

DISSERTATION

EXPLORATION OF PROTEIN ENGINEERING METHODS TOWARDS BIOMATERIALS,
THERAPEUTIC PROTEIN SCAFFOLDS, AND LOCALIZATION DETECTION WITHIN
MAMMALIAN CELLS

Submitted by

Jennifer N. Bjerke

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2019

Doctoral Committee:

Advisor: Alan Kennan

Christopher Snow
Robert M. Williams
Jennifer Nyborg

Copyright by Jennifer N. Bjerke 2019

All Rights Reserved

ABSTRACT

EXPLORATION OF PROTEIN ENGINEERING METHODS TOWARDS BIOMATERIALS, THERAPEUTIC PROTEIN SCAFFOLDS, AND LOCALIZATION DETECTION WITHIN MAMMALIAN CELLS

Proteins are large biomolecules entangled with complex chemistry that uniquely control the processes, function and efficiency of everyday life. These macromolecules are the final product of the central dogma, and as in any synthesis, the proper reactants must combine in order to produce the correct products. As technologies develop to understand their production and resulting structure, sequence and activities, they have increasingly become a bulk chemical platform in which nearly any desired application can be engineered- ranging from materials to therapeutics. The first half of this dissertation will discuss the potential to break down the central dogma in order to create new biological materials through incorporation of novel building blocks, known as amino acids. The second half will focus solely on the design and analysis of engineered proteins to create scaffolds for biological therapeutics that have the capability to bind almost any disease relevant target.

The second chapter of this dissertation describes the synthesis of novel non-canonical amino acids that add new chemistries to any protein of interest. Using amber codon suppression, and careful engineering of tRNA synthetase and cognant tRNA's, these new amino acids can be site selectively incorporated. To start, derivatives of phenylalanine are selected since this machinery has been successfully implemented for many new amino acids. The synthesis of

bithiophene explores the utilization of phase transfer catalysis and transition metal cross-coupling to deliver a racemic mixture of the non-canonical amino acid with a bithiophene moiety.

The third chapter of this dissertation will discuss the engineering potential of nanobodies, a monomeric protein responsible for all binding interactions, isolated from the variable heavy chain of camelid antibodies. We had previously reported a cationic resurfacing strategy that endowed mammalian cell penetration in three nanobody scaffolds. We have since explored and refined this strategy, along with endowing the capability to bind new targets. In the end we found that, when fused to sfGFP, a polyarginine version (PolyR-NB) increased internalization over the original cationic GFP nanobody (CatNB), but the original CatNB scaffold was more adaptable to extensive mutagenesis. Therefore, the PolyR nanobody scaffold can be utilized to aid the internalization of therapeutic cargo proteins that would otherwise not be able to cross the lipid bilayer, and the CatNB scaffold can become a therapeutic itself, modified to bind any intracellular target of disease relevant proteins. To further prove the later point, the complimentary determining regions (CDRs) of a separate nanobody, the BC2 nanobody, were grafted upon the CatNB framework, keeping all resurfacing and structure intact. Remarkably, the CatNB-BC2 CDR nanobody was able to maintain binding to its wild type partner.

The final chapter will showcase the efforts towards development of a facile luminescent assay that detects protein delivery to the cytosol of cells. Adapted from the NanoBit assay developed by Promega to identify protein-protein interactions, our assay utilizes the split nanoluciferase technology to produce a luminescent signal once an exogenous protein has recombined with its other half in the cytosol. This system, in theory, could be applied to identify endosomal release of virtually any therapeutically relevant protein.

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisors over the past 6 years and all of those who adopted me in to their labs for any period of time. Specifically, Dr. Nick Fisk, Dr. Eric Ferreira, Dr. Brian McNaughton and Dr. Alan Kennan. I greatly appreciate the time and effort that you put in to get me to this point, it certainly was valuable to receive your advice and mentorship. I definitely was not expecting to move around so much in my graduate degree, but I can say that I have learned a lot these past few years, am thankful for my diversity in experience, and have truly found a passion for the work I discuss in this dissertation.

Of course, I absolutely could not of handled all the ups and downs of these years without my family and friends. I want to thank my sisters Crystal Boyle and Shanon Bjerke, and my brother-in-law Marc Boyle. Crystal and Marc, you have no idea how refreshing it is to be able to come home and visit and feel like I'm hanging out with my best friends. You always make time to hang out with me and ask me how I'm doing, even if it's midnight when you're done putting my two incredibly amazing nephews to bed. You have gone out of your way to provide for me and take care of me while I'm home and I am forever grateful for that gift. To my big sister Shanon, thank you for visiting me in Colorado! Nothing makes me happier than showing my home, my work, and the places that give me joy to the ones I love the most. Your care packages and words of inspiration and wisdom have always been so thoughtful and have meant so much to me. I'm so excited to be done with this chapter in my life and be able to spend so much more time with my siblings and my nephews.

To my parents, I don't think anything I write here could thank you enough. Especially because it's not just the past 6 years that you have supported me and my dreams and goals. To

my father, you relentlessly continued to remind me that this was a goal I made, and no matter what was thrown at me I could handle it and complete my goal. You knew I could do it and I really needed that. It balanced perfectly with mom's complete and utter support for any decision I would make, which is an incredible gift in and of itself, even if it was to leave grad school. To my mother, you truly are my number one fan and even if I didn't receive it well all the time, it always pumped me up and made me continue to believe in myself. I really needed that as well. You both are such incredible parents and my biggest drive in life has always been to make you proud. A small effort to give back to you the love and support you have shown me.

To Madeline Renz, my best friend and also my cheerleader. Watching you face all the challenges you have and still be a ray of sunshine in everyone's life has been awe inspiring. Your energy always brightened up my days and made me know that I could do anything and you would be by my side. I look up to you, I admire you, I'm so impressed by you in all that you do and I am so blessed to have had you by my side for over 17 years. I'm so thankful for your support and encouragement through these years.

To Courtney Weeks, the most caring and thoughtful friend I have ever had. Undergrad would have not been the same without you and I'm even more thankful we have stayed close over the years and I've gotten to see you accomplish your dreams. Your drive, your passion for life, your absolute determination to earn your doctorate was, and always has been, so inspiring. I loved that we were able to share each other's grad school pains, even in separate states! Thank you for always treating me every moment you could during our visits, I'm excited to finally be able to treat you back!

Ok.... To Zachary Butz and Nicholas DeWeerd. I think you both know how much I love and appreciate you, so I will try and make a short summary. Zach, as much as I hated the Welch

house and living in that basement, nothing would make me trade that year we shared as mole people. From making breakfast with each other every morning, to watching EVERY episode of gossip girl, to you writing out a hundred name reactions on our white board (for a class you weren't in) to help me study for my organic exams... those are by far the best roommate memories I have. From hiking the Grand Canyon to running around New York City together, I'm so thankful we were able to continue making memories over the past 6 years and I look forward to many more. You challenged me to believe in myself, you fought for me to be strong and independent, you taught me how to let go of stress and be myself, you really dug into my thoughts and feelings more than any friend really has. I don't know how to thank you enough and I couldn't be more honored to be a groomsman in your wedding.

Nick... or should I say Nic? Although we were roommates our first year, I'm more thankful for these past three. You became my day to day best friend. I love the fact that we are constantly chatting about our thoughts about politics, random news articles, memes, dogs, cats, weather... really anything. You challenge me on an intellectual level and you always encourage any idea or anything I have to say. You've questioned me on why I would ever think I wasn't smart enough, or didn't accomplish enough, or why I wouldn't be proud of the work I've done. You've supported me and pushed me to be proud of myself and that is something I cherish deeply.

To my love, Kyle Pigeon. You certainly changed my life two years ago. It can be challenging to maintain a strong relationship in grad school, especially with someone in a city an hour away, but with you it has been incredibly easy. You taught me what it's like to be in a healthy, loving, supportive, fun, adventurous, honest, and trusting relationship. Most importantly, you gave me the freedom and support to chase every experience and build all of the relationships

I have today. I love that we both got to explore the best of life, music and Colorado together. I am so thankful for every one of our experiences and our many trips that we've had because of your support and because you wanted to share every memory possible with me. You've been there for me through some of my most challenging moments in grad school, and were "all-in" even when I was about to move across the country. It was a pure blessing that I didn't have to leave and got this bonus year with you in Colorado. You continue to surprise me with the effort you put in to make me happy and make sure that I always put myself first in many situations. I can't wait to see how we grow together now that this chapter is over. I love you endlessly.

