UNIVERSITY OF CALIFORNIA SAN DIEGO

Examining the Phage and Ribosome Display of Diversity-Generating Retroelement Variable Proteins

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Adrian Bahn

Committee in charge:

Professor Partho Ghosh, Chair
Professor James Wilhelm, Co-Chair
Professor Russell Doolittle

2018

The Thesis of Adrian Bahn is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2018

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

aa	Amino acid (number)
bp	Base pairs
Da	Daltons
F_{ab}	Antibody antigen-binding fragment
g	Gravitational acceleration (m/s^2)
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MW	Molecular weight
PCR	Polymerase chain reaction
PVDF	Polyvinylidene fluoride
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
V_HH	Camelid variable heavy chain antibody

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Partho Ghosh for his support and guidance in the lab and as the chair of my committee. In addition to all of the members of the Partho Ghosh Lab who have helped me over the years, I would particularly like to thank Dr. Cosmo Buffalo and Dr. Sumit Handa for their mentorship in my development as a researcher.

I would also like to acknowledge Dr. Candace Spier Bever of the Hammock Lab in UC Davis for her gift of the phagemid containing the V_HH gene used for the phage display positive control.

Finally, I would also like to acknowledge Dr. Takuya Ueda of the Ueda Lab in the University of Tokyo for his gift of the pET-17b plasmid, with which the majority of the DNA cloning work for ribosome display was done.

Chapter 1 is, in part, a reprint of my undergraduate Senior Honors Thesis at the University of California San Diego.

VITA

2012-2018	Research Assistant, University of California San Diego
2014-2017	Teaching Assistant, University of California San Diego
2016	Bachelor of Science, University of California San Diego
2018	Master of Science, University of California San Diego

PUBLICATIONS

Buffalo CZ, Bahn-Suh AJ, Hirakis SP, et al. Conserved patterns hidden within group A *Streptococcus* M protein hypervariability are responsible for recognition of human C4b-binding protein. *Nature microbiology*. 2016;1:16155. doi:10.1038/nmicrobiol.2016.155.

FIELDS OF STUDY

Major Field: Biology

Studies in Biochemistry

Professor Partho Ghosh

ABSTRACT OF THE THESIS

Examining the Phage and Ribosome Display of Diversity-Generating Retroelement Variable Proteins

by

Adrian Bahn

Master of Science in Biology

University of California San Diego, 2018

Professor Partho Ghosh, Chair
Professor James Wilhelm, Co-Chair

Diversity-generating retroelements (DGRs) are genetic modules with the ability to introduce massive sequence variation into specific target proteins. Such DGR variable proteins are able to specifically recognize and bind ligands, and their properties have several potential applications in molecular surface display technologies. The first step in such an endeavor is to show that the DGR variable proteins are compatible with molecular surface display systems. The M13 filamentous phage display system was tested for surface-display of DGR variable proteins. However, only truncated versions of the DGR variable proteins were detected in isolated phage. Further investigation strongly suggested that periplasmic proteolysis of the DGR variable proteins occurred during the phage maturation process. Next, ribosome display of DGR variable proteins

was tested. Intact DGR variable proteins were detected by polyacrylamide gel electrophoresis and western blot after *in vitro* transcription and translation. Affinity selection of a ribosome-displayed DGR variable protein was demonstrated, which verified the presence of accompanying mRNA. This work demonstrates the incompatibility of several DGR variable proteins for M13 phage display, but also demonstrates that ribosome display is a viable alternative for surface display of DGR variable proteins.

INTRODUCTION

Significance of the Project

Molecular surface display is a powerful tool used in a wide range of research topics, such as therapeutic drug development and structural biochemistry, and a major application in these areas is the selection of displayed recombinant molecules with desirable characteristics from combinatorial libraries (Hosse et al., 2006; Sergeeva et al., 2006). Selection criteria can vary, but one example is the selection of recombinant protein scaffolds with high binding affinities for target ligands from protein libraries.

One of the oldest and most popular molecular display scaffolds for identifying binders is one derived from a naturally occurring system of display: antibodies (Parmley and Smith, 1988). Some of the advantages of using antibody derivatives — F_{ab} fragments, single-chain variable fragments (scFvs), and single-domain antibodies (for example V_HH fragments) — for molecular display include large potential library complexity, strong and specific binding interactions, and relatively high overall stability (Plückthun, 2009). The robustness and versatility of antibody scaffolds have led to their success in both clinical and academic settings. Multiple FDA-approved drugs have been developed through antibody engineering with phage display alone, such as Adalimumab and Belimumab (Nixon et al., 2014). Furthermore, selection of antibody scaffolds as crystallization chaperones through molecular display has led to the crystallization and subsequent structural determination of a number of difficult targets, including the unstable addiction antidote MazE and the closed conformation of the potassium channel KscA (Loris et al., 2003; Uysal et al., 2009). However, antibody display is not without its limitations. For example, F_{ab} fragments are more difficult to produce due to the necessary assembly of two heavy and light chains (Bradbury and Marks, 2004). In addition, F_{ab} and V_HH fragments usually contain disulfide bridges that are necessary for proper folding, which can hamper their production or use in reducing environments.

ScFvs can be engineered without disulfides, but the absence of the stabilizing constant domains can result in unintended oligomerization, aggregation, and modified binding activity (Proba et al., 1998; Weatherill et al., 2012). While work is being done to improve upon current antibody display technologies, alternative display scaffolds are simultaneously being developed for molecular display applications.

Alternative binding scaffolds like affibodies, fibronectin type III (FN3) domains, designed ankyrin repeat proteins (DARPins), and others are diverse in origin, structure, and physical properties (Hosse et al., 2006). These alternative scaffolds may complement or prove superior to antibody scaffolds. DARPins for example are α-helical proteins engineered from the ubiquitous ankyrin repeat motif, with the number of ankyrin repeats determining the molecular weight. DARPins typically range from 14-18 kDa, lack disulfide bonds, express and fold well in *Escherichia coli* cytoplasm, and exhibit high thermodynamic stability (Binz et al., 2003; Kohl et al., 2003). In a proof of concept, crystallization chaperones to the transmembrane sodium citrate transporter CitS of *Klebsiella pneuomoniae* were selected from DARPin and F_{ab} fragment libraries using molecular display. Both the selected DARPins and F_{ab} fragments were shown to aid in crystallization of CitS, with evidence for different binding epitopes between the two scaffolds (Huber et al. 2007). Indeed, a comparison of the crystal structure of the receptor-binding protein (RBP) of the lactococcal phage TP901-1 between DARPin and V_HH-complexed crystals showed that the concave paratope of DARPins bound to a convex epitope on RBP, and the convex paratope on the V_HH interacted with a concave epitope on RBP (Veesler et al., 2009). The ability to target different binding epitopes can be thought of as a complementary or superior feature of DARPins. Thus, alternative scaffolds have the potential to compete with antibodies, although currently antibody display remains the most developed and successful technology.

A large portion of the success of antibodies as binding scaffolds can be attributed to the fact that nature has evolved a protein that is well suited to its function over millions of years. Therefore, it would be reasonable to explore the development of alternative display scaffolds that also come from natural display systems. In addition to antibodies, there are two other antigen-binding proteins in the vertebrate adaptive immune system: T cell receptors (TCRs) and variable lymphocyte receptors (VLRs). TCRs, however, recognize antigen epitopes presented by major histocompatibility complexes (MHCs), restricting their potential applications compared to antibodies and other scaffolds. Furthermore, while library-based molecular display of T cell receptors has been shown to be feasible, the production of soluble recombinant TCRs is laborious and challenging due to their origins as membrane-bound proteins (Kieke et al., 1999; Shusta et al., 2000; Tasumi et al., 2009). A more promising scaffold is the VLR from the jawless vertebrate adaptive immune system (Pancer et al., 2004). Three VLRs have been identified; VLRA and VLRC are T cell receptor-like, and VLRB can be secreted akin to antibodies (Boehm et al., 2012). Monoclonal VLRs have been selected through molecular display to bind specifically and strongly to a broad range of targets, but the technology is still in its infancy (Satoshi et al., 2009). Outside the vertebrate adaptive immune system, there is only one other naturally occurring system of display: diversity-generating retroelements.

Diversity-generating retroelements (DGRs) are a family of genetic modules whose functions are to diversify the DNA sequences of their (often surface-displayed) DGR variable proteins (VPs). DGRs are currently the only natural case of massive protein sequence variation outside of the vertebrate adaptive immune system and have been found in bacteria (*Treponema denticola*, *Legionella pneumophila,* and *Thermus aquaticus*), bacteriophage (*Bordetella* bacteriophage), archaea, and archaeal viruses (Arambula et al., 2013; Doulatov et al., 2004; Le Coq and Ghosh, 2011; Paul et al., 2015). Although the evolutionary functions of most DGRs

remain largely unknown, at least in *Bordetella* bacteriophage the massive protein sequence variation appears to be a mechanism to adapt to changes in expression patterns of potential *Bordetella* surface phage receptors (Liu et al., 2002; Miller et al., 2008). This strategy of displaying variants of a surface protein in the hopes of capturing variable ligands is analogous to that of the vertebrate adaptive immune system, and like antibodies DGR VPs also possess many attractive features for biotechnological development. DGR VPs have a conserved C-type lectin fold (a general ligand-binding motif), with the potential to target different epitopes compared to antibody loops (Le Coq and Ghosh, 2011). Furthermore, DGR VPs are relatively small proteins (~28-40 kDa) that are crystallizable, easy to produce in large quantities in bacteria, lack disulfide bonds, and their ligand-binding sites can potentially accommodate at least 10^{20} different sequences, which is much greater than the potential sequence diversity of antibodies (~10^{14-16}) (Miller et al., 2008; Le Coq and Ghosh, 2011). While the biotechnological development of DGR VPs as binding scaffolds has not yet been explored, their properties hold promise to compete with or complement antibody scaffolds in many of the same applications—crystallography, flow cytometry, and ELISA among others.

The biotechnological potential of DGR VPs in crystallography is particularly compelling because obtaining diffraction-quality crystals remains one of the primary challenges for structural biologists (Koide, 2009). Many biological macromolecules are not amenable to crystallization due to their inherent instability, conformational heterogeneity, or insolubility (McPherson and Gavira, 2014). Of the many developments made in an effort to overcome the problems of macromolecular crystallization, crystallization chaperones are one of the most successful. Crystallization chaperones are auxiliary molecules (often proteins) that interact with a target of interest to promote crystallization. They can provide polar interactions for crystal lattice formation, “lock” proteins into single conformations, and in some cases, crystallization is only possible when a target is

complexed with the chaperone (Bukowska and Grutter, 2013; Griffin and Lawson, 2011; Kovari, 1995; Ostermeier, 1997). Natural ligands with a propensity for crystallization have been used as crystallization chaperones to co-crystallize previously "uncrystallizable" target molecules, although this tactic is limited by the number and characteristics of known binding partners (Griffin and Lawson, 2011). Therefore, crystallization chaperones have been selected using molecular display, such as the previously mentioned DARPins and antibodies and antibody fragments among others (Binz et al., 2005; Koide, et al., 2007; Loris et al., 2003). DGR VPs are especially suitable for crystallization chaperone selection because of their high potential variability, thermostability, and crystallizability. High variability increases the probability of selecting a suitable binder, while thermostability and crystallizability are beneficial for promoting the crystallization of target molecules. However, before these properties can be exploited, the compatibility of DGR VPs with molecular surface display must first be tested.

