Development of a Live Cell Phage Display Screening Protocol

Sandra Sisko

A thesis submitted to the

Faculty of the 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

November 2022

Development of a Live Cell Phage Display Screening Protocol

Sandra Sisko

Advisor: Jianmin Gao, Ph.D.

Protein-protein interactions (PPIs) are essential for all biological functions. Developing peptides that disrupt these PPIs is an avid research effort, as peptides possess several advantages over small molecules and monoclonal antibodies. Peptide phage display is a useful tool in identifying peptides for targeting PPIs. This technology displays up to 10^{10} unique polypeptides on the surface of bacteriophage, which after several rounds of panning enriches high affinity peptide sequences towards a target protein. Phage display is classically done on immobilized discrete protein; however, we propose to use this technology to identify peptides ligands for overexpressed oncogenic proteins on live cells *in-vitro*. This is a more accurate representation of the therapeutic target landscape and resembles how the peptide will interact with the receptor *in-vivo*. Several groups have explored live cell panning, such as Ruoslahti *et al.* and Cieslewicz *et al.*, and while they demonstrate the capabilities of *in-vitro* style phage display, there are areas for improvement. We intend to improve on this previous work by 1. Identifying a peptide ligand against specific receptor/protein, and 2. By incorporating the use of covalent phage libraries to elucidate a high affinity binder.

This work will be accomplished using the mammalian epidermal oncogenic cell line, A431, that is known to overexpress epidermal growth factor receptor (EGFR). Epidermal growth factor receptor (EGFR) is responsible for cellular proliferation, survival, differentiation and metastasis, which makes it an attractive target to inhibit oncogenic

proliferation. Despite successfully marketed monoclonal antibodies and tyrosine kinase inhibitors, EGFR can mutate and develop resistance as diseases progress; this phenomenon, in addition to the benefits of peptides as therapeutics, are driving factors for pursuing this project. Despite our best efforts using non-covalent phage libraries to identify a viable ligand, screening against EGFR extracellular domain (ECD) has proven to be more difficult than anticipated. We hypothesize that non-covalent phage libraries do not possess any sequences with a high enough binding affinity for this protein, and that the use of covalent libraries will be needed to pull out a positive hit. Due to these findings, we have successfully constructed two phage libraries, a ACX_7C and a ACX_7C-TEV, where the latter introduced a TEV protease cleavage site on the C'-terminal side of the randomized amino acids suitable for covalent warhead modification and screening.

Further, we have begun work on constructing an EGF-displaying phage construct to aid in optimizing a live cell panning protocol. In the future, we plan to evaluate ligand affinity and protein density, as well as determine the optimal covalent warhead/peptide combination for live cell screenings. With this information, we intend to apply this to other oncogenic cell lines, such as MCF-10CA1a, to identify potent peptide ligands for overexpressed oncogenic proteins.

To my late dad:

I will love you forever and always

Thank you for your unconditional support

I am so proud to be your daughter

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Jianmin Gao, for his guidance, knowledge, and encouragement during my time in lab. He has helped me grow as a scientist and critical thinker in many ways. I would also like to thank the other members of my committee, Professor Abhishek Chatterjee and Professor Jia Niu, for providing helpful feedback during my qualifying exam and taking the time to read my thesis.

I would like to thank my fellow lab members for their advice and help with experiments. 209A will always hold a special place in my heart; your friendship made being in lab more enjoyable. I would especially like to thank my mom and brother for their unconditional support; I love you both dearly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

ATP	Adenosine triphosphate
bioEGFR	Biotinylated epidermal growth factor receptor
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
CBT-CA	Cyanobenzothiazole-chloroacetamide
diH2O	Deionized water
DIPEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
ECD	Extracellular domain
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
FAM	Carboxyfluorescein
FDA	Food and Drug Administration
FITC	Fluorescein
Fmoc-SPPS	Fluorenylmethoxycarbonyl-Solid phase peptide synthesis
HBTU	(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, Hexafluorophosphate Benzotriazole Tetramethyl Uronium)

HPLC-UV	High-performance liquid chromatography-ultra-violet spectroscopy
HPQ	Histidine-Proline-Glutamine
IDT	Integrated DNA Technologies
IPTG/XGal	Isopropyl β- d-1-thiogalactopyranoside/5-Bromo-4-Chloro-3-Indolyl-beta-D-Galactoside
LB	Luria-Bertani
LCMS-QTOF	Liquid chromatograph mass spectrometer-quadrupole time of flight
NEB	New England Biolabs, Inc.
NH_4HCO_3	Ammonium bicarbonate
NH_4OAc	Ammonium acetate
NHS	N-Hydroxysuccinimide
NSCLC	Non-small cell lung cancer
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline – Tween-20
PCR	Polymerase Chain Reaction
PEG/NaCl	Polyethylene glycol/sodium chloride
pfu	Plaque forming units
PPIs	Protein-protein interactions
RF	Replicative form
rpm	Revolutions per minute
SOB	Super optimal broth
ssDNA	Single-stranded deoxyribonucleic acid
TBS	Tris-buffered saline

TCEP	Tris(2-carboxyethyl)phosphine
TE	Transformation efficiency
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TGFα	Transforming growth factor alpha
TKIs	Tyrosine kinase inhibitors
TNBC	Triple negative breast cancer

CHAPTER 1: INTRODUCTION

1.1 PHAGE DISPLAY PLATFORM

George P. Smith and Sir Gregory P. Winter were awarded the Nobel Prize in Chemistry in 2018 for their work with phage display of peptides and antibodies.[1] This powerful application allows scientists to screen libraries of recombinantly expressed DNA on phage virions against a target of interest to evaluate receptor-ligand interactions.[1,2] The filamentous bacteriophage strains M13, f1, and fd are the most common vectors used for this technique.[2] Filamentous phage are malleable rods approximately 1 μm long and 6 nm in diameter with over 2,700 copies of its pVIII major coat protein helically oriented along the length of the virus. Several other coat proteins with only 3-5 copies per virion, including pIII, pIX, pVI, and pVII are located at the top and bottom of the virus, as illustrated in **Figure 1-1**.[2]

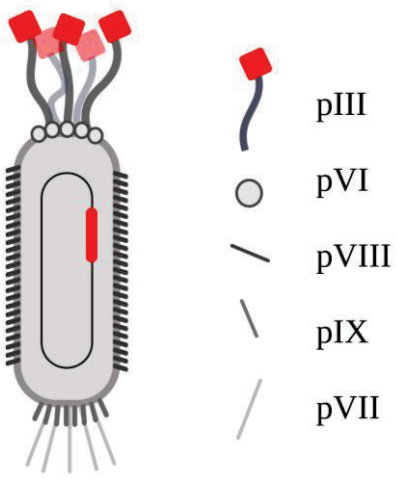


Figure 1-1. M13 filamentous phage structure highlighting its coat proteins. Created with BioRender.com.

Recombinant DNA, typically from chemically synthesized DNA, is most commonly displayed as peptides, proteins, or antibodies on the N'-terminus of the pIII

(illustrated as a red square in **Figure 1-1**) or pVIII coat proteins for high-throughput affinity screenings. The pIII minor coat protein offers several advantages compared to the pVIII major coat protein; first, the displayed peptide will be less sterically hindered, second, it preserves the functionality of the pIII protein, and third, it can identify high affinity binders.[1] Phage display using the pVIII protein is often used to enhance the detection signal and increase the efficiency of display, however, it is too sterically hindered to display proteins and antibodies.[1]

The displayed peptide libraries can consist of up to 10^{10} diverse sequences, depending on the length of the amino acid sequence. Traditional phage display is executed using a biotinylated target that is immobilized on streptavidin magnetic beads. The immobilized target is incubated with the phage library for a predetermined time, followed by washing away the non- and low-affinity binding phage, before eluting with a glycine pH 2.2 buffer. The elution of high affinity binding phage is then amplified in *E. coli*; the N-terminus of the pIII protein will attach to the F pilus which promotes the F pilus to retract, internalizing the phage DNA into the bacterial cell.[2,3] Once inside, the phage ssDNA will become RF dsDNA through bacterial machinery, various viral proteins will be expressed, and several hundred new phage per cell per division cycle will be packaged and exported from the cell.[2,3] This amplified output becomes the input for following screening rounds.

After several rounds of screenings and amplifications, high affinity binding phage for the target become enriched. Peptide(s) that exhibit preferential binding for the target can be easily identified through Sanger sequencing, as the peptide sequence(s) are

genetically encoded into the phage genome.[4] This pentavalent display increases binding affinity and potency against the target protein.[5]

1.2 BENEFITS TO LIVE CELL SCREENING

Protein-protein interactions (PPIs) comprise essentially all intercellular processes in biology. These interactions initiate, indirectly or directly, enzymatic reactions, post-translational modifications of proteins, and downstream signaling cascades, among other functions.[6] Developing peptides that disrupt these PPIs is an avid research effort, as peptides possess several advantages over small molecules and monoclonal antibodies. Peptides have a higher degree of conformational flexibility, which allows them to adapt more readily to irregularly shaped targets.[6,7] They have excellent safety profiles, and exhibit higher affinity and specificity than small molecules.[6,7] In addition, they are less immunogenic, membrane permeable and significantly cheaper to scale and manufacture compared to antibodies.[6,7]

Peptide phage display has given researchers a considerable advantage in identifying peptides for targeting PPIs. This technology displays up to 10^{10} diverse polypeptides on the surface of filamentous bacteriophage, which after several rounds of panning gives researchers peptide sequences with high affinity towards their target.[6] These genetically encoded peptide libraries are limited to linear or cyclic display of naturally occurring amino acids, but phage modifications such as cyclization linkers, fluorophores, small molecules, and cleavage sites have allowed researchers to expand upon this technology.[6]

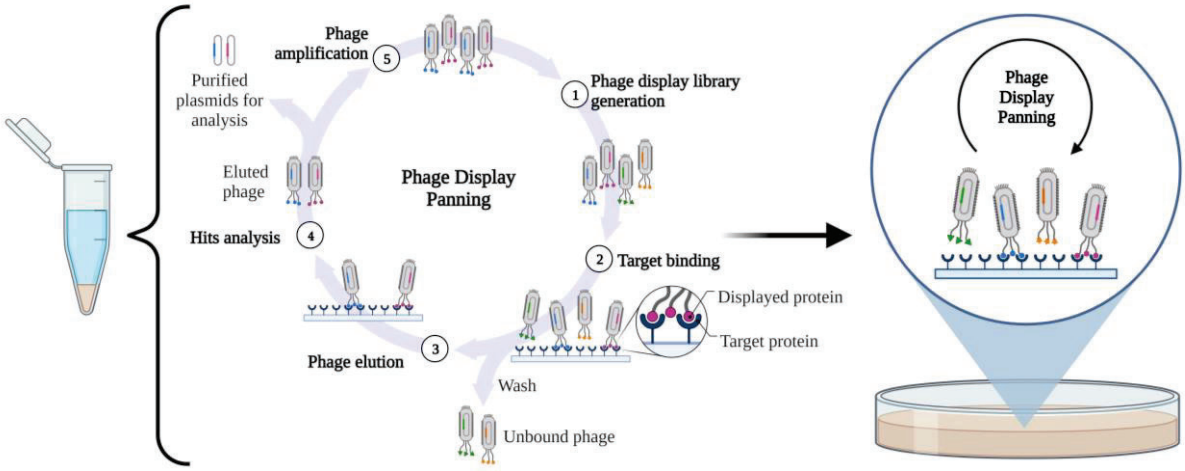


Figure 1-2. Utilizing the phage display platform for live cell screening. Adapted from "Phage Display Panning", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

We propose to use this technology to identify peptides ligands for overexpressed oncogenic proteins on live cells *in-vitro*. Panning and identifying peptide ligands on cells is a more accurate representation of the therapeutic target landscape and resembles how the peptide will interact with the receptor *in-vivo*. In addition, this protocol can be applied to other oncogenic cell lines with unknown protein expression levels to identify peptide ligand(s), expanding the ability to identify potential therapies for diseases.

1.3 PREVIOUS WORK EXPLORING *IN-VITRO* PHAGE DISPLAY

Phage display is traditionally done on immobilized discrete protein; however, several groups have explored live cell panning. Ruoslahti *et al.* investigated whether peptides can localize to specific tissues in mice.[8] They accomplished this by injecting phage at 10^9 plaque forming units (pfu) into nude mice with MDA-MB-435 breast carcinoma

xenografts. Following recovery of phage, they identified three distinct motifs: RGD, NGR, and GSL.[8] The RGD motif was identified in the RGD-4C peptide (CDCRGDCFC) which has shown selective binding towards α_v integrins.[8,9] They showed that RGD-4C phage selectively binds to the tumor vasculature over the control tissue. In addition, the researchers conjugated doxorubicin, a known antiangiogenic chemotherapy, to RGD-4C, and observed that mice who were administered this conjugate compared to doxorubicin alone had significantly smaller tumors, decreased lymph node metastasis, decreased toxicity to the liver and heart, and increased overall survival at different doses.[8] The researchers concluded that the RGD-4C phage selectively binds to human vasculature, as α_v integrins are known to be expressed in tumor blood vessels.[8]

Furthermore, Cieslewicz *et al.* discovered a peptide, YEQDPWGVKWWY (M2pep), that specifically targets M2 macrophages from subtractive live cell biopanning.[10] They were able to demonstrate selectivity over other leukocytes, specifically M1 macrophages, neutrophiles, and bone-marrow derived dendritic cells.[10] M2pep had a K_D of 90 μM against murine M2 macrophages, yet no activity against human M2 cells.[10] Despite this, M2pep exhibited *in-vivo* targeting and activity, improved survival times, and slowed tumor growth.[10]

While this literature demonstrates the capabilities of *in-vitro* phage display, there are areas for improvement. First, Ruoslahti *et al.* was unable to identify the target protein for which their phage was binding to, instead they made broad conclusions about receptor and cell type.[8] Second, Cieslewicz *et al.*., despite identifying a selective M2 peptide, it showed poor binding affinity to murine M2 cells and no affinity towards human M2 cells.[10] We intend to improve on this previous work by 1. Identifying a peptide ligand against a

specific receptor/protein, and 2. By incorporating the use of covalent phage libraries to elucidate a high affinity binder.

