A Mechanistic and Chemistry-Focused Approach Towards the Development of Novel Covalent Binding Cyclic Phage Libraries

A thesis by
Vincent Nobile

A thesis

submitted to the Faculty of

the Boston College department of Chemistry

in partial fulfillment

of the requirements for the degree of

Master of Science

Boston College
Morrissey College of Arts and Sciences
Graduate School
December 2022

A Mechanistic and Chemistry-Focused Approach Towards the Development of Novel Covalent Binding Cyclic Phage Libraries

Vincent Nobile

Advisor: Professor Jianmin Gao, Ph.D.

Covalent drugs present a unique situation in the clinical world. Formation of a covalent bond between a drug molecule and its target protein can lead to significant increases in a number of desirable traits such as residence time, potency, and efficacy of a drug. From a kinetic perspective, the formation of a covalent bond between a drug and its target functionally eliminates the dissociation rate (k_{off}) of the compound, ensuring that the compound will stay engaged with its target. However, development of covalent drugs has been met with caution and concern, as an irreversible covalent bond forming on the wrong target can have disastrous results, so specificity is of the utmost importance. One option for increasing specificity is by linking a covalent binding electrophile, or warhead, to a peptide. Peptide-based therapeutics have already been shown to serve as effective protein-targeting modalities with high specificity, a specificity that would greatly benefit covalent drugs.

Phage display is a powerful technique for the discovery of selective peptides which utilizes the screening of vast libraries of randomized peptides to identify strong binders. This technology has been used to discover a large number of protein-targeting peptides, but also a smaller number of cyclic, covalent binding peptides that function as enzymatic inhibitors. Herein, this study aimed to explore the idea of adding covalent-binding functionality to phage libraries in novel ways and expand upon the scope of proteins that can be targeted with phage libraries containing covalent libraries. We sought to develop a mechanistic and chemical understanding of the interactions between

bacteriophage and chemical warheads to best understand both the limits and the potential of this technology.

In order to best understand the relationship between chemical warhead and phage particle, a model system was developed based on the M13KE pIII protein. It was found that the extracellular N-terminal domains of this protein could be expressed and purified in low yields in bacterial cells and that these domains would behave similarly in solution as in the membrane of the M13KE bacteriophage. With this protein in hand, experiments previously performed using small, cysteine containing peptides, could be performed on a full protein to mimic the phage labeling environment. This protein was used to identify efficient cysteine crosslinkers, most notably dichloroacetone (DCA) and bis-chlorooxime (BCO). The pIII protein system was then used to study the viability of bifunctional warhead molecules containing a covalent warhead and a cysteine crosslinker.

Based on preliminary analyses with the pIII protein, aryl sulfonyl fluoride was chosen as a novel warhead candidate that warranted further pursuit. Kinetic NMR studies verified that aryl sulfonyl fluoride was capable of forming covalent bonds with phenols under phage labeling conditions. Labeling experiments analyzed with LC/MS seemed to indicate a degradation of the warhead. However, as the source of the degradation was not able to be determined, it was decided that various affinity assays would be used to identify if phage could be labeled with an aryl sulfonyl fluoride-DCA conjugate. Both streptavidin-bead pulldown assays and ELISA assays were used, however both assays yielded results that could not conclusively verify the integrity of the warhead.

During phage labeling experiments, a phenomenon was noted that phage titers after modification showed a 2-3 order of magnitude drop in phage count. Covalent

modification of phage beyond what is intended could have troubling consequences for all covalent phage libraries, and so a more in-depth approach was taken to identify and better understand phage toxicity as it relates to covalent warheads. As a model, a well-studied diazaborine-mediated warhead with a slow dissociation rate was selected and used in a range of phage toxicity screenings. Despite statistical fluctuations between trials, toxicity screenings using this warhead served to highlight a unique concern for bifunctional covalent warheads. A concentration-dependent toxicity can be seen in phage incubated with bifunctional small molecules that is not present when incubated with the monofunctional equivalents. The presence of this toxicity even towards a phage with no free thiols highlights a unique challenge of off-target labeling within phage particles that, if solved, could provide the next significant step towards developing novel covalent phage libraries.

To my family, friends, and Meg,
Thank you for your support when I needed it most,
I could not have done this without you.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Jianmin Gao for his support, effort, and commitment to seeing me grow as a scientist, as a student, and as a person. You have been a great mentor and role model who has allowed me to grow so much and for that I will always be grateful.

I would also like to thank the members of my committee, Dr. Jia Niu and Dr. Huiqing (Jane) Zhou for taking the time to read this thesis and for their continued support and feedback on my work.

Thank you to all the current and former members of the Gao lab. You all taught me so much I am so grateful that I could call you all coworkers and friends.

Thank you to my family, and especially my mom for always encouraging me to try my hardest and for teaching me to always keep learning.

I would especially like to thank the most important person in my life: Meg. You stuck with me for this entire journey, and I could not have done this without you. Thank you for everything you do for me, I love you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

ABBREVIATIONS AND ACRONYMS

AmpR	Ampicillin resistance
amu	Atomic mass units
APBA	2-acetylphenylboronic acid
BCO	Bis-chlorooxime
BSA	Bovine serum albumin
BSF	Benzene sulfonyl fluoride
BTyr	Biotin Tyramide
C, Cys	Cysteine
CT	Carboxyl terminal
DBM	Dibromomaleimide
DCA	Dichloroacetone
DMF	N, N-Dimethylformamide
DNA	Deoxyribonucleic acid
DPP	Diphenyl phosphonate
E. Coli	Escherichia coli
eGFP	Enhanced green fluorescent protein
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
FS	Fluorosulfate
G, Gly	Glycine
GGGS	Glycine-Glycine-Glycine-Serine

H, His	Histidine
hACE2	Human angiotensin converting enzyme 2
HMW	High molecular weight
HRP	Horseradish peroxidase
IA	Iodoacetamide
IDT	Integrated DNA Technologies
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K, Lys	Lysine
LB media	Luria-Bertani media
LC	Liquid chromatography
MeOH	Methanol
MS	Mass Spectrometry
NEB	New England Biolabs
Ni-NTA	Nickel (II) nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD	Optical Density
OSu	N-Hydroxysuccinimide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PD1	Programmed Death 1
PD-L1	Programmed Death Ligand 1
pfu	Plaque forming units

PPI	Protein-protein interaction
rpm	Revolutions per minute
S, Ser	Serine
SARS	Severe Acute Respiratory Syndrome
SDS	Sodium dodecyl sulfate
SF	Sulfonyl fluoride
SuFEX	Sulfur (VI) fluoride exchange
TAE	Tris base, acetic acid and EDTA
TCEP	Tris(2-carboxyethyl) phosphine
TCI	Targeted covalent inhibitor
TIC	Total ion chromatogram
TOF	Time of flight
UV	ultraviolet
Vis	Visible light
VS	Vinyl sulfone
XGal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside
Y, Tyr	Tyrosine

CHAPTER 1: INTRODUCTION

1.1 Covalent Drug Development And Usage

When the term "covalent drugs" pops up in talk of therapeutics, the concept is often met with both hopeful optimism but also serious concerns. Traditionally, the term refers to a small molecule drug that engages with its target via irreversible covalent bond formation between the two. The formation of this irreversible bond brings with it a number of advantages that could give a covalent drug candidate the edge over a noncovalent counterpart. Covalent drugs have been shown to exhibit enhanced biochemical efficiency and potency, enhanced residence time, an outlasting of drug pharmacokinetics, reduced necessary dosages, total inactivation of the targets and enhanced difficulty in developing drug resistance[1,2]. The enticing benefits of covalent drugs are not without their drawbacks, however. Perhaps the most notable drawback is also the factor that has caused the greatest hesitation in research and development of covalent drugs: idiosyncratic toxicity caused by off-target binding. Covalent bond formation significantly alters the binding kinetics of a drug in that it eliminates the dissociation rate (often depicted as k_{off} in protein binding). While the elimination of dissociation enhances engagement with the target, this effect is not selective and therefore will also enhance engagement with other, non-target proteins. This off -target binding can potentially lead to inhibition, degradation, or denaturation of proteins beyond the target which could have unprecedented detrimental effects on the health of the person receiving the drug[3]. As such, the specificity of a covalent drug is of the highest priority.

Despite the wariness of medicinal chemists, covalent drugs are remarkably prevalent within the sphere of clinically available drugs. To date, over 40 covalent drugs have been approved for public use by the Food and Drug Administration (FDA). Of this

collection, several are well-known household names such as aspirin, penicillin, and omeprazole. **Figure 1-1** depicts the structures and basic mechanisms of actions of these common drugs. Others such as afatinib and ibrutinib are lifesaving anti-cancer drugs[4]. The presence of these covalent compounds in the markets is still largely coincidental as the covalent mechanisms of these drugs are often discovered serendipitously with their beneficial effects[5].

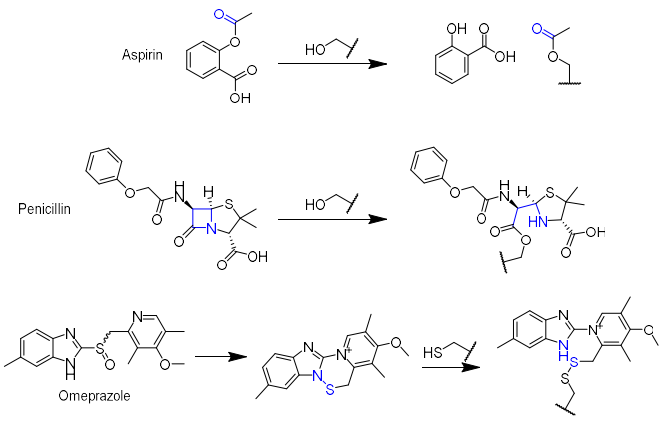


Figure 1-1: Common covalent drugs and their mechanisms of action. Covalent functionalities of each molecule are highlighted in blue.

Due to the irreversible nature of a covalent drug, development of them has been focused on creating targeted covalent inhibitors (TCIs) for clinically relevant proteins. As mentioned previously, covalent drugs form a permanent engagement with their target. TCIs boast a permanent inactivation limited only by the resynthesis rate of the targeted protein if implemented properly. Within the past decade, development of TCIs has seen a resurgence. New chemo-selective and biorthogonal reaction discovery coupled with enhanced technology for high-throughput screening puts medicinal chemists in a unique and advantages position for the development of covalent drugs and TCIs[3].

1.2 Covalent Warheads

In order to enhance selectivity to the highest degree, much consideration must be given to the reactive modality within the drug molecule, sometimes referred to as the covalent warhead. The design of a covalent warhead is therefore centralized around the selectivity

and reactivity of a reactive electrophile. Within the scope of TCIs, specificity of the warhead is a largely chemical property in that a covalent warhead candidate should only be able to form a covalent bond with specific residues. Reactivity of the warhead is unique in that there is no easy definition for desirable reaction kinetics. If covalent bond formation is not kinetically favorable, the reaction may be too slow to occur in a biological setting and may either be outcompeted, or the drug may be metabolized before a reaction can occur. On the other hand, if the reaction kinetics are too favorable and the warhead reacts spontaneously then the likelihood of off-target binding increases as well. As such, there would exist an ideal range of reactivity that a warhead should fall into in order to be considered for implementation within a drug molecule[6]. To date, this ideal range of reactivity has not been definitively observed, and most likely varies on a case-by-case basis.

Currently, one of the most dominant strategies in covalent warhead design has been the targeting of cysteine thiols with acrylamides and α, β-unsaturated carbonyls[7]. While these methods are proven to be selective and efficient, it could be argued that the scope of what can be accomplished with these warheads remains limited due to the fact that cysteine is the least abundant amino acid and those which are present in proteins are often structural via disulfide bond formation. A search has begun to identify the most capable covalent warheads for targeting novel residues beyond cysteine. **Figure 1-2** shows a number of covalent warheads that have been developed recently, as well as the amino acid residues each is designed to bond with. Epoxides[8,9] and aziridines[10,11] have already demonstrated their clinical viability as they have been implemented in commercially available products

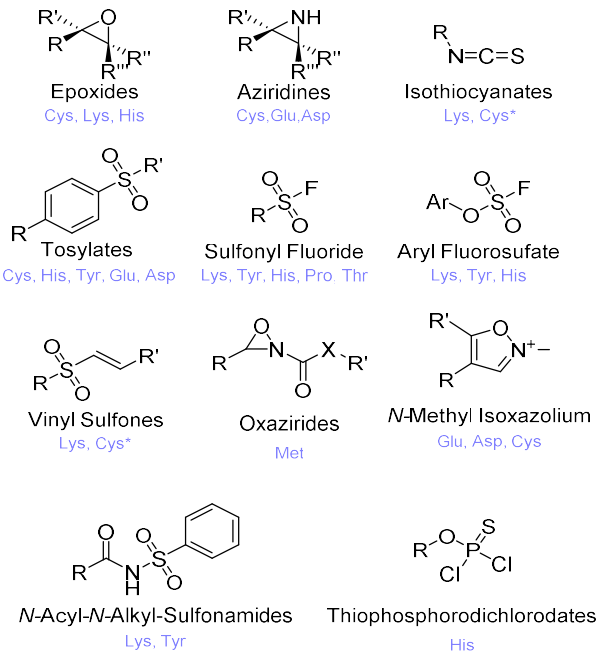


Figure 1-2: Documented covalent warhead catalogue. Listed below each warhead are the amino acid residues that warhead has been documented covalently binding to. Residues with an asterisk designate that reactivity has been observed but is not favored.

as covalent warheads targeting thiols. While thiol reactivity has been the focus of past research, it has been shown independently that both functional groups are capable for covalently binding to other residues, with epoxides being able to react with lysine and histidine while aziridines have been shown to react with carboxylates in glutamic acid and aspartic acid. Similarly, isothiocyanates[12] and vinyl sulfones[13–15] are capable of targeting thiols, but have been shown to react favorably with other motifs. Of particular interest in covalent warhead development are the select warheads which are specifically designed to target residues beyond cysteine including sulfonyl fluorides[16,17], aryl fluorosulfates[18–20], N-methyl isoxazolium salts[21], oxaziridines[22], and thiophosphorodichloridates[23].

1.3 Peptide Therapeutics

Peptide based therapeutics exist on the opposite end of the medicinal chemistry spectrum. Peptides are capable of fulfilling some of the same roles as small molecule-based TCIs, but instead of binding strength coming from an irreversible covalent bind it instead comes from a compounded effect of multiple noncovalent interactions such as hydrogen bonding, ionic interactions, and π-orbital stacking. **Figure 1-3a** shows a basic, cartoon

representation of this binding. Technological advances, namely the development of solid-phase peptide synthesis[24], have allowed peptide therapeutics to come to the forefront of drug development. Factors such as the significant potency, selectivity, low toxicity of peptides, and rapid clearance time combined with the fact that peptides are naturally used as vital mediators and signals in biological pathways has helped spur the drive to identify and synthesis peptide therapeutics, resulting in approximately 80 clinically approved peptide drugs in the market today with over 100 in clinical development and over 400 in preclinical trials[25].

It is important to note that, like TCIs, peptide therapeutics are not without drawbacks that complicate their development. Since peptides are derived from naturally occurring molecules, all but the most rigid and heavily modified peptides are readily metabolized by the human body. This has been observed to cause low oral bioavailability, low metabolic stability, short half-lives, and rapid renal clearance[26] which must be overcome to make a more successful drug candidate.

As the development of peptide therapeutics continues, it becomes important to make distinctions between types of peptides being investigated, as some may be more clinically relevant. Native peptide mimics represent the classical method of peptide therapeutics dating back to the first use of insulin to treat diabetes mellitus[27]. A native peptide mimic is simply a copy of a naturally occurring peptide. Native peptide analogs are similar to native peptides except that they are modified to enhance stability, selectivity, binding affinity, or some other desirable factor. The metabolically stable analog of somatostatin is a common example of this[28]. Lastly and most relevant are heterologous peptides. This class of peptides encompasses those which have structures identified

serendipitously with their function. Generally, this represents peptides that are pulled out of large-scale peptide libraries following stringent selection processes.

1.4 Innovations in Phage Display Technology

Since it was discovered that bacteriophage could display specific, genetically encoded antigens and peptides on their surface[29], phage display has become a versatile and high throughput methodology for the de novo discovery of heterologous peptide binders to proteins of interest. Though several methodologies of this technology exist, one of the most common options utilizes the filamentous bacteriophage M13, which is capable of expressing specific antigens on its surface in limited and controllable amounts. It was discovered that foreign and synthetic strands of DNA could be inserted into the M13 genome within the gene coding for the phage gene III protein (also referred to as gIII or pIII protein). Following incorporation, phage are capable of expressing the foreign peptide sequence on the N-terminal end of the pIII protein without interfering with phage infectivity. The resulting mutant phage can then be identified and amplified to create many more copies of the construct[30].

Phage libraries are created when the DNA fragment inserted into the phage genome contains some number of randomized sequences, such that an extremely large number of peptides can be displayed at once, with the complexity of the library being directly associated with the size of the insert[30]. **Figure 1-3b** shows a cartoon representation of one phage particle from such a library. Oftentimes, the inserted sequence will also contain discrete sequences that can serve as markers or reactive handles following phage production. For example, a common example of this is the inclusion of a cysteine residue on either side of the randomized sequence. In solution, these residues will form a disulfide

bond, cyclizing the displayed peptide and providing structural rigidity. If the disulfide is reduced, each thiol instead provides a unique reactive handle for modification of the phage. This motif is referred to as a CXC sequence, with X representing the number of random residues between these Cys residues.

Phage libraries with sufficient diversity can be systematically panned against immobilized protein targets to identify the strongest peptide binders of the protein within the pool. This is possible due to the distinct connection between phenotype (in most cases protein binding ability) and genotype that is created with the construction of phage libraries. As noted previously, peptides can serve as potent binders and inhibitors that can function with high degrees of specificity and as shown in **Figure 1-3c,** peptides displayed on a phage particle are capable of binding proteins in much the same way as independent peptides. This has allowed phage display to serve as a powerful tool for developing strong binding molecules like peptides, proteins, and target-specific antibodies[31]. The platform has even seen use in whole-cell screenings to select for peptides that can inhibit growth and proliferation of microbes and viruses[32].

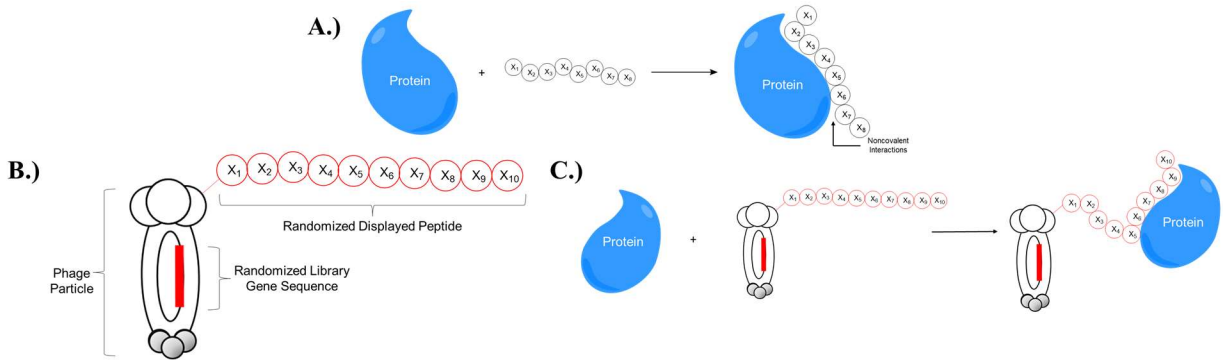


Figure 1-3: Cartoon representation of peptide and protein interactions involved in phage display. Represented here is a peptide binding to a protein surface (A), a peptide displayed on a phage particle (B), and a displayed peptide on a phage binding to a protein (C). All depicted interactions between peptides and proteins occur through noncovalent interactions such as H-bonding and salt bridge formation.

As noted previously, once reduced the two solvent exposed Cys thiols can serve as uniquely nucleophilic reactive handles on a phage, serving as a primary example of the next major steps in phage display technology: chemically modified phage. Chemical modifications of phage allow for screening of peptides beyond natural, homologous peptides and opens the door to discover protein binding peptides with higher binding affinity than a natural peptide may be able to achieve. It has been well documented that modification of thiols on the pIII protein of phage can be utilized to create cyclized and multicyclic libraries of various sizes, adding a determinable amount of structural rigidity to a peptide[33]. Reactive handles on phage proteins also allows for the incorporation of new chemical motifs that are not naturally present; chemical warheads can be affixed to the displayed peptides on the phage particle to potentially enhance binding, as depicted in **Figure 1-4**. The Gao lab has previously seen success in labeling phage with 2-acetylphenylboronic acid (APBA) warheads[34,35] and as such has begun moving into derivatives of this motif designed for longer residence time[36], as will be discussed later. This class of iminoboronate-forming warheads are capable of eliciting a reversible covalent bond with amines. This reversible covalent bond formation, coupled with the traditional

noncovalent interactions that allow peptides to bind a protein allows for the screening and identification of strong and selective binders to proteins and even whole bacterial cells[35].

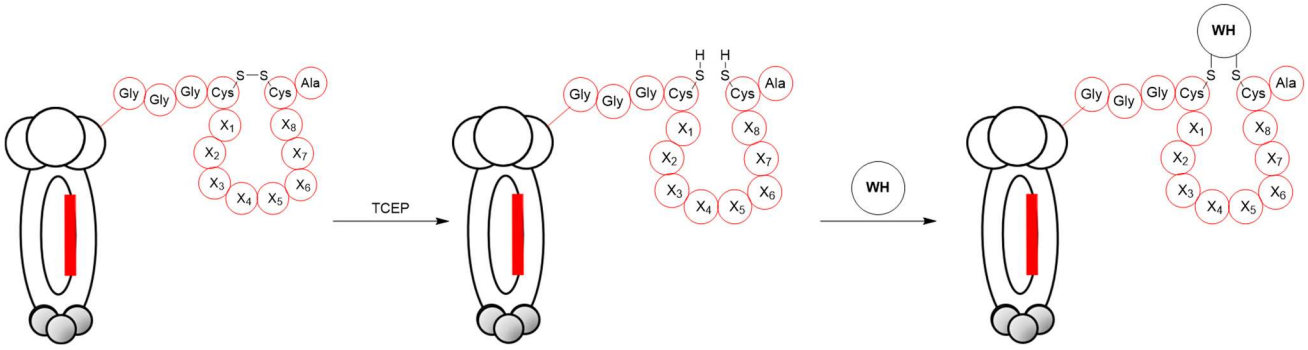


Figure 1-4: Reduction and cyclization of displayed peptide on a phage particle using a warhead molecule. Phage can be cyclized with any small molecule capable of selectively reacting with both displayed thiols. Warheads must be small molecules with this function and contain a second group capable of target engagement. Cyclization of phage provides enhanced rigidity to the structure which can contribute to binding ability.

In order to circumvent the reversibility of peptide binding, as well as the other downsides of peptide-based inhibitors discussed previously, there is a desire to implement irreversible covalent binding into peptides. Efficient incorporation of a covalent warhead onto a phage particle could allow for the discovery of protein-binding molecules with high selectivity due to the prevalent noncovalent interactions inherent to peptides combined with the higher efficiency of a covalent warhead. Some groups have begun investigating the possibility of this incorporation. A notable step in this field can be seen in the work of the Bogyo lab, who were successful in using vinyl sulfone (VS) and diphenyl phosphonate (DPP) based warheads to covalently target catalytic cysteine and serine residues respectively[37]. This discovery indicates that a weakly electrophilic covalent warhead can be incorporated into phage libraries and used to successfully screen for protein inhibitors, however there is the drawback that cysteine and activated serine are very uncommon residues. When these residues are in a form capable of being targeted, as seen by Chen et al., they are commonly found within the active site of enzymes. Thus, there is a desire to

discover novel covalent warheads with a wider array of targetable amino acids that can be implemented into phage display. Doing so would allow for potent, covalent binders to be discovered for a much wider array of proteins and create many new and exciting opportunities for therapeutic development.

1.5 PPI's and "Undruggable" Targets

Proper and efficient incorporation of covalent warheads onto a phage display platform creates many novel opportunities for drug design, with one of the most significant being the potential to interact with targets that were traditionally referred to as "undruggable," such as proteins sites that engage in protein-protein interactions (PPIs). Covalent modification of proteins is an attractive method for inhibiting PPIs since the thermodynamics of binding with a noncovalent molecule are often drastically disfavored[38]. In most biological cases, the protein surfaces that engage in PPIs consist of large portions of the protein surface (~1000-2000 $Å^2$), tend to be shallow, and often contain many non-reactive and hydrophobic residues[39]. It has been shown previously that peptide-like drugs can be utilized to inhibit PPI's in a specific and noncovalent manner[40], however the advantages of covalent drugs noted previously may serve to make these interactions even more potent and effective, something that is exceptionally important in today's clinical environment.

Many significant examples of clinically relevant PPIs exist, however one that is currently of particular relevance is the spread of the novel coronavirus SARS-CoV-2. The viral spike (S) glycoprotein of the SARS-CoV-2 virus engages in a PPI with the hACE2 receptor of human epithelial cells, and this interaction has been shown to have affinity in the nanomolar range[41].The nature of this interaction has been extensively studied and

documented[42], and while the binding relies on a number of complex factors what is relevant is the properties of the binding site. As shown in **Figure 1-5a** the binding site on the S glycoprotein contains a number of tyrosine residues, as well as some lysine residues. Inhibition of this interaction could potentially prevent infection of human cells by viral particles if drug binding to the spike protein is able to sufficiently outcompete hACE2 binding. Dysregulation of PPI's can also play a significant role in cancers, for example the interaction between the proteins Programmed Death 1 (PD1) and its ligand PD-L1 which is shown in **Figure 1-5b**. This set of membrane protein receptors facilitate the signaling between immune cells such as T cells, natural killer T cells, and dendritic cells along with most other hematopoietic and non- hematopoietic cells[43,44]. This interaction has a measured Kd of 8.2±0.1 μM[45] and is also facilitated via hydrophobic regions that contain Lys and

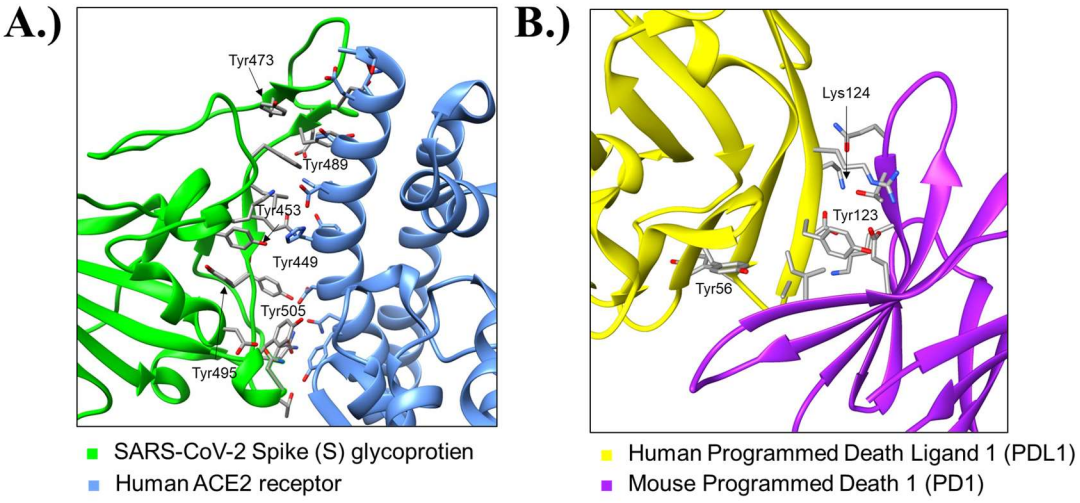


Figure 1-5: Models of Tyr-rich protein-protein interaction sites. Protein-protein interaction sites between the SARS-CoV-2 spike (S) glycoprotein with the hACE2 receptor (A) and the human PDL1 ligand with a mouse PD1 protein are examples of Tyr rich interactions sites. The mouse PD1 protein is a structural analog to human PD1, therefore binding is facilitated with human PDL1 in a similar manner. Both interactions sites have numerous solvent expose tyrosine and lysine residues at the sites of interaction.