Last but not least, to all my lab mates and friends that have mentored me the past couple of years. Rachel Tennyson, Susanne Walker, Patrick Beardslee, Angie Ta, David Crawford and Alex Hendricks, you guys made my time in the McNaughton lab special. Thank you for your council and support. Thank you for being my lab family. I had an absolute blast getting to know all of you and having friends that can directly relate and be there for me on a daily basis. And of course to all of my "Colorado Adventurers", I think we are all well aware of the special people we have found here. I have been blessed to know each and every one of you. You all repeatedly reminded me to focus on self-care and all my passions outside of research. When I thought I had to move across the country, leaving this incredible group of friends was one of the hardest things I had to face. Colorado would not have been the same home to me without you.

When I was a senior in undergrad, I went to a seminar on 'how to get into grad school' and everything about grad school. I don't remember anything from that seminar. All I remember is my professor stating at the end that grad school is extremely challenging and will really test you're limits... but you will also make some of the best friends you will have for a lifetime. Nothing in grad school turned out how I thought it would- except for that. I love you all so much.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES	x
LIST OF FIGURES.....	xi
LIST OF ABBREVIATIONS	xiii
CHAPTER 1: Protein Engineering Methods that Expand upon and Modify the Natural Biology of Proteins and their Applications.....1	
1.1 Introduction.....	1
1.2 Recombinant Protein Synthesis	3
1.2.1 The Genetic Code	4
1.2.2 Non-Canonical Incorporation Systems	6
1.2.3 tRNA/aaRS Orthogonality	8
1.2.4 Screen and Selection Assays for Incorporation of ncAAs	9
1.3 Engineering Proteins Towards Therapeutics in Mammalian Cells.....	12
1.3.1 Methods of Internalization via Endocytosis	13
1.3.2 Endosomal Technologies for Exogenous Protein Delivery	15
1.3.3 Cationic Resurfacing: TAT HIV to CPPs	15
1.3.4 Cationic Resurfacing: GFP, Miniature Proteins and Nanobodies	17
1.4 Assays to Detect Internalization and Protein-Protein Interactions	19
1.4.1 Cre-Recombinase.....	20
1.4.2 GIGI/GIGT Assay.....	21
1.4.3 Complimentary Split Reporter Assays for Cytosolic Detection.....	23
1.5 Conclusion	24
REFERENCES	27
CHAPTER 2: Synthesis of Non-Canonical Amino Acid Bithiophene Towards Proteins with Polymer-Like Qualities	
2.1 Overview.....	35
2.2 Introduction.....	36
2.3 Synthetic Plan	39
2.4 Results and Discussion	41
2.4.1 Transition Metal Cross-Coupling	41
2.4.2 Phase Transfer Catalysis.....	44
2.5 Future Directions	50
2.6 Conclusion	51
2.7 Materials and Methods.....	52
REFERENCES	59
CHAPTER 3: Uptake and Cellular Localization of Differentially Resurfaced Cell-Penetrating Nanobodies	
3.1 Overview.....	64
3.2 Introduction.....	64
3.3 Results and Discussion	68

3.3.1 Optimization of the CatNB Resurfaced Nanobody Scaffold for Delivery of Cargo	68
Proteins	68
3.3.2 Physical Characteristics of Newly Resurfaced Nanobody Scaffolds	72
3.3.3 Cytosolic Fractionation of PolyR-NB _{OFF} -GFP and CatNB _{OFF} -GFP	74
3.3.4 Cytosolic Fractionation with Endosomal Release Agents	76
3.3.5 Inhibition of Endocytosis Pathways	77
3.3.6 CDR Grafting of BC2 Nanobody on to Resurfaced Scaffolds	78
3.4 Conclusion	81
3.5 Material and Methods	82
3.5.1 Materials	82
3.5.2 Cloning	83
3.5.3 Protein Purification	83
3.5.4 Mammalian Cell Culture	84
3.5.5 Live Cell Confocal Microscopy	84
3.5.6 Flow Cytometry	84
3.5.7 Cytosolic Protein Extraction and Whole Cell Lysate Preparation for Fluorescent	
Microplate Reader Analysis	85
3.5.8 CDR Grafted ELISA	86
REFERENCES	87
CHAPTER 4: Split NanoLuciferase as a Tool for Identifying Protein Localization in Mammalian	
Cells	91
4.1 Overview	91
4.2 Introduction	92
4.3 Assay Design and Methodology	95
4.3.1 In Vitro Assay to Identify Successful Recombination	95
4.3.2 Transfection Conditions	98
4.3.3 High Copy Plasmid Conditions	101
4.3.4 Secretion Tests	104
4.3.5 Luminescence of Cells in Suspension	106
4.3.6 Comparison of Mammalian Expression Plasmids	107
4.3.7 Protein Interaction within Co-Transfected Cells	109
4.3.8 Luminescence of Total Cell Lysate	110
4.3.9 Endosomal Release of Cationic Resurfaced Proteins	112
4.4 Conclusion and Future Directions	114
4.5 Materials and Methods	117
4.5.1 Cloning and Protein Sequences	117
4.5.2 Protein Expression and Purification	120
4.5.3 In Vitro Assays	120
4.5.4 Mammalian Cell Culture	121
4.5.5 NanoBit Luminescence Assays	121
REFERENCES	122
Appendices	125
Appendix A: Supplemental Data for Chapter Two	125
Appendix B: Supplemental Data for Chapter Three	142
Appendix C: Supplemental Data for Chapter Four	151

LIST OF TABLES

Table 2.1- Reaction Series of Ethyl-4-Bromobutyrate and Aromatic Halides	43
Table 2.2- Synthetic Routes and Conditions to Alkyl Halide Bithiophenes.....	46
Table 2.3- Synthesis of Racemic Bithiophene Amino Acid with Various Schiff Bases	48

LIST OF FIGURES

Figure 1.1- Rational Design and Directed Evolution.....	2
Figure 1.2- Recombinant DNA Technology Overview	5
Figure 1.3- Peptide Synthesis with Amber Codon Suppression Highlighted	7
Figure 1.4- tRNA and tRNA aaRS Orthogonality	9
Figure 1.5- Common Selection and Screen Methods for ncAA Incorporation	10
Figure 1.6- Simplified Pathways of Endocytosis.....	14
Figure 1.7- Supercharged GFP Internalization Compared to Resurfaced Nanobodies	18
Figure 1.8- Well-known Reporter Assay Overview	21
Figure 2.1- Various non-canonical Amino Acid Side Chains	38
Figure 2.2- Transition Metal Cross Coupling Mechanism	39
Figure 2.3- Phase Transfer Catalysis Mechanism.....	40
Figure 2.4- Kumada-Corriu coupling for synthesis of 4-bromo-2,2'-bithiophene	41
Figure 2.5- Reaction Scheme for the Synthesis of Bithiophene ncAA.....	42
Figure 2.6- Crystal Structure of the Mutant Bipyridylalanine Synthetase.....	50
Figure 3.1- Cartoon Depiction of an IgG Antibody and Corresponding Nanobody	65
Figure 3.2- Ribbon Structure of Resurfaced GFP Binding Nanobody	66
Figure 3.3- Confocal Images of CatNB _{OFF} -GFP.....	69
Figure 3.4- Total Cellular Internalization of Optimized Resurfaced Nanobodies.....	70
Figure 3.5- Physical Characteristics of Resurfaced Nanobodies	73
Figure 3.6- Cytosolic Fractionation Assay of Normal Cells.....	75
Figure 3.7- Cytosolic Fractionation Assay of Cells with Chloroquine.....	76
Figure 3.8- Inhibition of Endocytosis Pathways	78
Figure 3.9- BC2 CDR grafted CatNB scaffold.....	81
Figure 4.1- Promega's NanoBit Reporter Protein.....	93
Figure 4.2- NanoBit Assay Design	94
Figure 4.3- In Vitro Analysis of new SmBit Proteins.....	97
Figure 4.4- Transfection Conditions and Initial Assay Luminescence.....	100
Figure 4.5- High copy Plasmid Transfections	103
Figure 4.6- Cells compared to Media Secretion of NanoBit	105
Figure 4.7- Relative Luminescence of Cells at Various Timelines	108
Figure 4.8- Co-Translation of NanoBit Proteins.....	110
Figure 4.9- Luminescence of Whole Cell Lysate	112
Figure A.1- ¹ H NMR of 4-bromo-2,2'-bithiophene	125
Figure A.2- ¹³ C NMR for 4-bromo-2,2'-bithiophene	126
Figure A.3- ¹ H NMR of N-Boc-Homoserine-Benzyl Ester	127
Figure A.4- ¹ H NMR of N-Boc-Iodohomoserine-Benzyl Ester	128
Figure A.5- ¹ H NMR of Ethyl-4-pyridin-2-yl) butanoate.....	129
Figure A.6- ¹ H NMR of Ethyl-4-(thiophen-2-yl) butanoate.....	130
Figure A.7- ¹ H NMR of Ethyl-4-([2,2'-bithiophen]-4-yl) butanoate	131
Figure A.8- ¹ H NMR for 2-([2,2'-bithiophen]-4-yl)ethyl-4-methylbenzenesulfonate	132
Figure A.9- ¹ H NMR of 2-([2,2'-bithiophen]-4-yl)ethan-1-ol.....	133
Figure A.10- ¹³ C NMR of 2-([2,2'-bithiophen]-4-yl)ethan-1-ol	134