1.4 MCF-10CA1a CELL SCREENING

Initial work previously done by Dr. Kaicheng Li in our lab investigated panning libraries against the MCF-10CA1a cell line. This breast cancer cell line is the most malignant in its series of KRAS-expressing isogenic cell lines, namely MCF-10A (normal), MCF-10AT (pre-malignant), and MCF-10CA1a (highly aggressive and malignant).[11,12] We aimed to identify high affinity binding peptides that selectively binds to overexpressed proteins on the oncogenic cell surface as compared to its benign counterparts.

Dr. Li attempted phage panning three times, twice with a Ph.D.-CX_7C library and once with our lab's CX_9C library. After sequencing select plaques from rounds 2 and 3, as well as calculating the output ratio (output/input x 10^{-6}) and comparing that to a negative control (panning conducted on a plate containing no cells), it was concluded that no prospective peptide hits could be pulled out, illustrated in **Tables 1-1 and 1-2**.

Table 1-1. Output Ratios (10^{-6}) of trials screening against the MCF-10CA1a cell line. The "Blank Plate" was used as a negative control that contained no adherent cells, yet still underwent the entire phage display screening.

	Output Ratio (10^{-6})			
	Trial 1 CX_7C	**Trial 2 CX_7C**	**Trial 3 CX_9C**	**Blank Plate**
Round 1	3.33	3.33	10	10
Round 2	5	4	25	50
Round 3	2.5	4	10	40

Table 1-2. Sanger sequencing results of rounds 2 and 3 for trials screening against the MCF-10CA1a cell line. Sequences colored blue were determined to be biased based on previous assays conducted in our lab.

Trial 1 CX_7C		Trial 2 CX_7C		Trial 3 CX_9C	
Round 2	**Round 3**	**Round 2**	**Round 3**	**Round 2**	**Round 3**
CLHSTKTSC	CDRSTTKIC	CMGGKPSTC (2)	CNFGKNAHC (2)	CENAPSFIKTC	CRPSFQSNLMC (3)
CDSRLLNGC	CSTLHQKLC	CNFGKNAHC	CSTLHQKLC (3)	CSGSLNKYTFC	CNTGGSLIKKC (2)
CDGRPDRAC (2)	CINGTHSQC	CSASFKTDC	CMKESIRGC	CSNTWPRPLYC	CTTGTYSERNC (3)
CTDKASSSC	CDGRPDRAC (2)	CNAAQHSDC	CSSRVFTSC	CEQGSTFTNDC	CSAKRYMMGKC
	CTDKASSSC (2)	CNTRSTELC	CDGRPDRAC (2)	CFSELKRGRWC	CSAAHRLTGHC
				CSVANITNPLC	CVRDLSNPSTC
				CHVPPSGFATC	CDNANRMGLSC
				CTPSPNSDRLC	CINKGQGHNYC
				CKHKMPYSKNC	CLVRXGDHNYC
				CSTPRKDTGRC	CGSHHNIGSLC
				CNFSHELSSTC	CKDGKRGTTQC
					CPSNDVHFKLC
					CFNKNINSSTC

With this information, we decided that a positive control protein that is overexpressed on an oncogenic cell line, and identifying a peptide ligand that discretely binds to it would be the most efficient way to optimize a live-cell protocol. We will be working with a mammalian epidermal oncogenic cell line, A431, that is known to overexpress epidermal growth factor receptor (EGFR)[13,14] to identify a high affinity peptide ligand. We sought to improve potency and selectivity, and identify a novel peptide from live cell panning.

1.5 EPIDERMAL GROWTH FACTOR RECEPTOR

Epidermal growth factor receptor (EGFR) is responsible for cellular proliferation, survival, differentiation and metastasis, which makes it an attractive target to inhibit oncogenic

proliferation.[15] EGFR is a member of the ErbB/HER tyrosine kinase family and its expression ranges from 40,000-100,000 receptors per cell on normal cells.[15] Overexpression of this protein can lead to up to $2x10^6$ receptors per cell and can occur in several cancer types, such as breast, lung, colorectal, and NSCLC.[15] Several ligands bind and activate EGFR, including epidermal growth factor (EGF), transforming growth factor-α (TGFα), and amphiregulin.[15] Upon ligand binding, EGFR extracellular domain (ECD) undergoes a conformational change that exposes a dimerization arm to form dimers with other EGFR molecules; this results in EGF being immobilized between domains I and III on the ECD, as illustrated in **Figure 1-3**.

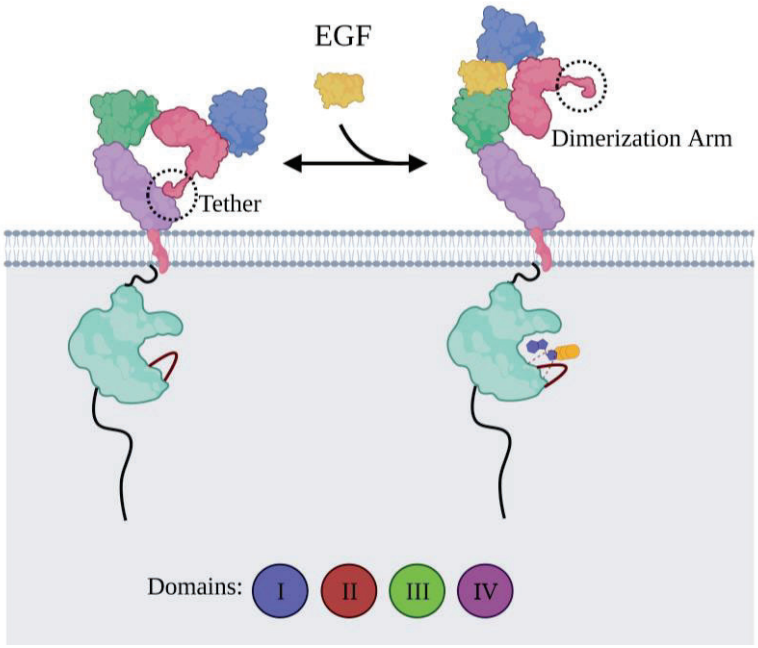


Figure 1-3. Epidermal growth factor receptor structure and subsequent conformational change from EGF binding. The extracellular domain (domains 1-4) is connected to the transmembrane domain, which is attached to the intracellular tyrosine kinase domain. Created with BioRender.com.

There have been several efforts to develop antagonists for EGFR signaling and to hinder oncogenic proliferation. FDA approved monoclonal antibodies, such as Cetuximab, Panitumumab, and Necitumumab competitively bind to domain III on the ECD to inhibit

EGF binding.[15,16] In addition, FDA approved small molecule tyrosine kinase inhibitors (TKIs) competitively bind to the ATP binding pocket on the kinase domains inhibiting ATP binding and downstream signaling.[15,17] TKIs have also been developed for common mutations of EGFR, such as T790M and L858R.[15,17] Despite these successfully marketed therapies, EGFR can mutate and develop resistance to other TKIs as diseases progress; this phenomenon, in addition to the benefits of peptides as therapeutics, are driving factors for pursuing this project.

1.5.1 Literature Reported Positive Controls

Our first effort was to examine the literature to see if there are any known reported EGFR peptide ligands. Li *et al.* from the Shanghai Cancer Institute, China, identified a 12-mer peptide sequence using phage display that is reported to bind with a K_D=22nM.[18] This peptide, GE-11, (YHWYGYTPQNVI) displayed on phage was able to competitively bind to EGFR with 0.5mM EGF present.[18] They also reported the internalization of FITC-labeled GE-11 at 5μM.[18] Researchers from Chapman University expanded upon this finding and evaluated the binding affinity of 29 analogues of GE-11 against triple negative breast cancer (TNBC) cell lines.[19] They reported that substituting Q9E resulted in a 2.8 fold increase in binding affinity to the TNBC cell lines as compared to GE-11; no K_D was reported.[19]

Similarly, Hamzeh-Mivehroud *et al.* from the Tabriz University of Medical Sciences, Iran, conducted subtractive biopanning against A431 cells using a 7-mer phage library and reported two distinct converging sequences after four rounds.[14] Labeled P1 and

P2, these sequences SYPIPDT and HTSDQTN, respectively, competitively bind to EGFR in the presence of EGF.[14] This was the only example of peptide phage panning against A431 cells we were able to find reported in the literature.

1.6 REFERENCES

(1) Bazan, J.; Całkosiński, I.; Gamian, A. *Human Vaccines & Immunotherapeutics* **2012**, *8* (12), 1817–1828.
(2) Smith, G. P.; Petrenko, V. A. *Chem. Rev.* **1997**, *97* (2), 391–410.
(3) Cwirla, S. E.; Peters, E. A.; Barrett, R. W.; Dower, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87* (16), 6378–6382.
(4) Scott, J. K.; Smith, G. P. *Science* **1990**, *249* (4967), 386–390.
(5) Instruction Manual, Ph.D. Phage Display Libraries, 2020.
(6) Lee, A. C.-L.; Harris, J. L.; Khanna, K. K.; Hong, J.-H. *IJMS* **2019**, *20* (10), 2383.
(7) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. *Sig Transduct Target Ther* **2022**, *7* (1), 48.
(8) Arap, W.; Pasqualini, R.; Ruoslahti, E. *Science* **1998**, *279* (5349), 377–380.
(9) Pasqualini, R.; Koivunen, E.; Ruoslahti, E. *Nature*. **1997**, *15*, 542-546.
(10) Cieslewicz, M.; Tang, J.; Yu, J. L.; Cao, H.; Zavaljevski, M.; Motoyama, K.; Lieber, A.; Raines, E. W.; Pun, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (40), 15919–15924.
(11) Imbalzano, K. M.; Tatarkova, I.; Imbalzano, A. N.; Nickerson, J. A. *Cancer Cell Int* **2009**, *9*, 7.
(12) Martinko, A. J.; Truillet, C.; Julien, O.; Diaz, J. E.; Horlbeck, M. A.; Whiteley, G.; Blonder, J.; Weissman, J. S.; Bandyopadhyay, S.; Evans, M. J.; Wells, J. A. *eLife.* **2018**, *7*, 1-26.
(13) Zhang, F.; Wang, S.; Yin, L.; Yang, Y.; Guan, Y.; Wang, W.; Xu, H.; Tao, N. *Anal. Chem.* **2015**, *87* (19), 9960–9965.
(14) Hamzeh-Mivehroud, M.; Mahmoudpour, A.; Dastmalchi, S. *Chemical Biology & Drug Design* **2012**, *79* (3), 246–259.
(15) Herbst, R. S. *International Journal of Radiation Oncology*Biology*Physics* **2004**, *59* (2), S21–S26.
(16) Cai, W.-Q.; Zeng, L.-S.; Wang, L.-F.; Wang, Y.-Y.; Cheng, J.-T.; Zhang, Y.; Han, Z.-W.; Zhou, Y.; Huang, S.-L.; Wang, X.-W.; Peng, X.-C.; Xiang, Y.; Ma, Z.; Cui, S.-Z.; Xin, H.-W. *Front. Oncol.* **2020**, *10*, 1249.
(17) Solassol, I.; Pinguet, F.; Quantin, X. *Biomolecules* **2019**, *9* (11), 668.
(18) Li, Z.; Zhao, R.; Wu, X.; Sun, Y.; Yao, M.; Li, J.; Xu, Y.; Gu, J. *FASEB H.* **2005**, *19* (14), 1978–1985.
(19) Hossein-Nejad-Ariani, H.; Althagafi, E.; Kaur, K. *Sci Rep* **2019**, *9* (1), 2723.

2.0 CHAPTER 2: PHAGE DISPLAY PROTOCOL OPTIMIZATION AGAINST EPIDERMAL GROWTH FACTOR RECEPTOR

2.1 UTILIZING LITERATURE REPORTED POSITIVE CONTROLS

2.1.1 Phage Retention

We cloned the reported sequences, YHWYGYTPQNVI (PQN), YHWYGYTPENVI (PEN), SYPIPDT (SYP), and HTSDQTN (HTS), into the M13KE phage vector following the NEB Cloning Peptide Display Libraries in M13KE (explained in detail in **Section 2.3.2** and **Figure 2-5**).[1] A handful of blue plaques (6-10) were picked the following day for DNA extraction and Sanger sequencing to confirm proper insertion. Results are shown in **Table 2-1**; although the correct inserts were not present in all plaques sequenced, we are still able to generate discrete phage stocks by collecting and amplifying the supernatant of the mini-prep culture.

Table 2-1. Sanger sequencing results of the cloned positive control sequences.

SYP	HTS	PQN	PEN
ASYPIPDTGGGS (4)	AHTSDQTNGGGS (6)	AYHWYGYTPQNVIGGGS (8)	AYHWYGYTPENVIGGGS (6)
LQHHQSSCGGGS		CSLLQHHQSSCGGGS (2)	CSLLQHHQSSCGGGS (2)
PRSSLMACGGGS			CALPRSSLMACGGGS (2)

Biotinylated epidermal growth factor receptor (bio-EGFR) extracellular domain (residues 25-645) was ordered from Acro Biosystems, and functionality was confirmed through agarose bead microscopy using EGF-FAM (Invitrogen). Streptavidin coated agarose beads immobilized the bio-EGFR through the binding interaction between

streptavidin and biotin, $K_d = 10^{-15}$ M. Microscopy results confirming protein functionality are shown in **Figure 2-1**.

