T). 1 ml of bacterial culture was collected in a 1.5 ml tube and centrifuged at maximum speed for 15-20 seconds in a microcentrifuge. Supernatant was discarded, and the

ZymoPURE Plasmid DNA Purification Centrifugation Protocol was followed at room temperature. The final elution of purified plasmid was done with sterile filtered ultrapure water.

2.3.2 pET transformation in expression hosts

Rosetta 2(DE3) Competent *E. coli* (Novagen #71397), SHuffle T7 Express Competent *E. coli* (New England Biolabs #C3029J), and Origami 2(DE3) Competent *E. coli* (Novagen #71346) were transformed in triplicate (n = 3) by the heat shock method with each pET plasmid [71]. For each transformation 50 µl of *E. coli* was thawed on ice. 1 µl of purified plasmid DNA was added to the *E. coli* transformation tube. The *E. coli* transformation tube was gently flicked two times to mix the *E. coli* and DNA. The transformation tubes were incubated on ice for 30 minutes followed by 30 seconds at 42 °C in a pre-warmed water bath. Reaction tubes were moved directly from the water bath to ice for a 2-minute incubation. 80 µl of room temperature SOC Medium was added to each tube. The *E. coli* transformations were incubated at 37 °C for 1 hour with 250 RPM shaking. Each biological replicate was plated on LB agar with the appropriate antibiotic selection and incubated overnight at 37 °C. pET30 and pET28 plasmids were plated on LB agar containing 50 µg/ml kanamycin antibiotic as a selection marker, and pET21 plasmid was plated on LB agar containing 100 µg/ml ampicillin antibiotic as a selection marker.

2.3.3 Variable Lymphocyte Receptor expression studies

Primary cultures of each VLR construct were prepared with appropriate antibiotics in both TB and LB media for initial expression tests. For LB cultures, one plate of colonies was scraped into 100 ml of selective media in 500 mL baffled shake flasks and grown at 37 °C with 200 RPM shaking to an OD_{600} of 0.7-0.9. The LB cultures were then cooled to 20°C, and IPTG was added to each LB culture at a final concentration of 1 mM. The induced LB cultures were grown overnight at 20 °C. For TB cultures, one plate of colonies was scraped into 200 ml of selective media in 500 mL baffled shake flasks and grown at 37 °C with 200 RPM shaking to an OD_{600} of 0.7-0.9. The TB cultures were cooled to 20 °C with 200 RPM shaking and grown for an additional 24 hours. After 24 hours, IPTG was added to each TB culture at a final concentration of 0.1 mM. The induced TB cultures were grown for an additional 18-24 hours at 20 °C.

2.3.4 Insoluble and soluble sample preparation for sodium dodecyl-sulfate polyacrylamide gel electrophoresis

Overnight cultures were harvested the next day by centrifugation. 15 ml samples were centrifuged at 3,500 × g for 10 minutes. The pellet was resuspended in 1.5 ml lysis buffer (50 mM sodium phosphate, 20 mM imidazole, 300 mM sodium chloride, 10% glycerol, pH 8). Samples were maintained on ice and sonicated with a microtip for two intervals of 15-20 seconds at 100% amplitude with 5 minutes of rest between each interval. A 500 µl sample of each lysate was collected for analysis. The 500 µl samples were centrifuged at 21, 330 × g on a benchtop centrifuge for 2 minutes. All supernatant from each sample was collected as the "soluble" fraction and brought to 1.5 ml in ultrapure water. The pellet was collected as the

“insoluble” fraction and resuspended in 1.5 ml of 8 M Urea. Soluble and insoluble samples were prepared for a reduced sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) NuPAGE 4-12% Bis-Tris gel (Invitrogen #NP0323BOX). Reduced samples were prepared with 2 µl of soluble or insoluble sample, 2.5 µl NuPAGE LDS Sample Buffer (4x) (Invitrogen # NP0007), 1 µl NuPAGE Sample Reducing Agent (10x) (Invitrogen # NP0004), and 4.5 µl deionized water. Samples were heated at 70°C for 10 minutes. Samples were loaded at 10 µl per lane. Gels were run with NuPAGE MES SDS Running Buffer (Invitrogen # NP0002) at 200 V for 35 minutes. All gels were run at ambient temperature.

2.3.5 Western blot quantification of insoluble and soluble fractions

NuPAGE gels were transferred to nitrocellulose membranes with 0.2 µm pore size (Invitrogen #LC2000) for western blot analysis [72]. The transfer was done on ice in cold NuPAGE™ Transfer Buffer (Invitrogen #NP0006) at 100 V for 1 hour with stirring. After transfer, membranes were washed with deionized water 4 times for 5 minutes each with agitation. After washing, membranes were blocked with 5% BSA in Tris-Buffered Saline with 0.1% Tween 20 (TBST) for 1 hour at ambient temperature with gentle shaking. Following blocking, the membranes were incubated in primary antibody (Invitrogen 6x-His Tag Monoclonal 4E3D10H2/E3, mouse IgG1) diluted 1:1000 in 5% BSA in TBST overnight at 2–8 °C with gentle shaking. The following day membranes were washed 3 times in TBST for 10 minutes with agitation. After washing the membranes were incubated in the secondary antibody solution (LI-COR IRDye 800CW Donkey Anti-Mouse IgG # 926-32212) in a 1:20,000 dilution in 5% BSA in TBST for 1 hour with agitation

at ambient temperature. After incubation with the secondary antibody, membranes were washed 3 times in TBST for 10 minutes with agitation. Membranes were scanned on a LI-COR scanner using the 800 nm channel. Soluble and insoluble fractions were quantified by densitometric measurement of the signal at 800 nm with background subtraction for each biological replicate.

2.3.6 Scale-up production of Variable Lymphocyte Receptors in SHuffle *E. coli*

SHuffle *E. coli* were transformed as previously described. P1C10 was grown in TB media, and RBC36 and VLRA.R2.1 were grown in LB media according to previous described conditions. Each construct was grown in twelve 1 L cultures in 2 L baffled shake flasks with vented lids. Cultures were harvested at 7,000 × g for 15 minutes. Supernatant was discarded, and the cellular pellets were flash frozen in liquid nitrogen for storage at -80 °C. 40 g of frozen cell pellets were resuspended in 120 ml of lysis buffer (50 mM sodium phosphate, 20 mM imidazole, 300 mM sodium chloride, 10% glycerol, pH 8). Samples were kept on ice and sonicated with a preparatory tip for five intervals of 1 minute each at 85% amplitude with 2 minutes of rest between each interval. Lysates were centrifuged at 45,000 × g for 30 minutes to pellet insoluble material. Supernatants were collected for VLR purification.

2.3.7 Nickel- Nitriloacetic acid purification of Variable Lymphocyte Receptors

VLRs were purified using peristaltic pump chromatography. Cleared supernatant was loaded onto a column packed with HisPur Ni-NTA resin (Thermo Scientific #88223). 10 ml of resin was used for each set of 12 L cultures. After two washes with wash buffer (20 mM sodium phosphate, 300 mM sodium chloride, 25 mM imidazole, pH 8), protein was eluted with a 0 to 300 mM gradient of imidazole in elution buffer (20 mM sodium phosphate, 300 mM sodium chloride, 300 mM imidazole, pH 8). Fractions were collected according to peak absorbances at A_{280} using a peristaltic pump with a 2 ml/minute flow rate. Fractions were assessed using SDS-PAGE, and fractions containing pure protein were pooled together for buffer exchange. All VLRs were dialyzed into 10 mM TRIS, 200 mM NaCl pH 8. Following dialysis, VLRs were concentrated using Amicon Ultra-15 Centrifugal Filter Units with molecular weight cutoff of 3 kDa (Amicon #UFC900396).

2.3.8 Size Exclusion Chromatography

Size exclusion chromatography (SEC) of purified VLRs was performed using an ÄKTA pure 25M for evaluation of purity. Samples were separated using a Superdex 200 Increase 10/300 GL column (Cytiva # 28990944) with 10 mM TRIS, 200 mM NaCl pH 8 mobile phase. 0.2 ml samples were injected and eluted with a flowrate of 0.75 ml/min. Spectra were analyzed in Unicorn 7.

2.3.9 Extracellular Matrix Binding Assay and K_D calculation of P1C10

bEnd.3 (ATCC #CRL2299) culture was stored in liquid nitrogen prior to use. Immediately prior to use, the culture vial was thawed in a 37 °C water bath. Using aseptic technique, the culture was transferred to conical tube with 9 ml prewarmed complete culture medium and centrifuged at 125 × g for 5 minutes. The cell pellet was resuspended in 9 ml of prewarmed complete medium. The culture was transferred to a T75 culture flask and incubated at 37 °C 5% CO_2 with medium renewal every 48 hours. When cells reached 80-90% confluence, 1,000 to 2,000 cells were seeded per well in a 96-well flat bottom plate (Thermo Scientific #165305) and incubated at 37 °C 5% CO_2 for 48 to 72 hours until 100% confluent. Media was removed, and the cells were incubated twice with either versene for 30 minutes at 37 °C or PBS + 1% Triton-X-100 for 5 minutes at 4 °C to remove cellular debris and maintain ECM. Removal of cellular debris was confirmed via microscopy. Cells were incubated with DNAse for 30 minutes at ambient temperature. Finally, cells were washed 3 times with PBS, and a bicinchoninic acid (BCA) assay (Thermo Scientific #23225) was performed to confirm ECM proteins remained in the wells. Next, VLRs were diluted in assay diluent (1% BSA + 1.5% goat serum in PBS), and 100 µl of sample at saturating concentration of 100 µM was added per well and incubated for 1 hour at ambient temperature. Following 5 washes, the plate was incubated for 1 hour at ambient temperature with mouse anti-6xHis diluted to 100 ng/ml in assay diluent (6x-His Tag Monoclonal Antibody 4E3D10H2/E3, Thermo Scientific #MA1-135). The plate was washed 5 times, labeled with goat anti-mouse diluted to 1 µg/ml in assay diluent (Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 488, Thermo Scientific # A32723), and incubated for an additional 1 hour at ambient temperature. Following the final

incubation, the plate was washed 5 times and read at excitation 493 nm and emission 518 nm on a BioTek SynergyH1 plate reader. All samples were assayed in 4 replicates.

2.3.10 Isothermal titration calorimetry and K_D calculation of VLRA.R2.1 and RBC36

A VP-ITC microcalorimeter (Microcalorimetry, Inc.) was used to generate binding data for VLRA.R2.1 and RBC36. Prior to Isothermal Titration Calorimetry (ITC), VLRA.R2.1 and RBC36 NT-6xHis constructs were incubated with TEV protease for 48 hours at 4 °C to remove the 6xHis epitope tags. After passage over Ni-NTA column to remove protease and uncleaved protein, the TEV-cleaved constructs were subsequently dialyzed into 10 mM TRIS, 200 mM NaCl pH 8 dialysis buffer. The same dialysis buffer was used to dialyze the ITC injectants, Hen Egg Lysozyme (HEL) (Roche #10837059001) and H-trisaccharide (Sigma #sc-300786). For VLRA.R2.1, 0.35 mM HEL was loaded into the injector and titrated at 3 µl injection volumes into the cell that contained 0.033 mM VLRA.R2.1 at 25 °C. For RBC36, 1.05 mM H-trisaccharide was loaded in the injector and titrated at 3 µl injection volumes into the cell that contained 0.11 mM RBC36 at 10 °C. All injections had a 20 second duration, 180 second spacing, and 307 RPM stir speed. K_D values were calculated using nonlinear least squares fits of ITC data with a single-site binding model in Origin 7.0 software.

2.4 Results

2.4.1 SHuffle *E. coli* produce soluble Variable Lymphocyte Receptors

We first set out to produce a VLR known as P1C10, which has been shown to have brain drug delivery capabilities by targeting pathologically-exposed brain extracellular matrix (ECM) [48]. As such, it was desired to produce P1C10 at large-scale for downstream characterization and use. Previously, P1C10 had been expressed at small scale via yeast (*S. cerevisiae)* and as Fc-fusions in mammalian culture (HEK293), both with limited yields [48]. To expand production capacity, we attempted to produce P1C10 in a standard *E. coli* protein production strain, Rosetta 2(DE3), via inclusion body refolding approaches following previously published production strategies [58, 61]. The Rosetta BL21(DE3) derivative is one of the most common bacterial strains used for recombinant protein production because in comparative studies with other standard *E. coli* strains, the Rosetta derivative consistently produces the highest recombinant protein titer, lowest byproduct, and the fewest proteases [73]. However, despite exploring several refolding conditions, purified inclusion bodies from the Rosetta expression system were never successfully refolded where the desired P1C10 monomers were the dominant species (Supplement Table S2 and Figure S1). Thus, we decided to explore SHuffle *E. coli* as an alternative bacterial expression system to produce soluble, properly folded, monomeric P1C10. Moving forward, we also wanted to validate any of our expression approaches and production workflow by evaluating two additional VLRs known as RBC36 and VLRA.R2.1. RBC36 is a well-characterized VLR with known binding to H-trisaccharide, and

monomeric RBC36 has been produced in insect cells [54] and yeast [48, 53] as well as a dimeric Fc fusion in mammalian cells [53]. VLRA.R2.1 binds to Hen Egg Lysozyme (HEL), and it has been produced from refolding inclusion bodies expressed in *E. coli* [58]. Thus, the set of 3 VLRs offer a representative sample set to validate a new VLR production strategy.

As an initial comparison, VLRs were expressed in both SHuffle and Rosetta bacterial strains using a cytoplasmic expression construct under the control of the T7 Promoter (Figure 1). Two different constructs with 6xHis tags included at either the N-terminus (NT) or C-terminus (CT) of the constructs for purification and detection were evaluated for each VLR. Both NT-6xHis and CT-6xHis constructs were assessed since protein solubility can be dependent on the location of the epitope tag [74]. The NT-6xHis constructs also included a TEV protease cleavage site for downstream uses that may require tagless VLRs. All cultures were induced at 20 °C, and bacteria recovered 18-24 hours after IPTG induction. Soluble and insoluble bacterial cell fractions were collected from each of the 6 VLR constructs for both the Rosetta and SHuffle strains to quantify the relative amounts of soluble VLR produced and the fraction of total VLR protein that was soluble. After bacterial lysis, the cell pellets were used to quantify insoluble VLR, and the lysate supernatant was used to quantify soluble VLR (see materials and methods for details).

Western blots were quantified for each VLR construct (Figure 2a-f). Total soluble VLR production was uniformly higher in SHuffle *E. coli* compared to Rosetta *E. coli* for all 6 constructs (Figure 2g-i). P1C10 NT-6xHis and CT-6xHis constructs yielded 7.3-fold and 3.9-fold, respectively, higher soluble protein amounts in SHuffle compared to Rosetta *E. coli*. RBC36 NT-

6xHis and CT-6xHis constructs yielded 9.4-fold and 5.7-fold, respectively, higher soluble protein amounts in the SHuffle strain. Finally, VLRA.R2.1 NT-6xHis and CT-6xHis constructs yielded 2-fold and 4.6-fold, respectively, higher soluble protein amounts in SHuffle compared to Rosetta *E. coli*. When comparing the yields for NT-6xHis and CT-6xHis, the P1C10 CT-6XHis construct produced slightly more soluble protein in the SHuffle strain (1.1-fold increase) compared to its NT-6xHis counterpart, but this improvement did not extend to RBC36 or VLR2.1.

We also assayed the fraction of the total VLR produced that was soluble to compare the cytoplasmic folding efficiency of the Rosetta and SHuffle expression systems (Figure 2j-l). As expected, given the differences in the cytoplasmic folding environment, the SHuffle strain led to uniformly higher fractions of soluble VLR. The P1C10 NT-6xHis and CT-6xHis constructs had 2.4-fold and 1.8-fold, respectively, higher fractions of soluble material in the SHuffle strain compared to Rosetta. Similarly, RBC36 NT-6xHis and CT-6xHis constructs had a 5.8-fold and 2.9-fold increase in the fraction of soluble VLR in SHuffle compared to Rosetta. VLRA.2R.1 performed similarly with a higher fraction of soluble VLR in SHuffle compared to Rosetta for the NT-6xHis and CT-6xHis constructs with 1.6-fold and 3.3-fold increases, respectively. Taken together, the major improvements in soluble production were a result of the SHuffle strain with more minor effects of the 6xHis epitope tag location. In addition to the Rosetta and SHuffle strains, we tested the Origami strain for its ability to express VLR P1C10 given its similarity to SHuffle with an oxidizing cytoplasm and since it was previously used to produce a highly engineered VLR scaffold [7]. Western blot analysis of P1C10 CT-6xHis showed no significant difference of total and fraction soluble VLR in SHuffle compared to Origami (Figure S2).

However, since the SHuffle strain has DsbC for disulfide bond isomerization, this strain was chosen for our subsequent scale-up.

2.4.2 Variable Lymphocyte Receptors produced in the SHuffle strain can be purified as monomeric protein preparations

SHuffle-produced VLRs were purified using Ni-NTA resin (see materials and methods). VLR purity was assessed by reduced SDS-PAGE with Coomassie staining (Figure 3). A major band for each VLR construct was observed at the correct molecular weight. Purified VLRs were then assessed by liquid chromatography-mass spectrometry (LC-MS) to confirm that all purified VLRs were the correct proteins and identified peptides covered 99-100% of the amino acid sequences for the VLRs (Table 1). Finally, purified VLRs were analyzed by SEC. All VLRs produced a dominant, symmetric peak with purity ranging from 65-96% (Figure 4a). Moreover, SEC chromatograms showed that the dominant VLR peaks for RBC36 and VLRA.R2.1 were at the expected retention time for the theoretical molecular weights of the monomers relative to the molecular weight standards (Figure 4b-d). P1C10 VLR constructs both eluted later than would be expected from the molecular weight standards. However, because the SDS-PAGE analysis of P1C10 does not show any low molecular weight species, we attributed the longer retention time of P1C10 on the SEC column to hydrophobic interactions between P1C10 and the stationary phase of the SEC column which are known to lead to increased retention times in SEC [75]. In sum, the SHuffle-produced VLRs can be purified as monomeric, soluble protein products.

2.4.3 Variable Lymphocyte Receptor production using the SHuffle strain is scalable

For many applications, it is desirable to produce large amounts of VLR protein. Thus, once production of soluble, monomeric VLRs in the SHuffle strain was verified, protein production was scaled-up to generate bulk protein for downstream analyses. Twelve, 2 L shake flasks with 1 L of VLR-producing SHuffle cultures were used for each of the 6 VLRs. Following induction of protein expression, bacterial pellets containing the expressed VLR were recovered by centrifugation and flash frozen in liquid nitrogen for further processing. VLRs were purified as described above with HisPur Ni-NTA purification. The yields of purified protein for each VLR are listed in (Table 2) and ranged from 20-38 mg/L for VLRs P1C10 and RBC36 with lower levels of 3.3-6.3 mg/L for VLRA.R2.1. In addition, all proteins could be concentrated to $\geq$ 15 mg/ml for downstream evaluation.

2.4.4 SHuffle-produced P1C10 is properly folded and active

The fidelity of purified P1C10 produced in the SHuffle strain was verified through binding to bEnd.3 cell produced ECM, the substrate which was used to identify the P1C10 VLR clone [48]. In an ELISA format, both purified P1C10 NT-6xHis and P1C10 CT-6xHis were tested for binding to ECM derived from decellularized bEnd.3 cell cultures. RBC36 NT-6xHis was used as a non-ECM targeted negative control in this assay. Both P1C10 constructs exhibited binding to the bEnd.3 ECM when compared to the RBC36 NT-6xHis negative control (Figure 5). Additionally,

the purified P1C10 product was subsequently used for x-ray structure determination (PDB ID: 9CJ0). Thus, SHuffle-produced P1C10 VLR was demonstrated to be correctly folded and active.

2.4.5 SHuffle-produced RBC36 and VLRA.R2.1 are properly folded and active

Finally, to verify the fidelity of purified RBC36 and VLRA.R2.1, we performed isothermal titration calorimetry (ITC) with their known binding partners. RBC36 has been reported to bind H-Trisaccharide with a published K_D of 8.3 µM [54]. For RBC36 ITC, H-trisaccharide was injected into an RBC36 containing solution (Figure 6a) [54], and the resulting ITC data yielded a K_D of 2.6 µM for RBC36 produced in SHuffle *E. coli* (Figure 6b). VLRA.R2.1 is known the bind HEL with a published K_D of 7 nM by ITC [28]. HEL was injected into a VLRA.R2.1 containing solution (Figure 6c) [28], and the resulting ITC data yielded a K_D of 2 nM for VLRA.R2.1 (Figure 6d). These results indicate that the VLRA.R2.1 and RBC36 proteins produced and purified from SHuffle bacteria were correctly folded and active.

2.5 Discussion

VLRs are an emerging class of proteins useful in many biomedical applications. The moderate to large-scale production of functional, monomeric VLRs will be crucial for characterization and analysis of VLRs used in diagnostics and therapeutics. In the present study, we have demonstrated the ability to produce a purified yield of as much as ~40 milligrams of VLR per liter of SHuffle *E. coli* culture. VLRs P1C10, RBC36, and VLRA.R2.1 were chosen for the expression studies in SHuffle as these are well-characterized VLRs but are subject to challenging production methods such as production as fusion proteins, production in insect cells and inclusion body refolding [48, 54, 58].

Of the three VLRs tested, VLRA.R2.1 was the only VLR previously produced in *E. coli.* Previously, VLRA.R2.1 was produced in BL21-CodonPlus (DE3)-RIL *E. coli* as an inclusion body [58]. The BL21-CodonPlus (DE3)-RIL *E. coli* production method required inclusion body separation, washing, resuspension, solubilization and refolding [58]. Following refolding, the protein required dialysis, concentration and two forms of purification [58]. The labor-intensive example for VLRA.R2.1 demonstrates the utility of the SHuffle strain for VLR production. In contrast to the inclusion body expression approach, VLRA.R2.1 can be simply produced and purified from the SHuffle strain at titers of 3-6 mg/L. Interestingly, VLRA.R2.1 had the lowest yield of the three VLRs produced, 3.3-6.3 mg of protein per liter of culture compared to 18-38 mg of protein per liter of culture for the other VLRs. This difference could be due to the biophysical differences in VLR isoforms, VLRB and VLRA. RBC36 and P1C10 are both in the VLRB family of

proteins, which are secreted from B cell-like lymphocytes [76]. In contrast, VLRAs are cell surface receptors that are displayed on the surface of T Cell-like lymphocytes [58, 76]. By analogy, native T Cell Receptors (TCRs) are known to be unstable with poor solubility when removed from the T Cell environment as they have evolved as transmembrane proteins that rely on the T Cell membrane for stabilization [77]. Thus, it is also possible that VLRAs may similarly be less soluble when removed from the cell surface. Perhaps because of the instability of VLRAs, there are far fewer solved protein structures of VLRAs compared to VLRBs. The ability to produce soluble VLRAs in SHuffle strain will likely help enable future work to better understand this type of VLR.

Origami *E. coli* previously was used to express an engineered VLR. The engineered VLR, a scaffold VLR protein termed “repebody,” was engineered for production in Origami *E. coli [7].* The repebody VLR required N-terminal domain engineering using the internalin-B cap to promote solubility when produced in the Origami strain. Following the redesign, the yield of the repebody in Origami *E. coli* was 2 mg of protein per liter of culture [7]. Here, with different VLR proteins, the SHuffle strain demonstrated higher yields even in the absence of engineering the constructs. However, when we directly compared the production capacity of the SHuffle and Origami strains with the same P1C10 CT-6xHis VLR, total soluble yields were the same. Thus, while the SHuffle strain has a putative advantage as the result of cytoplasmic DsbC expression to help promote correct disulfide pairing, both strains could be used successfully to produce VLRs.

The location of epitope tags can affect solubility, and these effects have been demonstrated to be sequence-dependent [74]. However, for the three VLRs expressed in *E. coli*, the location of the 6xHis epitope tag had only a minor effect on soluble yields of P1C10 and no effect for RBC36 or VLRA.R2.1 in the SHuffle strain. By contrast, the Rosetta strain was more significantly affected by a nonoptimal tag, varying by as much as 2-fold in soluble yields . The VLR solubility being less affected by the location of the epitope tag when produced in SHuffle further supports the use of the SHuffle strain for VLR expression.

Finally, we compared the VLR yields to those reported for disulfide bonded IgGs that were also expressed using the SHuffle strain. Production titers for IgGs have been reported to be as low as 0.1 mg of protein per culture liter, with other antibodies such as Humira reported yields as high as 168 mg/L or 427 mg/L when co-expressed with PDI-GPx7 fusion in high density fermentation [78]. As other examples, Trastuzumab (3.7 mg/L), Cetuximab (4.5 mg/L) and Ranibizumab (10 mg/ml) were also produced in the SHuffle strain [78]. Interestingly, Tastuzumab also was produced in SHuffle as an scFv [79]. The scFv production was compared between BL21(DE3) and SHuffle [79]. While the BL21(DE3) strain produced the Trastuzumab scFv at 214 mg/L compared to the SHuffle yield of 147 mg/L, a solubility assessment showed that the SHuffle strain fraction of soluble yield was approximately twice that of the BL21(DE3) soluble fraction, making the SHuffle production more suitable for soluble Trastuzumab scFv production [79]. Therefore, compared to mammalian IgGs and scFvs, the VLRs tested here were produced in similar amounts.

2.6 Conclusion

In conclusion, we have demonstrated the utility of SHuffle bacteria for the facile expression and scale-up of VLRs. Importantly, we were able to produce three different VLRs without the need to engineer VLR mutations to assist production, add fusion partners, or refold from insoluble inclusion bodies. As such, this study provides a method to produce high quality VLRs using the simple, scalable SHuffle bacterial host.

2.7 Tables

Table 2-1. Liquid Chromatography – Mass Spectrometry Results for Purified Variable Lymphocyte Receptors

VLR	Location of 6xHis-tag	Unique Peptides	Unique Spectra	Total Spectra	Amino Acid Coverage (%)
P1C10	N-Terminus	4	10	1130	100
	C-Terminus	4	11	991	100
RBC36	N-Terminus	32	166	1292	100
	C-Terminus	31	148	1022	99
VLRA.R2.1	N-Terminus	36	105	848	100
	C-Terminus	37	107	858	99

Table 2-2. Scaled-up production of Variable Lymphocyte Receptors in the SHuffle strain

VLR	Location of 6xHis-tag	Total Culture Volume (L)	Wet Cell Pellet Mass (g)	Purified Protein Mass (mg)	Purified Protein Concentration (mg/ml)	Purified Protein Yield per Culture (mg/L)
P1C10	N-Terminus	12	30	240	18	20
	C-Terminus	12	40	450	18	38
RBC36	N-Terminus	12	40	240	20	20
	C-Terminus	12	40	216	18	18
VLRA.R2.1	N-Terminus	12	40	40	16	3.3
	C-Terminus	12	40	75	15	6.3

2.8 Figures

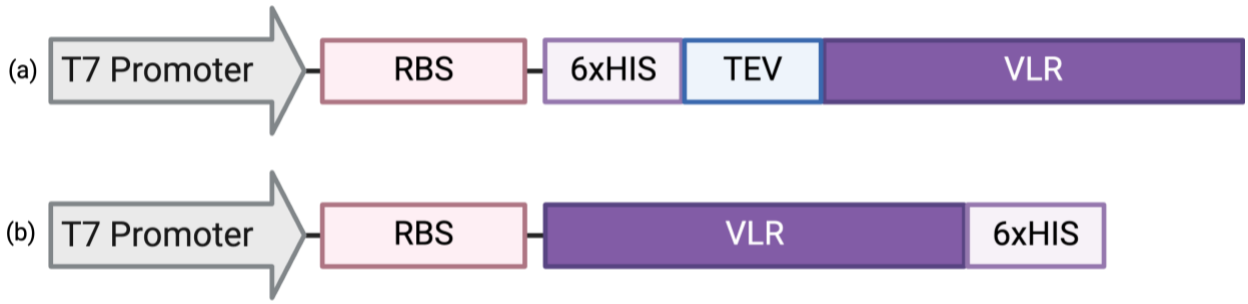


Figure 2-1. Variable Lymphocyte Receptor expression constructs in pET vectors with T7 promoters. (a) NT-6xHis constructs were designed with a 6xHis tag and TEV cleavage site at the N-terminus of the VLR. (b) CT-6xHis constructs were designed with a 6xHis tag at the C-terminus of the VLR.

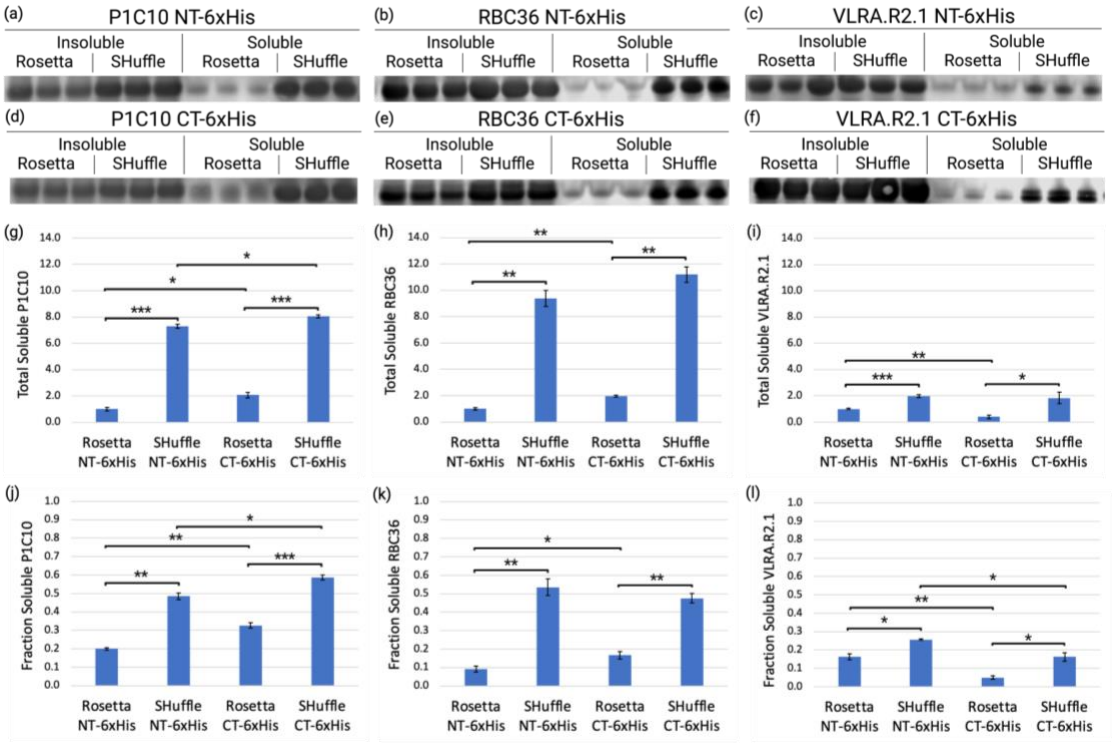


Figure 2-2. Comparison of soluble Variable Lymphocyte Receptors yields using the SHuffle and Rosetta bacterial strains. (a-f) Western blot comparison for 6 VLR constructs. Bacterial lysates from Rosetta and SHuffle cultures were separated into insoluble and soluble fractions via centrifugation. Samples were run on reduced 4-12% Bis-Tris gels, transferred to a nitrocellulose membrane, and probed with anti-6xHis antibody. (g-i) Densitometric quantification of the of total soluble VLR. Data were normalized to Rosetta NT-6xHis for each VLR. (j-l) Densitometric quantification of the fraction soluble was assessed by dividing the total soluble material by the sum of the insoluble and soluble material. Data in all panels represent the mean and standard error of 3 independent bacterial transformants (n=3). Data are expressed as mean ± S.D. and statistical significance was determined by a one-way ANOVA with Tukey's post-hoc test. * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$, all comparisons with no p-value indicated were not statistically significant.

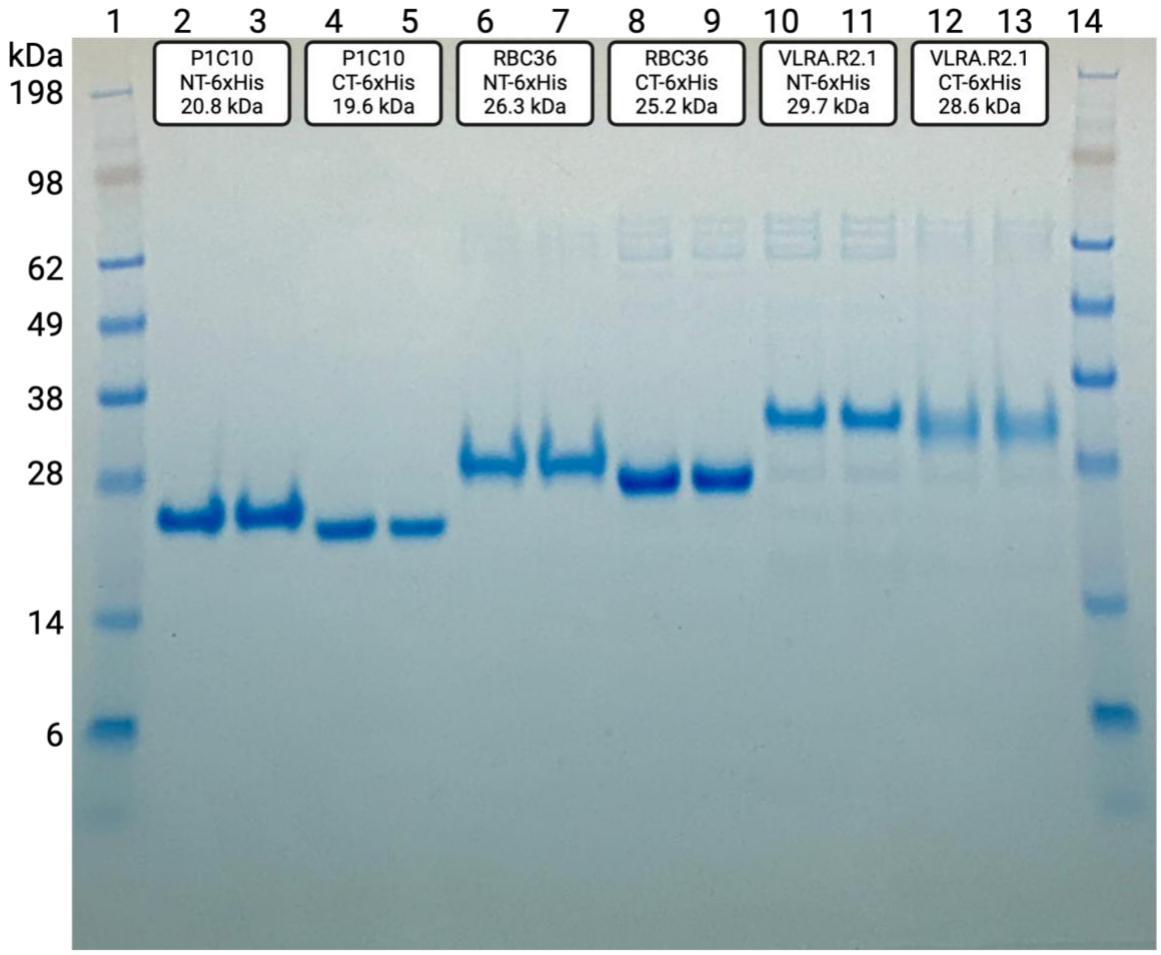


Figure 2-3. Reducing SDS-PAGE assessment of purified SHuffle-produced Variable Lymphocyte Receptors. Ni-NTA-purified VLR preparations were resolved on a reducing 4-12% Bis-Tris gel and visualized by Coomassie staining. Purified VLRs were loaded in duplicate lanes as depicted on the top of the gel. Molecular weight standards were loaded in lanes 1 and 14. Theoretical molecular weights for each construct are noted at the top of the gel.