Candidate DGR VPs

Candidate DGR VPs for biotechnological development include the *Bordetella* bacteriophage major tropism determinant (Mtd), the *Treponema denticola* variable protein (TvpA), and the archaeal variable protein (AvpA) (Handa, et al., 2016; Le Coq and Ghosh, 2011; Liu, et al., 2002). The *Thermus aquaticus* variable protein TaqVP is another candidate that has been structurally characterized by the Partho Ghosh lab (S. Handa, manuscript in preparation). Polymeric proteins are difficult to use in phage and ribosome display, so the monomeric AvpA, TvpA, and TaqVP were chosen in favor of the trimeric Mtd for this project. AvpA, TvpA, and TaqVP have masses of 28 kDa, 36 kDa, and 40 kDa, respectively. TvpA can accommodate the most variability, as many as ~10^{20} sequences (LeCoq and Ghosh, 2011). The other DGR VPs can accommodate at least 10^9 different sequences. AvpA is thermostable up to 70 °C, as is TaqVP

(Paul et al., 2015). All three proteins lack disulfide bonds and can be expressed in the *E. coli* cytoplasm as soluble proteins in large quantities.

Introduction to Molecular Display and Affinity Selection

Molecular display can be accomplished in a variety of ways, but certain core principles remain the same across all molecular display methods. Beginning with a "template" protein sequence, a genetic library of mutant sequences is constructed and expressed on the surface of its carrier (i.e. the cell surface, phage surface, etc.). Each mutant protein is linked to its specific genetic sequence (phenotype-genotype linkage) in various ways depending on the display method. The surface displayed proteins then undergo a selection step based on their binding affinities to a target molecule. Typically, this selection is done by immobilizing the target on a solid substrate, introducing the displayed proteins, and washing away nonbinding mutants. The iterative process of washing away non-binding proteins and enriching binders is called "affinity selection" (McCafferty et al., 1990). After washing, the genetic information linked to any binders is retrieved and re-used in more cycles of display and affinity selection, which is called "panning." Panning essentially results in the directed evolution of the proteins by enriching the displayed population with binding phenotypes (Levin and Weiss, 2005). This thesis examines the feasibility of the phage and ribosome display of DGR VPs, but other molecular display methods exist as well (Galán, et al., 2016).

CHAPTER 1: PHAGE DISPLAY

1.1 INTRODUCTION

Phage display is the oldest and most widespread method for displaying recombinant proteins, with hundreds of phage display publications a year (Sergeeva et al., 2006). In phage display, the to-be-displayed protein's gene is inserted into a phage coat gene for expression as a fusion coat protein. The phenotype-genotype linkage is accomplished via the phage's replication cycle. During phage capsid assembly, the individual fusion proteins and their DNA sequences are assembled into phage particles. The recombinant proteins are thus exposed on the surface of the phage and their corresponding genes are encapsulated by the phage capsid (Smith, 1985).

There are multiple bacteriophage, vectors, and phage coat proteins to consider when performing phage display. Phage display can be accomplished with lytic (T7, T4, lambda) phages or non-lytic filamentous (f1, fd, M13) phages (Bradbury and Marks, 2004). A major consideration for choosing which phage to work with is the eventual construction of the phage library. For example, construction of a T7 phage library using the Novagen T7Select® kit requires cloning of linear phage DNA vectors and packaging the DNA into phage (T7 Select System Manual). Creating libraries in such a fashion can be difficult, expensive, and result in variable quality and complexity. Filamentous phage libraries on the other hand can be constructed quite easily and inexpensively with phagemid templates and bacterial transformation. Therefore, for the phage display of DGR VPs, M13 phage display using the 3+3 vector system was chosen for its established history, ease of use, and low avidity of displayed proteins as explained below (Barbas et al., 2001).

M13 is a non-lytic filamentous phage; it only infects bacteria containing the F episome (fertility factor) and does not kill host cells during replication. Infected bacterial cultures continuously produce phage particles and the phage can be purified from the culture supernatant.

The 3+3 vector system is a system that consists of a "helper" phage and a phagemid vector containing the sequence (gene III) encoding the C-terminal residues of the phage coat protein pIII (aa 230-406). Bacteria containing the phagemid cannot normally produce phage because the phagemid lacks other necessary phage genes. However, infection of the bacteria with wild-type helper phage supplements the genes needed for phage production. A portion of the newly assembled phage will package phagemid and, if a gene has been inserted upstream of gene III, the pIII fusion protein. pIII is the protein responsible for infecting bacteria through the F-pilus, and each phage particle only contains 1-5 copies of pIII. Approximately 1-10% of phages will contain one copy of fusion pIII, and an extremely low population of phages will have multiple copies of fusion pIII. The inefficient incorporation of fusion pIII is beneficial for future affinity selections steps because selected phages will most likely display proteins with strong single interactions rather than multiple weak interactions (avidity), and phages with too many fusion proteins exhibit reduced infectivity (Barbas et al., 2001; Bradbury and Marks, 2004).

However, the major shortcoming of filamentous phage display rests on the assumption that the recombinant protein is compatible with the phage assembly process. Non-lytic filamentous phage assembles on the plasma membrane before it buds out of the bacterium. Therefore, the recombinant protein must undergo periplasmic secretion before it is incorporated into phage. Some proteins cannot complete secretion or survive periplasmic proteases, and these requirements can be affected by the sizes, sequences, and folding characteristics of different proteins. Thus, successful filamentous phage display is largely protein-dependent, and the compatibility of DGR VPs with phage display must be experimentally determined.

1.2 MATERIALS AND METHODS

Phage Display Vector

The phage display vector used was the pComb3xSS phagemid — a gift from the Barbas lab at the Scripps Research Institute. The phagemid cannot produce viable phage, and helper phage is required for phage production. Antibiotic selection for the phagemid is accomplished through its amp^R gene. The "SS" in pComb3xSS refers to its two stuffer regions (double stuffer) used for F_{ab} cloning, a 1200 bp stuffer bounded by SacI and XbaI restriction sites, and a 300 bp stuffer bounded by XhoI and SpeI restriction sites (Figure 1.1). An *ompA* signal sequence upstream of the double stuffer directs the export of fusion proteins to the periplasm for phage display. Downstream of the double stuffer is the sequence encoding the C-terminal residues of pIII. Two peptide tags, a hexa-histidine (6xHis) tag directly followed by a hemagglutinin (HA) tag, are present in a linker region between the stuffer and gene III. The double stuffer can be removed by SfiI digestion for cloning of non-F_{ab} genes. Therefore, replacement of the double stuffer with a DGR VP gene allows for the expression of DGR VP-pIII fusion protein. Transcription of cloned genes is under the control of the *lac* operon. While transcription is normally repressed in the absence of lactose or lactose-homologues, "leaky" repression of the *lac* operon can lead to low level production of fusion protein that is sufficient for phage display (Barbas et al., 2001).

An important feature of pComb3xSS is the presence of an amber stop codon after the sequence encoding the HA tag. Certain nonsense suppressor strains of *E. coli* will translate through the amber stop codon and produce periplasmic pIII-fusion protein anchored to the inner membrane. Strains lacking the correct nonsense suppressor gene will not read through the amber stop and will instead produce protein targeted to the periplasm and lacking a transmembrane anchor. A suppressor strain is necessary for phage display using pComb3xSS, though production of soluble protein can be accomplished by the same phagemid simply by switching the strain of *E. coli*

(Barbas et al., 2001). The *E. coli* strain used for phage display was *E. coli* K12 ER2738 (New England Biolabs). Important features of *E. coli* K12 ER2738 are *glnV*, which allows the amber stop codon to be read as a glutamine, and the F' episome, which is necessary for infection by filamentous phage. *E. coli* K12 ER2738 cells must be plated or grown in the presence of tetracycline for maintenance of the F' episome.

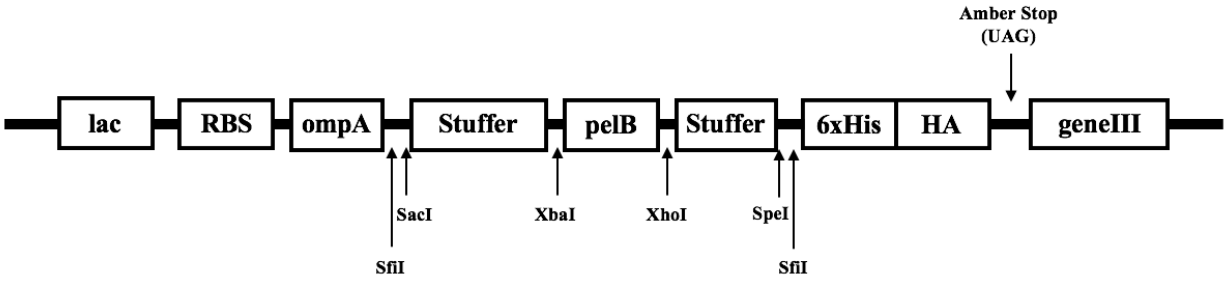

Figure 1.1 Schematic of the pComb3xSS cloning region. Cloning via the SfiI restriction sites results in the expression of pIII-fusion protein in suppressor strains of *E. coli*.

DGR VP Cloning

The DGR VP genes were kindly provided by Sumit Handa of the Partho Ghosh Lab. The genes were amplified by PCR using primers containing SfiI restriction sites flanking the 5' and 3' ends of the genes. PCR reactions consisted of 1 ng/μL of template DNA, 0.4 μM of forward and reverse primers, 5% DMSO, and in-house prepared Pfu DNA polymerase in 1x Pfu buffer. The annealing step used a thermal gradient from 50-68 °C for thirty cycles. The resulting PCR products were electrophoresed through a 1% agarose gel, stained with GreenGlo™ Safe DNA Dye, and gel purified using the Wizard® SV Gel and PCR Clean-Up System kit. The PCR products and the pComb3xSS phagemid were separately digested with 20 units of SfiI (New England Biolabs) for 16 h. Digestion products were gel purified and ligated overnight at 16 °C. Ligation reactions contained 1:1-1:10 molar ratios of digested vector and insert, 1 mM ATP, and 0.5 μL T4 DNA

ligase (New England Biolabs) in 1x T4 DNA ligase buffer (New England Biolabs). The ligation products were transformed into *E. coli* BL21-Gold (DE3) (Agilent Technologies) chemically competent cells through heat shock at 42 °C for 30 sec. After a 1 h recovery period in SOC media, the transformed cells were plated onto 100 μg/mL carbenicillin Lysogeny Broth (LB) agar plates. A few of the resulting colonies were picked and cultured overnight with shaking at 37 °C in a few mL of LB media. Phagemids from the cultures were purified using the Wizard® Plus SV Minipreps DNA Purification System. Phagemids containing the correct inserts were determined by DNA sequencing. The phagemids were then re-transformed into *E. coli* K12 ER2738 chemically competent cells.