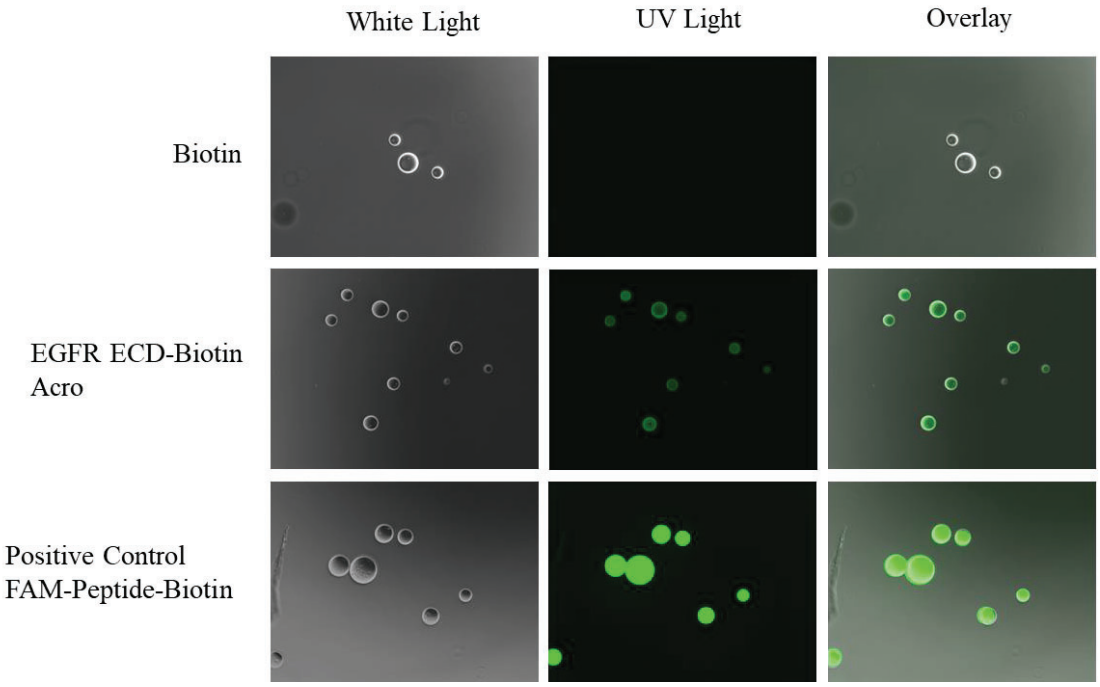


Figure 2-1. Agarose bead microscopy of the biotinylated EGFR-ECD from Acro Biosystems using EGF-FAM. Fluorescence is observed on the surface of the agarose beads indicated functional protein.

We conducted phage retention assays against EGFR ECD in triplicate. Results are shown in **Table 2-2** and **Figure 2-2**. Our negative control was the M13KE phage template with no displayed peptide on the pIII coat protein, labeled "Template" below. We concluded that the literature phage positive controls were irreproducible because the output ratio (10^{-6}) was less than our negative control template construct.

Table 2-2. Phage retention, in triplicate, of the literature reported positive control peptides. The "Template" was used as a negative control, which was the M13KE phage virion with no insert displayed on the pIII coat protein.

	Replicate #	Input Average (PFU)	Output Average (PFU)	Out/In (10^{-6})	Average Ratio (10^{-6})	Standard Deviation
PQN	**1**	$1.3x10^{11}$	$1.04x10^{6}$	8.00	7.45	2.51
	2	$2.0x10^{11}$	$1.93x10^{6}$	9.65		
	3	$1.7x10^{11}$	$8.0x10^{5}$	4.70		
PEN	**1**	$1.6x10^{11}$	$1.22x10^{6}$	7.63	6.59	1.18
	2	$1.7x10^{11}$	$1.16x10^{6}$	6.82		
	3	$1.6x10^{11}$	$8.5x10^{5}$	5.31		
SYP	**1**	$2.8x10^{10}$	$1.1x10^{5}$	3.93	2.33	2.26
	2	$4.1x10^{10}$	$3.0x10^{4}$	0.732		
	3	$3.2x10^{10}$	NA	NA		
HTS	**1**	$3.9x10^{10}$	$6.0x10^{4}$	1.54	0.843	0.627
	2	$3.1x10^{10}$	$1.0x10^{4}$	0.323		
	3	$3.0x10^{10}$	$2.0x10^{4}$	0.667		
Template	**1**	$1.2x10^{11}$	$1.58x10^{6}$	13.2	11.1	5.54
	2	$1.2x10^{11}$	$1.84x10^{6}$	15.3		
	3	$1.1x10^{11}$	$5.3x10^{5}$	4.82		

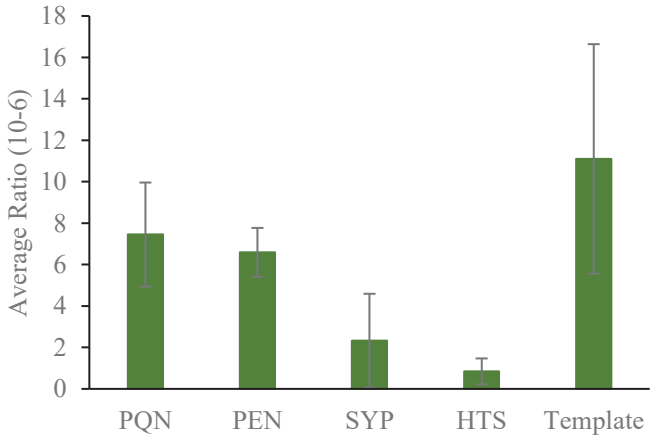


Figure 2-2. Graphical representation of the average output ratio (10^{-6}) vs. literature reported positive control phage clone against EGFR-ECD.

2.1.2 PQN/PEN Microscopy

To further corroborate these results, we also synthesized FITC-labeled PQN and PEN peptides in accordance with the reported protocol.[2] We conducted agarose bead microscopy to assess binding, and observed no fluorescence on EGFR ECD (**Figure 2-3**).

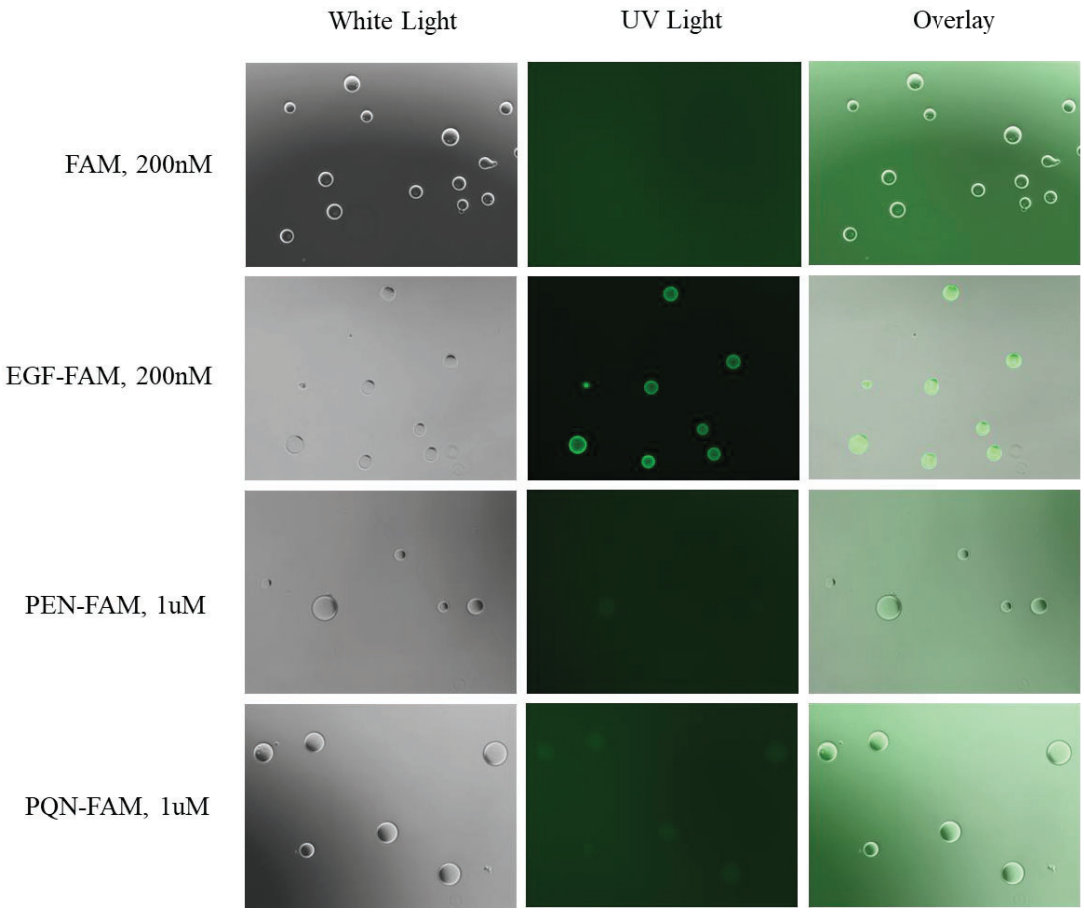


Figure 2-3. Fluorescent microscopy of FITC-labeled PQN and PEN peptides from literature.[2] EGFR-biotin (2.5ug) was immobilized on agarose-streptavidin beads and incubated with 200nM FAM, 200nM EGF-FAM, 1uM FITC-PQN, or 1uM FITC-PEN at room temperature for 30 minutes and then imaged.

We show that these results are irreproducible and propose this to be the case as the original authors[3] conducted all binding assays against the SMMC-7721 human hepatoma cell line, not discrete EGFR ECD. It is likely that their peptides were binding to other

component(s) of the cell, skewing the observed results. With this information in hand, our efforts shifted to identify a novel peptide ligand for EGFR ECD using phage display.

2.2 LIBRARY PANNING

2.2.1 CX_9C and 12-mer Libraries

The Ph.D-12-mer Phage Display Library (Lot: 10111203) was ordered from NEB on the notion that PQN and PEN were originally identified from this library. In addition, our lab had a CX_9C phage library with an N-terminal HA-tag and Factor Xa cleavage site, and this was screened against EGFR ECD in parallel with the 12-mer library. Panning results are shown in **Table 2-3** for both trials.

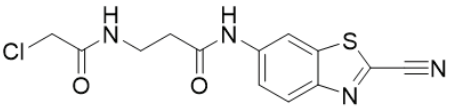
Table 2-3. Input, output, output ratio, and Sanger sequencing results of rounds 3-5 of the screenings with CX_9C and 12-mer libraries against EGFR-ECD. Repeating sequences are color-coded.

	CX_9C			12-mer		
	Round 3	**Round 4**	**Round 5**	**Round 3**	**Round 4**	**Round 5**
Input (PFU)	$2.8x10^{10}$	$5.5x10^{10}$	$3.0x10^{10}$	$3.5x10^{10}$	$2.1x10^{10}$	$5.0x10^{10}$
Output (PFU)	$9.5x10^{6}$	$3.1x10^{8}$	$1.1x10^{8}$	$1.3x10^{8}$	$3.7x10^{7}$	$1.1x10^{7}$
Out/In (10^{-6})	339	5640	3670	3710	1760	220
	CSQYSTGTGTC	CSQYSTGTGTC	CMAFEKSRPRC (3)	QLDSRHFGWFMP (3)	QLDSRHFGWFMP	QLDSRHFGWFMP (3)
	CNIITKAPRNC	CIDRYSREPFC	CNIITKAPRNC	VVPSEYDRHSFS	QSVKTSNNWWLF	LLGMADTNHNHW
	CPSTKLTAGLC	CSLFDSYMRNC	CPSTKLTAGLC	MLNHPAYGIRLT	YSVNGLKHTGVV	NGYQVHPATSPP
	CSSRLNTNPTC	Template (8)	CFVSRDKTASC	SGTTSMYVSWTR	GIHASILPEVRE	QLVKTXNNGGPX
	CVFAGSHTNRC		CTRNTPMRWQC	ANLDLSPHEQWS	YATKNMQTPVSL	
	CTSDHLRRGHC		CSGTLSNSLLC	ASDSFVLLSRGS	NGYQVHPATSPP	
	CPAAQKSNTPC		CDRFNSNQPFC	VNSSSAFTDDGR	WDPGTFSYMLGA	
	CTHDLNVSGMC		CNKTSLYHKSC	SPNGDHWPRITV	TETAASHHAQRM	
	CPTRKEGSMIC		CSLKLNNNNYC		SDLYPGRSVNTN	
	CMDNGDAVNNC					
	CQHVPFRNGTC					

Although repeating sequences were observed across rounds 3-5, no convergence was noted as the rounds progressed, which would indicate an enriched binder for EGFR ECD. We concluded that these sequences are biased, and not true binders. No phage retention was done for these sequences.