Covalent Drugs		Peptide-Based Drugs	
Pros	Cons	Pros	Cons
✓ Enhanced potency	x Potential haptenization	✓ Broad target range	x Limited oral bioavailability
✓ Enhanced residence time	x Idiosyncratic reactions	✓ Highly variable	x Short half-lives
✓ Outlasting pharmacokinetics	x Toxicity	✓ High potency	x Rapid clearance
✓ Lower dosage	x Off target labeling	✓ High Selectivity	x Low metabolic stability
✓ Total inactivation of target		✓ Low toxicity	
✓ Resistance is difficult		✓ Low tissue accumulation	

Table 1-1: Pros and cons of covalent drugs and peptide-based drugs. Many of the pros of covalent drugs are capable of overcoming the cons of peptide-based drugs and vice versa. Thus, creation of a covalent, peptide-based drug is highly desirable.

Tyr residues[46]. PD-L1 is frequently seen overexpressed on the surface of tumor cells, leading to inhibited proliferation of T lymphocytes as well as cytokine release. This, in turn, can lead to exhaustion and apoptosis of tumor-specific T-cells and allow tumor cells to escape immune detection[47,48]. Inhibition of this interaction has been shown to reduce the physiological burden of tumors and serves as a favorable method of cancer therapy[49]. Both of these PPIs serve as strong candidates for drug development and are particularly popular targets for monoclonal antibody development. However, antibodies often have poor tissue penetrance and adverse immune responses, sometimes even leading to permanent tissue damage[20]. As such, there has been recent developments into peptide based therapeutics for both cases[50–53], however to date no covalent binding peptides have been discovered.

1.6 Research Goals

Implementation of covalent warheads into phage display systems, such as the general method outlined in **Figure 1-6** which uses PD1 and the S-glycoprotein as example targets, would serve as a unique platform for the discovery of PPI inhibitors. **Table 1-1** lists the advantages of disadvantages of both covalent drugs and peptide therapeutics

discussed thus far. It can be seen that many of the disadvantages for one motif can be accounted for via the advantages of the other. As such, covalent binding peptides hold great potential for drug development, and phage display is a powerful tool for peptide discovery. Paradoxically, however, is the issue that phage is also made of protein, and so to add in the more promiscuous binding that would be necessary to properly form a covalent bond within the site of a PPI there is great risk involved. It therefore becomes a necessity to gain further knowledge and understanding of the relationships between bacteriophage, their displayed peptides, and the chemical warheads that can provide covalent binding.

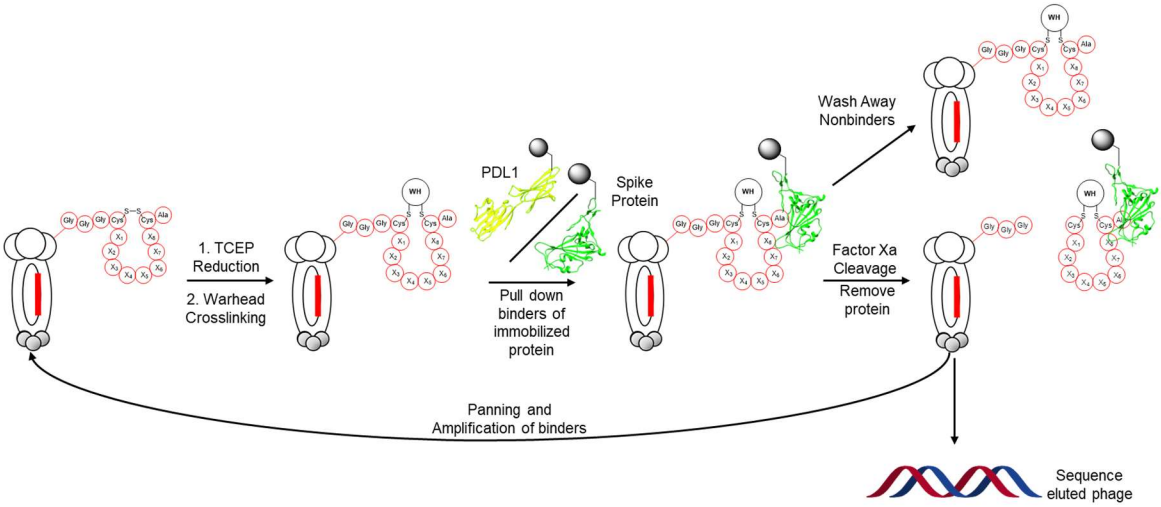


Figure 1-6: Proposed experimental flowthrough for discovering covalent protein inhibitors using phage display. The depicted experimental process would require the successful creation of a covalent library using a suitable covalent warhead. In this model, the proteins PDL1 and the SARS-CoV-2 spike protein are used as example proteins that can be screened against to discover modified peptides that act as surface binding covalent inhibitors.

1.7 References

(1) Mah, R.; Thomas, J. R.; Shafer, C. M. Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2014**, *24* (1), 33–39.
(2) Lonsdale, R.; Ward, R. A. Structure-Based Design of Targeted Covalent Inhibitors. *Chem. Soc. Rev.* **2018**, *47* (11), 3816–3830.

(3) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat Rev Drug Discov* **2011**, *10* (4), 307–317.
(4) Bauer, R. A. Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discovery Today* **2015**, *20* (9), 1061–1073.
(5) Gao, J.; Nobile, V. Chemistry Perspectives of Reversible Covalent Drugs. In *Annual Reports in Medicinal Chemistry*; Elsevier, 2021; Vol. 56, pp 75–94.
(6) Gehringer, M.; Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2019**, *62* (12), 5673–5724.
(7) Jackson, P. A.; Widen, J. C.; Harki, D. A.; Brummond, K. M. Covalent Modifiers: A Chemical Perspective on the Reactivity of α,β-Unsaturated Carbonyls with Thiols via Hetero-Michael Addition Reactions. *J. Med. Chem.* **2017**, *60* (3), 839–885.
(8) Evans, M. J.; Morris, G. M.; Wu, J.; Olson, A. J.; Sorensen, E. J.; Cravatt, B. F. Mechanistic and Structural Requirements for Active Site Labeling of Phosphoglycerate Mutase by Spiroepoxides. *Mol. BioSyst.* **2007**, *3* (7), 495.
(9) Morgen, M.; Jöst, C.; Malz, M.; Janowski, R.; Niessing, D.; Klein, C. D.; Gunkel, N.; Miller, A. K. Spiroepoxytriazoles Are Fumagillin-like Irreversible Inhibitors of MetAP2 with Potent Cellular Activity. *ACS Chem. Biol.* **2016**, *11* (4), 1001–1011.
(10) Rempel, B. P.; Withers, S. G. Covalent Inhibitors of Glycosidases and Their Applications in Biochemistry and Biology. *Glycobiology* **2008**, *18* (8), 570–586.
(11) McGregor, L. M.; Jenkins, M. L.; Kerwin, C.; Burke, J. E.; Shokat, K. M. Expanding the Scope of Electrophiles Capable of Targeting K-Ras Oncogenes. *Biochemistry* **2017**, *56* (25), 3178–3183.
(12) Nakamura, T.; Kawai, Y.; Kitamoto, N.; Osawa, T.; Kato, Y. Covalent Modification of Lysine Residues by Allyl Isothiocyanate in Physiological Conditions: Plausible Transformation of Isothiocyanate from Thiol to Amine. *Chem. Res. Toxicol.* **2009**, *22* (3), 536–542.
(13) Li, M.; Cheng, F.; Li, H.; Jin, W.; Chen, C.; He, W.; Cheng, G.; Wang, Q. Site-Specific and Covalent Immobilization of His-Tagged Proteins via Surface Vinyl Sulfone–Imidazole Coupling. *Langmuir* **2019**, *35* (50), 16466–16475.
(14) Anscombe, E.; Meschini, E.; Mora-Vidal, R.; Martin, M. P.; Staunton, D.; Geitmann, M.; Danielson, U. H.; Stanley, W. A.; Wang, L. Z.; Reuillon, T.; Golding, B. T.; Cano, C.; Newell, D. R.; Noble, M. E. M.; Wedge, S. R.; Endicott, J. A.; Griffin, R. J. Identification and Characterization of an Irreversible Inhibitor of CDK2. *Chemistry & Biology* **2015**, *22* (9), 1159–1164.
(15) Palmer, J. T.; Rasnick, D.; Klaus, J. L.; Bromme, D. Vinyl Sulfones as Mechanism-Based Cysteine Protease Inhibitors. *J. Med. Chem.* **1995**, *38* (17), 3193–3196.
(16) Mukherjee, H.; Debreczeni, J.; Breed, J.; Tentarelli, S.; Aquila, B.; Dowling, J. E.; Whitty, A.; Grimster, N. P. A Study of the Reactivity of S $^{(VI)}$ –F Containing Warheads with Nucleophilic Amino-Acid Side Chains under Physiological Conditions. *Org. Biomol. Chem.* **2017**, *15* (45), 9685–9695.
(17) Narayan, S.; Muldoon, J.; Finn, M. G.; Fokin, V. V.; Kolb, H. C.; Sharpless, K. B. "On Water": Unique Reactivity of Organic Compounds in Aqueous Suspension. *Angew. Chem. Int. Ed.* **2005**, *44* (21), 3275–3279.

(18) Chen, W.; Dong, J.; Plate, L.; Mortenson, D. E.; Brighty, G. J.; Li, S.; Liu, Y.; Galmozzi, A.; Lee, P. S.; Hulce, J. J.; Cravatt, B. F.; Saez, E.; Powers, E. T.; Wilson, I. A.; Sharpless, K. B.; Kelly, J. W. Arylfluorosulfates Inactivate Intracellular Lipid Binding Protein(s) through Chemoselective SuFEx Reaction with a Binding Site Tyr Residue. *J. Am. Chem. Soc.* **2016**, *138* (23), 7353–7364.
(19) Fadeyi, O. O.; Hoth, L. R.; Choi, C.; Feng, X.; Gopalsamy, A.; Hett, E. C.; Kyne, R. E.; Robinson, R. P.; Jones, L. H. Covalent Enzyme Inhibition through Fluorosulfate Modification of a Noncatalytic Serine Residue. *ACS Chem. Biol.* **2017**, *12* (8), 2015–2020.
(20) Li, Q.; Chen, Q.; Klauser, P. C.; Li, M.; Zheng, F.; Wang, N.; Li, X.; Zhang, Q.; Fu, X.; Wang, Q.; Xu, Y.; Wang, L. Developing Covalent Protein Drugs via Proximity-Enabled Reactive Therapeutics. *Cell* **2020**, *182* (1), 85-97.e16.
(21) Martín-Gago, P.; Fansa, E. K.; Winzker, M.; Murarka, S.; Janning, P.; Schultz-Fademrecht, C.; Baumann, M.; Wittinghofer, A.; Waldmann, H. Covalent Protein Labeling at Glutamic Acids. *Cell Chemical Biology* **2017**, *24* (5), 589-597.e5.
(22) Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; Chang, C. J. Redox-Based Reagents for Chemoselective Methionine Bioconjugation. *Science* **2017**, *355* (6325), 597–602.
(23) Jia, S.; He, D.; Chang, C. J. Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation. *J. Am. Chem. Soc.* **2019**, *141* (18), 7294–7301.
(24) Merrifield, R. B. **Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide**. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.
(25) Muttenthaler, M.; King, G. F.; Adams, D. J.; Alewood, P. F. Trends in Peptide Drug Discovery. *Nat Rev Drug Discov* **2021**, *20* (4), 309–325.
(26) Lau, J. L.; Dunn, M. K. Therapeutic Peptides: Historical Perspectives, Current Development Trends, and Future Directions. *Bioorganic & Medicinal Chemistry* **2018**, *26* (10), 2700–2707.
(27) Banting, F. G.; Best, C. H.; Collip, J. B.; Campbell, W. R.; Fletcher, A. A. Pancreatic Extracts in the Treatment of Diabetes Mellitus. *Can Med Assoc J* **1922**, *12* (3), 141–146.
(28) Fricker, G.; Bruns, C.; Munzer, J.; Briner, U.; Albert, R.; Kissel, T.; Vonderscher, J. Intestinal Absorption of the Octapeptide SMS 201–995 Visualized by Fluorescence Derivatization. *Gastroenterology* **1991**, *100* (6), 1544–1552.
(29) Smith, G. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **1985**, *228* (4705), 1315–1317.
(30) Matthews, D.; Wells, J. Substrate Phage: Selection of Protease Substrates by Monovalent Phage Display. *Science* **1993**, *260* (5111), 1113–1117.
(31) Hamzeh-Mivehroud, M.; Alizadeh, A. A.; Morris, M. B.; Bret Church, W.; Dastmalchi, S. Phage Display as a Technology Delivering on the Promise of Peptide Drug Discovery. *Drug Discovery Today* **2013**, *18* (23–24), 1144–1157.
(32) Huang, J. X.; Bishop-Hurley, S. L.; Cooper, M. A. Development of Anti-Infectives Using Phage Display: Biological Agents against Bacteria, Viruses, and Parasites. *Antimicrob. Agents Chemother.* **2012**, *56* (9), 4569–4582.
(33) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-Encoded Combinatorial Chemical Libraries Based on Bicyclic Peptides. *Nat Chem Biol* **2009**, *5* (7), 502–507.

(34) Bandyopadhyay, A.; McCarthy, K. A.; Kelly, M. A.; Gao, J. Targeting Bacteria via Iminoboronate Chemistry of Amine-Presenting Lipids. *Nat Commun* **2015**, *6*, 6561.
(35) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; van Opijnen, T.; Gao, J. Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (19), 6137–6145.
(36) Reja, R. M.; Wang, W.; Lyu, Y.; Haeffner, F.; Gao, J. Lysine-Targeting Reversible Covalent Inhibitors with Long Residence Time. *J. Am. Chem. Soc.* **2022**, *144* (3), 1152–1157.
(37) Chen, S.; Lovell, S.; Lee, S.; Fellner, M.; Mace, P. D.; Bogyo, M. Identification of Highly Selective Covalent Inhibitors by Phage Display. *Nat Biotechnol* **2020**.
(38) Way, J. Covalent Modification as a Strategy to Block Protein–Protein Interactions with Small-Molecule Drugs. *Current Opinion in Chemical Biology* **2000**, *4* (1), 40–46.
(39) Hwang, H.; Vreven, T.; Janin, J.; Weng, Z. Protein-Protein Docking Benchmark Version 4.0: Protein-Protein Docking Benchmark Version 4.0. *Proteins* **2010**, *78* (15), 3111–3114.
(40) Arkin, M. R.; Tang, Y.; Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chemistry & Biology* **2014**, *21* (9), 1102–1114.
(41) Chen, Y.; Guo, Y.; Pan, Y.; Zhao, Z. J. Structure Analysis of the Receptor Binding of 2019-NCoV. *Biochemical and Biophysical Research Communications* **2020**, *525* (1), 135–140.
(42) Valdes-Balbin, Y.; Santana-Mederos, D.; Paquet, F.; Fernandez, S.; Climent, Y.; Chiodo, F.; Rodríguez, L.; Sanchez Ramirez, B.; Leon, K.; Hernandez, T.; Castellanos-Serra, L.; Garrido, R.; Chen, G.-W.; Garcia-Rivera, D.; Rivera, D. G.; Verez-Bencomo, V. Molecular Aspects Concerning the Use of the SARS-CoV-2 Receptor Binding Domain as a Target for Preventive Vaccines. *ACS Cent. Sci.* **2021**, *7* (5), 757–767.
(43) Okazaki, T.; Honjo, T. PD-1 and PD-1 Ligands: From Discovery to Clinical Application. *International Immunology* **2007**, *19* (7), 813–824.
(44) Han, Y.; Liu, D.; Li, L. PD-1/PD-L1 Pathway: Current Researches in Cancer. *Am J Cancer Res* **2020**, *10* (3), 727–742.
(45) Cheng, X.; Veverka, V.; Radhakrishnan, A.; Waters, L. C.; Muskett, F. W.; Morgan, S. H.; Huo, J.; Yu, C.; Evans, E. J.; Leslie, A. J.; Griffiths, M.; Stubberfield, C.; Griffin, R.; Henry, A. J.; Jansson, A.; Ladbury, J. E.; Ikemizu, S.; Carr, M. D.; Davis, S. J. Structure and Interactions of the Human Programmed Cell Death 1 Receptor. *Journal of Biological Chemistry* **2013**, *288* (17), 11771–11785.
(46) Zak, K. M.; Grudnik, P.; Magiera, K.; Dömling, A.; Dubin, G.; Holak, T. A. Structural Biology of the Immune Checkpoint Receptor PD-1 and Its Ligands PD-L1/PD-L2. *Structure* **2017**, *25* (8), 1163–1174.
(47) Sun, Q.; Chen, X.; Zhou, Q.; Burstein, E.; Yang, S.; Jia, D. Inhibiting Cancer Cell Hallmark Features through Nuclear Export Inhibition. *Sig Transduct Target Ther* **2016**, *1* (1), 16010.
(48) Sharma, P.; Allison, J. P. The Future of Immune Checkpoint Therapy. *Science* **2015**, *348* (6230), 56–61.
(49) Sharpe, A. H.; Pauken, K. E. The Diverse Functions of the PD1 Inhibitory Pathway. *Nat Rev Immunol* **2018**, *18* (3), 153–167.

(50) Sasikumar, P. G. N.; Ramachandra, M.; Vadlamani, S. K.; SHRIMALI, K. R.; Subbarao, K. Therapeutic Compounds for Immunomodulation. WO2012168944A1, December 13, 2012.
(51) Miller, M. M.; Mapelli, C.; Allen, M. P.; Bowsher, M. S.; Boy, K. M.; Gillis, E. P.; Langley, D. R.; Mull, E.; Poirier, M. A.; Sanghvi, N.; Sun, L.-Q.; Tenney, D. J.; Yeung, K.-S.; Zhu, J.; Reid, P. C.; Scola, P. M. Macrocyclic Inhibitors of the Pd-1/Pd-L1 and Cd80(B7-1)/Pd-L1 Protein/Protein Interactions, September 25, 2014.
(52) Gurung, S.; Khan, F.; Gunassekaran, G. R.; Yoo, J. D.; Poongkavithai Vadevoo, S. M.; Permpoon, U.; Kim, S.-H.; Kim, H.-J.; Kim, I.-S.; Han, H.; Park, J.-H.; Kim, S.; Lee, B. Phage Display-Identified PD-L1-Binding Peptides Reinvigorate T-Cell Activity and Inhibit Tumor Progression. *Biomaterials* **2020**, *247*, 119984.
(53) Magiera-Mularz, K.; Skalniak, L.; Zak, K. M.; Musielak, B.; Rudzinska-Szostak, E.; Berlicki, Ł.; Kocik, J.; Grudnik, P.; Sala, D.; Zarganes-Tzitzikas, T.; Shaabani, S.; Dömling, A.; Dubin, G.; Holak, T. A. Bioactive Macrocyclic Inhibitors of the PD-1/PD-L1 Immune Checkpoint. *Angew. Chem. Int. Ed.* **2017**, *56* (44), 13732–13735.

CHAPTER 2: DEVELOPING A pIII PROTEIN-BASED SCREENING PLATFORM FOR COVALENT WARHEADS

2.1 The M13KE pIII protein

As noted in Chapter 1, incorporation of covalent warheads into phage libraries is paradoxical in nature. Covalent warheads are designed to irreversibly bind a protein target, yet a phage particle is no more than a small genome encapsulated by a shell of proteins. **Figure 2-1a** depicts an illustration of the potential outcomes of modifying phage via a bifunctional molecule containing a covalent warhead. As shown, three distinct possibilities emerge, however only one is desired. Therefore, in order to have the greatest chances of success it is desirable to create a model system. This was determined to be an ideal place to begin investigating the development of covalent phage libraries, and so the M13KE phage pIII protein was selected as the model.

The M13KE pIII protein, sometimes referred to as the gIII protein, is a minor coat protein expressed on the surface of filamentous bacteriophage like M13KE. Typically, five copies of the protein are expressed on one end of the phage and are used to facilitate infection of bacteria cells via recognition of the *E. Coli* F-pilus. Each copy of the protein consists of three major domains: Two solvent exposed N-terminal domains, referred to as N1 and N2, and a hydrophobic carboxy-terminal, or CT domain. The N1 and N2 domains both serve as the binders to the F-pilus of *E. coli* and are connected and stabilized via a number of interior disulfide bonds. A glycine rich linker chain connects the N2 domain to the CT domain, which is embedded within the membrane of phage[1]. **Figure 2-1b** depicts a representation of the M13KE phage, with the pIII protein and its structure highlighted. In

this example, the pIII protein has been modified to incorporate a C7C sequence, such as one that would be found in a phage library.

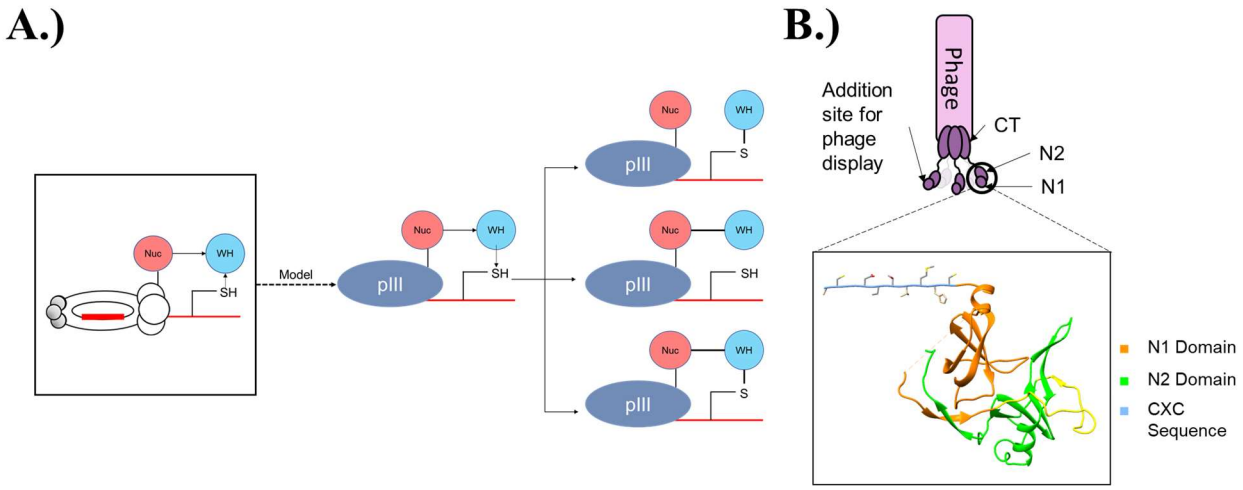
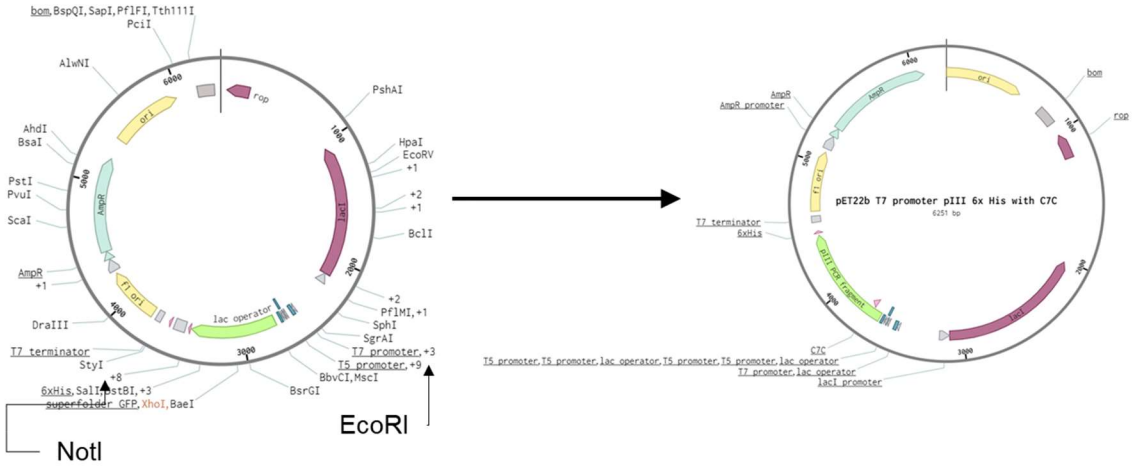


Figure 2-1: Labeling functionality and structure of the pIII protein. Labeling of the pIII protein has two routes of engagement between the exposed thiol and other nucleophilic residues, leading to three possible end results for labeling. The pIII protein can be used to model these modes of labeling and determine if a warhead is engaging with cysteine, another amino acid, or both (A). The structure desired for expression is the N-terminal domains N1 and N2 domains of the pIII protein, as this fragment is solvent exposed, loses many of the hydrophobic residues expressed on the C-terminal (CT) domain of the protein, and contains the displayed CXC peptide sequence.

2.2 Plasmid Construction and Transformation into *E. Coli*

Previous literature has shown that the N1 and N2 domains of the pIII protein, as well as nonnatural derivatives, can be expressed in bacteria cells[2]. With this in mind, the first goal became to express a correctly assembled pIII protein from a phage library that could be used as a model. *E. coli* of the strain ER2738 were infected with a commercially available C7C M13KE phage library. Following phage tittering, plaques formed from single bacteriophage were used to infect bacterial subcultures, and phagemid DNA containing library sequences were extracted, purified, and sequenced. Positive sequencing results and a lack of internal Lys or Cys residues led to the pIII protein displaying the peptide sequence CGDPTTMHC to be selected for expression.

As a template, a pET22b vector was provided by the members of the lab of Dr. Abhishek Chatterjee. **Figure 2-2** shows a representation of the original and desired plasmid product following standard digestion and ligation protocols. Also shown is the desired final protein product, with a 6x His tag on the C-terminus of the protein, where the glycine rich linker chain would be in the natural protein. PCR amplified fragments of the desired phagemid as well as the template plasmid were both digested with the restriction enzymes EcoRI and NotI and analyzed via gel electrophoresis, as shown in **Figure 2-3**. Following verification of digestion, fragments were combined and ligated before directly being transformed into electrocompetent *E. coli* cell line TOP10 for expression. Sequencing of



Desired Protein Sequence

ACGDPTTMHCGGGSAETVESCLAKSHTENSFTNVWKDDKTLDRYANYE
GCLWNATGVVVCTGDETQCYGTWVPIGLAIPENEGGGSEGGGSEGGG
SEGGGTKPPEYGDTPIPGYTYINPLDGTYPPGTEQNPANPNPSLEESQPL
NTFMFQNNRFRNRQGALTVYTGTVTQGTDPVKTYYQYTPVSSKAMYD
AYWNGKFRDCAFHSGFNEDLFVCEYQGQSSDLPQPPVNAGGGSGGGS
GGAAALEHHHHHH

Figure 2-2: Cloned plasmid product containing the gene for the pIII protein fragment. Plasmid was derived from a pET22b vector. NotI and EcoRI enzymes were used to replace a gene coding for eGFP with the PCR cloned fragment of pIII. The desired protein sequence, post-expression and following post-translational modifications is shown, with the displayed library sequence highlighted.