(a)

VLR	Theoretical MW (kDa)	Purity (%)
P1C10 NT-6xHis	20.8	96
P1C10 CT-6xHis	19.6	85
RBC36 NT-6xHis	26.3	92
RBC36 CT-6xHis	25.2	80
VLRA.R2.1 NT-6xHis	29.7	65
VLRA.R2.1 CT-6xHis	28.6	65

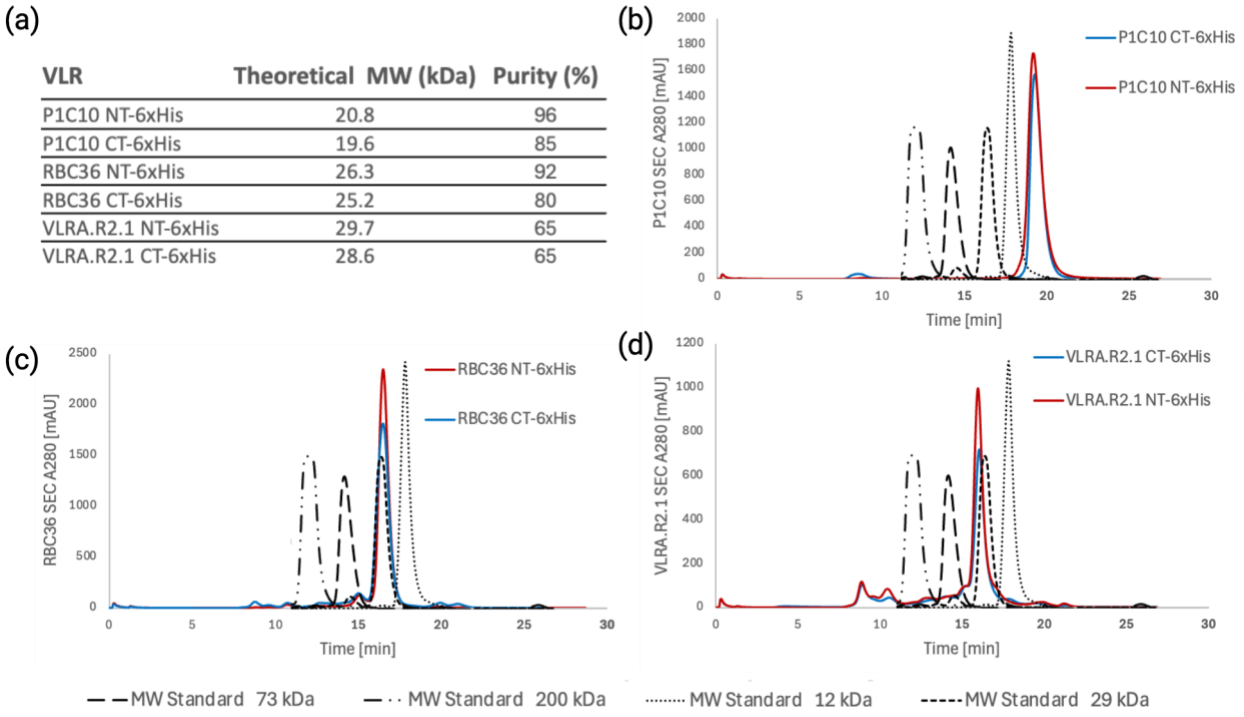


Figure 2-4. Analysis of purified SHuffle-produced Variable Lymphocyte Receptors by size exclusion chromatography. (a). The monomeric purity for each VLR construct was assessed by SEC by integrating the largest peak and comparing to the total area under the elution curve. (b-d) Purified VLR preparations were separated on a Superdex 200 Increase 10/300 GL column. The SEC chromatograms for each VLR were overlaid with that for the molecular weight standard to compare the monomeric peak retention times relative to the standard.

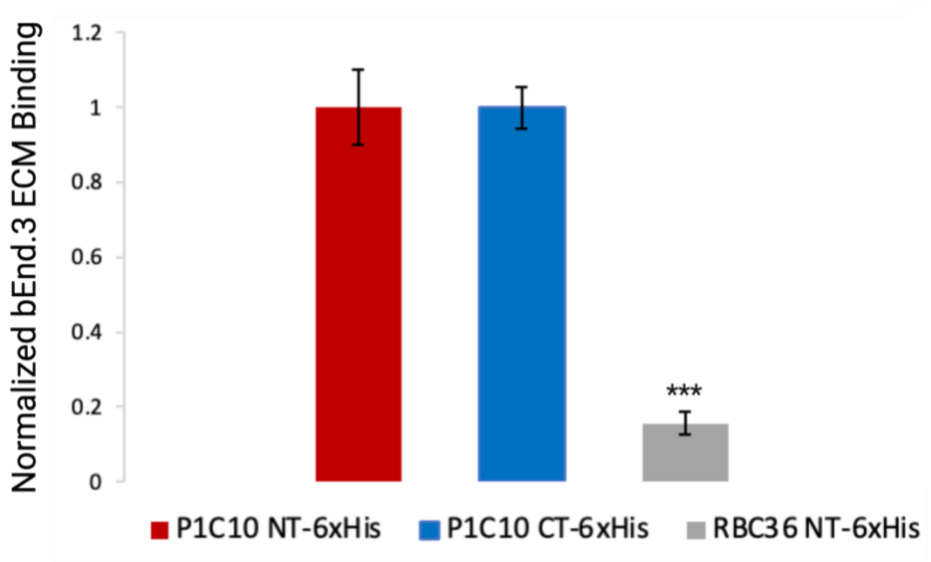


Figure 2-5. bEND.3 Extracellular Matrix ELISA to confirm activity of P1C10 Variable Lymphocyte Receptors. P1C10 NT-6xHis and P1C10 CT-6xHis binding was assayed at saturating concentrations with decellularized bEnd.3 ECM. RBC36 NT-6xHis was used as a non-binding control. Data represent the mean ± S.D. of assay replicates (n=4). Statistical significance was determined by a one-way ANOVA with Tukey's post-hoc test. *** represents $p < 0.001$, all comparisons with no p-value present were not statistically significant.

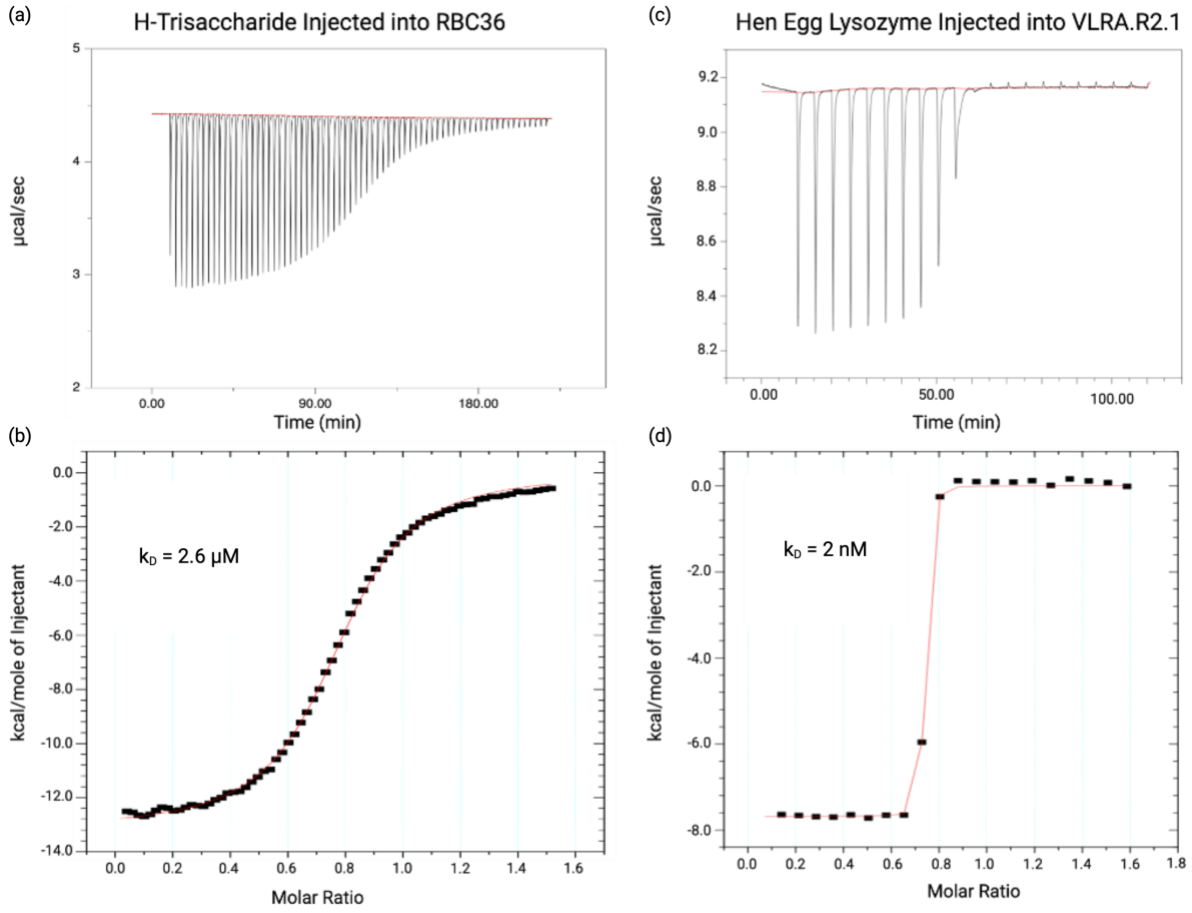


Figure 2-6. ITC with SHuffle-produced RBC36 and VLRA.R2.1 and their known binding partners to confirm activity. (a) H-trisaccharide was titrated into TEV-cleaved RBC36 (no 6xHis tag). (b) ITC data was fit to a nonlinear least squares binding model producing a k_D of 2.6 µM for RBC36. (c) HEL was titrated into TEV-cleaved VLRA.R2.1 (no 6xHis tag). (d) ITC data was fit to a nonlinear least squares binding model producing a k_D of 2 nM for VLRA.R2.1.

2.9 Supplement

VLR	Sequence
P1C10 NT6xHis	HHHHHHSSGENLYFQGHMSACPSQCSCDQTTVKCHSRRLTSVPAGIPTTTKILR LYSNQITKLEPGVFDHLVNLEKLYISWNQLSALPVGVFDKLTKLTHLSLGYNQLKS VPRGAFDNLKSLTHIWLLNNPWDCECSDILYLKNWIVQHASIVNLQGHGGVDN VKCSGTNTPVRAVTEASTSPSKCP
P1C10 CT6xHis	HMSACPSQCSCDQTTVKCHSRRLTSVPAGIPTTTKILRLYSNQITKLEPGVFDHL VNLEKLYISWNQLSALPVGVFDKLTKLTHLSLGYNQLKSVPRGAFDNLKSLTHIW LLNNPWDCECSDILYLKNWIVQHASIVNLQGHGGVDNVKCSGTNTPVRAVTE ASTSPSKCPLEHHHHHH
RBC36 NT6xHis	HHHHHHSSGENLYFQGGGSCPSQCSCSGTTVDCRSKRHASVPAGIPTNAQILYL HDNQITKLEPGVFDSLINLKELYLGSNQLGALPVGVFDSLTQLTVLDLGTNQLTV LPSAVFDRLVHLKELFMCCNKLTELPRGIERLTHLTHLALDQNQLKSIPHGAFDRL SSLTHAYLFGNPWDCECRDIMYLRNWVADHTSIAMRWDGKAVNDPDSAKCA GTNTPVRAVTEASTSPSKCP
RBC36 CT6xHis	HMGSCPSQCSCSGTTVDCRSKRHASVPAGIPTNAQILYLHDNQITKLEPGVFDS LINLKELYLGSNQLGALPVGVFDSLTQLTVLDLGTNQLTVLPSAVFDRLVHLKELF MCCNKLTELPRGIERLTHLTHLALDQNQLKSIPHGAFDRLSSLTHAYLFGNPWD CECRDIMYLRNWVADHTSIAMRWDGKAVNDPDSAKCAGTNTPVRAVTEAST SPSKCPLEHHHHHH
VLRA.R2.1 NT6xHis	HHHHHHSSGENLYFQGGGSTCETVTGCTCNEGKKEVDCQGKSLDSVPSGIPAD TEKLDLQSTGLATLSDATFRGLTKLTWLNLDYNQLQTLSAGVFDDLTELGTLGLA NNQLASLPLGVFDHLTQLDKLYLGGNQLKSLPSGVFDRLTKLKELRLNTNQLQSI PAGAFDKLTNLQTLSLSTNQLQSVPHGAFDRLGKLQTITLFGNQFDCSRCETLYL SQWIRENSNKVKDGTGQNLHESPDGVTCSDGKVVRTVTNETLKYECPFVE
VLRA.R2.1 CT6xHis	HMGSKTCETVTGCTCNEGKKEVDCQGKSLDSVPSGIPADTEKLDLQSTGLATLS DATFRGLTKLTWLNLDYNQLQTLSAGVFDDLTELGTLGLANNQLASLPLGVFDH LTQLDKLYLGGNQLKSLPSGVFDRLTKLKELRLNTNQLQSIPAGAFDKLTNLQTLS LSTNQLQSVPHGAFDRLGKLQTITLFGNQFDCSRCETLYLSQWIRENSNKVKDG TGQNLHESPDGVTCSDGKVVRTVTNETLKYECPFVELEHHHHHH

Table 2-S1

The amino acid sequences for each Variable Lymphocyte Receptor encoded by the expression construct shown in Figure 1.

Condition	Base Refolding Buffer (900 µL)	100 mM EDTA (µL)	200 mM GSH (µL)	100 mM GSSG (µL)	MilliQ (µL)	Protein (µL)
1	55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	10	2	28	50
2	440 mM L-Arg, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	10	4	26	50
3	880 mM L-Arg, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	5	10	25	50
4	550 mM Guanidine, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	10	4	26	50
5	550 mM Guanidine, 440 mM L-Arg, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	5	10	25	50
6	550 mM Guanidine, 880 mM L-Arg, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	10	2	20	50
7	1.1 M Guanidine, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	5	10	25	50
8	1.1 M Guanidine, 440 mM L-Arg, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	10	2	28	50
9	1.1 M Guanidine, 880 mM L-Arg, 55 mM Tris, 21 mM NaCl, 0.88 mM KCl, pH 8.2	10	10	4	26	50

Table 2-S2. Refolding conditions for P1C10 expression as inclusion bodies in Rosetta *E. coli*.

The Pierce Protein Refolding Kit (Thermo Scientific #89867) was used for a broad assessment of refolding conditions for VLR P1C10 inclusion bodies. Inclusion bodies were isolated from *E. coli* lysates and purified using denaturing conditions. Inclusion bodies were denatured in guanidine hydrochloride and reduced with beta-mercaptoethanol before being buffer exchanged into Tris-NaCl for Ni-NTA purification. Purified inclusion bodies were concentrated to 0.8 mg/ml for refolding using the buffers noted in the table.

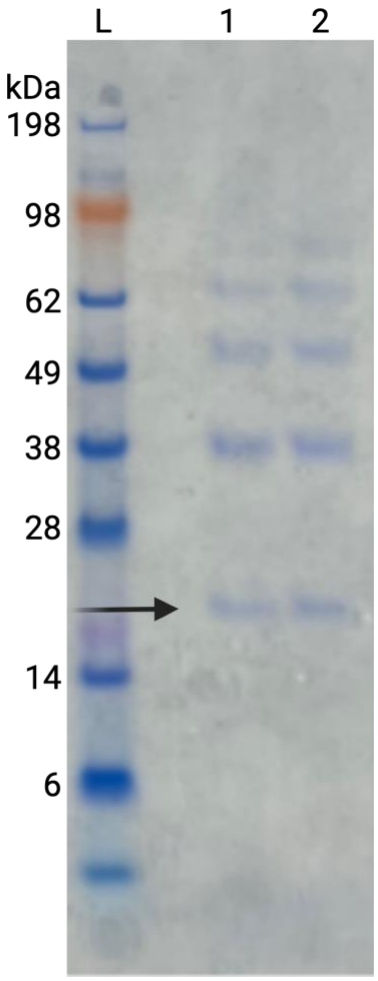


Figure 2-S1. A representative gel of P1C10 following refolding experiments. Non-reduced SDS-PAGE gels were used to assess the purity of monomer recovered from the refolding experiments. The size of the desired monomeric material is denoted with the arrow and incorrectly folded multimers can be seen at higher molecular weights. The samples in this gel are from Condition 9 in Table S2 (n = 2).

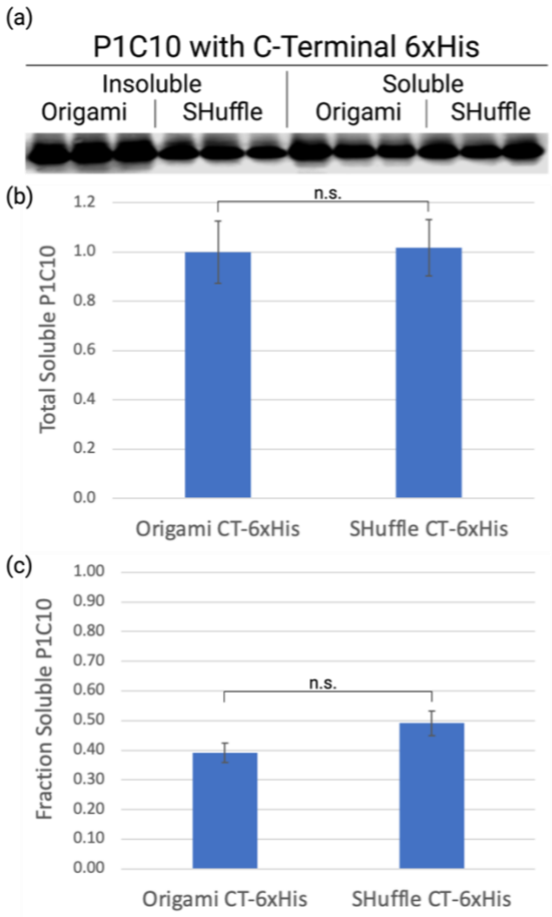


Figure 2-S2. Origami and SHuffle soluble Variable Lymphocyte Receptor western blot comparison for P1C10 CT-6xHis. (a) Lysates from Origami and SHuffle cultures were separated into insoluble and soluble fractions via centrifugation. Sample were run on reduced 4-12% Bis-Tris gels, transferred to a nitrocellulose membrane, and probed with anti-6xHis. (b) Densitometric quantification of the western blotting of soluble VLR with data was normalized to the lowest quantity for each data set (c) Densitometric quantification of the fraction soluble was assessed by dividing the total soluble material by the sum of the insoluble and soluble material. Data in all panels represent the mean and standard error of 3 independent bacterial transformants (n=3). Data are expressed as mean ± S.D. and statistical significance was determined by a Student's t-test.

CHAPTER 3 – The High-Resolution Structure of a Variable Lymphocyte Receptor from *Petromyzon marinus* Capable of Binding to the Brain Extracellular Matrix

This chapter is adapted from our submitted manuscript “The High-Resolution Structure of a Variable Lymphocyte Receptor from *Petromyzon marinus* Capable of Binding to the Brain Extracellular Matrix.” This chapter was completed with the support and guidance of Prof. Hazel Holden. Thank you so much for mentoring and educating me on the complexities of crystal structures. Dr. Jim Thoden provided additional laboratory expertise for producing the protein, setting up the crystal screens and diffracting the crystals.

3.1 Abstract

Variable lymphocyte receptors (VLRs) are antigen receptors derived from the adaptive immune system of jawless vertebrates such as lamprey (*Petromyzon marinus*). First discovered in 2004, VLRs have been the subject of numerous biochemical and structural investigations. Due to their unique antigen binding properties, VLRs have been leveraged as possible drug delivery agents. One such VLR, previously identified and referred to as P1C10, was shown to bind to the brain extracellular matrix. Here, we present the high-resolution X-ray crystal structure of this VLR determined to 1.3 Å resolution. The fold is dominated by a six-stranded mixed β-sheet which provides a concave surface for possible antigen binding. Electron density corresponding to a HEPPS molecule was found in this region. By comparing the P1C10 molecular architecture and its buffer binding residues with those of other VLRs previously reported, it was possible to illustrate how this unique class of proteins can accommodate diverse binding partners. Additionally, we provide an analysis of the experimentally determined structure compared to the models generated by the commonly used AlphaFold and iTASSER software packages.

3.2 Introduction

Variable lymphocyte receptors (VLRs) are a type of antigen receptor derived from the adaptive immune system of jawless vertebrates such as lamprey (*Petromyzon marinus*).[80] These vertebrates diverged from mammals more than 500 million years ago, and as a consequence, it is possible for lamprey VLRs to target mammalian antigens that traditional immunoglobulins (Igs) fail to recognize as antigenic due to self-tolerance.[80, 81] It has been demonstrated that lamprey immunized against mammalian antigens generate a robust adaptive immune response, producing a full repertoire of VLRs with unique specificity for brain vasculature,[82, 83] mammalian sugars,[84, 85] cell markers,[86] and many other conserved mammalian proteins.[80-83, 87, 88]

The diverse specificity of VLRs is derived from their unique genetic assembly and structure. Like Igs, VLRs are expressed in the lymphocytes of their host.[88] Five different types of VLRs have been identified: VLRA, VLRB, VLRC, VLRD, and VLRE.[89] VLRAs, VLRCs, VLRDs, and VLREs are expressed in T cell-like lymphocytes of jawless vertebrates whereas VLRBs are expressed in lymphocytes that are functionally similar to B-cells.[88] VLRBs are the most studied as these are expressed in a soluble form by the activated lymphocytes of immunized lamprey.[89] Structurally, VLRBs are known to adopt solenoid-type architectures that are dominated by leucine-rich repeats (LRRs), and they range in size from 15,000 to 35,000 Da.[81] LRR proteins contain signature sequence motifs typically composed of the following: xLxxLxxLxLxxNQLxxVFD, where “x” represents highly variable surface residues that contribute to antigen specificity and “L” represents buried leucines (or isoleucines) that form the hydrophobic core of the VLRB.[90, 91] The parallel b-strands of the LRRs wrap around to form a concave β-sheet that performs a

similar functional role to the complementarity-determining regions (CDRs) of Igs that are involved in antigen binding.[92] Indeed, it has been shown that the antigen specificity of the lamprey VLRBs arise from the variable surface residues in the LRRs of the β-sheet.[81]

The LRRs assemble according to a combinatorial process which further augments the possible antigen-targeting diversity of the VLRBs.[93, 94] The LRRs exist as genetic cassettes within the lamprey genome and can assemble in as many as 10^{17} different combinations of unique VLRBs.[95] The VLRB assembly is initiated with an N-terminus capping domain, referred to as the LRRNT. After the LRRNT, there is the LRR1 domain which is followed by as many as nine variable LRR domains referred to as LRRVs. The last LRR domain, referred to as LRRVe, is followed by a connecting peptide (CP), and a C-terminus capping domain (LRRCT). Like the LRRs of the β-sheet, the LRRNT and LRRCT both contain characteristic signature motifs, specifically Cx_nCxCx_nC and $CxCx_nCx_nC$, respectively. In both of these sequences, the cysteine residues form disulfide bonds contributing a total of four linkages in each VLR.[87] In their native forms, following the LRRCT, the VLRBs contain an invariant stalk region rich in threonines and prolines,[91] a glycosylphosphatidylinositol (GPI) anchor,[91] and a cysteine-rich hydrophobic tail.[88, 91] The stalk, GPI anchor, and hydrophobic tail allow the VLRBs to exist as multimers of dimers in the lamprey immune system.[81, 83, 96] When investigated as recombinant proteins for biotechnological applications, the stalk, the GPI anchor, and the tail are removed from the expression gene to eliminate undesired multimerization and to produce high affinity monomers.[83, 97]

Not surprisingly, due to their unique antigen binding properties, VLRBs have been investigated for their potential use as drug delivery vehicles in biomedical applications.[82-87, 90, 96, 98-100]

In previous investigations, the Shusta laboratory identified a VLRB capable of binding to the brain extracellular matrix (ECM).[83] Indeed, the brain ECM is a target of interest in diseases of the central nervous system. In instances of both acute insult such as trauma and stroke as well as chronic pathologies such as multiple sclerosis and tumor, the brain ECM is exposed at the site of the pathology [101-104]. Thus, being able to target the brain ECM, a ubiquitous source of antigens present in many disease states, allows for a potential general drug delivery system that can be adapted to include many different therapeutics. This VLRB, hereafter referred to as P1C10, was identified from a library of VLRs derived from a lamprey immunized with murine brain microvessel plasma and enriched for ECM binding.[83] Utilizing murine brain tumor models, subsequent *in vivo* studies demonstrated that systemically administered P1C10 showed selective accumulation in the brain at the site of the tumor.[83] When P1C10 was conjugated to doxorubicin-loaded liposomes and administered to tumor-bearing mice, there was a significant increase in survival of those dosed with the P1C10 therapeutic conjugate compared to the no treatment cohort.[83]

Curious to the fine structural details of P1C10, we initiated an X-ray crystallographic analysis of this VLR. Herein, we describe the molecular structure of P1C10 determined at 1.3 Å resolution with a bound buffer molecule and compare its molecular architecture and binding residues to those of previously reported VLRs. We also compare the X-ray crystallographic structure of P1C10 to models generated with iTASSER and AlphaFold to show the utility of experimentally determined structures. On the basis of this analysis, it is possible to speculate on the identity of the ECM-binding residues of P1C10 and to demonstrate its sequence variability that may be responsible for its unique binding properties relative to other VLRs.

3.3 Materials and Methods

3.3.1 Protein expression and purification

Two vectors for the expression of P1C10 from *P. marinus* were purchased from Genscript. The first was a pET28a(+) construct and contained an N-terminal poly-histidine tag and an rTEV cleavage site (MHHHHHHSSGENLYFQGHMS). The second was a pET21b(+) construct that produced a protein with a non-cleavable C-terminal poly-histidine tag (LEHHHHHH).

The expression plasmids were used to transform SHuffle® T7 Express competent cells (New England Biolabs), an *Escherichia coli* cell line optimized for disulfide bond formation.[105] Transformations were carried out according to the manufacturer's instructions. Following overnight growth on selective plates, the colonies were suspended in media and used to inoculate terrific broth (TB) for growth and protein expression. The TB was supplemented with either kanamycin (50 mg/L concentration) for the pET28a(+) vector or ampicillin (100 mg/L) for the pET21(b+) vector. The cultures were grown at 37ºC with shaking until an optical density of 0.4 was reached at 600 nm. The flasks were subsequently moved to room temperature, and the cells were allowed to grow for an additional two days.

All cultures were harvested by centrifugation and flash frozen in liquid nitrogen for storage at -80ºC. The pellets were subsequently resuspended in lysis buffer (50 mM sodium phosphate, 20 mM imidazole, 10% glycerol, and 300 mM NaCl (pH 8.0)) and sonicated on ice. Soluble lysate was separated by centrifugation, and the proteins were purified at 4^{o}C with Ni-NTA resin (Thermo Scientific #88223) according to the manufacturer's instructions. All buffers were titrated to pH 8.0 and contained 50 mM sodium phosphate, 300 mM NaCl, and imidazole concentrations of 20 mM for the wash buffer and 300 mM for the elution buffer. The purified P1C10 protein obtained from the pET28a(+) construct was split into two fractions, one to maintain the N-terminal tag and the other to remove the N-terminal tag via rTEV protease. The rTEV protease and remaining tagged proteins were separated by passage over Ni-NTA resin and the unbound fraction was collected as the tag-free P1C10 protein.

Following the purification procedures, all samples were dialyzed against 10 mM Tris-HCl (pH 8.0) and 200 mM NaCl before concentration. The proteins were concentrated to approximated 18 mg/mL based on an extinction coefficient of 0.69 $(mg/mL)^{-1}cm^{-1}$.

3.3.2 Crystallization

Crystallization trials for all proteins were initially conducted at both 20°C and 4°C using the hanging drop method of vapor diffusion with an in-house crystallization screen composed of 144 conditions.

Crystals of the tag-free P1C10 were grown via the hanging drop vapor diffusion method at 20°C from 2.5 – 2.8 M ammonium sulfate, 2% hexylene glycol, and 100 mM HEPPS (pH 8.0). They belonged to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 45.5 Å, b = 53.1 Å, and c = 60.3 Å. The asymmetric unit contained one subunit. For X-ray data collection, the crystals were transferred to a cryo-protectant solution composed of 3.4 M ammonium sulfate, 200 mM NaCl, 2% hexylene glycol, 20% sucrose, and 100 mM HEPPS (pH 8.0) and flash frozen in liquid nitrogen before X-ray data collection.

3.3.3 X-ray data collection and processing

X-ray data were collected using at 100K BRUKER D8-VENTURE sealed tube system equipped with HELIOS optics and a PHOTON II detector. The X-ray data set was processed with SAINT and scaled with SADABS (Bruker AXS). Relevant X-ray data collection statistics are provided in Table 1.

3.3.4 Structure solution and model refinement

The structure of P1C10 was solved via molecular replacement with Phaser[106] using PDB entry 3G39[91] as the search probe. Iterative cycles of model-building with COOT[107, 108] and refinement with REFMAC[109] led to a final X-ray model with an overall R-factor of 18.6%. Relevant refinement statistics are listed in Table 1.

3.4 Results and Discussion

3.4.1 X-ray crystallographic structure of P1C10 determined at 1.3 Å resolution

Two recombinant P1C10 constructs were expressed in SHuffle *E. coli* cells for use in the crystallization screens. One P1C10 construct contained an N-terminal poly-histidine tag with a cleavable rTEV site. The other construct contained a C-terminal poly-histidine tag. The N-terminal poly-histidine P1C10 protein, the cleaved and tagless P1C10 protein (rTEV-cleaved), and the C-terminal poly-histidine protein were screened for crystal formation side-by-side on the same coverslips. Interestingly, only the tagless rTEV-cleaved protein produced crystals. These crystals were grown from ammonium sulfate (pH 8.0) (see Materials and Methods for details) and belonged to the space group $P2_12_12_1$ with unit cell dimensions of a = 45.5 Å, b = 53.1 Å, and c = 60.3 Å and an asymmetric unit containing a single molecule. The model was refined at 1.3 Å resolution to an overall R-factor of 18.6% (Table 1). The electron density for the polypeptide chain was continuous from Ala 1 to Pro 168, albeit was weaker with higher temperature factors from Gly 159 to the C-terminus.

A ribbon representation of P1C10 is presented in stereo in Figure 1a. The fold is dominated by a six-stranded mixed β-sheet, which provides a concave surface for antigen binding, in this case the buffer molecule. The first two β-strands run antiparallel and are connected by a classical Type II' turn delineated by Asp 9 to Thr 12. There are other classical reverse turns that define the structural architecture of P1C10 including: Glu 45 to Val 48 (Type II), Leu 52 to Leu 55 (Type I), Pro 69 to Val 72 (Type II), Leu 76 to Leu 79 (Type I), Pro 93 to Ala 96 (Type II) and Leu 100 to Leu 103 (Type I). The four disulfide bridges formed between Cys 2 and Cys 8, Cys 6 and Cys 15, Cys 115 and Cys 147, and Cys 117 and Cys 167 serve to stabilize the fold.

As described in the Introduction, the VLRs are divided into distinct regions, referred to as the LRRNT, the LRR1, up to nine LRRVs, the LRRVe, the CP, and the LRRCT. These regions are colored coded in Figure 1a. In P1C10, the LRRNT domain is formed by Ala 1 to Thr 31. The LRR1 domain is defined by Lys 32 to Asp 50,

which is subsequently followed by a single LRRV motif delineated by His 51 to Pro 69. The LRRVe, CP, and LRRCT domains are formed by Lys 75 to Pro 93, Asn 99 to Ile 106, and Pro 112 to the C-terminus, respectively. An amino sequence alignment of these repeats is provided in Figure 1b.