2.2.2 CX_9C-Luciferin Cyclized Library

To determine the most suitable library to use next, we looked at the intermolecular forces between EGF and EGFR ECD in an attempt to disrupt this interaction. There are three binding sites between EGF and EGFR ECD. The first between EGF and domain I, and the following two between EGF and domain III.[4,5] Hydrophobic interactions make up the majority of contacts between the two proteins, followed by salt-bridges.[5] We reasoned that by increasing the hydrophobicity of our phage library, we would have a better chance of pulling out a positive it. This was accomplished by Factor Xa cleavage followed by reducing the disulfide bond of the CX_9C library, then cyclizing with cyanobenzothiazole-chloroacetamide generating a luciferin moiety, illustrated in **Figure 2-4**.



cyanobenzothiazole-chloroacetamide (CBT-CA)

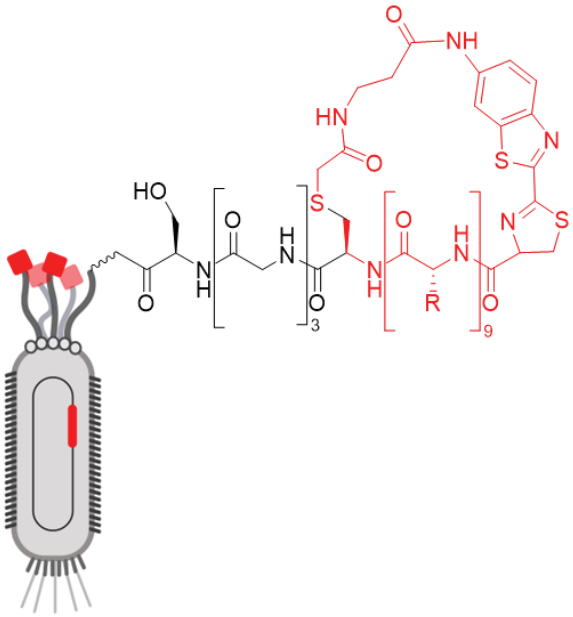


Figure 2-4. Cyanobenzothiazole-chloroacetamide (CBT-CA) structure and subsequent chemically modified luciferin cyclized phage construct structure. Created with BioRender.com.

Phage panning was done against EGFR ECD, and sequencing results are shown in **Table 2-4**. Again, while repeating sequences were observed, there was a lack of convergence as the rounds progressed.

Table 2-4. Input, output, output ratio, and Sanger sequencing results of rounds 2-4 of the screening with CX_9C-Luciferin cyclized library against EGFR-ECD. Repeating sequences are color-coded.

	CX_9C-Luciferin Cyclized Library		
	Round 2	**Round 3**	**Round 4**
Input (PFU)	$1.2x10^{11}$	$7.2x10^{10}$	$2.4x10^{10}$
Output (PFU)	$2.4x10^{7}$	$8.1x10^{7}$	$3.3x10^{7}$
Out/In (10^{-6})	200	1130	1380
	CSGMLPAQRTC (3)	CSGMLPAQRTC (5)	CSGMLPAQRTC (5)
	CERATLSNVAC (2)	CERATLSNVAC (2)	CERATLSNVAC (2)
	CRPTGSAPKKC	CRPTGSAPKKC (2)	CRPTGSAPKKC (2)
	CHKAVGALMEC	CMQGPTSGYLC	CDLFGGMAPHC (2)
	CPYPDSRSHSC	CDGFNNTDRGC	CPKPNGSSSPC
	CQGLRSTYPTC	CSLATRDVNTC	
	CARHAGSELTC		
	CNSTSAMYLTC		
	CTLDTYPSPSC		

To verify that these sequences were biased, discrete stocks of the CSGMLPAQRTC and CDLFGGMAPHC phage were made and phage retention was conducted. The output ratio was less than our template phage, indicating that these sequences were indeed biased (results not shown).

2.2.3 Panning Protocol Optimization

After several failed attempts to identify a high affinity binding peptide for EGFR ECD using different libraries, we decided to reevaluate the display protocol to ensure that was not the issue. Several parameters were assessed including: 1. BSA vs dry non-fat milk as a blocking buffer, 2. Increased concentrations of BSA as a blocking buffer, 3. Increased amount of Tween in wash buffer, 4. Use of urea during washing to remove non-specific binders, 5. Increasing the number of washes, and 6. Adding Tween-20 to the incubation solution.

Of all the parameters tested, only increasing the number of washes and adding Tween-20 to the incubation solution improved non-specific binding and decreased background noise. Phage retention with M13KE phage containing no peptide insert was used. Results are provided in **Table 2-5**.

Table 2-5. Output ratio and fold decrease of the phage retention using optimized conditions. All retentions were conducted with M13KE phage virions displaying no peptide on the pIII coat protein to improve non-specific binding. The smaller the output ratio, the better = less non-specific binding.

		Output Ratio (10^{-6})	Fold Decrease
	5x	429	NA
	5x-Tween	1270	
Trial 1	**10x**	93	132
	10x-Tween	0.705	
Trial 2	**10x**	1200	58.3
	10x-Tween	20.6	
Trial 3	**10x**	1230	67.2
	10x-Tween	18.3	

Once the protocol was optimized, we conducted phage retention assays with discrete phage stocks, PQN and AVRGD (pulled out from live-cell A431 screening), and the CX_9C library to verify that this new protocol works on discrete phage stocks and libraries, **Table 2-6**, which we show that it does.

Table 2-6. Output ratio and fold decrease of the phage retention with discrete and library stocks using optimized conditions.

	Output Ratio (10^{-6})	Fold Decrease
10x PQN	353	192
10x-Tween PQN	1.84	
10x AVRGD	1670	118
10x-Tween AVRGD	14.1	
10x CX_9C Library	1420	2.99
10x-Tween CX_9C Library	475	

2.2.4 Optimized CX_9C-Luciferin Cyclized and CX_7C Libraries

Phage panning using the luciferin-cyclized CX_9C was done again using the optimized protocol. In addition, we also decided to pan using a smaller sized library, namely the Ph.D. CX_7C Phage Display Library from NEB (Lot: 10081106). Results from those pannings are provided in **Table 2-7**. Again, no sequence convergence was noted, indicating no high affinity binding ligands for EGFR ECD.

Table 2-7. Input, output, output ratio, and Sanger sequencing results of rounds 2-4 of the screenings with CX_9C-Luciferin cyclized and 12-mer libraries against EGFR-ECD using the optimized protocol. Repeating sequences are color-coded. Sequences colored blue were determined to be biased based on previous assays conducted in our lab.

	CX_9C Luciferin Cyclized Library			**CX_7C**		
	Round 2	**Round 3**	**Round 4**	**Round 2**	**Round 3**	**Round 4**
Input (PFU)	$2.5x10^{11}$	$6.3x10^{11}$	$2.6x10^{11}$	$6.5x10^{10}$	$8.0x10^{10}$	$4.1x10^{10}$
Output (PFU)	$6.2x10^{6}$	$2.7x10^{8}$	$5.0x10^{5}$	$6.4x10^{7}$	$6.2x10^{6}$	$4.7x10^{5}$
Out/In (10^{-6})	24.8	429	1.92	985	77.5	11.5
	CTQDSKTPTKC	CQTYNQSSSMC (2)	CPSSRHSQTEC (2)	CDHAPPLSC	CPFSPMSHC (2)	CTPRSANYC
	CSPAGQTYPLC	CFMSGSREMNC	CLITQNKTTDC	CSRSMDSTC	CSGLHDRSC	CNDTISMKC
	CTQKTLPHSEC	CKSGRSYNTIC	CSSVSAYHPSC	CPGLTPERC	CEPRSLANC	CSSNTVPAC
	CHNWNGSHIVC	CPTNPSKTETC	CDLLNHFNEKC	CTIERAKVC	CHPTLKYRC	CMWLAAATC
	CSQLPSGKWTC	CTSFVSFEHKC	CYTNTYMQNTC	CTDYTNKSC	CGEGEADVC	CDASKIFIC
	CKASTVTNVTC	CRSADSNKHMC	CDVSPSVAKSC	CHHLMYHLC		CTDKASSSC (2)
	CYHKHDSDLVC	CPQTSTHHKTC	CDKIHSRLGNC	CTVRDMVGC		CGEGEADVC
	CTSDAMHRTPC	CDKIHSRLGNC		CTPRSANYC		
				CDGRPDRAC		

It appears that despite our best efforts using non-covalent phage libraries to identify a viable ligand, screening against EGFR ECD has proven to be more difficult than anticipated. We hypothesize that non-covalent phage libraries do not possess any sequences with a high enough binding affinity for this protein, and that the use of covalent libraries will be needed to pull out a positive hit.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 General Methods

Chemicals and reagents were commercially purchased from Fisher Scientific and Sigma-Aldrich. Fmoc-protected amino acids were purchased from Chem-Impex. Cloning primers were purchased from IDT. Restriction enzymes and buffers, ligation buffers, and Ph.D. phage libraries were purchased from NEB. Biotinylated EGFR-ECD was purchased from Acro Biosystems. Sanger sequencing was done at Azenta.

Absorbances were taken on a Nanodrop 2000 Spectrometer. Ligation reactions were done with T4 DNA ligase in its respective buffer in a MiniAMP Plus Thermal Cycler, Thermo Fisher. Electrocompetent cells were prepared using the Beckman Coulter Avanti J.E. centrifuge. Electroporations were done on Bio-Rad Gene Pulser: 25 μFD capacitance, 200 Ω resistance, and 1.8 kVolts. Streptavidin agarose beads (Thermo Scientific, Lot: WC318678) were imaged on a Zeiss Observer AI microscope at 10x magnification on this FITC setting; the UV lamp, X-Cite Series 120 Q was on the lowest intensity. Phage retention and pannings were done using Dynabeads M-280 Streptavidin (Thermo Fisher, 10mg/mL).

Peptides were synthesized on a Tribute Peptide Synthesizer from Protein Technologies Inc. Peptides were purified on a Water HPLC (1525 Binary Pump, and 2489 UV/Vis Detector) with a Jupiter 10 μm C18 300 Å, 250x10 mm column, followed by 1260 Agilent Infinity Series HPLC/6230 Agilent TOF Mass Spectrometer with an Infinity Lab Poroshell 120 EC-C18, 3.0x50mm, 2.7-micron column.

2.3.2 New England Biolabs Cloning Protocol

Phage cloning was done following the New England Biolabs Cloning Protocol,[1] with modifications. An overview of the protocol is illustrated in **Figure 2-5**.

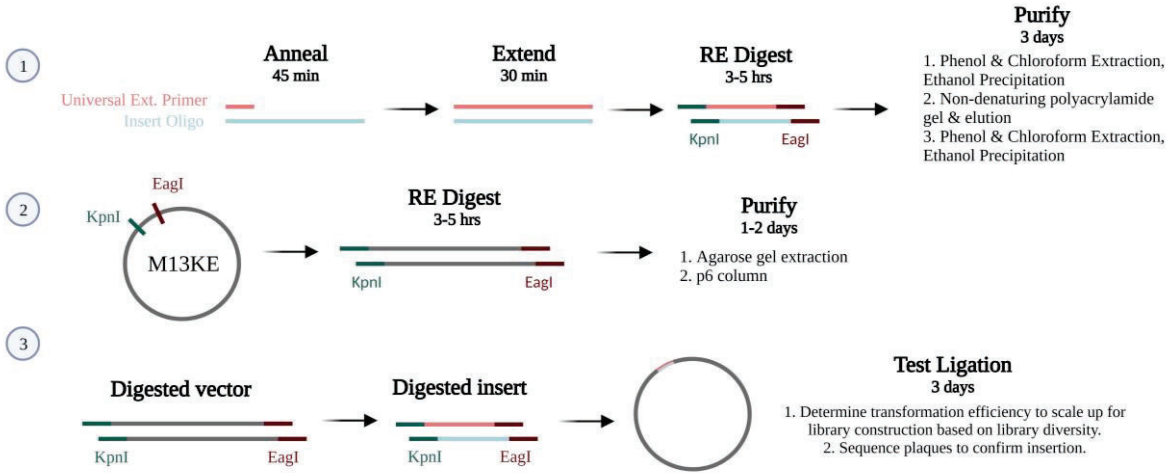


Figure 2-5. A visual representation of the NEB cloning protocol, taking into account times for each step. Created with BioRender.com.

The following oligomer inserts (~5µg), **Table 2-8**, were annealed to 3 molar equivalents of the universal extension primer in Tris-EDTA with 100 mM NaCl. The reaction was heated to and held at 95°C for 5 minutes and decreased by 5°C every 2 minutes until 30°C, and held idle at 4°C. The annealed duplex was extended following the protocol before being digested via KpnI-HF (20 Units/µL) and EagI-HF (20 Units/µL) restriction enzymes for 3.5 hours at 37°C. The inserts were purified via phenol/chloroform extraction, chloroform extraction and ethanol precipitation (**Section 2.3.3**) before being gel-purified on an 8% non-denaturing polyacrylamide gel. The gel was visualized via ethidium bromide staining and the band was excised, eluted, and purified following the protocol. The insert DNA was resuspended in Tris-EDTA and quantitated spectrophotometrically.

Table 2-8. Oligomer sequences used for the literature reported positive controls, and the universal extension primer.