Bacterial Expression of Proteins

Starter cultures of *E. coli* were grown overnight from freshly streaked plates. These cultures were diluted 1:1000 in fresh LB media (e.g. 10 μL of culture to 10 mL of LB) containing 100 μg/mL carbenicillin and grown to an OD_{600} of 0.4-0.6. After reaching the desired optical density, protein expression was induced with a final concentration of 1 mM IPTG. The cultures were then shaken overnight at 37 °C, after which they were pelleted by centrifugation at 4,000 x *g* for 5-10 min. The pellets were resuspended in 1/10 of the original culture volume of phosphate-buffered saline (PBS), and lysed by sonication on ice using a duty cycle of 30% for 3 cycles of 30 sec on and 1 min off.

Plaque Assay and Preparation of Helper Phage Stock

The helper phage used was the M13 derivative VCSM13 helper phage (Agilent Technologies). VCSM13 helper phage preferentially packages phagemid and contains kan^R for selection of infected bacteria. In order to generate a working stock of helper phage, a plaque assay

was performed using serial dilutions of commercially available VCSM13 helper phage. The plaque assay was performed as follows: *E. coli* K12 ER2738 without phagemid were grown to an OD_{600} of 0.5-0.6 in LB (20 μg/mL tetracycline and 100 μg/mL carbenicillin) at 37 °C with shaking. One hundred μL of bacteria was mixed with 5 μL of various dilutions of helper phage and incubated at RT for 15 min. After infection, 3 mL of warm LB top agar (0.7% agar in LB media) was mixed with each tube, and the mixtures were plated onto pre-warmed LB agar plates. The top agar was allowed to solidify, and the plates were then incubated overnight at 37 °C to develop plaques. The goal was to obtain a plate with well-separated plaques, which appear as regions of slow bacterial growth.

Helper phage stock was prepared the day following the plaque assay. Ten mL of 2x YT media (Sigma-Aldrich) was inoculated with 10 μL of *E. coli* K12 ER 2738 and incubated for 1 h at 37 °C with shaking. A single plaque was picked from the plates with a sterile pipette tip and used to infect the 10 mL of *E. coli* K12 ER2738. After incubating the infected cells for a further 2 h at 37 °C with shaking, the culture was then added to a 1-liter Erlenmeyer flask containing 250 mL of 2x YT with 70 μg/mL of kanamycin and 20 μg/mL of tetracycline. After growing overnight at 37 °C with shaking, the bacteria were pelleted at 2,500 x *g* for 15 min. The phage-containing supernatant was decanted and incubated in a 70 °C water bath for 20 min to kill residual bacteria, and pelleted again at 2,500 x *g* for 15 min. The resulting supernatant is the crude helper phage stock. The stock has been observed to be stable for months at 4 °C, though the liquid may become turbid over time from the growth of surviving bacteria. Filtering the stock through a 0.22-micron membrane delayed the onset of turbidity, although the stock's titer decreased after filtration. The stock's titer was periodically determined through plaque assays, and if the titer fell significantly during storage, this preparation was repeated.

M13 Phage Production and PEG/NaCl Purification

Five mL of *E. coli* K12 ER2738 cells containing pComb3xSS was grown to an OD_{600} of 0.5-0.6 in 10 mL of 2x YT with 20 μg/mL of tetracycline and 100 μg/mL of carbenicillin at 37 °C with shaking. For production of wild-type phage (i.e., not containing pIII fusion proteins), *E. coli* K12 ER2738 not containing phagemid were used. Cultures were infected at a multiplicity-of-infection (MOI) of 5 plaque forming units (pfu) from the phage stock and incubated for 30 min at 37 °C with shaking. Infected cultures were then transferred to separate 1 L flasks containing 150 mL of 2x YT media with 20 μg/mL tetracycline, 100 μg/mL carbenicillin, and 25 μg/mL of kanamycin. The phages were allowed to propagate for 18 h at 37 °C with shaking.

For subsequent PEG/NaCl precipitation, the bacteria were pelleted from the cultures by centrifuging at 2,400 x *g* for 10 min at 4 °C. The supernatants were decanted into new tubes, and more bacteria were pelleted by centrifuging the supernatants again at 6,200 x *g* for 10 min. The supernatants were decanted into new tubes, and 0.15 volumes of PEG/NaCl solution (16.7% polyethylene glycol 8000/3.3 M NaCl) was added to each sample. The phages were allowed to precipitate overnight at 4 °C and pelleted from the solutions by centrifuging at 6,200 x *g* for 40 min at 4 °C. The supernatants were discarded, and the pellets were resuspended in 2.5 mL Tris-buffered saline, pH 8.0 (TBS, 50 mM Tris and 150 mM NaCl) and transferred to Oak Ridge tubes for centrifugation at 22,700 x *g* for 10 min at 4 °C. A dark pellet may be visible, though the phages should remain in solution. The supernatants were then transferred to fresh tubes for a second round of overnight precipitation at 4 °C with 0.15 volumes of PEG/NaCl solution. Precipitated phages were pelleted again at 10,100 x *g* for 40 min at 4 °C, supernatants discarded, and resuspended in 2 mL TBS. The Oak Ridge tubes were centrifuged one last time for 10 min at 22,700 x *g*. Another pellet may be visible. The supernatants were transferred to fresh tubes, and these samples

correspond to the final purified phages. Phage yield was determined through a plaque assay. One pfu was considered to correspond to one phage particle.

SDS-PAGE and Western Blot

Samples were prepared for SDS-PAGE by boiling for 2 minutes in 1x SDS-PAGE loading buffer (62.5 mM Tris-HCl, pH 6.8, 2.5% SDS, 0.002% bromophenol blue, 5% β-mercaptoethanol, and 10% glycerol) and electrophoresed through in-house made 10% SDS-polyacrylamide gels. After SDS-PAGE, gels were incubated in Towbin buffer (25 mM Tris, 192 mM glycine, and 20% methanol), pH 8.3 for 5-10 min. PVDF membranes were cut to size, activated with 100% methanol, and also incubated in Towbin buffer. A wet transfer was then performed in Towbin buffer using 100 V for 30 min onto the membranes. The membranes were then blocked for 1 h at RT with blocking buffer (5% nonfat milk in TBS with 0.05% Tween 20, or TBST). After discarding the blocking buffer, the membranes were probed with primary antibody diluted in 1% nonfat milk in 1x TBST overnight at 4 °C. The next day, the membranes were washed with 1x TBST for 5 min three times and probed with secondary antibody diluted in 1% nonfat milk in 1x TBST for 1 h at RT. After washing three more times, the membranes were developed using 1 mL of SuperSignal™ Pico (Thermo Fisher Scientific) and 0.2 mL of SuperSignal™ Femto (Thermo Fisher Scientific) and visualized. Primary antibodies used were mouse monoclonal HA antibody (Thermo Fisher Scientific) and mouse anti-M13 pIII monoclonal antibody (New England Biolabs). The secondary antibody was goat anti-mouse HRP-conjugated sc-2005 (Santa Cruz Biotechnology). Antibody dilutions were according to manufacturer's recommendations.

1.3 RESULTS

Bacterial Expression of TaqVP and TaqVP-pIII Fusion

Before proceeding to phage display, the expression of a DGR VP in the phagemid vector was tested to determine if the membrane-bound pIII fusion protein could be detected in bacterial samples. Confirmation of the production of fusion protein would be supportive of future phage display experiments. The gene for TaqVP was cloned into pComb3xSS, and *E. coli* K12 ER2738 cells containing TaqVP in pComb3xSS were induced with IPTG to produce protein overnight. *E.* coli BL21-Gold (DE3) cells containing TaqVP in pET-28a (provided by Sumit Handa) were used as a positive control for induction. Whole cell extracts were then analyzed by SDS-PAGE (Figure 1.2). The production of TaqVP was confirmed in the cell extract of induced *E. coli* BL21-Gold (DE3) cells, but the TaqVP-pIII fusion protein was not observed in the cell extract of induced *E. coli* K12 ER2738 cells. Two bands with apparent molecular weights of ~48-51 kDa were apparent, but their molecular weights do not correspond to the weight of the TaqVP-pIII fusion protein (MW ~58 kDa).

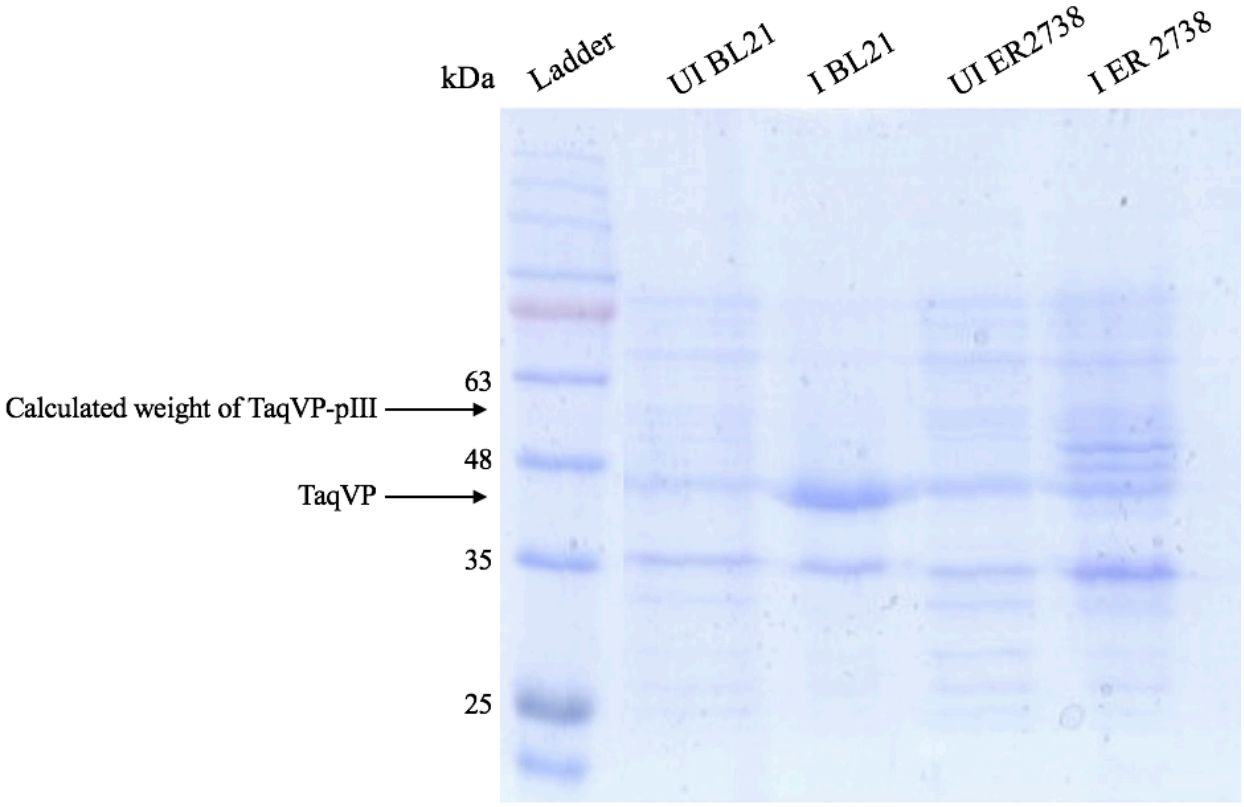

Figure 1.2 SDS-polyacrylamide gel electrophoresis of uninduced (UI) and induced (I) cultures of *E. coli* BL21-Gold (DE3) cells expressing cytoplasmic TaqVP and *E. coli* K12 ER2738 cells expressing periplasmic TaqVP-pIII fusion protein. TaqVP (calculated molecular weight of 40 kDa) can be seen in the induced BL21 lane. Two additional bands can be seen between 48-50 kDa in the induced ER2738 lane.