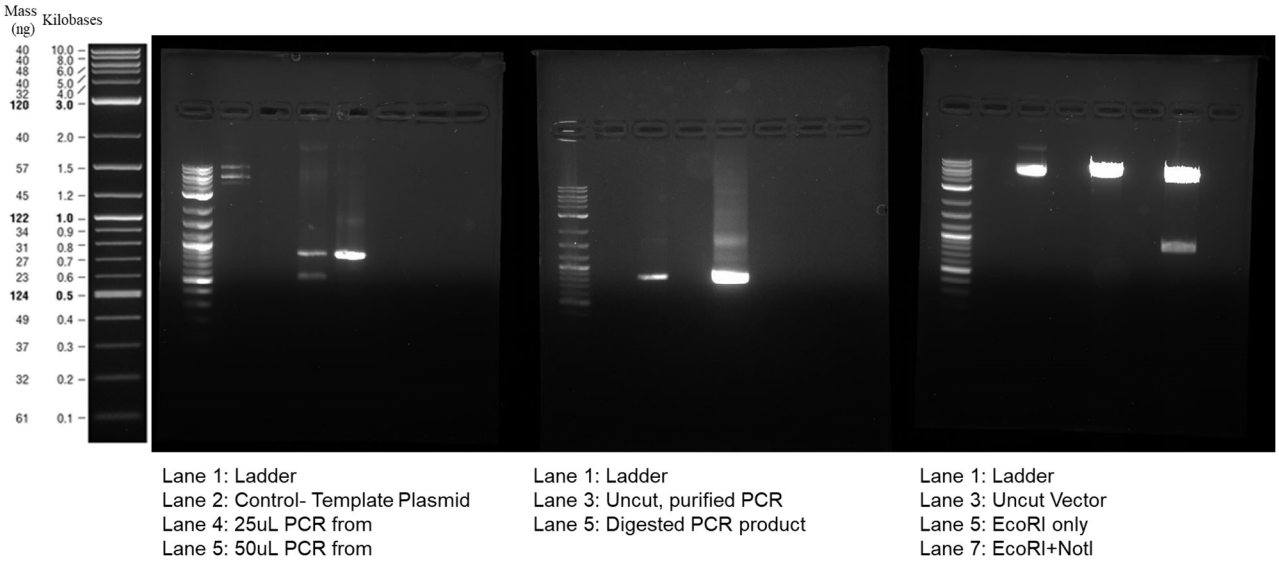


Figure 2-3: Verification of plasmid component creation using agarose gels. Depicted are 1% agarose gels stained with 1 mg/mL ethidium bromide used to verify completion of the PCR reaction (left), digestion of the PCR product (center), and digestion of the template plasmid (right) In each case a 1 kb ladder was used as a reference and an undigested control was used to verify that a change in size had occurred.

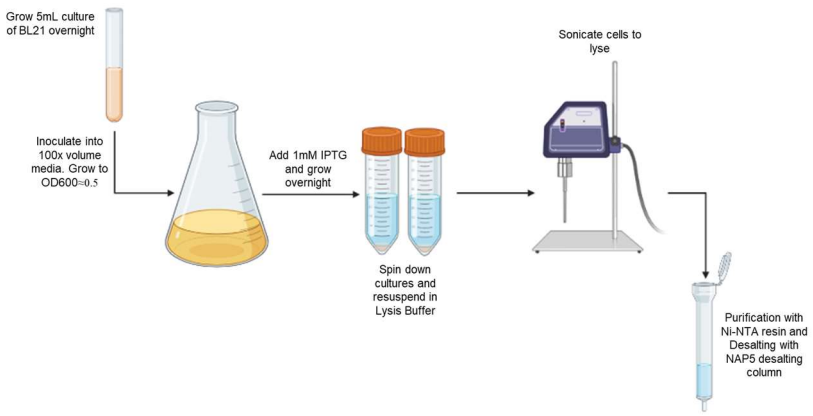
plasmid DNA following transformation of ligated product confirmed that the *E. coli* contained a *lac* inducible gene for the pIII protein fragment.

2.3 Protein Extraction and Purification

In order to express and purify the cloned fragment of the pIII protein, plasmid DNA containing the protein expression was purified from TOP10 cells, sequenced, and transformed into the chemically competent cell line *E. coli* BL21 (DE3). Plasmid DNA was then extracted, purified, and sequenced from this cell line as well to ensure proper cellular uptake of plasmid DNA.

Following successful incorporation of plasmid construct into BL21 (DE3), frozen stocks were prepared from single colonies containing the plasmid and were then used to carry out large scale expression of the protein. To simplify purification protocols as well as to best ensure proper folding and behavior of the protein, an alternative expression

method was used compared to the literature noted previously. **Scheme 2-1** shows a representation of the process used. While this method does not account for insoluble portion of the total protein product, it was selected in an attempt to minimize the amount of misfolded protein that may be purified, as it was theorized that misfolded protein should be less soluble than the native pIII fragment which already expresses poor water solubility. **Figure 2-4** shows example SDS-PAGE protein gels used to identify overexpression of the protein.



Scheme 2-1: Experimental process of protein expression. BL21 (DE3) cells contained cloned plasmid with pIII protein fragment. IPTG was used to induce *lac* operon and induce expression of protein on the pET22b vector. Ni-NTA resin was used to bind 6x His tag expressed on the C-term of the protein. Only 3 fractions of 1 mL with strongest bands on SDS-PAGE gel corresponding to the pIII protein were pooled and desalted.

PAGE gel analysis of the crude lysis, as shown in **Figure 2-4** in the presence and absence of the expression promoter IPTG indicated that protein of the approximate expected size was being overexpressed, as indicated by the presence of new and darker bands. The inclusion of a 6xHis tag on the C-terminal end of the protein fragment allowed for purification using Ni-NTA resin. The PAGE gel shown in **Figure 2-4** shows comparisons of column flowthroughs, compiled washes, and individual elutions during the purification process. It can be noted that elution fractions are intended to contain only a single band indicative of the desired protein product, however each elution was seen to contain several bands indicating protein impurity. Bands are not present when IPTG is not

added to incubating cell cultures and consecutive runs through new resin failed to isolate a single band. This indicated that each protein side product was being expressed in tandem with the pIII protein and likely contained a poly-His tag as well. With PAGE gel results conflicting regarding the nature of protein purification, new methods of protein identification were sought out.

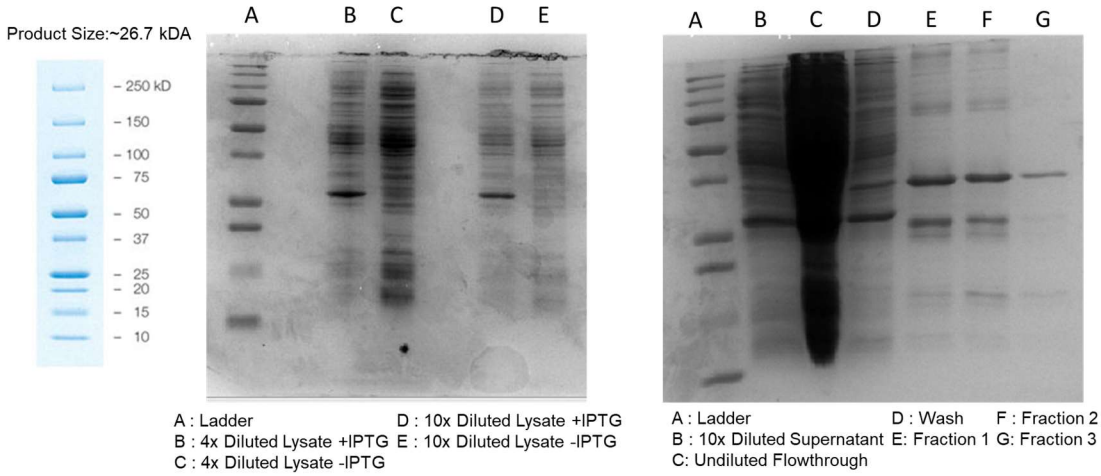


Figure 2-4: SDS-PAGE Gel analysis of pIII protein expression. All proteins run on 15% acrylamide stacking gel. Voltage was slowly ramped over ~1 hour from 80 V up to 120 V. Shown are analysis of crude cell lysate +/- 1 mM IPTG to verify overexpression of protein (A) and comparison of several stages of protein expression, with the fractions containing most intense pIII bands being displayed (B). Based on the high mw ladder used the dark band above the 25 kDA standard was concluded to be pIII, which has an unmodified mass of 26772.

Due to expected low yields of protein sample, total protein concentration was determined using by measuring the A280 of pooled elutions. Based on the calculated extinction coefficient for the protein fragment, total protein concentration was determined for each sample prior to further testing. It would be later determined that UV absorbance at 280 nm was only resulting from the pIII protein, so total protein concentrations were expected to be indicative of pIII protein concentration. It was determined that the average yield per expression was approximately 1mg/L ± 0.05mg. As expected, protein yields were low from this method, but yield was determined to be sufficient enough for modeling

purposes, as actual concentrations of pIII protein in a phage labeling experiment are equally low.

2.4 Confirming Protein Mass and Conformation Using LC/MS

To best understand the complicated mixture of proteins that seemed to be present, it was determined that LC/MS analysis would be used to determine the identity of proteins present. To account for potential contaminants, protein load was determined using the approximate concentration derived from the UV/Vis analysis, which would represent total protein concentration. **Figure 2-5a** shows the spectra obtained from such an analysis, showing the total ion chromatogram (TIC) trace as well as the UV absorbance at 220, 254, and 280 nm.

As was expected, multiple peaks can be seen in the TIC trace, with 3 major peaks which all elute between 14 and 16 minutes with the method used. It is important to note that, while three major peaks can be seen, this does not correlate to three products, as some peaks in the TIC contain multiple protein envelope signals. Perhaps the most important facet of this data is the UV absorbance seen in these samples. Contrary to the TIC trace, **Figure 2-5a** shows that a single peak is seen at all three wavelengths tested within the 14–16-minute range. Based on the elution time, it can be determined that this single UV absorbance peak correlates to the A peak seen in the TIC trace. As shown in **Figure 2-5b**, the A peak contained the envelope of only a single protein, which allowed for deconvolution of the masses. This deconvolution yielded a mass of 26772 amu, the expected mass for the pIII fragment with all structural disulfide bonds formed. Therefore, it was concluded that the pIII fragment was expressed and present in the protein mixture. Additionally, this protein was the major contributor to UV absorbance at 280 nm, so UV

analysis is therefore indicative of pIII protein concentration. The lack of absorbance peaks from other products indicates that all protein co-eluants and potential degradation products are lower in concentration than the pIII protein. However, for future experimentation these additional proteins must be considered as unknown factors as an efficient method of removing them from this specific protein mixture was not found.

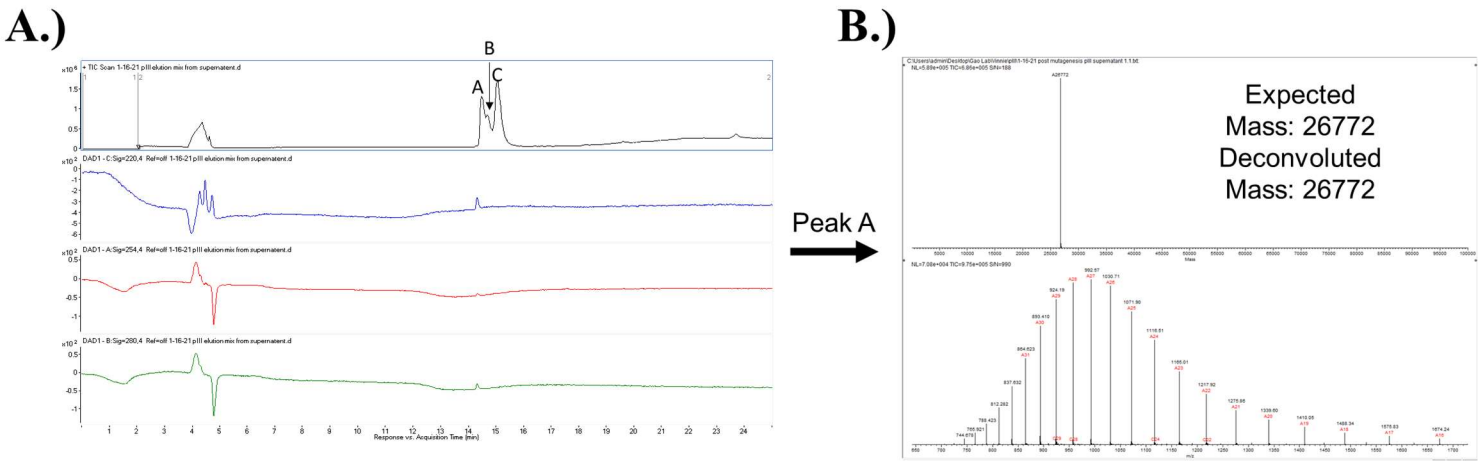


Figure 2-5: LC/MS analysis of purified pIII protein. The raw spectra obtained (A) for the purified pIII protein showing the total ion chromatogram (TIC) labeled. Below the TIC are the UV spectra showing absorbance at 220, 254, and 280 nm respectively. Also shown is the protein envelope and deconvoluted envelope for the "A" peak in the TIC (B). Identical match to the expected mass indicated that the A peak was the pIII protein after proper post-translational modification. The A peak was the only signal to have a corresponding absorbance peak at any wavelength measured.

LC/MS analysis proved that the protein identity and sequence was correct, however the question of protein structure remained. In lieu of traditional structural determination methods, the pIII protein was unique in that structure can also be determined using LC/MS. As noted earlier, the pIII protein cloned was that of a C7C library. Therefore, a properly folded copy of the pIII protein fragment cloned would contain a total of 4 disulfide bonds; one solvent exposed and three internal and structural. A sterically hindered reducing agent, such as the selected TCEP, will only reduce the solvent exposed disulfide bond which cyclizes the library sequence. Phage display itself verifies this in that TCEP is used to selectively reduce this disulfide bond. **Figure 2-6** shows the protein mass envelope and

deconvoluted mass obtained after running an identical method as the one shown in **Figure 2-5** following a 1-hour incubation with 1 mM TCEP. As shown in this figure, only a single signal was observed, with the mass 26774. This directly correlates to the addition of two protons, indicative of a reduced disulfide bond. This confirmed that that only one disulfide bond is solvent exposed and that this bond is capable of being reduced as it would be in a phage display experiment.

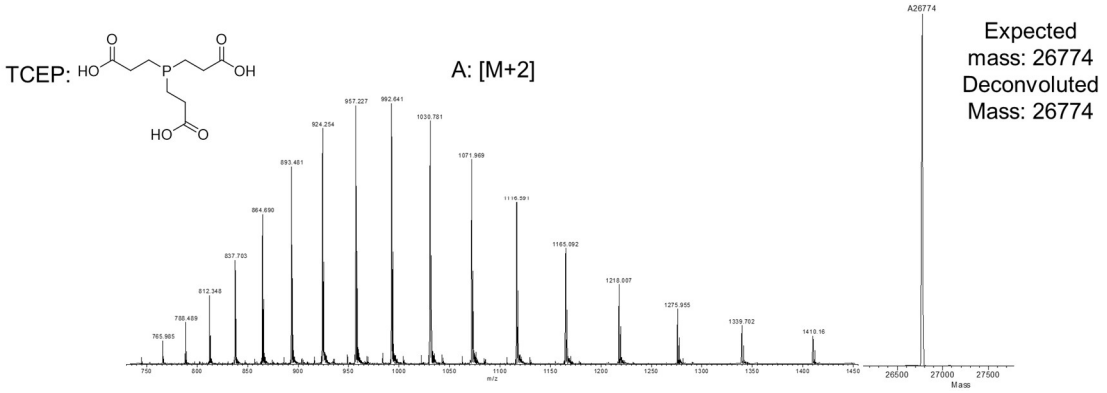


Figure 2-6: Deconvoluted LC/MS analysis of TCEP reduced pIII protein. Shown is the deconvolution of the A peak in the TIC spectra. Crude TIC is not shown for deconvoluted samples due to the chromatogram being indistinguishable from unmodified protein. After incubation with 1 mM TCEP, a 2-mass unit increase corresponding to the reduction of the disulfide bond was seen with no side products seen from reaction with TCEP.

To conclusively verify that the structure of the pIII protein was the same as the native protein, a method was developed to mimic phage labeling conditions with a simple one-step S_N2-type substitution reaction between an exposed thiol and the alpha carbonyl chloride. For this crosslinking experiment, the simplest crosslinker was chosen: dichloroacetone (DCA). Since a disulfide bond selectively reduced using TCEP would yield two free thiols in close proximity, a reduced pIII protein of proper structure will only see an increase in mass equal to a single labeling from DCA. **Figure 2-7** shows the protein envelope obtained from the A peak of a pIII protein sample exposed to 100 μM DCA following a TCEP reduction. Adjacent to the envelope within **Figure 2-7** is the single

deconvoluted mass obtained from the A peak. The deconvoluted mass indicated a protein mass of 26827, 55 amu larger than the unmodified pIII construct or [M+55]. This coincides with the addition of the carbonyl from DCA cyclizing the exposed thiols on the C7C sequence. Based on these results, it can be concluded that the pIII construct created folds and behaves in the same manner as the native pIII protein, and therefore is sufficient for suitability testing for phage display.

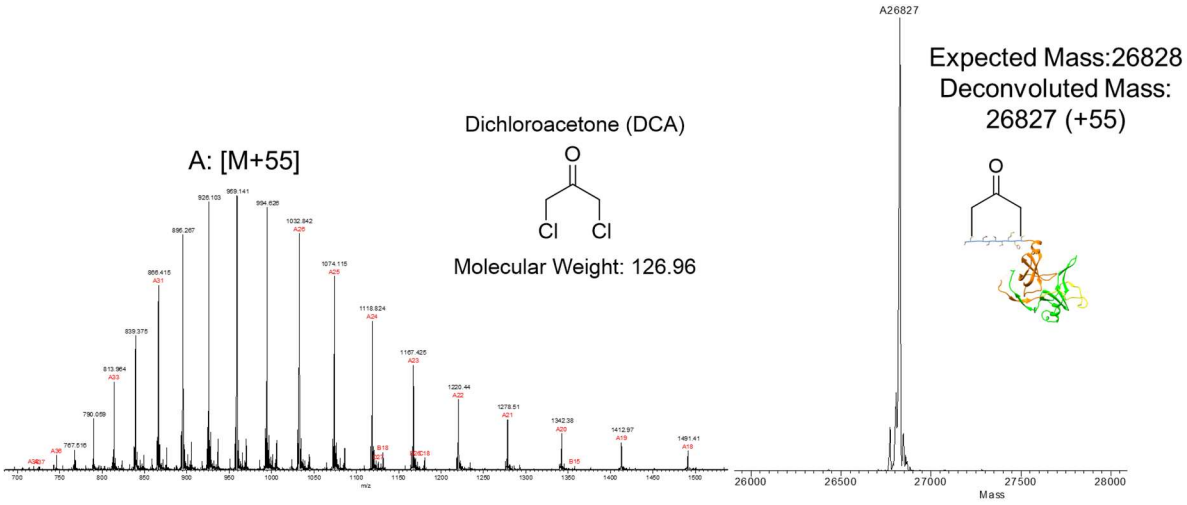


Figure 2-7: Deconvoluted LC/MS analysis of DCA labeling of the pIII protein. Following TCEP reduction, pIII protein was incubated with 100 µM DCA for 1 hour at RT. Only a single product was seen of the expected mass. Change from expected mass of ± 2 mass units was considered to be desired product and change was attributed to an artifact of deconvolution.

2.5 Comparing Cysteine Crosslinkers Using the pIII Based Platform

In order to assess the functionality of the pIII protein as a platform to study small molecule suitability for phage display, the first set of analyses focused on comparing thiol-reactive crosslinkers. Since each bifunctional molecule used for phage labeling contains a headgroup that is thiol reactive, it is desirable to find the most optimal crosslinker. A crosslinker with a faster reaction rate can be used at lower concentrations to yield sufficient phage labeling. While this is valuable for the simple reason of conserving warhead supply, there is the added benefit that lower concentrations of warhead molecule may allow for more potent warheads to be used. A lower concentration of a potent warhead would

decrease the likelihood of side reactions occurring before proper labeling of the displayed peptide. Thus, utilizing the most efficient crosslinker allows for the most expansive library of potential warhead candidates for phage display incorporation.

With DCA having already been evaluated, this molecule was used as a baseline for the comparison of other crosslinker molecules. **Figure 2-8** shows the comparison of two independent DCA labeling experiments with varying concentration. When concentration of the DCA crosslinker was reduced to 50 μM, multiple peaks were deconvoluted from the A peak, corresponding to both unlabeled and labeled but not crosslinked pIII protein. As such the conditions for this labeling were set as the "minimum requirements" that a crosslinker had to meet to be considered. Since DCA is commercially available and easily obtainable, significant advantages in necessary concentration or reaction time were sought after. **Table 2-1** shows all of the crosslinker structures that were evaluated in this experiment and their corresponding molecular masses. As with DCA, labeling from each crosslinker was observed using LC/MS analysis based on the change in mass of the A peak from the protein mixture containing the pIII protein.

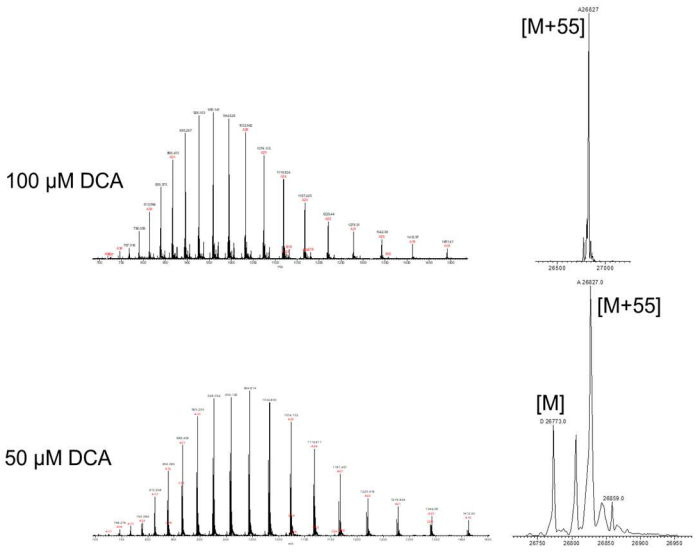


Figure 2-8: Deconvoluted LC/MS analysis of DCA concentration dependence of labeling. When 50 μM DCA was used under identical conditions, a peak was seen in deconvoluted spectra corresponding to unlabeled pIII, which was not seen when 100 μM DCA was used.

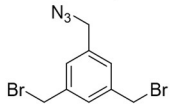
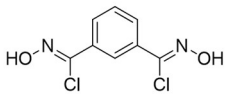
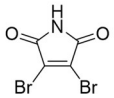
1,3 Dichloroacetone (DCA)  Molecular Weight: 126.96	1-(azidomethyl)-3,5-bis (bromomethyl)benzene (ABBM)  Molecular Weight: 319.00
1,3 bis-Chloro oxime (BCO)  Molecular Weight: 233.05	2,3 Dibromomaleimide (DBM)  Molecular Weight: 254.87

Table 2-1: Structures and masses of potential crosslinkers.

Dibromomaleimide (DBM) serves as a clear example of a crosslinker that was removed from the pool quickly. **Figure 2-9a** shows the protein envelope and deconvoluted masses of a labeling experiment with DBM that was identical to the one performed with DCA. Under identical conditions, DBM was not capable of sufficiently crosslinking the pIII protein to the same degree as DCA, as unlabeled protein was still seen along with the labeled protein indicated by the [M+94] peak. The bis-bromo benzene based crosslinker, provided by members of Dr. Eranthie Weerapana's lab, proved to be a more complicated decision. This motif was selected has it has been shown by Heinis et al. to be an efficient and nontoxic phage crosslinker[3]. As shown in **Figure 2-9b** this molecule had a labeling capacity similar to that of DCA, as indicated by the [M+159] peak. Shown in the figure is the reduction of the concentration in the labeling to 50 μM. As with DCA, this concentration reduction yielded an increase in the amount of unlabeled pIII protein present in the solution. Not shown is the data corresponding to timepoints less than one hour, as no labeling was seen at the time points examined. Lastly is the interesting case of bis-chlorooxime based crosslinkers. As shown in **Figure 2-9c**, pIII crosslinking with the pIII protein was incredibly efficient. The minimum conditions in which 100% crosslinking was observed was found to be 10 μM for 10 minutes, a significant improvement over previously tested crosslinkers. **Table 2-2** shows the compiled minimum labeling requirements determined for each warhead. It is

important to note that an additional parameter was added to this table: phage toxicity. Prior testing by other members of the Gao Lab indicated that phage exposed to the same bis-chlorooxime crosslinker under the same conditions used in this experiment saw >90% reduction in phage output in titers following modification. This would seem to indicate the

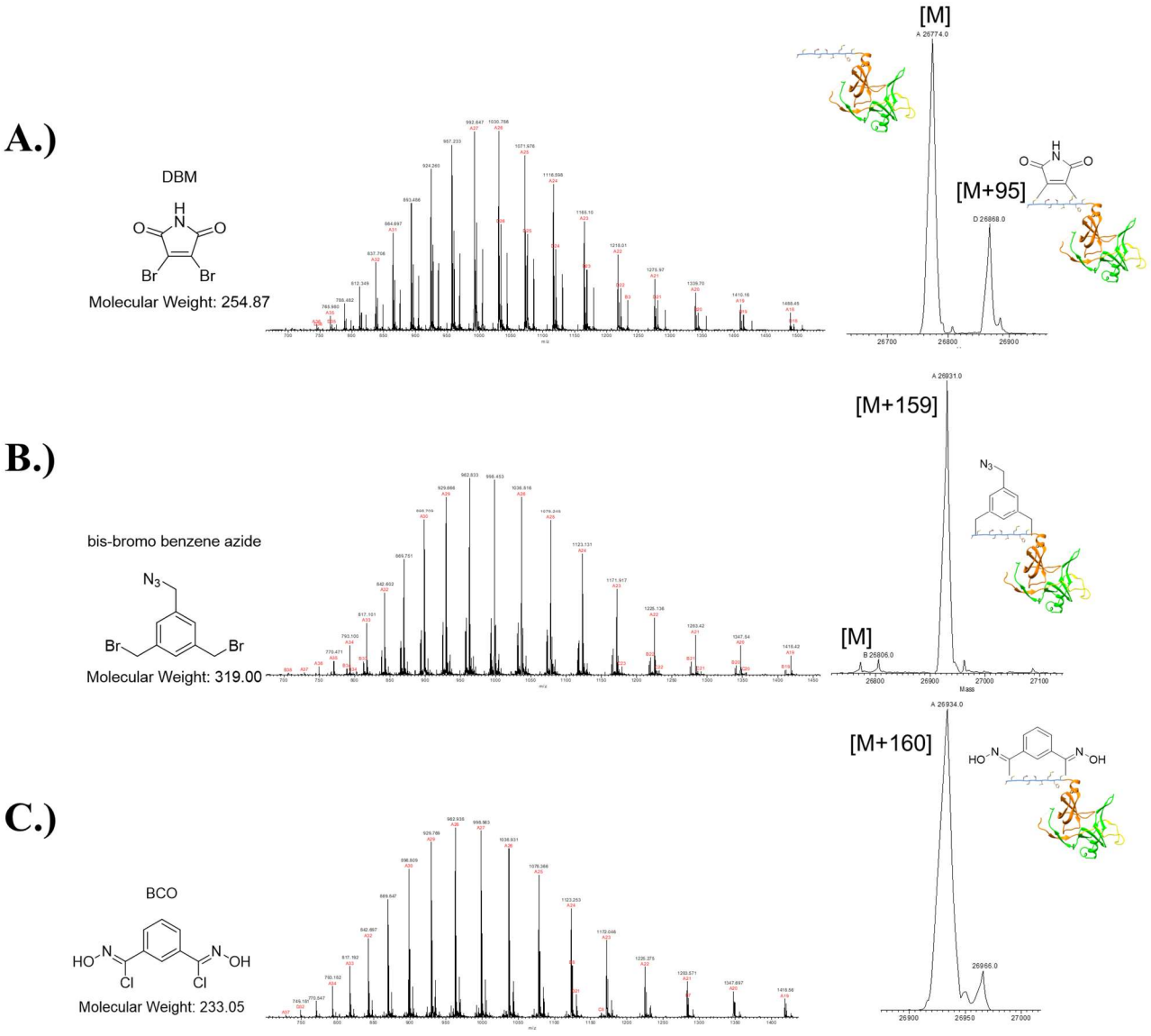


Figure 2-9: Deconvoluted LC/MS analysis of crosslinker comparison test. Dibromo maleimide (DBM) was the only crosslinker to yield the unlabeled product as the dominant product under identical conditions to DCA (A). The bis-bromo benzene azide tested exhibited similar labeling efficiency to DCA, with a less than 5% unlabeled pIII remaining after a 1-hour incubation a RT with 100 µM crosslinker (B). Bis-chlorooxime (BCO) was capable of crosslinking with higher efficiency than DCA. Shown in (C) is the labeling reaction of pIII with BCO for 10 minutes at RT with 10 µM BCO. Under these conditions, no unlabeled pIII protein was seen.

bis-chlorooxime at this concentration is reactive enough that it may be crosslinking other proteins beyond the pIII protein or crosslinking the pIII protein in ways that cannot be studied with this platform. As such, while bis-chlorooxime appears to be a superior crosslinker on paper, it is clear that further optimization and consideration must be taken before it can properly be used in phage display.