Sequence	Oligomer 5' - 3' (Sequence insert in lowercase)
ASYPIPDT	CATGTTTCGGCCGAACCTCCACCggtatccggaatcggatagctcgcAGAGT GAGAATAGAAAGGTACCCGGG
AHTSDQTN	CATGTTTCGGCCGAACCTCCACCgttggtctgatcgctggtatgcgcAGAGTG AGAATAGAAAGGTACCCGGG
AYHWYGYTPQNVI	CATGTTTCGGCCGAACCTCCACCaatcacgttctgcggggtatagccataccaatga tacgcAGAGTGAGAATAGAAAGGTACCCGGG
AYHWYGYTPENVI	CATGTTTCGGCCGAACCTCCACCaatcacgttttccggggtatagccataccaatgat acgcAGAGTGAGAATAGAAAGGTACCCGGG
Universal Extension Primer	5′-CATGCCCGGGTACCTTTCTATTCTC-3′

The M13KE vector, provided by NEB (#N3541S), was digested using the same restriction enzymes as the insert for 4.5 hours at 37°C in a total volume of 400µL. The vector was gel purified on a 1% agarose gel and the Monarch DNA Gel Extraction Kit, NEB #T1020, and quantitated spectrophotometrically. An overnight ligation (100ng) at 10:1 insert:vector at 16°C and subsequent heat inactivation was completed before a 5ng electroporation into electrocompetent TOP10 cells (**Section 2.3.4**).

The cells were allowed to outgrow following the protocol for one hour before being plated/titered on IPTG/XGal plates. In summary, 10µL of cells diluted in LB were added to 200µL mid-log ER2738 *E. coli* and mixed by pipetting up and down; 790µL of heated top agar was added to the tube, mixed briefly and spread on IPTG/XGal plates. The plates were incubated at 37°C overnight and blue plaques were counted and sent out for Sanger sequencing the following day.

2.3.3 Phenol/Chloroform Extraction & Ethanol Precipitation

Phenol/Chloroform extraction and ethanol precipitation was done following the protocol provided on the ThermoFisher Scientific website with modifications.[6] In summary, the sample was divided into 200µL aliquots and one volume (200µL) of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The sample was shaken vigorously and centrifuged at room temperature for 2 minutes at 14,000 rpm. The upper aqueous layer was carefully transferred to a new Eppendorf tube, and the phenol/chloroform extraction was repeated.

Following the two phenol/chloroform extractions, the following reagents were added in order to the final aqueous layer:

Reagent	Volume
Glycogen (20µg/µL)	1 µL
7.5M NH_4OAc	0.5 x volume of sample
100% ethanol (chilled)	2.5 x (volume of sample + NH_4OAc)

The tube was placed at -80°C for 1 hour to precipitate the DNA from the sample, and then centrifuged at 4°C for 30 minutes at 14,000 rpm to pellet the DNA. The supernatant was carefully removed, and the DNA pellet was washed with 150µL of 70% ethanol. The sample was centrifuged at 4°C for 2 minutes at 14,000 rpm, the supernatant was removed, and the DNA was washed a second time. Following the final removal of the supernatant, the Eppendorf tube was left open on the benchtop covered with a Kimwipe to allow all ethanol to evaporate. The DNA pellet was then resuspended in Tris-EDTA buffer and quantitated spectrophotometrically.

2.3.4 Electrocompetent Cell Preparation

Solutions: 1L SOB media (20g Tryptone, 5g Yeast extract, and 0.5g NaCl) in a 4L Erlenmeyer flask and 1L of 15% glycerol, and JA-10 centrifuge tubes were prepared and autoclaved the day before. A 10mL overnight bacterial culture was also prepared the day before.

The 10mL overnight culture was inoculated into the 1L SOB media and allowed to grow at 37°C on a shaker set at 250 rpm for 2-3 hours until the OD_{600} is between 0.5-0.7. Once at the desired OD_{600}, the culture was divided into three pre-chilled/sterile JA-10 tubes and kept on ice for 15 minutes. The centrifuge was pre-chilled at 4°C and the tubes were centrifuged at 4°C for 15 minutes at 6500 rpm. The supernatant was decanted off and the cells were resuspended with 10mL 15% glycerol by pipetting up and down. The resuspended cells were transferred to the subsequent tubes and those cells were resuspended until combined. The cells were washed with ~300mL 15% glycerol and centrifuged again with the same conditions. The cells were washed for a total of three times. Following the last wash, the supernatant was almost completely decanted off, leaving only ~2mL remaining. The cell pellet was resuspended in the remaining solution and divided into 50µL aliquots before being flash frozen in liquid nitrogen. The aliquots were labeled accordingly and stored at -80°C until use.

2.3.5 Electroporation

A sterile electroporation cuvette (0.1cm gap width), DNA sample, and electrocompetent cell aliquot were chilled and kept on ice until use. The DNA sample was mixed well by

pipetting up and down, and 1µL was added between the plates in the cuvette. The 50µL electrocompetent cell aliquot was then added, and the cuvette was tapped on the benchtop to remove any air pockets. The cuvette was inserted into the slider and the sample was electroporated using the conditions listed in **Section 2.3.1**. Prewarmed LB (250µL) was immediately added, the sample was mixed by gently pipetting up and down, and transferred to a culture tube which was allowed to outgrow for 1 hour before plating.

2.3.6 DNA Extraction and Purification

DNA extraction and purification was done following the Thermo Scientific GeneJET Plasmid Miniprep Kit with modification.[7] In summary, a single plaque was picked off of an IPTG/XGal plate and added to 5mL LB in a culture tube. The culture is allowed to grow for 3-4.5 hours at 37°C before being harvested by centrifugation at 5,000 rpm for 5 minutes at room temperature. The supernatant can either be collected to create phage stocks, or is discarded into an appropriate waste container.

The pelleted cells were resuspended completely in 250µL of the Resuspension Solution and transferred to a microcentrifuge tube. An equal volume (250µL) of the Lysis Solution was added and was mixed thoroughly by inverting the tube 10-12 times, ensuring the sample was not incubated for more than 5 minutes to avoid denaturation. Neutralization Solution (350µL) was added and was mixed thoroughly and immediately by inverting the tube 10-12 times before centrifugation for at room temperature for 5 minutes at 14,000 rpm.

The supernatant was then transferred to the provided GeneJET spin column, and centrifuged for 1 minute (room temperature, 14,000 rpm). The flow-through was discarded

before washing twice by adding 400µL of the Wash Solution, centrifuging for 30 seconds, and discarding the supernatant. The column was spun again for 1 minute to remove any residual Wash Solution and was transferred to a 1.5mL microcentrifuge tube. 50µL of Elution Buffer was added, allowed to incubate in the membrane for 2 minutes at room temperature before centrifuging for 2 minutes. The purified plasmid DNA was then quantitated spectrophotometrically and sent out for Sanger sequencing.

2.3.7 Agarose Bead Microscopy

Pierce™ Streptavidin Agarose Resin (#20349) with a capacity of up to 3mg protein/mL resin was used. An aliquot of agarose beads (20µL) was washed with 480µL PBS. The beads were centrifuged at room-temperature for 2 minutes at 7,000 rpm, and washed two more times in 500µL PBS. Following the last centrifugation, the beads were resuspended in 500µL PBS. Each sample will include 20µL of washed beads, which after taking into consideration the dilutions, results in a binding capacity of 2.4µg protein/20µL washed beads.

In 1.5mL microcentrifuge tubes, 20µL washed beads, 2.5ug of EGFR ECD or NHS-biotin, and 200nM EGF-FAM (Invitrogen) were added in a final volume of 200µL PBS. For the positive control, 20µL washed beads and 200nM biotinylated FAM-peptide construct were added in a final volume of 200µL. All samples were placed on a rotator for 30 minutes at room-temperature protected from light, and then imaged on the microscope.

2.3.8 Fmoc-based Solid Phase Peptide Synthesis and Characterization

We synthesized FITC-labeled PQN and PEN peptides following the protocol reported by Hossein-Nejad-Ariani et al. in order to reproduce their results.[2] In summary, PQN and PEN were synthesized via Fmoc-solid phase peptide synthesis (Fmoc-SPPS) on preloaded Fmoc-isoleucine Wang resin (0.05 mmol scale) using automated peptide synthesizer (Tribute, Protein Technology Inc.). Preloaded Fmoc-isoleucine Wang resin (0.05 mmol) was added to the plastics reaction vessel and the resin was allowed to swell in DMF under nitrogen with mechanical shaking. All amino acids (5 equiv**)** were coupled for 1 hour in sequence using HBTU (4.75 equiv) for each coupling. Fmoc was removed by 20% piperidine in DMF. β-alanine was added at the N-terminal of each peptide sequence as a spacer before FITC coupling. FITC (0.15 mmol) in DMF (1.5 mL) with DIPEA (0.075 mmol) was added to the resin, and the mixture was incubated in the dark for 20 h. The peptide was cleaved from the resin using the cleavage cocktail (1 mL) of trifluoroacetic acid (TFA)/triisoproylsilane/ ultra-pure water (95:2.5:2.5) for 2 hours. Cold diethyl ether (Et_2O, 5 mL) was added to the filtered resin followed by centrifugation for 10 min in order to precipitate and collect the crude peptide. The cleaved peptides were purified and characterized using HPLC-UV and LCMS-QTOF.

2.3.9 Phage Display/Retention

A visual representation of the phage display and retention (only one round) panning is illustrated in **Figure 2-6**.

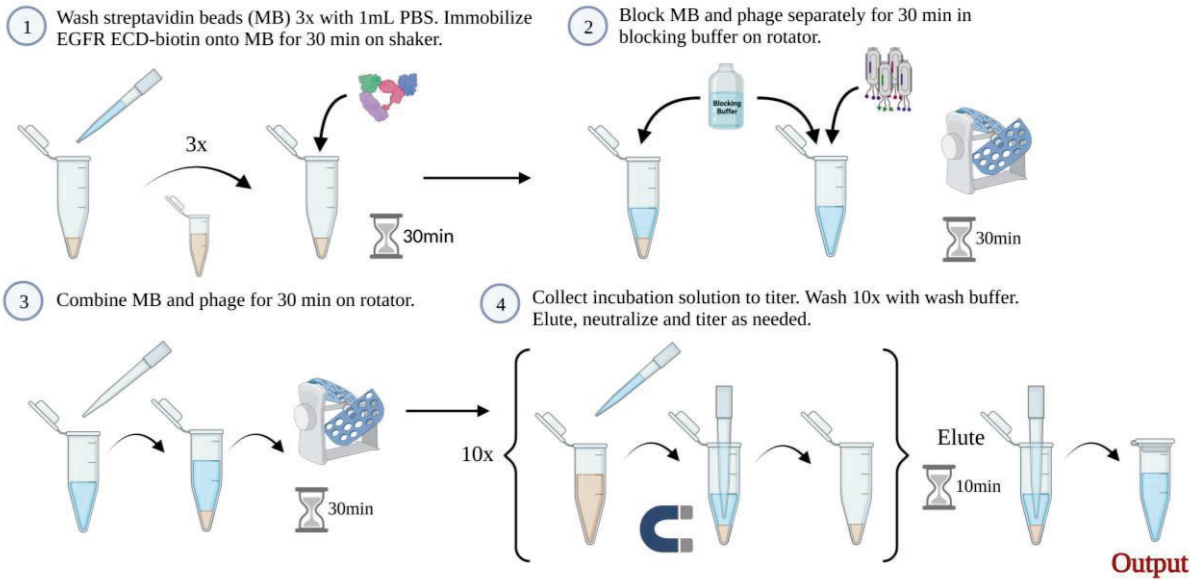


Figure 2-6. A visual representation of the phage display screening protocol against immobilized protein on streptavidin coated magnetic beads. Created with BioRender.com.

Dynabeads M-280 Streptavidin magnetic beads (#11205D) were vortexed briefly to mix and 10µL was aliquoted to a microcentrifuge tube. The beads were washed 3x by adding 1mL PBS, vortexing briefly, placing the tube on the magnetic rack and allowing the beads to congregate to the side of the tube, and removing the supernatant. Biotinylated EGFR ECD (0.35µg/10µL beads, Acro Biosystems) was added in a total volume of 100µL PBS and placed on a shaker for 30 minutes at room temperature to immobilize the protein. The beads were washed 3x with PBS to remove any excess protein. The protein and phage (~10^{10}) pfu were blocked separately with 500µL of 1mg/mL BSA in 0.1% PBST on a rotator for 30 minutes at room temperature to minimize non-specific binding.

Following blocking, the beads were mixed with the phage and allowed to incubate on the rotator for 30 minutes at room temperature. The incubation solution/supernatant was collected for titering (input). The beads were washed ten times by pipetting up and down with 0.1% PBST to remove unbound phage, and the bound phage were eluted for <10

minutes with 200µL 0.2M glycine-HCl pH 2.2 buffer before the supernatant was transferred to a new tube. The eluted phage (output) were neutralized with 800µL 1M Tris pH 9 buffer and mixed well. The input and output fractions were then titered to approximate the concentration of phage (**Section 2.3.2**) and/or amplified for subsequent rounds.

2.3.9.1 Phage Amplification

A 500µL aliquot of output phage was added to a sterile 250mL Erlenmeyer flask containing 20mL LB and 1% ER2738 *E. coli* from an overnight culture containing 0.1% Tetracycline. The cells were allowed to grow for up to 4.5 hours at 37°C, 250 rpm. The cells were harvested by centrifugation and the supernatant was collected in a 50mL Falcon tube. Approximately 5mL (~1/6 volume) of sterile PEG/NaCl was added and the phage were allowed to precipitate overnight at 4°C.

The following day, the phage were pelleted by centrifuging at 4°C for 20 minutes at 7830rpm. The supernatant was decanted, and the phage were gently resuspended in 1mL PBS and transferred to a microcentrifuge tube. The phage were reprecipitated by adding 200µL PEG/NaCl and left on ice for 1 hour. Following reprecipitation, the phage were pelleted again by centrifuging at 4°C for 20 minutes at 14,000rpm. The supernatant was removed and the phage were resuspended in 50-200µL PBS depending on the size of the pellet. The amplified phage was then titered and used as the input in subsequent rounds of panning.