A few possibilities exist for why the TaqVP-pIII fusion protein was not observed. One possibility is that the TaqVP-pIII fusion protein is detrimental or toxic to host cells and undetectable amounts of protein were produced. Another possibility is that the TaqVP-pIII fusion protein does not adequately tolerate periplasmic secretion or misfolds in the cytoplasm or periplasm, in which case the misfolded protein could be targeted by cytoplasmic or periplasmic proteases. It is possible that the ~48-51 kDa bands for the induced *E. coli* K12 ER2738 cells could be products from the proteolysis of TaqVP-pIII. However, despite the failure to detect fusion protein in bacterial cell lysates, this result did not necessarily rule out the possibility of the phage display of TaqVP. If the protein is secretion-competent, assembly of the fusion protein into phage

capsids could rescue the protein. Therefore, phage produced from infected *E. coli* K12 ER2738 cells containing TaqVP in pComb3xSS were purified to determine if the peptide tags on the fusion protein were present on the phage.

M13 Phage Display of TaqVP

Phage display of TaqVP was examined by anti-HA and anti-pIII western blot of phage samples purified from K12 *E. coli* ER2738 cultures containing no phagemid (wild-type phage) or either of TaqVP or BDE-47 VHH in pComb3xss (Figure 1.3). BDE-47 VHH was used as a positive control for production of phage containing fusion pIII, and it is a 15 kDa V_HH fragment that was previously selected to bind BDE-47 (2,2`, 4,4`-tetrabromodiphenyl ether) using pIII phage display (Bever, et al., 2014). Each western blot sample represents approximately 1 x 10^9 pfu as determined by plaque assay.

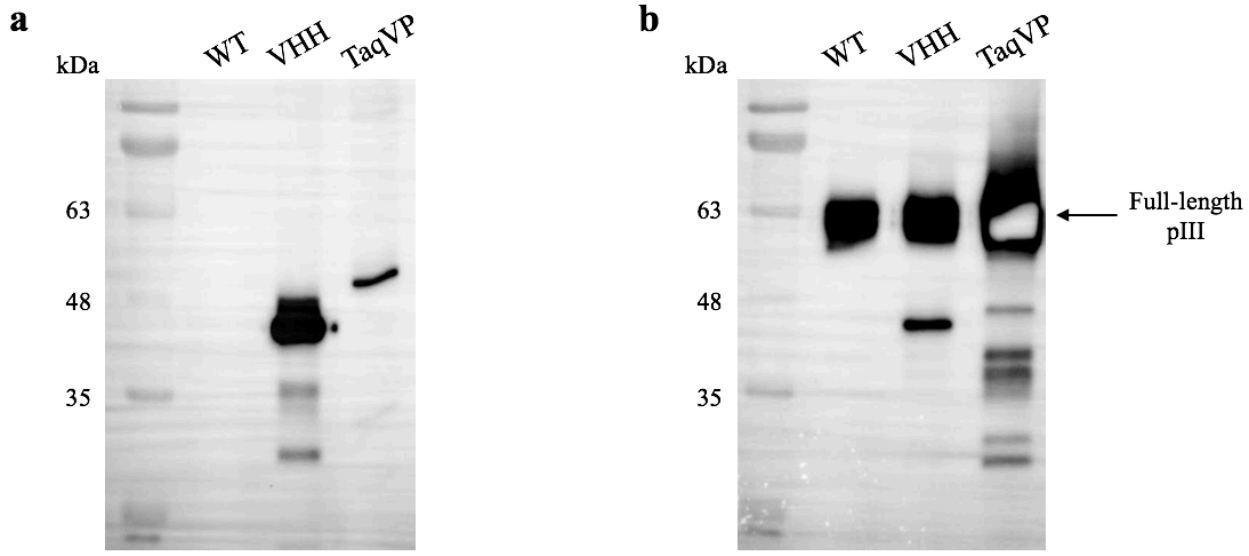

Figure 1.3 Western blots of VCSM13 phage purified from bacteria expressing TaqVP-pIII fusion protein (TaqVP), V_HH-pIII fusion protein (VHH), or wild-type (WT) phage. **a.** Anti-HA western blot. No signal was detected in wild-type phage, and the V_HH fusion appears to run at 43-45 kDa. A single band was detected for TaqVP fusion. **b.** Anti-pIII western blot. Full-length pIII is 42 kDa but can be seen to run at 60-65 kDa and is the major species in all the phage sample. The V_HH fusion was detected at the same size as in the anti-HA western blot, but multiple low molecular weight bands were detected for TaqVP fusion,

In Figure 1.3a, the HA tag was not detected in the anti-HA western blot of wild-type phage samples, which demonstrates that there is no cross-reactivity of the antibody. The VHH sample shows detection of the HA tag on a ~43-45 kDa protein, which is corroborated by the detection of a similar band by anti-pIII western blot in Figure 1.3b. VHH-pIII fusion has a calculated molecular weight of 33 kDa, of which 18 kDa is contributed by the C-terminal residues of pIII. Figure 1.3b provides an explanation as to why a 33 kDa band is not detected in the VHH lane. Full-length pIII has a calculated molecular weight of 42 kDa but is observed in Figure 1.3b as running at ~60-63 kDa. pIII is known to run at a higher apparent molecular mass in SDS-PAGE, possibly due to its glycine repeats (Wezenbeek et al., 1980). As a result, the pIII fusion proteins may also run higher than their calculated molecular weights, explaining why VHH-pIII is running at a molecular mass of ~43-45 kDa. The TaqVP-pIII fusion protein may also be expected to run higher than its calculated molecular weight of 58 kDa, although Figure 1.3a. shows detection of a single ~48-51 kDa band, which is significantly smaller than the calculated weight of TaqVP-pIII. A similar band is detected in Figure 1.3b, in addition to revealing multiple lower mass bands. The detection of similarly sized proteins with the anti-HA and anti-pIII antibodies in the TaqVP lanes strongly suggests that the TaqVP-pIII fusion protein is secreted to the periplasm and assembles into phage. However, the low apparent molecular masses for the bands in the TaqVP lanes in Figure 1.3 indicate that the fusion protein is degraded.

Troubleshooting the Degradation of TaqVP-pIII Fusion

Degradation of the fusion protein would most likely be caused by proteolysis, but proteolysis of the fusion protein once it is assembled into phage should not occur unless there was contamination or mishandling of the samples. Indeed, these concerns were ruled out by the absence of degraded VHH-pIII fusion protein in Figure 1.3. Therefore, proteolysis was most likely

occurring before phage assembly from bacterial cytoplasmic or periplasmic proteases. This hypothesis draws support from the absence of full-length TaqVP-pIII and the detection of smaller than expected proteins in the bacterial extracts (Figure 1.2) and the phage samples (Figure 1.3).

TaqVP in pET-28a was shown to express well in *E. coli* BL21-Gold (DE3) (Figure 1.2), which is protease-deficient compared to *E. coli* K12 ER2738 (specifically, loss of the periplasmic protease OmpT and cytoplasmic protease Lon). Therefore, the TaqVP-pIII fusion may escape proteolysis if it is expressed in *E. coli* BL21-Gold (DE3). *E. coli* BL21-Gold (DE3) lacks the fertility factor necessary for infection by filamentous phage, so experiments were done to compare the expression of TaqVP in *E. coli* BL21-Gold (DE3) and *E. coli* K12 ER 2738 cells with phagemid pComb3xSS instead of purified phages. Furthermore, western blotting of the *E. coli* K12 ER2738 cell extracts would rule out the possibility that fusion protein was produced in small amounts that were not detectable by Coomassie staining in Figure 1.2. If the experiments indicate that intact fusion protein can be produced by *E. coli* BL21-Gold (DE3), then the fertility factor can be introduced to the strain for subsequent analysis of purified phage.

E. coli BL21-Gold (DE3) does not have an amber stop codon suppressor gene and expresses TaqVP in pComb3xSS only as unfused TaqVP in the periplasm. *E. coli* K12 ER2738 has the suppressor gene and expresses TaqVP in pComb3xSS only as the membrane-bound TaqVP-pIII fusion protein. However, it would be beneficial to test the expression of either TaqVP or the TaqVP-pIII fusion protein in both two strains to determine if the C-terminal residues of pIII have an impact on protein expression and degradation. Therefore, in order to simulate the suppressor and nonsuppressor phenotypes in *E. coli* BL21-Gold (DE3) and *E. coli* K12 ER2738 respectively, pComb3xSS vectors were created with the amber stop codon mutated either to a

glutamine (GLN) or an ochre stop codon (OCH) using the QuikChange Site-Directed Mutagenesis Kit (Agilent).

Total cell extracts of IPTG-induced *E. coli* BL21-Gold (DE3) and K12 *E. coli* ER2738 cells expressing TaqVP in various pComb3xSS vectors were analyzed by western blot (Figure 1.4). In Figure 1.4, BL21:pC3xSS is from bacteria that should produce the unfused TaqVP. A single strong band at ~42 kDa would be expected due to IPTG induction, but the presence of multiple faint bands suggests that the TaqVP was degraded. BL21:pC3xSS-GLN should produce the pIII-TaqVP fusion protein, and two major protein bands were observed. However, these ~130 kDa and ~53 kDa bands are too large or too small to be the correct TaqVP-pIII fusion protein, suggesting that expression of fusion protein was also unsuccessful. ER2738:pC3xSS and ER 2738:pC3xSS-OCH are from bacteria that should produce the pIII-TaqVP fusion protein and unfused TaqVP, respectively, but the detected bands are similar to those seen in BL21:pC3xSS. Overall, the results support the hypothesis that TaqVP-pIII fusion degradation occurs in *E. coli* before phage assembly occurs. Furthermore, expression of unfused or fused TaqVP had no impact on the protein's survival in *E. coli* K12 ER2738, suggesting that degradation is dependent on only TaqVP. The expression tests in *E. coli* BL21-Gold (DE3) also indicate that, even though TaqVP can be produced in the cytoplasm when using the pET-28a vector (Figure 1.2), periplasmic expression with pComb3xSS is unsuccessful for unknown reasons. Although *E. coli* BL21-Gold (DE3) is

deficient in the proteases OmpT and Lon, other proteases and variables may be inhibiting the periplasmic expression of TaqVP and the fusion protein.