Figure A.11- ^1H NMR of 4-(2-iodoethyl)-2,2'-bithiophene	135
Figure A.12- ^1H NMR of 4-(2-bromoethyl)-2,2'-bithiophene.....	136
Figure A.13- ^1H NMR of 2-([diphenylmethylene]amino)-4-phenylbutanenitrile	137
Figure A.14- ^1H NMR of Tert-butyl-2-([diphenylmethylene]amino)-4-phenylbutanoate	138
Figure A.15- ^1H NMR of protected racemic bithiophene amino acid	139
Figure A.16- ^{13}C NMR of protected racemic bithiophene amino acid	140
Figure A.17- ^1H NMR of deprotected Bithiophene amino acid	141
Figure B.1- Co-Purification of GFP binding nanobody.....	143
Figure B.2- Crystal Structures Showing Native Binding of GFP nanobody	144
Figure B.3- Expression and Co-Purification Assay of Optimized Nanobodies.....	145
Figure B.4- Fluorescent Standard Curves for Cytosolic Fractionation Assay.....	146
Figure B.5- Fluorescent Standard Curves for Cytosolic Fractionation Assay.....	147
Figure B.6- Coomassie and Fluorescence of FITC Labeled Nanobodies.....	148
Figure B.7- Sequences of BC2 Grafted Nanobodies	149
Figure B.8- MTT Assay	150
Figure C.1- Sequences and Binding Affinities of Promega's Engineered NanoBit.....	151
Figure C.2- PAGE-SDS gel of Ni-purified SmBit and LgBit Proteins	152
Figure C.3- Varying NanoGlow Reagents Amounts used in NanoBit Assay	153
Figure C.4- NanoBit Assay of Cells in Suspension.....	154
Figure C.5- In Vitro Assay of Proteins with Added Endosomal Escape Peptide	155

LIST OF ABBREVIATIONS

Abbrev.	Full Name
GFP	Green Fluorescent Protein
tRNA	Transfer RNA
aaRS	Aminoacyl tRNA Synthetase
ncAA	Non-canonical amino acid
mRNA	Messenger RNA
<i>Mj</i>	<i>Methanococcus jannachii</i>
PTD	Protein Transduction Domain
CPP	Cell Penetrating Peptide
sfGFP	Super Folder GFP
CDR	Complementary Determining Region
NB	Nanobody
Wt	Wildtype
Cat	Cationic
PolyR	Poly Arginine
SmBit	Small Bit, 1.3KDa Peptide
LgBit	Large bit, 18KDa Protein
PPI	Protein-Protein Interaction
Blac	B-lactamase protein
POI	Protein of Interest
PBS	Phosphate Buffered Saline

CHAPTER ONE

PROTEIN ENGINEERING METHODS THAT EXPAND UPON AND MODIFY THE NATURAL BIOLOGY OF PROTEINS AND THEIR APPLICATIONS

1.1 Introduction

Proteins are responsible for all processes in life that determine cellular function and fate. Proteins that have undergone mutations that alter their original form or activity can result in a variety of diseases, cancer, autoimmune and inflammatory responses, and even genetic disorders.¹⁻⁵ Understanding the lineage between DNA, RNA and proteins can help scientists build a map to probe their activity and gain control over exact characteristics through protein engineering.⁶⁻⁸ By mimicking evolution, protein engineering methods have produced proteins with improved binding affinities to block or disrupt disease relevant proteins, proteins that can replace and improve synthetically made materials, and proteins that even contain synthetically made amino acids to expand upon their native chemistry.⁹⁻¹⁴ With directed evolution and rational design, the capabilities of these macromolecules almost become limitless. This research will exemplify the use of proteins with new chemical capabilities beyond their natural biology, the ability of proteins to become new therapeutics, and the development of assays that help detect proteins within mammalian cells.

Directed evolution is a process by which natural selection is accelerated and steered in a defined direction. Typically, the gene of a protein is mutated to create a library of variant genes that will essentially result in a library of variant proteins (**Figure 1.1**). These variants will then range in the proportion of functional or dysfunctional variants. This function is a phenotype that

can be used to apply selection pressure against the library and separate out variants that reveal beneficial mutations.^{15,16} Proteins are created with 20 small molecule building blocks called amino acids. These amino acids hold all the chemical capabilities of every protein, but their range is rather limited. To expand upon this chemistry, the Schultz lab utilized directed evolution to create tRNA synthetases capable of binding and incorporating a non-natural amino acid in a desired protein. Chapter two of this dissertation will discuss the methods by which new chemical abilities can be added to proteins, and the synthesis of a novel amino acid to be incorporated using such methods.

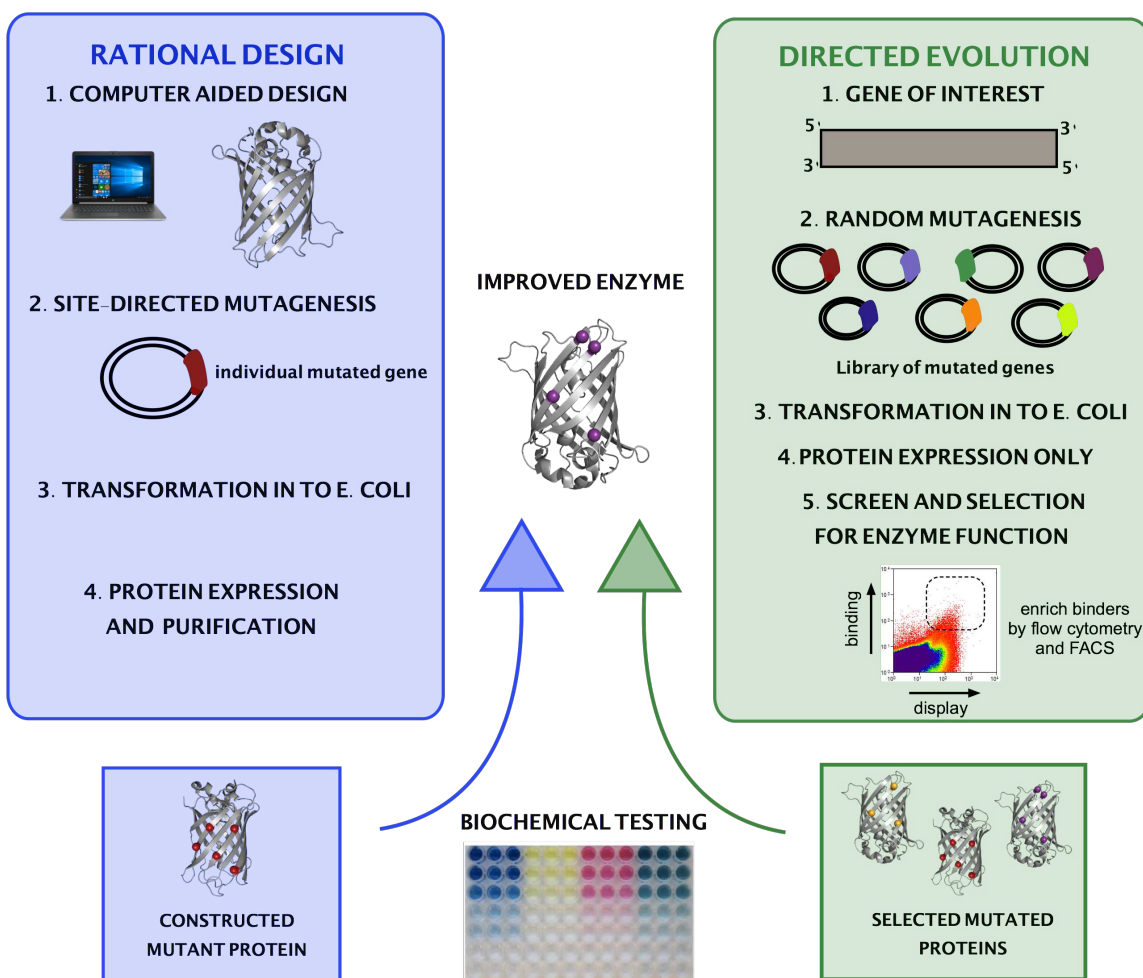


Figure 1.1: Rational design and directed evolution methodology towards engineered proteins