2.3.10 Lucifierin Cyclization

Chemical modification was conducted on our lab's CX_9C phage library. On the N'-terminal of the pIII protein, in order, there is an HA-tag, a Factor Xa cleavage site, the randomized amino acids CX_9C, a GGGS spacer, followed by the rest of the phage virion. Chemical modification was done in three steps: 1. Factor Xa cleavage, 2. TCEP Reduction, and 3. CBT-CA cyclization.

For Factor Xa cleavage, ~10^{11} pfu phage was suspended in a total volume of 100µL cleavage buffer (20mM Tris, 50mM NaCl, 1mM $CaCl_2$ in sterile water), with 1µL of Factor Xa protease (NEB, #P8010S). The reaction was allowed to mix at room temperature for 6 hours on a benchtop shaker. The phage was precipitated on ice for 1 hour by adding 20µL PEG/NaCl. The phage was centrifuged at 14,000 rpm for 20 minutes at 4°C, and the supernatant was pipetted off.

The phage was then resuspended in 98µL of reaction buffer (20mM NH_4HCO_3, 5mM EDTA, pH 8.0 in sterile water) for TCEP reduction. To the sample, 2µL of 50mM TCEP (#77720) was added for a final concentration of 1mM. The reaction was allowed to mix at room temperature, protected by light, for 1 hour on a benchtop shaker. The phage was then precipitated as stated above.

Next, the phage pellet was resuspended in 99µL PBS, pH 7.4, with 1µL of 10mM CBT-CA (in DMF) for a final concentration of 100µM. The reaction was allowed to mix at room temperature for 3 hours on a benchtop shaker, and the phage was precipitated as stated above. The final modified phage was resuspended in 100µL PBS, with a final concentration of ~10^{10} pfu. Panning against the protein could then be conducted, followed

by amplification of the output. This chemical modification was repeated before each round after the amplification step.

2.4 REFERENCES

(1) Instruction Manual, Ph.D. Phage Display Libraries, 2020.
(2) Hossein-Nejad-Ariani, H.; Althagafi, E.; Kaur, K. *Sci Rep* **2019**, *9* (1), 2723.
(3) Li, Z.; Zhao, R.; Wu, X.; Sun, Y.; Yao, M.; Li, J.; Xu, Y.; Gu, J. *FASEB J.* **2005**, *19* (14), 1978–1985.
(4) Ogiso, H.; Ishitani, R.; Nureki, O.; Fukai, S.; Yamanaka, M.; Kim, J.-H.; Saito, K.; Sakamoto, A.; Inoue, M.; Shirouzu, M.; Yokoyama, S. *Cell.* **2002**, *110*, 775-787.
(5) Shao, Q.; Zhu, W. *Phys. Chem. Chem. Phys.* **2019**, *21* (15), 8141–8151.
(6) *How to Use Phenol/Chloroform for DNA Purification*. https://www.thermofisher.com.
(7) *Thermo Scientific GeneJET Plasmid Miniprep Kit*. https://www.thermofisher.com.

3.0 CHAPTER 3: PEPTIDE PHAGE LIBRARY CONSTRUCTION SUITABLE FOR COVALENT MODIFICATION

3.1 ACX_7C LIBRARY CONSTRUCT

Creating our own libraries offers many advantages; first, we can design the peptide insert to suit the needs of our project, such as introducing protease cleavage sites, bicyclic constructs and varied ring sizes. Second, depending on the number of randomized amino acids, there would be up to 10^{10} diverse sequence combinations for screening. Third, we can design our peptides to either be displayed on the major coat pVIII protein, or minor coat pIII protein. With only five copies, the pIII protein offers advantages over its 2,700-copy major coat counterpart, including increased binder potency and less steric bulk between displayed peptides.[1] However, generating said libraries is difficult and poses many challenges.

Before progressing to more novel library constructs, we began our library construction with a ACX_7C library to become proficient in this skill. The cloning protocol described in **Section 2.3.2** was used with the oligomer insert, 5' – CATGTTTCGGCCGAACCTCCACCACAMNNMNNMNNMNNMNNMNNMNNAC AAGAGTGAGAATAGAAAGGTACCCGGGCATG, where N = A, C, T or G, and M = A or C. The maximum library diversity for displayed amino acids is $1.28x10^9$ pfu.

Our first attempt at test ligations yielded a transformation efficiency of $1.50x10^6$ pfu/μg DNA. This would mean that we would need approximately 850 μg of DNA to cover the maximum diversity of the library. In addition, eight of the twelve plaques sent out for

sequencing had no insert present, indicating that the vector was re-ligating. Despite our best efforts, our transformation efficiency and library diversity were not suitable for pannings. As a result, significant troubleshooting was needed to improve both parameters.

3.1.1 Troubleshooting

The first issue we tackled was the re-ligation of the vector. We suspected that this was due to incomplete digestion from the restriction enzymes. We originally digested the vector with EagI-HF and KpnI-HF at the same time for 3.5 hours at 37°C. Multiple variables were examined: discrete vs. parallel digestion and 1x vs. 2x restriction enzymes. With that, we digested the midi-prepped M13KE vector with four different conditions: 1. 1x enzymes, digest with KpnI-HF first, then EagI-HF, 2. 1x enzymes, digest with enzymes at the same time, 3. 2x enzymes, digest with KpnI-HF first, then EagI-HF, and 4. 2x enzymes, digest with enzymes at the same time. All digestions lasted 5 hours at 37°C, followed by heat inactivation at 65°C for 20 minutes only after EagI-HF. In addition, due to poor DNA purity, we purified each condition with a p6-column (BioRad, #732-6221). Yields are shown in **Table 3-1**.

Table 3-1. Vector DNA concentration, purity, and yield post p6-column for the four restriction enzyme conditions.

		Post-p6 Column			% yield
	Sample	**[X] ng/uL**	**260/280**	**260/230**	
1	**1x K then E**	4.2	1.7	-1.79	13.1
2	**1x K+E**	8.9	1.81	-39.05	22.3
3	**2x K then E**	10.5	1.94	4.73	23.6
4	**2x K+E**	14.5	1.83	2.13	26.1

Test ligations (50ng) were done with a 10:1 insert:vector ratio, followed by a 2.5ng electroporation into Top10 cells. Transformation efficiency and sequence results are shown in **Table 3-2**. Conditions 1-3 had an improved TE, however, no library insert was present in any of the plaques sequenced, indicating re-ligation was still occurring.

Table 3-2. Test ligation results for the four restriction enzyme conditions. No condition resulted in proper insert ligation.

	Transformation Efficiency (pfu/ug)	Diversity (pfu)	Ligation required (µg)	Sequence Results
1	1.32x10^8	$1.28x10^9$	9.7	Template (10)
2	4.08x10^7		31.4	Template (10)
3	9.00x10^7		14.2	Template (10)
4	2.4x10^5		5330	NA

Next, we decided to incubate our digested vectors with Quick CIP, a phosphatase that removes the 5'- and 3'- phosphates of DNA thereby inhibiting re-ligation. Due to the amount of vector generated from each condition, Quick CIP was only added to conditions 3 and 4. In a PCR tube, 250ng vector, 4 uL 10x rCutSmart Buffer, 1 uL of 20x QuickCIP enzyme, and up to 40 uL diH_2O was added; the reaction was incubated at 37°C for 10 minutes, followed by heat inactivation at 80°C for 2 minutes. A 50ng test ligation was done with no insert present before running on a 1% agarose gel, **Figure 3-1**. We observed that conditions 5 and 6 do not re-ligate after the addition of QuickCIP, indicated by the travel distance compared to the digested vector control.

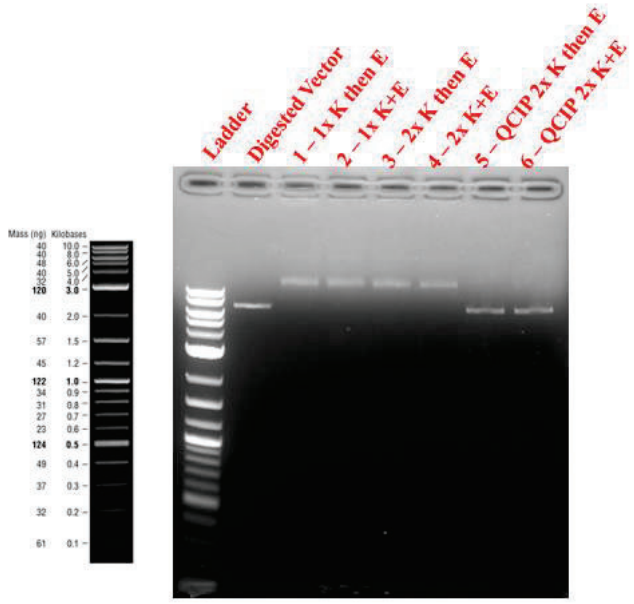


Figure 3-1. DNA post test ligation with no insert present, with and without a Quick-CIP (QCIP) digestion. Conditions 5 and 6 show no re-ligation as indicated by the distance traveled as compared to the control digested vector not subjected to ligation.

We also wanted to evaluate whether our midi-prepped vector decreased TE and insert ligation. M13KE vector from NEB was digested with the same conditions as our midi-prepped vector; both were digested with 20 units of enzyme in parallel per µg DNA for 5 hours at 37°C, followed by a PCR clean-up step, then QuickCIP digestion before a p6-column for DNA purification. Six ligation conditions (100ng) were tested to evaluate insert:vector ratios and re-ligation, listed in **Table 3-3**.

Table 3-3. Restriction enzyme digestion of midi-prepped vs. commercially available M13KE vector. The midi-prepped vector with a ligation condition of 10:1 insert:vector (I:V) resulted in the greatest TE.

Sample	Description	Transformation Efficiency (pfu/µg)	Diversity (pfu)	Ligation required (µg)
1	Midi 5:1 I:V	7.02×10^6	1.28×10^9	182
2	Midi 10:1 I:V	1.19×10^9		1.08
3	Midi, No Insert	NA		NA
4	NEB 5:1 I:V	8.16×10^8		1.57
5	NEB 10:1 I:V	6.90×10^8		1.86
6	NEB, No Insert	NA		NA

Conditions 1, 2, 4, and 5 were transformed into electrocompetent cells, while conditions 3 and 6 were run on an agarose gel to assess re-ligation, **Figure 3-2**.

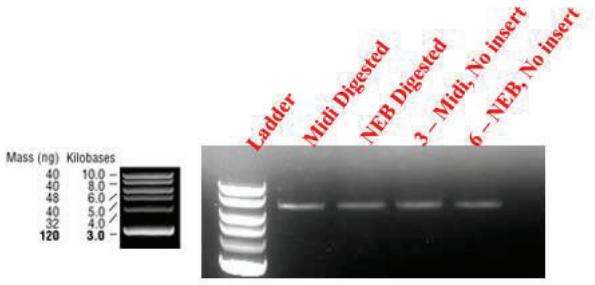


Figure 3-2. Conditions 3 and 6 post test ligations with no insert present. Distance traveled indicated no re-ligation as compared to the control digested vectors not subjected to ligation.

We can see from **Table 3-3** that condition 2 had the best transformation efficiency, requiring only 1.08 µg of vector DNA to cover the entire diversity of the library, a significant improvement from our initial >850 µg vector DNA. In addition, we selected 12 plaques from conditions 2 and 5 to send out for Sanger sequencing, and observed no blanks and no repeating sequences, **Table 3-4**. With this information in hand, we felt confident moving forward scaling up the library.

Table 3-4. Sanger sequencing results of conditions 2 and 5 indicating no repeating or blank sequences.

Condition 2	Condition 5
CPLGKFTRC	CDLFHNSTC
CPPRSNPVC	CSQFQLEAC
CGFMPFSDC	CDSTLSGAC
CPTLSFDPC	CSLMLSWFC
CHPLYNNFC	CMSNHTLDC
CPMPGKLSC	CDPDFTNMC
CGAANELVC	CSWHYTRAC
CHHNHDVPC	CSDKEVVTC
CADKQVMVC	CYTSLSASC
CTTGPLWAC	CYLEDSTMC
CSKNVTWHC	CDFGHPTAC
CDWNTDRVC	CSWDKRSAC

3.1.2 Protocol Optimization

Due to limited digested vector from troubleshooting, we conducted a large-scale vector digestion using the same conditions from before; 120 μg of midi-prepped M13KE vector was digested in six aliquots with 20 units of enzyme in parallel per μg DNA for 5 hours at 37°C, followed by a PCR clean-up, QuickCIP digestion before a p6-column for DNA purification. With a combined total volume of 350 uL at 15.2 ng/uL, we had a final yield of 5.32 ug = 4.52% yield. While this is not ideal, a low yield after vector digestion is a common bottleneck for library construction. Before setting up a large-scale ligation, we did a test ligation (50ng) of the new digested vector, with a 10:1 insert:vector based on results during troubleshooting. Our TE was $5.4x10^9$ pfu/μg DNA, resulting in only 0.237 μg vector DNA needed to achieve maximum library diversity. In addition, we sent 20 plaques out for sequencing, and results are shown in **Table 3-5**.

Table 3-5. Sanger sequencing results of the small-scale ACX_7C library test ligation.