There was clear electron density for the buffer, HEPPS, that was used in the crystallization trials as shown in Figure 2a. A close-up view of the binding pocket for the buffer is displayed in Figure 2b. The sulfonic acid moiety of the buffer is wedged into place via hydrogen-bonding interactions with the side chains of Ser 61 and Trp 62 in the LRRV domain, and Ser 83 in the LRRVe domain. Both the side chains of Ser 61 and Ser 83 adopt two distinct conformations. There is also a water molecule that lies within 3.0 Å from a nitrogen atom in the piperazine ring. Additionally, there is a water molecule that bridges O8 of the buffer to a sulfate anion which is anchored into place via the side chains of Arg 35 in the LRR1 domain and Lys 57 and Tyr 59 found in the LRRV domain. Given the formation of a clear binding pocket for the buffer molecule, it is possible that the target ECM antigen, which is currently unknown, could also interact in this region. Indeed, the observation of bound buffer molecules to proteins is common in many X-ray crystallographic structures determined to high resolution, and the positions of these small molecules are often located in physiologically relevant positions.[110][-35]

3.4.2 X-ray crystallographic model of P1C10 shows errors in predicted models from AlphaFold and iTASSER

As has become common practice, molecular modeling was utilized in an effort to gain insight into the three-dimensional structure of P1C10 prior to its successful X-ray crystallographic analysis. Both AlphaFold[111] and iTASSER[112] were used for generating homology models of P1C10. AlphaFold, iTASSER, and the Dali Server[113] all identified Variable Lymphocyte Receptor 4 (VLR4) (PDB ID: 3TWI)[114] as having the highest structural and sequence homology to P1C10 (81% amino acid sequence identity). Thus, VLR4 was used as the threading template for the P1C10 models. The top-scored AlphaFold model aligned to the P1C10 crystal structure with a root-mean-square deviation of 0.8 Å for 1254 atoms

(according to the LSQ algorithm in COOT[107, 108]). The residues identified in the buffer binding pocket, speculated to be involved in antigen binding, were among the first set we compared between the X-ray crystallographic and the AlphaFold models. The side chains of Lys 57, Tyr 59, Ser 61, Trp 62, and Ser 83 had very similar rotamer positions between the AlphaFold model and the crystal structure (Figure 3a). The side chain of Arg 35 in the AlphaFold model, however, was rotated away such that $N^{\eta 1}$ was situated beyond ~5 Å from the sulfate oxygen atoms (Figure 3a).

In addition to the buffer binding site of P1C10, we also compared the LRRCTs between the experimentally determined and the AlphaFold models. The LRRCTs in the VLRs are known to be highly variable and are often implicated in binding.[87] In the P1C10 crystal structure, the side chain $N^{\varepsilon 2}$ of Gln 137, which is part of the LRRCT, lies within 3.4 Å of O8 of the buffer (Figure 3b). In the AlphaFold model, the carboxamide group of Gln 137 adopts a different orientation such that $N^{\varepsilon 2}$ now lies within 4.7 Å of O8. Additionally, the positions of the α-carbons for Gln 137 in the crystal structure and the computer model differed by 0.9 Å. There are additional places where the crystal structure and AlphaFold model deviate including the N-terminus, the side chains of Gln 10, Lys 14, Arg 18, Lys 43, Glu 45, Ile 60, Gln 64, Val 92, Arg 94, Leu 109, Glu 116, Lys 124, Gln 129, Lys 146, and Arg 155 (Figure 3c). Whereas these residues are not implicated in buffer binding, some (Gln 10, Lys 14, Leu 109) are surface residues in LRR regions of the VLR, so they are of potential importance for binding putative antigens (Figure 3c). Finally, the backbone atoms between Thr 150 to the C-terminus do not align as well as might be expected between the computer generated and the experimentally determined models.

In contrast to the AlphaFold model, the iTASSER model showed significant, and in some cases serious, deviations from the actual P1C10 structure. Indeed, there were 17 residues whose ϕ/ψ angles were outside of the allowed regions of the Ramachandran plot in the iTASSER model compared to the AlphaFold structure were there were none. Some of the more serious ϕ/ψ angles are present in b-strand five where many of the buffer binding residues are located including those for Ser 83 (ϕ = -71°, ψ = -118°)

and Leu 84 (ϕ = 124°, ψ = -15°). As a consequence of these outliers, the carbonyl oxygens of Leu 82 and Ser 83 are separated by only 2.7 Å and the backbone amide nitrogens of Leu 84 and Gly 85 lie within 2.7 Å of one another. Not surprisingly, the experimentally determined structure of P1C10 and the iTASSER-generated model align with a high root-mean-square deviation of 1.8 Å for 1254 atoms (with a maximum deviation of ~10 Å). Importantly, the iTASSER model did not predict the correct geometry for any of the disulfide bridges whereas the AlphaFold model did. There were at least 43 side chains whose conformations were not modeled correctly. The puckers for the side chains of Pro 28, Pro 46, Pro 93, and Pro 153 were incorrect. The Type II' turn formed by Asp 9 to Thr 12 was predicted incorrectly (Figure 4a). Perhaps the most serious error was in the region defined by Val 134 to Gly 141 as can be seen in Figure 4b. This is part of the LRRCT region that contains Gln 137 which, in the actual P1C10 model, lies within 3.4 Å of the bound buffer molecule. The ϕ/ψ angles for His 139 in the model, 135° and 151°, respectively, are particularly problematic as they are outside of the allowed regions of the Ramachandran plot. As noted above, the side chain of Trp 62 anchors the buffer molecule to the protein. In the iTASSER model, the ϕ/ψ angles for Trp 62 were in the disallowed regions of the Ramachandran plot (ϕ = 46°, ψ = 72°) and the torsional angle defined by N-Cα-Cβ-Cγ was -158° rather than -62° as observed in the experimentally determined model of P1C10. Other significant errors in side chain conformations include those for His 81 and Tyr 122 (Figure 4c). These concerns yet again emphasize the continued need for experimentally determined high-resolution protein structures.

3.4.3 P1C10 has structural homologs with diverse antigen targets as a result of the LRR residues

VLR4, which was used as the template for both the AlphaFold and iTASSER models of P1C10, is known to bind *Bacillus anthracis* spores and specifically to the immunodominant glycoprotein, BclA.[114] The $\alpha-$carbons for the P1C10 and VLR4 (subunit D) structures align with a root-mean-square deviation of 0.6 Å for 168 atoms. As shown in Figure 5, those residues that line the *B. anthracis* spore binding pocket in VLR4 include Asn 79, His 105, Ile 106, Gly 158, Asn 159, and Asn 164. In P1C10, these residues correspond to

Tyr 59, Gly 85, Tyr 86, Gly 138, His 139, and Asn 144 which abut the observed buffer molecule. Despite 81% sequence homology between VLR4 and P1C10, these proteins have dramatically different binding partners largely because of the sequence variability in their concave surfaces as can be seen in Figure 5.

A search of the Protein Data Bank using the Dali Server[113] also revealed a close structural correspondence of P1C10 to a VLR-like protein, referred to as Bf66946, from *Branchiostoma floridae*, a Florida lancelet.[115] A superposition of the α-carbons for Bf66946 and P1C10 is shown in Figure 6a. Binding assays and site-directed mutagenesis studies suggest that Bf66946 interacts with the surfaces of the Gram-positive bacteria *Staphylococcus aureus* and *Streptococcus pneumonia* via the side chains of Asp 105 and Asp 107.[115] These two aspartates correspond to His 81 and Ser 83, respectively, in the buffer binding pocket of P1C10. In Bf66946 there is a five-residue insertion lying between Val 29 and Met 35 (Figure 6a). Other than this region, the two proteins are remarkably similar up to the beginning of the CP domain where the structures diverge slightly. In P1C10 there is a loop delineated by Leu 100 to Leu 103 which lies in the region connecting the fifth and sixth β-strands. In Bf66946 there is a seven-residue insertion in this region. The polypeptide chains diverge again at Ser 148 (P1C10 numbering) where there is a five-residue insertion in Bf66946 relative to P1C10 (Figure 6b). We speculate that these subtle structural deviations between P1C10 and Bf66946 contribute to their vastly different binding specificities.

3.5 Conclusion

The structure of P1C10 described here demonstrates that it adopts the classical solenoid shape of the VLRBs. We were able to identify the similarities and differences of the crystal structure with two computational models of P1C10 generated using AlphaFold and iTASSER. The observed differences clearly emphasize the importance of determining structures experimentally in order to gain detailed molecular insight. In addition, based on the position of a bound buffer molecule and comparisons with homologous structures of other VLRs/binding partner complexes, we were able to further demonstrate that the variable residues in VLRs can accommodate diverse antigen binding profiles.

3.6 Tables

Table 3-1 X-ray data collection and model refinement statistics.

PDB code	9CJ0
Resolution limits (Å)	40.0-1.30 (1.40 – 1.30)[b]
Number of independent reflections	36207 (6984)
Completeness (%)	98.0 (94.0)
Redundancy	8.1 (5.0)
avg I/avg σ(I)	20.1 (3.3)
R_{sym} (%)[a]	5.1 (42.2)
[c]*R*-factor (overall)%/no. reflections	18.6/36207
R-factor (working)%/no. reflections	18.4/34487
R-factor (free)%/no. reflections	22.2/1720
number of protein atoms	1362
number of heteroatoms	205
average B values	
protein atoms (Å^2)	16.5
solvent (Å^2)	28.0
weighted RMS deviations from ideality	
bond lengths (Å)	0.008
bond angles (º)	1.64
planar groups (Å)	0.006
Ramachandran regions (%)[d]	
most favored	96.4
additionally allowed	3.6
generously allowed	0.0

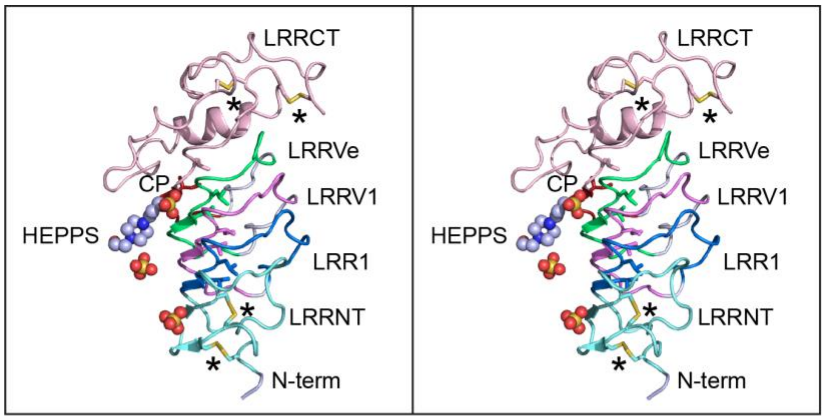
[a]$R_{sym} = (\sum|I - \overline{I}|/ \sum I) \times 100$.

[b]Statistics for the highest resolution bin.

[c]*R*-factor = $(\Sigma|F_o - F_c| / \Sigma|F_o|) \times 100$ where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.

[d]Distribution of Ramachandran angles according to PROCHECK.[116]

3.7 Figures

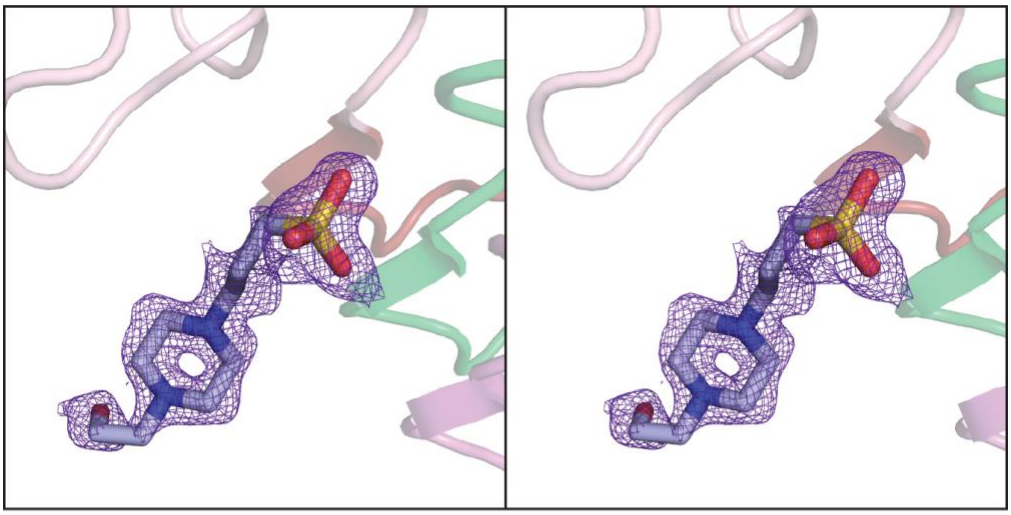


(a)

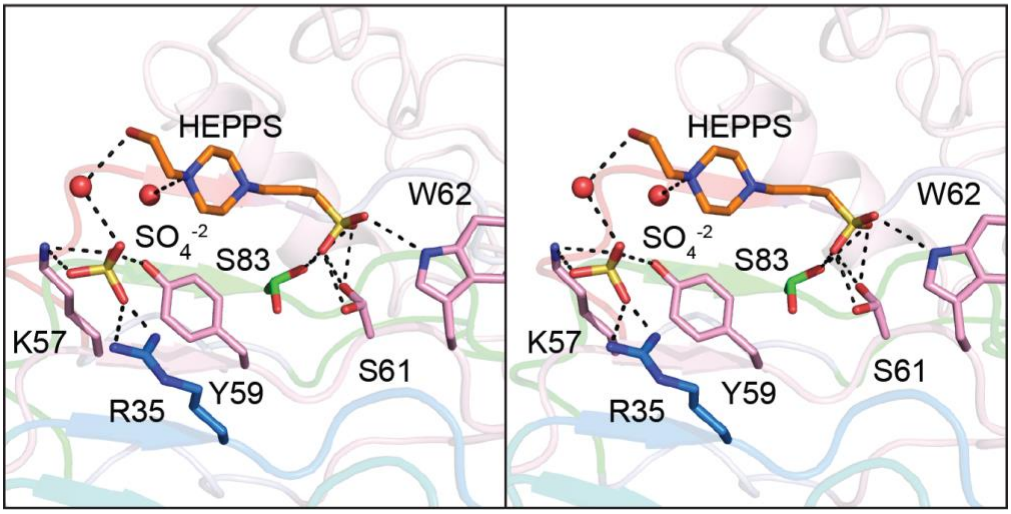
```
xLxxLxxLxLxxNQLxxVF          consensus sequence
    * *
KILRLYSNQITKLEPGVFD          LRR1
       * * **
HLVNLEKLYISWNQLSALP          LRRV
       * *  *
KLTKLTHLSLGYNQLKSVP          LRRVe

NLKSLTHIWLLNLKSLTHI          CP
```

(b)

Figure 3-1 Structure of P1C10. A ribbon representation of its molecular architecture is presented in stereo in (a). The bound buffer, HEPPS, and the two ammonium sulfate ions in the model are displayed in sphere representation. The LRRNT, LRR1, LRRV, LRRVe, CP, and LRRCT domains are highlighted in teal, marine, pink, green, red, and light pink, respectively. The positions of the disulfide bridges are marked by the black asterisks. An amino acid sequence alignment of the LRR1, LRRV, LRRVe, and CP domains is provided in (b). Those amino acids marked by blue asterisks in the alignment provide side chains that line the region responsible for the binding the buffer. Structural representations in this Figure, and all other figures were prepared with PyMOL.[117]

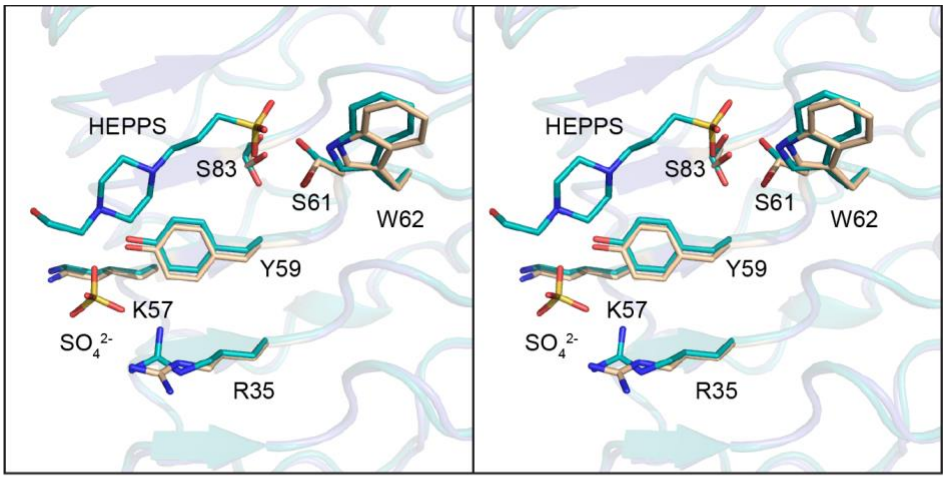


(a)

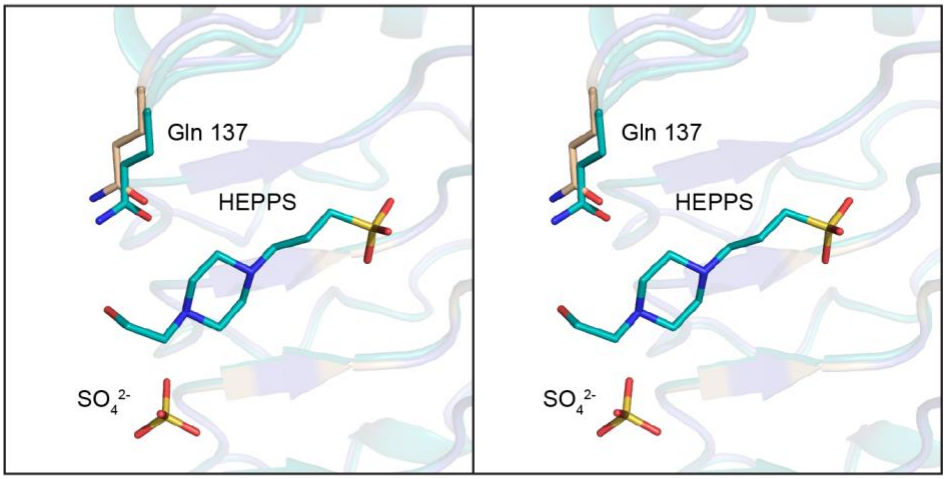


(b)

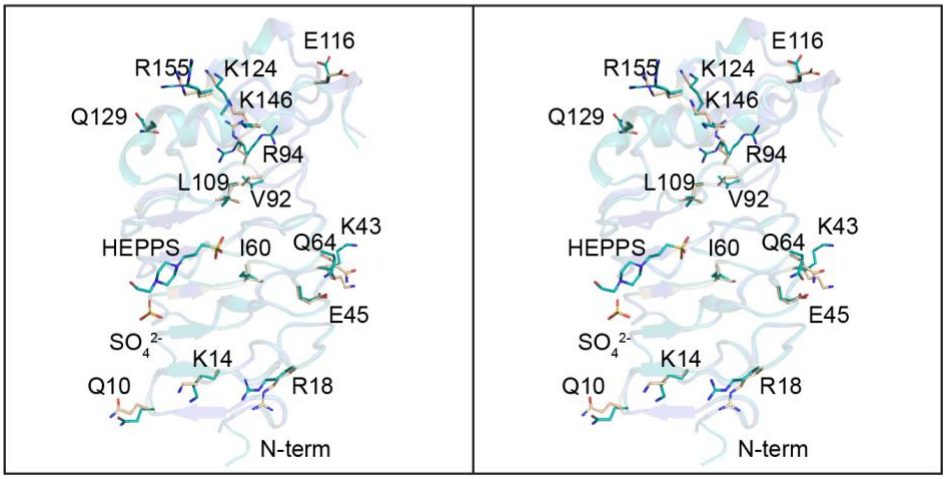
Figure 3-2 The HEPPS binding site. Electron density corresponding to the buffer molecule is displayed in stereo in (a). The electron density map was calculated with (F_o-F_c) coefficients and contoured at 3σ. The HEPPS ligand was not included in the X-ray coordinate file used to calculate the omit map, and thus there is no model bias. A closeup stereo view of the region surrounding the buffer molecule is presented in (b). Ordered water molecules are displayed as red spheres. The dashed lines indicate possible hydrogen-bonding interactions within 3.2 Å.


HEPPS
S83
S61
W62
Y59
K57
SO_4^{2-}
R35
HEPPS
S83
S61
W62
Y59
K57
SO_4^{2-}
R35

(a)

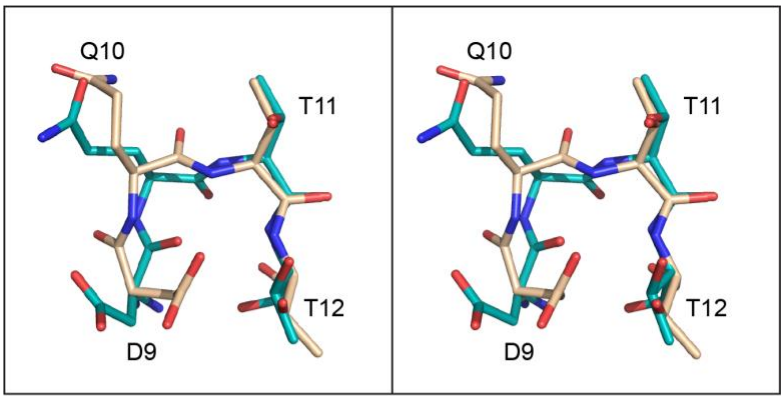

Gln 137
HEPPS
SO_4^{2-}
Gln 137
HEPPS
SO_4^{2-}

(b)

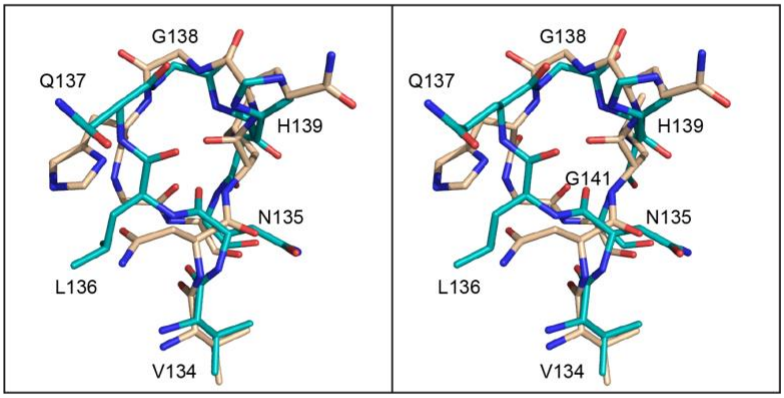

E116
R155
K124
K146
Q129
R94
L109
V92
K43
HEPPS
I60
Q64
E45
SO_4^{2-}
K14
Q10
R18
N-term
E116
R155
K124
K146
Q129
R94
L109
V92
K43
HEPPS
I60
Q64
E45
SO_4^{2-}
K14
Q10
R18
N-term

(c)

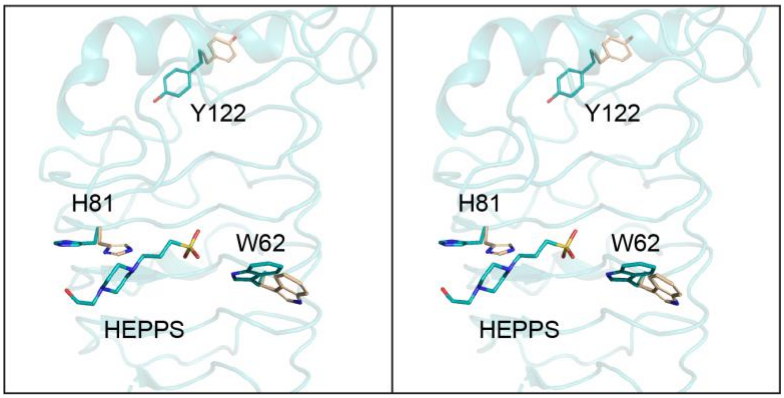
Figure 3-3 Comparison of the experimentally determined P1C10 structure and the AlphaFold generated model. Shown in stereo in (a) are the positions of the buffer binding residues in P1C10 versus those predicted in the AlphaFold model. The side chain of Gln 137, which is part of the LRRCT, lies within 3.4 Å of O8 of the buffer in the P1C10 structure. As can be seen by the overlay in (b), the AlphaFold model positions the carboxamide group of Gln 137 farther away. Finally, displayed in stereo in (c) are the positions of the other side chains that display significant variations between the two models. In all panels, the P1C10 structure is highlighted in teal whereas the predicted model is shown in wheat.



(a)



(b)



(c)

Figure 3-4 Comparison of the experimentally determined P1C10 structure and the iTASSER generated model. Shown in stereo in (a), (b), and (c) are regions where the experimentally derived and the computer-generated models show significant, and in some cases, serious deviations. As in Figure 3, the crystal structure of P1C10 is highlighted in teal whereas the computer-generated model is displayed in wheat.

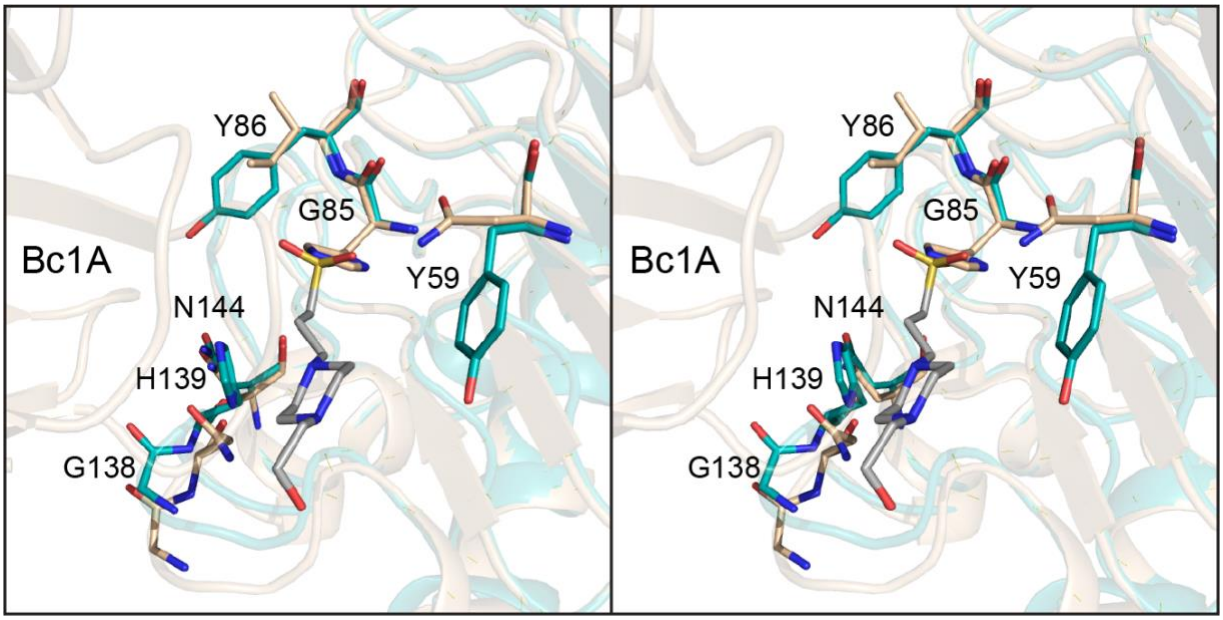
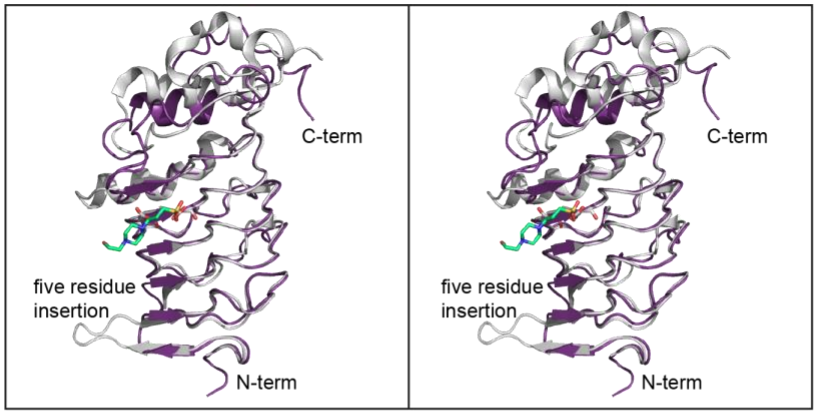


Figure 3-5 Comparison of the P1C10 structure to the VLR4/BclA complex from *B. anthracis*. Shown in stereo is a closeup superposition of the binding face between P1C10 (highlighted in teal) and the buffer molecule onto the VLR4/BclA complex (displayed in wheat). The residue labels correspond to those in P1C10.



(a)

```
                    LRRNT                                LRR1          LRRV1
P1C10   1   ACPSQCSCD-----QTTVKCHSRRLTSVPAGIPTTTKILRLYSNQITKLEPGVFDHLVNLEKLYISWNQLSALPVGVFDK  75
4XSQ    1   MCPSACKCTVSLYGEMVVACGGMGLTEIPEDIPHRAVYLVLKDNNITKITSYSFKGLRNLQGIDLSNNKINHISSAALRH  80
3TWI    1   ACPSQCSCS-----GTTVNCQERSLASVPAGIPTTTQVLHLYINQITKLEPGVFDSLTQLTYLNLAVNQLTALPVGVFDK  75

                    LRRVe
P1C10   76  LTKLTHLSLGYNQLKSVPRGAFDNLKSLT-------HIWLLNNPWDCECSDILYLKNWIVQH-ASIVNLQGHGGVDNVKC  147
4XSQ    81  LGHLDDIDLSRNELTSVSEKLFDFPISSAKAQGRRFVYLANNPWGCDCR-----MAWLAQElAGGSKTFGDR---HMEC  152
3TWI    76  LTKLTHLALHINQLKSIPMGVFDNLKSLT-------HIYLFNNPWDCECSDILYLKNWIVQH-ASIVNPLGNGGVDNVKC  147


P1C10   148 SGTNTPV----RAVTEASTSPSKCP---------  168
4XSQ    153 A---TPAALAGRGLSEIPQTSFVCTGRDISFDSD  183
3TWI    148 SGTNTPV----RAVTEASTSPSKCP---------  168
```

(b)

Figure 3-6 Comparison of P1C10 to Bf66946 from *B. floridae*. Superposition of the P1C10 a-carbons onto those of Bf66946 are shown in stereo in (a). The ribbon representation for P1C10 is highlighted in dark purple. An amino acid sequence alignment for P1C10, Bf66946, and VLR4 is provided in (b) where 4XSQ and 3TWI correspond to the PDB coordinates for Bf66946 and VLR4, respectively. The alignment was prepared with the software COBALT.[43] The boundaries for the LRRNT, LRR1, LRRV, and LRRVe domains in the three proteins are indicated by the pink brackets.

CHAPTER 4 – Evaluation of the Putative Binding Partners of Variable Lymphocyte Receptor P1C10

In this chapter the putative binding targets of variable lymphocyte receptor P1C10 are investigated. This chapter was completed with initial data from Dr. Benjamin Umlauf who performed the LC-MS/MS pulldown of putative binding targets for P1C10, identified the putative binding targets from the MS peptide results, and performed initial binding studies with the putative binding targets. Dr. Jim Thoden assisted in the production of all in-house carboxypeptidase expression tests. Dr. Ivan Rayment advised on the construct design of all in-house carboxypeptidases. Dr. Peter Lyons provided plasmid and advised on the mammalian expression of CPA6.

4.1 Abstract

The Shusta lab previously identified variable lymphocyte receptor P1C10 from a library of variable lymphocyte receptors generated from a lamprey immunized with mouse brain endothelial cell plasma membrane. The library was enriched for brain extracellular matrix binders, and P1C10 was identified as a candidate variable lymphocyte receptor capable of drug delivery to the brain in instances of pathologically-disrupted blood-brain barrier. Because of the exciting drug delivery potential of P1C10, an investigation was pursued to find the binding partner at the brain extracellular matrix. Previously, the Shusta lab designed a mass spectrometry pulldown method to identify putative binding targets of P1C10. The Shusta lab identified carboxypeptidases as a class of enzymes that could act as potential binding partners to P1C10 at the extracellular matrix of the brain. In the present study, the putative binding of P1C10 and carboxypeptidases is investigated. A comprehensive study of all commercially available carboxypeptidases is presented with binding data from ELISAs, ITC, AlphaFold Multimer, and crystal screens. Analysis of carboxypeptidases is performed using SEC and NativePAGE. Results suggest that carboxypeptidases are not the putative binding targets of P1C10.

4.2 Introduction

Determining the binding target of therapeutic proteins is a complex component of pharmaceutical development. It is often debated when to pursue target identification (TID) in the discovery timeline of a new therapeutic, and if TID is necessary for the efficacy of the drug. Interestingly, the Food & Drug Administration (FDA) does not require TID nor Mechanism of Action (MOA) for drug approvals [119, 120]. Furthermore, 10-20% of FDA-approved drugs have no known target or MOA [121]. Therefore, it is necessary to determine the utility to the program when considering a TID/MOA study. Deciding whether to pursue TID/MOA is most important for therapeutics identified from phenotypic screens [119]. In instances of phenotypic screens, a general indication *in vivo* is used to screen a library without knowing the specific binding target of the positive binders in the screen. Thus, questions arise after a positive phenotypic binder is identified as to what the binding interactions are *in vivo*.

There are several arguments both for and against TID/MOA studies for pharmaceuticals identified from phenotypic screens. Arguments against TID/MOA studies cite therapies such as aspirin and lithium which have been commercially available for decades yet do not have a known MOA nor target, and still these therapies have minimal adverse effects in patients [122, 123]. Furthermore, TID/MOA studies can divert limited funds away from primary research goals of developing a new medicine to move into the clinic [122]. However, some instances of pharmaceutical development benefit from TID/MOA studies. Larger pharmaceutical companies pursuing expensive late-stage phase III trials often pursue TID/MOA studies to minimize the risks associated with therapeutics of unknown TID/MOA [122]. TID/MOA studies enable investigators

to comprehensively understand molecular targets and potential off-target effects, mitigating the risk of costly adverse events in phase III trials [122]. In contrast, academic labs and smaller biotechnology startups move therapeutics into clinical trials without pursuing TID/MOA studies due to limited funding and lower risk associated with early-stage programs. TID/MOA studies also must be considered relative to the type of therapy being pursued [122]. In the case of novel therapeutics indicated for diseases without any existing disease-modifying therapies, TID/MOA may be less important than efficiently moving the therapy to the clinic, so patients have access to a potentially lifesaving or life-altering therapy. In other cases, when a therapeutic is targeting a disease with existing therapies, understanding the TID/MOA may determine optimal use-cases for each therapy. Thus, the decision to pursue TID/MOA studies is dependent on the available resources to pursue the TID/MOA and the overall cost-benefit to the pharmaceutical program.

TID of variable lymphocyte receptor (VLR) P1C10 was pursued for several reasons. P1C10 demonstrated exciting *in vivo* drug delivery properties in mouse glioblastoma (GB) models [48]. However, P1C10 was identified from a yeast surface display (YSD) library screen enriched for bEnd.3 ECM-binding VLRs [48]. Therefore, P1C10 was known to bind the ECM of mouse brain endothelial cells, but the specific antigen in the ECM was unknown. Furthermore, VLRs have not been dosed in humans, and there are no FDA-approved VLRs. Thus, the potential immunogenicity and off-target effects of lamprey-derived VLRs used as therapeutics in humans remains unknown. Therefore, understanding the binding partner of P1C10 was a prudent first step in understanding the potential efficacy of VLRs as a human drug delivery moiety.