Crosslinker	Optimal Conc.	Optimal Time	TCEP Inhibited	Phage Toxic
DCA	100uM	1hr	No	No
Bis-bromo benzene	100uM	1hr	Yes	No
Bis-chloro Oxime	10uM	10 minutes	Yes	Yes
Dibromomaleimide	>100uM	N/A	Yes	N/A

Table 2-2: Compiled minimum labeling requirements determined for each potential crosslinker. Listed conditions indicate reactions where less than 5% pIII protein was seen.

2.6 Conclusions

In this section the functionality of the pIII protein to serve as a platform to test the suitability of small molecules for incorporation into phage display was assessed. The pIII protein has valuable potential to serve as a model of phage labeling, but there are clear and present drawbacks that must be assessed. Future work on this protein platform could yield a very powerful method of assessing warhead and crosslinker viability before moving on to phage work.

The first point that must be addressed is the yield and efficiency of purification. As seen earlier, the yield for this protein is exceptionally low and as such purification was incredibly challenging. Alternative methods such as periplasmic space extraction or inclusion body purification as seen in the literature[2], while unsuccessful in our experiments, may be optimized to yield a pIII product of higher purity. While the pIII peak seen in the LC trace earlier was distinguishable from other proteins and the UV signal indicated that

the pIII protein was the only protein present at quantifiable concentrations, it is undeniable that the presence of these unidentified contaminants could have unknown effects on crosslinker efficiency as well as small molecule reactivity. As such, this method currently serves as a qualitative analytical method.

The results of the bis-chlorooxime crosslinker testing also highlight a drawback of this platform the must be addressed as well. Based on the results shown here, bis-chlorooxime appears to be the superior crosslinker, however phage testing indicates that this crosslinker is challenging to implement. Current work is being done in the Gao lab to optimize bis-chlorooxime as a phage crosslinker[4] and it may yet prove to be an efficient crosslinker, however the complications associated were a factor that could not be identified using the pIII platform. For the future experiments discussed in this work, DCA was selected as the small molecule crosslinker of choice. This work, as well as previous successful work, indicates that DCA is a readily available, efficient, and nontoxic toward the M13 bacteriophage. As will be discussed, the DCA motif is synthetically stable and can be incorporated into small molecule scaffolds with relative ease.

Ultimately, the pIII protein was capable of serving its intended purpose as a phage display model, though drawbacks in the system were revealed when pushed in new directions. As noted previously, this platform does not account for any potential reactivity with other phage proteins. However, the pIII protein does contain a wide array of solvent exposed amino acids, so highly reactive molecules should be identifiable using the methods discussed here. What may prove to be an indeterminable factor with this method is the possibility of inter-pIII protein crosslinking. The M13KE bacteriophage used in phage display expresses 5 copies of the pIII protein in close proximity, though the exact

conformation is highly variable and changing. As such, a bifunctional molecule may be more likely to react with two separate pIII proteins in phage labeling. Once the first labeling occurs, it may serve to direct reaction of the molecule with another protein, as the localized concentration of pIII protein is now much higher than the solvent concentration of protein. This explanation serves to explain the apparent toxicity of bis-chlorooxime crosslinker despite no excess reactivity seen on the pIII protein. Despite these factors, pIII protein labeling can still serve as a useful metric and a type of "early-warning" system for phage display. As shown, since the protein is properly folded and functions as a normal pIII protein, it can be a model of "ideal labeling." As such, it can still serve as tool to compare and contrast crosslinkers and warheads and identify challenging properties of these molecules to support conclusions seen in phage-based experiments.

2.7 Experimental Methods

General Methods

All chemicals were purchased from Fisher Scientific or Sigma-Aldrich. All Nano Pure water was purified and dispensed from a Thermo Barnstead Nanopure water dispenser Sequencing samples were sent to Eton Biosciences for Sanger sequencing. All samples under 500 mL volume were centrifuged using Eppendorf 5424 R model centrifuges with appropriate rotors. Samples of 500 mL or larger were centrifuged using a Beckman Coulter Avanti J-E centrifuge with appropriate rotor. Primers for PCR reactions were purchased from IDT. All PCR and Ligation reactions were performed using available protocols from NEB. Restriction enzymes were purchased from NEB. All commercially available cell lines were purchased from NEB. All cells were grown in autoclaved LB media prepared from protocols available from Cold Spring Harbor protocols. All reactions performed in

PBS buffer used autoclaved buffer prepared from protocols available from Cold Spring Harbor protocols and were adjusted to desired pH using 1M HCl or 1M NaOH. All plasmid DNA was purified using a GeneJET Plasmid Miniprep Kit using the protocols provided within the kit. All DNA concentrations were determined using a Thermo Fisher Nanodrop 2000 UV/Vis Spectrophotometer. All proteins containing a 6xHis tag were purified using manually packed columns containing 0.5-1mL HisPur Ni-NTA resin using protocols available from Thermo Scientific. Following purification, all protein samples were desalted using PD-10 desalting columns for volumes >0.5mL. All samples of 0.5mL or less were purified using NAP5 desalting columns. Protocols were followed based on those provided with columns. Elutions of final protein mixtures were performed with 5x PBS buffer pH 8.0 unless otherwise noted.

All protein samples analyzed with LC/MS used an Agilent 6230 LC TOF mass spectrometer. LC/MS data were processed using Agilent MassHunter software package. Deconvoluted mass data was processed using mMass 5.5.0 and MagTran 1.03 b3 software. Mobile phases for LC/MS analysis were kept constant for all experiments as follows: 0.1% formic acid, 5% acetonitrile in H_2O (Solvent A) and 0.1% formic acid, 5% H_2O in acetonitrile (Solvent B).

All phage tittering discussed in this document was performed based on protocols provided in the Ph.D. Phage Display Library Instruction Manual, Version 3.0 provided by NEB. Recipes for all buffers and media used are contained within the referenced document.

Plasmid Construction and Transformation into E. Coli

A PET22b plasmid containing a reporter gene for eGFP was generously provided by the lab of Dr. Abhishek Chatterjee, the structure can be found in **Figure 2-2**. 5 mL

overnight cultures of ER2738 cells were inoculated with 1 µL of a phage library from a Ph.D.™-C7C Phage Display Peptide Library Kit from NEB and the tittered. Individual plaques were amplified and sequenced as discussed in the General Methods. The following primers were ordered from IDT:

Forward-5'ATTATTGAATTCTTTTTTGGAGATTTTCAACGTG 3'

Reverse- 5'ATTATTGCGGCCGCCACCAGAACCACCACCAGAG 3'.

Primers were used to amplify the pIII region from purified phagemid DNA in 50 µL reactions using *Taq* DNA polymerase PCR protocol available from NEB and resulting DNA construct was purified using a Monarch PCR & DNA Cleanup Kit. To summarize, in 1x *Taq* reaction buffer 0.2 mM dNTPs, 0.2 µM forward primer, 0.2 µM reverse primer, 500-1000 ng of template DNA, and 1.25 U of *Taq* polymerase were reacted in a 50 µL reaction mix. Reaction mixture was heated at 95°C for 30 seconds to denature template DNA. The reaction mixture then underwent 30 cycles consisting of heating at 95°C for 30 seconds, 58°C for 60 seconds, and 68°C for 7 minutes. After 30 cycles were completed, reaction mixture was kept at 68°C for 5 minutes before holding at 4°C. PCR reaction mixture was mixed with NEB Gel loading dye and was run on a 1% agarose gel containing 0.5ug/mL ethidium bromide at 100V until visible bands from gel loading dye were 1-2 cm from the end of the gel (~35 min). All agarose gels were run in 1x TAE buffer.

Following successful PCR, 17 µL of cleaned PCR product (~63.0 ng/µL) and 10 µL of plasmid template (~120.8 ng/µL) were digested with EcoRI and NotI restriction enzymes (~1 U of each) in NEB CutSmart buffer. Reaction was quenched using NEB Purple Gel Loading Dye. The online tool NEBcloner provided by NEB was used to calculate appropriate concentrations of restriction enzymes. To facilitate complete

digestion of both the insert and vector, DNA was digested at 37°C for 1 hour in commercially available Cutsmart Buffer. Digestions were confirmed and purified by running samples on a 1% agarose gel containing 0.5ug/mL ethidium bromide as mentioned previously, and then excising bands and purifying with a Zymo Research Gel Extraction kit.

Ligation was performed with 2 μL digested vector (21.6 ng/ μL) and 1 μL digested insert (41.7 ng/ μL) for a target 7:1 molar ration of insert: vector in 1x NEB Ligase Buffer. Ligation was performed using T4 ligase based on T4 Ligase protocol (M0202) for sticky end DNA from NEB. For this set of ligations, mixture was incubated at 16°C overnight before heat shocking at 65°C for 10 minutes to inactivate T4 Ligase. Ligation was immediately quantified and used for electroporation.

Electrocompetent TOP10 cells were freshly prepared as follows. 10mL overnight cultures of TOP10 cells were used to inoculate 1L of LB media. Inoculated media was incubated at 37°C while shaking until OD was measured to be ~0.5-0.6 at 600 nm. The entire culture was centrifuged at 6300 rpm for 10 minutes at 4°C. Cell pellets were resuspended in total 15mL of 15% glycerol and then washed 3x with 300mL of cold 15% glycerol. After the final wash, cell pellet was resuspended in ~2mL cold 15% glycerol. 50 μL aliquots of this resuspension were flash frozen and stored at -80°C until needed. A pUC19 vector containing the AmpR gene was purchased from NEB and used to confirm electrocompetence of cells and to determine the transformation efficiency of cells.

All electroporation of electrocompetent cells utilized a BioRad system consisting of a BioRad Capacitance Extender, BioRad Pulse Controller, and BioRad GenePulser. 50 μL aliquots of TOP10 cells were diluted up to 100 μL with 15% glycerol and electroporated

with 5 ng of ligated DNA. Electroporation utilized a pulse of 1500 V, 200 Ω resistance, and 25 μF capacitance. Electroporated cells were diluted up to 1 mL with LB media and incubated at 37°C for 1 hour before plating 100 μL on 10% agar plates containing 100mg/mL ampicillin. DNA was extracted from resulting colonies a GeneJET plasmid miniprep kit and samples were sequenced. One sequence was selected from sequencing results and used to transform BL21 (DE3) cells using NEB transformation protocol (C2527). To summarize, BL21 (DE3) cell aliquots were thawed on ice for 10 minutes and 100ng of miniprepped DNA was added to 50 μL of thawed cells. The mixture was incubated on ice for 30 minutes before heat shocking at 42°C for 10 seconds. The mixture was returned to an ice bath for 5 minutes before diluting up to 1mL with LB media. Cell outgrowth in LB was incubated at 37°C for 1 hour before Transformed BL21 (DE3) were subject to identical plating techniques as previously noted. Plasmid DNA extraction and sequencing using previously noted methods was used to verify the presence of the plasmid. Successfully transformed colonies were amplified and 1mL aliquots were stored at -80°C until needed.

Protein Extraction and Purification

The pIII protein was purified from BL21 (DE3) cells. 10mL of overnight cultures in LB of BL21 (DE3) containing the plasmid construct with the pIII gene were diluted 100x in LB media and grown for ~2 hour until OD_{600} was ~0.5-0.7. A small volume of IPTG in DMF was added to a final concentration of 1 mM IPTG in the culture and culture was incubated overnight (~12 hours) while shaking at room temperature. Following incubation, cell cultures were spun down at 5000 rpm for 10 minutes and 4°C. Resulting cell pellets were resuspended in 20-40 mL of cell lysis buffer containing 20 mM tris base,

300 mM NaCl, 10 mM imidazole pH 7.4 and divided evenly between two 50mL Falcon tubes. Resuspended cells were then lysed at ~4°C using a probe sonicator. After saving <1 mL of crude lysate, cell lysate was centrifuged at ~7000 rpm for 40-60 minutes at 4°C and supernatant was collected.

Supernatant was filtered using 0.45 μm syringe filters and ran through columns containing HisPur Ni-NTA resin. The column was equilibrated with 20 mM sodium phosphate, 300 mM sodium chloride (PBS) with 10 mM imidazole; pH 7.4 before loading samples. The sample was washed initially with 1 mL buffer containing 300 mM NaCl, 20 mM tris base and 25 mM imidazole; pH 7.4. All washes and elutions used 1 mL aliquots of buffer. The column was then washed with the same buffer 3x before gradually increasing concentrations of imidazole, from 25 mM up to 250 mM. The concentration of imidazole in PBS was increased by 15 mM with each wash. The final elution buffer was the same as the wash buffers except with 250 mM. The elution buffer was used to wash the column 3 times.

Column fractions as well as diluted lysate were analyzed via SDS-PAGE analysis and were run on 15% polyacrylamide gels containing 10% SDS with a 4% polyacrylamide stacking gel also containing 10% SDS. Gels were made using 150 M Tris buffer pH 8.8. Gels were run in 250 mM Tris buffer with 10% SDS. All gels were run using a voltage gradient, starting at 80 V before being brought to 120 V by increasing the voltage 10 V every 15-20 minutes until bands were visibly ~2 cm from the bottom of the gel. Gels were stained with Coomassie Blue dye and destained with ~20 MeOH 10% acetic acid in H_2O to identify fractions with strongest bands corresponding to the pIII protein. All SDS-PAGE gels referenced in this document utilized this protocol. Three fractions containing the most

intense bands were then pooled and desalted using a PD10 desalting column. Eluent from the desalting column was tested for protein concentration by measuring the A280 using a Thermo Fisher Nanodrop 2000 UV/Vis spectrophotometer.

Confirming Protein Mass and Conformation Using LC/MS

All protein samples were diluted ~50 ng/µL with PBS pH 7.5 before being loaded onto the LC. All LC was performed on an Agilent Technologies 1260 series LC setup containing a high performance degasser (G4225A), binary pump (G1312B), autosampler (G1329B), thermostatted column compartment (G1316A), a diode array/multiple wavelength detector (G1315D), and an Agilent Technologies 6230 Time of Flight (TOF) mass analyzer (G230)LC utilized a Phenomenex Aeris 3.6 µm wide pore XB-C8 20, 100x46 mm column for all protein samples. The table below shows the solvent gradient used for all protein samples studied via LC/MS. Samples were ionized in positive ion mode using electrospray ionization (ESI), a 5 µL injection volume with 200 µL/min draw and eject speed, solvent flow rate of 0.200 mL/min, extended dynamic mode in mass range was 200 – 2000 m/z, temperature of drying gas = 325 °C, flow rate of drying gas = 8 L/min, pressure of nebulizer gas = 35 psi, capillary = 4.286 µA, fragmentor = 175 V, and octupole rf voltages = 750. Data was collected at 1 spectra/s.

Time (min)	% Solvent A	% Solvent B	Flow Rate (mL/min)
0.00	95.0	5.0	0.200
3.00	95.0	5.0	0.200
10.00	5.0	95.0	0.200
18.00	5.0	95.0	0.200
20.00	95.0	5.0	0.200
25.00	95.0	5.0	0.200

Table 2-3: LC/MS conditions for analysis of pIII protein samples.

UV absorbance was monitored simultaneously at 280, 254.4, and 220 nm. Only peaks in the total ion count (TIC) chromatogram that had a corresponding absorbance peak at 280 nm and 220 nm were extrapolated for deconvolution.

In this document, all observed mass spectra for the pIII protein experienced a potential ± 1 difference in mass from the expected value.

Comparing Cysteine Crosslinkers Using the pIII Based Platform

Stock solutions of DCA (100 mM), bis-bromo benzene (21 mM), and dibromomaleimide (100 mM) were prepared in pure CH_3CN. Stock solutions of each compound were diluted to final concentrations of ~2 mM for working stocks. A 100 mM stock of TCEP was prepared in Milli Q water. Protein modification was designed to simulate phage modification protocols found in the Ph.D. Phage Display Library Instruction Manual provided by NEB. To facilitate removal of TCEP, 0.1-0.5 mL aliquots of ~10 μM protein stock were reduced with a final concentration of 1mM TCEP added from the working stock. Reduction took place at 37 °C for 1 hour to facilitate total reduction. Following TCEP reduction, the reduced protein aliquot was cleaned via filtration through a NAP5 desalting column to remove TCEP and prevent interactions between TCEP and crosslinker molecules. Elutions from the column requires excess buffer, so final concentrations of ~5 μM were used as working concentrations for modification.

Modifications using DCA, bis-bromo benzene azide, and dibromo maleimide were all performed in 5% CH_3CN. Reactions all took place at room temperature. Concentration and time of reaction were varied as needed for each experiment. Regardless of concentration used per experiment, 5% CH_3CN was kept constant, adding pure CH_3CN as needed.

Modification with bis-chlorooxime required a separate protocol due to poor water solubility and negligible solubility in CH_3CN. Samples of bis-chlorooxime crosslinkers were synthesized in house and provided by Dr. Fa-jie Chen. A 60 mM stock of bis-chlorooxime was provided in acetone, which was diluted to 2 mM in acetone. 5 µL of the 2 mM stock was added to a 1.5 mL centrifuge tube and all solvent was removed with a Heidolph Laborota 4001 efficient rotary evaporator followed by incubation under vacuum for 20 minutes. 100 µM of reduced protein stock was then added to the tube containing solid bis-chlorooxime and the small amount of bis-chlorooxime was resuspended in PBS buffer pH 8.0 only. It is important to note that acetone was not believed to impact the reaction of the bis-chlorooxime, however acetone was to be avoided in phage modifications and so it was removed from this experiment to best model a phage modification. Volume of bis-chlorooxime evaporated was reduced when lower concentrations were needed. Time point was varied as necessary per experiment, but all modifications occurred at room temperature.

For all modifications, no quenching reaction was used to stop reaction between the pIII protein fragment and desired crosslinkers. Instead, protein modifications were carried out at small scale (100 µL reactions) and injected directly into the LC/MS upon completion of the reaction to minimize the number of peaks present in the chromatogram. This method was used for all crosslinker candidates. LC/MS methods followed those shown in the previous table.

2.8 References

(1) Deng, L.-W.; Perham, R. N. Delineating the Site of Interaction on the PIII Protein of Filamentous Bacteriophage Fd with the F-Pilus of Escherichia Coli. *Journal of Molecular Biology* **2002**, *319* (3), 603–614.

(2) Martin, A.; Schmid, F. X. Evolutionary Stabilization of the Gene-3-Protein of Phage Fd Reveals the Principles That Govern the Thermodynamic Stability of Two-Domain Proteins. *Journal of Molecular Biology* **2003**, *328* (4), 863–875.
(3) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-Encoded Combinatorial Chemical Libraries Based on Bicyclic Peptides. *Nat Chem Biol* **2009**, *5* (7), 502–507.
(4) Chen, F.; Zheng, M.; Nobile, V.; Gao, J. Fast and Cysteine-Specific Modification of Peptides, Proteins and Bacteriophage Using Chlorooximes. *Chemistry A European J* **2022**, *28* (20).

CHAPTER 3: EVALUATING ARYL-SULFONYL FLUORIDE AS A TYROSINE TARGETING COVALENT WARHED FOR USE IN COVALENT PHAGE LIBRARIES

3.1 Aryl-sulfonyl fluoride and SuFEX based warheads

Pivotal work from Bogyo et al. showed that vinyl sulfone could be successfully implanted onto phage platforms[1]. Based on these results, it became desirable to investigate more versatile sulfur-based warheads. Sulfur (VI) Fluoride Exchange is a well-documented and powerful class of biorthogonal click reactions[2]. The work in this document was particularly influenced by an in-depth kinetic study of sulfur based warheads and their reactivity with amino acids from Neil Grimster[3]. Aryl sulfonyl fluorides were found to be weak electrophiles stable to a wide range of synthetic transformations. These compounds were also found to react with a wide range of amino acids, with tyrosine being of particular interest in this work. The sulfonate product formed is stable while the fluoride leaving group is relatively benign. As shown By Mukherjee et al. these SuFEX based warheads have a wide range of reactivity with tyrosine (164-1902 $M^{-1}hr^{-1}$) that is tunable via modification of the aromatic ring.

Figure 3-1 below shows the structures of two SuFEX warhead-containing molecules synthesized in house by Dr. Fa-Jie Chen. A sulfonyl fluoride and fluorosulfate based warhead, hereby referred to as SF1 and FS1 respectively, were both provided as solid compound and resuspended to ~50 mM working stocks in pure CH_3CN. Sulfonyl-fluoride was the warhead of interest in this experiment, as its highly reactivity and favorability towards tyrosine was desirable to implement into a phage library. Fluorosulfate was chosen as reference point since it is less reactive than sulfonyl fluoride

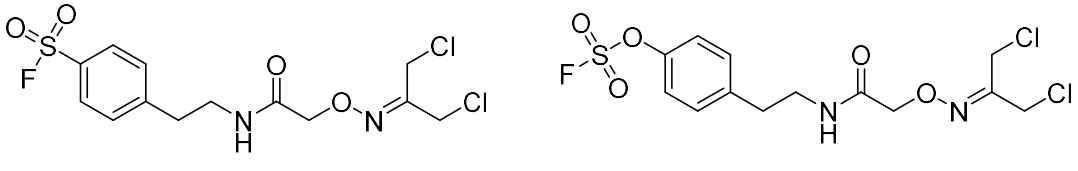


Figure 3-1: Molecular structures of SF1 and FS1.

and has been successfully implemented in biological systems already. Fluorosulfate has seen some use as a genetically incorporated non-canonical amino acid[4] and has even been used in early phage display experiments[5]. As such, the goal of the experiments detailed here was to assess the viability of aryl-sulfonyl fluoride as a potential warhead candidate to be implemented into phage display, and yield the desired product shown in **Figure 3-1.**

3.2 Peptide and Protein Modification with aryl-SF warhead

As shown in Chapter 2, the pIII protein was shown to be a valuable tool for developing a qualitative understanding between a labeling molecule and the phage protein. As mentioned previously, this method is not an exact model of the phage labeling environment, however it still can serve as strong starting point for warhead suitability studies. Therefore, this platform was used to begin the suitability study of SF1.

The bifunctional molecule SF1 was synthesized to contain two major components: a DCA-based crosslinker and an SF based warhead. The reactivity of the DCA-based crosslinker was well studied, however the SF component remained a mystery. As such, it was necessary to determine the potential intermolecular reactivity that could occur before crosslinking. **Figure 3-2** shows two SuFEX warhead-containing

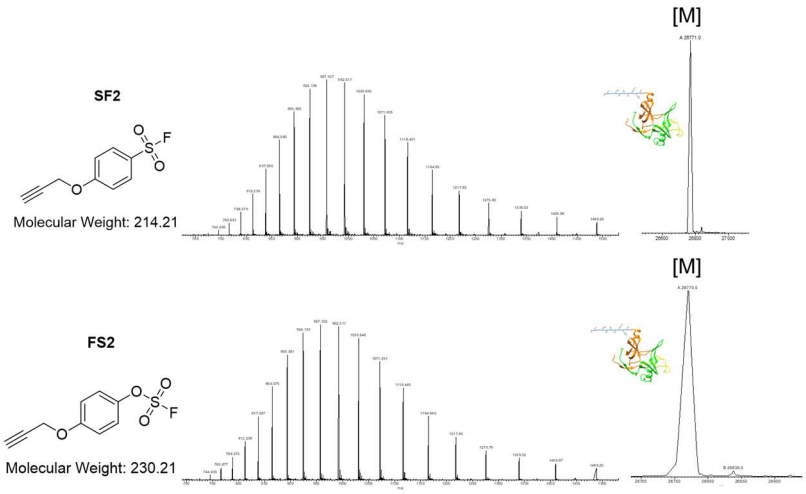


Figure 3-2: Deconvoluted LC/MS analysis of intermolecular reactions between warheads and pIII protein. 1 mM sulfonyl fluoride (top) and fluorosulfate (bottom) molecules without a crosslinker were incubated with 4 μM pIII protein for 1 hour at 37°C. In both instances, only a single signal was extracted from the A peak in the TIC corresponding to an unmodified pIII protein.

molecules used to test these interactions. The sulfonyl fluoride and fluorosulfate molecules shown, hereby referred to as SF2 and FS2 respectively, both contained a nonreactive alkyne group in place of a DCA motif and were provided by Dr. Emma Ste. Marie from the lab of Dr. Eranthie Weerapana. **Figure 3-2** also shows the deconvoluted masses of the A peak following a 1-hour reaction of the pIII protein fragment with 1mM of each respective molecule. In both cases, no clear evidence was seen to indicate any kind of labeling had occurred. In each, small higher-molecular weight (HMW) peaks were seen, however the masses were too small to correspond to a singly labeled protein and therefore were considered contaminate proteins. With these HMW peaks accounted for, each incubation yielded only a single peak corresponding to the unreduced protein, which is conclusively indicative of no reaction occurring.

Before reactions were carried out it was necessary to validate fluctuations in small molecule concentration would not influence labeling. The goal was to ensure that at least 500 μM solutions of both warheads could be made in 20% CH_3CN solutions. Again, this was done to best mimic the conditions that could be used in a phage labeling experiment.

To best determine this, H^1 NMR was used with an internal standard in place to determine solvating concentrations of both molecules in 20% CH_3CN. As shown in **Figure 3-3** under these conditions SF1 had a solvating concentration of ~1.2 mM while FS1 had a solvating concentration of ~520 µM. Both of these values were deemed sufficient for protein and phage labeling conditions.