Rational design differs from directed evolution in the sense that creation of a library of proteins is unnecessary. The process relies on in-depth knowledge of a proteins structure (such as a crystal structure) and function in order to design thoughtful alterations (**Figure 1.1**).^{17,18} Small molecules have dominated the therapeutic field due to their ability to cross the lipid bilayer of mammalian cells to reach intracellular targets. Proteins, despite their large binding surface and increased ability to disrupt disease relevant protein-protein interactions, contain surface chemistry that typically does not allow passage into cells. Chapter three covers the rational design of proteins to restructure their surface chemistry to open the intracellular binding capabilities for protein therapeutics.

Finally, when a new function is created for a protein, there is a need for systems that can detect the successful activity and separate it from those that fail. Green fluorescent protein (GFP), and luminescent protein nanoluciferase, have pioneered the field of reporter assays that detect protein-protein interactions.¹⁹ Chapter four will discuss the manipulation of split nanoluciferase to detect the successful delivery of therapeutically relevant proteins into cells.

1.2 Recombinant Protein Synthesis

DNA is a double helix structure held together through the hydrogen bonding of nucleotide bases that have optimal Watson-Crick base pairing partners (Adenine with Thymine, Guanine with Cytosine.)²⁰ Genomic DNA contains all protein genes necessary for cellular health and function and is maintained with the nucleus of a cell.^{21,22} The level of which a protein is expressed and downstream regulation is largely dependent upon the amount of RNA transcripts.²³ RNA transcripts are the single stranded complimentary partners (with the exception of Uracil in place of thymine) of genes that are exported out of the nucleus to the cytosol and translated into proteins. This is done when the nucleotides are read by the RNA ribosome in a

triplicate fashion, called a codon.²⁴ These codons tell the ribosome which amino acid is to be added next to the polypeptide chain, the primary structure of a protein. Generally, a transfer RNA (tRNA) molecule carries the correct amino acid to the ribosome and is matched to the RNA transcript through its anticodon. An enzyme known as the aminoacyl-tRNA synthetase contains a binding pocket that is unique to an amino acid and catalyzes a reaction that acylates the amino acid on to its cognate tRNA partner.²⁵ This translational machinery greatly differs between species. Eukaryotes contain translational machinery that is able to perform RNA splicing of transcripts that can produce proteins with different functions from the same gene. Eukaryotic cells also perform post-translational modifications to many proteins that are necessary for their biological activity. Genetic engineering has been able to remove components from eukaryotic DNA and recombine these genes into DNA vectors that can be delivered to new hosts, such as bacteria, yeast and mammalian cells.^{26,27} These vectors, known as plasmids, can be transfected into a bacterium and adopt the host transcription/translation machinery without integrating into the host genome as seen in **Figure 1.2**. The resulting recombinant proteins have thus spearheaded the endless possibilities of protein engineering. So much so that researchers have begun to push the boundaries and expand upon the natural chemistry the native 20 amino acids provide.²⁸

1.2.1 The Genetic Code

Incorporation of non-canonical amino acids (ncAAs) consists of three key players; an mRNA transcript of the protein of interest, a tRNA bearing the correct anticodon, and the aminoacyl-tRNA synthetase capable of acylating the amino acid to the cognant tRNA.^{29,30} Starting with the mRNA component, permutations of the four possible DNA nucleotides in each of the three positions of a codon create a genetic code consisting of 64 codons as seen in **Step 2**

of **Figure 1.2**. Of these, 61 are called sense codons, which base pair with the anticodon of a tRNA molecule which contain an amino acid to ligate into the growing polypeptide chain.³¹ The remaining three are nonsense codons that encode the termination of protein synthesis, and are recognized by protein release factors instead of amino acid bearing tRNAs. With 20 natural amino acids and 61 codons, this means there are multiple RNA codons that code for the same amino acid. These codons vary in frequency between different host organisms, and the number of tRNA molecules for each codon can vary as well.³²

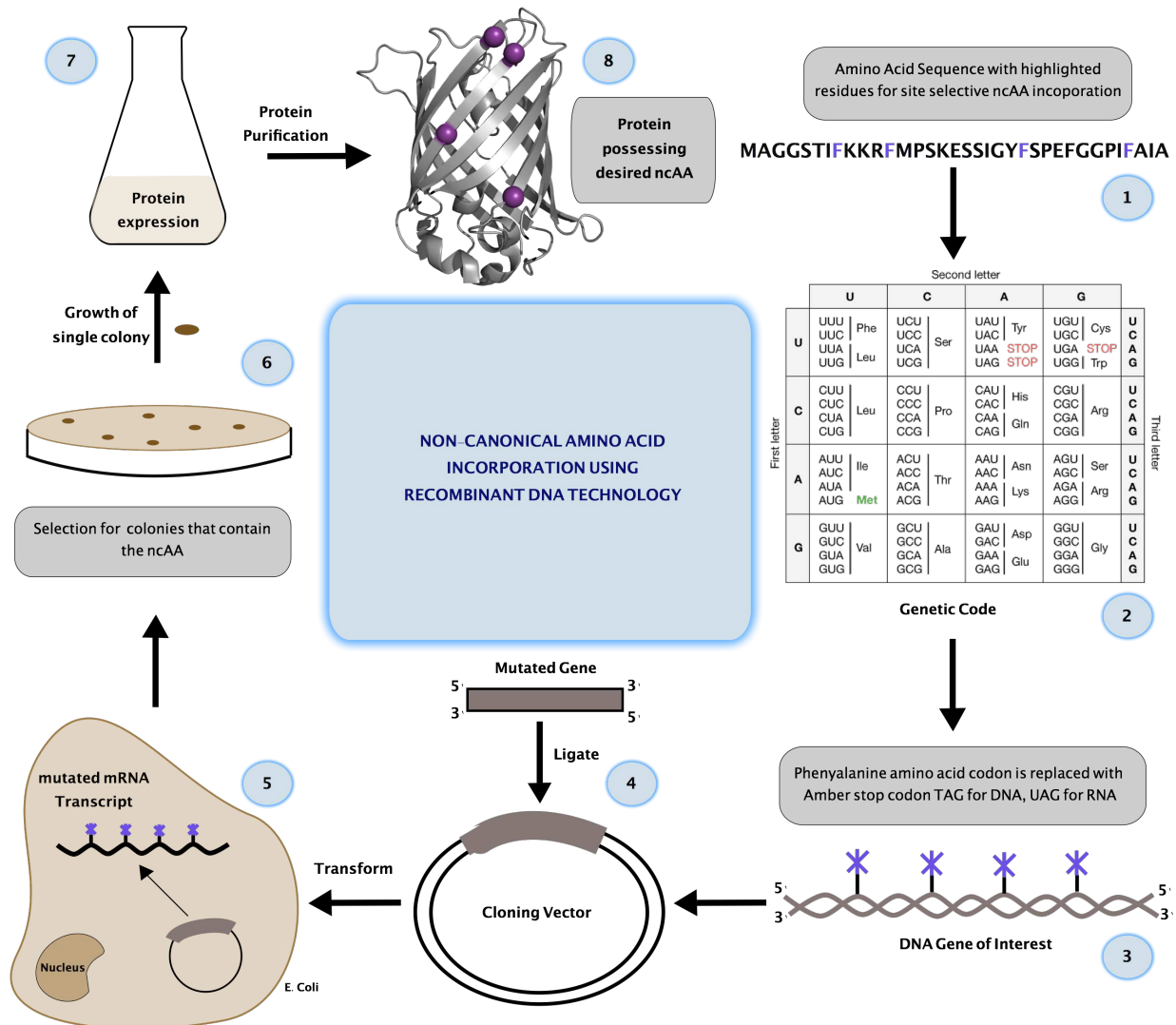


Figure 1.2: Recombinant DNA technology is utilized for all protein engineering methodologies including non-canonical amino acid incorporation