Test Ligation	
CDMPQNNRC (5)	CHAQPTLCC
CRGSFMPGC (2)	CAGAYRGPC
CDRVNGSTC	CDQPNLFGC
CTLTPENIC	CGVNGQQPC
CNLQAAYQC	CTLPWTNKC
CVTATSNPC	CTWGTGMSC
CAPRVPALC	CQEKSRPWC
CTSISRPSC	

Despite repeat sequences, we opted to set up the large-scale ligation. It was noted that these sequences may be biased in future assays. Per the protocol,[1] the large-scale ligation needed to be purified via phenol/chloroform extraction, chloroform extraction and ethanol precipitation before transforming into cells. We expected some sample loss from this step, so we set up a 500ng ligation, with a 10:1 insert:vector into two aliquots in a total volume of 50uL each. The combined purified ligation was resuspended in 10µL of Tris-EDTA for 10, 1µL transformations. We pooled the outgrowths in groups of five, allowed them to outgrow, aliquoted 10µL from each to determine diversity and ligation insertion, before continuing on with the remainder of the protocol. Maximum diversity was achieved and sequencing results for the two pools after ~45 minutes of outgrowth are shown in **Table 3-6**.

Table 3-6. Sanger sequencing results of the large-scale ACX_7C library pools pre-amplification.

Pool 1		Pool 2	
ACILQSPAYC (2)	ACPEVSSLIC	ACPAGLFCLC (2)	ACKILNTISC
ACGAGEDNRC	ACYPLTAGSC	ACSSPVDNLC (2)	ACTLRTWQAC
ACSLYAPMEC	ACPNIDSYMC	ACVGITKTAC	ACKADMFRMC
ACESWQRNLC	ACTIREPGLC	ACEPGGALLC	ACDDLVINTC
ACDTYTYSWC	ACSFPDVNHC	ACDLFQTSMC	ACWHPHAQEC
ACNDAPTVLC	ACCANRLELC	ACRDAVLERC	ACRINTEVSC
ACASESQSAC	ACAVLSESSC	ACSAVALLPC	

The final library was resuspended in 20mL TBS and titered and sequenced. The library was re-titered for reproducibility, then precipitated and resuspended in a smaller volume (10mL) to accommodate an equal volume of glycerol for long-term storage. Each step was titered and sequenced, results in **Table 3-7**. We observed no repeating sequences throughout this process, nor from the test ligation and outgrowth pools. With a final ACX_7C library with a library titer of ~$1x10^{13}$ pfu/mL, we conducted library validation against streptavidin coated magnetic beads to pull out the HPQ motif, explained in **Section 3.3.1**.

Table 3-7. Sanger sequencing results of the large-scale ACX_7C library post-amplification and subsequent steps for long-term storage with glycerol.

Post Amplification and Titer		Re-titer	Post Precipitation	Final w/ Glycerol
ACTVGHFRVC	ACLNNDSMLC	ACNSDPSMVC	ACLLKILEHC	ACKPLEQLLC
ACARSTSGLC	ACHAALYPAC	ACSPLTFQRC	ACWTDRERSC	ACADSSFDTC
ACCKTFPNKC	ACSHAAKHFC	ACPAHFAHQC	ACLSGSSAHC	ACLSHITNKC
ACIPSMMRTC	ACMQPAQLLC	ACHNMFSATC	ACLDGRHLHC	ACVTTGHRVC
ACMPFPSRIC	ACTSTPIYIC	ACHTSTTRSC	ACHSVRQDKC	ACPRWEERNC
ACSDHTTEAC	ACVSSRLNDC	ACAVWVGLSC	ACTHSSLNNC	ACVSPSRDVC
ACTTVGAADC	ACDTPTKVEC	ACHQTAFLAC	ACLPGATNTC	ACQPLTQHHC
ACMGHHHYLC	ACVHKKMWSC	ACRLPGHLGC	ACEWTSVSVC	ACQDRQMFLC
ACTLLWMNGC	ACAENPVGLC	ACEQHALRAC	ACNNGWPMFC	ACAADANPEC
Blank	ACGLTPLHTC	ACEADDRFYC	ACELFSKADC	ACGPSTPKQC

3.2 ACX$_7$C-TEV LIBRARY CONSTRUCT

Following the success of the ACX$_7$C library, we started on our second library construct, ACX$_7$C with a TEV protease cleavage site on the C-terminal side of the randomized amino acids to be suitable for covalent warhead modifications, illustrated in **Figure 3-3**. TEV Protease is a highly specific cysteine protease with a recognition sequence with the highest catalytic efficiency is ENLYFQ^{+}S; however, the amino acid in the P1' position can also be G, A, M, C, or H.[2]

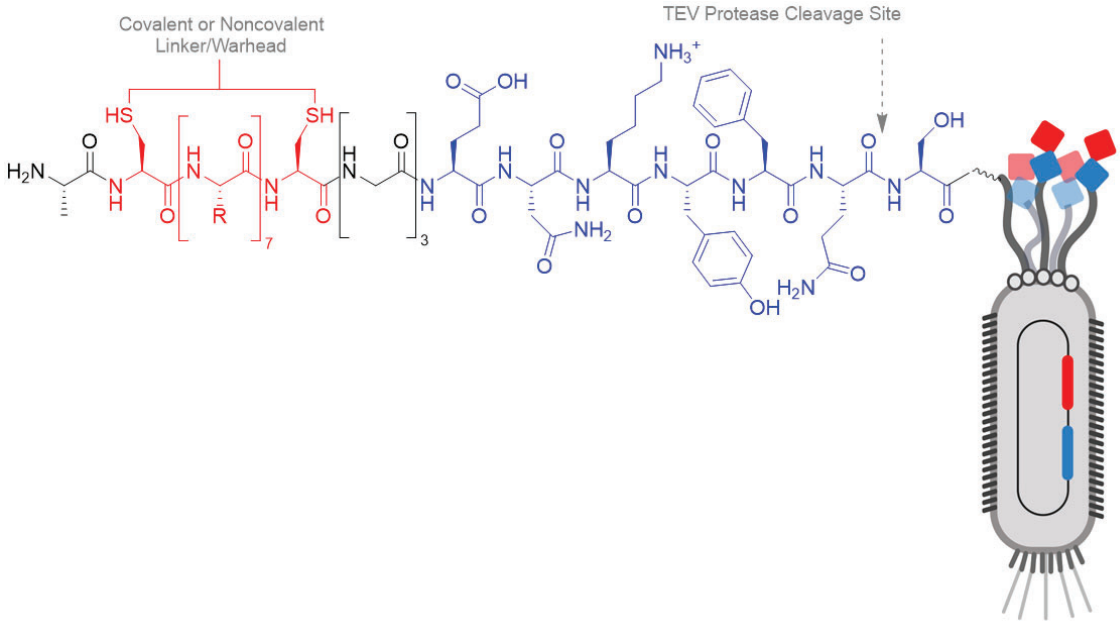


Figure 3-3. ACX$_7$C-TEV phage construct, highlighting key residues on the pIII coat protein. Created with BioRender.com.

The oligomer insert, 5'-CATGTTTCGGCCGACGACTGAAAATACAGGTTTT CACCTCCACCACAMNNMNNMNNMNNMNNMNNMNNACAAGCAGAGTGAGA ATAGAAAGGTACCCGGG-3', where N = A, C, T or G, and M = A or C, will display the amino acids ACXXXXXXXCGGGENLYFQS on the N-terminal of the pIII protein,

where X = a random amino acid. The maximum library diversity for displayed amino acids is $1.28x10^9$ pfu.

The insert was annealed, extended, and digested and two test ligations (50ng) with 10:1 insert:vector was done using the large-scale vector described in **Section 3.1.2**. The transformation efficiencies for both were ~$1.4x10^8$ pfu/µg DNA, which will require 9.1µg vector DNA to cover the maximum diversity of the library. Sequencing results showed proper insertion with some repeating sequences, **Table 3-8**. Since these ligations were relatively reproducible, we decided to scale-up the library construction.

Table 3-8. Sanger sequencing results of the 2 small-scale ACX_7C-TEV library test ligations. Repeating sequences are color-coded.

Small Scale Test Ligation #1	Small Scale Test Ligation #2	
ACTHNTSMTCGGGENLYFQS (3)	ACDNRKTQMCGGGENLYFQS (3)	ACAASPKVQCGGGENLYFQS
ACNTYRTAMCGGGENLYFQS	ACQNKSQSICGGGENLYFQS	ACDHAQWRECGGGENLYFQS
ACNNVNLPSCGGGENLYFQS	ACLNSKMDMCGGGENLYFQS	ACTFENTHSCGGGENLYFQS
ACETEPEDKYGGGENLYFQS	ACNRLSPWQCGGGENLYFQS	ACIPRQHNFCGGGENLYFQS
ACDPWHSGQCGGGENLYFQS	ACSMGNDARCGGGENLYFQS	ACQTEEMKWCGGGENLYFQS
ACNDEVSLRCGGGENLYFQS	ACREQIDVLCGGGENLYFQS	ACYNYANYPCGGGENLYFQS
ACTKFELLNCGGGENLYFQS	ACPPNIEHNCGGGENLYFQS	ACTNQLLRACGGGENLYFQS
ACEKQFVEWCGGGENLYFQS	ACPIQQESLCGGGENLYFQS	ACLHEGICGGGENLYFQS

Due to a limited quantity of digested vector, our large scale library ligation was done with the remaining 4µg vector DNA with 10:1 insert:vector. In a total volume of 350µL, 10.2µL insert (37.1ng/µL), 263µL vector (15.2ng/µL), 35µL of 10x ligation buffer, and 8.75µL T4 DNA ligase was combined and divided into seven 50µL aliquots for overnight ligation. The following day, the aliquots were combined and underwent phenol/chloroform extraction and ethanol precipitation.[3] The DNA was allowed to precipitate overnight at -20°C to improve yield, and then resuspended in a final volume of 15µL for 15, 1µL transformations. We pooled the outgrowths in groups of five, allowed

them to outgrow, aliquoted 10uL from each to determine diversity and ligation insertion, before continuing on with the remainder of the protocol. Approximate diversity of $1x10^8$ pfu was achieved, and sequencing results for the three pools after ~45 minutes of outgrowth are shown in **Table 3-9**.

Table 3-9. Sanger sequencing results of the large-scale ACX_7C-TEV library pools pre-amplification.

Pool 1	Pool 2	Pool 3
ACTFQMDFRCGGGENLYFQS	ACLTPPVYFCGGGENLYFQS (5)	ACPAPQDQLCGGGENLYFQS (3)
ACQLSLSQICGGGENLYFQS	ACSIVQXQGCGGGENLYFQS	ACWSIVENACGGGENLYFQS
ACEVKGHERCGGGENLYFQS	ACQSAWKPRCGGGENLYFQS	ACNSYKAIICGGGENLYFQS
ACGTSGNHMCGGGENLYFQS	ACSDPSSRCCGGGENLYFQS	ACMWNTQQPCGGGENLYFQS
ACSTENSDVCGGGENLYFQS	ACGDSLPYICGGGENLYFQS	ACTMKGYFMCGGGENLYFQS
ACDLRPRTDCGGGENLYFQS		ACAPSLAEPCGGGENLYFQS
ACSTENSDVCGGGENLYFQS		ACDKLYIPRCGGGENLYFQS
ACPGMIKYRCGGGENLYFQS		ACHTDWLFECGGGENLYFQS
ACGTSGNHMCGGGENLYFQS		

The library was resuspended in 20mL TBS, titered and sequenced. The library was precipitated and resuspended in a smaller volume (10mL) to accommodate an equal volume of glycerol for long-term storage. The final library was titered and sequenced, results in **Table 3-10**.

Table 3-10. Sanger sequencing results of the final large-scale ACX_7C-TEV library post-amplification already prepared to long-term storage with glycerol.

Final w/ Glycerol	
ACKHGTHKECGGGENLYFQS	ACINYRATTCGGGENLYFQS
ACHQSHAHMCGGGENLYFQS	ACSTSQDTECGGGENLYFQS
ACVMWGSNSCGGGENLYFQS	ACVIQMLRNCGGGEKPVFQS
ACKNFESLNCGGGENLYFQS	ACTSERSRLCGGGENLYFQS
ACKNSMPVDCGGGENLYFQS	ACLRTNMMQCGGGENLYFQS
ACRTTTAMNCGGGENLYFQS	ACTRFTNAFCGGGENLYFQS
ACFYNYKYDCGGGENLYFQS	ACKSSESNACGGGENLYFQS
ACNPKNNNLCGGGENLYFQS	ACTHSMRTLCGGGENLYFQS
ACPSLKESPCGGGENLYFQS	ACTFNTSGWCGGGENLYFQS
	Blank (7)

Similar to the ACX_7C library, we observed no repeating sequences that were originally present in the outgrowth pools. With a final ACX_7C-TEV library with a library titer of ~$2x10^{13}$ pfu/mL, we conducted library validation against streptavidin coated magnetic beads to pull out the HPQ motif, explained in **Section 3.3.2**.

3.3 LIBRARY VALIDATION

3.3.1 ACX_7C Library

Two trials were done in parallel to increase the chances of pulling out the motif and to save time. The HPQ motif was pulled out in Trial 1B, round 2, with 7/10 sequences being CHPQNNRFC. We successfully constructed and validated a ACX_7C phage library with a titer of ~$1x10^{13}$ pfu/mL in 20mL.

Table 3-11. Titer results for the validation of the ACX_7C library.

Trial	Round #	Amplification Plaque # 10^{11}	Amplification Plaque # 10^{12}	Amplification Average (PFU)	Input Plaque # 10^8	Input Plaque # 10^9	Input Average (PFU)	Output Plaque # 10^3	Output Plaque # 10^4	Output Average (PFU)	Output Ratio (10^{-6})
A	1	NA	NA	NA	147	NA	$1.5x10^{10}$	161	12	$1.3x10^5$	8.67
	2	422	53	$4.8x10^{13}$	NA	219	$2.2x10^{11}$	farm	~700	$\sim 7.0x10^6$	31.8
B	1	NA	NA	NA	210	NA	$2.1x10^{10}$	31	3	$3.1x10^4$	1.48
	2	480	57	$5.3x10^{13}$	NA	293	$2.9x10^{11}$	193	18	$1.9x10^5$	0.655

Table 3-12. Round 2 Sanger sequencing results for the validation of the ACX_7C library.