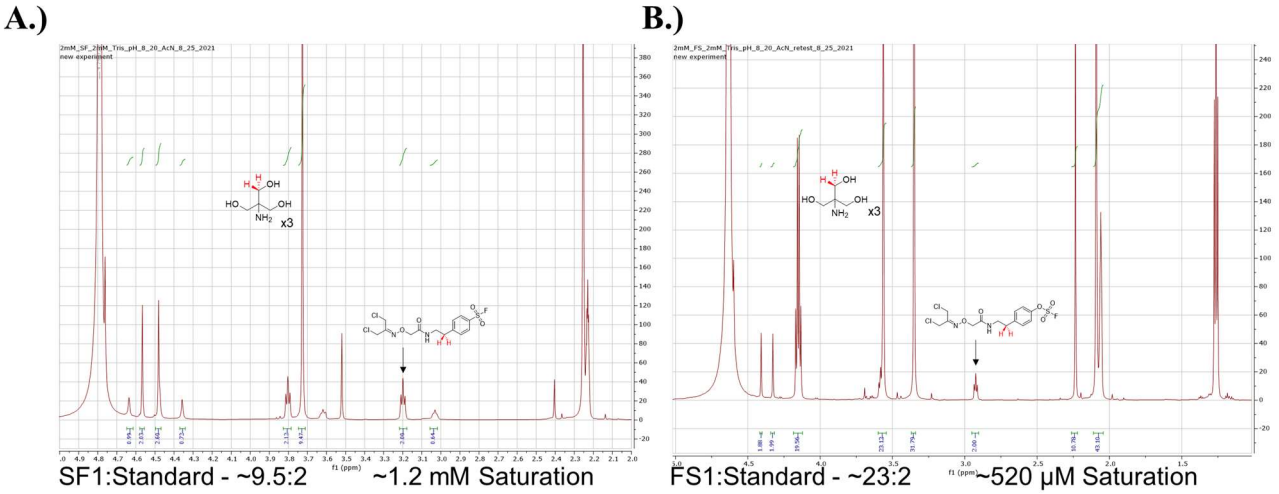


Figure 3-3: Crude NMR results determining warhead solubility in 20% CD_3CN. Signals used as reference points in SF1 (A), FS1 (B), and a Tris-base internal standard correspond to the protons highlighted in red on each structure. Warhead was added in excess such that if fully soluble the concentration would be expected to be 2 mM, any actual concentration beneath this point would be the saturating concentration. The concentration of tris base in solution was known based on the mass of tris base used to make a stock solution and the ratio between the integrations for the tris base signal and the warhead signal was used to determine the concentration of warhead in solution. Listed beneath each spectrum is the ratio of tris base to warhead protons based on the integrated peaks as well as the calculated saturating concentration based on this ratio.

Labeling reactions were conducted on the pIII protein using SF1 and FS1 and were analyzed using the same methods that were used to analyze the crosslinker optimization studies. The goal of these sets of experiments was to identify potential intramolecular interactions that may occur between the SuFEX warheads and the pIII protein following a successful labeling. As such, to facilitate the highest degree of labeling possible, each warhead was used in heavy excess by labeling with concentrations

close to the solvation point since even a 1 mM would not yield intermolecular reactivity, as shown in **Figure 3-2**.

In order to best decipher the relationship between protein and warhead molecule, pIII protein samples were incubated with the appropriate warhead for 8 hours, with timepoints being taken every hour and subjected to LC/MS analysis. Since all masses were extrapolated from the A peak in the TIC spectra, each was represented by a single peak in the UV absorbance at 280 nm. As such, the best estimates for quantification were derived from the crude deconvolution peak areas. These values were not used to determine definitive concentrations, but instead as a helpful metric to observe the relationship between starting material and product in each reaction. In this instance, product refers to the combined integrations of all peaks where single labeling of the protein via the DCA motif on the warhead molecule occurred, regardless of the state of the SuFEX warhead. **Figure 3-4** shows the documented relative relationship between the

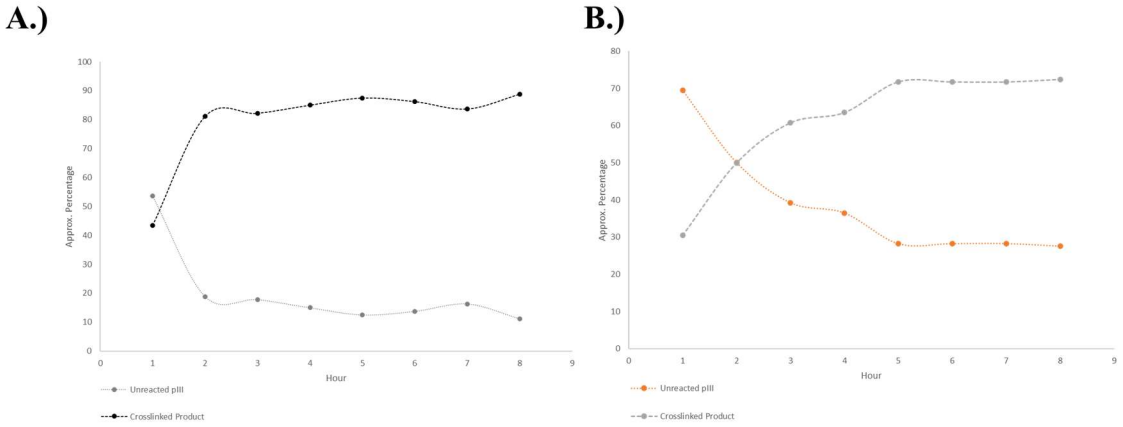


Figure 3-4: SuFEX warhead pIII protein labeling kinetics. 4 μM pIII protein was labeled in 5% CH_3CN, 95% PBS pH 8.0 with a saturated solution SF1 (A) or FS1 (B). All reactions were run at room temperature, pH 8 and aliquots of the reaction mixture were injected onto the LC/MS at each time point for analysis. To determine approximate percentage of each product, the deconvoluted peak for each respective mass was integrated and the relationship between the labeled and unlabeled protein peak integrations was used to determine the percent labeled. As noted, it was concluded that reactions run in 20% CH_3CN would be faster and yield higher percentages of labeled product than what is displayed.

unlabeled pIII protein and the modified protein over the 8-hour reaction period. Both cases showed a maximum labeling of 70-80% of protein in solution, however since the M13KE phage displays 5 copies of the pIII protein this was deemed acceptable to label all phage particles. As expected, FS1 did experience a slower rate of labeling than SF1, but this can simply be attributed to the lower concentration that was used. It should be noted that this experiment was conducted in only 5% CH_3CN instead of the more optimal 20% CH_3CN. Since both molecules experience more favorable solubility at higher concentrations of CH_3CN, it would be expected that labeling yields would be higher and occur at a faster rate under these conditions. Since 5% CH_3CN yielded acceptable results, it was concluded that 20% CH_3CN would be more than sufficient for the purposes of pIII labeling.

Figure 3-5 shows the deconvoluted A peaks from two time points within this range for each warhead. The time points shown in **Figure 3-5** serve as representations of the general trends noted across the 8 hours for each. FS2 was slow reacting but yielded results on par with the expected assumptions made based on literature studies. After 3 hours, the most significant protein signal withing the A peak was an [M+329] peak, which was indicative proper crosslinking without intramolecular reactivity. A smaller peak equating to [M+308] did indicate that some intramolecular reactivity may have occurred, however the intensity of this signal did not increase over time, and it never became the major product observed. The labeling data for SF1 was not as concise as for FS1. As early as 1 hour following small molecule addition, two definitive HMW product peaks were observed. The expected [M+312] signal was present in the deconvoluted spectra; however, it was not the dominant HMW product. Instead, a [M+295] signal was

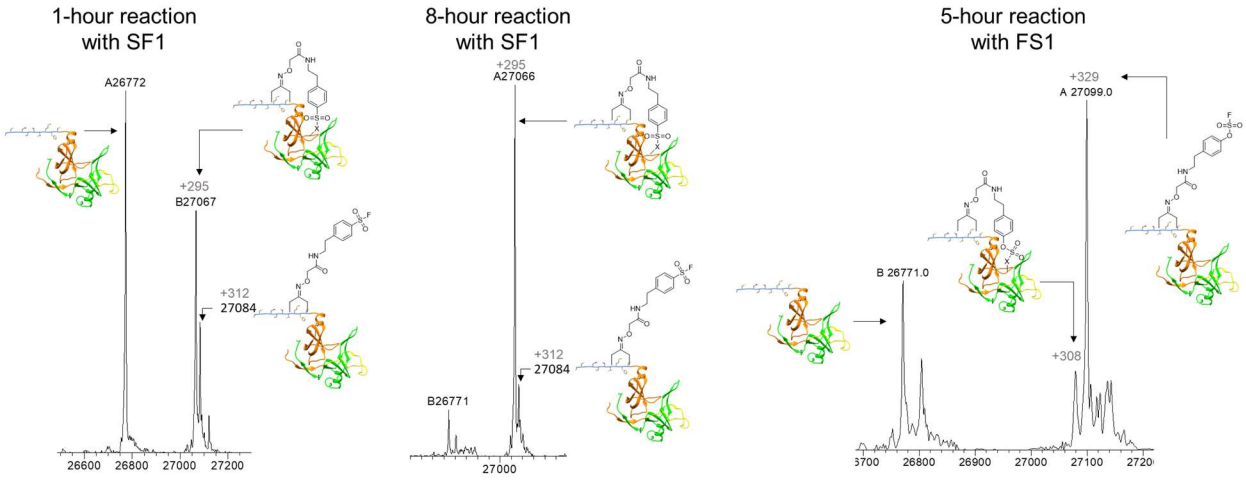
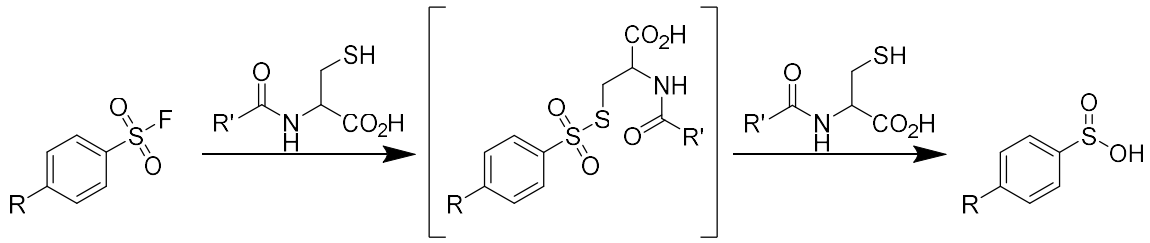


Figure 3-5: Deconvoluted LC/MS analysis of intramolecular reactions between SuFEX warheads and the pIII protein. Shown are isolated time points from the experiment detailed in Figure 3-4. Shown are the 1-hour timepoint for SF1 (left), the 8-hour timepoint for SF1 (center) and the 5-hour time point for FS1. For SF1 an incorrectly labeled pIII protein was the dominant product, whereas the correctly labeled product was dominant for FS1 at all time points. It was initially believed that the incorrectly labeled product resulted from an intramolecular reaction between the SuFEX warhead and a nucleophile on the pIII protein. However, with the observed mass often being ± 2 mass units from the expected due to overlapping signals, it was also determined that sulfinic acid formation stemming from a reduction of the warhead would yield a similar change in mass.

the dominant signal in each time point. This signal also appears to increase in intensity over time as the [M+312] signal decreased in intensity. This signal was initially believed to represent the intramolecular reaction of the SF warhead with an amino acid within the pIII sequence, however this was expected to result in a [M+292] peak. It was determined that a second alternative was possible: reduction of the SF to sulfinic acid. This would yield an [M+294] signal, which is in line with the results seen. **Scheme 3-1** shows a potential mechanism of SF reduction; however, this mechanism requires a free thiol. It was seen that following the 8-hour reaction, all unlabeled pIII protein reformed the disulfide bond, cyclizing the displayed peptide and so the source of the reducing agent was believed to have come from one of the unidentifiable contaminant proteins in solution. As such, a different approach would need to be taken to study SF reactivity.



Scheme 3-1: Proposed reaction scheme for the reduction of SF to sulfinic acid. In the reaction scheme, the addition of a second thiol would result in a second disulfide bond forming between the two thiols as well. In the case of the pIII protein, it was proposed that each thiol in the reaction could correspond to one thiol on either end of the C7C sequence. Based on the amount of oxidized pIII protein with a reformed disulfide bond being lower than expected, it was concluded that a second, unidentified reducing agent had to present in the solution.

To ensure that intramolecular cyclization was not occurring with the N-terminus and to help confirm the identity of the improperly labeled product in the SF1 incubation, a peptide labeling experiment was performed. In place of the pIII protein, a synthesized peptide with the sequence ACPFPASWC and a variant with an acylated N-terminus were generously provided by Dr. Rahi Masoom Reja for test labeling. **Figure 3-6** shows the comparison between the reduced, non-acylated and acylated peptides following incubation with 1 mM SF1. While much of the peptide was oxidized, two distinct peaks were seen in each case, equating to an [M+310] peak of the labeled peptide, and an [M+292] peak equating to a sulfinic acid product. Since no crosslinking was seen between peptide molecules, and the reduced mass product was seen with the N-terminus acylated peptide as well, this strengthened the theory that SF1 was not reacting intramolecularly with the pIII protein, but instead being reduced by an unknown source.

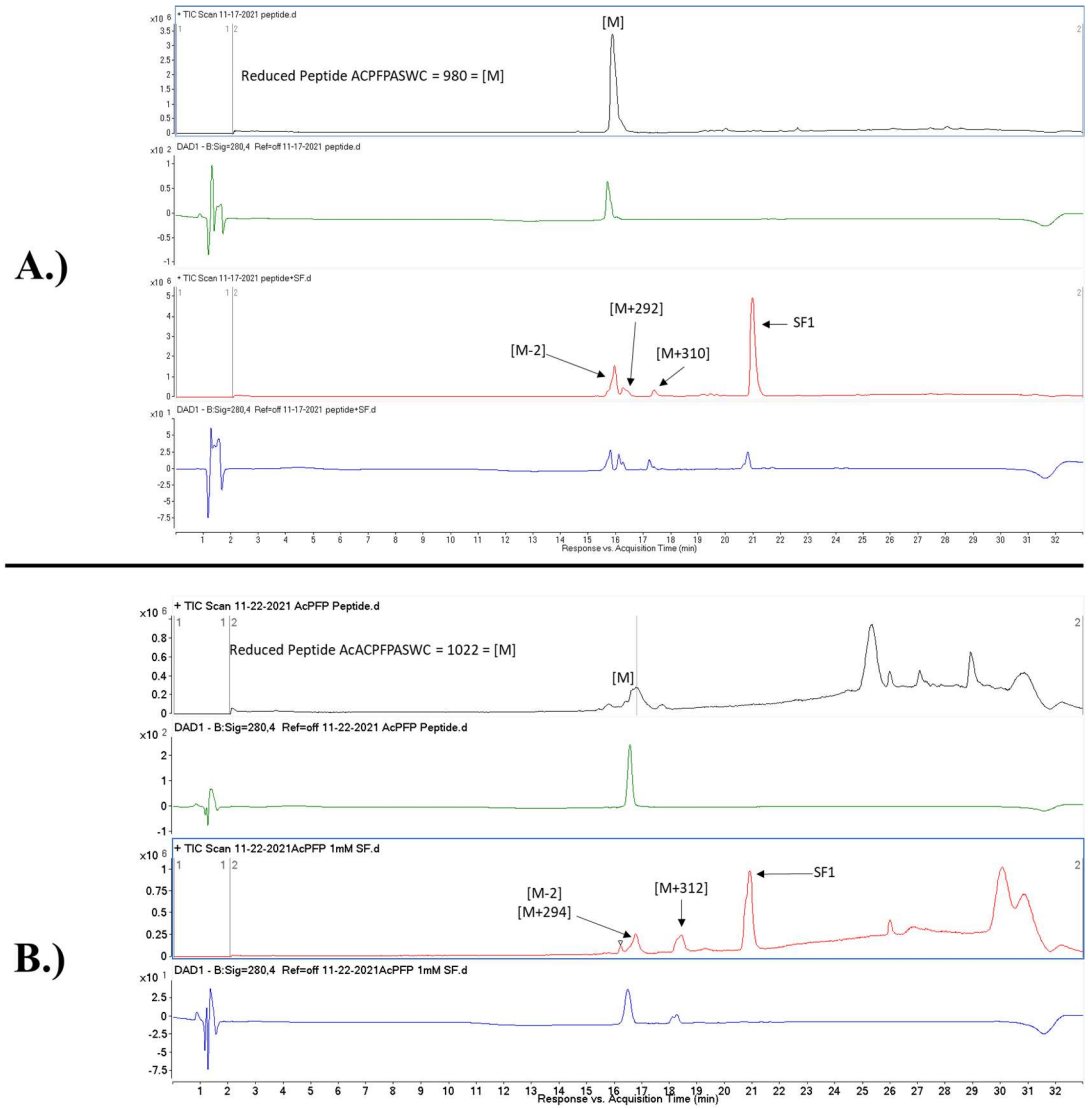


Figure 3-6: Labeling of acylated and non-acylated peptide with SF1 to identify intramolecular reactivity. SF1 was used to label the peptide ACPFPASWC (A) as well as the N-Acyl derivative of this peptide (B). Shown are the spectra for TIC and absorbance at 280 nm for the unlabeled PFP peptide (top) and the labeled PFP peptide (bottom). In each reaction 200 μM peptide was either injected directly onto the LC/MS in PBS pH 7.5 or was reacted with 1 mM SF1 first. In both cases, two masses were seen corresponding to the labeled product: one for an intact SF warhead and one for a reduced warhead. The presence of this second mass after reacting with the Ac-PFP peptide verified that the source of this unexpected decrease in mass was not due to intramolecular reactions, as this derivative had not available nucleophiles that the SF warhead could react with.

3.3 Aryl-SF Reactivity Profiling via NMR

With the stability of the intact warhead becoming an unforeseen complication, a better understanding of the kinetic factors associated with the reaction of aryl-SF and tyrosine was desired. To best understand this interaction H^1 NMR was used to monitor the progress of the reaction. Two commercially available molecules were chosen to model this interaction: benzene sulfonyl fluoride (BSF) and biotin tyramide (BTyr). Approximately equimolar amounts of both molecules were reacted in 20% CD_3CN and time points were taken hourly to monitor the reaction progress. To simplify data interpretation and best ensure minimization of signal overlap, the aromatic region of each spectrum (6.8-8.3 ppm) was analyzed for each. **Figure 3-7** shows the aromatic region of the obtained NMR spectra in this experiment at t=0 hrs. and t=22 hrs. The timepoint t=22 hrs. yielded more signals than expected for the fully formed product, therefore a second product must have been formed. Since the components of this reaction were extensively

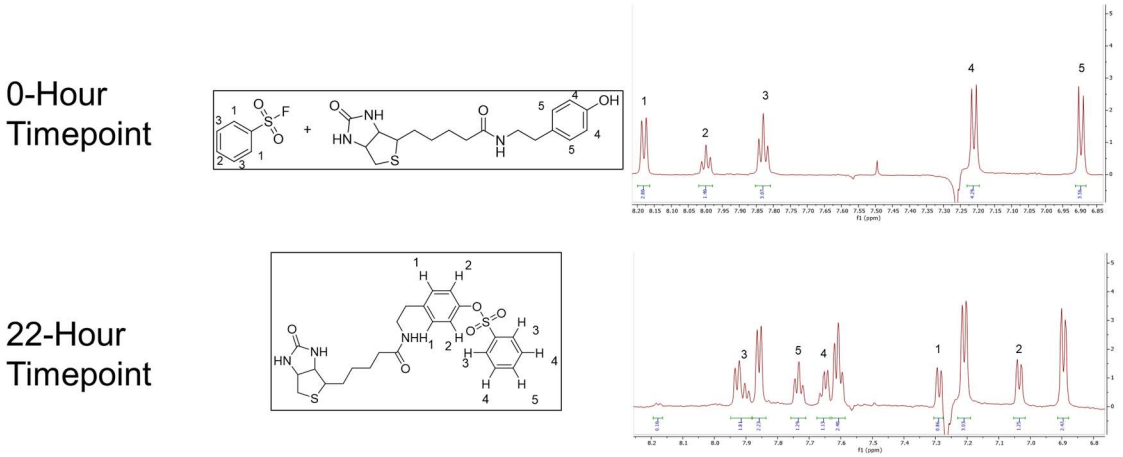


Figure 3-7: NMR analysis of benzene sulfonyl fluoride (BSF) reaction with biotin tyramide (BTyr). 840 μM BSF was incubated with 1mM BTyr at 37°C for 22 hours, with time points taken every hour for 8 hours and the final timepoint taken at 22 hours. Only the aromatic region from 6.8 to 8.2 ppm is shown to easily distinguish product and starting material peaks more easily. For quantification of product, the peak labeled Peak 2 in the 22-hour timepoint was compared to an internal standard of 600μM tris base to determine concentration of the product shown in solution. Unlabeled peaks in the 22-hour timepoint represent peaks from the hydrolyzed BSF, as determined by controls without BTyr.

controlled, unlike with the protein modifications, it was theorized that the likelihood of reduction would be minimal and that the second product seen was most likely a hydrolysis product.

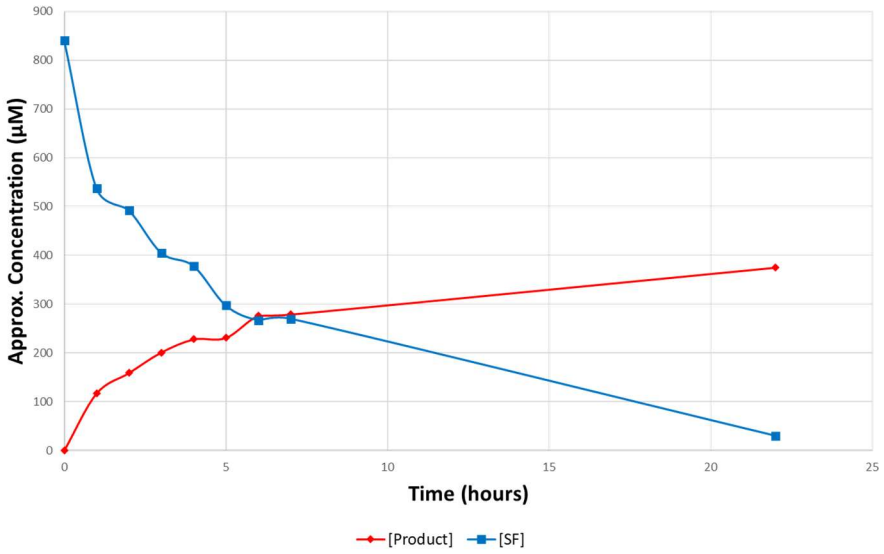


Figure 3-8: Approximating reaction speed of BSF with BTyr via NMR. Peaks corresponding to BSF and the product formed from the reaction of BSF with BTyr shown in Figure 3-7 were integrated at each time point and compared to the integrated peak of a tris base internal standard. 600 µM tris base was included in all reactions as a reference point. 840 µM BSF was incubated with 1mM BTyr at 37°C for 22 hours, with time points taken every hour for 8 hours and the final timepoint taken at 22 hours. The $t_{1/2}$ of the reaction was determined to be approximately 4 hours; however, the final concentration was determined to be only 375 µM, despite starting with 840 µM BSF. Remaining BSF was converted into nonreactive, hydrolyzed product instead of the proper reacted product.

Figure 3-8 shows the concentration of only the starting material and intact product over the course of the reaction, as determined by reference to a nonreactive internal standard. Here, it becomes clear to see that, despite the literature values, only a small amount of intact warhead remained at each time point. With the concentration of intact product increasing slower than the concentration of starting material decreased, it became evident that warhead hydrolysis was occurring at a far more rapid rate than anticipated. When incubated under identical conditions but without the addition of BTyr the hydrolysis of SF appears to occur at a rapid rate at 37 °C, with a $t_{1/2}$ of only ~4 hours

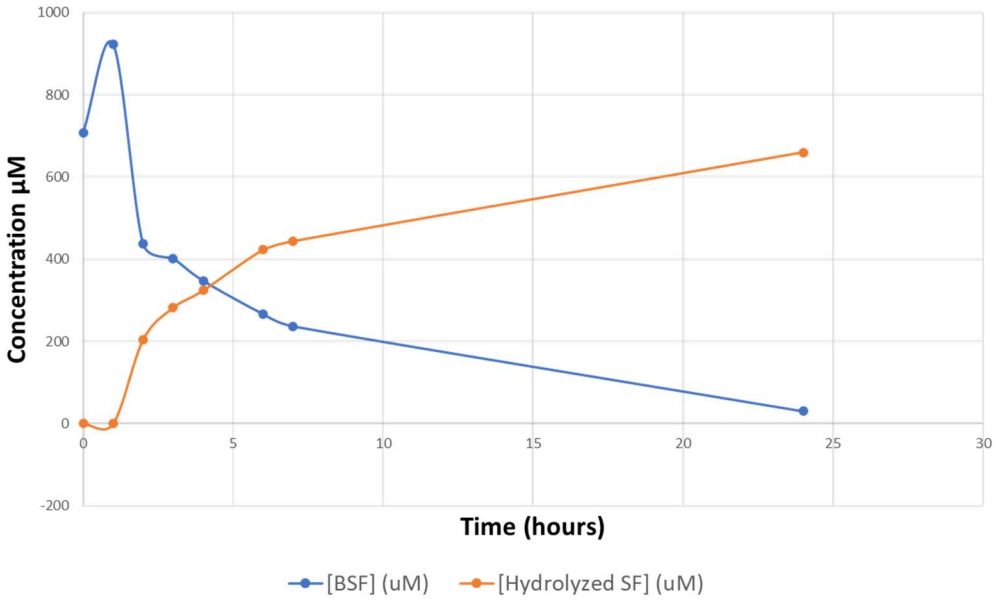


Figure 3-9: Approximating hydrolysis speed of SF warhead using NMR. BSF was incubated without BTyr under the same conditions as the experiment detailed in Figure 3-8. 600 µM Tris base was included in all reactions as a reference point. ~708 µM BSF was incubated at 37°C for 24 hours, with time points taken every hour for 7 hours and the final timepoint taken at 24 hours. Signals associated with BSF and the hydrolyzed product were selected and integrated, with the same signal integrated at each time point. The rate of hydrolysis appeared to have a similar $t_{1/2}$ compared to the reaction with BTyr, despite being the major product when the reactions are competing. It was thus concluded that the speed of each reaction was comparable, though the reaction conditions favored the formation of the hydrolyzed product.

as shown in **Figure 3-9**. However, this is fairly comparable to the $t_{1/2}$ seen for the reaction with BTyr even though the hydrolyzed product is dominant when the two are competing. Therefore, the reaction rates between the nucleophilic attack of the SF and its hydrolysis are likely comparable, with the reaction conditions favoring hydrolysis. However, this would go against what can be found in the literature, which would indicate that aryl-SF is actually resistant to hydrolysis. While the observed hydrolysis rate may have been enhanced by the increased temperature which samples were incubated at, the possibility of both hydrolysis and reduction of the warhead proved to be concerning nonetheless for the future of SF as a suitable warhead. While this proved problematic, these factors decreased concerns that the warhead may react with the phage particle, and while a

definitive conclusion could not be drawn for the suitability of SF1 yet, it was determined that moving to labeling of phage particles could yield better results.

3.4 Phage Labeling with Aryl-SF

To best determine the effect SF1 labeling would have on phage, a set of labeling experiments using the M13KE phage were designed. To best account for all possible amino acid reactivities, a phage library was used for labeling experiments instead of one discrete construct. For the purposes of SF1 labeling, a C5C M13KE phage library provided by Dr. Mengmeng Zheng was used as it was readily available and contained a high diversity of displayed sequences to best account for potential changes in reactivity. Since it was known that water solubility of SF1 was problematic, phage labeling reactions had to occur in a mixture of water and organic solvent. Based on previous successes, CH_3CN was chosen as the solvent of choice. In order to accurately determine phage output following modification, solvent susceptibility of phage was tested. Reduced phage were incubated at 30 °C for 3 hours in both 5% and 20% CH_3CN. Phage tittering following modification indicated that for 5% and 20% CH_3CN the average output of C5C phage was $1.29\pm0.15x10^{12}$ pfu/mL and $1.31\pm0.32x10^{12,}$ respectively. This verified that there was no significant drop in phage population with increasing concentrations of CH_3CN up to 20%, and thus this was set as the organic solvent limit in labeling experiments. These values also serve as a background to compare other labeling experiments.