1.2.2 Non-Canonical Incorporation Systems

The two components, the tRNA molecule and aa-tRNA synthetase, require the most genetic engineering for *in vivo* ncAA incorporation (**Figure 1.3**). One of the frontier technologies in the incorporation of new amino acids is global amino acid replacement (GAAR), developed by Cohen and Cowie in the 1950's.^{33,34} It was proven that some aminoacyl-tRNAs are promiscuous with the binding of amino acids with side chains that are structurally similar to their wild type binding partner. This promiscuity could simply be controlled by concentration of the desired amino acid and the decrease in concentration, or total elimination of its wild type binder. Methionine is an amino acid that binds methionyl-tRNA synthetase, and has been efficiently replaced (up to 98%) by its non-canonical analogs.^{35,36} For the best incorporation efficiency, GAAR requires the host cells line to be auxotrophic for the amino acid that is to be replaced to eliminate any competitive binding of the natural amino acid and the ncAA to the endogenous synthetase. For this reason, methionine is commonly the amino acid of choice to completely reassign since, in *E.Coli.*, it only has one mRNA codon that codes for one tRNA. Amino acids that are encoded by multiple codons and tRNAs would drastically increase the complexity of GAAR, but has been done by Merkle and co-workers.³⁷ They were able to create a polyauxotrophic cell line to simultaneously incorporate proline, phenylalanine and tryptophan with fluorinated ncAA's into a protein of interest to study it's structure by fluorine NMR.

Rather than total elimination of a natural amino acid, researchers have been able to hijack codons with multiplicity for a single amino acid in order to incorporate their desired ncAA, creating a 21 amino acid genetic code.³⁰ Also, codon usage heavily varies in the frequency of use depending on the organism of interest. Due to this degeneracy in the genetic code, sense codon and nonsense codon reassignment have been developed in order to incorporate a number of

ncAAs, however; only nonsense codon suppression will be discussed here. Quickly, the nonsense codon suppression system is the efficient incorporation of a ncAA into a protein in response to a nonsense codon in the mRNA transcript using suppressor tRNAs. Amber (UAG), Opal (UGA) and Ochre (UAA) are the three nonsense codons that encode for protein synthesis termination, with amber occurring at the lowest frequency (0.02%) in *E. Coli*.³⁸ Pioneered by the Schultz lab, the Amber codon was globally replaced by one of the other two stop codons, leaving the UAG codon unique to the genome (**Figure 1.2: steps 2-5**).^{35,39,40} With the amber codon now unique to the host DNA, it can be placed anywhere within a gene to provide site-specific incorporation of a non-canonical amino acid. Next, the wild type suppressor tRNA bearing the Amber anticodon (CUA) would need to be amino acylated with the ncAA by the cognate tRNA synthetase. The codon-anticodon binding relationship is independent from the amino acylation mechanism between the tRNA synthetase and the suppressor tRNA. Therefore, incorporation of each new ncAA must be met with the engineering of the tRNA/tRNA synthetase pair.

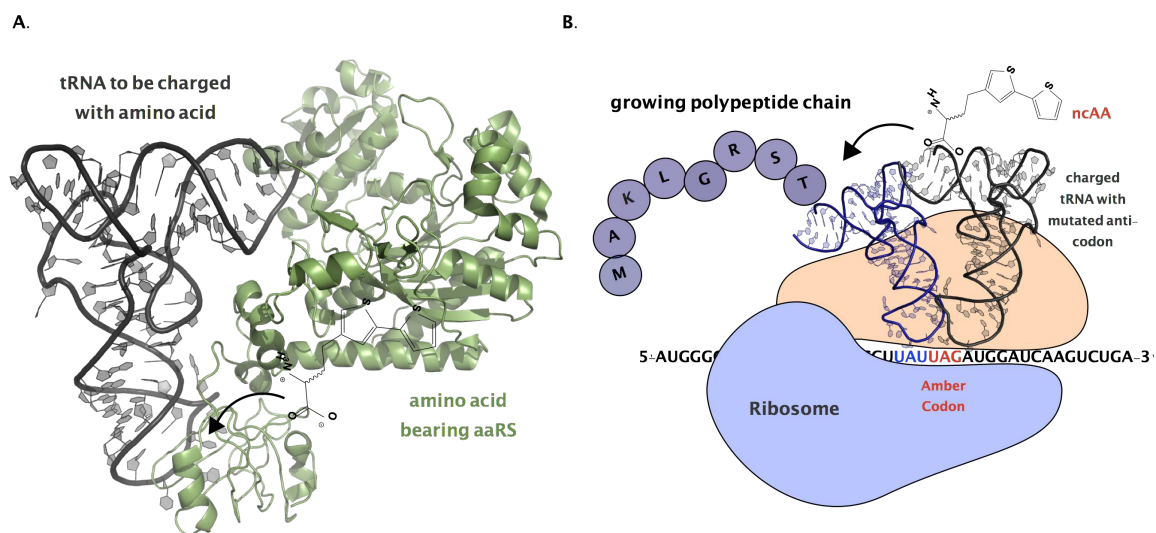


Figure 1.3: Peptide synthesis with amber codon suppression highlighted. A. Aspartyl-tRNA synthetase complexed with tRNA_{asp} (PDB: 1EFW). B. The tRNA charged with the desired amino acid will enter the ribosome bearing the mRNA transcript and bind through the anti-codon/codon relationship to transfer the amino acid to the growing polypeptide chains.

1.2.3 tRNA/aaRS Orthogonality

Usually, ncAA's are incorporated by exogenously adding a high concentration of the ncAA directly to the growth media. The small molecule permeates the cell wall and hopefully find its way to the synthetase that could normally bind its natural structural analog. This aminoacyl-tRNA synthetase would then load the suppressor tRNA with the ncAA. In early developments of incorporation systems, the wild-type synthetase could still bind its cognate amino acid, along with the ncAA, and the resultant proteins would show incorporation of both small molecules. To optimize incorporation efficiency of the ncAA, creating orthogonal systems became a critical development. Orthogonality is the ability of tRNA/tRNA synthetase pairs from one organism to be functional in a different host organism, without cross reactivity of the translational machinery (**Figure 1.4**).^{41–43} This means that a suppressor tRNA from yeast would not be aminoacylated by any tRNA synthetases from a host bacterium, but would still be accepted into the ribosome for translation.⁴⁴ Likewise, the cognate yeast tRNA synthetase would not perform acylation of any bacteria tRNA molecules. If the bacteria's tRNA synthetase were to recognize the exogenous yeast tRNA molecule (or vice-versa), this would subject the tRNA to proof-reading mechanisms that could result in the deacylation of the ncAA. The Schultz lab demonstrated the first successful orthogonal pair tRNA₂^{Gln}/GlnRS for the amino acid glutamine derived from the yeast *Saccharomyces cerevisiae* to be functional with *E. coli* translational machinery *in vitro*.⁴⁵ Although 138 ncAA's with structural similarity to glutamine and glutamic acid were incorporated with this orthogonal pair *in vitro*, it was not as effective *in vivo*. Unfortunately, the yeast derived GlnRS was found to be incapable of charging ncAA's onto the tRNA^{Gln} in *E. Coli*, due to the low expression of a eukaryotic GlnRS protein in the prokaryotic host.⁴⁶ Archaeal aminoacyl-tRNA synthetases also resemble eukaryotic machinery, however, the

synthetases were shown to have elevated levels of expression within *E. Coli*. A tyrosyl orthogonal pair was derived from archaeal bacteria *Methanococcus jannachii*. The amber suppressor *Mj* tRNA tyr and *Mj* TyrRS naturally incorporates the tyrosine amino acid and were a revolutionary discovery for their functionality in *E. Coli*.^{47,48} This orthogonal pair became widely popular due to the synthetases wide promiscuity for aromatic side chains in ncAAs, however, complete orthogonality was not achieved. Initial applications of this pair discovered that endogenous *E.Coli*. synthetases were capable of aminoacylating the *Mj* tRNA tyr. Again, researchers looked to protein engineering methods to improve the cross reactivity of the *Mj* TyrRS and identify evolved variants with selected characteristics for ncAAs of interest.