Previous work by a former post-doctoral researcher in the Shusta Lab, Dr. Benjamin Umlauf, had been done to elucidate the binding target of P1C10. Liquid Chromatography with tandem Mass Spectrometry (LC-MS/MS) was performed according to previously published methods for TID [124]. In this method, P1C10 was conjugated to a capture moiety to bind and pulldown the antigen on the bEnd.3 ECM. Among many peptides, a peptide from Carboxypeptidase E (CPE) was identified from the pulldown. CPE is an enzyme with known tissue distribution in the extracellular space of the brain [125]. Physiologically, CPE was a relevant putative binding target of P1C10 to investigate. In the initial ELISAs of P1C10 with CPE performed in the Shusta lab, the K_D was found to be 181.6 nM. However, this was worse affinity than the previously published 48.4 nM P1C10 binding to bEnd.3 ECM [48]. Thus, the Shusta lab decided to explore other enzymes from the carboxypeptidase (CP) family as potential binding partners of P1C10. The other CPs tested by the Shusta lab included Carboxypeptidase A (CPA), Carboxypeptidase B (CPB), and Carboxypeptidase A6 (CPA6). While CPA, CPB, and CPA6 were not identified in the LC-MS/MS pulldown, they were investigated as potential binding partners to P1C10 since they had structural similarities to CPE, and they have distribution in the central nervous system [126, 127]. Of the CPs that were tested in early studies, Human CPA6 bound the tightest to P1C10. CPA6 was found to have a K_D = 23.6 nM. Therefore, CPA6 was pursued as the putative binding target of P1C10.

Putative binding targets must always be validated to confirm real bimolecular interactions. The continuous improvements of MS sensitivity produces extensive lists of putative binding partners [128]. In the case of P1C10, each LC-MS/MS experiment produced >60,000 unique peptides. Most of the peptides identified in MS pulldowns are nonspecific binders that are copurified with

specific interacting peptides [128]. Thus, any potential hits that are not eliminated by a negative control in the LC-MS/MS experiment must be validated with downstream assays.

In the present work, the apparent binding between CPA6 and P1C10 was investigated. To lower the cost of TID studies, CPA6 in-house production was pursued. CPA6 and other CPs previously identified as putative binding partners were analyzed by Size Exclusion Chromatography (SEC) and NativePAGE. Binding was assessed using crystal screens, ELISAs, Isothermal Calorimetry (ITC), and computational models in AlphaFold Multimer. The results from the subsequent studies indicated CPA6 and the other CPs investigated are unlikely to be the binding target of P1C10.

4.3 Materials and Methods

4.3.1 Carboxypeptidase A6 Production Attempts

CPA6 production was attempted in HEK293 cells according to previously published methods [126]. pcDNA3.1(-)hCPA6-HA-H6 plasmid was generously provided by Professor Peter Lyons. The plasmid was used to transfect Freestyle 293-F cells (Thermo Scientific #R79007). Cell media was recovered in the presence of 400 μg/ml heparin and purified according to Ni-NTA purification methods (Thermo Scientific #K95001).

CPA6 production was attempted in Rosetta B21(DE3) and SHuffle cells. pET plasmids were used to transform *E. coli.* Plasmids were transformed into Rosetta Competent Cells (Novagen #70954) and SHuffle Competent Cells (New England Biolabs #C3029J) by the heat shock method according to the manufacturers' protocols. Each sample was plated on LB agar with the appropriate antibody selection and incubated overnight according to the manufacturer's recommendations. Cultures of each CPA6 construct were prepared with appropriate antibiotics in both TB and LB media for initial expression tests.

4.3.2 Carboxypeptidase Computational Models

The sequences of P1C10 [48] with Carboxypeptidase B1 (UniProt #P15086) as well as P1C10 with Carboxypeptidase A6 (UniProt # Q8N4T0) were run on AlphaFold Multimer via COSMIC². Both sets of sequences were run on all 5 machines of AlphaFold, and the top predicted structure from each machine was evaluated in PyMOL. InterfaceResidues.py was run in PyMOL to identify

binding residues in P1C10 predicted by each model. The predicted binding residues were compared across all models to identify most-predicted binding residues in P1C10.

4.3.3 P1C10 Alanine Scan

The most-predicted binding residues in P1C10 from the computational models were mutated to Alanine residues to knockout binding and validate the results of the computational screen. The alanine mutants were purchased in pET vectors from Genscript. All alanine mutants were expressed in SHuffle cells (New England Biolabs #C3029J) according to the protocol described in Chapter 2. Alanine mutants were purified by Ni-NTA purification (Thermo Scientific #K95001) and validated by LC-MS analysis at the UW – Madison Biotechnology Center.

4.3.4 Size Exclusion Chromatography of Carboxypeptidases

SEC of all purchased carboxypeptidases was performed using an ÄKTA pure 25M. CP samples included Human CPA6 (MyBioSource.com #MBS1330838) and Rat CPB (MyBioSource.com #MBS143539). Both samples were suspended in 10 mM TRIS, 200 mM NaCl pH 8 at 5 mg/ml immediately before being loaded onto the ÄKTA. Samples were separated using a Superdex 200 Increase 10/300 GL column (Cytiva # 28990944) with 10 mM TRIS, 200 mM NaCl pH 8 mobile phase. All samples were injected at 0.2 ml with a flowrate of 0.75 ml/min. Spectra were analyzed in Unicorn 7.

4.3.5 NativePage of Carboxypeptidases

NativePAGE™ 4 to 16%, Bis-Tris, 1.0 mm, Mini Protein Gels (Thermo Scientific # BN1004BOX) were run according to the manufacturer recommendations. CP samples included Human CPA1 (Sino Biological # 10504-H08H), Human CPA2 (Sino Biological #10499-H08H), Human CPA6 (MyBioSource.com #MBS1330838), Human CPB1 (MyBioSource.com # MBS8120318), Human CPE (Sino Biological # 10069-H08H), Mouse CPA1 (Sino Biological # 50448-M08H), Rat CPB (MyBioSource.com #MBS143539). NativeMark™ Unstained Protein Standard (Thermo Scientific #LC0725) was used as a ladder on the gel.

4.3.6 Crystal Screen of Carboxypeptidase B

Rat CPB (MyBioSource.com #MBS143539) was screened for crystals in complex with P1C10. Crystallization trials were conducted at ambient temperature using the hanging drop method of vapor diffusion with an in-house crystallization screen composed of 144 conditions. CPB, P1C10, and CPB-P1C10 mixture were dropped for all 144 conditions. Crystals were allowed to grow for 1-2 weeks.

4.3.7 Isothermal Calorimetry of Carboxypeptidases

Human CPA6 (MyBioSource.com #MBS1330838) and Rat CPB (MyBioSource.com #MBS143539 were evaluated with P1C10 titration on ITC. A VP-ITC microcalorimeter (Microcalorimetry, Inc.) was used to generate binding data. P1C10 NT-6xHis construct was incubated with TEV protease for 48 hours at 4 °C to remove the 6xHis epitope tags. The TEV-cleaved construct was dialyzed into 10 mM TRIS, 200 mM NaCl pH 8 dialysis buffer. The same dialysis buffer was used to dialyze the ITC CP samples, CPB and CPA6. 1.1 mM P1C10 was titrated at 3 µl injection volumes into the

cell that contained 0.1 mM CPA6 or CPB at 25 °C. The experiment was repeated at 10 °C for both CPA6 and CPB. All injections had a 20 second duration, 180 second spacing, and 307 RPM stir speed. K_D values were calculated when possible using nonlinear least squares fits of ITC data with a single-site binding model in Origin 7.0 software.

4.3.8 Enzyme Linked Immunosorbent Assays

CP samples tested in ELISA included Human CPA1 (Sino Biological # 10504-H08H), Human CPA2 (Sino Biological #10499-H08H), Human CPA6 (MyBioSource.com #MBS1330838), Human CPE (Sino Biological # 10069-H08H), Mouse CPA1 (Sino Biological # 50448-M08H), Mouse CPB2 (Sino Biological # 50963-M08H), Human CPE (Abcam #ab139787). A maxisorp plate (Thermo Scientific #442404) was coated with P1C10 or RBC36 tagless monomer at 7 µg/ml overnight at 4 °C. The plate was washed with wash buffer, PBS+0.05% tween 20 (Thermo Scientific #28352), twice. The plate was blocked in blocking buffer (Thermo Scientific #37525) with 300 µl per well for 60 minutes at ambient temperature on a plate shaker at 350 RPM. The plate was washed 4 times with wash buffer. Samples were transferred to the plate 100 µl per well and incubated at ambient temperature on a plate shaker at 350 RPM for 90 minutes. Again, the plate was washed 4 times with wash buffer. Mouse anti-6x Histidine tag antibody-HRP (Invitrogen #MA1-80218, clone AD1.1.10) was diluted 1:500 in assay diluent (0.1% BSA + 0.045% tween 20) and was added 100 µl per well and incubated at ambient temperature on a plate shaker at 350 RPM for 60 minutes. The plate was washed an additional 4 times with wash buffer. TMB (brought to room temperature for 1 hour, Thermo Scientific #34022) was added 100 µl per well and incubated at ambient temperature for 20 minutes. Stop solution (Invitrogen #SS04) was added 100 µl per well

and incubated at ambient temperature for 5 minutes. The plate was read at 450nm. For instances of VLR-Fc fusions, the plate was coated with CPs at 7 μg/ml overnight at 4 °C and titrated with VLR-Fcs. All subsequent steps were processed the same as the monomers on the same plate except the detection method which used mouse anti-rabbit Fc antibody-HRP diluted 1:20,000 (Genscript #14H9H10).

4.4 Results

4.4.1 Carboxypeptidase A6 cannot be produced as a soluble recombinant monomer

To further analyze the apparent binding of P1C10 and CPA6, in-house production of CPA6 was attempted. MyBioSource.com, the supplier of CPA6 (#MBS1330838), charges $1,435 per 1 milligram of CPA6. Thus, purchasing CPA6 for potential structural analysis and kinetic binding experiments was not feasible. Initial attempts to produce CPA6 were based on previously published methods in HEK cells [126]. The plasmid from previously published CPA6 studies was gifted by Dr. Peter Lyons [126]. The plasmid was used to transfect HEK cells for soluble protein expression and subsequent purification using Ni-NTA. Unfortunately, the soluble expression was negligible. The previously published production method used an adherent HEK293 culture for cell-based analysis, and soluble expression and purification of CPA6 was never attempted from the method. Therefore, it was unsurprising that CPA6 expression was negligible.

CPA6 expression tests were further attempted in Rosette and SHuffle cells. All constructs were developed in pET vectors, and the constructs are listed in Table 4-1. The expression tests followed manufacturer recommendations for transformation and culture conditions (Novagen #70954 and NEB # C3029J). While most constructs expressed soluble material, the material was never successfully purified via IMAC (attempts using GST, Ni-NTA, and Cobalt resin were made for all relevant constructs). The inability to purify the soluble material suggested that the material was not correctly folded and poorly soluble when detached from its native environment of the cellular ECM. Indeed, Dr. Lyons confirmed that his lab was never able to make soluble CPA6 for recombinant purification. Interestingly, one of the constructs was the exact sequence used by

MyBioSource.com (for their *E. coli* expressed product), yet we were still unable to purify soluble material with that sequence.

4.4.2 Computational models reveal putative binding residues of P1C10

Given the challenge of purifying soluble CPA6, the other putative binding partners that were previously identified in the Shusta lab were investigated to confirm binding. Of the four carboxypeptidases that were tested, CPB was the only carboxypeptidase with a solved crystal structure (PDB ID: 1XLI). Thus, 1XLI was used for a docked model with P1C10 (PDB ID: 9CJ0) since previous binding studies in the Shusta lab demonstrated 46.3 nM affinity between P1C10 and CPB. P1C10 and CPB sequences were run as a complex with AlphaFold Multimer via COSMIC[2]. All 5 machines from AlphaFold produced models of the docked P1C10 and CPB complex (Figure 4-1). All the models predicted the same interface for binding of P1C10 with slight variations of the interface that would be on CPB. The models were ranked 0-4 with 0 being the best. Each model is assessed using predicted local distance difference test (pLDDT) for each residue in the model. A perfect pLDDT is 100%. Any value < 100% indicates a lower probability that the given residue is in the correct position. While most residues in the CPB-P1C10 models had high pLDDT scores, some of the positions dipped below 70% indicating there are likely errors in the models.

Since CPA6 was identified as a putative binding partner of P1C10, the CPA6 sequence was run in complex with the P1C10 sequence of AlphaFold Multimer. All 5 machines of AlphaFold Multimer produced top-predicted models of the complex (Figure 4-2). While the binding interface of P1C10 was constant for all 5 models, the predicted binding interface of CPA6 had variability between the models. The CPA6 docked models had less potential accuracy compared to the CPB models

as seen with the pLDDT scores. CPA6 does not have a solved crystal structure, so we would expect the CPA6-P1C10 models to have worse accuracy than the CPB-P1C10 models. Before docking attempts, the AlphaFold machines must make a homology model for CPA6 monomer using threading templates from other solved crystal structures in the Protein Data Bank. Thus, the model of the docked complex, P1C10-CPA6, is based on a model of CPA6 which adds compounding layers of inaccuracy compared to the P1C10-CPB model that uses two solved crystal structures to model the complex.

The python script InterfaceResidues.py was run in PyMOL for all ten models in figures 4-1 and 4-2. The script was used to identify residues in P1C10 that might be implicated in binding based on the models. While the models can be sufficient predictors of how molecules could bind, they do not guarantee binding. AlphaFold Multimer will always return docked molecules in putative binding complexes regardless of whether binding is real. Thus, AlphaFold Multimer was not used to confirm binding, but rather it was used to assess putative binding residues of P1C10. As such, the binding residues identified in the models were used to create a heat map of P1C10 residues that were identified as potential binding residues in the CPB and CPA6 models (Table 4-2). The heat map was augmented with the residues identified as buffer binders from the crystal studies of P1C10 in Chapter 3. Residues were assigned a score based on the number of times they were implicated in binding with a maximum score of 11 (10 models and 1 crystal structure). Interestingly, the LRRs of P1C10 were implicated in binding with a very clear pattern as seen in the heat map. This is in accordance with what is expected of VLR binding. Residues of particular interest scored 11 (highest possible score) being implicated in buffer-binding and binding in all ten models. These residues of interest include: W62 and Q137. 6 residues were assigned 10

points each for their implication in binding, and these residues included: R35, Y37, Y86, N110, G138, and H139.

While AlphaFold multimer can produce a probability score for residue positions, it does not provide the probability of true docking. Any two proteins submitted to the AlphaFold server will return “docked” models with no indication of how likely the two proteins are to dock in nature. As suggested by other authors, the docked complexes can be used as a hypothesis to test with biochemical experimentation, but should never be used for definitive analysis of a binding complex [129]. Therefore, to evaluate the binding of P1C10 with CPB and CPA6 as observed in the models, an alanine scan was performed.

4.4.3 P1C10 Alanine Scan shows no statistical significance between putative dissociation constants

The highest scoring residues of P1C10 predicted to be involved in binding were used to engineer alanine mutants. Alanine mutants were engineered by mutating the predicted binding residues to alanine in an attempt to knockout binding. 6 clones were engineered in pET vectors for monomeric expression in SHuffle cells (Table 4-3) as described in the SHuffle protocol of Chapter 2. All clones were soluble and able to be purified by Ni-NTA except for clone A3. The Y86A mutation in clone A3 led to insolubility in the molecule, and the clone was unable to be purified. Interestingly, clone A6 that included all five mutations, including Y86A, was soluble and able to be purified. The subsequent 5 clones were evaluated for binding in the CPA6 ELISA. The residues that were mutated are highlighted in Figure 4-3.

Clones A1, A2, A4, A5, A6 were evaluated in binding ELISAs with CPA6. P1C10 wildtype monomer (P1C10 WT) was used as a positive control on each assay play. RBC36 monomer was used as a non-binding control for each assay. VLRs were coated on the maxisorp plates and CPA6 (#MBS1330838) was titrated to assess binding via anti-6xHis-HRP. A four-parameter logistic (4PL) regression was fit to all absorbance readings from the ELISA plates to compare binding curves of each alanine mutant with the positive and negative control on each plate (Figures 4-4, 4S-1). Surprisingly, all alanine mutants had binding curves very similar to the P1C10 WT, suggesting that the mutated binding residues did not affect binding. Furthermore, the non-binding control, RBC36, bound at the same affinity as P1C10 WT and all alanine mutants. In many instances, the alanine mutants and RBC36 bound with a stronger dissociation constant than the P1C10 WT, a stark contrast to what was expected. This finding prompted investigation into the MyBioSource.com CPA6 product, and subsequent ELISAs were performed to confirm this unusual finding.

4.4.4 Commercially available Carboxypeptidase A6 has poor purity by Size Exclusion Chromatography

Following the unexpected alanine scan results, it was desired to assess the purity of the CPA6 material that was purchased. SEC was used to evaluate CPA6 (#MBS1330838) and rat CPB (#MBS143539) purchased from MyBioSource.com. SEC traces demonstrated mixed purity of CPA6 (Figure 4-5a). The CPA6 SEC trace showed high molecular weight species as well as low molecular weight species in the solution with no prominent monomer suggesting that the material was misfolded. In contrast, Rat CPB (Figure 4-5b) had high monomeric purity with a

prominent monomer peak. Both samples had been reconstituted in a neutral mobile phase according to the manufacturer recommendations immediately prior to injecting.

4.4.5 NativePage of Carboxypeptidase A6 further validates poor purity

Following the concerning SEC results of CPA6, the carboxypeptidases were further evaluated by NativePAGE. CP samples included Human CPA1 (Sino Biological # 10504-H08H), Human CPA2 (Sino Biological #10499-H08H), Human CPA6 (MyBioSource.com #MBS1330838), Human CPB1 (MyBioSource.com # MBS8120318), Human CPE (Sino Biological # 10069-H08H), Mouse CPA1 (Sino Biological # 50448-M08H), Rat CPB (MyBioSource.com #MBS143539) (Figure 4-6a). All samples produced clear bands indicative of single species except for CPA6 and CPB1, both from MyBioSource.com. CPA6 and CPB1 produced smears without a prominent band suggesting these are mixed species samples of aggregates. A second sample of CPA6 was reconstituted from the lyophilized material and run on another NativePAGE (Figure 4-6b). The second NativePAGE of CPA6 demonstrated the same results as the first gel, confirming there is not a prominent single species in the CPA6 sample.

4.4.6 Crystal Screen of Carboxypeptidase B shows no complexing with P1C10

The rat CPB (#MBS143539) showed better SEC and NativePage results than CPA6. Both SEC and NativePage suggested rat CPB was present with high purity monomer. Since CPB was identified as a putative binding partner of P1C10, a crystal screen was attempted with rat CPB. A crystal screen was desired to attempt to definitively solve the crystal structure of P1C10 in complex with a carboxypeptidase. Rat CPB was chosen as it was significantly cheaper than the human

carboxypeptidases (50 milligrams for $625). The commercial material was cheap enough to purchase for a crystal screen. Furthermore, rat CPB (UniProt ID P19223) has 89% homology to human CPB (UniProt ID P15086). Additionally, the original LC-MS/MS pulldown method used bEND.3 cells, a murine cell line. Therefore, if CPB was a true binding partner of P1C10, we would expect the rodent isoforms to bind P1C10. Unfortunately, the crystal screen of rat CPB and P1C10 did not produce any crystals of a complex. Both rat CPB and P1C10 produced crystals on their own in many of the conditions, but no crystals were generated when P1C10 was combined with carboxypeptidase.

4.4.7 Isothermal titration calorimetry of Carboxypeptidases with P1C10 show no binding

Following the unsuccessful crystal screen of rat CPB in complex with P1C10, the binding was further investigated. CPB and CPA6 were available in enough quantity to perform ITC trials with P1C10. Both CPB and CPA6 were tested at different temperatures and maximal concentrations. CPA6 did show any heat spikes above the background. Rat CPB showed minimal heat spikes above the background. The rat CPB ITC data was fit to a binding curve that produced a K_D = 100 µM (Figure 4-7). While the ITC results fit a K_D = 100 µM, the VP-ITC microcalorimeter upper limit of quantitation is 10 µM. Thus, a K_D = 100 µM is outside the range of quantitation for this instrument and ITC in general. Furthermore, a K_D = 100 µM does not have biological significance. This binding constant was much worse than the binding constant previously observed in the Shusta lab and does not correlate with what was observed with previous *in vivo* studies in mouse models with dosages well below 100 µM [48] suggesting that rat CPB is not a binding partner of P1C10.

4.4.8 Enzyme Linked Immunosorbent Assays of P1C10 and Carboxypeptidase A6 suggest false positive binding

Following the unexpected ITC results, it was desired to re-evaluate the binding of P1C10 with all the carboxypeptidases. CP samples tested in ELISA included Human CPA1 (Sino Biological # 10504-H08H), Human CPA2 (Sino Biological #10499-H08H), Human CPA6 (MyBioSource.com #MBS1330838), Human CPE (Sino Biological # 10069-H08H), Mouse CPA1 (Sino Biological # 50448-M08H), Mouse CPB2 (Sino Biological # 50963-M08H), Human CPE (Abcam #ab139787). All samples were tested with monomeric P1C10. No carboxypeptidases demonstrated any positive binding except for CPA6 which produced a K_D = 32.1 nM (Figure 4-8a). These results contradicted previous findings in the Shusta lab which suggested P1C10 bound CPE, CPB, and CPA. To confirm these results were not an artifact of the P1C10 monomer having worse binding than P1C10-rFc, the experiment was repeated for human CPE (the peptide identified from the LC-MS/MS pulldown) as well as mouse CPB2 with both monomeric P1C10 and P1C10-rFc (P1C10 conjugated to rabbit Fc, the construct from the previous binding studies in the Shusta lab). Neither monomeric P1C10 nor P1C10-rFc produced saturated binding curves (Figure 4-8b) suggesting CPE and CPB2 are not the binding partners of P1C10. CPA6 was further evaluated to determine if the positive binding observed by ELISA was true biomolecular binding. Monomeric RBC36 and RBC36-rFc were tested as non-binding controls. P1C10 monomer, P1C10-rFc, and Alanine mutant A6 were tested for binding to CPA6. All samples, including non-binding controls, produced dissociation constants < 70 nM demonstrating that neither P1C10 construct bound with greater affinity than the non-binding controls (Figure 4-8c).

4.5 Discussion

To further validate and characterize previously conducted binding studies between CPA6 and P1C10, it was desired to produce CPA6 at large-scale. The commercially available CPA6 was cost prohibitive for many analytical assays. In an attempt to minimize cost and make CPA6 in-house, 11 different CPA6 constructs in pET vectors along with a CPA6 construct in pcDNA3.1 were tested for expression, correct folding and solubility in Rosetta, SHuffle and HEK293 expression tests. None of the constructs were able to be purified suggesting CPA6 was improperly folded when expressed as a recombinant protein. This assumption was supported by previous authors who suggested CPA6 is ECM-associated and insoluble when removed from the ECM [126]. One of the pET constructs tested for *E. coli* expression, 6xHis tag with Small Ubiquitin-like Modifier (SUMO) fusion, was the exact genetic sequence of the protein purchased from MyBioSource.com, yet this protein was not correctly folded in our in-house attempts. This was concerning since the commercially available CPA6 from MyBioSource.com was expressed in *E. coli,* as well. Thus, the commercially available CPA6 was suspected to be misfolded. NativePAGE and SEC results confirmed that the CPA6 purchased from MyBioSource.com indeed was a heterogenous mixture of misfolded aggregates. Thus, the other CPs identified as putative binding targets were investigated.

The binding of P1C10 to CPE overserved previously in the Shusta lab was unable to be replicated in the present study. Binding studies presented in this chapter represent comprehensive evaluation of all commercially available carboxypeptidases. The peptide previously identified from the LC-MS/MS pulldown was GNETIVNLIHSTR. This peptide is present in both human and

mouse CPE. Thus, it would not be expected that we observe a difference in binding between human or mouse CPE isoforms. Binding was not observed using P1C10-rFc nor P1C10 monomer. CPB and CPA were tested for binding to P1C10, as well. CPB and CPA binding could not be replicated with any commercially available isoforms. CPA6 purchased from MyBioSource.com bound P1C10 monomer and P1C10-rFc. Unfortunately, CPA6 also demonstrated positive binding to the A6 mutant and the non-binding controls, RBC36 monomer and RBC36-rFc. Thus, the P1C10 binding to CPA6 was not above the background for the assay. Nanomolar affinity is unquestionably high, and it was reasonable to assert CPA6 as a putative binding partner of P1C10 given the initial ELISA results. However, aggregates are known to produce false positives in ELISAs due to their higher absorbance readings [130-132]. Given the highly heterogenous composition of CPA6 as observed by SEC and NativePAGE, it is unsurprising that this product is “sticky” with high absorbance readings in ELISAs. Furthermore, the CPA6 binding could not be replicated with any other assays such as ITC. Taken together, these results suggest CPA6 is not the binding target of P1C10. Additionally, binding to CPE could not be replicated, and it is of note that only one peptide from CPE was identified in any of the LC-MS/MS pulldowns. A robust LC-MS/MS pulldown would anticipate high peptide coverage of the putative binding target. Thus, it is unlikely that any of these CPs are the binding target of P1C10 given the ELISA results.

4.6 Tables

Table 4-1 Carboxypeptidase A6 recombinant protein constructs in pET vectors. 11 constructs were designed for expression tests in SHuffle and Rosetta cells. All constructs were trialed with both cobalt and nickel purification methods. Constructs containing GST tags were also attempted to be purified by GST resin.

Gene Name	Fusion Partner	Tags	Cleavage Site	MW (kdA)
Signal-pro-TEV-CPA6mutant-6xHis	signal/pro	C-term 6XHis	TEV	53
CPA6mutant-6xHis	-	C-term 6XHis	-	37
6xHis-TEV-CPA6mutant	-	N-term 6XHis	TEV	38
6xHis-TEV-CPA6	-	N-term 6XHis	TEV	38
6xHis-SUMO-TEV-CPA6	SUMO	N-term 6XHis	TEV	50
6xHis-SUMO-TEV-CPA6mutant	SUMO	N-term 6XHis	TEV	50
GST-Thrombin-CPA6	GST	-	Thrombin	63
8xHis-MBP-Linker-TEV-CPA6	MBP	N-term 8XHis	TEV	80
8xHis-GST-Linker-TEV-CPA6	GST	N-term 8XHis	TEV	65
6xHis-TEV-CPA6	-	N-term 6XHis	TEV	38
CPA6-6xHis	-	C-term 6XHis	-	36

Table 4-2 Heatmap of putative P1C10 binding residues. Binding residues identified from each of the AlphaFold Multimer models along with the buffer-binding residues from the P1C10 crystal structure were compared and assigned scores. 'P1C10-CPB Complex' Models 0-4 correspond to residues identified from the AlphaFold Models in Figure 4-2. 'P1C10-CPA6 Complex' Models 0-4 correspond to residues identified from the AlphaFold Models in Figure 4-3. 'P1C10 Buffer Binders Crystal Structure' correspond to the residues bound to buffer molecules identified in the solved crystal structure of P1C10. Scores are a sum of the total number of instances the residue was implicated in binding with a maximum score of 11 (ten models and one crystal structure). The darker green shades correlate with higher scores indicating those residues are most likely to be implicated in binding.

P1C10			P1C10-CPB Complex					P1C10-CPA6 Complex					P1C10 Buffer Binders	Points
Position	Residue	SS	Model 0	Model 1	Model 2	Model 3	Model 4	Model 0	Model 1	Model 2	Model 3	Model 4	Crystal Structure	
1	A	Loop												0
2	C	Loop												0
3	P	Loop												0
4	S	Loop												0
5	Q	Loop												0
6	C	Loop												0
7	S	β Strand					X					X		2
8	C	β Strand												0
9	D	β Strand		X		X	X		X	X	X	X		7
10	Q	Loop												0
11	T	Loop												0
12	T	β Strand				X	X		X	X	X	X		6
13	V	β Strand												0
14	K	β Strand		X		X	X		X	X	X	X		7
15	C	Loop												0
16	H	Loop		X	X	X	X	X	X	X	X	X		9
17	S	Loop		X	X	X	X	X	X	X	X	X		9
18	R	Loop					X							1
19	R	Loop				X			X	X				3
20	L	Loop												0
21	T	Loop												0
22	S	Loop												0
23	V	Loop												0
24	P	Loop												0
25	A	Loop												0
26	G	Loop												0
27	I	Loop												0
28	P	Loop												0
29	T	Loop												0
30	T	Loop												0
31	T	Loop												0
32	K	Loop									X	X		2
33	I	β Strand		X		X	X		X	X		X		6
34	L	β Strand												0
35	R	β Strand	X	X	X	X	X	X	X	X	X	X		10
36	L	Loop												0
37	Y	Loop	X	X	X	X	X	X	X	X	X	X		10
38	S	Loop		X	X	X	X	X	X	X	X	X		9
39	N	Loop							X					1
40	Q	Loop							X					1
41	I	Loop												0
42	T	Loop												0
43	K	Loop												0
44	L	Loop												0
45	E	Loop												0
46	P	Loop												0
47	G	Loop												0
48	V	Helix												0
49	F	Helix												0
50	D	Helix												0
51	H	Loop												0
52	L	Loop												0
53	V	Loop												0
54	N	Loop												0
55	L	Loop												0
56	E	Loop										X		1
57	K	β Strand		X	X	X	X	X	X	X	X	X		9
58	L	β Strand												0
59	Y	β Strand		X	X	X	X	X	X	X	X	X		9
60	I	Loop												0
61	S	Loop		X	X		X	X	X	X	X	X	X	9
62	W	Loop	X	X	X	X	X	X	X	X	X	X	X	11
63	N	Loop							X					1
64	Q	Loop							X					1
65	L	Loop												0
66	S	Loop												0
67	A	Loop												0
68	L	Loop												0
69	P	Loop												0
70	V	Loop												0
71	G	Loop												0
72	V	Loop												0
73	F	Loop												0
74	D	Loop												0
75	K	Loop												0
76	L	Loop												0
77	T	Loop												0
78	K	Loop												0
79	L	Loop												0
80	T	Loop												0
81	H	β Strand		X	X	X	X		X	X	X			7
82	L	β Strand												0
83	S	β Strand		X	X		X		X	X	X	X	X	8

84	L	Loop												0
85	G	Loop												0
86	Y	Loop	X	X	X	X	X	X	X	X	X	X		10
87	N	Loop							X					1
88	Q	Loop							X	X	X			3
89	L	Loop												0
90	K	Loop												0
91	S	Loop												0
92	V	Loop												0
93	P	Loop												0
94	R	Loop												0
95	G	Loop												0
96	A	Loop												0
97	F	Loop												0
98	D	Loop												0
99	N	Loop												0
100	L	Loop												0
101	K	Loop												0
102	S	Loop												0
103	L	Loop												0
104	T	Loop												0
105	H	β Strand		X	X		X	X			X	X		6
106	I	β Strand												0
107	W	β Strand	X	X	X	X	X		X	X		X	X	9
108	L	Loop												0
109	L	Loop	X	X			X	X	X	X	X	X		8
110	N	Loop	X	X	X	X	X	X	X	X	X	X		10
111	N	Loop												0
112	P	Loop												0
113	W	Loop												0
114	D	Loop												0
115	C	Loop												0
116	E	Loop												0
117	C	Loop												0
118	S	Helix												0
119	D	Helix												0
120	I	Helix												0
121	L	Helix												0
122	Y	Helix												0
123	L	Helix												0
124	K	Helix												0
125	N	Helix												0
126	W	Helix												0
127	I	Helix												0
128	V	Helix												0
129	Q	Helix												0
130	H	Helix												0
131	A	Loop	X											1
132	S	Loop	X											1
133	I	Loop												0
134	V	Loop												0
135	N	Loop				X								1
136	L	Loop	X		X			X		X				4
137	Q	Loop	X	X	X	X	X	X	X	X	X	X	X	11
138	G	Loop	X	X	X	X	X	X	X	X	X	X		10
139	H	Loop	X	X	X	X	X	X	X	X	X	X		10
140	G	Loop	X			X	X	X						4
141	G	Loop	X			X		X						3
142	V	Loop	X											1
143	D	Loop	X					X						2
144	N	Loop	X			X	X	X	X	X	X	X		8
145	V	Loop												0
146	K	Loop	X					X						2
147	C	Loop												0
148	S	Loop												0
149	G	Loop												0
150	T	Loop												0
151	N	Loop	X				X	X				X		4
152	T	Loop	X											1
153	P	Loop	X											1
154	V	Helix												0
155	R	Helix	X					X						2
156	A	Helix	X											1
157	V	Loop												0
158	T	Loop												0
159	E	Loop												0
160	A	Loop												0
161	S	Loop												0
162	T	Loop												0
163	S	Loop												0
164	P	Loop												0
165	S	Loop												0
166	K	Loop												0
167	C	Loop												0
168	P	Loop												0

Table 4-3 Alanine mutants of P1C10. All clones had single point mutations except A6 which had all 5 mutations of clones A1-A5. A3 was poorly soluble and unable to be purified. A1, A2, A4, A5, A6 were soluble and purified by Ni-NTA purification. All purified clones were confirmed by LC-MS analysis.