Trial 1AR2	Trial 1BR2
CLSIGGYSC (4)	CHPQNNRFC (7)
CSGTALLSC (3)	CLSIGGYSC (2)
CYPSMNRMC (2)	CSGLHQKLC
CSGLHQKLC	

3.3.2 ACX_7C-TEV Library

Similarly, two trials were done in parallel to increase the chances of pulling out the motif and to save time. The HPQ motif was pulled out in both trials in the second round, with 3/10 sequences in Trial 1A containing the HPQ motif, and 9/10 sequences in Trial 1B containing the HPQ motif. We successfully constructed and validated a ACX_7C phage library with a titer of $\sim 1x10^{13}$ pfu/mL in 20mL. Future efforts for determining the TEV protease cleavage efficiency will need to be done.

Table 3-13. Titer results for the validation of the ACX_7C-TEV library.

Trial	Round #	Amplification Plaque # 10^{11}	Amplification Plaque # 10^{12}	Amplification Average (PFU)	Input Plaque # 10^{9}	Input Average (PFU)	Output Plaque # 10^{3}	Output Plaque # 10^{4}	Output Average (PFU)	Output Ratio (10^{-6})
A	1	NA	NA	NA	50	$5.0x10^{10}$	>2000	321	$3.2x10^{6}$	64
	2	NA	20	$2.0x10^{13}$	81	$8.1x10^{10}$	570	67	$6.2x10^{5}$	7.65
B	1	NA	NA	NA	34	$3.4x10^{10}$	566	NA	$5.7x10^{5}$	16.8
	2	282	27	$2.8x10^{13}$	106	$1.1x10^{11}$	~700	64	$6.7x10^{5}$	6.10

Table 3-14. Round 2 Sanger sequencing results for the validation of the ACX_7C-TEV library. Repeating sequences containing the HPQ motif are color-coded.

Trial 1AR2	Trial 1BR2
ACHPQGDPNCGGGENLYFQS (2)	ACHPQFPRYCGGGENLYFQS (6)
ACHPQGPLMCGGGENLYFQS	ACHPQGDPNCGGGENLYFQS (2)
ACSIWALMHCGGGENLYFQS	ACHPQNPHSCGGGENLYFQS
ACNYELTDNCGGGENLYFQS	ACLSTNGEDCGGGENLYFQS
ACNETYVKYCGGGENLYFQS	
ACYVPSQQPCGGGENLYFQS	
ACTPSPTNTCGGGENLYFQS	
ACLQMMHITCGGGENLYFQS	
ACIPAWDPRCGGGENLYFQS	

3.4 EXPERIMENTAL PROCEDURES

3.4.1 General Methods

Chemicals and reagents were commercially purchased from Fisher Scientific and Sigma-Aldrich. Cloning primers were purchased from IDT. Restriction enzymes and ligation buffers were purchased from NEB. Sanger sequencing was done at Azenta.

Absorbances were taken on a Nanodrop 2000 Spectrometer. Ligation reactions were done with T4 DNA ligase in its respective buffer in a MiniAMP Plus Thermal Cycler,

Thermo Fisher. Electroporations were done on Bio-Rad Gene Pulser: 25 μFD capacitance, 200 Ω resistance, and 1.8 kVolts. Phage retention and pannings were done using Dynabeads M-280 Streptavidin (Thermo Fisher, 10mg/mL, # 11205D).

Methods detailing NEB Cloning Protocol, phenol/chloroform extraction and ethanol precipitation, electroporation, DNA extraction and purification, and phage display/retention/amplification can be found in Chapter 2.

3.4.2 Midiprep M13KE gIII Phage Vector

The QIAGEN® Plasmid Midi Kit (#12143) was used to generate ample M13KE vector for library construction. The protocol was followed with some modifications.[4] In summary, M13KE (NEB #N3541S) was transformed into electrocompetent cells and plated on IPTG/XGal plates. The following day, one plaque was picked and allowed to grow in a 5mL culture for ~6 hours. One mL was inoculated in 100mL LB culture and allowed to grow overnight at 37°C at 250 rpm. The following day, the bacterial culture was harvested by centrifugation into 3 sterile JA-17 tubes at 6,000 x rpm for 15 minutes at 4°C.

The pellets were resuspended in 4mL Buffer P1 and combined. Buffer P2 (4mL) was added and mixed thoroughly by inverting 10 times and allowed to incubate at room temperature for 5 minutes. Prechilled buffer P3 (4mL) was added and mixed by inverting 10 times until colorless and allowed to incubate on ice for 15 minutes. The sample was then centrifuged at 6,000 rpm for 30 minutes at 4°C. The supernatant was transferred to a new, sterile JA-17 tube, and re-centrifuged for another 15 minutes using the same conditions.

A QIAGEN 100-tip was equilibrated by adding 4mL Buffer QBT and allowed to empty by gravity flow. The supernatant was added to the equilibrated tip, and was then washed twice with 10mL Buffer QC. The DNA was eluted with 5mL Buffer QC into a 15mL Falcon tube. The DNA was precipitated by adding 3.5mL of room-temperature isopropanol and inverted to mix before centrifuging at 7,830 rpm for 30 minutes at 4°C. The supernatant was decanted and the DNA pellet was washed twice with 2mL prechilled 70% ethanol. After the final 10-minute centrifugation at 7,830 rpm at 4°C, the pellet was allowed to air-dry before being resuspended in 100-500µL TE buffer.

3.4.3 PCR Clean-up

DNA cleanup was done following the Monarch PCR & DNA Cleanup Kit (NEB #T1030).[5] In summary, the sample was diluted with the DNA Cleanup Binding Buffer at a 2:1 buffer:sample ratio due to the size of the vector (7.2kb > 2kb), and mixed well by pipetting up and down. The sample was then loaded into the provided column/collection tube and spun at 14,000 rpm for 1 minute. The flow-through was discarded and the sample was washed twice with 400µL DNA Wash Buffer and spun for 1 minute at 14,000 rpm. Following the last centrifugation, the column was transferred to a clean 1.5mL microcentrifuge tube, and the DNA was eluted by adding 10-100µL DNA Elution Buffer to the membrane. The buffer was allowed to incubate in the membrane for 1 minute before centrifuging for 1 minute to elute the DNA.

3.5 REFERENCES

(1) Instruction Manual, Ph.D. Phage Display Libraries, 2020.
(2) *TEV Protease*. https://www.neb.com.
(3) *How to Use Phenol/Chloroform for DNA Purification*. https://www.thermofisher.com.
(4) Quick-Start Protocol QIAGEN® Plasmid Mini, Midi and Maxi Kits, 2016.
(5) Monarch® PCR & DNA Cleanup Kit (5 Mg) Protocol.

4.0 CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS

In parallel with the efforts described in **Chapter 3.2** and **3.3** above, we also started preliminary work on constructing a phage construct with the epidermal growth factor (EGF) displayed on the pIII coat protein as another avenue towards optimizing a phage display protocol against cell *in-vitro*.

4.1 EGF-PHAGE CONSTRUCT

Due to the size of EGF (53 amino acids residues)[1] we will need to utilize helper phage to construct this phage virion, as it is too large to insert into the M13KE vector. The plasmid pHen2 Avi-tag-linker-p3 from Addgene (#119819) will be used to conjugate the EGF oligomer to the pIII coat protein before transduction with helper phage. Initial efforts have begun with this construct. **Table 4-1** details the oligomer sequences to be used.

Table 4-1. Oligomer sequences used for generating the EGF insert. The annealing portion of each sequence is denoted as lowercase.

Sequence	Oligomer 5' - 3' (Annealing portion in lowercase)
EGF-Forward	CATGCCCGCCATGGAACAGCGATAGCGAATGCCCGCTGAGCCATGATGGC TATTGCCTGCATGATGGCGTGTGCATGTATATTGAAgcgctggataaata
EGF-Reverse	CATGTTTGCTAGCGCGCAGTTCCCACCATTTCAGATCGCGATACTGGCAGC GTTCGCCAATATAGCCCACCACGCAGTTGCACGCAtatttatccagcgc

Due to the limitation of oligomer size, we decided to design two oligomers that will anneal at the 3'-ends, followed by extension and digestion, similar to library construction detailed in **Chapter 3**. **Figure 4-1** is a visual representation of this process.

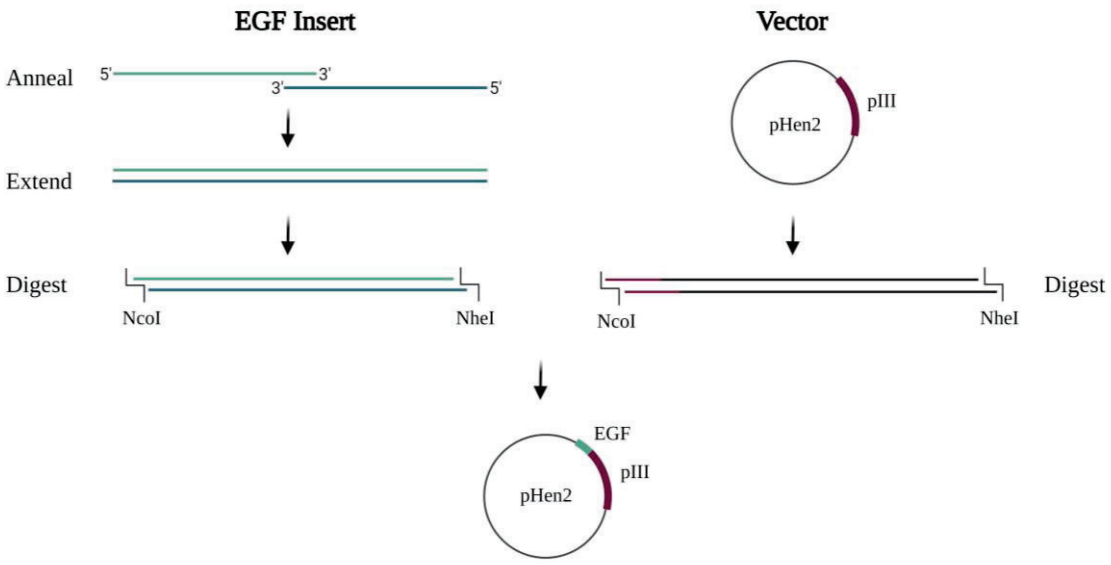


Figure 4-1. A visual representation to generate the EGF-pHen2-pIII plasmid for the EGF-phage virion transduction. Created with BioRender.com.

We intend to use the CM13[2] and CM13d3 pIII-defective[3] helper phages to generate EGF-phage constructs with varying copies of EGF on the pIII coat protein. CM13 is a derivative of M13KO7 containing an interference resistant ir3B A→G mutation at position 8418 of M13KO7.[2] CM13-infected cells produce more phage virions and are very efficient at a small scale. We predict that using different concentrations of CM13 helper phage we can produce EGF-phage constructs will varying copies of displayed EGF (1-5 copies). Similarly, to ensure that up to all 5 copies of the pIII protein contain the conjugated EGF protein, we will use the CM13d3 pIII-defective helper phage, which is specifically designed for multivalent phage display. CM13d3 is also a derivative of M13KO7 containing a wild-type pIII phenotype allowing efficient bacterial transduction yet lacks a functional pIII gene, thereby increasing phagemid multivalent display by up to a 100-fold.[3] General protocols for bacterial transduction with these helper phages are included.[4,5]

4.2 FUTURE DIRECTIONS

After constructing the EGF-phage virion, we will use that to begin on the protocol optimization against cells *in-vitro.* We have already confirmed that EGFR is overexpressed on A431 cells as compared to other cells lines, Skov3 and BT20, via western blot (results not provided). We will start screening efforts using the previous protocol[6] and optimize it as needed. From here, we will evaluate the relationship between ligand affinity and protein density. This will be accomplished by using EGF-phage virions with varying copies of displayed EGF and lentiviral transduced HEK293 cells with different expression levels of EGFR, which have already been made and validated.

In parallel, we will continue our efforts to identify a novel peptide ligand for EGFR ECD using covalent phage libraries. It is anticipated that peptide ligands with varying affinities will be identified, which can be used to assess the relationship between ligand affinity and protein density. In addition, we intend to assess which covalent warhead and peptide library construct is best for inhibiting PPIs. Finally, we intend to apply this data (*in-vitro* protocol optimization, the relationship between ligand affinity and protein density, and optimal covalent warhead/library construct) to other oncogenic cell lines, such as MCF-10CA1a, to identify potent peptide ligands for overexpressed oncogenic proteins.

4.3 REFERENCES

(1) Ogiso, H.; Ishitani, R.; Nureki, O.; Fukai, S.; Yamanaka, M.; Kim, J.-H.; Saito, K.; Sakamoto, A.; Inoue, M.; Shirouzu, M.; Yokoyama, S. *Cell.* **2002**, *110*, 775-787.
(2) *CM13 - Interference-Resistant Helper Phage*. https://www.abdesignlabs.com.
(3) *CM13d3 pIII-Defective Helper Phage*. https://www.abdesignlabs.com.

(4) CM13 Interference-Resistant Helper Phage INSTRUCTION MANUAL, 2016.
(5) CM13d3 PIII-Defective Helper Phage INSTRUCTION MANUAL, 2020.
(6) McGuire, M. J.; Li, S.; Brown, K. C. *Methods Mol Biol.* **2009**, *504*, 291–321.