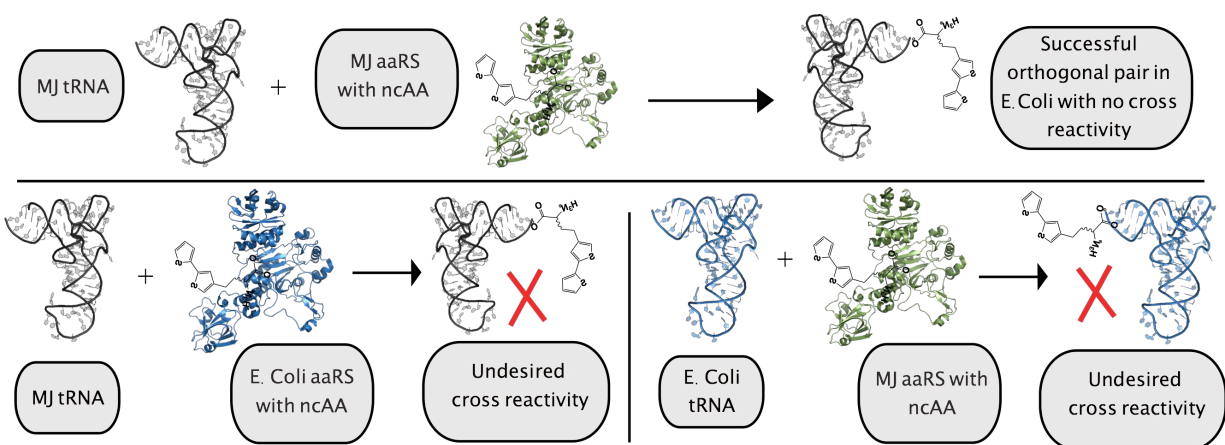


Figure 1.4: Diagram of a successful orthogonal tRNA and tRNA synthetase when applied in an *E.Coli* system. If the *Mj* tRNA and *Mj* tRNA synthetase charge amino acids using *E. Coli* machinery then the pair is not orthogonal and will reduce ncAA incorporation efficiency in *E. Coli*

1.2.4 Screen and Selection Assays for Incorporation of ncAAs

Thus far, the development of incorporation systems relied on the natural promiscuity of wild type synthetases for ncAA's that closely resemble analogs of the natural 20 amino acids. Directed evolution accelerated the discovery of new synthetases with improved binding

efficiencies for ncAAs. Using sense and nonsense codon suppression systems, engineers could site selectively place an open codon in any protein of interest that would drive a desired cellular outcome. This freedom advanced the creation of high throughput selection and screen assays in order to verify incorporation success and efficiency. It is known that with every new amino acid to be incorporated, a new orthogonal tRNA/aaRS pair must be engineered to bind the ncAA.^{36,42,44} Thus, libraries of mutational variants of wild-type synthetases are created to screen and select for enzymes with the best fit for the ncAA. In brief, screen assays will place the available codon in a critical position of a reporter gene (generally fluorescent or luminescent protein) that will result in a visible signal in living cells only if incorporation is successful. Selection assays operate similarly, except the reporter gene is one that will cause death of cells that do not contain the incorporated ncAA, thus “selecting” for positive hits. Here, a selection assay that utilizes amber codon suppression to verify incorporation, and a recently developed screen assay using sense codon reassignment will be discussed.

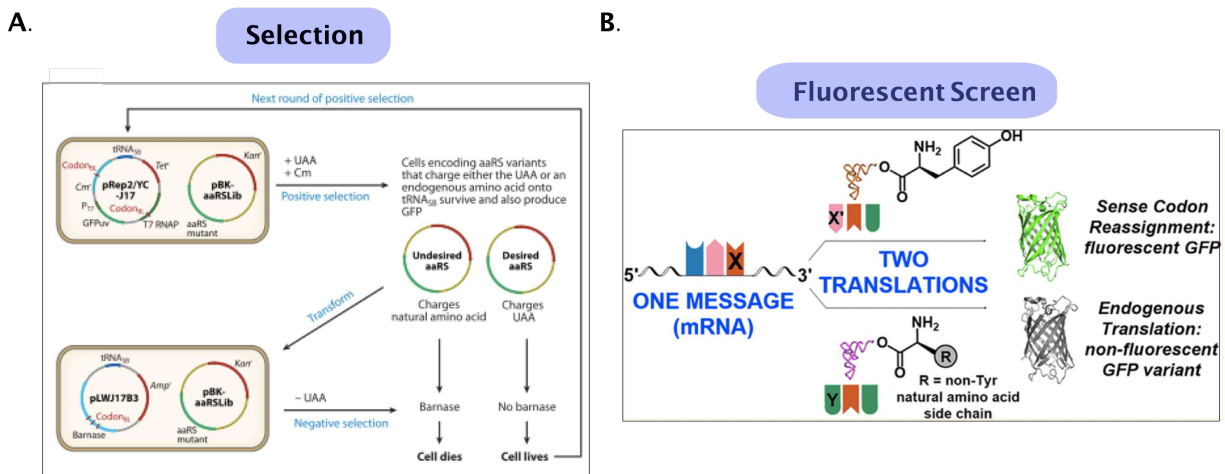


Figure 1.5: Common selection and screen methods developed by the Schultz and Fisk lab respectively.^{28,49} A. Positive and negative selection with a toxic selection marker such that cell viability is dependent upon desired protein function. B. Simple screen for a facile approach to visualize incorporation of ncAA's

The Schultz lab established a combination of positive and negative selection assays to evolve for an optimal mutant aaRS and mutant orthogonal tRNAs capable of charging ncAAs in *E. Coli* (**Figure 1.5.**) To establish this protocol, they started with libraries of the *Mj* tRNA^{Tyr} / tyrRS pair to discover the most efficient variants of each.⁵⁰ These selection assays can be used to sort through libraries of either tRNAs or tRNA synthetase to observe both orthogonality and ability to incorporate ncAAs. Initially, an amber codon is placed in a non-essential position in the gene of an essential protein for cell growth. The barnase protein is extremely cytotoxic to the cell and expression results in complete cell death.^{50,51} If the evolution of tRNAs is desired, the negative selection system will strike out tRNAs from the library that are aminoacylated from *E. Coli* machinery, indicating unwanted cross reactivity. Only the cells containing orthogonal tRNAs that did not incorporate any amino acid in response to the amber codon will survive because the barnase gene will not be expressed. All tRNAs extracted from surviving clones are then subjected to a round of positive selection. Positive selection will provide clones that only contain tRNAs that are successfully aminoacylated by its orthogonal synthetase and incorporate amino acids in response to the amber codon. In this case, the amber codon is placed within an antibiotic resistance gene such that, when the cells are grown in the presence of that antibiotic, the cells survive. Adaptations of this positive and negative selection system using the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS pair and amber codon suppression have successfully incorporated nearly 100 ncAAs.²⁸

In all incorporation methods discussed thus far, each require either the total removal of a natural amino acid in GAAR, or the total removal of a codon from the genome in amber codon suppression. These genome wide modifications can have cytotoxic effects on the cell and result in poor protein expression. Concomitantly, researchers have begun to strive for multisite

incorporation of ncAAs. To satisfy both needs, the degeneracy of the genetic code must be broken a step further. Sense codon reassignment takes advantage of rare sense codons and/or codons that bind their cognant tRNA through wobble base pairing. Rare sense codons are those occurring at the lowest frequency within *E. Coli.*, while wobble base pairing codons are instances where at least one of the three nucleotides do not bind the tRNA anticodon through Watson-Crick interactions.^{32,52} Watson-Crick base pairing is known to exhibit more favorable interactions with higher binding free energy than wobble base pairs. Therefore, in practice, tRNA anticodons with Watson-Crick pairing should be able to outcompete a tRNA with an anticodon that binds through wobble base pairing. For example, the first sense codon reassignment came with the exploitation of two codons that code for phenylalanine (Phe), UUU and UUC. Both codons are decoded by a single tRNA species with the anticodon GAA. The UUC codon would bind the GAA anticodon through Watson crick pairing, while the UUU codon exhibits a wobble base pair in the first position. Therefore, a tRNA bearing the anticodon AAA was engineered to incorporate a 21st amino acid, while leaving the UUC codon available for phenylalanine.^{32,52} Recently, the Fisk lab developed a facile fluorescent screen utilizing the essential tyrosine residue in the fluorescent chromophore to identify successful sense codon reassignment and evaluated four other wobble codon-anticodon pairs.⁴⁹ To date, nonsense and sense codon suppression has been widely implemented *in vivo* to study hot spots in protein interactions, protein structure and folding mechanisms, and even mimic post-translational modifications to elucidate their biological role.