Following solvent susceptibility testing, the next experiment was aimed at determining the presence of detrimental intermolecular interactions, as was done with the pIII protein. As a test molecule, BSF was used. **Figure 3-10** shows the results of phage

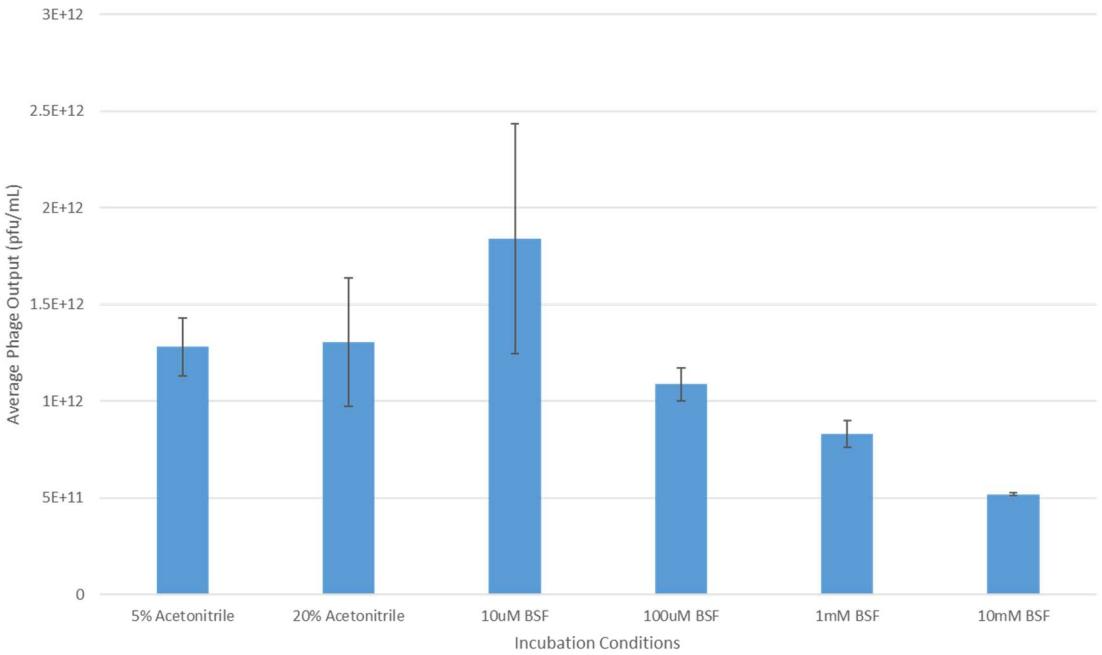


Figure 3-10: Concentration dependence of phage toxicity associated with intermolecular reactivity between M13KE phage and aryl sulfonyl fluoride. 10mM sample was used to indicate a fully saturated solution of BSF, as the saturation point was expected to be below this limit. All phage labeling reactions occurred for 3 hours at 30°C. The most significant decrease in phage survival was seen with the saturated solution of phage, however this drop was still less than one order of magnitude compared to solvent only incubations. While there is some toxicity associated with BSF which most likely comes from intermolecular reactivity, the decrease in phage population was not considered significant enough to prevent further testing.

incubation with increasing concentrations of BSF. Based on the data obtained, a drop in phage viability is not seen until 1mM of BSF and a more significant drop is seen at 10 mM, representing a fully saturated solution. In phage labeling experiments, the concentration of small molecule warhead added rarely exceeds 0.5 mM, and this it was concluded that intermolecular interactions did not have a significant impact on phage viability.

With these results in hand, the labeling of phage with SF1 could be tested. In addition, labeling with FS1 was also investigated. The intermolecular interactions of a free FS were not tested, as FS is less reactive than SF, and since free SF was determined to be nonreactive intermolecularly with phage, it was concluded that FS would have to yield the same results. To best represent successful phage labeling performed previously

in the Gao lab, it was decided that labeling reactions would be run using 300 µM of the respective small molecule in 20% CH_3CN for 3 hours at 30 °C. The resulting labeled phage was tittered to determine the concentration of surviving phage and these results are shown in **Figure 3-11.** Here, the unreduced phage (n=6) serves as control as these phage were exposed to CH_3CN only and saw no significant difference in phage before or after addition of organic solvent. As a reference, the 100 µM BSF (n=3) is also shown. Both SF1 incubation (n=9) and FS1 incubation (n=3) yielded a significant drop in phage, approximately 2 orders of magnitude lower than the starting concentration, represented by the unmodified phage. Based on this data, in conjunction with the lack of effects from

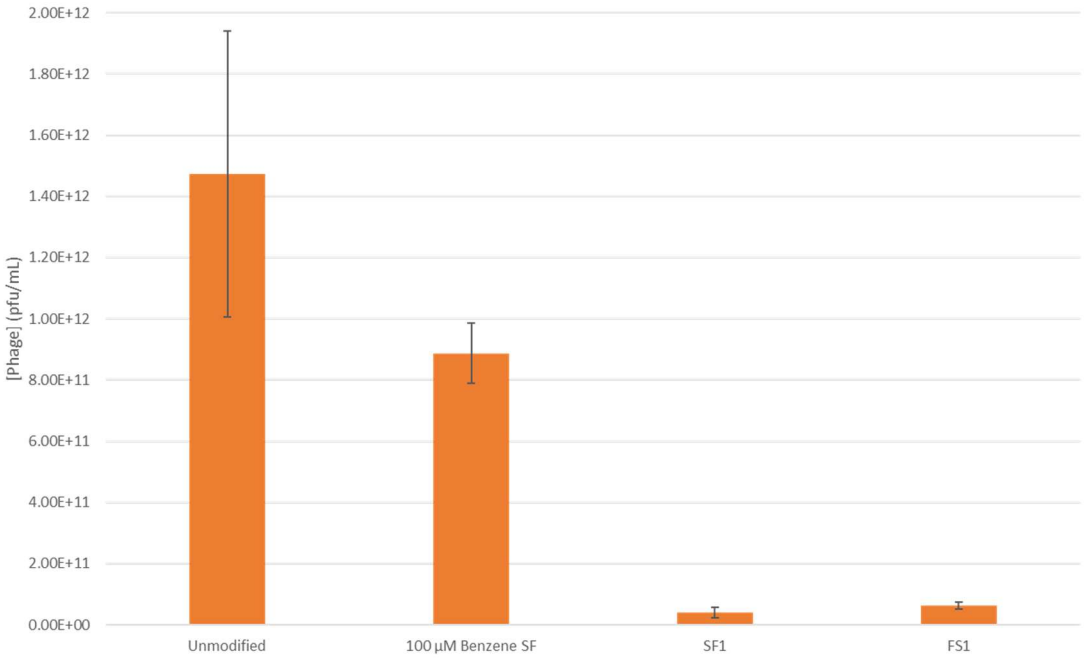


Figure 3-11: Identifying M13KE phage toxicity associated with SuFEX warhead labeling. Phage were reduced with 1 mM TCEP, which was the removed via phage precipitation prior to incubation with indicated molecule. All phage were incubated for 3 hours at 30°C. Unmodified phage were incubated in 20% CH_3CN only. SF1 and FS1 labeling both used 300 µM of the respective warhead. Both SuFEX-based warhead molecules saw a decrease in phage population of about 2 orders of magnitude. While significant, this still kept phage population high enough to maintain full diversity of a C7C library, and thus was not detrimental enough to deter further testing.

BSF incubation, it can be concluded that there is a form of labeling-induced toxicity experienced by phage.

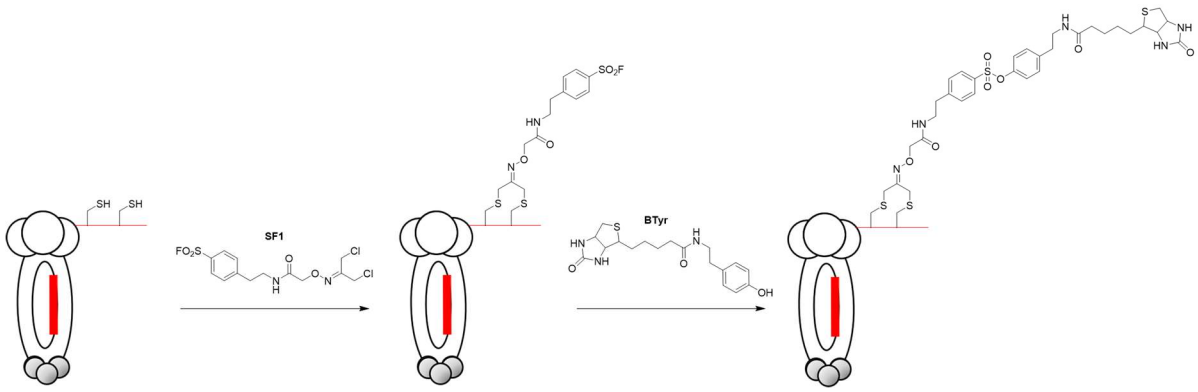
The topic will be expanded upon in the next chapter; however, it is important to identify what phage toxicity is. In this document phage toxicity simply refers to a drop in phage population between the input and output of a reaction, such as that seen with SF1. An important distinction to make is that phage toxicity may not directly correlate to phage death or lysis. An observable drop in phage concentration seen in tittering could represent lysed or "dead" phage but can also represent phage which have lost infectivity. As noted earlier, the phage protein labeled in these experiments is the pIII protein, which is also responsible for host recognition and attachment in the M13KE phage. Therefore, a significant disruption, such as a change in conformation or size, could inhibit the infectivity of the phage particle.

In the case of SF1 and FS1 it is difficult to determine what the source of toxicity is on the M13KE phage. Two points can be concluded though that are sufficient for progression with SF1. First, the toxicity seen by both SF1 and FS1 provides an interesting frame of reference. It is well documented the fluorosulfate has much lower reactivity than sulfonyl fluoride, so for the decreased phage count to be comparable between the two allows for the hypothesis that toxicity from these molecules is not due to intramolecular reactivity. However, this cannot be concluded immediately as the environment in which each warhead exists in post-labeling is drastically different from that which it was in before. The local concentration of amino acids is significantly increased when the warhead is no longer free floating in solution. While the fluorosulfate warhead may be less reactive, if it is held in place next to a potential reactive partner, such as a tyrosine

side chain, the proximity may induce reactivity on par with that of sulfonyl fluoride. The second significant factor is that the phage output population is still acceptable for library diversity. Using the C5C library as an example, each phage displays 5 randomized amino acids and with 20 possible amino acids in each position, the maximum number of phage needed for complete diversity is 20^5 or $3.20x10^6$ phage particles. Therefore, if the surviving phage are labeled, it would allow for libraries with up to 7 randomized amino acids to be created and still maintain full diversity after SF1 labeling. With these two hypotheses in mind, the goal became to identify if the remaining phage still had an intact sulfonyl fluoride following labeling.

3.5 ELISA based determination of warhead presence on phage

To determine if any C5C phage were being successfully labeled and contained an intact sulfonyl fluoride warhead a dual labeling experiment was proposed. The biotin-streptavidin interaction is a well-studied and potent biorthogonal reaction, and as shown previously it was already determined that sulfonyl fluoride can react with BTyr and leave the biotin intact. As such, an enzyme-linked immunosorbent assay (ELISA) utilizing this reactivity was developed. The logic of this experiment is that if a high concentration of BTyr is pulsed into a phage sample immediately following modification with SF1 that the phenol of the BTyr will react with both free SF1 still in solution as well as any intact SF1 warhead on a phage particle, thereby covalently creating a biotinylated phage, as shown in **Scheme 3-2**. Following a tittering for quantification, potentially biotinylated phage can be added to wells coated in Streptavidin. Following a stringent washing regimen, anti-M13KE coat protein- HRP conjugate antibody can be introduced followed by the introduction of HRP substrate. Phage should only bind to a well if biotin is present, the



Scheme 3-2: Double-labeling scheme to biotinylate phage via reactive SF warhead. Phage was reduced with 1mM TCEP before addition of 300 µM SF1. Phage were precipitated and an excess of BTyr was used to label to SF1 labeled phage in 5% DMF. Final phage should be a stable, biotinylated construct.

anti-M13KE will only remain in the well if there is phage present, and the introduction of the HRP substrate will only yield a visible color change if the antibody conjugate is present.

For ELISA experiments, a double labeling experiment was conducted, where phage were labeled with SF1 using previously discussed methods, followed by a precipitation and then modification with BTyr. BTyr labeling experiments were conducted using the conditions shown in **Table 3-1.** Before addressing ELISA results, one factor worth noting is the titer results from BTyr incubation. **Figure 3-12** shows the initial results for phage tittering following dual modification with SF1 and BTyr. What is

[BTyr] (µM)	Incubation Time (hr.)	Temperature	Avg. Output (pfu/mL)
600	6	37°C	3.05E+10
600	6	RT	2.77E+10
600	18	37°C	2.24E+10
600	18	RT	2.21E+10
0	6	37°C	0.00E+00
0	6	RT	2.05E+09
0	18	37°C	0.00E+00
0	18	RT	1.05E+09

Table 3-1: Incubations conditions for BTyr pulse labeling experiment. All reactions shown were run in 5% DMF. All samples were previously incubated with 300µM SF1 for 3 hours at 30°C. Phage were precipitated and refrigerated after the designated incubation time to prevent further loss of phage from SF1 reactivity.

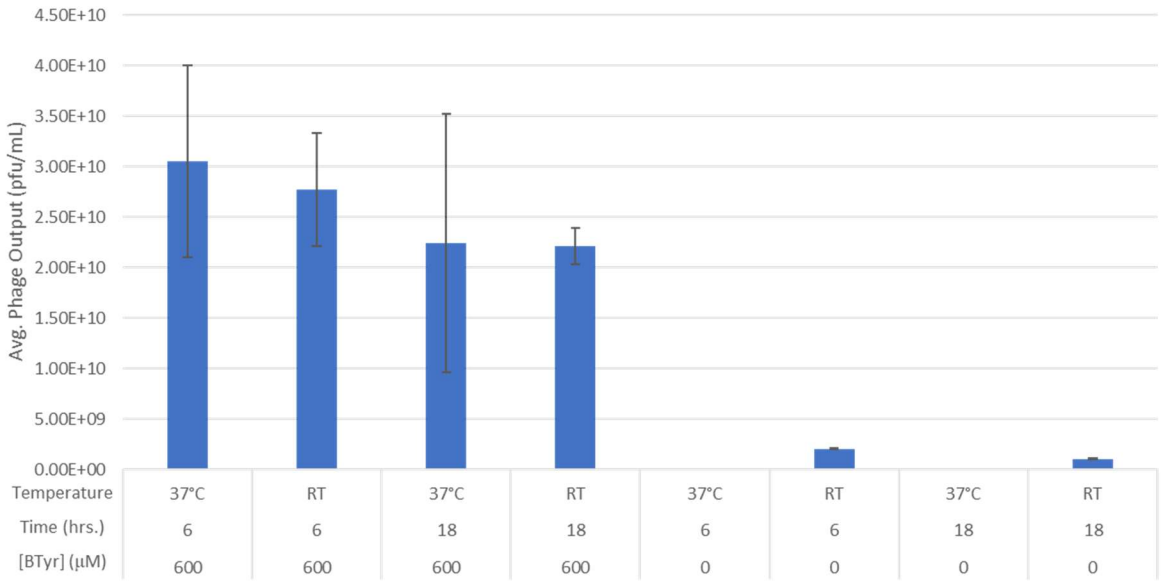


Figure 3-12: Dual labeling experiment using SF1 followed by BTyr labeling. Shown are the average M13KE phage outputs of the C5C library tested after incubating in the conditions listed in Table 3-1. All reactions shown were run in 5% DMF. All samples were previously incubated with 300 µM SF1 for 3 hours at 30°C. Phage were precipitated and refrigerated after the designated incubation time to prevent further loss of phage from SF1 reactivity before tittering. Samples incubated with SF1 and BTyr exhibited a phage output comparable to labeling with SF1, indicating that BTyr did not had have a significant negative impact on the phage. However, phage incubated for this extended period of time without BTyr were exposed to SF1 for a longer period of time, and therefore a more significant detrimental effect is seen on the resulting phage population.

important to distinguish in this table is the contrast between phage incubations which did and did not contain BTyr. All samples incubated in the presence of BTyr yielded the expected phage output based on previous experiments (~2 order of magnitude decrease). However, phage that were not exposed to BTyr experienced a more significant decrease in phage population, with those being incubated at 37 °C showing no phage on any dilutions assessed. Since these samples were kept under identical conditions and run in parallel, the only significant difference is the presence of BTyr, which would indicate that over time BTyr is rescuing phage from loss. The most likely hypothesis for this is that exposed and intact sulfonyl fluoride on the pIII protein is able to engage in intramolecular interactions with other phage protein residues leading to innately detrimental effects for

the phage, but the introduction of a high concentration of BTyr is able to block this effect and prevent further loss of phage over time. This would also explain why phage incubated at a higher temperature had a more significant loss in phage, as the increased temperature would increase the thermodynamic favorability of the reaction. While this hypothesis appears to be supported by the data presented, more testing will be required to confirm these results.

Figure 3-13 depicts the ELISA results for the dual labeled phage, as well as C5C phage labeled with biotin iodoacetamide as a positive control and C5C phage labeled with DCA as the negative control. Very large fluctuations in data for each dual modification (n=3 for each) yielded very large error bars, making data analysis difficult. More concerning, though, is the absorbance measured in wells that did not contain streptavidin. These wells were coated only in bovine serum albumin (BSA) from the blocking solution. The results from the negative controls wells without streptavidin are indistinguishable from those which were coated in streptavidin. Due to this unforeseen complication, it becomes impossible to determine if the labeled phage is biotinylated, as biotinylated pIII appears to have little effect on the binding ability of the phage to each well. The phage modified with biotin iodoacetamide were aliquoted from the sample library, so it is evident that SF1 incubation is responsible the apparent nonselective

binding, or “stickiness”, of these phage particles, and so the next challenge became identifying the source of this stickiness.

Condition 1: Labeled library incubated with Biotin Tyramide for 6 hours at 37°C.

Condition 2: Labeled library incubated with Biotin Tyramide for 18 hours at 37°C.

Condition 3: Labeled library incubated with Biotin Tyramide for 6 hours at RT.

Condition 4: Labeled library incubated with Biotin Tyramide for 18 hours at RT.

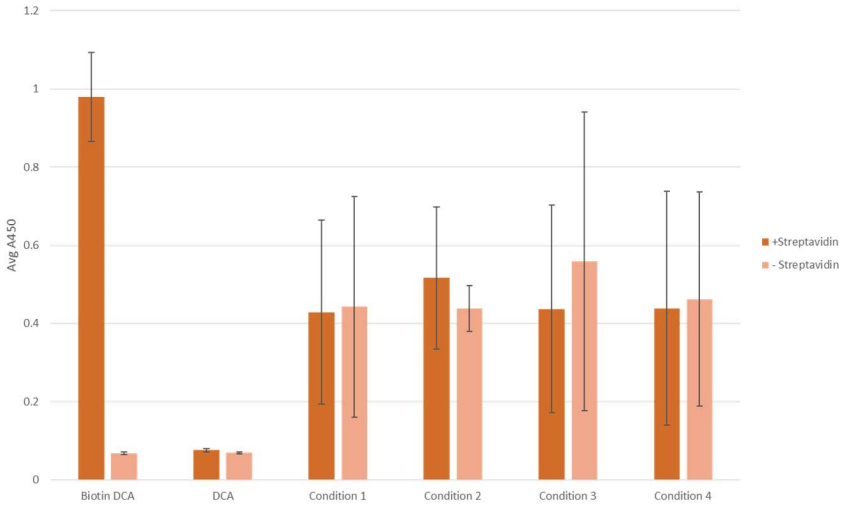


Figure 3-13: ELISA affinity assay used to determine biotinylation of double labeled phage. Phage tested in this assay were from the labeling experiment detailed in Figure 3-12. Phage labeled with 2mM Biotin-DCA served as a positive control to ensure biotinylation. Phage labeled with 100µM DCA served as a negative control, as this would prevent binding to any wells. Every sample had a corresponding negative control well that was coated in BSA instead of Streptavidin. Ideally, biotinylated phage should only remain in wells containing Streptavidin. The presence of phage in BSA coated wells prevents an accurate conclusion to be drawn about the biotinylation status of the phage but does indicate that a source of non-specific binding is present in all double labeling experiments tested.

To ensure that stickiness was derived from SF1 and not from BTyr, new aliquots of C5C library were incubated with SF1 using an identical protocol but were subjected to ELISA before incubation with BTyr. As was expected, the phage incubated with SF1 still experienced the same stickiness as seen previously, as shown in **Figure 3-14a**. Therefore, the stickiness to the BSA coated wells had to be attributed to the SF warhead. Based on the properties noted earlier in this document, it was determined that the most likely cause of this nonselective binding was nonspecific binding activity of the sulfonyl fluoride. Therefore, it was hypothesized that using a lower activity warhead, such as fluorosulfate, would reduce or potentially even negate the nonspecific binding. However, as seen in **Figure 3-14b,** following a labeling with FS1, the fluorosulfate-containing phage

exhibited the same non-specific binding as SF1 labeled phage. Based on these results, as well as those obtained previously from pIII labeling experiments, it was determined to be highly unlikely that pIII labeling was the root cause of this nonspecific binding. In order to prove this hypothesis, a similar phage labeling experiment was carried out with SF1, however no TCEP was added so that the exposed disulfide on the pIII protein would remain intact. In this experiment, the phage should remain unlabeled and behave identically to the wild type, unlabeled phage. As shown in **Figure 3-14c,** however, this was again not the case. The lack of selectivity between streptavidin and BSA was clear evidence that a second, unaccounted for reaction was occurring between the phage particles and these small molecules, though the mechanisms of this reaction remain a mystery.

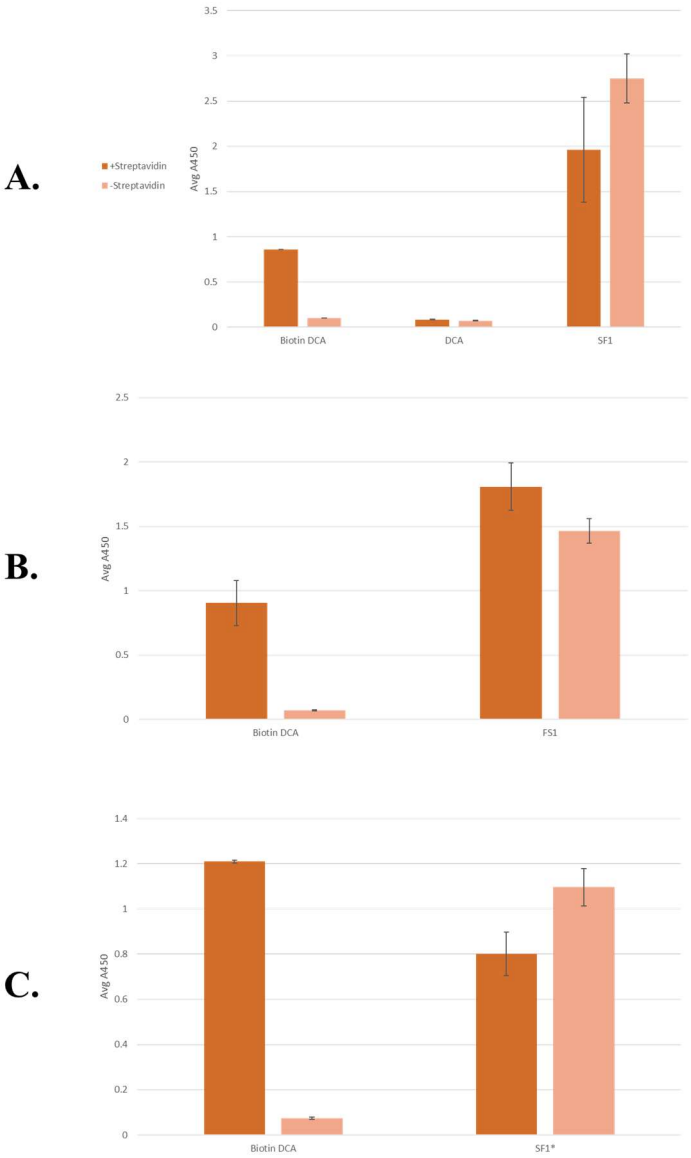


Figure 3-14: Identifying sources of nonspecific binding of SF1 labeled phage using ELISA. Phage library labeled with 2 mM Biotin-DCA or 100 µM DCA served as controls. Every sample had a corresponding negative control coated in BSA. Phage library incubated for 3 hours at 30°C with 300 µM SF-DCA was precipitated, tittered, and then subjected to ELISA analysis with no BTyr labeling step (A). In this instance, nonselective binding was seen comparable to the positive control, indicating that BTyr was not the source. FS1 labeling of the phage library was carried out under identical conditions as the SF1 (B). The FS1 labeled phage experienced the same nonselective binding issue, meaning that the nonselective binding may come from a source beyond warhead reactivity. Lastly, unreduced phage were incubated with SF1 using an identical protocol (C). Under these conditions, SF1 cannot label phage since the disulfide is still intact. Despite this, this phage sample still experienced nonselective binding comparable to the positive control.

3.6 Conclusions

The nonselective binding seen in nonreduced phage indicated two possibilities. Either the labeling process was damaging to the phage in some way that was causing it to non-selectively bind to BSA and streptavidin, or there is a second labeling site somewhere on the phage that has not been accounted for. While the exact mechanics of this nonspecific binding are not known at the time, both potential scenarios are challenging for the future of SF1 as a warhead. The goal of this set of experiments was to serve as a case study for studying covalent warhead suitability for implementation into phage display. While aryl sulfonyl fluoride is a good candidate on paper, the experiments detailed here would indicate that the phage labeling process and covalent phage library creation is far more nuanced than previously thought. Shown here is the fact that a strong candidate does not always make a good warhead, and that any covalent warhead is going to require extensive and detailed analysis before being properly implemented into phage libraries. As for SF1 and aryl sulfonyl fluoride-based warheads, it is currently concluded that this class of covalent warhead is not a suitable candidate for covalent library creation. While some results are hopeful and other experiments can be run, several factors point towards SF being unsuitable.

First is the concern of the warhead's stability. Thus far, it has never been concretely proven that the sulfonyl fluoride warhead remains intact on the pIII protein for any significant portion of time. LC/MS analysis indicated that in the presence of some proteins, the aryl sulfonyl fluoride warhead can be reduced down to nonreactive sulfinic acid. Similarly, NMR analysis indicated that the warhead could also hydrolyze over time

to the sulfonic acid, with this reaction being expediated at higher temperatures. With two potential routes of degradation even having five copies of the pIII protein to label may not be sufficient without definitive evidence confirming that the warhead remains intact long enough for target engagement to occur.

Second is the clear propensity for off-target, non-specific labeling. Many of the results seen in this set of experiments could be attributed to improper labeling of the phage by SF1. No adverse or unexpected effects are seen with DCA labeling or following incubation with free sulfonyl fluoride and therefore it seems that the unexpected reactivity requires a bifunctional molecule such as SF1 to exhibit such effects. This is an idea that will be further explored in the next chapter. While the exact mechanism is not known, it is undeniably problematic, and this factor alone is significant enough to determine that aryl-SF may not be a suitable warhead for phage display.