Label	Mutation
A1	Y37A
A2	W62A
A3	Y86A
A4	N110A
A5	Q137A
A6	Y37A, W62A, Y86A, N110A, Q137A

4.7 Figures

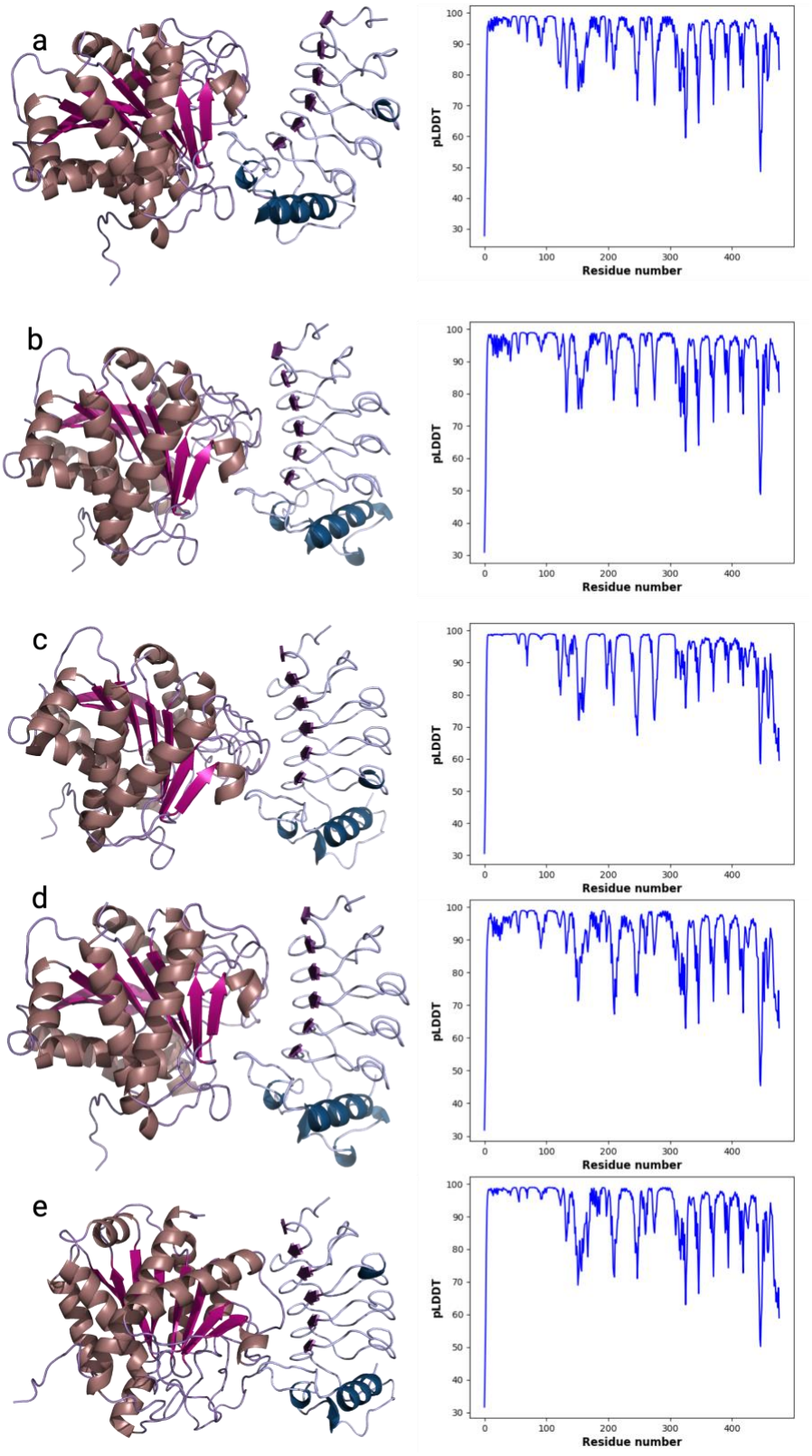

a
b
c
d
e
pLDDT
Residue number
100
90
80
70
60
50
40
30
0
100
200
300
400

Figure 4-1 AlphaFold Multimer models of Carboxypeptidase B (brown/magenta) docked with P1C10 (blue/purple). The 5 machines from AlphaFold all produced models of CPB docked with P1C10. Each machine produced a top-predicted model and corresponding predicted local distance difference test (pLDDT) for each model. Each machine produced a top-predicted model. (a) is model 0, (b) is model 1, (c) is model 2, (d) is model 3, (e) is model 4.

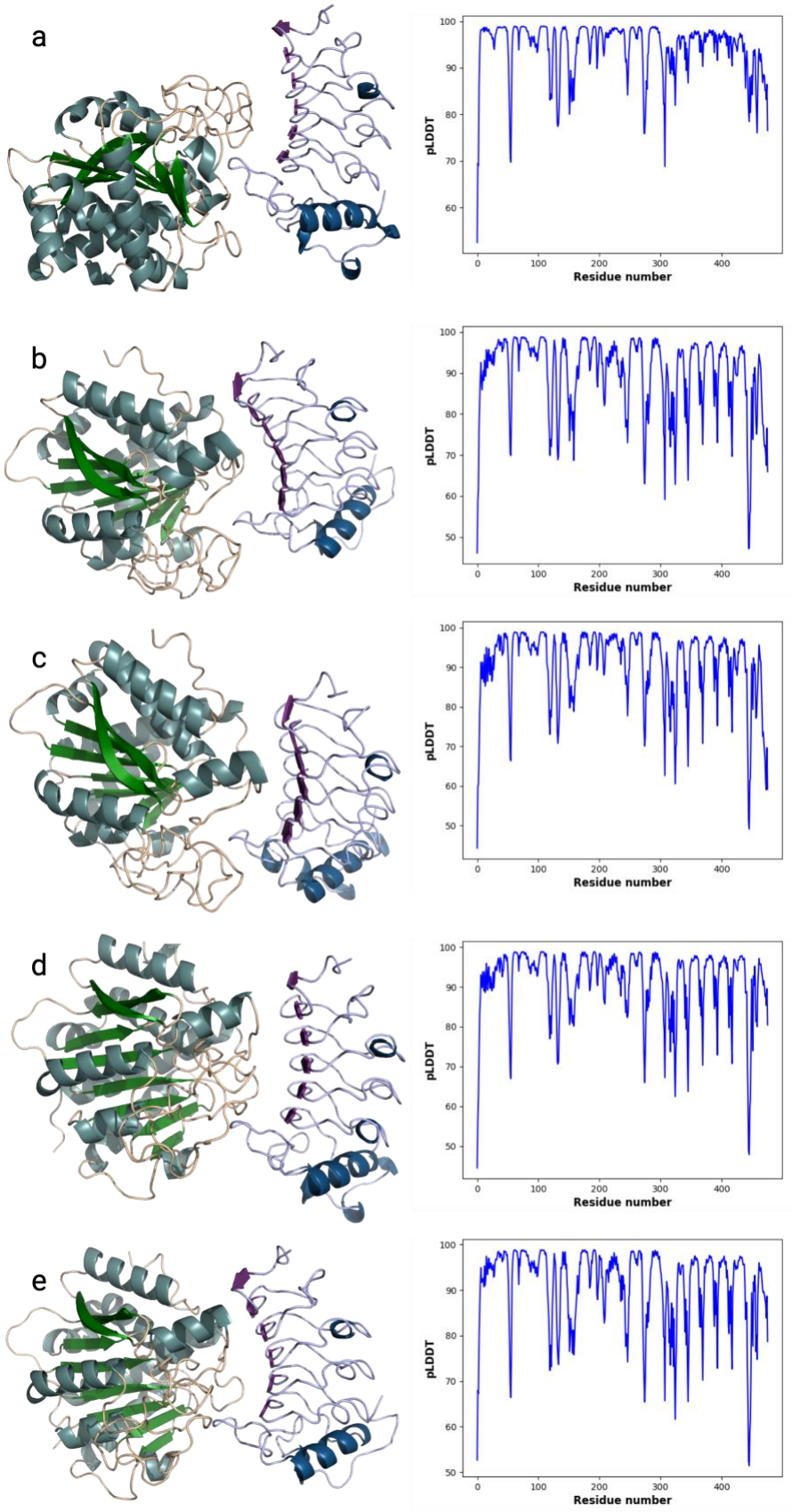

a
b
c
d
e
pLDDT
Residue number

Figure 4-2 AlphaFold Multimer models of Carboxypeptidase A6 (teal/green) docked with P1C10 (blue/purple). The 5 machines from AlphaFold all produced models of CPB docked with P1C10. Each machine produced a top-predicted model and corresponding predicted local distance difference test (pLDDT) for each model. (a) is model 0, (b) is model 1, (c) is model 2, (d) is model 3, (e) is model 4.

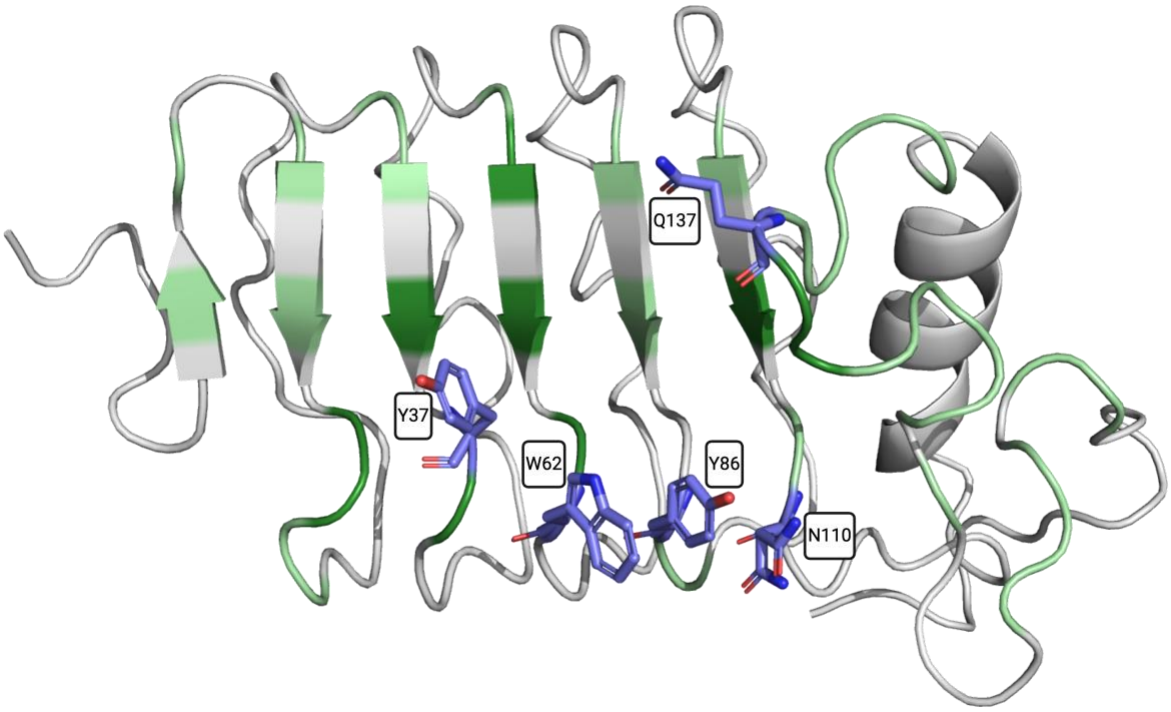


Figure 4-3 Structure of P1C10 with Suspect Binding Residues Highlighted. The residues identified as putative binding residues in Table 4-2 are highlighted in green. The residues with higher scores in Table 4-2 are highlighted in darker green (more likely to be involved in binding). Of the dark green residues, the residues chosen for alanine mutation (Table 4-3) are depicted as sticks with purple carbon chains.

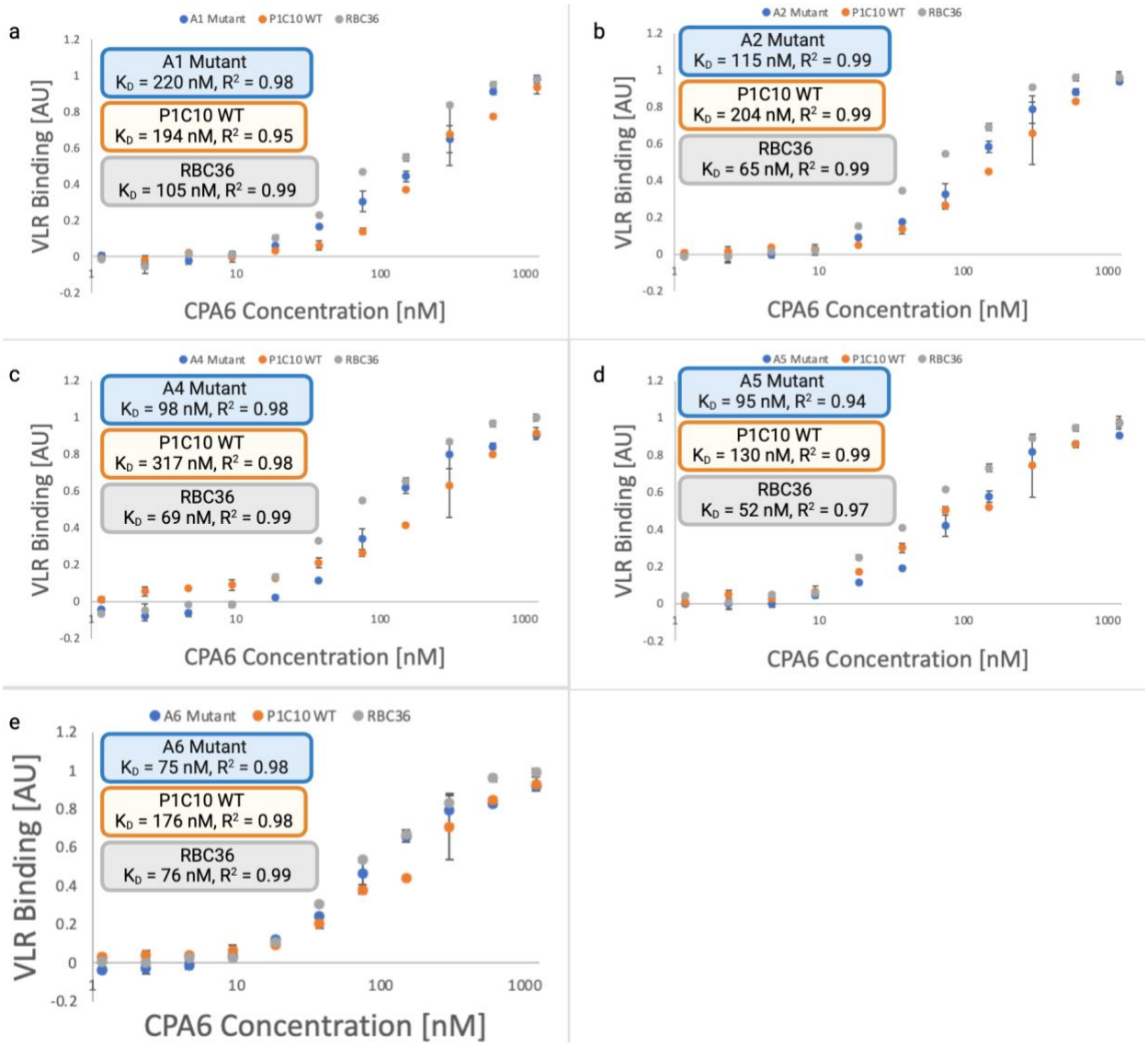


Figure 4-4 ELISA results of Alanine mutants titrated with Carboxypeptidase A6. All plates were analyzed at 450 nm. The absorbance readings were normalized and fit using a four-parameter logistic regression using the MyAssays.com Four Parameter Logistic Curve program which generated the values displayed for K_D and R^2. Binding curves represent averages (Alanine mutants, n = 3; P1C10 WT, n = 3; RBC36, n = 2) with ± standard error. The curves from MyAssays.com do not allow standard error bars to be added, so these are available in the supplement (Figure 4S-1).

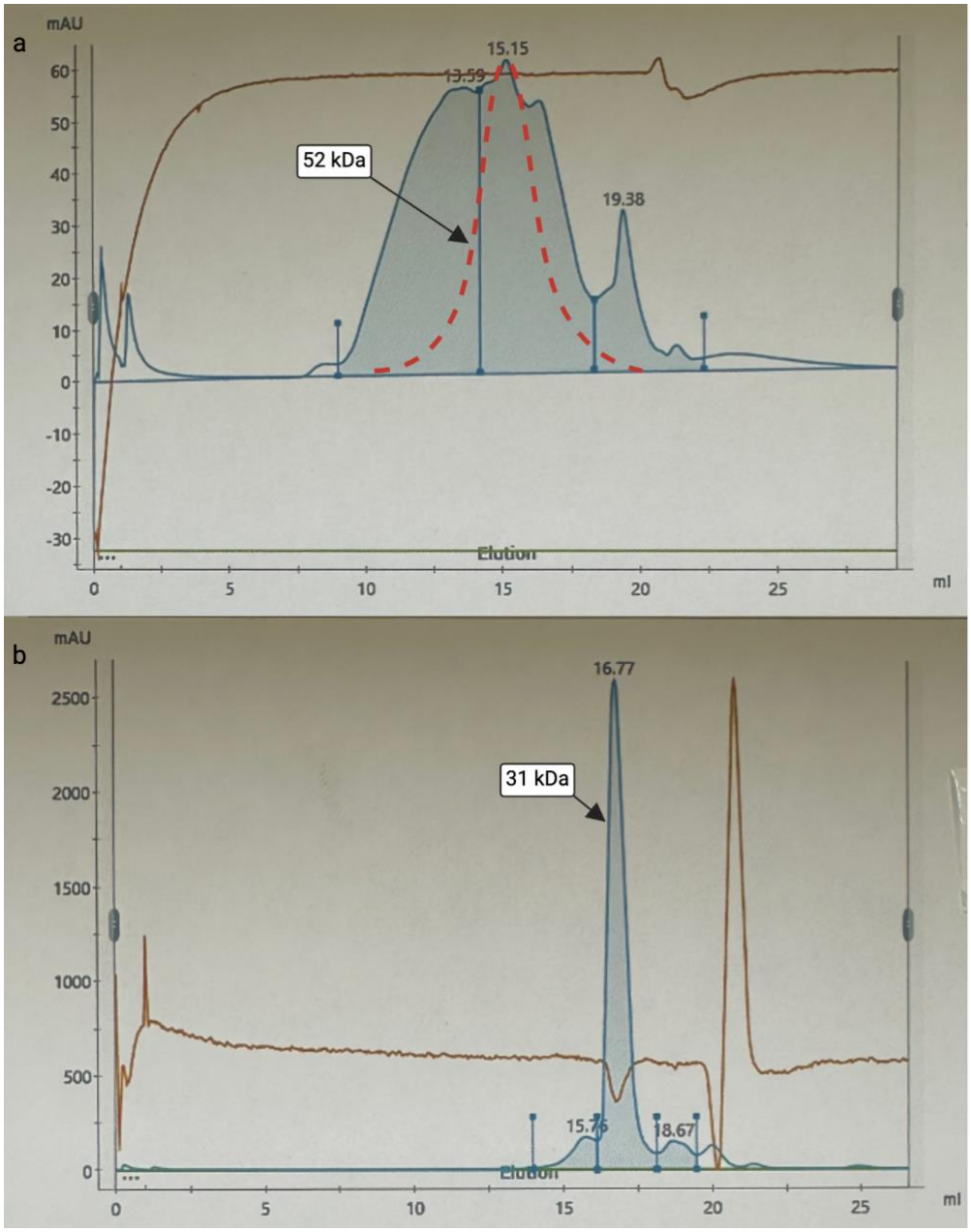


Figure 4-5 Size Exclusion Chromatography results from Carboxypeptidase A6 and Carboxypeptidase B. (a) CPA6 SEC trace with the theoretical monomer depicted with the dashed red line (MW = 52kDa for the recombinant CPA6-SUMO fusion) and (b) CPB SEC trace with the arrow pointing to the main peak and monomer (MW = 31kDa for the recombinant mature enzyme).

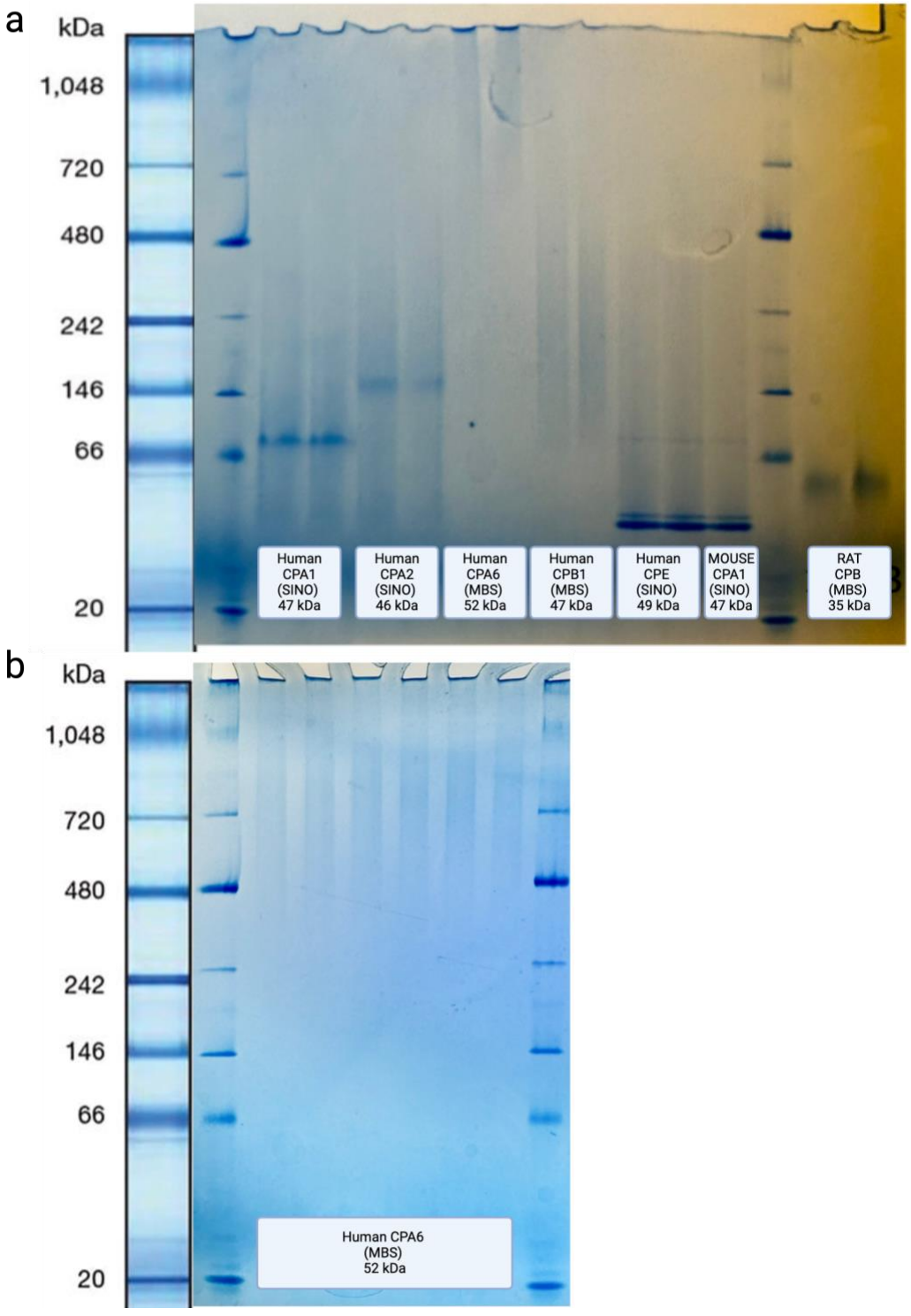


Figure 4-6 NativePAGE of Commercial Carboxypeptidases. (a) Carboxypeptidases were run in duplicate. Mouse CPA1 was only run once due to limited material availability. (b) CPA6 was run on a second NativePAGE to confirm results (n = 6).

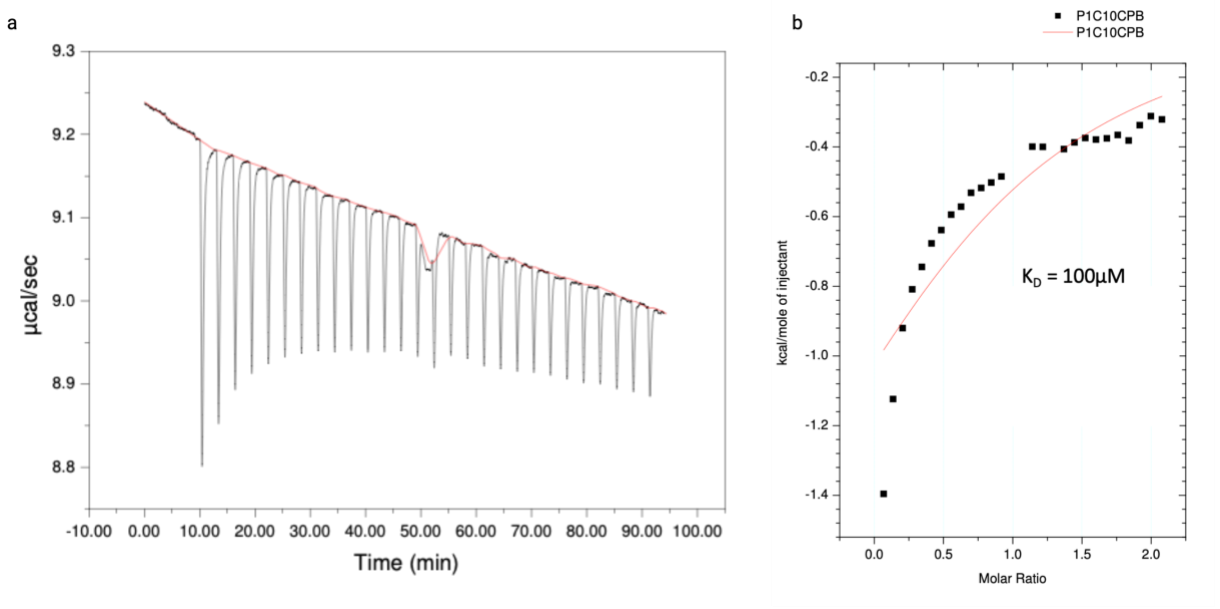
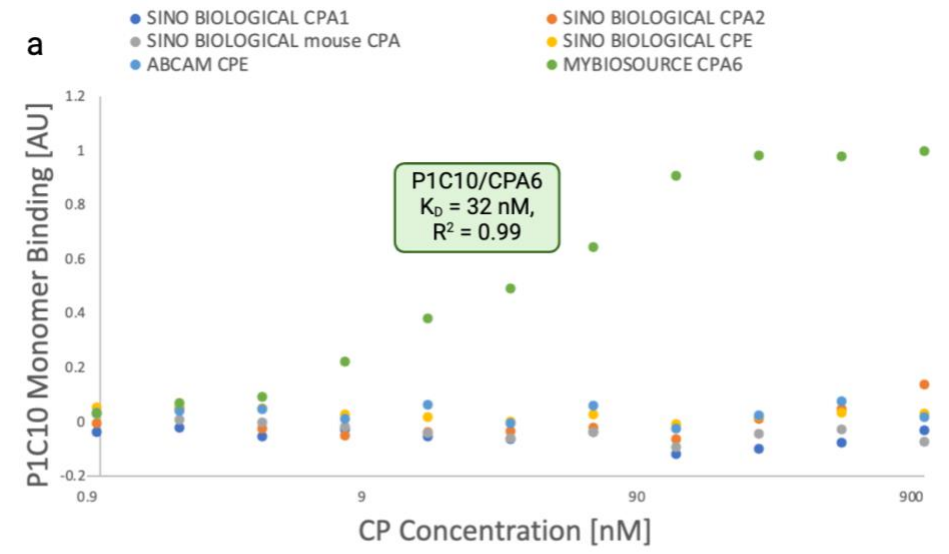
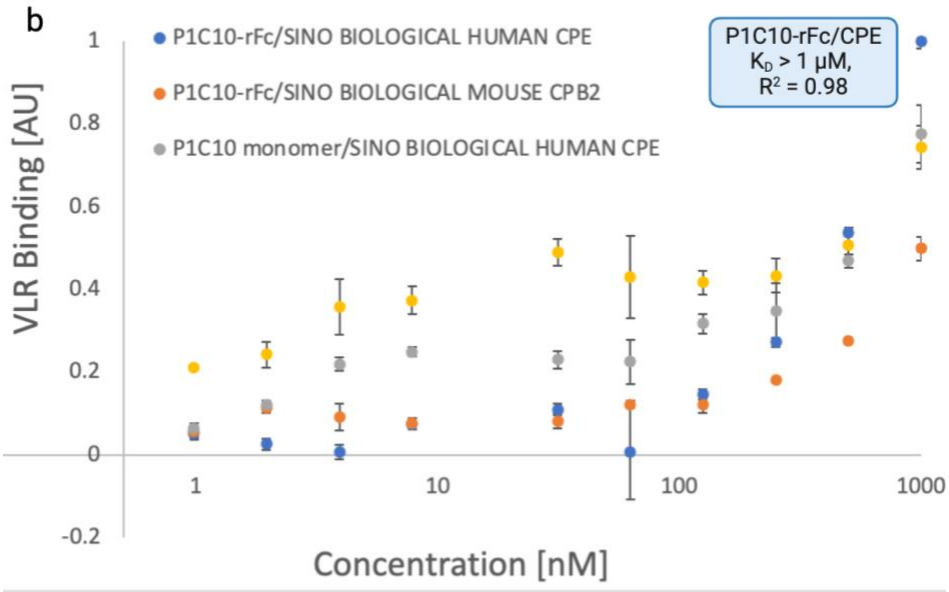


Figure 4-7 ITC of rat Carboxypeptidase B and P1C10. (a) P1C10 was titrated into rat CPB. (b) ITC data was fit to a nonlinear least squares model that produced a k_D of 100 µM for monomeric P1C10 produced in SHuffle cells.


a
SINO BIOLOGICAL CPA1
SINO BIOLOGICAL CPA2
SINO BIOLOGICAL mouse CPA
SINO BIOLOGICAL CPE
ABCAM CPE
MYBIOSOURCE CPA6
P1C10/CPA6
K_D = 32 nM,
R² = 0.99
P1C10 Monomer Binding [AU]
CP Concentration [nM]


b
P1C10-rFc/SINO BIOLOGICAL HUMAN CPE
P1C10-rFc/SINO BIOLOGICAL MOUSE CPB2
P1C10 monomer/SINO BIOLOGICAL HUMAN CPE
P1C10-rFc/CPE
K_D > 1 µM,
R² = 0.98
VLR Binding [AU]
Concentration [nM]

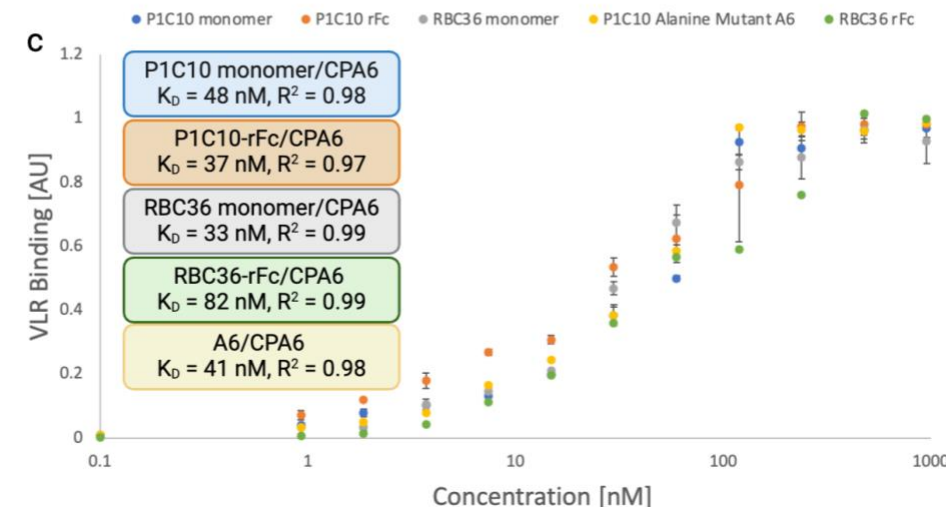

c
P1C10 monomer
P1C10 rFc
RBC36 monomer
P1C10 Alanine Mutant A6
RBC36 rFc
P1C10 monomer/CPA6
K_D = 48 nM, R² = 0.98
P1C10-rFc/CPA6
K_D = 37 nM, R² = 0.97
RBC36 monomer/CPA6
K_D = 33 nM, R² = 0.99
RBC36-rFc/CPA6
K_D = 82 nM, R² = 0.99
A6/CPA6
K_D = 41 nM, R² = 0.98
VLR Binding [AU]
Concentration [nM]

Figure 4-8 ELISAs of carboxypeptidases with P1C10. (a) Monomeric P1C10 was evaluated for binding with all commercially available carboxypeptidases. (b) P1C10 monomer and P1C10-rFc were tested together for binding to human CPE and mouse CPB2. Data represent averages (n=3) with standard error. (c) P1C10 monomer, P1C10-rFc, RBC36 monomer, RBC36-rFc, and alanine mutant A6 were evaluated for binding to CPA6. VLR monomers (no epitope tag) were used to coat the ELISA plate and titrated with CPA6 (6xHis-HRP/TMB detection). VLR-rFc was used as a titrant (rabbit Fc-HRP/TMB detection) in wells coated with CPs. Data represent averages (n=3) with standard error. All plates were analyzed at 450 nm. The absorbance readings were fit using a four-parameter logistic regression.