1.3 Engineering Proteins Towards Therapeutics in Mammalian Cells

Beyond engineering of proteins within bacterial hosts, of greater interest is to engineer proteins for therapeutic relevance in mammalian cells. Previously, the development of biological

drug candidates had been restricted due to large differences between translational machinery in eukaryotes versus the simpler prokaryotic organisms. Thus, small molecules have consistently dominated the field, covering over 71% of total FDA approved drugs in the past two years.⁵⁴ However, most cellular functions are through proteins interacting with other proteins, and intercepting these interactions with small molecules presents many challenges. Luckily, genetic and protein engineering techniques have aided in surpassing some of biologics biggest challenges; purification, half-life activity, and intracellular delivery.^{55–58}

The first major advancement in biological therapeutics came when genetic engineering made it possible to place eukaryotic human genes of interest into prokaryotic host organisms. Recombinant human insulin was the first human protein therapeutic approved by the FDA that could be produced in *E. coli*.⁵⁹ Thanks to recombinant technology, protein fusions to super folder GFP, human serum and even conjugations to synthetic peptides and polymers aid in purification efficiency and potentially the biological half-life.^{57,60–62} The entirety of this review cannot begin to contain all the advancements that have been made to improve purification and half-life of biologics. Therefore, this review will solely discuss methods in which protein engineering has aided in the delivery of biologics into mammalian cells.

1.3.1 Methods of Internalization via Endocytosis

Endocytosis is the process by which extracellular material is engulfed by the cell membrane, creating a vacuole, which is then internalized in mammalian cells. The major pathways of endocytosis include receptor-mediated endocytosis (also known as clathrin-mediated endocytosis), pinocytosis, macropinocytosis, phagocytosis, and caveolin dependent (**Figure 1.6**).⁶³ Pinocytosis is the formation of the smallest vesicles in order to deliver extracellular fluid to the cell while macropinocytosis involves a large engulfment of extracellular

fluid, particles and sometimes macromolecules. Although some pathogens enter the cell through caveolar endocytosis, clathrin dependent endocytosis is by far the most common pathway for intracellular delivery of proteins. Clathrin coated pits are able to engulf a large variety of cell surface receptors that bind an array of extracellular proteins and molecules. Early endosomes are formed once the vesicle has pinched off from the external lipid membrane, creating a new internal lipid pocket (**Figure 1.6**).⁶⁴ Endosomal release is essential for the successful delivery of content to the cytosol of the cell where most cellular protein activity lies. Content that remains trapped within these internal vacuoles will eventually turn into lysosomes, which then signal for degradation by the cell. It has become critical for protein therapeutics to not only be able to cross the external lipid bilayer, but also show successful escape from endosomal entrapment.

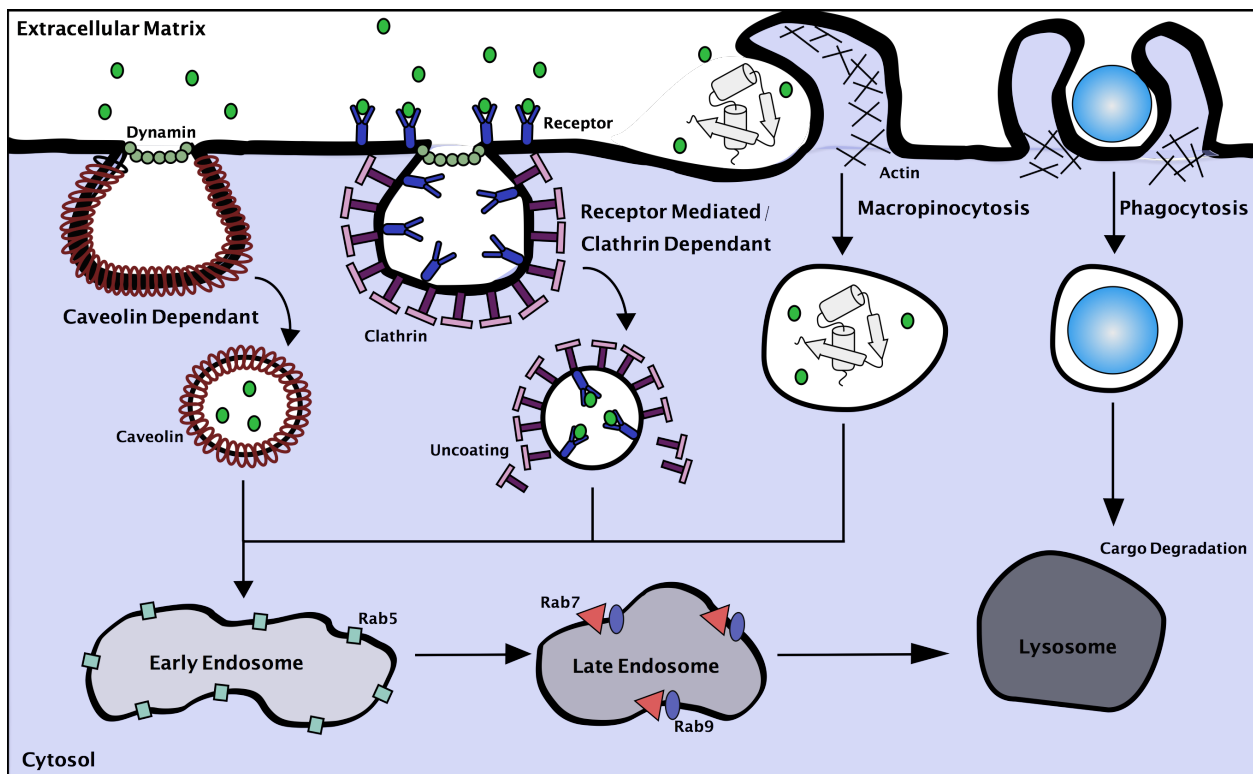


Figure 1.6: Simplified pathways of endocytosis. The majority of pathways shown become early endosomes. Successful protein delivery depends upon endosomal release before the vesicle is degraded as a lysosome.

1.3.2 Endosomolytic technologies for exogenous protein delivery

Anthrax, Diphtheria and Botulinum are common plant and bacterial toxins that mainly utilize clathrin-mediated endocytosis for the delivery of their cargo into the cytosol of mammalian cells. These toxins consist of two protein units, the A unit being the internalized toxic protein while the B unit is responsible for cell membrane fusion. Recently, the anthrax toxin protein translocation machinery has been redesigned to remove cytotoxicity of the subunit A, known as the lethal factor protein (LF). At the same time, subunit B was engineered for increased recognition of cell surface receptors known as a protective antigen (PA) that assemble within the cell surface and create a PA pore.⁶⁵ Therapeutically relevant protein cargo can be fused to the LF, which then recognizes the PA pore and is internalized into early endosomes.⁶⁶ Minimalized versions the LF (LF_N) have been engineered to recognize HER2 receptors expressed on the surface of breast cancer cells and deliver treatment to induce apoptosis.^{67,68} The PA/LF_N delivery system has been used to deliver over 30 macromolecules and polypeptides. Although the PA/LF pathway allows for high levels of cytosolic delivery, the requirement of multiple protein domain fusions increases the difficulty of protein expression and thereby practical use. Also, it has been proven that not every desirable protein cargo is able to translocate into cells by the PA/LF_N system. However, these delivery systems have enhanced our knowledge about the surface composition of cells and spearheaded relative research in to other variants of protein transduction domains.