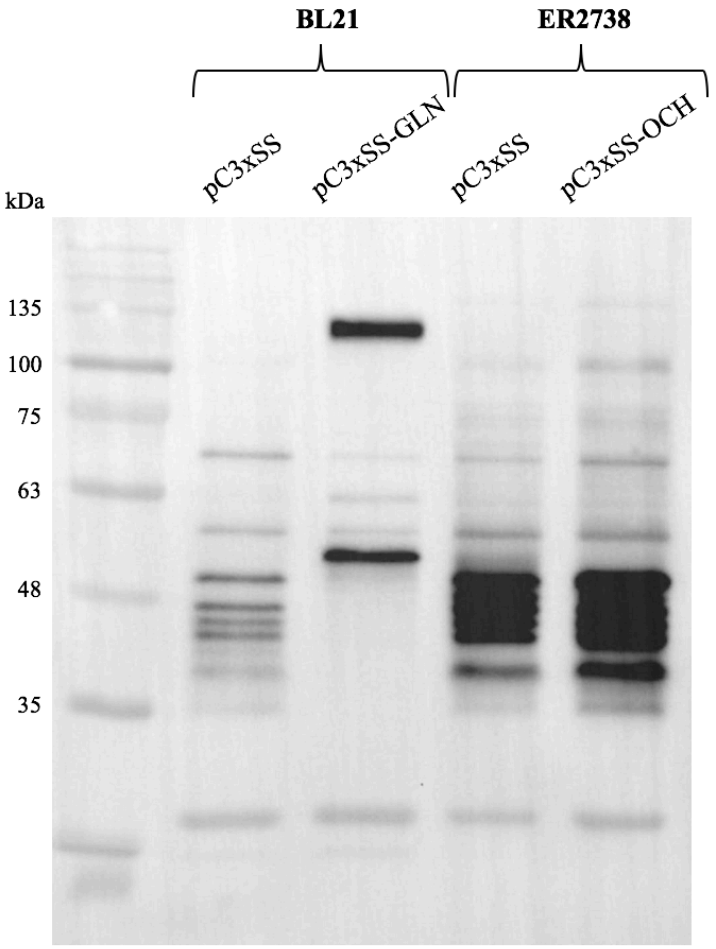

Figure 1.4 Anti-HA western blot of *E. coli* BL21-Gold (DE3) and *E. coli* K12 ER2738 cell lysates expressing periplasmic TaqVP and TaqVP-pIII fusion.

M13 Phage Display of Other DGR VP-pIII Fusions

While difficulties remained in displaying TaqVP-pIII fusion protein on M13 phage, there were two other options for the phage display of a DGR VP: AvpA and TvpA. The genes for AvpA and TvpA were cloned into pComb3xSS phagemids and transformed into *E. coli* K12 ER2738. Phages were purified from infected cultures of *E. coli* K12 ER2738 expressing either AvpA or TvpA in pComb3xSS. The phages were then analyzed by western blot using VHH and WT controls

as previously described (Figure 1.5). As with TaqVP, low molecular weight proteins rather than intact AvpA or TvpA were detected by western blot, indicating that AvpA and TvpA had undergone protein degradation as well.

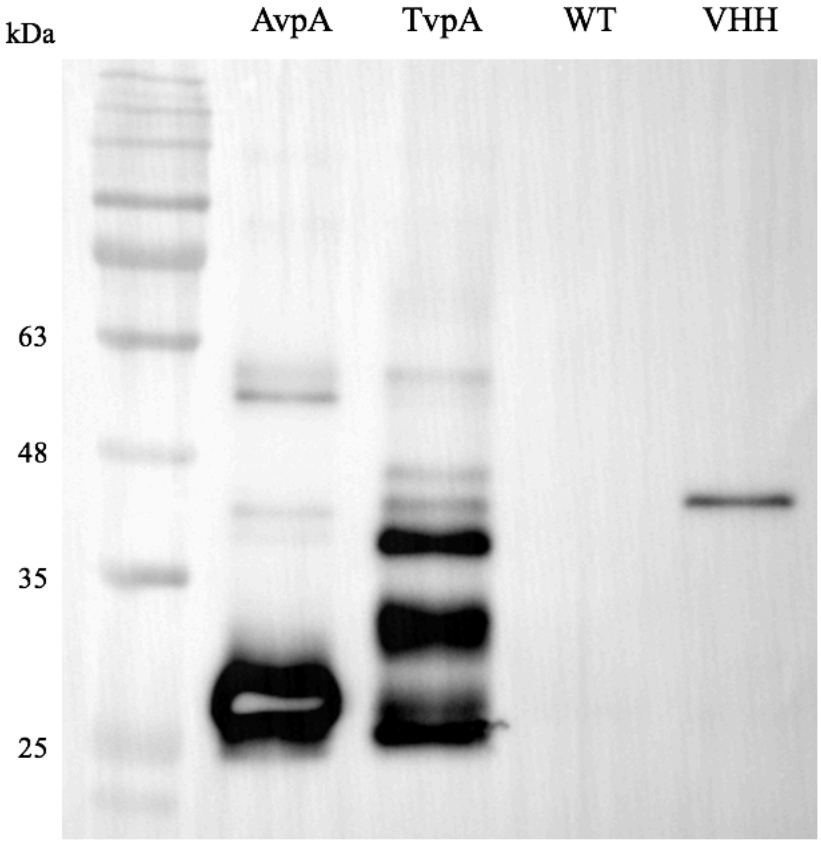

Figure 1.5 Anti-HA western blot of VCSM13 phage containing AvpA-pIII fusion (AvpA), TvpA-pIII fusion (TvpA), VHH-pIII fusion (VHH), or wild-type phage (WT). Multiple low molecular weight bands are apparent for AvpA and TvpA.

1.4 DISCUSSION

In order for the DGR VP filamentous phage display to succeed, the proteins must successfully be translated in the cytoplasm, complete translocation to the periplasm, fold and survive in the periplasmic environment, and finally integrate into phage capsids. The DGR VPs can be easily expressed in the *E. coli* cytoplasm, but it is unknown if they can be expressed in the periplasmic environment.

The first obstacle to periplasmic expression is translocation across the inner membrane. The two major translocation systems in *E. coli* are the general secretion pathway (Sec) and the twin-arginine translocation pathway (Tat). The Sec pathway primarily secretes unfolded proteins, while the Tat pathway primarily secretes folded proteins (Green et al., 2016). Secretion of pIII-fusion protein by the OmpA signal sequence occurs through the Sec system, presumably in a secretion-competent unfolded state (Barbas et al., 2001; Natale et al., 2008). Secretion may be hindered if a protein folds into stable conformations quickly before it completes the secretion process (Huber et al., 2005). However, the detection of the HA tag in DGR-VP phage samples (Figure 1.3 and Figure 1.5) suggests that the DGR-VP fusion proteins are secretion-competent and are being integrated into phage capsids. If the fusion proteins were not secretion-competent then the HA tag would not be detected on the phages, as seen with the wild type phage samples. Thus, if translocation and integration are occurring, the only obstacle remaining is folding and surviving in the periplasm.

After the pIII fusion protein is secreted to the periplasm, the OmpA signal sequence is cleaved off and the N-terminus of the fusion (DGR VP) is exposed to the periplasmic space while the C-terminus (pIII) remains anchored to the inner membrane (Barbas et al., 2001; Davis et al., 1985). The DGR VPs must then fold correctly in the periplasmic environment and avoid being targeted by periplasmic proteases. Production of high levels of recombinant protein in the periplasm is known to cause misfolding and proteolysis (Miot and Betton, 2004). However, phage display using pComb3xSS relies on the production of low amounts of protein. As mentioned in the description of the pComb3xSS vector in the Materials and Methods, transcription of the fusion protein is regulated in the pComb3xSS phagemid by the *lac* operon. The *lac* operon is "leaky," and basal transcription levels of the operon yield sufficient fusion protein production for phage display. Relying on the leaky transcription of the *lac* operon is beneficial because it avoids

overwhelming the Sec translocational machinery with abundant amounts of protein and prevents overcrowding of the periplasmic space. Both of these factors decrease the likelihood of the fusion proteins misfolding, eliciting a bacterial stress response, or recruiting periplasmic quality control proteases (Barbas et al., 2001). Despite these considerations, all of the DGR VP phage samples exhibit fusion protein degradation, as seen in Figure 1.3 and Figure 1.5.

The results for TaqVP suggest that the protein degradation occurred in the periplasm. Whether in *E. coli* K12 ER2738 or in the protease-deficient (Δ*ompT* and Δ*lon*) *E. coli* BL21-Gold (DE3) strain, periplasmic expression of unfused and fused TaqVP failed to produce a large amount of protein at the correct molecular weight, despite the previously successful cytoplasmic expression of TaqVP in *E. coli* BL21-Gold (Figure 1.2 and Figure 1.4). These results also indicate that the difficulty in expressing TaqVP in the periplasm is not OmpT- or Lon-dependent and is not affected by the presence or absence of a membrane anchor (in pIII). Since full-length protein was not observed in either induced (Figure 1.2 and Figure 1.4) or non-induced (Figure 1.3) samples, the results also suggest that protein expression levels did not affect the periplasmic expression of TaqVP. TaqVP could exist in three different states in the periplasm: soluble folded protein, insoluble inclusion bodies, or proteolytically degraded protein, with the relative proportions being dependent on the kinetics of folding, aggregation, and degradation (Betton et al., 1998) Overall, it remains unknown whether the degradation is a result of TaqVP misfolding in the periplasm, if TaqVP's sequence is targeted by periplasmic proteases regardless of its folded state, or some other variable. OmpT is only one of many periplasmic proteases identified in *E. coli*, and it may not have a major role in the proteolysis of TaqVP compared to other proteases, such as DegP (Miot and Betton, 2004). *E. coli* strains deficient in multiple proteases have been developed and could be tested for DGR VP phage display, but deficiencies in multiple proteases negatively impact bacterial growth and survival and promote the formation of inclusion bodies; these disadvantages

eliminate any advantages such strains may have for phage display (Barbas et al., 2001). Thus, the filamentous phage display of DGR VPs currently appears to be unfeasible.

CHAPTER 2: RIBOSOME DISPLAY

2.1 INTRODUCTION

Compared to filamentous phage display, ribosome display is a more recently developed technology with some potential advantages over the former. In ribosome display, the to-be-displayed protein's gene is inserted upstream of a flexible linker for expression as a ribosome-protein-mRNA ternary complex. The ternary complex represents the phenotype-genotype linkage required for molecular display, and it is formed during translation. When the ribosome translates the nascent protein from its mRNA, the ribosome undergoes translational arrest near the C-terminal end of the protein sequence. The C-terminus of the nascent protein and its mRNA remain attached to the ribosome, and the flexible linker allows the recombinant protein to clear the ribosome exit tunnel and fold. Thus, the basic components for ribosome display are genetic template, transcriptional and translational machinery, and a mechanism to induce translational arrest in the ribosome (Hanes and Plückthun, 1997).