4.8 Supplement

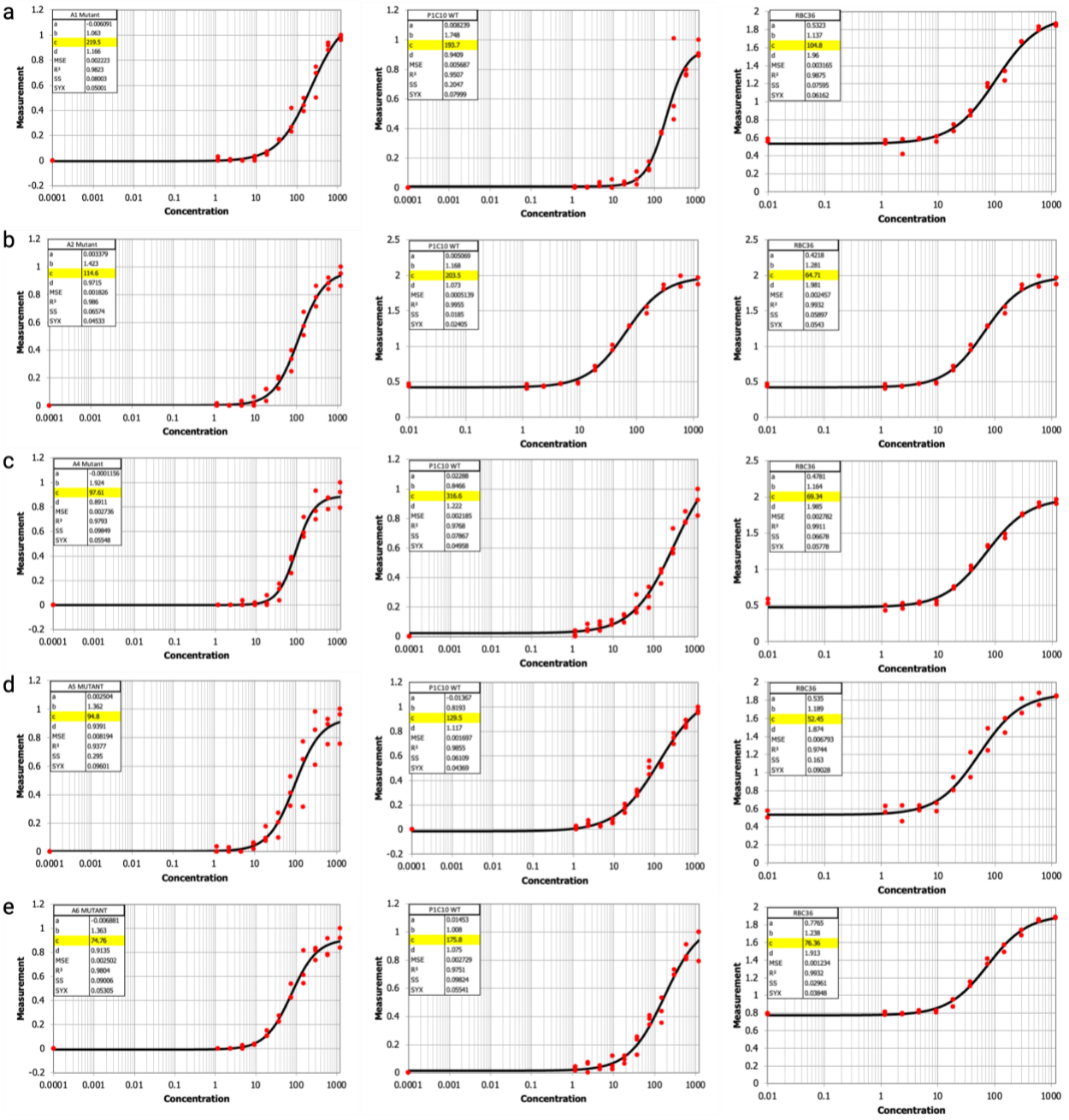


Figure 4S-1 Four Parameter Logistic Fits of Alanine Scan ELISAs. The 4PL curves were fit and parameters generated using MyAssays.com. The K_D values are highlighted in yellow. Each Alanine mutant was run on its own plate along with P1C10 WT and RBC36. All dilutions were run in

triplicate. (a) Alanine mutant A1 (Figure 4-4 a) (b) Alanine mutant A2 (Figure 4-4 b) (c) Alanine mutant A4 (Figure 4-4 c) (d) Alanine mutant A5 (Figure 4-4 d) (e) Alanine mutant A6 (Figure 4-4 e). All concentrations in are in nM. The 'Measurement' axis represents the raw data from the plate reader at 450 nm absorbance.

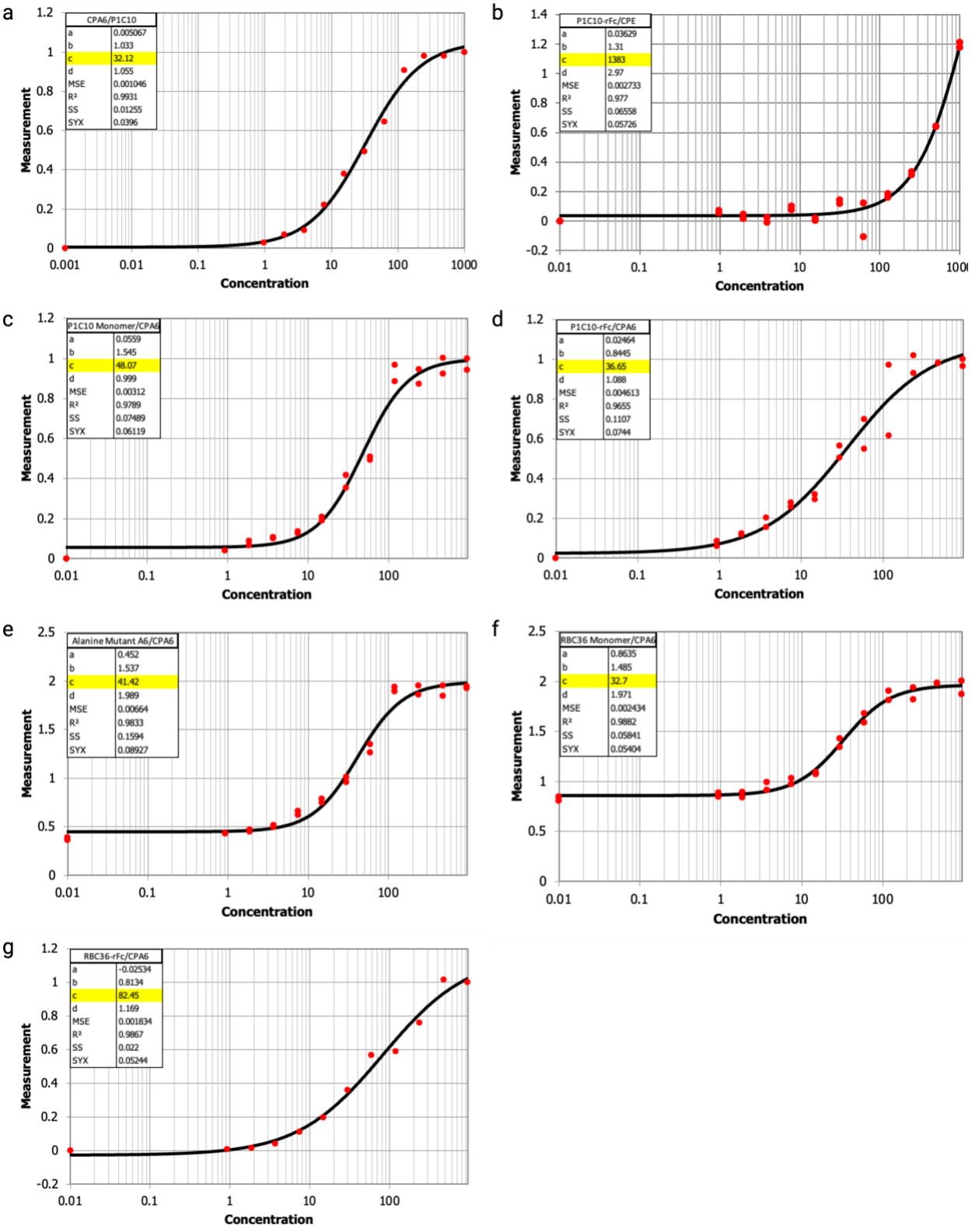


Figure 4S-2 Four Parameter Logistic Fits of P1C10 and RBC36 constructs with Carboxypeptidase

A6 ELISAs. The 4PL curves were fit and parameters generated using MyAssays.com. The K_D values are highlighted in yellow. (a) CPA6/P1C10 monomer (Figure 4-8 a) (b) CPE/P1C10-rFc (Figure 4-8 b) (c) CPA6/P1C10 monomer (Figure 4-8 c) (d) CPA6/P1C10-rFc (Figure 4-8 d) (e) CPA6/P1C10 Alanine Mutant A6 (Figure 4-8 c) (f) CPA6/RBC36 monomer (Figure 4-8 c) (g) CPA6/RBC36-rFc (Figure 4-8 c). All concentrations in are in units of nanomolar. The 'Measurement' axis represents the raw data from the plate reader at 450 nm absorbance.

CHAPTER 5 – Yeast Surface Display Screening against Non-traditional Cellular Antigens

This chapter was adapted from our contribution to "Protein Engineering and High-Throughput Screening by Yeast Surface Display: Survey of Current Methods" published in *Small Science.*[133] This chapter was completed with the help of Professor Shusta (edits and revisions) and Professor Michael Nash who invited us to contribute to the publication.

5.1 Abstract

Yeast surface display has become a powerful tool for high-throughput expression and screening of proteins. Its ability to display proteins with mammalian post-translational modifications without the need for purification and downstream processing has made yeast surface display a chosen method for large libraries of unique classes of proteins with biomedical applications. As biomedical disease targets become better understood, the antigens with which we screen large libraries of proteins have become more complex. In this review we discuss the advantages of yeast surface display for screening non-traditional antigens as it pertains to our use with variable lymphocyte receptors. Among the methods discussed is the brain extracellular matrix screen of the yeast surface display library that produced variable lymphocyte receptor P1C10.

5.2 Introduction

Yeast Surface Display (YSD) is a high-throughput method for expression and characterization of proteins. In YSD platforms *Saccharomyces cerevisiae* expresses a recombinant protein that is tethered to the cell wall via a disulfide linkage between a cell wall protein and its binding partner fused to the recombinant protein. The cell wall protein, often Aga1p, is incorporated in the mannoprotein layer of the cell wall via a glycosylphosphatidylinositol (GPI) anchor, and when bound to its binding partner, Aga2p, provides a stable display system for the recombinant protein fusion [134]. The tethered platform for protein expression in YSD has several advantages.

YSD provides a high-throughput method for expressing a broad range of different proteins. Unlike other display platforms, *S. cerevisiae* natively accommodates mammalian post-translation modifications (PTMs) and has been shwen to stabilize proteins that are otherwise poorly soluble [135]. YSD has been used for G protein-coupled receptor (GPCRs) [136], class II major histocompatibility complex (MHCII) [137], and anti-CD3ε scFv [138] mammalian protein expression. In addition to mammalian proteins, YSD has been used to express other challenging proteins containing disulfide bonds including Variable Lymphocyte Receptors (VLRs) [48, 139, 140].

YSD also has been utilized for its ease of use in library generation [141, 142]. *S. cerevisiae* can be electroporated for efficient transformation of libraries with authors citing as much as 10^9 diversity [143]. Libraries can be further diversified in yeast with homologous recombination [134]. *S. cerevisiae* naturally can recombine homologous DNA such that mutated genes with overlapping DNA segments can be assembled *in vivo* by the cell. Homologous recombination

methods in YSD libraries have been used to impart binding specificity and increase binding affinity [134].

The methods by which YSD libraries are tested for positive binders has evolved with time. As disease pathologies and cellular targets become further elucidated, novel therapeutic antigens continue to be discovered and types of YSD screens become more nuanced. Newly discovered antigens from targeted cell types or microenvironmental targets are often not soluble proteins for which binding proteins can be identified and engineered using standard yeast and phage display techniques. As a result, screens against such non-traditional cellular antigens using YSD are becoming a more prominent approach to developing new therapeutic binding moieties. Here, we discuss YSD approaches using cell lysates and extracellular matrices as antigen sources for lead binding protein identification and engineering (Figure 5-1).

5.3 Cell Lysates

YSD libraries are often used to identify therapeutically relevant binding proteins. For such applications, it is beneficial to screen binding protein libraries against targets that are presented in a form that best mimics their native fold and environment. As opposed to recombinant antigen screens, detergent-solubilized cell lysates using non-denaturing detergents can offer the advantage of presenting target antigens in a native or near native form in the background of the complete repertoire of antigens. Several studies established the usefulness of cell lysates in YSD antibody screens and affinity maturation [144-146]. This collection of techniques is broadly known as yeast display immunoprecipitation (YDIP), and here we will focus on the most recent adaptations of these methods.

Recent work demonstrated the utility of YDIP screening using an *in vivo*-sourced cell lysate antigen pool which holds the promise of increased *in vivo* relevance of the output binding pool [140]. A YSD library of variable lymphocyte receptors (VLRs) was generated from lamprey that were immunized with brain endothelial cell plasma membranes directly fractioned from mouse brain microvessels [140]. To enrich the library for brain endothelial cell membrane targeted VLRs, YDIP performed using an antigen pool comprising detergent-solubilized, biotinylated lysates of the same mouse brain microvessel plasma membranes that were used for lamprey immunization [140]. YDIP with one round of magnetic activated cell sorting (MACS) followed by one round of fluorescence activated cell sorting (FACS) leading to a ten-fold enrichment in yeast-displayed VLRs capable of binding brain microvessel plasma membrane compared to the starting library [140]. Subsequent biopanning of the library for VLRs that could bind extracellular epitopes was

performed by biopanning against mouse brain endothelial cell cultures [140]. While biopanning for enrichment of VLRs binding to extracellular targets was performed using cell cultured cell substrates rather than *in vivo*- sourced cell lysates, the initial enrichment of the library using the brain endothelial plasma membranes ensured that the resulting pool of VLRs had many VLRs (16 of 26 clones tested) that could bind in vivo-relevant antigens in brain microvessels. Therefore, the naturally occurring protein expression profiles in lysates can offer a more physiologically relevant array of therapeutic targets for YSD screens of challenging cell types.

Another recent study directly compared YSD screens using a recombinant antigen versus cell lysate [147]. A YSD library was generated using paired scFvs from immunization of the Trianni humanized mouse with recombinant OX40 [147]. The subsequent YSD library was screened by flow cytometry with either recombinant OX40 or OX40-containing cell lysate [147]. The scFvs identified as positive OX40 binders were reformatted as mAbs for subsequent binding characterization to assess therapeutic potential. The mAbs produced from scFvs identified in the cell lysate screen all bound cell surface OX40 in a cell-based binding assay [147]. However, of the scFvs identified in the recombinant OX40 screen, only 39% of the reformatted mAbs bound cell surface OX40 in the same cell-based binding assay [147]. The 61% false positive rate identified in the recombinant OX40 screen was attributed to the monomeric form of OX40 that differs from physiological OX40 which exists as a trimer [147]. Multimeric differences between recombinant and native protein states for certain antigens further support the benefits for lysate screening.
While lysate-sourced antigens have proven useful for many YSD library screens, lysates are not efficient screening strategies for all antigens as demonstrated in another recent study [148].

Mutagenic libraries of a fibronectin domain and an affibody were expressed in YSD and screened for binders to CD276 or Thy1 [148]. Screen workflows for the YSD libraries included an initial step of magnetic bead sorting with recombinant extracellular domains followed by i) FACS with recombinant extracellular domains, (ii) FACS with detergent solubilized cell lysate or (iii) cellular panning [148]. For the CD276 antigen, FACS with recombinant extracellular domains did not produce fibronectin or affibody binders [148]. Using CD276-containing cellular lysates also did not enrich the libraries for specific binders [148]. However, when cell panning screens were included, CD276-binding proteins in both the fibronectin and affibody YSD libraries were identified [148]. A possible explanation for not observing lysate enrichment in this instance is that the initial FACS enrichment against recombinant epitopes does not share the same conformations as epitopes found in the lysate, as was observed by the OX40 study. In instances when lysate enrichment is minimal, it is possible to augment YSD libraries through lysate depletion strategies if an appropriate cell line is available. Previous work showed that cell lines absent of the library target can be used in lysate form as a polyspecificty reagent for library depletion and to measure the cross-reactivity of candidate clones discovered in lysate enrichment screens [149]. Such studies demonstrate the need for employing a multi-pronged approach to screening YSD libraries depending on the desired therapeutic target.

5.4 Membrane Proteins

Membrane proteins are a particularly challenging set of targets for the identification of binding proteins as they are poorly soluble and recombinant ectodomains are commonly misfolded or do not result in productive identification of binding proteins. As noted in the previous section, one strategy is to deploy cell lysates in combination with YSD screening. To address this challenge, recent studies have provided alternate methods for screening YSD libraries specifically towards membrane protein targets.

One such study using YDIP leveraged the fact that detergent solubilized complexes can preserve protein-protein interactions much like classic immunoprecipitation procedures. Hence, one can screen for binding protein interactions with cell surface receptors that are in turn associated with key cellular machinery or processes. For example, such “functional YDIP” (fYDIP) was used to enrich for binders to blood-brain barrier membrane protein complexes involved in endocytocytic processes [139]. In this study, brain microvessels were isolated from bovine or rat brains and the membrane protein fraction biotinylated [139]. First, a YSD library of nonimmune scFvs was screened for a pool of binders to the detergent-solubilized, biotinylated membrane proteins using MACS [139]. The second round screen employed a FACS protocol to isolate scFv that bound to biotinylated membrane protein complexes that contained AP-2, a key adaptor protein in clathrin-mediated endocytosis [139]. A total of 31 unique clones were identified from the fYDIP screen. Of the 31, 26 were confirmed to bind proteins in detergent-solubilized blood-brain barrier membrane protein preparations, and 22 of the binding scFvs were confirmed to bind a

membrane protein complex that contained AP-2 [139]. For validation, 10 of the clones were produced as soluble scFv-Fc fusions, and all 10 clones demonstrated positive staining of blood-brain barrier membranes in rat brain tissue sections, indicating the *in vivo* relevance of such a screen [139].

G protein-coupled receptors (GPCRs) are another class of membrane proteins that are notoriously difficult to process through binding protein identification campaigns [150]. This next study deployed a variation of YSD biopanning where the yeast displaying mutagenic YSD antibody library was enriched for antibodies with improved binding to the mu opioid GPCR [150]. In a bit more detail, Mu opioid GPCRs were overexpressed in mammalian cells that were then incubated with the mutagenic YSD antibody library [150]. Yeast cell (green reporter)- Mu opioid expressing mammalian cell (red reporter) conjugates were then isolated by dual color FACS [150]. Following four rounds of selection for the highest YSD binding population, 30 clones were chosen for binding the mu opioid enriched cells [150]. Two lead clones were shown to bind the GPCR with affinities of 76 nM and 280 nM [150]. Subsequently, a kinetic dissociation competition screen using an excess of mu opioid receptor expressing cells (yellow) followed by FACS to identify the persisting YSD (green)-mammalian cell (red) conjugates was successful in further maturing the antibody affinity. The results from the cell-based kinetic screen were similar to those when a traditional YSD kinetic dissociation screen was performed with mu opioid receptor ectodomain.

Another approach for screening YSD libraries for membrane protein binders was developed by using YSD for presentation of both the binder library and the cognate membrane protein antigen [151]. In this case, the membrane proteins, ectodomains or fragments are co-expressed in YSD format along with an iron oxide-binding protein [151]. When this population of yeast is incubated with iron-oxide nanoparticles, the yeast cells are magnetized for facile magnetic sorting [151]. The binder library YSD pool is then contacted with the target protein-expressing yeast population, and binder-target YSD complexes isolated by MACS. The method was validated using the two targets, the cytosolic domain of the mitochondrial membrane protein TOM22 and the extracellular domain of the c-Kit receptor [151]. YSD libraries based on the Sso7d binding scaffold and nanobody scaffolds were screened against the target expressing yeast resulting in lead clones binding their respective membrane proteins on the order of nanomolar affinity [151]. Taken together, difficult targets like membrane proteins can be amenable to YSD techniques when the antigens are presented as detergent-solubilized lysates, overexpressing cell lines or if they themselves are expressed in YSD formats.

5.5 Extracellular Matrices

In addition to cell lysates or membrane proteins, the extracellular matrix (ECM) is a growing therapeutic target that can be explored using YSD screens. Target cells, including tumor cells, secrete ECM proteins that can make for highly specific drug delivery targets. Several studies have demonstrated the advantage of therapeutic targeting of the ECM [104, 152]. Thus, using YSD libraries to identify and engineer binders to ECM offers another application of YSD.

For example, YSD has been used to identify biners to brain ECM for the targeting of glioblastoma tumors [48]. In this study, a variable lymphocyte receptor (VLR) library was generated from lamprey immunized with murine brain microvessel plasma membranes and their associated brain ECM [48]. The resultant YSD library was enriched by two rounds of biopanning on decellularized mouse brain endothelial cell ECM [48]. Following the library enrichment, 285 VLR clones from the YSD library were screened using a high throughput method where candidate VLRs were released from the surface of individual YSD clones and subjected to a comparative binding ELISA against brain ECM or fibroblast ECM. Ten of the 285 clones showed a minimum 2.5-fold preference for brain ECM binding [48]. The lead clone for selective binding to brain ECM, P1C10, was produced in mammalian cells, and demonstrated selective binding at the site of pathologically-exposed brain ECM in mouse glioblastoma models [48]. When conjugated to chemotherapy, P1C10 resulted in increased survival in murine glioblastoma survival studies [48].

Another study demonstrated the utility of using YSD to identify improved nanobodies that target ECM for tumor therapy [153]. In this study, soluble, biotinylated EIIIB domain of fibronectin, a known tumor-specific ECM protein, was used as the antigen to enrich a YSD library of mutagenic nanobodies using standard YSD protocols [153]. The resultant library produced nanobody variants with picomolar affinity for EIIIB [153]. The high affinity nanobody clones were conjugated to immunocytokines and intratumoral administration resulted in improved survival in murine tumor models [153].

5.6 Conclusion

YSD provides a facile method for high-throughput expression of proteins with biomedical relevance. YSD can express complex proteins with mammalian PTMs that other display systems cannot express, and the tethered fusion to the cell surface stabilizes proteins that are otherwise poorly soluble. The natural ability of *S. cerevisiae* to recombine homologous DNA lends itself to diverse library generation on the order of 10^9. Taken together, these attributes make YSD a robust, high-throughput screening platform for expressing proteins with even the most complex of targets. As highlighted in this review, YSD can be used for many non-traditional antigen targets, and the list continues to grow to meet clinical indications.

5.7 Figures

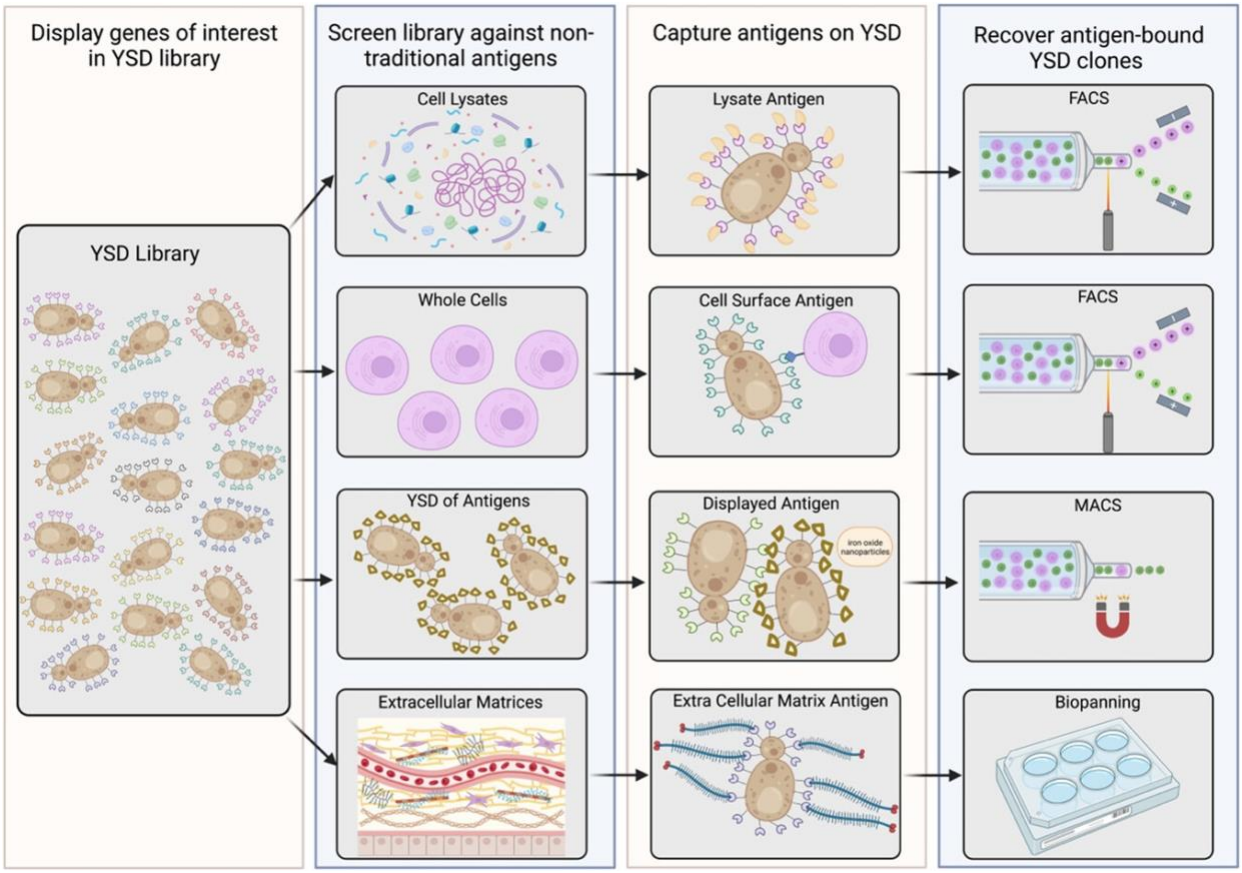


Figure 5-1 Using nontraditional antigen sources and screens to identify binding proteins. YSD can be leveraged in screens using cell lysates, whole cells, YSD-displayed antigens, and extracellular matrices to identify and engineer binding proteins.

CHAPTER 6 – Variable Lymphocyte Receptor-like Mutagenic Yeast Surface Display Library

In this chapter a human-derived protein structurally similar to variable lymphocyte receptor P1C10 is investigated as a variable lymphocyte receptor-like scaffold protein. This chapter was completed with the help of Seth Gehrke. Seth helped with library design and performed much of the lab work necessary for this project including Gibson assembly, plasmid preps, yeast electroporation, yeast culture and flow cytometry.

6.1 Abstract

Variable lymphocyte receptor P1C10, previously identified in the Shusta lab, has exciting potential therapeutic properties as a drug delivery moiety capable of targeting the brain extracellular matrix in instances of pathologically-disrupted blood-brain barrier. However, the potential immunogenicity of P1C10 if dosed in humans is of concern because the variable lymphocyte receptor was derived from lamprey. The desire to create a humanized version of P1C10 motivated the generation of a variable lymphocyte receptor-like scaffold library derived from humans to potentially mimic the binding properties of variable lymphocyte receptors while minimizing immunogenicity. SLIT2 was chosen as a human-derived protein alternative for its homology to P1C10 and ability to act as a variable lymphocyte receptor-like scaffold protein. A yeast surface display library of mutant SLIT2 was generated to sort for SLIT2 mutants with variable lymphocyte receptor-like binding properties demonstrating the potential of a scaffold library. The SLIT2 library demonstrates the potential to engineer variable lymphocyte receptor-like scaffold proteins.

6.2 Introduction

Previously the Shusta lab published exciting data regarding the drug-delivery potential of lamprey-derived variable lymphocyte receptor (VLR) P1C10 [48]. In glioblastoma (GB) mouse models, P1C10 had selective distribution at the site of the tumor [48]. Furthermore, when P1C10 was conjugated with doxorubicin-loaded liposomes and dosed to GB-bearing mice, the animals had a statistically improved survival time compared to the control cohorts [48]. Thus, P1C10 has been pursued as a potential drug delivery vehicle for diseases of the central nervous system (CNS).

P1C10 owes its unique binding properties at the extra-cellular matrix (ECM) of the brain to its native host species. P1C10 is a VLR that was derived from lamprey. Because lamprey diverged from humans > 500 million years ago, the lamprey immune system can produce antigen receptors (VLRs) against human-derived antigens [154]. Indeed, when lamprey are immunized with human antigens or proximal-to-human antigens, the lamprey produce a robust immune response with a full repertoire of antigen-binding VLRs [155]. As was the case with VLR P1C10, lamprey immunized with mouse brain microvessel plasma membranes produced a library of VLRs capable of binding the ECM of the brain, from which P1C10 was identified as a top binder of the ECM [48].

Despite P1C10's unique binding specificity, P1C10 has yet to be dosed in humans. To date, no VLRs have had first-in-human studies. This is largely due to the potential immunogenicity of VLRs in human. While the phylogenic divergence of lamprey from human is exploited for unique antigen-binding receptors, the same divergence is what could be responsible for VLR

immunogenicity in humans. Similar to how human-derived antigens produce a robust immune response in lamprey, it is conceivable that a lamprey-derived VLR could produce an immune response in humans.

Huma immune responses to therapeutics are well-documented [156]. The extensive history of mAb therapeutic development indicates that most non-human-derived antibodies are highly immunogenic in humans [157]. Non-human antibodies induce anti-idiotypic immune responses leading to anti-drug antibodies (ADAs) [158]. ADAs may have neutralizing effects on the therapeutic causing premature clearance and reducing the efficacy of the drug [159]. In more severe cases, ADAs may lead to inflammatory responses including cytokine storm and anaphylaxis [160]. Other ADAs have been demonstrated to attack native hormones including granulocyte/macrophage colony-stimulating factor (GM-CSF), erythropoietin (EPO), and thrombopoietin/megakaryocyte-derived growth factor (TPO/MGDF) which have led to severe auto-immune disorders in patients [161, 162]. Thus, humanization to reduce immunogenicity is a critical step in therapeutic antibody engineering [163].

Humanization is an essential component of therapeutic protein engineering. The long history of murine-derived antibodies in the pharmaceutical industry has given rise to 45 FDA-approved, humanized therapeutics between 1998 and 2020 [164, 165]. A common approach of mAb humanization is complementary determining region (CDR) grafting [166-169]. CDRs consist of the paratopes, or antigen binding residues, that are responsible for epitope specificity. Thus, it is common in mAb humanization to mutate the CDRs from a non-human antibody onto a human protein with structurally similar CDR regions [170]. CDR grafting was first used in clinical

development in 1988 for Mabcampath, the anti-CD52 mAb for T and B cell cancers [171]. Recent humanization efforts continue to use CDR grafting as the initial step for humanizing mAbs in early stage research [166-168]. In 2020 alone 7 CDR-grafted, humanized mAbs were FDA-approved [164].

CDR grafting offers a potential parallel for VLR humanization. Like CDRs, leucine rich repeats (LRRs) of VLRs determine the antigen binding specificity of the VLRs [20, 59]. Therefore, it is reasonable to consider LRR grafting with structurally similar human scaffolds as a potential method for humanizing VLRs. The extensive history and pitfalls of humanizing mAbs via CDR grafting can guide the initial approach for humanizing VLRs. Of note, Vernier zone residues have presented a unique challenge [157, 172, 173]. Vernier zone residues are not antigen contact residues, but rather structural supports to the paratopes. Therefore, if the Vernier residues are not grafted with the CDRs, the CDRs will not be positioned to bind the epitopes. Indeed, failure to graft Vernier residues in humanized mAbs has resulted in complete loss of antigen binding in many humanization studies as reviewed by Pavlinkova, et al [174]. Previously, authors have suggested a trial-and-error approach to identify residues in the Vernier zone through a series of back-mutations in the humanized protein [175]. More recently, computational approaches have been used to identify the Vernier zones of proteins with homology modeling followed by iterative back-mutations in silico and comparison of native binding affinity with a known ligand in a docking protocol [176-178]. Such approaches can be adapted using VLR homology models and in silico mutations of LRR residues.

In the present study, the potential to humanize VLRs was investigated using a combination of mAb-inspired humanization methods. VLR P1C10 was first investigated for a structural homolog derived from *Homo sapiens* to act as a non-immunogenic scaffold protein to mutate with P1C10 LRR residues, thus minimizing the potential immunogenicity of P1C10 while imparting its binding specificity on a human homolog. SLIT2 Domain 2 (S2D2) was identified as a human LRR protein with structural similarity to P1C10. S2D2 is a human LRR protein native to the CNS with a known binding partner, Roundabout Homolog 1 (ROBO1) [179]. S2D2 has been solved in complex with ROBO1 (PDB ID: 2V9T), and the binding residues of S2D2 have been well-characterized [179]. Thus, it is known what residues in the LRRs of S2D2 are most important to native binding. Therefore, S2D2 was an ideal candidate to act a scaffold protein to receive the LRRs from P1C10. LRR-grafting of P1C10 residues onto binding residues of S2D2, in theory, could eliminate native binding and impart P1C10-like binding onto S2D2.

Unfortunately, the LRR-grafting technique produced an unstable S2D2 mutant protein that had negligible expression and display yeast surface display (YSD) when compared to wild type (WT) S2D2 and P1C10. This suggested that the LRR-grafting technique led to instability in the S2D2 protein. Without the known binding residues of P1C10, it was required to graft all P1C10 LRRs to cover the full binding pocket of the VLR. To find a combination of LRRs that maintain binding specificity while not impairing the stability of S2D2, a mutagenic library was developed.

The YSD library utilized S2D2 as a scaffold protein for mutations along the known binding residues. The theoretical library size was 7.9e15, with 10 mutation sites at the binding residues of WT S2D2. To determine the feasibility of humanizing VLRs through a mutagenic library

approach, the library must be sorted for binders to known VLR antigens. If a mutagenic S2D2 library clone is able to bind a VLR antigen, this suggests that humanizing VLRs is possible through a library design in which a human protein such as S2D2 is mutated at its binding residues to create an S2D2 mutant that is primarily human with VLR-like binding specificity.

The S2D2 scaffold library provides a preliminary proof of concept for creating a VLR-like mutagenic library. In theory, an S2D2 library in YSD can be enriched for binders to known VLR antigens. Humanization of P1C10 would require an S2D2 scaffold library to be enriched for binders to the brain ECM. More specifically, the YSD library would require enrichment against the specific antigen target at the brain ECM, which at this time remains unknown. A candidate clone from a YSD S2D2 library could be selected for binding specificity, and downstream analysis would be required to confirm humanization. Such analysis would include binding studies and GB models that mimic the initial studies presented for VLR P1C10 [48]. Additional immunogenicity studies would be required to confirm the “humanized P1C10” does not illicit an immune response.

In sum, the mutagenic S2D2 library provides early insight into the potential of a VLR-like scaffold library and its uses. Our design suggests the possibility to engineer VLR-like proteins derived from humans. Further work is required to see the full potential of such a library.

6.3 Methods and Methods

6.3.1 Protein Data Bank search for scaffold protein

The Dali Server[180] was used to compare the P1C10 structure with the closest structural homologs published in the Protein Data Bank. The P1C10 crystal structure (PDB ID: 9CJ0) was used for the data input into the Dali server. All structural homologs were reviewed for percent homology to P1C10. Proteins with homology to P1C10 were further reviewed for their native host organism with preference given to proteins derived from *Homo sapiens*.