The combination of warhead stability and reactivity issues has led to the conclusion that SF1 is not a suitable warhead for phage display. While this molecule is problematic, there are still plenty of opportunities in phage display with sulfonyl fluorides as a class of warhead. As Grimster and co. showed, this modality is highly tunable, and so more rigorous testing and continued experimentation could yield a successful and stable covalent phage library containing a sulfonyl fluoride motif.

3.7 Experimental Methods

General Methods

All listed chemicals not synthesized in house by members of the Gao lab were purchased from Fisher Scientific or Sigma-Aldrich unless otherwise specified. All cells were grown in a low salt LB media, made as specified in the NEB Ph.D.™ Phage Display

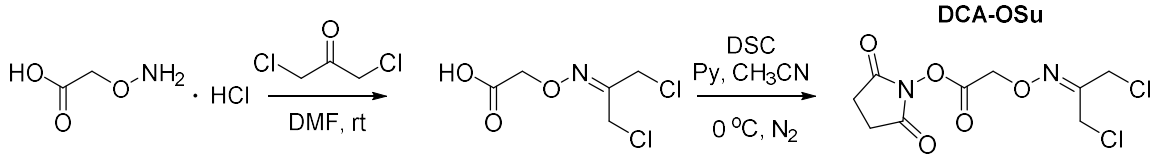
Libraries Instruction Manual. All reactions performed in PBS buffer used autoclaved buffer prepared from protocols available from Cold Spring Harbor protocols and were adjusted to desired pH using 1M HCl or 1M NaOH. All proteins containing a 6xHis tag were purified using manually packed columns containing 0.5-1mL HisPur Ni-NTA resin using protocols available from Thermo Scientific. Following purification, all protein samples were desalted using PD-10 desalting columns for volumes >0.5mL. All samples of 0.5mL or less were purified using NAP5 desalting columns. Protocols were followed based on those provided with columns. Elutions of final protein mixtures were performed with 5x PBS buffer pH 8.0 unless otherwise noted. Protein and peptide concentration was determined using a Thermo Scientific NanoDrop 2000 spectrophotometer. All solvent was evaporated using a Heidolph Laborota 4001 efficient rotary evaporator. LC/MS methods and instruments used can be found in Chapter 2.

Samples of SF1 and FS1 were provided as solid samples, the synthetic route for each is summarized below. SF1 was resuspended in CH_3CN to create a working solution of concentration ~60 mM. FS1 was resuspended in $CHCl_3$ to create a working solution of concentration ~50 mM. Aliquots of working solution were further diluted with respective solvent to create sub-stocks of desirable concentrations.

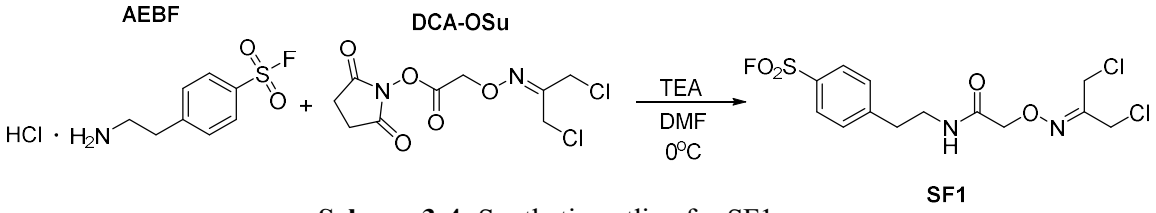
Peptide and Protein Modification with aryl-SF warhead

pIII protein fragment purification and usage is detailed in Chapter 2. All protein labeling experiments were performed in 20% CH_3CN in PBS buffer pH 8.0. Reactions were run at 100 µL scale. Both SF1 and FS1 were synthesized in house by Dr. Fa-Jie Chen. Both molecules were created from a succinate-DCA compound, referred to as DCA-OSu, the synthetic scheme of which is outlined in **Scheme 3-3** below. SF1 was

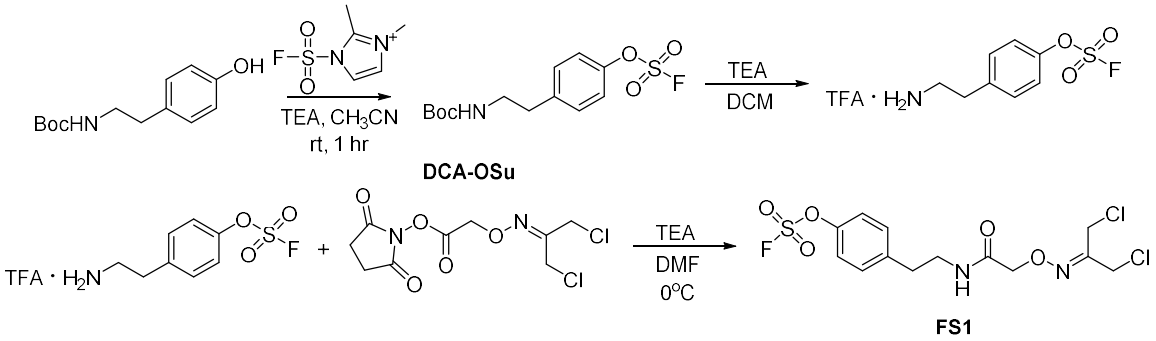
carried out in a one-step reaction using DCA-OSu and the commercially available sulfonyl fluoride amine 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBF). The reaction conditions are outlined below in **Scheme 3-4**. FS1 required a more elaborate synthetic route, with the outline of the two-step synthetic scheme shown below in **Scheme 3-5**.



Scheme 3-3: Synthetic outline for DCA-OSu warhead precursor molecule.



Scheme 3-4: Synthetic outline for SF1.



Scheme 3-5: Synthetic outline for FS1. Performed as a 3-step reaction.

Stock solution of SF1 was diluted down to 1.5 mM in CH_3CN so that 20 µL of this solution could be added to a desalted stock of pIII protein for final SF1 concentration of 1 mM. To account for the poor solubility of FS1, the stock solution in $CHCl_3$ was aliquoted into 1.5mL Eppendorf centrifuge tubes and solvent was evaporated to result in small amounts of solid FS1 as desired. PBS/CH_3CN mixture containing the pIII protein desalt was then added directly to tubes containing the solid FS1. Final concentration of

FS1 was intended to be 1 mM but due to lower solubility of FS1 the actual concentration of FS1 in solution was found to be ~520 μM based on the solubility in 20% CH_3CN. The solubility point of both molecules was determined using NMR following the protocols detailed in the next subsection. Protein was reduced using methods detailed in Chapter 2. All labeling was conducted at 30°C for 3 hours before being injected onto LC/MS for analysis. LC/MS analysis of proteins followed that detailed in Chapter 2.

Reduced peptide samples were provided as solids and resuspended in Nanopure water. Since the peptide used contained a phenylalanine residue, concentration could be determined using UV/Vis spectroscopy. Solid peptide was added to 100 μL of water until the concentration of the peptide was at least 200 μM. CH_3CN was then added containing SF1 such that the final concentration of SF1 in solution would be 1 mM. Labeling reaction was conducted at 37 °C for 3 hours before injecting sample onto the LC/MS for analysis. Peptide samples were run through an Agilent Poroshell 120 EC-C18 2.7 mm 3.0x50 mm column for separation. The table below shows the solvent gradient used for all protein samples studied via LC/MS. Samples were ionized in positive ion mode using electrospray ionization (ESI), a 2.5 μL injection volume with 200 μL/min draw and eject speed, solvent flow rate of 0.200 mL/min, extended dynamic mode in mass range was 200 – 1000 m/z, temperature of drying gas = 325 °C, flow rate of drying gas = 8 L/min, pressure of nebulizer gas = 35 psi, capillary = 4.286 μA, fragmentor = 175 V, and octupole rf voltages = 750. Data was collected at 1 spectra/s.

Time (min)	% Solvent A	% Solvent B	Flow Rate (mL/min)
0.00	95.0	5.0	0.200
5.00	95.0	5.0	0.200
20.00	5.0	95.0	0.200
25.00	5.0	95.0	0.200
26.00	95.0	5.0	0.200
33.00	95.0	5.0	0.200

Table 3-2: LC/MS Conditions used for analysis of peptide samples

For analysis of protein samples, the A peak of each spectrum was deconvoluted and analyzed. LC/MS conditions for protein samples can be found in Chapter 2. Deconvolution protocol is detailed in Chapter 2 as well. Deconvolution was not necessary for peptide samples.

Aryl-SF Reactivity Profiling via NMR

All ^{1}H NMR spectra were obtained using a VNMRS 500 or 600 MHz NMR spectrometer. NMR data was processed using MestReNova software. All deuterated solvents were purchased from Cambridge Isotope Labs. Deuterated PBS pH 8.0 (dPBS) was created using publicly available protocols from Cold Spring Harbor replacing H_2O with D_2O. Appropriate pH was achieved via addition of 1mM HCl or 1 mM NaOH as necessary. All NMR samples contained 600 μM Tris HCl as an internal standard. Concentrations for molecules of interest were determined by comparing integrated peak area of selected molecule peaks with the 6 proton -CH_2 peak present in the Tris HCl signal. Due to the presence of large solvent peaks, only peaks with chemical shifts in the aromatic region (6.8-8.3 ppm) were integrated with the exception of the Tris -CH_2 peak.

Pure benzene sulfonyl fluoride was weighed and CD_3CN was added to resuspend the benzene sulfonyl fluoride and make a working stock. This working stock was added

to final reaction mixture to a final concentration of ~600 μM. Biotin Tyramide was added to reaction vial until saturation occurred in 20% CD_3CN, which resulted in a concentration of ~1 mM. Reactions between benzene sulfonyl fluoride and Biotin Tyramide were carried out overnight at 37°C. Timepoints were taken every hour for 7 hours and a final timepoint was taken after 22 hours. For all reactions 600 μL volume reactions were run in capped NMR vial. For saturation verification using NMR varying volumes of the SF1 stock were diluted up to 120 μL with CD_3CN. The spectra shown in this document represents the maximum concentrations seen in the experiment sets tested. For FS1, stock solutions in $CHCl_3$ were aliquoted into 1.5mL Eppendorf tubes and solvent was evaporated off. The remaining solid was resuspended in 120 μL of CD_3CN before diluting up to 600 μL with dPBS.

Phage Labeling with Aryl-SF

A C5C phage library was created and provided by Dr. Mengmeng Zheng. Phage library creation and bacteriophage maintenance and tittering protocols can be found in the Ph.D. Phage Display Library Instruction Manual, Version 3.0 provided by NEB. Recipes for all solvents and buffers used in phage tittering can also be found in this manual. All phage labeling reactions occurred in 100 μL aliquots of approximately 10^{12} pfu/mL phage concentration. 10 μL of concentrated phage stock (10^{13} pfu/mL) was diluted up to 100 μL with 2 0mM ammonium bicarbonate pH 8.0 to create each aliquot. Phage that was reduced prior to reaction were exposed to 1 mM TCEP for 1 hour at room temperature. 1 μL of a 100 mM stock prepared from adding solid TCEP to Nanopure water was used for these reductions. Phage aliquots that were not reduced were exposed to the same conditions but 1 μL of 20 mM ammonium bicarbonate buffer was added instead of TCEP

stock. All precipitations of phage were carried out on ice for at least 1 hour before spinning down at 14,000 rpm to pellet bacteriophage. Supernatant was removed following centrifugation via pipetting to avoid disturbing pellet. All labeling reactions with SF1 and FS1 occurred in 20% CH_3CN, 80% 20 mM ammonium bicarbonate pH 8.0 for 3 hours at 30°C in a warm water bath. Stock preparation for both molecules was as described previously. Labeled phage were precipitated using the previously described method and resuspended in 100 µL PBS pH 8.0. Modified phage stocks were immediately tittered on IPTG/XGal plates following modification. ER2738 strain *E.coli* maintenance and usage is described in Chapter 2. Tittering outputs were determined by counting individual phage plaques by hand.

ELISA based determination of warhead presence on phage

For ELISA assay studies, phage first underwent a double labeling procedure. For all phage used in these assays, SF1 and FS1 labeling was carried out as described previously. Labeling with Biotin Iodoacetamide (IA) for positive controls was performed using the same protocol except 1mM Biotin IA was used and no organic solvent was used. DCA labeling followed this same method with 1 mM DCA instead of Biotin IA. Doubly labeled phage were precipitated after modification with SF1 and resuspended in 95 µL 20 mM ammonium bicarbonate pH 8.0. Labeling with BTyr occurred in 5% DMF using 500 µM BTyr in DMF. This was done by diluting from a 10 mM stock of BTyr in DMF. These reactions were incubated at the time and temperatures listed previously. Phage reactions were precipitated upon completion of reaction and resuspended in 100 µL PBS pH 8.0.

Prior to running ELISA assays, a 1 mg/mL stock of Streptavidin was prepared from solid streptavidin from Invitrogen in PBS pH 7.5 and then diluted to 0.1 mg/mL in 0.1M $NaHCO_3$ pH 8.0. For ELISA assays all controls and experimental groups were performed in parallel triplicate on 96 well plates from Fisher Scientific. 100 µL of 1mg/mL streptavidin was added per well for each sample being run (3 wells per sample). For each well containing streptavidin, an adjacent was filled with 100 µL of 0.1M $NaHCO_3$ pH 8.0 only as a negative control (-streptavidin trials). To coat wells in streptavidin, 96 well plates were sealed with parafilm and incubated overnight at 4°C. Supernatant was removed from 96 well plate by blotting on paper towels. All wells were then washed 1x with 100 µL PBS pH 8.0. Each well was blocked with 5mg/mL BSA, prepared from solid BSA in PBS pH 8.0, and shaken on a shaker plate at 30rpm for 1 hour. PBST buffer was prepared by adding 0.1% Tween to PBS pH 8.0 buffer. Following blocking, wells were washed 6x with PBST pH 8.0. Based on tittering results, phage stock was diluted to ~10^9 pfu/mL using PBS pH 8.0 containing 5 mg/mL BSA. 100 µL of phage dilution was added to each well and the plate was shaken for 1 hour at room temperature. All wells were then washed 6x with PBST pH 8.0. A commercial stock of anti-M13KE major coat protein antibody-HRP conjugate was diluted 10,000x in PBST pH 8.0 and mixed thoroughly before adding 100 µL to each well and shaking at room temperature for 1 hour. Wells were washed again 6x with PBST pH 8.0. 100 µL of commercially available tetramethylbenzidine (TMB) solution was added to all wells. The plate was then covered with aluminum foil and gently shaken for 30 minutes before quenching the reaction with 100 µL of 2 mM H_2SO_4. The presence of phage turned

solutions in well a visible yellow color. The absorbance of each well was then measured at 450 nm using a SpectraMax M5, SoftMax Pro 5 Plate Reader.

3.8 References

(1) Chen, S.; Lovell, S.; Lee, S.; Fellner, M.; Mace, P. D.; Bogyo, M. Identification of Highly Selective Covalent Inhibitors by Phage Display. *Nat Biotechnol* **2020**.

(2) Dong, J.; Krasnova, L.; Finn, M. G.; Sharpless, K. B. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angew. Chem. Int. Ed.* **2014**, *53* (36), 9430–9448.

(3) Mukherjee, H.; Debreczeni, J.; Breed, J.; Tentarelli, S.; Aquila, B.; Dowling, J. E.; Whitty, A.; Grimster, N. P. A Study of the Reactivity of S $^{(VI)}$ –F Containing Warheads with Nucleophilic Amino-Acid Side Chains under Physiological Conditions. *Org. Biomol. Chem.* **2017**, *15* (45), 9685–9695.

(4) Liu, J.; Cao, L.; Klauser, P. C.; Cheng, R.; Berdan, V. Y.; Sun, W.; Wang, N.; Ghelichkhani, F.; Yu, B.; Rozovsky, S.; Wang, L. A Genetically Encoded Fluorosulfonyloxybenzoyl- L -Lysine for Expansive Covalent Bonding of Proteins via SuFEx Chemistry. *J. Am. Chem. Soc.* **2021**, *143* (27), 10341–10351.

(5) Tabuchi, Y.; Watanabe, T.; Katsuki, R.; Ito, Y.; Taki, M. Direct Screening of a Target-Specific Covalent Binder: Stringent Regulation of Warhead Reactivity in a Matchmaking Environment. *Chem. Commun.* **2021**, *57* (44), 5378–5381.

CHAPTER 4: A MECHANISTIC STUDY OF PHAGE TOXICITY ASSOCIATED WITH LABELING USING DIAZABORINE-FORMING WARHEADS

4.1 Diazaborine Forming Warhead RMR1

In the previous chapter the suitability of aryl sulfonyl fluoride as a covalent warhead for phage display platforms was studied and discussed. While it was ultimately determined that the tested molecule, SF1, would most likely not be suitable for creating covalent phage libraries, the experiments presented created many questions about the relationship between covalent warheads and bacteriophage. One of these questions that arose during this study relates to the idea of small molecule mediated phage toxicity. As noted earlier, discussing toxicity in relation to bacteriophage is complicated in that it is debatable of bacteriophage are even alive to begin with. However, as seen with SF1, there is clear and present evidence that incubation with some small molecules has a direct and adverse effect on phage growth and viability. This is primarily noted in instances where there is a significant decrease in phage concentration following incubation with a molecule of interest. The mechanisms for this loss of phage viability still remained a mystery during SF1 testing. It remained unknown if this toxicity would be uniform across all covalent warheads or if this was unique to SF1. As such, a new set of experiments were designed to further our understanding of warhead mediated phage toxicity.

To better study phage toxicity, a new candidate molecule was to be chosen. The diazaborine forming warhead hereby referred to as RMR1 was selected as the candidate. Previous work in the Gao lab provided a well-studied mechanism and a strong understanding of the reactivity of this warhead[1]. The design of this lysine targeting warhead is derived strongly from earlier work from the Gao lab on the APBA warhead[2], however a more stable six-membered ring formation allowed for a much slower off rate compared to APBA. **Figure 4-1** displays this idea and shows the basic structure of the

RMR1 warhead. While this warhead is not an irreversible covalent warhead, as SF1 and FS1 were intended to be, it was believed that the slow off rate, with a $t_{1/2}$ on the hour scale, could appropriately mimic the effect of an irreversible warhead on a phage.

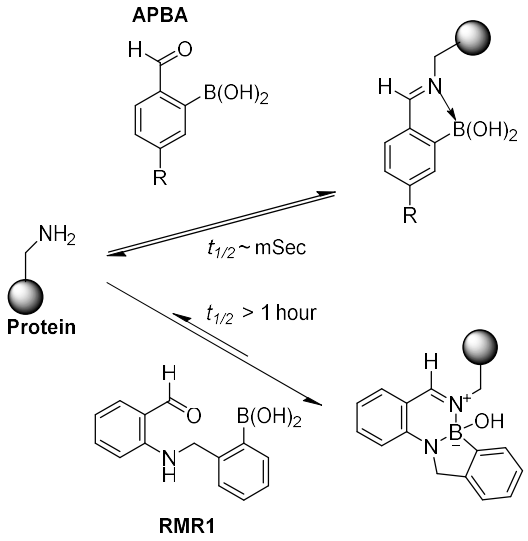
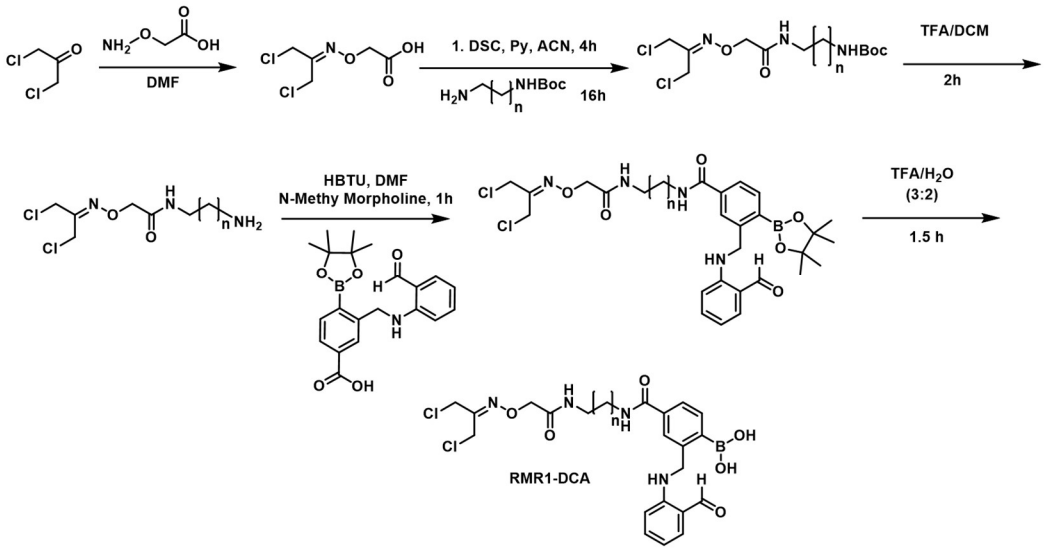


Figure 4-1: Comparison of APBA and RMR1 warheads. Both warheads are characterized by their formation of a reversible covalent bond with amines. RMR1 is unique in that the enhanced stability afforded by the formation of a six-membered diazaborine ring yields a much slower dissociation rate compared to APBA, making the warhead more similar in nature to an irreversible covalent warhead.

In order to construct bicyclic phage as SF1 did, a bifunctional molecule was needed. As such, RMR1-DCA was synthesized in house and would be used for phage labeling experiments in the same manner that SF1 was. **Scheme 4-1** shows the simplified reaction scheme carried out to synthesize RMR1-DCA, courtesy of Dr. Rahi Masoom Reja. As with other warheads seen previously, the DCA motif would serve to quickly conjugate to the exposed thiols on the pIII protein, while the RMR1 warhead would ideally be left available to engage with a target protein. The RMR1 was independently shown to be hydrolytically stable, and therefore was proposed to avoid the potential degradation issues faced by SF1. An additional note that must be addressed is the solubility of RMR1-DCA. It was previously found that RMR1-DCA was not soluble in CH_3CN, however it was found to be soluble in DMF. It was determined that phage experienced no significant

detrimental effects from being incubated with 5% DMF, and so this was deemed an acceptable parameter to use for labeling.



Scheme 4-1: Synthetic outline of the RMR1-DCA bifunctional molecule. Original design of this molecule as well as synthetic pathway was developed by Dr. Rahi Masoom Reja.

4.2 Preliminary Phage Toxicity Screening using C5C Phage Library

Before analyzing data relevant to phage labeling with RMR1-DCA an important preliminary note must be made. For many of the experiments detailed in the section, despite large sample sizes, average error was often exceedingly large due to high variations in data. As such, the discussion around this data will be primarily qualitative in nature. The data obtained can be used to observe trends and patterns within sample sets, however with consistently large error values quantification of the data becomes increasingly challenging and unreliable.

The first step in studying phage toxicity related to the RMR1 warhead was to determine an approximate degree of toxicity and validate that the presence of RMR1 was indeed detrimental to phage growth. The first set of experiments utilized the same C5C library discussed in the previous chapter due to its availability. To gauge approximate toxicity, M13KE phage were reduced and then incubated with varying concentrations of

RMR1-DCA or Biotin-DCA as a negative control. Biotin-DCA serves as a negative control in this experiment because, as seen previously with Biotin-IA, the Biotin motif is not known to have a detrimental effect on phage viability. Since the parameter being investigated was phage death, the negative control was therefore expected to yield no drop in phage. For this experiment, and all those going forward aimed at assessing phage toxicity, the most important effect to be studied was the immediate impact of small molecule incubation on the phage. As such, data is derived from a comparison of phage populations immediately before addition of small molecule (inputs) and immediately after incubation with small molecules (output).

Figure 4-2 shows the results of this initial experiment comparing Biotin-DCA (n=3), 300 µM RMR1-DCA (n=4), and 100 µM RMR1-DCA (n=3). To better visualize larger differences in the data, fold differences are shown as the log of the obtained value. Biotin-DCA yielded a drop in phage population of less than one order of magnitude, with the decrease most likely stemming from loss of phage in precipitation. In direct contrast to this minor decrease was

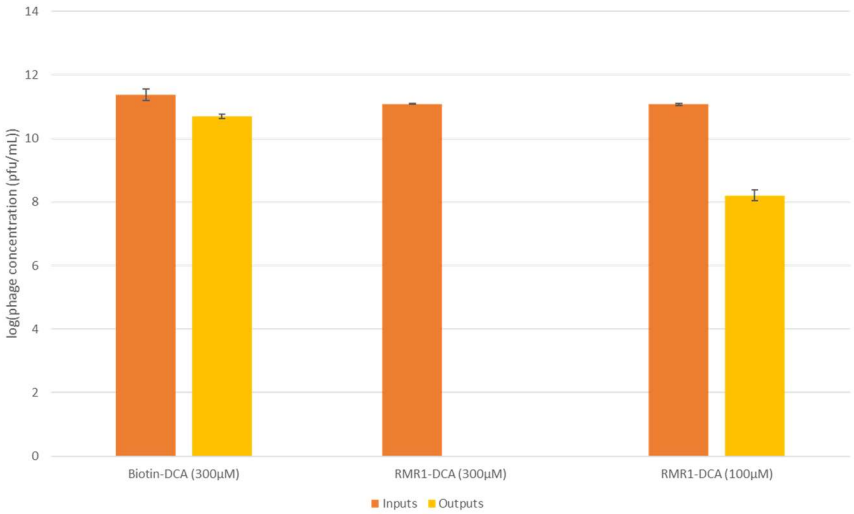


Figure 4-2: Preliminary C5C library toxicity screening with RMR1-DCA. Inputs refers to aliquot of the phage library collected immediately before incubation with small molecule whereas Outputs refers to phage stock following incubation and labeling with the designated molecule. Biotin-DCA was previously shown to non-toxic and is used as a positive control. To better depict larger differences in data, the log of average phage concentration at each stage is depicted.

the complete lack of phage at all tested dilutions of phage incubated with 300 μM RMR1-DCA. Based on the dilutions tested, this would mean that if any of the labeled phage remained viable that the minimum decrease in surviving phage would be approximately 5 orders of magnitude, or a 100,000-fold decrease from the input. Lastly, the phage incubated with 100 μM RMR1-DCA had a significant but observable decrease, corresponding to a 3 order of magnitude drop. Going forward, a 100 μM RMR1-DCA incubation would be used as a reference point for future experiments.