Ribosomal stalling for ribosome display can be accomplished by omitting a stop codon in the genetic template, excluding translational release factors in the reaction, or inserting a stalling peptide sequence at the C-terminal end of the nascent protein (Hanes and Plückthun, 1997; Evans et al., 2005; Shimizu et al., 2001). Any single one of the aforementioned methods is sufficient for stalling, although redundant stalling mechanisms may increase the likelihood of successful ribosome display (Ohashi et al., 2007). Multiple known peptide sequences can induce ribosomal stalling, although canonically the stalling sequence (aa 150-166) of the *E. coli* secretion monitor protein (SecM) has been used in ribosome display experiments (Nakatogawa and Ito, 2002; Tenson

and Ehrenberg, 2002). Presumably it is because SecM-mediated ribosomal stalling is accomplished by the peptide sequence itself, whereas other stalling peptide sequences require a co-effector to accomplish stalling, such as arginine with the *Saccharomyces cerevisiae* arginine attenuator peptide, or tryptophan with the *E. coli* tryptophanase leader peptide (Gong et al., 2001; Nakatogawa and Ito, 2001; Wang et al., 1997). In addition to a stalling mechanism, the length of the linker sequence can affect ribosome stalling as well. A short linker sequence may not provide enough distance from the ribosome exit tunnel for the nascent protein to fold without hindrance. Furthermore, if the nascent protein can fold despite a short linker sequence, the mechanical forces applied by the nascent protein folding against or near the ribosome exit tunnel can release the nascent chain from the stalled ribosome (Goldman et al., 2015). The exact length of the linker sequence required to prevent this outcome is most likely dependent on individual protein sequences and folding kinetics, but a length of about 40 amino acids (including the length of a stalling sequence) was sufficient to prevent release of Top7, a protein that folds rapidly and stably (Goldman et al., 2015).

In vitro ribosome display can be accomplished using *E. coli* S30 extract or the PURE system. Ribosome display was first accomplished using cellular extracts from *E. coli,* but these can be problematic as they run the risk of contamination with inhibitory DNases, RNases, proteases, translational release factors, and transfer messenger RNA (Hanes and Plückthun, 1997). In bacteria, transfer messenger RNA (tmRNA) is a product of the *ssrA* gene, and it rescues and recycles stalled ribosomes (Keiler et al., 1996). Synthetic oligonucleotides that are anti-sense to and neutralize the tmRNA must be added to cell extracts to prevent stalled ribosome rescue (Hanes et al., 2000). A more recent method uses individually purified recombinant components, known as the Protein synthesis Using Recombinant Elements (PURE) system (Shimizu et al., 2001). Ribosome display using the PURE system minimizes the risks associated with using cell extracts

and allows for ribosome display in a more controlled environment. Depending on the quality of the purified components, PURE ribosome display should be devoid of DNases, RNases, proteases, release factors, and tmRNA. Furthermore, the exact concentrations of individual components can be optimized and scaled. Therefore, for the ribosome display of DGR VPs, the PURE system was chosen for its lower risk of contamination and greater customizability.

The New England Biolabs PURExpress® system is a commercial transcription and translation kit based on the original PURE system. The kit contains two reaction mixtures, mixtures A and B, and can be used with release factors, amino acids and tRNA, and ribosomes separate from the mixtures for customizable reactions. Circular DNA, linear DNA, or mRNA can be used as genetic templates for ribosome display, as long as the template contains a T7 promoter, ribosome binding site, start codon, linker sequence, and ribosome stalling mechanism(s).

There are a few significant advantages of PURE ribosome display compared to phage display. Since PURE ribosome display is completely *in vitro*, it eliminates the obstacle of periplasmic secretion in phage display. Furthermore, ribosome display in general can sample a significantly larger number of distinct sequences from a library compared to what phage display can accomplish. For example, construction of a phagemid library involves a bacterial transformation step, which limits the number of distinct sequences sampled in a round of phage display to around 10^9-10^{10} distinct transformants (Schaffitzel et al., 1999). Ribosome display on the other hand is limited only by the number of ribosomes in a reaction. The concentration of ribosomes in a PURExpress® reaction is approximately 2 μM, which translates to about 10^{12} ribosomes per μL of reaction, making ribosome display orders of magnitudes more efficient than phage display (New England Biolabs, Schaffitzel et al., 1999). Thus, ribosome display of DGR VPs holds greater potential than phage display.

2.2 MATERIALS AND METHODS

Construction of the Ribosome Display Genetic Template

The DGR VP genes were PCR amplified (see: Chapter 1.2 DGR VP Cloning) and inserted into pET-17b using NdeI and BamHI restriction sites. The sequence encoding amino acids 220-326 of pIII was PCR amplified from whole M13 phages from the phage display project to be used as the spacer sequence for ribosome display to replicate previous PURE ribosome display experiments (Ohashi et al., 2007). The SecM stalling sequence (aa 148-170) was amplified from *E. coli* BL21-Gold (DE3) using colony PCR. The spacer and SecM sequences were then spliced using overlap extension PCR and cloned into pET-17b using BamHI and XhoI restriction sites. The structure of the resulting genetic templates are outlined in Figure 2.1, and the templates are called "DGR VP-GS" for the inserted gene III and SecM sequences. For Ni-NTA affinity selection, a sequence encoding a 6xHis tag was inserted before the sequence encoding the N-terminus of the DGR VP through PCR.

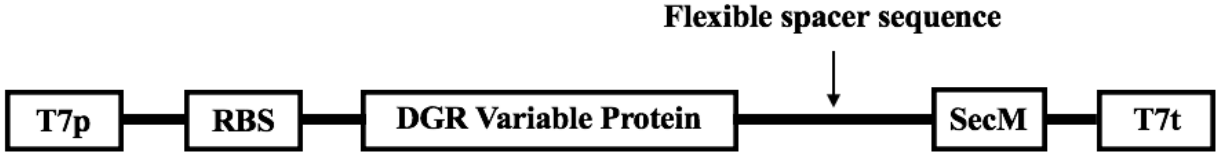

Figure 2.1 Schematic of the genetic template structure used for the *in vitro* transcription/translation of DGR VPs. T7p and T7t are the T7 promoter and terminator sequences, respectively, and RBS is the ribosome binding site. Original figure adapted from Fig. 2A of "Efficient protein selection based on ribosome display system with purified components." 2006. Ohashi et al.

Preparation of the Genetic Template

High-quality DNA templates are necessary for successful *in vitro* transcription/translation. Plasmid templates used in PURExpress reactions were purified using the transfection-grade QIAGEN Plasmid Mini kit to minimize RNase and DNase contamination. Linear DNA templates used in PURExpress reactions were PCR amplified from plasmid using T7-promoter and T7-

terminator primers and electrophoresed through a 2% agarose gel. If the PCR resulted in single amplicons, the original PCR samples were purified using the Wizard® SV Gel and PCR Clean-Up System kit and used directly for *in vitro* transcription/translation. In the case of multiple amplicons, gel extraction of the correct band using the same Wizard® kit was sufficient.

In Vitro Transcription/Translation Using the PURE System

In vitro transcription/translation reactions were performed using the PURExpress® ΔRF123 Kit (NEB). Unless otherwise stated, typical reactions did not contain the release factors RF 1, 2, or 3. DNA template concentrations in the reactions were 10 ng/μL for both circular and linear templates. Reactions were carried out according to the kit's specifications and supplemented with 2 units/μL of Murine RNase inhibitor. Samples for subsequent Ni-NTA affinity purification were immediately quenched after reacting for 30 min by diluting to 100 μL with ice-cold dilution buffer (1x TBS, pH 7.5 with 50 mM magnesium acetate and 0.1% Tween 20). Otherwise, reactions were allowed to complete for 2 h and analyzed by SDS-PAGE and western blotting (see Chapter 1.2 SDS-PAGE and Western Blot).

Ni-NTA Affinity Purification

Two hundred μL of a 1:1 slurry of HisPur Ni-NTA Resin (Thermo Fisher) equilibrated in washing buffer (1x TBS, pH 7.5 with 50 mM magnesium acetate and 10 mM imidazole) was added to the diluted reactions and nutated at 4°C for 1 h. The mixture was then applied to empty Micro Bio-Spin™ Columns (Bio-Rad) and centrifuged for 1 min at 800 x *g*. The column was washed four times with 500 μL of washing buffer and eluted with 200 μL of elution buffer (1x TBS, pH 7.5 with 50 mM magnesium acetate and 300 mM imidazole), centrifuging for 1 min at 800 x *g* each

time. During each step, the beads were resuspended by aspirating 3-4 times with a pipette to ensure adequate equilibration before centrifuging.

Reverse Transcription PCR

Fifty μL of the eluted samples were purified using the PureLink RNA Mini Kit (Thermo Fisher Scientific). The purified mRNA samples were treated with 2 units of TURBO™ DNase (Thermo Fisher Scientific) in 1x TURBO DNase buffer for 30 minutes at 37°C. DNase treated samples were purified again using the PureLink RNA Mini Kit and subjected to one step RT-PCR using the PrimeScript™ One Step RT-PCR Kit Ver. 2 (Takara Bio). Cycling conditions were according to the kit's standard recommendations.

2.3 RESULTS

SDS-PAGE and Western Blot of *In Vitro* Transcribed/Translated DGR VPs

In order to determine if DGR VPs could be produced using the PURE system, 10 μL PURExpress reactions were performed using DGR VPs in pET-28a. Five μL of the reactions were analyzed through SDS-PAGE (Figure 2.2). The pET-28a did not include the gene III or SecM sequences. Release factors 1, 2, and 3 were included in these reactions to maximize protein production and visualization. Upon staining the gel with Coomassie Brilliant Blue, TvpA and TaqVP can be seen as faint bands in their respective lanes alongside the PURExpress kit components. AvpA did not appear to be produced in amounts detectable by Coomassie staining. In order to detect AvpA by western blot, another set of 10 μL PURExpress reactions were carried out using linear templates containing gene III and SecM (AvpA-GS and TaqVP-GS). The inclusion of gene III allows for immunodetection with anti-pIII antibody. The western blot showed that AvpA can be produced by the PURE system (Figure 2.3). Having tested the expression of all three

DGR VP candidates, TaqVP was chosen for testing the affinity selection and RT-PCR steps of ribosome display because TaqVP appears to be produced at greatest abundance in the PURE system.