6.3.2 Yeast Surface Display of SLIT2 Domain 2

S2D2 was cloned into the pCT plasmid using NhEI (Thermo Scientific #FD0974) and BamHI (Thermo Scientific #FD0054) restriction enzyme cloning methods. The plasmid sequence was confirmed via Sanger Sequencing at the UW – Madison Biotechnology Center. The pCT plasmid containing S2D2 (pCT-S2D2) was transformed into AWY101 *Saccharomyces cerevisiae* using the previously published lithium acetate transformation protocol [181]. The transformed AWY101 was cultured at 30 °C with 250 RPM shaking in synthetic defined (SD) media deficient in uracil, leucine and tryptophan for selection according to previously published methods [182]. Display was induced using the galactose promoter and incubation temperature was reduced to 20 °C. After 20 hours of induction, the yeast was harvested and labeled for flow cytometry. Labels for cytometry included mouse anti-myc clone 9E10 (Invitrogen #MA1-980-1MG), goat anti-mouse Alexa 488 (Invitrogen #A11001), rabbit anti-HA clone RM305 (Invitrogen #MA5-27915), and goat anti-rabbit Alexa 647 (Invitrogen #A21244). The S2D2 sequence is the following:

LHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGAFSPYKKLRRIDLSNNQISELAPDAFQGL RSLNSLVLYGNKITELPKSLFEGLFSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQT MHLAQNPFICDCHLKWLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSA (UniProt ID: O94813).

6.3.3 Slit2 Domain 2 Mutant in Yeast Surface Display

A mutated S2D2 gBlock gene fragment was purchased from IDT. Mutations in the gBlock were made at the surface residues in the LRRs. The WT surface residues of the LRRs were mutated to the surface residues of P1C10. The following represent P1C10 surface LRR residues grafted on the positions of the S2D2 surface LRR mutations: R330T, D332K, S334H , N335S, S354I, V356R, G359S, L378K, L380Y, N382S, A383W, L402H, Y406G, D407Y, T426H, H428W, A430L, Q431N. The gene was cloned into the pCT plasmid using restriction digest cloning techniques as described above. The pCT plasmid with the S2D2 mutant was transformed into AWY101 and induced according to the previously described method. The S2D2 mutant was labeled for full-length display with C-Myc epitope and analyzed by flow cytometry. The S2D2 mutant sequence (mutations in red) was the following: LHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPGAFSPYKKLR**T**I**K**L**HS**NQISELAPDAFQGL RSLN**I**L**R**LY**S**NKITELPKSLFEGLFSLQ**K**L**Y**L**SW**NKINCLRVDAFQDLHNLN**H**LSL**GY**NKLQTIAKGTFSPLRAIQ **H**M**W**L**LN**NPFICDCHLKWLADYLHTNPIETSGARCTSPRRLANKRIGQIKSKKFRCSA

6.3.4 SLIT2 Domain 2 Mutant Library Design

Gene fragments were designed using the WT pCT-S2D2 plasmid and fragments purchased with IDT Duplex Ultramar technology. In brief, the pCT-S2D2 was digested at BamHI and CsiI (within

the S2D2 gene) restriction sites. The cut DNA was run on a 1% Agarose gel and purified using the Zymo Clean Large Fragment DNA Recovery Kit (Zymo #D4046) and stored with Antarctic Phosphatase. The cut plasmid was used as the backbone for the library. The Duplex Ultramers were ordered as 3 fragments to span the length of S2D2 between BamHI and CsiI. All fragments were designed with 20 bp of overhang. The fragments were designed with hydrophobic mutations (NTK codon) at VAL354, ALA381, LEU376, LEU378, LEU400, TYR404, TYR356. ARG328, SER402, HIS426, SER453 allowed all other mutations (NVK codons).

6.3.5 HiFi Assembly of Gene Fragments

The DNA was assembled using the NEB HiFi DNA Assembly Cloning Kit (NEB #E5520S). The assembly reaction followed the manufacturer protocol. The reaction included the cut pCT-S2D2 backbone and the three IDT Duplex Ultramers (Fragments 1, 2 and 3). Including the backbone, 4 fragments were used for the assembly. Thus, the protocol for 4-fragments was used, and a molar ratio of 1:1 was used for the fragments and cut backbone. Following the assembly reaction, NEBuilder® HiFi DNA Assembly Chemical Transformation Protocol (NEB #E5520S) was used for transformation of the included chemically competent cells. After the outgrowth, 100 µl of cells was streaked on an LB Ampicillin agar plate to determine colony growth. The remaining cells were maintained in 250 ml LB at 37 °C 250 RPM overnight, and the plasmid was recovered from the cells using the ZymoPure Plasmid MaxiPrep Kit (Zymo #D4203). The pool of recovered plasmid along with 16 colonies from the plate were sent to Plasmidsaurus for sequencing.

6.4 Results

6.4.1 Protein Databank Search for Scaffold Protein identifies SLIT2 Domain 2 as human homolog to Variable Lymphocyte Receptors

The Dali Server[180] was used to identify structural homologs to P1C10 to act as potential grafting scaffolds for humanization. Homologs were prioritized for being structurally similar to P1C10 such that the putative binding residues in the surface LRRs of P1C10 could be grafted onto corresponding LRR sites of the homolog. The homologs were further prioritized for derivatives from *Homo sapiens* to address the concerns of P1C10 immunogenicity in humans. The search identified SLIT2 domain 2 (SLIT2D2) (PDB ID: 2V9T) [179] as the human protein having the highest homology to P1C10 (Figure 6-1). SLIT2D2 is a LRR protein like P1C10. Like P1C10, S2D2 has an LRRNT, LRR1, LRRVs, CP, and LRRCT. S2D2 has 2 more LRRVs than P1C10.

6.4.2 SLIT2 Domain 2 can be expressed and displayed in Yeast Surface Display

SLIT2D2 was tested for display in YSD to confirm compatibility for high-throughput screening. VLR RBC36 was used as a reference control to compare the percent of the induced population. HA labeling was used to quantify induction. C-myc labeling was used to quantify full-length display of the protein. S2D2 produced a 74.5 ± 4% population of both induced protein and full-length display indicating that S2D2 was stable in the YSD format (Figure 6-2 a,b,c) with 100% of the expressed protein being stable for full-length display. Furthermore, the S2D2 display population was comparable to the RBC36 display population (76.8 ± 4%) validating the use of YSD for S2D2 display (Figure 6-2 a,d,e).

6.4.3 SLIT2 Domain 2 Mutant engineered by Leucine Rich Repeat grafting does not display in Yeast Surface Display

To determine the feasibility of LRR-grafting, an S2D2 mutant was designed with point mutations along the WT S2D2 LRRS V1, V2, V3, Ve, and CP. Point mutations were made at the surface residues of the LRRs in the "XLXLXX" motif where the surface mutations are represented by the X. The positions were mutated to the P1C10 surface residues in the LRR motif representing surface residues in the LRRs from LRRNT, LRR1, LRRv1, LRRv3, and CP from P1C10 (Figure 6-3). This "LRR-grafting" technique was meant to mimic the CDR-grafting technique of IgGs. The surface residues mutated in S2D2 LRRs included: R328T, D330K, S332H , N333S, S352I, V354R, G357S, L376K, L378Y, N380S, A381W, L400H, Y404G, D405Y, T424H, H426W, A428L, Q429N.

The P1C10 LRR-grafted S2D2 mutant was tested for stability in YSD. WT P1C10 and WT S2D2 demonstrated expected induction populations, 65% and 67% respectively (Figure 6-4a,b) via flow cytometry analysis. However, induction was diminished to <5% in the mutant S2D2 indicating a lack of stability in the mutant S2D2 (Figure 6-4c). The inability to display an LRR-grafted S2D2 suggested LRR-grafting was not a feasible method. This finding motivated developing a mutant S2D2 library rather than relying on individual point mutations.

6.4.4 SLIT2 Domain 2 Mutant Library Design

The S2D2 mutant library was engineered to insert mutations at the known binding residues of S2D2. PDB ID 2V9T shows S2D2 in complex with t's native ligand, Roundabout homolog 1 (ROBO1). In this structure, VAL354, ALA381, LEU376, LEU378, LEU400, TYR404 and TYR356 are

identified as having hydrophobic interactions with ROBO1 [179]. ARG306, GLU304, ARG328, ARG287, SER402, HIS426 and SER453 are identified as having electrostatic interactions with ROBO1 [179]. Thus, these residues were mutated to impart diversity in the S2D2 library at the known binding positions of WT S2D2 (Figure 6-5).

To minimize the complexity of the gene assembly, ARG287, GLU304, ARG306, and SER453 were kept constant. This enabled the library to be engineered with 3 variable fragments as opposed to 5 (Figure 6-6). The resulting 3 fragments spanned S2D2 LRRv1, LRRv2, LRRv3, LRRve, and CP with mutations at the binding residues within the LRRs. VAL354, ALA381, LEU376, LEU378, LEU400, TYR404, and TYR356 bind ROBO1 with hydrophobic interactions, so mutations at these positions used the NTK codon to encompass hydrophobic residues. NTK codes hydrophobic residues F, L, M, I, V. ARG328 , SER402, HIS426 and SER453 bind with electrostatic interactions, so NVK codon was used at these positions to allow any non-hydrophobic residues to be inserted at these positions. NVK codes residues S, P, T, A, Y, H, Q, N, K, D, E, C, W, R, G. The theoretical library diversity was calculated according to the following equation: $(7^5)*(3^{15})$= 7.9e15. The calculation comes from 7 positions having 5 potential amino acids with NTK codons and 5 positions having 15 potential amino acids with NVK codons.

6.4.5 HiFi Assembly of Gene Fragments

The fragments were assembled into the cut pCT-S2D2 backbone and the pool from the liquid culture plasmid prep along with 16 unique colonies from the plate were sequenced for diversity. Theoretically, we would expect 7.9e15 unique sequences assembled in this library design. Unfortunately, all sequencing results have produced pCT-S2D2 WT with no mutants identified.

This indicated that the backbone has not been fully purified, and the fragments are not successfully assembling. Subsequently, the Duplex Ultramer gene fragments from IDT were sequenced. Sequence results of the Ultramers showed mixed purity with insertions and deletions at the site of the mixed base codons. Since this finding, we have started to explore alternative fragments for a library design.

6.5 Discussion

S2D2 is an exciting human homolog to P1C10. The human-derived protein offers potential to act as a scaffold protein to receive P1C10 binding residues and minimize the immunogenicity of WT P1C10. Because direct LRR-grafting of all the P1C10 LRRs resulted in diminished stability of the S2D2 mutant, there are two approaches that can be taken for alternative humanization strategies. First, P1C10 binding residues could be identified to enable more targeted mutations of the S2D2 scaffold. Currently, P1C10's binding partner is unknown, so we are unable to graft the specific binding residues of P1C10 onto the S2D2 scaffold. Secondly, a library could be designed to encompass a large subset of S2D2 mutants with mutations at the binding residues of WT S2D2 to impart binding diversity into WT S2D2. The second approach was taken in this study.

The first step in developing an S2D2 library was validating the YSD platform for expression and characterization. S2D2 demonstrated display properties on the same scale as that of VLR control proteins, indicating YSD was a suitable expression platform for S2D2. Following this confirmatory step, we sought to create a mutagenic library of S2D2 with mutations targeting the binding residues of S2D2 to impart binding diversity in our library.

The library design for S2D2 mutants was theoretically 7.9e15. In practice, we were unable to achieve this diversity. We speculate that the diversity was not achieved due to inefficiencies in the HiFi assembly method because of the poor purity of the Duplex Ultramer fragments that we received. An alternative approach to HiFi assembly could be PCR assembly of single strand fragments to avoid the need for the double-stranded Duplex Ultramers. Once the library has been successfully assembled, the plasmid can be electroporated into AWY101 for antigen testing by

FACS of the YSD library. The library could be screened against known VLR antigens. Positive binders to a VLR antigen will suggest that the VLR-like scaffold library can adopt binding specificity like that of VLRs. A VLR-like scaffold library could be used for possible efforts to humanize VLRs. Since the VLR-like scaffold is derived from the human S2D2, it is feasible to consider VLR antigen binders from the library as “humanized VLRs.” However, further validation of candidate clones from the VLR-like scaffold library would be required for a “humanized VLR” to be confirmed.

6.6 Figures

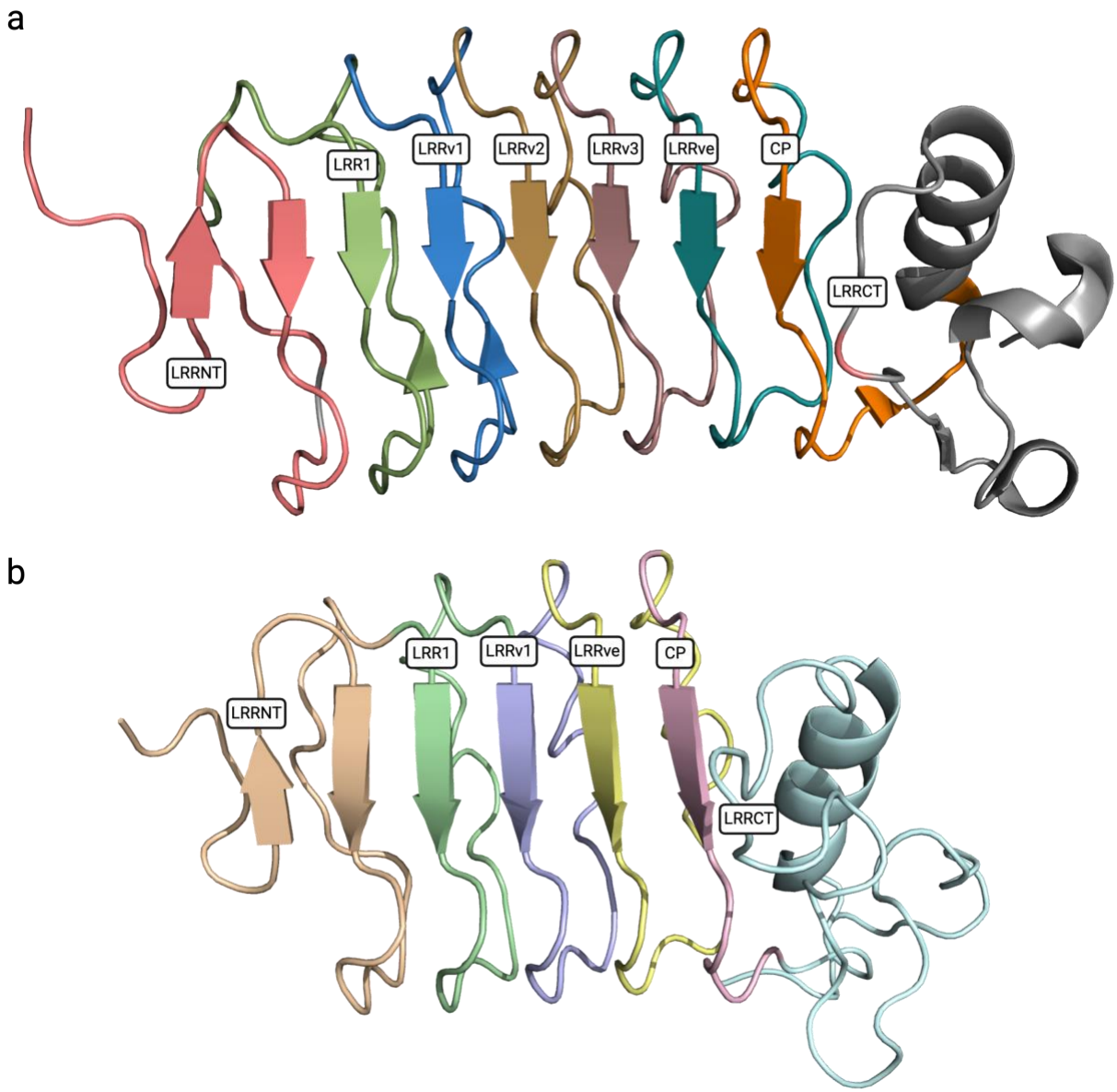


Figure 6-1 Structure comparison of P1C10 and SLIT2D2. (a) SLIT2 Domain 2 (PDB ID: 2V9T) was evaluated for LRR composition like P1C10. (b) P1C10 crystal structure (PDB ID: 9CJ0) highlighting the LRR composition.

a

Protein	Label	% Population Induced	Average	SD
Slit2-D2	HA	77.8%	74.5%	4%
		70.6%		
		75.2%		
	C-myc	79.4%	74.5%	5%
		69.6%		
		74.5%		
RBC36	HA	80.4%	77.5%	3%
		75.0%		
		77.0%		
	C-myc	80.4%	76.8%	4%
		72.5%		
		77.5%		

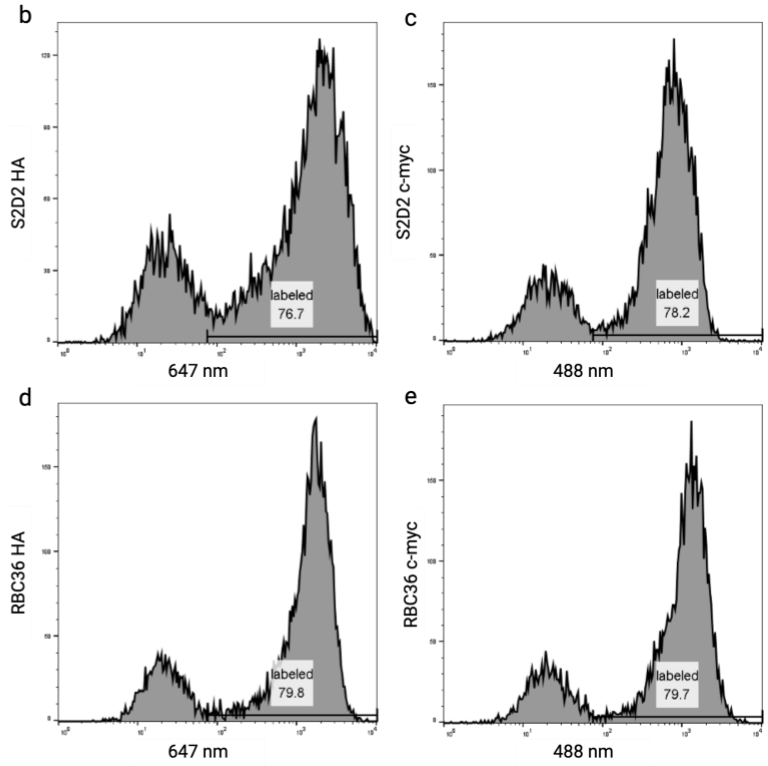


Figure 6-2 Yeast Surface Display Flow Cytometry of SLIT2 Domain 2 and RBC36. (a) The population of S2D2 and RBC36 was evaluated for display (HA) and full-length expression (C-myc). The average and standard deviation (SD) of the induced populations comes from biological triplicates (n = 3). (b) Example flow plot of the non-labeled and labeled populations of S2D2 HA. (c) Example flow plot of the non-labeled and labeled populations of S2D2 C-myc. (d) Example flow plot of the non-labeled and labeled populations of RBC36 HA. (d) Example flow plot of the non-labeled and labeled populations of RBC36 C-myc.

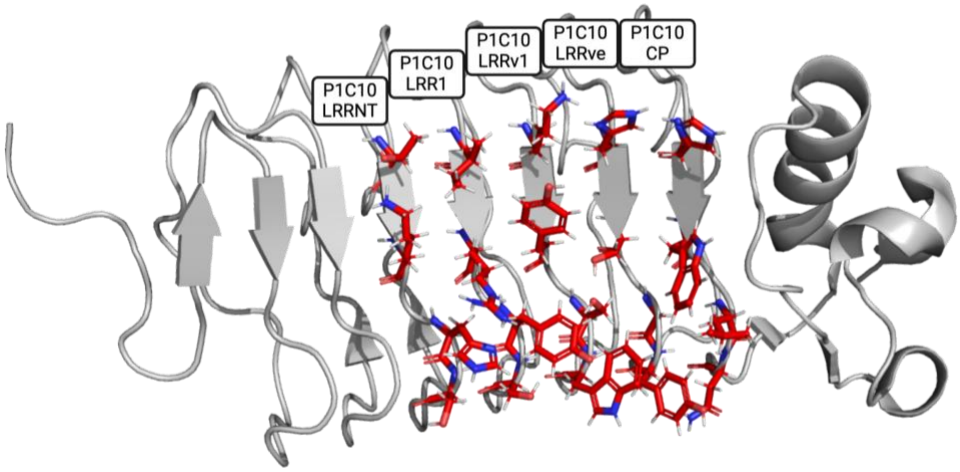


Figure 6-3 Model of SLIT2 Domain 2 with Leucine Rich Repeat mutations. The residues depicted as sticked color-coded by atom represent the mutations made to WT S2D2 to correlate with the LRR surface residues from P1C10 in those positions.

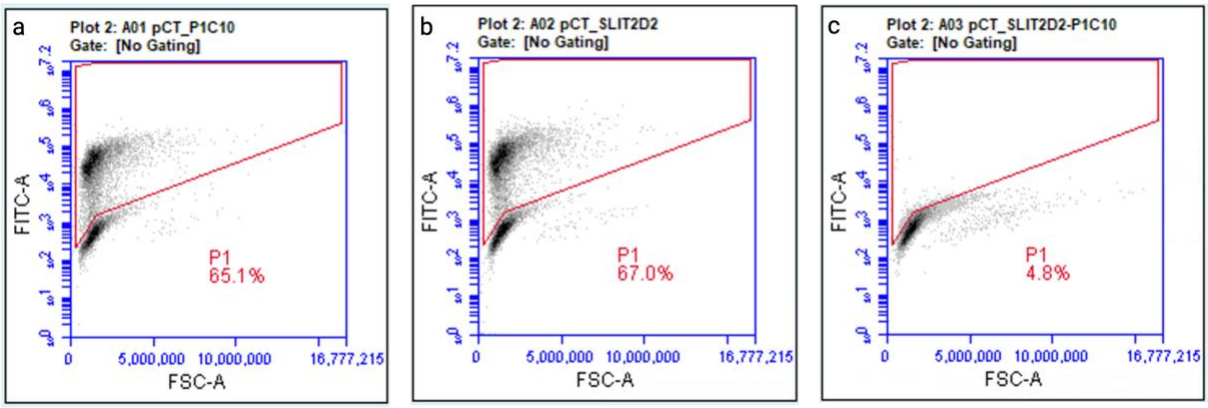


Figure 6-4 Yeast Surface Display Flow Cytometry of Leucine Rich Repeat-grafted SLIT2 Domain 2 mutant. (a) WT P1C10 is used as a positive control for full-length display labeled for the C-myc epitope. (b) WT S2D2 is used as a positive control for full length display labeling the C-myc epitope. (c) Full-length expression of the LRR-grafted S2D2 mutant is tested via the C-myc epitope.

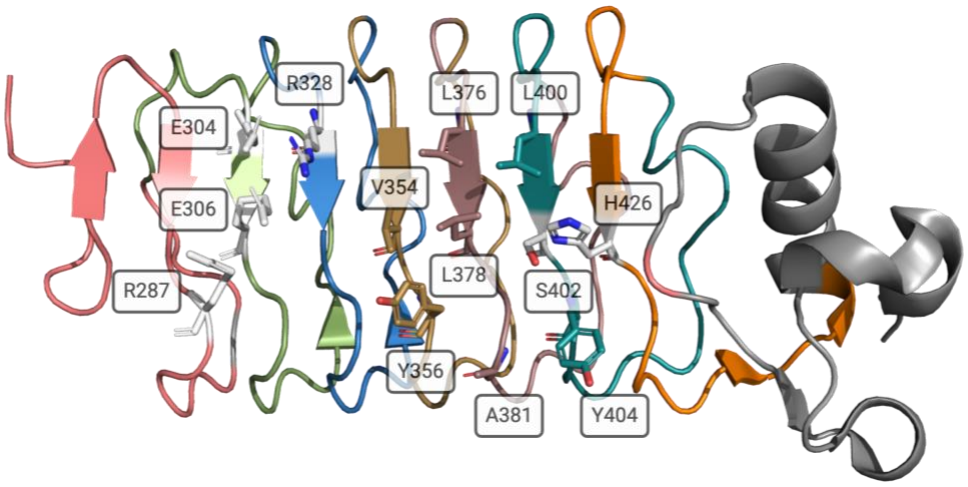


Figure 6-5 SLIT2 Domain 2 binding residues. The native binding residues of WT S2D2 were chosen as the locations to insert genetic diversity for the S2D2 library to change the binding specificity of the S2D2 mutants. The binding residues of S2D2 are highlighted as sticks. R287, E304, and E306 are highlighted in white because these are known binding residues of S2D2, buy they were kept constant for the library to minimize the number of fragments required for assembly.

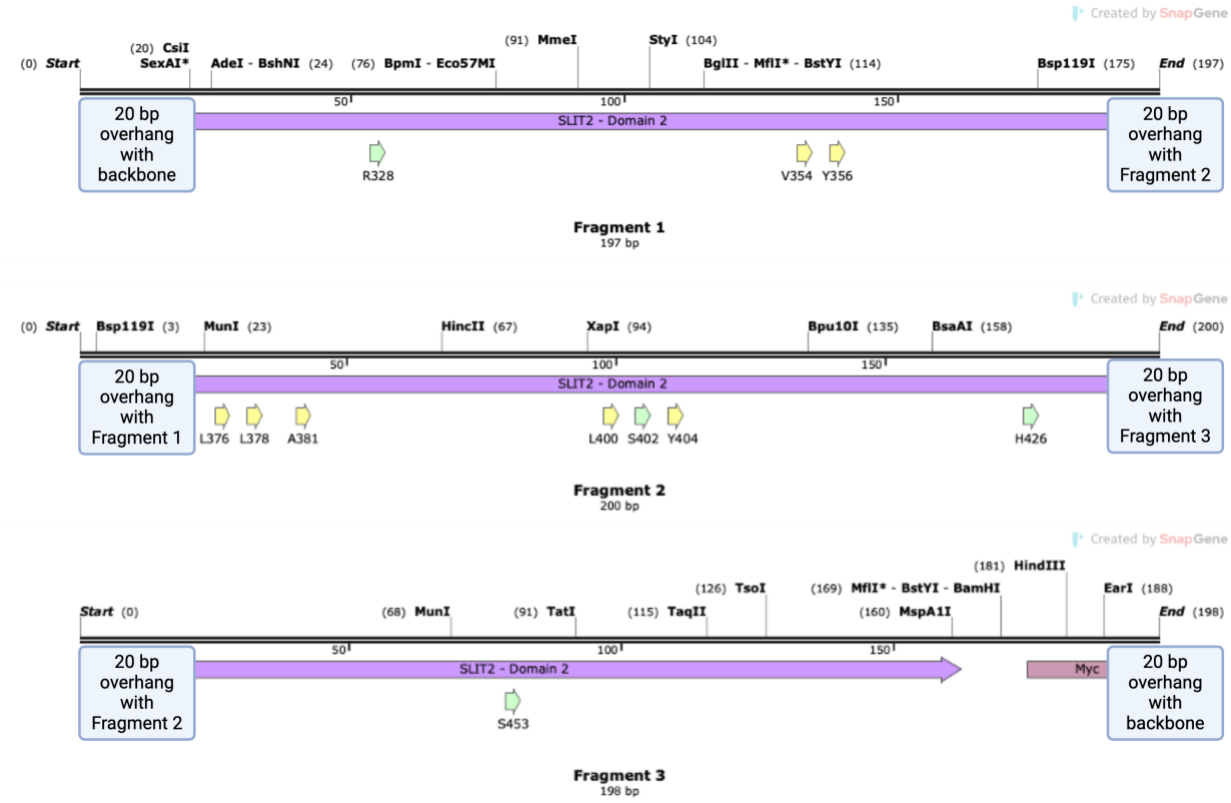


Figure 6-6 Ultramer Duplex fragment design for SLIT2 Domain 2 library. 3 gene fragments were purchased from IDT as Duplex Ultramers for library assembly. All were designed with 20 bp of overlap to meet the specifications of assembly. NTK mutations are annotated in yellow. NVK mutations are annotated with green. S453 is annotated because it is a native binding residue in WT S2D2. However, this position was left constant in the library. Thus, Fragment 3 was a constant fragment.

CHAPTER 7 – Best Wishes, Warmest Regards: Conclusions and Future Directions

7.1 Variable Lymphocyte Receptor P1C10

Variable lymphocyte receptors (VLRs) are an emerging class of antigen receptors with potential biomedical relevance. In the presence of immune challenge, the alternative adaptative immune system of jawless vertebrates, like lamprey, produces VLRs instead of traditional immunoglobulins. VLRs are biomedically relevant for their ability to target human-derived antigens that traditional immunoglobulins do not recognize because of self-tolerance. As a result of their unique binding specificity, VLRs have been explored as possible drug delivery moieties.

Notably, previous research in the Shusta lab identified VLR P1C10 as an exciting potential drug delivery vehicle [48]. P1C10 was identified from a library of VLRs generated from a lamprey immunized with mouse brain microvessel plasma membrane and enriched for extracellular matrix (ECM) binding. P1C10 has demonstrated binding to brain ECM which is especially relevant in diseases of pathologically-disrupted blood-brain barrier where the brain ECM is exposed at the site of insult, as is the case with glioblastoma (GB). In GB-bearing mouse models, P1C10 demonstrated selective labeling at the site of the tumor [48]. When conjugated to doxorubicin-loaded liposomes, P1C10 significantly increased the survival time of GB-bearing mice [48]. Thus, in the present work, the desire to understand the mechanism of action for P1C10 binding the brain ECM motivated further characterization of the VLR.

7.2 Producing Variable Lymphocyte Receptors in SHuffle *E. coli*

P1C10 production limited its downstream characterization. VLRs are challenging to produce at large scale due to their conserved 4 disulfide bonds. Historically, P1C10 was only produced at small scale in mammalian and yeast cells. Here, SHuffle *E. coli* was investigated and validated as a platform method to produce P1C10 among other VLRs. The high yield (as much as 38 mg/L), stability and confirmed activity of VLRs produced in SHuffle *E. coli* make the SHuffle method a preferred method for producing VLRs. The SHuffle production method enabled further downstream characterization of P1C10 such as crystallography that would not have been possible with the poor yield and instability of previously described production methods [48]. Furthermore, the SHuffle production method can be used for other VLRs such as VLRAs which are less soluble and perhaps as a result less studied than VLRBs. The SHuffle *E. coli* method can be used in future studies to produce challenging VLRs, like VLRAs, to better understand their structure and function and possible biomedical applications.

7.3 Structure of Variable Lymphocyte Receptor P1C10

The crystal structure of P1C10 was made possible with the SHuffle production method. Enough P1C10 was purified and concentrated (18 mg/ml) for a robust crystal screen that resulted in a high- resolution structure (1.3 Å). The structure of P1C10 can be used to guide future engineering and production initiatives involving P1C10. For example, should divalent P1C10 fusions be engineered, we can use the structure to guide appropriate linkages given the fold at the LRRCT. Furthermore, we can use this structure to evaluate how P1C10 might bind its binding partner at the ECM. At the time of this writing, the binding partner of P1C10 is unknown. If future work elucidates the P1C10 binding partner, binding can be modeled in AlphaFold Multimer with the solved crystal structure of P1C10 and either a solved structure or model of the binding partner to elucidate binding residues and validate the model with biochemical assays. The solved structure of P1C10 will improve any future models of P1C10 in complex with a binding partner.

7.4 Putative Binding Partners of Variable Lymphocyte Receptor P1C10

Previous work in the Shusta lab identified carboxypeptidases as a class of enzymes that bound P1C10. In the present work, we investigated carboxypeptidase binding to P1C10. While ELISAs presented positive binding between CPA6 and P1C10, no other carboxypeptidases showed positive binding in ELISA or any other assays including ITC, crystal screens, and SEC. The CPA6 was analyzed for purity by SEC and NativePage. The purity analysis demonstrated the commercially available CPA6 was a highly heterogenous aggregate. In further ELISA testing, CPA6 showed "positive binding" to non-binding controls, suggesting the CPA6-P1C10 binding was a false positive due to the CPA6 aggregated composition.

Future work should investigate the true binding partner of P1C10. It is recommended that the TRICEPS method, previously published [124], be adapted for P1C10 investigation. There are several potential pitfalls to this method that must be addressed in the assay design. Soluble ECM isolation for the screen must be confirmed. Many ECM proteins, including CPA6, are insoluble when removed from their cellular environment [183]. Thus, isolating the ECM but preventing insoluble aggregation is imperative to preventing false-binders in the pulldown screen. Previous authors report methods for isolating soluble ECM and confirming isolation [184]. The bEnd.3 cells should be transfected with a fluorescently-tagged ECM protein for confirmation of ECM, and the cells should be decellularized using ammonium hydroxide [184]. This method should provide a way to quantify the amount of ECM remaining post-decellularization. Furthermore, a non-binding control must be designed to remove false-positive peptides from the LC-MS/MS results. The non-binding control must have no binding to the brain ECM, or if it has background binding,

this should be considered when comparing relative abundance of peptides between the binding and non-binding TRICEPS molecules. Furthermore, the TRICEPS method requires soluble P1C10 to be coupled to a capture and detection moiety. The SHuffle-produced P1C10 should be used in this construct. The previous P1C10 constructs expressed in yeast and mammalian cells had poor solubility and were unable to be concentrated beyond 3 mg/ml. The improved stability of the SHuffle-produced P1C10 (concentrated up to 18 mg/ml) would make the SHuffle-produced construct more amenable to post-expression conjugations as is necessary in TRICEPS. If the previous P1C10-TRICEPS molecule had poor stability, this would have affected the ECM capture and subsequent LC-MS/MS analysis.

7.5 Yeast Surface Display of Non-traditional Antigens

A review of yeast surface display (YSD) methods for screening non-traditional antigens was presented. We discussed the use of YSD for screening cell lysates, membrane proteins, and ECM. The ECM screen described the method used to identify VLR P1C10. A YSD screen of decellularized ECM could be used for future libraries to find ECM-binders with drug delivery properties to the brain that are like VLRs but not derived from lamprey.

7.6 Variable Lymphocyte Receptor-like Mutagenic Library

Despite the exciting drug delivery potential of VLR P1C10, the potential immunogenicity of P1C10 in humans is of concern since the VLR was derived from lamprey. For this reason, methods to humanize P1C10 were explored. Without the known binding partner of P1C10, it was challenging to adopt a parallel method to CDR-grafting, so a mutagenic library was designed.