The next step in assessing RMR1-DCA toxicity was to look at the individual components of the small molecule itself, much the same as done previously with SF1. Experiments discussed in previous chapters sufficiently determined that DCA is not a significant detractor of phage viability. Therefore RMR1-DCA was directly compared to two separate molecules: RMR1 and RMR2-DCA. RMR1 served to identify any toxicity related to the intermolecular interactions RMR1 warhead on its own prior to phage labeling. RMR2-DCA is a structural analog to RMR1-DCA where the amine-reactive boronic acid motif was replaced with an inert hydroxyl group. This molecule would serve to identify any potential steric clashes that could hinder phage proliferation. As noted earlier, covalent labeling of the M13KE phage occurs on the pIII protein that the phage uses for host recognition, and so RMR2-DCA serves to verify that this covalent labeling action itself is not inherently harmful to phage populations. **Figure 4-3** shows the results of this initial experiment set between the three small molecules (n=3 for each). To compare the degree of toxicity going forward the ratio of the output to the input measured in folds of difference (hereby referred to as X-Folds) was used in lieu of plotting the inputs and outputs separately. For example, a decrease in phage population equal to one

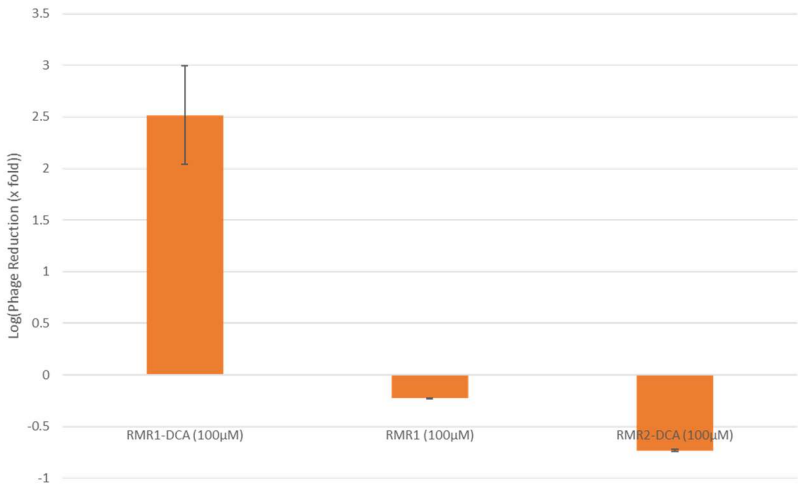


Figure 4-3: Verifying labeling-induced toxicity of RMR1-DCA using C5C phage library. Phage were labeled using identical methods as previous labeling experiment. Phage samples from the same stock were also incubated with RMR1 and RMR2-DCA. Data is shown as the log of average phage reduction in X-fold for each sample. The reduction of phage in X-folds was determined by taking the ratio of input to output phage populations. In this instance, a negative value indicates a higher output population than input population.

order of magnitude would equate to a 10-fold decrease in phage population. It should be noted that based on these parameters a negative fold difference is indicative of an increase in phage output populations compared to the input. Once again, incubation of the C5C phage library with 100 µM RMR1-DCA yielded a 3 order of magnitude decrease between the input and output populations. For both RMR1 and RMR2-DCA, however, a small (< 1 order of magnitude) increase was actually seen in the phage population. The small nature of this increase most likely does not represent an increase in phage viability caused by small molecule incubation, but instead is representative of statistical fluctuations inherent to phage tittering experiments, a concept that will be discussed in more detail later in this chapter. This experiment did serve to prove a distinct difference between the RMR1-DCA molecule and its individual components, confirming that toxicity is derived from the labeling event of RMR1-DCA.

4.3 RMR1 Phage Toxicity Screening using Discrete Phage Constructs

To allow for higher trial numbers as well as reduce variables in toxicity testing, a discrete phage construct was chosen for future experiments over the phage library. Previously, an M13KE phage was isolated from a C7C library displaying the sequence CRGDLASLCGGGS between the Cys residues on the pIII protein. This construct was selected as it was well studied, easily amplified, and had existing derivatives that were also created, as will be discussed later in this section.

Two experiments were conducted concurrently using this construct that aimed at deciphering the root causes of phage toxicity. However, as noted at the start of this chapter, data consistency proved to be problematic for the experiments discussed here. The lack of consistency between data points is in and of itself a factor worth noting in the discussion of phage toxicity, however it still leads to complications in data analysis. As such, graphical representations of data are also presented with corresponding data tables to serve as additional points of reference in discussion.

The first set of experiments was intended to validate the concentration dependence of phage toxicity. As had been seen with the C5C library, an increase from 100 μM to 300 μM RMR1-DCA yielded a significant impact of phage viability. To better assess the impact concentration had on phage survival, a smaller scale was used, increasing RMR1-DCA concentration in samples by 50 μM increments up to 200 μM. Some phage aliquots were also left unreduced, having only buffer added instead of 1 mM TCEP, to serve as a negative control. **Figure 4-4** shows the graphical representation of the trends observed in the data, which is shown in **Table 4-1**. As can be seen in **Figure 4-4**, there are several important conclusions that can be determined from this data set.

First is the significant variation between data points. Despite the large sample size that the shown data is pulled from, there are still drastic variations in the range of values observed for many of the conditions tested. For example, 100 μM RMR1-DCA incubation with reduced phage yielded outputs ranging from 1.5 orders of magnitude up to 4 orders of magnitude. The significant difference in values leads to large error bars that go beyond the shown data for some experiments. The large sample size used decreases the likelihood of this variance coming from experimental error, as does the low variance in control samples not exposed to RMR1-DCA. Therefore, the phenomenon of high variance seen here may be representative of the testing platform, which would mean that phage toxicity is not as consistent as previously thought. While the exact sources causing the data

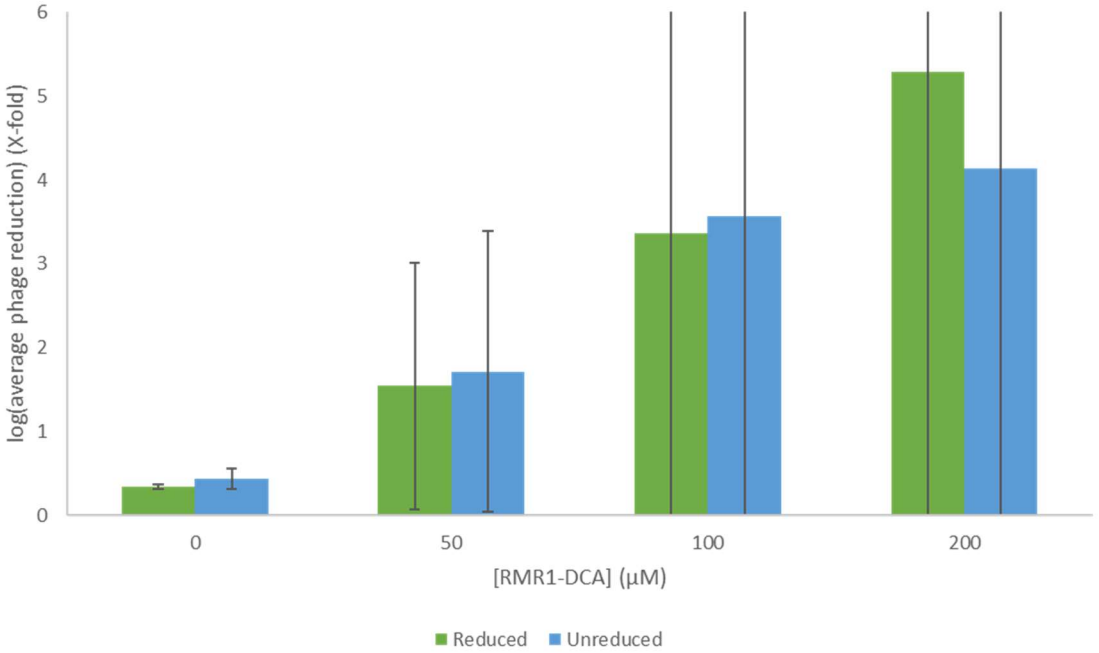


Figure 4-4: Concentration dependence of phage toxic effect of RMR1-DCA on discrete C7C phage constructs. Data is shown as the log of average phage reduction in X-fold for each sample. A definitive concentration-dependent effect can be seen on both reduced and unreduced phage, as the average reduction of phage increases with each concentration. However, as phage reduction increases so does the estimated error in each sample, despite large sample sizes (>20). A direct trend could not be drawn due to error, however what could be concluded was the comparable results between reduced and unreduced phage in each trial. Data in X-folds with minimum and maximum values is depicted in Table 4-1.

[TCEP] (mM)	[RMR1-DCA] (μM)	Average Phage Reduction (X-fold)	Standard Deviation	Minimum Value (X-fold)	Maximum Value (X-fold)	n
1	0	2.18	1.07	0.36	4.61	24
	50	35.12	29.60	9.98	94.49	7
	100	2315.89	3071.49	56.31	10117.65	32
	200	190825.87	254600.11	1957.14	612500.00	6
0	0	2.73	1.33	1.61	4.60	3
	50	51.91	47.44	8.16	131.47	5
	100	3663.22	7340.68	28.75	25350.00	19
	200	13624.95	22301.25	505.14	52250.00	4

Table 4-1: Phage reduction values for concentration dependence testing of RMR1-DCA toxicity testing. The reduction of phage in X-folds was determined by taking the ratio of input to output phage populations. The log of the average phage reductions is plotted in Figure 4-4.

variation seen in these toxicity studies remains unknown at the time, this data does show that there are more factors involved in the display sequence-independent toxicity of phage than small molecule concentration alone.

Second, while difficult to concretely determine due to the data variance mentioned above, the concentration dependence appears to be corroborated within this data set. As shown in **Table 4-1**, the average values as well as the minimum and maximum values observed for the fold decrease in phage all increase as the concentration of RMR1-DCA increases with only minor overlap in the ranges observed. Therefore, it can be concluded that small molecule concentration, while not the only factor involved, does play a major role in phage viability. As seen previously with SF1, this concentration dependence is not universal, as SF1 was used at 300 μM to obtain observable results, a feat that was not possible with RMR1-DCA under the same labeling and tittering parameters.

Third is the unexpected results seen in unreduced samples. Despite having never been exposed to TCEP, phage that were unreduced also experienced a similar degree of

concentration dependent toxicity. While the large variation makes it difficult to compare to the reduced phage, the differences in the average value at each concentration do not appear to be significantly different. Based on the original model of phage toxicity, the unreduced phage at all concentrations should yield data similar to that for phage incubated with no RMR1-DCA. However, the change from the expected means that a change to the model that the knowledge of phage toxicity is based off of may require a change as well. A concentration mediated effect on unreduced phage means that there is another component to phage toxicity beyond the labeling of the pIII protein. Since the concentration of small molecules in solution is most often guaranteed to be significantly higher than the concentration of individual phage proteins in solution, this secondary reaction is likely to happen with both reduced and unreduced phage. Additionally, the data discussed previously indicates that this secondary reaction does not occur with only DCA or the RMR1 warhead but instead requires both parts of the small molecule to be covalently attached and intact.

Concurrently with the previous experiment set, a second set of experiments was conducted with a derivative of the discrete C7C construct used. This second construct was selected as it was already available in house and also contained a desired sequence: an identical displayed peptide with two extra GGGS linkers following the displayed sequence. This construct, hereby referred to as C7C Linker, had 8 additional amino acid residues between the displayed peptide and the C-terminus of the pIII protein, creating a larger distance between the displayed sequence that would be labeled and the main body of the phage particle. **Figure 4-5** shows the graphical representation of the labeling of this construct compared to identical labeling experiments of C7C shown previously.

Table 4-2 also depicts the unmodified values obtained for these experiments as an additional point of reference. Unfortunately, the phenomenon of highly variable data persisted across trials with the C7C linker construct as well, and as such it makes it hard to directly compare it to the C7C construct. While the average fold decrease in phage

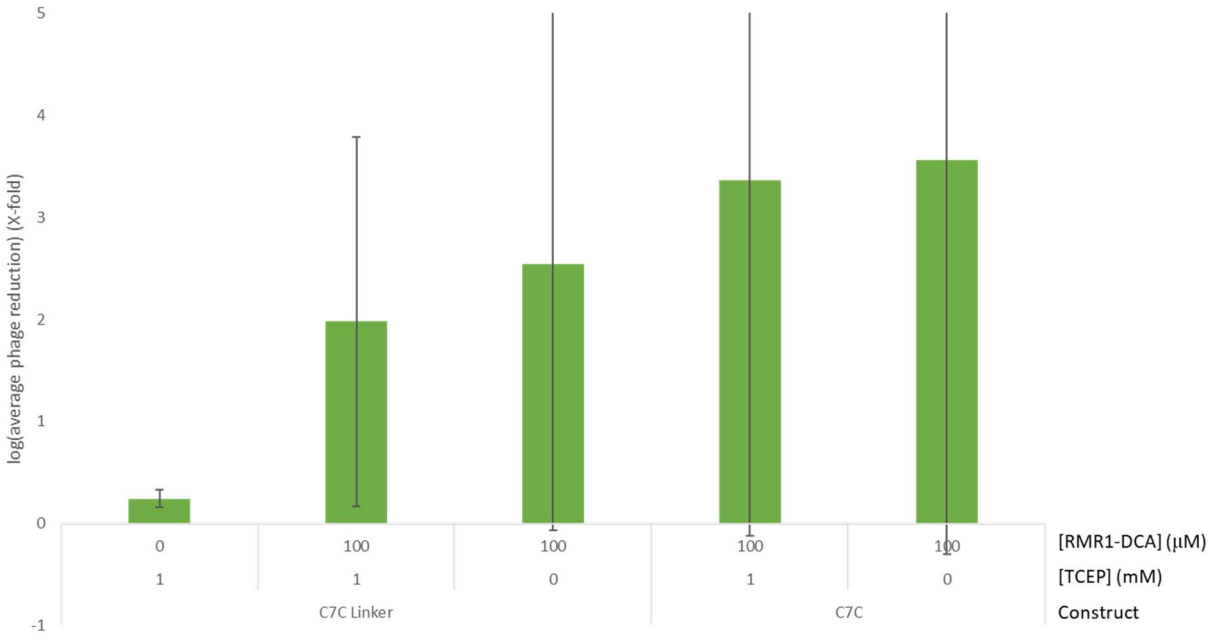


Figure 4-5: Comparison of RMR1-DCA toxicity vs. C7C and C7C Linker constructs. C7C Linker is an identical phage to the C7C construct used except it has two additional GGGS spacers between the pIII N-terminal domains and the CT domain, putting the N-terminal domains further from the main body of the phage. Data is again shown as the log of average phage reduction in X-fold for each sample. Similar to Figure 4-4, error determined from standard deviations yielded large values making conclusions difficult to make. The average reduction of phage for reduced C7C Linker phage exposed to RMR1-DCA appears to be slightly lower than other constructs, but due to large standard deviations, this determination could not be made accurately. Data in X-folds with minimum and maximum values is depicted in Table 4-2.

[TCEP] (mM)	[RMR1-DCA] (μM)	Average Phage Reduction (X-fold)	Standard Deviation	Minimum Value (X-fold)	Maximum Value (X-fold)	n
1	100	1.75	0.82	9.64	211.11	10
1	0	95.62	64.39	0.42	3.26	10
0	100	316.57	387.54	17.74	1112.86	6

Table 4-2: Phage reduction values due to RMR1-DCA incubation for C7C vs. C7C Linker testing. The reduction of phage in X-folds was determined by taking the ratio of input to output phage populations. The log of the average phage reductions is plotted in Figure 4-5.

output does appear to be lower with C7C linker than C7C, it is impossible to tell if it is a real result of fewer phage becoming unviable or if it is simply statistical fluctuation. While the maximum value observed for 100 μM RMR1-DCA incubation with C7C Linker is significantly lower than the average phage decrease seen with C7C, the decreased observed with unreduced C7C Linker is much closer to the values observed for C7C, which points in favor of the differences between the two constructs being purely due to statistical fluctuation. The C7C Linker construct did serve to corroborate the data seen with the C7C construct, in that a similar degree of toxicity was observed between reduced and unreduced phage samples. This helped to further support the theory that the current model of phage toxicity may need to be reassessed.

4.4 Conclusions

Conclusions were predicably difficult to draw based on the results shown here due to the large amount of variation present within all data sets. However, as noted previously, this variation is a factor worth addressing on its own. With the sample sizes of the performed experiments being considerably large (>20 trials for some experiments), the likelihood of variation coming from experimental error decreases, and therefore this variation may be derived directly from the type of experiment performed. This strongly supports the hypothesis that there are many contributing factors to phage toxicity that often go unaccounted for, however thus far a conclusive list of factors is difficult to determine. While factors such as small molecule concentration and linker length appeared to have an impact on phage toxicity, it is evident that these are not the only factors involved. Additionally, the degree of toxicity seen in unreduced phage samples clearly indicates a second, previously unaccounted for reactivity of the RMR1 warhead. **Figure**

4-6 shows a potential explanation for this unexpected toxicity. This hypothesis attempts to explain why the bifunctional RMR1-DCA molecule leads to such varying degrees of toxicity, but its individual components, being RMR1 and DCA, do not have a negative impact on phage. Though it is not clearly understood yet, this factor may pose a crucial barrier that must be considered in all covalent warhead design, as a second reactive group present on the surface of a phage particle would severely limit the potential for covalent phage library creation if ignored.

Based on the data obtained, it is suggested that if future work is to be continued towards developing covalent phage libraries, that the secondary reactivity of warhead molecules and phage particles be investigated further. As it stands, the number of unknown factors barring the way for covalent phage library creation appears to be great,

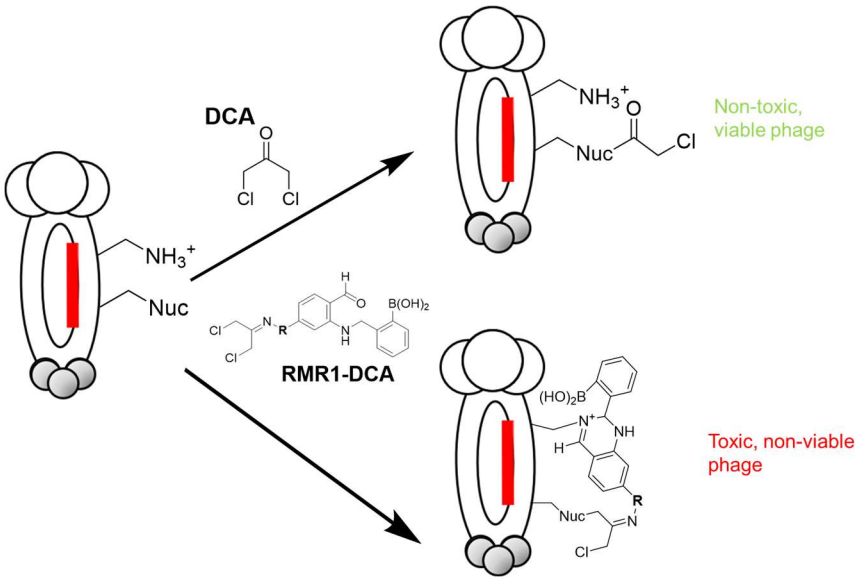


Figure 4-6: Hypothesized source of toxicity due to RMR1-DCA seen in unreduced phage. RMR1-DCA clearly showed toxicity towards unreduced phage that was comparable to reduced phage, meaning that labeling on the pIII protein is most likely not the only cause of toxicity. Since RMR1 and DCA are independently not toxic, it is hypothesized that the bifunctional molecule specifically can cyclize a second protein or proteins on the phage other than the pIII protein. When only DCA or RMR1 are included, this secondary reactive site may only be labeled but not cyclized and therefore toxicity may be derived from this cyclization. To date, this secondary reactive site has not been identified.

however a systematic and thorough investigation of these causes may allow for preventative action to be taken which, in turn, could allow for warhead implementation beyond luck. The distinct relationship between phage and covalent warheads still contains many unknowns, but if these factors are addressed and accounted for then there is hope of developing a more diverse and potent catalog of covalent libraries.

4.5 Experimental Methods

General Methods

All listed chemicals not synthesized in house by members of the Gao lab were purchased from Fisher Scientific or Sigma-Aldrich unless otherwise specified. Aliquots of *E. coli* strain ER2738 used for all tittering procedures was obtained from NEB. All M13KE strains originated from template M13KE plasmid which is commercially available from NEB as well. All media and buffers were prepared according to the recipes found in the NEB Ph.D.™ Phage Display Libraries Instruction Manual and all were sterilized via autoclave prior to use. All phage dilutions were mixed with Top Agar and plated on top of agar plates containing IPTG/XGal (~0.2 mM and 0.1 mM respectively), as denoted by the NEB Ph.D.™ Phage Display Libraries Instruction Manual. All plating was performed in a Biosafety Cabinet or in the presence of a flame. For all phage toxicity experiments described, results were determined by counting blue plaques present in plated phage dilutions following incubation for 12-16 hours at 37°C. All plaques were counted manually. For all tittering experiments, a minimum of three dilutions were plated to obtain an average concentration of phage in both input and output samples.

Preliminary Phage Toxicity Screening using C5C Phage Library

All RMR1 derivatives were synthesized in-house by Dr. Rahi Masoom Reja. Synthetic details can be found in the referenced literature. Aliquots of solid small molecules were obtained and resuspended in 500-1000 µL of pure DMF. Concentration of DMF solutions containing RMR1 derivatives were determined using a Thermo Scientific NanoDrop 2000 spectrophotometer measuring absorbance at 380 nm. An extinction coefficient of 5480 $M^{-1}cm^{-1}$ for the RMR1 warhead was determined using UV/Vis spectroscopy on the same instrument at 380 nm. Solid RMR1 was used to make a high concentration stock (10 mM) in pure DMF and then diluted with DMF. **Figure 4-7** shows the plot used to determine the extinction coefficient of the warhead in pure DMF. Stock solutions of all RMR1 warheads were diluted to appropriate concentration before addition to phage labeling reactions. All phage labeling reactions occurred in 100 µL reactions containing 5% DMF. Samples labeled with commercially available Biotin Iodoacetamide were also performed in 5% DMF. Stock solutions of Biotin Iodoacetamide were prepared by measuring mass of the compound and adding pure DMF to reach desired concentrations.

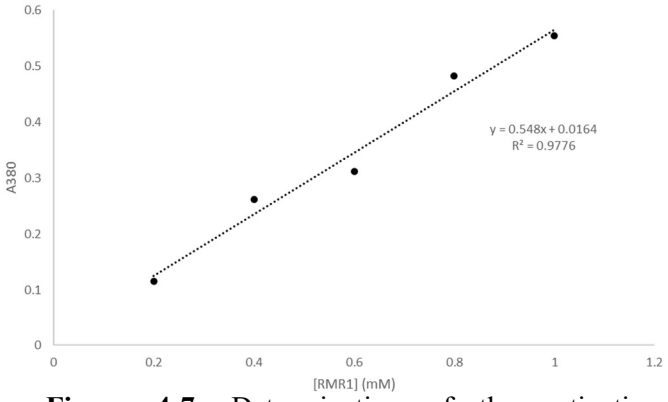


Figure 4-7: Determination of the extinction coefficient of RMR1-DCA in 100% DMF. UV absorbance at 380 was measured for samples containing known concentrations of RMR1-DCA (as determined by mass) and the linear formula of the plot of Absorbance vs. Concentration was used to determine an extinction coefficient of 5480 $M^{-1}cm^{-1}$ for RMR1-DCA.

All phage labeling experiments were performed using the same methodology as follows. 10 µL phage stock solutions of either phage library or discrete construct were

diluted up to 100 µL with 20 mM ammonium bicarbonate buffer pH 8.0. 1µL of either 100 mM TCEP solution in pure water or ammonium bicarbonate buffer was added to this mixture. The final concentration of TCEP in samples containing TCEP was 1 mM. Selective disulfide bond reduction was carried out at room temperature for 1 hour before precipitating phage using 1/6th volume (20 µL) 20% (w/v) PEG 2.5M NaCl on ice, as described in the NEB Ph.D.™ Phage Display Libraries Instruction Manual. Phage were precipitated on ice for 1 hour before centrifugation at max speed for 20 minutes to pellet phage. Supernatant was removed and the pellet was resuspended in 97 µL 20 mM ammonium bicarbonate buffer pH 8.0. 2 µL of this mixture were removed and saved to serve as input samples. Then 5% DMF containing small molecule to reach the desired final concentration was added. The resulting phage labeling mixtures were incubated in a water bath at 37°C for three hours before undergoing an identical precipitation. The final labeled phage product was then resuspended in 1x PBS buffer pH 8.0 before performing serial dilutions in low salt LB media for tittering.

Following modification, input and final output samples of phage were tittered in parallel. All diluted phage samples regardless of dilution factor were tittered using an identical method. 10 µL of phage dilution was added to 200 µL of *E. Coli* strain ER2738 liquid culture grown to mid-log phase in LB media ($OD_{600}\approx0.5$). This mixture was vortexed thoroughly to mix before adding 790 µL of molten top agar before mixing again. This agar-cell mixture was then spread on top of a 1% agar plate containing IPTG/XGal, as denoted in the NEB Ph.D.™ Phage Display Libraries Instruction Manual. The top agar was allowed to cool for 10 minutes before incubating for 12-16 hours at 37°C. Following incubation, plaques were counted to determine the plaque forming units

(pfu)/mL of each dilution and determine the average phage count in each sample. The ratio of average input:output phage count was then used to determine the average fold reduction of the phage solution due to small molecule incubation.

4.6 References

(1) Reja, R. M.; Wang, W.; Lyu, Y.; Haeffner, F.; Gao, J. Lysine-Targeting Reversible Covalent Inhibitors with Long Residence Time. *J. Am. Chem. Soc.* **2022**, *144* (3), 1152–1157.

(2) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; van Opijnen, T.; Gao, J. Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (19), 6137–6145.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Warhead Selection for Phage Display

The experiments detailed in this work were aimed at exploring and expanding the scope of covalent warheads that can be implemented into the phage display platform. The smaller scale experiments performed can be used to tell a larger story about covalent phage library design. Here it has been detailed that the pIII protein has the potential to be a valuable first assessment for warhead design. With the free protein behaving similarly to the native phage protein, crosslinkers and warheads can be assessed and modeled to identify adverse reactivity prior to phage labeling. As a test case study on warhead design, aryl-sulfonyl fluorides were examined and found to be unfit for phage display. While the aryl-SF motif passed testing with the pIII protein, phage labeling experiments yielded problematic results due to phage toxicity as well as potential warhead degradation. The nonspecific binding seen in ELISA results was the final point used to determine the unsuitability of this warhead. Finally, a more thorough investigation of warhead related toxicity was started and has thus far proven that a large number of factors beyond what was previously expected contribute to phage survival and potency.

For warhead selection as a whole, what has been shown here is that there are a number of parameters that must be accounted for in order for a warhead to be successful. Labeling capacity, warhead concentration, warhead stability, reactivity, and phage toxicity are all factors that have been explored here and shown to be vital in identifying new warheads. SF1 and RMR1 have both shown that even mildly electrophilic and reversible warheads can be challenging factors for phage labeling, and therefore successful warheads are likely to be derived from even milder reactive groups.

5.2 Conclusions for Covalent Library Design

The literature has shown that developing covalent phage libraries is possible, however expanding the scope of covalent phage libraries remains a challenge that has yet to be efficiently and definitively overcome. The complications and difficulties associated at various levels in the experimental process of warhead testing discussed here are a testament to the complex task that is creating a covalent phage library. While it is desirable to create a covalent phage library with a wider scope of reactivity to broaden the range of targetable proteins, it has become evident that the phage itself must be treated as a reactive body as well. Thus, expanding the scope of covalent phage library design becomes a careful balancing act: finding a balance between general warhead reactivity to identify meaningful and novel protein binders and warhead selectivity to avoid adverse reactions within the phage particle. While SF1 yielded unsuccessful results, continued study on RMR1 as well as other covalent warheads may still yield more success if subjected to similar route of trials as the warheads discussed here.

5.3 Future Directions

Despite the seemingly discouraging results of SF1, numerous opportunities exist to continue this work into the future. The pIII protein has shown to be a valuable tool in modeling phage labeling, so continued optimization of the expression methods may yield an even more valuable and quantifiable tool for warhead assessment. While SF1 was deemed ineligible for library creation, there is more work that can be done with aryl sulfonyl fluorides. Literature has shown that derivatization of the phenyl ring has a significant impact on sulfonyl fluoride reactivity, and thus SF1 derivatives may exist that could prove to be strong warhead candidates. Similarly, identification of the source of

nonspecific binding of SF1 on phage could allow for a potential elimination of this reactivity through phage mutations. Elimination of this reactivity may yield phage with intact aryl sulfonyl fluoride that could yield potent protein surface binders. Further work conducted with RMR1 and similar warheads may also yield more specific and valuable information regarding sources of phage toxicity. The experiments discussed here were not successful in identifying a warhead suitable for the expansion of covalent phage libraries, but the methodology utilized can still pave the way for larger-scale warhead screening that may yet yield a strong candidate warhead for the creation of covalent phage libraries.