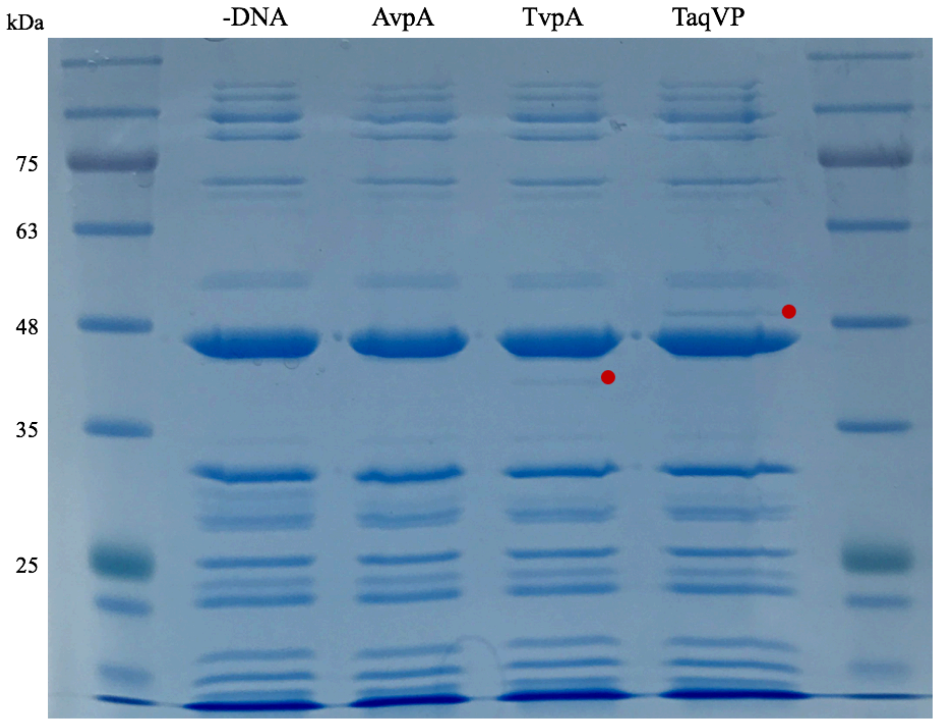

Figure 2.2 SDS-polyacrylamide gel electrophoresis of the *in vitro* transcription/translation of DGR VPs in pET-28a. The leftmost lane is a reaction that did not contain plasmid DNA. The bands corresponding to TvpA and TaqVP are designated with a red dot in their respective lanes.

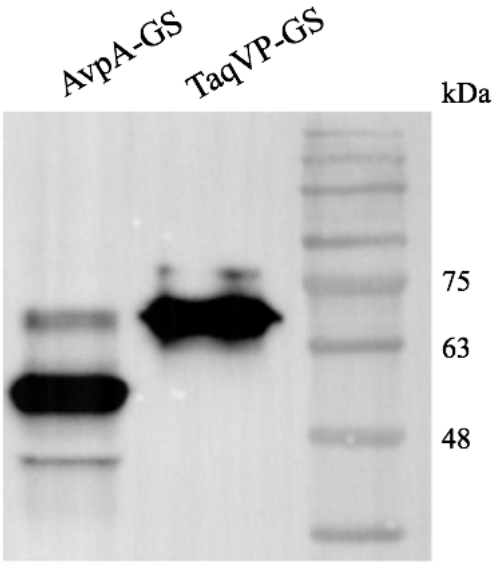

Figure 2.3 Anti-pIII western blot of *in vitro* transcribed/translated AvpA-geneIII-SecM (AvpA-GS) and TaqVP-geneIII-SecM (TaqVP-GS).

Affinity Purification of 6xHis-Tagged TaqVP-GS

Twenty-five μL PURExpress reactions of 6xHis-tagged TaqVP-GS and non-tagged TaqVP-GS were affinity selected using Ni-NTA agarose columns and analyzed by western blot (Figure 2.4). Since there is currently no target ligand to select for binding with TaqVP, this affinity purification is a stand-in for the affinity selection step in molecular display. In other words, the Ni-NTA column is the target ligand, and the 6xHis tag is the selection factor. The western blot demonstrates that background binding of non-tagged TaqVP-GS to the column was eliminated while retaining 6xHis-tagged TaqVP-GS.

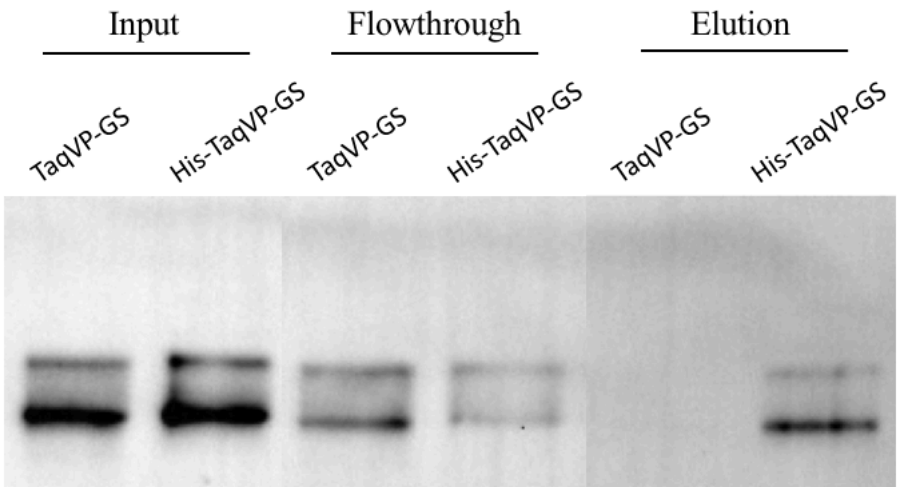

Figure 2.4 Western blot of the affinity-selected TaqVP-GS. Ten percent of the input, flow through, and elution volumes were loaded.

Reverse Transcription PCR of the TaqVP Gend

Eluted samples of non-tagged TaqVP-GS and 6xHis-tagged TaqVP-GS were serially diluted and the TaqVP gene was amplified by RT-PCR using gene-specific primers. As a control for DNA contamination from the Ni-NTA agarose column, duplicate reactions were done with DNase-treated and PureLink RNA Mini Kit-purified samples. PCR-amplified samples were electrophoresed through a 2% agarose gel (Figure 2.5). Amplification of the TaqVP gene (1.2 kb) can be seen in DNase treated and untreated samples. In addition, more amplification was observed for the His-tagged sample compared to the non-tagged sample, suggesting that greater quantities of His-tagged TaqVP mRNA were recovered from the Ni-NTA affinity purification step than the non-tagged TaqVP mRNA. This result demonstrates that the amount of recovered mRNA is dependent on the protein sequence it encodes, specifically the 6xHis-tag, which indicates the presence of a phenotype-genotype linkage and the successful ribosome display of TaqVP.

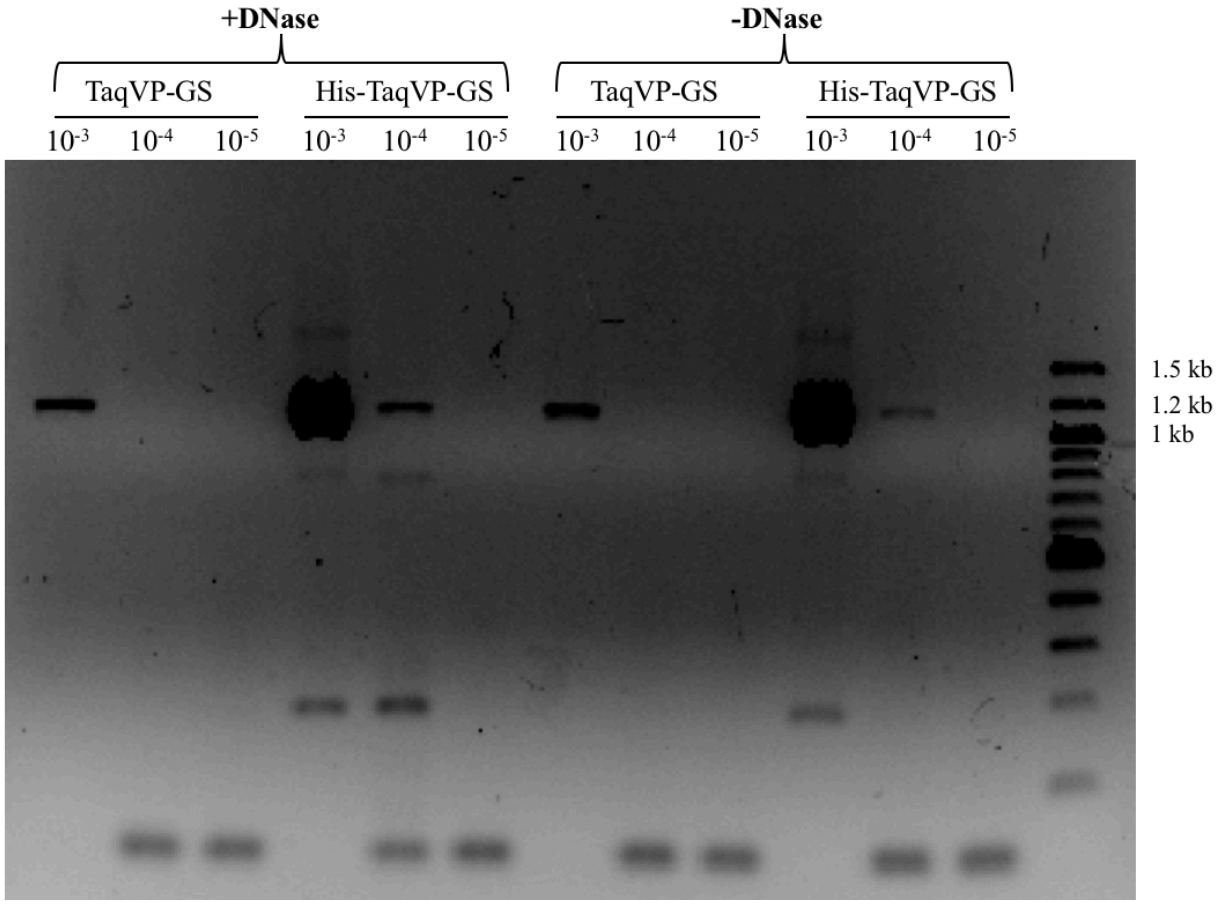

Figure 2.5 Reverse-transcription PCR of DNase-treated and -untreated affinity-purified ternary complexes.

2.4 DISCUSSION

While the phage display of DGR VPs did not bear fruit, ribosome display shows promise in developing a high-throughput selection method of DGR VPs. As seen in Figure 2.2 and Figure 2.3, all three DGR VP candidates can be expressed through *in vitro* transcription/translation. Furthermore, Figure 2.4 and Figure 2.5 together suggest that the ribosome display of TaqVP was successful. The Ni-NTA affinity purification resulted in the recovery of greater amounts of 6xHis-tagged TaqVP protein and its mRNA compared to non-tagged TaqVP protein and its mRNA. This relationship suggests that the protein and the mRNA are physically linked to each other, which can

only be accomplished by the formation of the ribosome-nascent protein-mRNA ternary complex. Thus, the Ni-NTA affinity purification of the 6xHis-tagged TaqVP protein and mRNA indicates that the phenotype-genotype linkage of TaqVP was successful and that TaqVP can be used for ribosome display.

The next step in the biotechnological development of DGR VPs is the construction of a DNA library of DGR VP mutants. Since ribosome display can use linear DNA as the genetic template, a DGR VP library can be constructed through PCR. DGR VPs contain naturally variable residues that are the result of adenine-specific mutations introduced by the DGR. Replicating the adenine mutations in these naturally variable residues is advantageous for generating a DGR VP library because it lowers the risks of introducing deleterious substitutions to the protein sequence. These mutations can be done through PCR by using oligonucleotide primers containing A->N substitutions at the relevant codons.