SLIT2 Domain 2 (S2D2) was chosen as a human-derived protein with the most structural homology to P1C10. A library was designed with mutations along the known binding residues of S2D2. The library design suggested the possibility of creating a human-derived scaffold protein that could adopt the binding specificity of VLRs. Unfortunately, poor DNA fragment purity resulted in the inability to assemble the library.

It's recommended that the library design stay the same, but the type of fragments change. Single-stranded fragments should be purchased and assembled via PCR assembly to eliminate the purity problems of the Duplex Ultramers. The PCR assembled product can be used to transform AWY101 via electroporation for YSD of the library. The library can be screened against known antigens of VLRs such as hen egg lysozyme or H-trisaccharide with rounds of FACS enrichment. The ability to find positive binders in a VLR-like mutagenic library capable of binding VLR antigens would suggest the ability to create recombinant VLR-like scaffold proteins. It would be possible to screen the library against decellularized ECM, as well, to attempt to find a humanized VLR-like

version of P1C10. To develop a true, humanized version of P1C10, the ECM binding partner must be identified and used for the library screen.

For any candidate clones identified as positive binders in the library screen, the soluble monomeric constructs should be produced in SHuffle for binding analysis in the appropriate binding assay provided the antigen. To make the claims that the VLR is “humanized,” the VLR-like protein should demonstrate the same binding specificity and affinity as the VLR, and it should also undergo a series of immunogenicity testing. Such assays could include PBMC immunogenicity screening [185] or a humanized mouse model [186].

7.7 Conclusion

In sum, we present a method for producing VLRs at large scale, a high-resolution structure of P1C10, and a theoretical library design for a VLR-like scaffold library. While we would have liked to see a solved crystal structure of P1C10 in complex with a confirmed binding partner, the work presented here makes VLR research more accessible to future researchers. The SHuffle production method enables easy and fast production of VLRs that can be scaled and concentrated with high stability. Such a method enables high-throughput screening of soluble, monomeric VLRs to be possible. Hopefully, with the methods presented here, further investigations of P1C10 and its putative binding partners are made more readily available to the next generation of VLR researchers in the Shusta lab. Good luck!

REFERENCES

1. Osório, J. and S. Rétaux, *The lamprey in evolutionary studies.* Development Genes and Evolution, 2008. **218**(5): p. 221-235.
2. Marlétaz, F., et al., *The hagfish genome and the evolution of vertebrates.* Nature, 2024. **627**(8005): p. 811-820.
3. Grillner, S. and B. Robertson, *The Basal Ganglia Over 500 Million Years.* Current Biology, 2016. **26**(20): p. R1088-R1100.
4. Pancer, Z., et al., *Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey.* Nature, 2004. **430**(6996): p. 174-180.
5. Amemiya, C.T., N.R. Saha, and A. Zapata, *Evolution and development of immunological structures in the lamprey.* Current Opinion in Immunology, 2007. **19**(5): p. 535-541.
6. Zapata, A. and C. Amemiya, *Phylogeny of lower vertebrates and their immunological structures.* Origin and evolution of the vertebrate immune system, 2000: p. 67-107.
7. Lee, S.-C., et al., *Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering.* Proceedings of the National Academy of Sciences, 2012. **109**(9): p. 3299-3304.
8. Das, S., et al., *Evolution of two distinct variable lymphocyte receptors in lampreys: VLRD and VLRE.* Cell Reports, 2023. **42**(8): p. 112933.
9. Kasamatsu, J., et al., *Identification of a third variable lymphocyte receptor in the lamprey.* Proceedings of the National Academy of Sciences, 2010. **107**(32): p. 14304-14308.
10. Li, J., et al., *Definition of a third VLR gene in hagfish.* Proceedings of the National Academy of Sciences, 2013. **110**(37): p. 15013-15018.
11. Rogozin, I.B., et al., *Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase.* Nature immunology, 2007. **8**(6): p. 647-656.
12. Boffa, G., et al., *Immunoglobulins and transferrin in marine lamprey sera.* Nature, 1967. **214**(5089): p. 700-702.
13. Fujii, T., H. Nakagawa, and S. Murakawa, *Immunity in lamprey I. Production of haemolytic and haemagglutinating antibody to sheep red blood cells in Japanese lampreys.* Developmental & Comparative Immunology, 1979. **3**: p. 441-451.
14. Pollara, B., et al., *The evolution of the immune response: VII. Antibody to human "O" cells and properties of the immunoglobulin in lamprey.* The Journal of Immunology, 1970. **105**(3): p. 738-745.
15. Hagen, M., M. Filosa, and J. Youson, *The immune response in adult sea lamprey (Petromyzon marinus L.): the effect of temperature.* Comparative Biochemistry and physiology. A, Comparative Physiology, 1985. **82**(1): p. 207-210.
16. Alder, M.N., et al., *Diversity and Function of Adaptive Immune Receptors in a Jawless Vertebrate.* Science, 2005. **310**(5756): p. 1970-1973.
17. Pancer, Z., et al., *Variable lymphocyte receptors in hagfish.* Proceedings of the National Academy of Sciences, 2005. **102**(26): p. 9224-9229.
18. Alberts B, J.A., Lewis J, et al., *Molecular Biology of the Cell. 4th edition*. 2002: Garland Science.
19. Waters, E.A. and E.V. Shusta, *The variable lymphocyte receptor as an antibody alternative.* Current Opinion in Biotechnology, 2018. **52**: p. 74-79.
20. Velikovsky, C.A., et al., *Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen.* Nature structural & molecular biology, 2009. **16**(7): p. 725-730.

21. Yu, C., et al., *Identification of human plasma cells with a lamprey monoclonal antibody.* JCI insight, 2016. **1**(3).
22. Herrin, B.R., et al., *Structure and specificity of lamprey monoclonal antibodies.* Proceedings of the National Academy of Sciences, 2008. **105**(6): p. 2040-2045.
23. Velásquez, A.C., et al., *Leucine-rich-repeat-containing variable lymphocyte receptors as modules to target plant-expressed proteins.* Plant Methods, 2017. **13**: p. 1-16.
24. Im, S.P., et al., *Investigation of variable lymphocyte receptors in the alternative adaptive immune response of hagfish.* Developmental & Comparative Immunology, 2016. **55**: p. 203-210.
25. Alder, M.N., et al., *Antibody responses of variable lymphocyte receptors in the lamprey.* Nature immunology, 2008. **9**(3): p. 319-327.
26. Hong, X., et al., *Sugar-binding proteins from fish: selection of high affinity "lambodies" that recognize biomedically relevant glycans.* ACS chemical biology, 2013. **8**(1): p. 152-160.
27. Han, B.W., et al., *Antigen Recognition by Variable Lymphocyte Receptors.* Science, 2008. **321**(5897): p. 1834-1837.
28. Tasumi, S., et al., *High-affinity lamprey VLRA and VLRB monoclonal antibodies.* Proceedings of the National Academy of Sciences, 2009. **106**(31): p. 12891-12896.
29. Courtemanche, N. and D. Barrick, *The leucine-rich repeat domain of Internalin B folds along a polarized N-terminal pathway.* Structure, 2008. **16**(5): p. 705-714.
30. Lee, J.j., et al., *Enzymatic prenylation and oxime ligation for the synthesis of stable and homogeneous protein–drug conjugates for targeted therapy.* Angewandte Chemie, 2015. **127**(41): p. 12188-12192.
31. Yun, M., et al., *A High-Affinity Repebody for Molecular Imaging of EGFR-Expressing Malignant Tumors.* Theranostics, 2017. **7**(10): p. 2620-2633.
32. Pyo, A., et al., *(64)Cu-Labeled Repebody Molecules for Imaging of Epidermal Growth Factor Receptor-Expressing Tumors.* J Nucl Med, 2018. **59**(2): p. 340-346.
33. Kim, T.Y., et al., *Prolonged half-life of small-sized therapeutic protein using serum albumin-specific protein binder.* Journal of Controlled Release, 2019. **315**: p. 31-39.
34. Mester, S., et al., *Extended plasma half-life of albumin-binding domain fused human IgA upon pH-dependent albumin engagement of human FcRnin vitroandin vivo.* mAbs, 2021. **13**(1): p. 1893888.
35. Lajoie, J.M., et al., *Identification of lamprey variable lymphocyte receptors that target the brain vasculature.* Scientific Reports, 2022. **12**(1).
36. Persidsky, Y., et al., *Blood–brain Barrier: Structural Components and Function Under Physiologic and Pathologic Conditions.* Journal of Neuroimmune Pharmacology, 2006. **1**(3): p. 223-236.
37. Wu, D., et al., *The blood–brain barrier: structure, regulation, and drug delivery.* Signal Transduction and Targeted Therapy, 2023. **8**(1).
38. Yager, J.Y., *Chapter 9 - Glucose and Perinatal Brain Injury—Questions and Controversies*, in *Neurology (Third Edition)*, J.M. Perlman and M.R. Cilio, Editors. 2019, Elsevier: Philadelphia. p. 141-161.
39. Koizumi, T., et al., *Vessel-Associated Immune Cells in Cerebrovascular Diseases: From Perivascular Macrophages to Vessel-Associated Microglia.* Frontiers in Neuroscience, 2019. **13**.
40. Kaal, E.C.A. and C.J. Vecht, *CNS Complications of Breast Cancer.* CNS Drugs, 2007. **21**(7): p. 559-579.
41. Yau, T., et al., *Incidence, pattern and timing of brain metastases among patients with advanced breast cancer treated with trastuzumab.* Acta Oncologica, 2006. **45**(2): p. 196-201.
42. Demeule, M., et al., *Involvement of the low-density lipoprotein receptor-related protein in the transcytosis of the brain delivery vector Angiopep-2.* Journal of Neurochemistry, 2008. **106**(4): p. 1534-1544.

43. Ulbrich, K., et al., *Transferrin- and transferrin-receptor-antibody-modified nanoparticles enable drug delivery across the blood-brain barrier (BBB).* Eur J Pharm Biopharm, 2009. **71**(2): p. 251-6.
44. Ulbrich, K., T. Knobloch, and J. Kreuter, *Targeting the insulin receptor: nanoparticles for drug delivery across the blood–brain barrier (BBB).* Journal of Drug Targeting, 2011. **19**(2): p. 125-132.
45. Christensen, S.C., et al., *Characterization of basigin monoclonal antibodies for receptor-mediated drug delivery to the brain.* Scientific Reports, 2020. **10**(1).
46. Chiduza, G.N., et al., *LAT1 (SLC7A5) and CD98hc (SLC3A2) complex dynamics revealed by single-particle cryo-EM.* Acta Crystallographica Section D Structural Biology, 2019. **75**(7): p. 660-669.
47. Harding, C., J. Heuser, and P. Stahl, *Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes.* Journal of Cell Biology, 1983. **97**(2): p. 329-339.
48. Umlauf, B.J., et al., *Identification of variable lymphocyte receptors that can target therapeutics to pathologically exposed brain extracellular matrix.* Science Advances, 2019. **5**(5): p. eaau4245.
49. Ostrom, Q.T., et al., *CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012.* Neuro-oncology, 2015. **17**(suppl_4): p. iv1-iv62.
50. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA: a cancer journal for clinicians, 2015. **65**(1).
51. Herrin, B.R. and M.D. Cooper, *Alternative adaptive immunity in jawless vertebrates.* The Journal of Immunology, 2010. **185**(3): p. 1367-1374.
52. Pancer, Z., et al., *Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey.* Nature, 2004. **430**(6996): p. 174-180.
53. Lajoie, J.M., et al., *Identification of lamprey variable lymphocyte receptors that target the brain vasculature.* Scientific Reports, 2022. **12**(1): p. 6044.
54. Collins, B.C., et al., *Structural Insights into VLR Fine Specificity for Blood Group Carbohydrates.* Structure, 2017. **25**(11): p. 1667-1678.e4.
55. Yu, C., et al., *Identification of human plasma cells with a lamprey monoclonal antibody.* JCI Insight, 2016. **1**(3).
56. Gunn, R.J., et al., *VLR recognition of TLR5 expands the molecular characterization of protein antigen binding by non-Ig-based antibodies.* Journal of molecular biology, 2018. **430**(9): p. 1350-1367.
57. Moot, R., et al., *Genetic engineering of chimeric antigen receptors using lamprey derived variable lymphocyte receptors.* Molecular Therapy-Oncolytics, 2016. **3**.
58. Deng, L., et al., *A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey.* Proceedings of the National Academy of Sciences, 2010. **107**(30): p. 13408-13413.
59. Luo, M., et al., *Recognition of the Thomsen-Friedenreich Pancarcinoma Carbohydrate Antigen by a Lamprey Variable Lymphocyte Receptor.* Journal of Biological Chemistry, 2013. **288**(32): p. 23597-23606.
60. Lee, J.Y., et al., *Overexpression, crystallization and preliminary X-ray crystallographic analysis of the variable lymphocyte receptor 2913 ectodomain fused with internalin B.* Acta Crystallographica Section F Structural Biology and Crystallization Communications, 2013. **69**(1): p. 39-41.
61. Kanda, R., et al., *Crystal Structure of the Lamprey Variable Lymphocyte Receptor C Reveals an Unusual Feature in Its N-Terminal Capping Module.* PLoS ONE, 2014. **9**(1): p. e85875.
62. Cooper, M.D. and M.N. Alder, *The Evolution of Adaptive Immune Systems.* Cell, 2006. **124**(4): p. 815-822.
63. Bhatwa, A., et al., *Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications.* Frontiers in Bioengineering and Biotechnology, 2021. **9**.

64. Grose, C., Z. Putman, and D. Esposito, *A review of alternative promoters for optimal recombinant protein expression in baculovirus-infected insect cells.* Protein Expression and Purification, 2021. **186**: p. 105924.
65. Stepanenko, A.A. and H.H. Heng, *Transient and stable vector transfection: Pitfalls, off-target effects, artifacts.* Mutat Res Rev Mutat Res, 2017. **773**: p. 91-103.
66. Rosano, G.L., E.S. Morales, and E.A. Ceccarelli, *New tools for recombinant protein production in Escherichia coli: A 5-year update.* Protein Science, 2019. **28**(8): p. 1412-1422.
67. Lobstein, J., et al., *SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm.* Microbial Cell Factories, 2012. **11**(1): p. 56.
68. Singh, P., et al., *Effect of Signal Peptide on Stability and Folding of Escherichia coli Thioredoxin.* PLoS ONE, 2013. **8**(5): p. e63442.
69. Schlegel, S., et al., *Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels.* Microbial Cell Factories, 2013. **12**(1): p. 24.
70. Malik, A., *Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of E. coli.* 3 Biotech, 2016. **6**(1).
71. Froger, A. and J.E. Hall, *Transformation of Plasmid DNA into E. coli Using the Heat Shock Method.* Journal of Visualized Experiments, 2007(6).
72. Sule, R., G. Rivera, and A.V. Gomes, *Western Blotting (immunoblotting): History, Theory, Uses, Protocol and Problems.* BioTechniques, 2023. **75**(3): p. 99-114.
73. Marisch, K., et al., *A Comparative Analysis of Industrial Escherichia coli K–12 and B Strains in High-Glucose Batch Cultivations on Process-, Transcriptome- and Proteome Level.* PLoS ONE, 2013. **8**(8): p. e70516.
74. Woestenenk, E.A., et al., *His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors.* Journal of Structural and Functional Genomics, 2004. **5**(3): p. 217-229.
75. Hong, P., S. Koza, and E.S.P. Bouvier, *A REVIEW SIZE-EXCLUSION CHROMATOGRAPHY FOR THE ANALYSIS OF PROTEIN BIOTHERAPEUTICS AND THEIR AGGREGATES.* Journal of Liquid Chromatography & Related Technologies, 2012. **35**(20): p. 2923-2950.
76. Cao, D.-D., et al., *Structure of a variable lymphocyte receptor-like protein from the amphioxus Branchiostoma floridae.* Scientific Reports, 2016. **6**(1): p. 19951.
77. Robinson, R.A., et al., *Engineering soluble T-cell receptors for therapy.* The FEBS Journal, 2021. **288**(21): p. 6159-6173.
78. Eaglesham, J.B., A. Garcia, and M. Berkmen, *Chapter Five - Production of antibodies in SHuffle Escherichia coli strains*, in *Methods in Enzymology*, Z. Kelman and W.B. O'Dell, Editors. 2021, Academic Press. p. 105-144.
79. Ahmadzadeh, M., et al., *Anti-HER2 scFv Expression in Escherichia coli SHuffle®T7 Express Cells: Effects on Solubility and Biological Activity.* Molecular Biotechnology, 2020. **62**(1): p. 18-30.
80. Herrin, B.R. and M.D. Cooper, *Alternative adaptive immunity in jawless vertebrates.* J Immunol, 2010. **185**(3): p. 1367-74.
81. Herrin, B.R., et al., *Structure and specificity of lamprey monoclonal antibodies.* Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2040-5.
82. Lajoie, J.M., et al., *Identification of lamprey variable lymphocyte receptors that target the brain vasculature.* Sci Rep, 2022. **12**(1): p. 6044.
83. Umlauf, B.J., et al., *Identification of variable lymphocyte receptors that can target therapeutics to pathologically exposed brain extracellular matrix.* Sci Adv, 2019. **5**(5): p. eaau4245.
84. Collins, B.C., et al., *Structural Insights into VLR Fine Specificity for Blood Group Carbohydrates.* Structure, 2017. **25**(11): p. 1667-1678 e4.

85. Hong, X., et al., *Sugar-binding proteins from fish: selection of high affinity "lambodies" that recognize biomedically relevant glycans.* ACS Chem Biol, 2013. **8**(1): p. 152-60.
86. Yu, C., et al., *Identification of human plasma cells with a lamprey monoclonal antibody.* JCI Insight, 2016. **1**(3).
87. Han, B.W., et al., *Antigen recognition by variable lymphocyte receptors.* Science, 2008. **321**(5897): p. 1834-7.
88. Waters, E.A. and E.V. Shusta, *The variable lymphocyte receptor as an antibody alternative.* Curr Opin Biotechnol, 2018. **52**: p. 74-79.
89. Das, S., et al., *Evolution of two distinct variable lymphocyte receptors in lampreys: VLRD and VLRE.* Cell Rep, 2023. **42**(8): p. 112933.
90. Luo, M., et al., *Recognition of the Thomsen-Friedenreich pancarcinoma carbohydrate antigen by a lamprey variable lymphocyte receptor.* J Biol Chem, 2013. **288**(32): p. 23597-606.
91. Velikovsky, C.A., et al., *Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen.* Nat Struct Mol Biol, 2009. **16**(7): p. 725-30.
92. Ng, A.C., et al., *Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**(Suppl 1): p. 4631-8.
93. Rogozin, I.B., et al., *Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase.* Nat Immunol, 2007. **8**(6): p. 647-56.
94. Pancer, Z., et al., *Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey.* Nature, 2004. **430**(6996): p. 174-80.
95. Alder, M.N., et al., *Diversity and function of adaptive immune receptors in a jawless vertebrate.* Science, 2005. **310**(5756): p. 1970-3.
96. Tasumi, S., et al., *High-affinity lamprey VLRA and VLRB monoclonal antibodies.* Proc Natl Acad Sci U S A, 2009. **106**(31): p. 12891-6.
97. Lee, S.C., et al., *Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering.* Proc Natl Acad Sci U S A, 2012. **109**(9): p. 3299-304.
98. Gunn, R.J., et al., *VLR Recognition of TLR5 Expands the Molecular Characterization of Protein Antigen Binding by Non-Ig-based Antibodies.* J Mol Biol, 2018. **430**(9): p. 1350-1367.
99. Lee, J.Y., et al., *Overexpression, crystallization and preliminary X-ray crystallographic analysis of the variable lymphocyte receptor 2913 ectodomain fused with internalin B.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2013. **69**(Pt 1): p. 39-41.
100. Moot, R., et al., *Genetic engineering of chimeric antigen receptors using lamprey derived variable lymphocyte receptors.* Mol Ther Oncolytics, 2016. **3**: p. 16026.
101. Abbott, N.J., et al., *Structure and function of the blood-brain barrier.* Neurobiol Dis, 2010. **37**(1): p. 13-25.
102. Minagar, A. and J.S. Alexander, *Blood-brain barrier disruption in multiple sclerosis.* Mult Scler, 2003. **9**(6): p. 540-9.
103. Holodinsky, J.K., et al., *History, Evolution, and Importance of Emergency Endovascular Treatment of Acute Ischemic Stroke.* Curr Neurol Neurosci Rep, 2016. **16**(5): p. 42.
104. Lockman, P.R., et al., *Heterogeneous blood-tumor barrier permeability determines drug efficacy in experimental brain metastases of breast cancer.* Clin Cancer Res, 2010. **16**(23): p. 5664-78.
105. Lobstein, J., et al., *SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm.* Microb Cell Fact, 2012. **11**: p. 56.
106. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
107. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

108. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
109. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
110. Crennell, S.J., G.O. Hreggvidsson, and E. Nordberg Karlsson, *The structure of Rhodothermus marinus Cel12A, a highly thermostable family 12 endoglucanase, at 1.8 A resolution.* J Mol Biol, 2002. **320**(4): p. 883-97.
111. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
112. Zhou, X., et al., *I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction.* Nat Protoc, 2022. **17**(10): p. 2326-2353.
113. Holm, L., *Dali server: structural unification of protein families.* Nucleic Acids Res, 2022. **50**(W1): p. W210-W215.
114. Kirchdoerfer, R.N., et al., *Variable lymphocyte receptor recognition of the immunodominant glycoprotein of Bacillus anthracis spores.* Structure, 2012. **20**(3): p. 479-86.
115. Cao, D.D., et al., *Structure of a variable lymphocyte receptor-like protein from the amphioxus Branchiostoma floridae.* Sci Rep, 2016. **6**: p. 19951.
116. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures.* Acta Crystallogr D Biol Crystallogr, 1993. **26**: p. 283-291.
117. Delano, W.R., *Unraveling hot spots in binding interfaces: progress and challenges.* Curr Opin Struct Biolo, 2002. **12**: p. 14-20.
118. Papadopoulos, J.S. and R. Agarwala, *COBALT: constraint-based alignment tool for multiple protein sequences.* Bioinformatics, 2007. **23**(9): p. 1073-9.
119. Zheng, W., N. Thorne, and J.C. McKew, *Phenotypic screens as a renewed approach for drug discovery.* Drug Discovery Today, 2013. **18**(21-22): p. 1067-1073.
120. Brown, D.G. and H.J. Wobst, *Opportunities and Challenges in Phenotypic Screening for Neurodegenerative Disease Research.* Journal of Medicinal Chemistry, 2020. **63**(5): p. 1823-1840.
121. Moffat, J.G., et al., *Opportunities and challenges in phenotypic drug discovery: an industry perspective.* Nature Reviews Drug Discovery, 2017. **16**(8): p. 531-543.
122. Davis, R.L., *Mechanism of Action and Target Identification: A Matter of Timing in Drug Discovery.* iScience, 2020. **23**(9): p. 101487.
123. Szałach, Ł.P., et al., *The immunomodulatory effect of lithium as a mechanism of action in bipolar disorder.* Frontiers in Neuroscience, 2023. **17**.
124. Frei, A.P., et al., *Direct identification of ligand-receptor interactions on living cells and tissues.* Nature Biotechnology, 2012. **30**(10): p. 997-1001.
125. Ji, L., et al., *Dissecting carboxypeptidase E: properties, functions and pathophysiological roles in disease.* Endocrine Connections, 2017. **6**(4): p. R18-R38.
126. Lyons, P.J., M.B. Callaway, and L.D. Fricker, *Characterization of Carboxypeptidase A6, an Extracellular Matrix Peptidase.* Journal of Biological Chemistry, 2008. **283**(11): p. 7054-7063.
127. Papp, H., et al., *Expression and distribution of carboxypeptidase B in the hippocampal subregions of normal and Alzheimer's disease brain.* Acta Biologica Hungarica, 2003. **54**(1): p. 55-62.
128. Trinkle-Mulcahy, L., et al., *Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes.* The Journal of Cell Biology, 2008. **183**(2): p. 223-239.
129. Terwilliger, T.C., et al., *AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination.* Nature Methods, 2024. **21**(1): p. 110-116.
130. Weids, A.J., et al., *Distinct stress conditions result in aggregation of proteins with similar properties.* Scientific Reports, 2016. **6**(1): p. 24554.

131. Pignataro, M.F., M.G. Herrera, and V.I. Dodero, *Evaluation of Peptide/Protein Self-Assembly and Aggregation by Spectroscopic Methods.* Molecules, 2020. **25**(20): p. 4854.
132. Mitragotri, S., P.A. Burke, and R. Langer, *Overcoming the challenges in administering biopharmaceuticals: formulation and delivery strategies.* Nature Reviews Drug Discovery, 2014. **13**(9): p. 655-672.
133. Lopez-Morales, J., et al., *Protein Engineering and High-Throughput Screening by Yeast Surface Display: Survey of Current Methods.* Small Science, 2023. **3**(12).
134. Shusta, E., et al., *A Decade of Yeast Surface Display Technology: Where Are We Now?* Combinatorial Chemistry & High Throughput Screening, 2008. **11**(2): p. 127-134.
135. Schreuder, M.P., et al., *Targeting of a heterologous protein to the cell wall of Saccharomyces cerevisiae.* Yeast, 1993. **9**(4): p. 399-409.
136. Silverman, L., R. Campbell, and J.R. Broach, *New assay technologies for high-throughput screening.* Current Opinion in Chemical Biology, 1998. **2**(3): p. 397-403.
137. Boder, E.T., et al., *Yeast surface display of a noncovalent MHC class II heterodimer complexed with antigenic peptide.* Biotechnology and bioengineering, 2005. **92**(4): p. 485-491.
138. Wang, Z., et al., *A new yeast display vector permitting free scFv amino termini can augment ligand binding affinities.* Protein Engineering Design and Selection, 2005. **18**(7): p. 337-343.
139. Lajoie, J.M., et al., *A yeast display immunoprecipitation screen for targeted discovery of antibodies against membrane protein complexes.* Protein Engineering, Design and Selection, 2019. **32**(5): p. 219-230.
140. Lajoie, J.M., et al., *Identification of lamprey variable lymphocyte receptors that target the brain vasculature.* Scientific Reports, 2022. **12**(1): p. 1-21.
141. Swers, J.S., B.A. Kellogg, and K.D. Wittrup, *Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display.* Nucleic acids research, 2004. **32**(3): p. e36-e36.
142. Cochran, J.R., et al., *Improved mutants from directed evolution are biased to orthologous substitutions.* Protein Engineering, Design and Selection, 2006. **19**(6): p. 245-253.
143. Feldhaus, M.J., et al., *Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library.* Nature Biotechnology, 2003. **21**(2): p. 163-170.
144. Tillotson, B.J., Y.K. Cho, and E.V. Shusta, *Cells and cell lysates: A direct approach for engineering antibodies against membrane proteins using yeast surface display.* Methods, 2013. **60**(1): p. 27-37.
145. Cho, Y.K. and E.V. Shusta, *Antibody library screens using detergent-solubilized mammalian cell lysates as antigen sources.* Protein Eng Des Sel, 2010. **23**(7): p. 567-77.
146. Cho, Y.K., et al., *A yeast display immunoprecipitation method for efficient isolation and characterization of antigens.* Journal of Immunological Methods, 2009. **341**(1-2): p. 117-126.
147. Medina-Cucurella, A., et al., *Preferential Identification of Agonistic OX40 Antibodies by Using Cell Lysate to Pan Natively Paired, Humanized Mouse-Derived Yeast Surface Display Libraries.* Antibodies, 2019. **8**(1): p. 17.
148. Stern, L.A., et al., *Cellular-Based Selections Aid Yeast-Display Discovery of Genuine Cell-Binding Ligands: Targeting Oncology Vascular Biomarker CD276.* ACS Combinatorial Science, 2019. **21**(3): p. 207-222.
149. Xu, Y., et al., *Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool.* Protein Engineering, Design and Selection, 2013. **26**(10): p. 663-670.
150. Yang, Z., et al., *A cell–cell interaction format for selection of high-affinity antibodies to membrane proteins.* Proceedings of the National Academy of Sciences, 2019. **116**(30): p. 14971-14978.

151. Bacon, K., et al., *Screening Yeast Display Libraries against Magnetized Yeast Cell Targets Enables Efficient Isolation of Membrane Protein Binders.* ACS Combinatorial Science, 2019. **21**(12): p. 817-832.
152. Woodworth, G.F., et al., *Emerging insights into barriers to effective brain tumor therapeutics.* Front Oncol, 2014. **4**: p. 126.
153. Lutz, E.A., et al., *Intratumoral nanobody–IL-2 fusions that bind the tumor extracellular matrix suppress solid tumor growth in mice.* PNAS Nexus, 2022. **1**(5).
154. Smith, J.J., et al., *Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution.* Nature Genetics, 2013. **45**(4): p. 415-421.
155. Guo, P., et al., *Dual nature of the adaptive immune system in lampreys.* Nature, 2009. **459**(7248): p. 796-801.
156. Harding, F.A., et al., *The immunogenicity of humanized and fully human antibodies.* mAbs, 2010. **2**(3): p. 256-265.
157. Baker, D., et al., *The Irony of Humanization: Alemtuzumab, the First, But One of the Most Immunogenic, Humanized Monoclonal Antibodies.* Frontiers in Immunology, 2020. **11**.
158. Krishna, M. and S.G. Nadler, *Immunogenicity to Biotherapeutics – The Role of Anti-drug Immune Complexes.* Frontiers in Immunology, 2016. **7**.
159. Vaisman-Mentesh, A., et al., *The Molecular Mechanisms That Underlie the Immune Biology of Anti-drug Antibody Formation Following Treatment With Monoclonal Antibodies.* Frontiers in Immunology, 2020. **11**.
160. Vaisman-Mentesh, A., et al., *Molecular Landscape of Anti-Drug Antibodies Reveals the Mechanism of the Immune Response Following Treatment With TNFα Antagonists.* Frontiers in Immunology, 2019. **10**.
161. Shankar, G., C. Pendley, and K.E. Stein, *A risk-based bioanalytical strategy for the assessment of antibody immune responses against biological drugs.* Nature Biotechnology, 2007. **25**(5): p. 555-561.
162. Schellekens, H., *Immunogenicity of therapeutic proteins: Clinical implications and future prospects.* Clinical Therapeutics, 2002. **24**(11): p. 1720-1740.
163. Clark, M., *Antibody humanization: a case of the 'Emperor's new clothes'?* Immunology Today, 2000. **21**(8): p. 397-402.
164. Administration, U.S.F.a.D. *Drug Databases*. 2020 [cited 2020; Available from: https://www.accessdata.fda.gov/scripts/cder/daf/.
165. Lu, R.-M., et al., *Development of therapeutic antibodies for the treatment of diseases.* Journal of Biomedical Science, 2020. **27**(1).
166. Liu, X., et al., *Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping.* Scientific Reports, 2017. **7**(1): p. 41306.
167. See, K., et al., *Reconstitution of an Anti-HER2 Antibody Paratope by Grafting Dual CDR-Derived Peptides onto a Small Protein Scaffold.* Biotechnology Journal, 2020. **15**(12): p. 2000078.
168. Sun, F., et al., *Engineering a high-affinity humanized anti-CD24 antibody to target hepatocellular carcinoma by a novel CDR grafting design.* Oncotarget, 2017. **8**(31): p. 51238-51252.
169. Sun, W., et al., *Improvement in affinity and thermostability of a fully human antibody against interleukin-17A by yeast-display technology and CDR grafting.* Acta Pharmaceutica Sinica B, 2019. **9**(5): p. 960-972.
170. Safdari, Y., et al., *Antibody humanization methods – a review and update.* Biotechnology and Genetic Engineering Reviews, 2013. **29**(2): p. 175-186.
171. Hale, G., et al., *Remission induction in non-Hodgkin lymphoma with reshaped human monoclonal antibody CAMPATH-1H.* Lancet, 1988. **2**(8625): p. 1394-9.

172. Foote, J. and G. Winter, *Antibody framework residues affecting the conformation of the hypervariable loops.* Journal of Molecular Biology, 1992. **224**(2): p. 487-499.
173. Makabe, K., et al., *Thermodynamic Consequences of Mutations in Vernier Zone Residues of a Humanized Anti-human Epidermal Growth Factor Receptor Murine Antibody, 528.* Journal of Biological Chemistry, 2008. **283**(2): p. 1156-1166.
174. Pavlinkova, G., et al., *Effects of humanization and gene shuffling on immunogenicity and antigen binding of anti-tag-72 single-chain Fvs.* International Journal of Cancer, 2001. **94**(5): p. 717-726.
175. Dondelinger, M., et al., *Understanding the Significance and Implications of Antibody Numbering and Antigen-Binding Surface/Residue Definition.* Frontiers in Immunology, 2018. **9**.
176. Olimpieri, P.P., P. Marcatili, and A. Tramontano, *Tabhu: tools for antibody humanization.* Bioinformatics, 2015. **31**(3): p. 434-435.
177. Choi, Y., et al., *Antibody humanization by structure-based computational protein design.* mAbs, 2015. **7**(6): p. 1045-1057.
178. Kurella, V.B. and R. Gali, *Antibody Design and Humanization via In Silico Modeling*, in *Antibody Engineering*. 2018, Springer New York. p. 3-14.
179. Morlot, C., et al., *Structural insights into the Slit-Robo complex.* Proceedings of the National Academy of Sciences, 2007. **104**(38): p. 14923-14928.
180. Holm, L., *Using Dali for Protein Structure Comparison*. 2020, Springer US. p. 29-42.
181. Gietz, R.D. and R.H. Schiestl, *Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method.* Nature Protocols, 2007. **2**(1): p. 1-4.
182. Kurtzman, C.P., C. Guthrie, and G.R. Fink, *Guide to Yeast Genetics and Molecular Biology.* Mycologia, 1993. **85**(4): p. 714.
183. Yue, B., *Biology of the extracellular matrix: an overview.* J Glaucoma, 2014. **23**(8 Suppl 1): p. S20-3.
184. Hellewell, A.L., S. Rosini, and J.C. Adams, *A Rapid, Scalable Method for the Isolation, Functional Study, and Analysis of Cell-derived Extracellular Matrix.* J Vis Exp, 2017(119).
185. Joubert, M.K., et al., *Use of In Vitro Assays to Assess Immunogenicity Risk of Antibody-Based Biotherapeutics.* PLoS One, 2016. **11**(8): p. e0159328.
186. Wada, Y., et al., *A humanized mouse model identifies key amino acids for low immunogenicity of H7N9 vaccines.* Scientific Reports, 2017. **7**(1): p. 1283.