In addition to the construction of a DGR VP library, further optimization of the ribosome display process is necessary. For example, Figure 2.4 suggests that nonspecific binding of non-tagged TaqVP protein is eliminated after washing. However, the more sensitive RT-PCR in Figure 2.5 reveals that nonspecific binding of non-tagged TaqVP nucleic acid remains, approximately at a level one-tenth that of the 6xHis-tagged construct. These nucleic acids are most likely mRNA transcripts produced during the *in vitro* transcription/translation reaction that are unassociated with a ternary complex. Ideally, the background should be reduced as much as possible to maximize the enrichment of interacting protein sequences during repeated ribosome display cycles. There are some concerns that more stringent washing conditions may destabilize the protein-ribosome-mRNA ternary complex, though the ternary complex is surprisingly stable in the PURE reaction, and experiments have shown that the complex is stable from 4-50 °C (Matsuura et al., 2007).

This project's goal was to examine the feasibility of the phage and ribosome display of diversity-generating retroelement variable proteins. In conclusion, phage display of DGR VPs was found to be problematic due to the degradation of DGR VPs, while ribosome display offers a viable strategy for the molecular display of DGR VPs.

REFERENCES

Arambula D, Wong W, Medhekar BA, Guo H, Gingery M, Czornyj E, Liu M, Dey S, Ghosh P, Miller JF. Surface display of a massively variable lipoprotein by a Legionella diversity-generating retroelement. PNAS. 2013;110(20):8212-8217.

Barbas CF, Burton DR, Scott JK, Silverman GJ. Phage Display: A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 2001.

Betton JM, Sassoon N, Hofnung M, Laurent M. Degradation versus aggregation of misfolded maltose-binding protein in the periplasm of *Escherichia coli*. Journal of Biological Chemistry. 1998; 273:8897-8902.

Bever CR, Majkova Z, Radhakrishnan R, Suni I, McCoy M, Wang Y, Dechant J, Gee S, Hammock BD. Development and utilization of camelid VHH antibodies from alpaca for 2,2',4,4'-tetrabrominated diphenyl ether detection. Analytical Chemistry. 2014;86(15):7875-7882.

Binz HK, Amstutz P, Plückthun A. Engineering novel binding proteins from nonimmunoglobulin domains. Nature Biotechnology. 2005;23(10):1257-1268.

Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. Journal of Molecular Biology. 2003;332(2):489-503.

Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD. VLR-based adaptive immunity. Annual Review of Immunology. 2012;30:203-220.

Bradbury ARM, Marks JD. Antibodies from phage antibody libraries. Journal of Immunological Methods. 2004;290(1-2):29-49.

Bukowska MA, Grutter MG. New concepts and aids to facilitate crystallization. Current Opinion in Structural Biology. 2013;23(3):409-416.

Davis NG, Boeke JD, Model P. Fine structure of a membrane anchor domain. Journal of Molecular Biology. 1985;181:111-121.

Doulatov S, Hodes A, Dai L, Mandhana N, Liu M, Deora R, Simons RW, Zimmerly S, Miller JF. Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. Nature. 2004;431(7007):476-81.

Evans MS, Ugrinov KG, Frese MA, Clark PL. Homogenous stalled ribosome nascent chain complexes produced *in vivo* or *in vitro*. Nature Methods. 2005;2:757-762.

Galán A, Comor L, Horvatić A, Kuleš J, Guillemin N, Mrljak V, Bhide M. Library-based display technologies: Where do we stand? Molecular BioSystems. 2016;12:2342-2358.

Gong F, Ito K, Nakamura Y, Yanofsky C. The mechanism of tryptophan induction of tryptophanase operon expression: Tryptophan inhibits release factor-mediated cleave of TnaC-peptidyl-tRNA^{pro}. PNAS. 2001;98(16):8997-9001.

Green ER, Mecsas J. Bacterial secretion systems: An overview. Microbiology Spectrum. 2016;4(1):VMBF-0012-2015.

Griffin L, Lawson A. Antibody fragments as tools in crystallography. Clinical & Experimental Immunology. 2011;165(3):285-291.

Hanes J, Jermutus L, Plückthun A. Selecting and evolving functional proteins *in vitro* by ribosome display. Methods in Enzymology. 2000;328:404-430.

Hanes J, Plückthun A. In vitro selection and evolution of functional proteins by using ribosome display. PNAS. 1997;94(10):4937-4942.

Hosse RJ, Rothe A, Barbara PE. A new generation of protein display scaffolds for molecular recognition. Protein Science. 2006;15(1):14-27.

Huber D, Cha M, Debarbieux L, Planson AG, Cruz N, López G, Tasayco ML, Chaffotte A, Beckwith J. A selection for mutants that interfere with folding of *Escherichia coli* thioredoxin-1 *in vivo*. PNAS. 2005;102(52):18872-18877.

Huber T, Steiner D, Röthlisberger D, Plückthun A. *In vitro* selection and characterization of DARPis and Fab fragments for the co-crystallization of membrane proteins: The Na^+-citrate symporter CitS as an example. Journal of Structural Biology. 2007;150:206-221.

Keiler KC, Waller PR, Sauer RT. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science. 1996;271(5251):990-993.

Kieke MC, Shusta EV, Boder ET, Teyton L, Wittrupt KD, Kranz DM. Selection of functional T cell receptor mutants from a yeast surface-display library. PNAS. 1999;96(10):5651-5656.

Kohl A, Binz HK, Forrer P, Stumpp MT, Plückthun A, Grütter MG. Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. PNAS. 2003;100(4):1700-1705.

Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S. High-affinity single-domain binding proteins with a binary code. PNAS. 2007;104(16):6632-6637.

Koide S. Engineering of recombinant crystallization chaperones. Current Opinion in Structural Biology. 2009;19(4):449-457.

Kovari LC, Momany C, Rossmann MG. The use of antibody fragments for crystallization and structure determinations. Structure. 1995;3:1291–1293.

Le Coq J, Ghosh P. Conservation of the C-type lectin fold for massive sequence variation in a Treponema diversity-generating retroelement. PNAS. 2011;108(35):14649-14653/

Levin AM and Weiss GA. Optimizing the affinity and specificity of proteins with molecular display. Molecular Biosystems. 2006;2:49-57.

Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, Miller JF. Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. Science. 2002;295(5562):2091-2094.

Loris R, Marianovsky I, Lah J, Laeremans T, Engelberg-Kulka H, Glaser G, Muyldermans S, Wyns L. Crystal structure of the intrinsically flexible addiction antidote MazE. Journal of Biological Chemistry. 2003;278:28252-28257.

Matsuura T, Yanagida H, Ushioda J, Urabe I, Yomo T. Nascent chain, mRNA, and ribosome complexes generated by a pure translation system. Biochemical and Biophysical Research Communications. 2007;352(2):372-377.

McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: Filamentous phage displaying antibody variable domains. Nature. 1990;348(6301):552-554.

McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallographica Section F, Structural Biology Communications. 2014;70(Pt 1):2-20.

Miller JL, Le Coq J, Hodes A, Barbalat R, Miller JF, Ghosh P. Selective ligand recognition by a diversity-generating retroelement variable protein. PLOS Biol. 2008;6(6):e131.

Miot M, Betton JM. Protein quality control in the bacterial periplasm. Microbial Cell Factories. 2004.

Nakatogawa H, Ito K. Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. Molecular Cell. 2001;7(1):185-192.

Nakatogawa H, Ito K. The ribosomal exit tunnel functions as a discrimination gate. Cell. 2002;108(5):629-636.

Natale P, Brüser T, Driessen AJM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. Biochimica et Biophysica Acta – Biomembranes. 2008;1778(9):1735-1756.

Ohashi H, Shimizu Y, Ying BW, Ueda T. Efficient protein selection based on ribosome display system with purified components. Biochemical and Biophysical Research Communications. 2007;352:270-276.

Ostermeier C, Harrenga A, Ermler U, Michel H. Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody F_V fragment. PNAS. 1997;94(20):10547-10553.

Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature. 2004;430(6996):174-180.

Parmley SF and Smith GP. Antibody-selectable filamentous fd phage vectors: Affinity purification of target genes. Gene. 1988;73(2):305-318.

Paul BG, Bagby SC, Czornyj E, Arambula D, Handa S, Sczyrba A, Ghosh P, Miller JF, Valentine DL. Targeted diversity generation by intraterrestrial archaea and archaeal viruses. Nature Communications. 2015;6:6585.

Plückthun A, Alternative Scaffolds: Expanding the Options of Antibodies. In: *Recombinant Antibodies for Immunotherapy*, Melvyn Little, Cambridge University Press, New York (2009), p. 244-271.

Proba K, Wörn A, Honegger A, Plückthun A. Antibody scFv fragments without disulfide bonds made by molecular evolution. Journal of Molecular Biology. 1998;275(2):245-253.

Schaffitzel C, Hanes J, Jermutus L, Plückthun A. Ribosome display: An in vitro method for selection and evolution of antibodies from libraries. Journal of Immunological Methods. 1999;231:119-135.

Sergeeva A, Kolonin MG, Molldrem JJ, Pasqualini R, Arap W. Display technologies: application for the discovery of drug and gene delivery agents. Advanced Drug Delivery Reviews. 2006;58(15):1622-1654.

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T. Cell-free translation reconstituted with purified components. Nature Biotechnology. 2001;19:751-755.

Shusta EV, Holler PD, Kieke MC, Wittrup KD. Directed evolution of a stable scaffold for T-cell receptor engineering. Nature Biotechnology. 2000;18(7):754-759.

Smith GP. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the viral surface. Science. 1985;228(4705):1315-1317.

Tasumi S, Velikovsky AC, Xu G, Gai AS, Wittrup DK, Flajnik MF, Mariuzza RA, Pancer Z. High-affinity lamprey VLRA and VLRB monoclonal antibodies. PNAS. 2009;106(31):12891-12896.

Tenson T, Ehrenberg M. Regulatory nascent peptides in the ribosomal tunnel. Cell. 2002;108(5):591-594.

Uysal S, Vásquez V, Tereshko V, Esaki K, Fellouse FA, Sidhu SS, Koide S, Perozo E, Kossiakoff A. Crystal structure of full-length KcsA in its closed conformation. PNAS. 2009;106(16):6644-6649.

Van Wezenbeek PM, Hulsebos TJ, Schoenmakers JG. Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: Comparison with phage fd. Gene. 1980;11(1-2):129-148

Veesler D, Dreier B, Blangy S, Lichière J, Tremblay D, Moineau S, Spinelli S, Tegoni M, Plückthun A, Campanacci V, Cambillau C. Crystal structure and function of a DARPin neutralizing inhibitor of lactococcal phage TP901-1: Comparison of DARPin and camelid VHH binding mode. Journal of Biological Chemistry. 2009;284(44):30718-30726.

Wang Z and Sachs MS. Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*. Molecular and Cellular Biology. 1997;17(9):4904-4913.

Weatherill EE, Cain KL, Heywood SP, Compson JE, Heads JT, Adams R, Humphreys DP. Towards a universal disulphide stabilised single chain FV format: Importance of interchain disulphide bond location and vL-vH orientation. Protein Engineering, Design and Selection. 2012;25(7):321-329.