1.3.3 Cationic resurfacing: TAT HIV → CPPs

Carrier proteins, termed protein transduction domains (PTDs), are protein or peptide units that have been identified to aid in protein internalization. Possibly the most significant discovery

was of the TAT protein of human immunodeficiency virus 1 (HIV-1.)⁶⁹ The native TAT protein is 86 amino acids long, however recombinant genome engineering and subsequent transactivation assays discovered that TAT amino acids 38-58 were sufficient for uptake of exogenous delivery. Interestingly, this region of the TAT protein is highly basic (consisting of lysines and arginines), and does not seem to bind any sort of cell surface receptor.^{70,71} The peptide was also found to penetrate a variety of mammalian cells, leading to the assumption that it is interacting with the negatively charged surface chemistry of the cells themselves. Since the discovery of the TAT peptide, it has been proven that highly dense cationic regions in peptides ranging from 4-24 amino acids in length are sufficient for protein transduction, now known as cell-penetrating peptides.⁷⁰

A shocking number of CPPs have been derived from other naturally occurring protein domains such as the Xentry peptide from Hepatitis B virus, penetratin from the third helix of the homeodomain in Antennapedia, and a variety derived from snake, scorpion and spider toxins.⁷²⁻⁷⁴ Although discovering new CPPs from nature can teach us mechanistic details about these viruses and toxins, a more efficient strategy to discover the most effective CPP may be to apply rational design with our pre-existing knowledge. Arginine's and lysine's are basic amino acids that are able to interact with the negatively charged, cell surface glycoaminoglycans.⁷⁵⁻⁷⁷ Poly-arginine peptides of six amino acids or longer have been proven to penetrate cells at a larger rate than comparable peptides of poly-lysine.⁷⁸ As of today, over 1700 new CPP's have been identified, with many of them being used towards medical imaging, anti-cancer agents and delivery systems of other protein therapeutics up to 120KDa.⁷⁹ Although CPPs remain the most commonly used method of protein transduction, these small α -helical structures are susceptible to degradation from proteases present in the extracellular matrix. One method to combat

proteolytic degradation is to fuse an anionic peptide that blocks the degradation of the cationic CPP through electrostatic interactions connected by a cleavable linker.⁸⁰ This hairpin-like structure is completely protected from degradation until the presence of extracellular proteases cleaves the linker, exposing the CPP and thus initializing cellular uptake.⁸⁰ Other shielding domains can be used to protect the CPPs from degradation once in the cytosol and increase the drugs half-life.⁷⁹ Another method to protect from proteolytic degradation is the addition of a synthetic hydrocarbon “staple” within the α -helical structure, called a helical mimetic. Recently, an amphipathic histidine containing stapled peptide LKH-stEK complexed with siRNA cargo was discovered to have high cellular potency and efficient endosomal release when delivered at 500nM to mammalian cells.⁸¹ However, even the most popular CPPs built to deliver protein therapeutic cargo suffer from high dosage concentrations (typically up to 5uM) and frequent administrations to combat this stability challenge. Protein engineering has therefor looked for ways to apply a cationic charge to a more robust structure, without the need for α -helical stability. The ideal candidate would minimally change the overall size of the cargo to be delivered and could not affect the native protein function.

1.3.4 Cationic resurfacing: GFP, miniature proteins, and nanobodies

A major successful strategy was the supercharging of GFP to endow cell penetration. The Lui lab took super folder GFP (sfGFP) and used site-directed mutagenesis to completely alter the surface chemistry to a net negative or net positive charge.⁸² Although supercharged sfGFP is a rather large protein, fusion to this fluorescent macromolecule provides a dual functionality, cellular penetration and visualization of protein trafficking in mammalian cells (**Figure 1.7**).⁸³ The -34 or +36 sfGFP was also shown to be resistant to aggregation, a characteristic that is important for therapeutic cargo to avoid.⁸⁴ Fusing +36GFP to cre recombinase, an enzyme

known to catalyze the recombination of reporter genes, showed a 10-fold higher cellular penetration over TAT and penetratin CPPs at 500nM.⁸⁵ While some CPPs can avoid entering the cell through endocytotic pathways, these larger macromolecules are shown to be delivered through clathrin coated pits and macropinocytosis, resulting in endosomal entrapment. Addition of the small molecule chloroquine enhances endosomal release and proved the high levels of endosomal sequestration of +36GFP. More importantly, cationic landscaping of surface-exposed residues is not a transferrable method to most proteins without altering their fold and native function. This in turn would require the fusion of any biologic to this rather large (30KDa) protein.

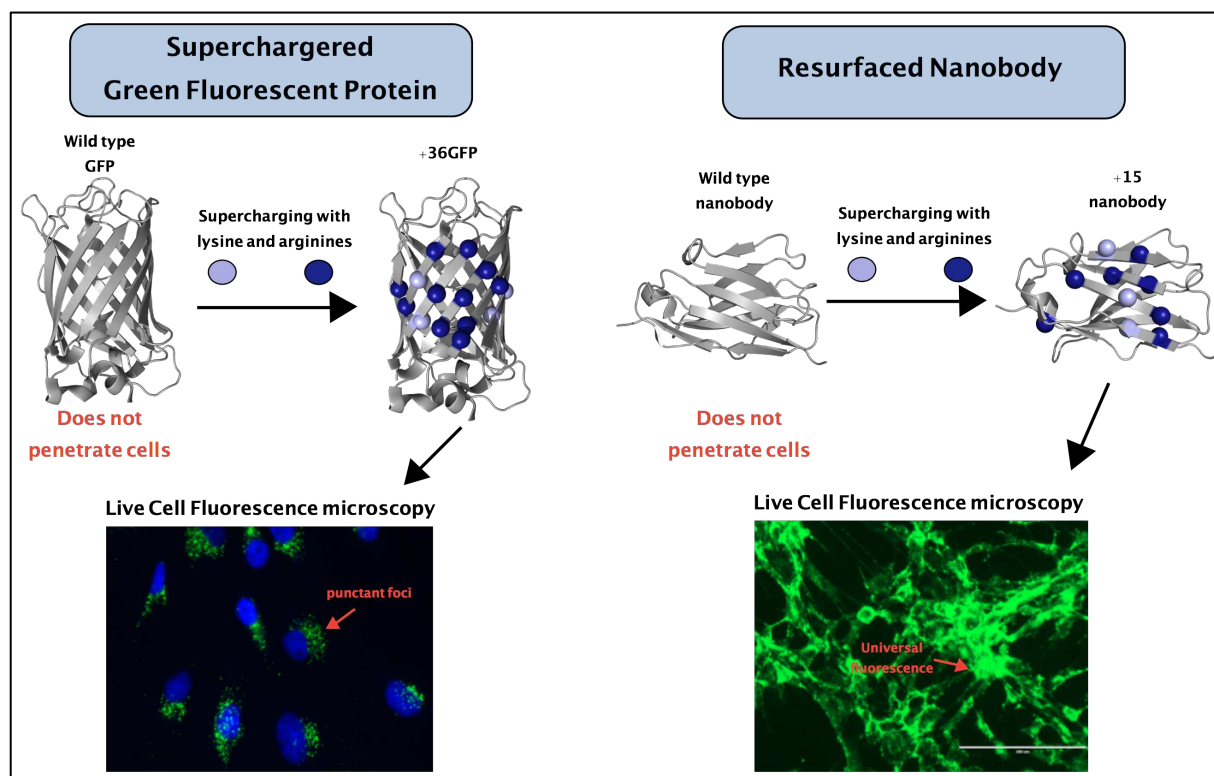


Figure 1.7: Supercharged GFP is able to enter cells, but the punctate foci observed in the fluorescence microscopy suggests that the protein remains trapped in the endosomes.⁸³ By applying the same cationic charging to nanobodies, the protein enters the cells and spreads throughout, suggesting the protein exists in the cytosol.⁸⁶

Recently, the Schepartz lab has engineered cell-permeant miniature proteins (CPMP) consisting of dense penta-arginine grafted regions on a hairpin like α -helical protein scaffold.⁷⁸ The hairpin like fold was an important feature in maintaining the stability of the miniature protein and protecting it from degradation, much like the method developed by the Tsein lab. These proteins, derived from avian polypeptide (aPP) and zinc finger domains, aPP5.3 and ZF5.3 were shown to deliver their approximately 250nM of cargo to cytosol exceeding efficiencies by 6-fold of other common CPPs.⁸⁷ Miniature proteins are likely the most efficient fusion agents for delivery of therapeutic proteins, however, cationic resurfaced nanobodies combine delivery and potential elimination of cargo proteins for therapeutic relevance. Nanobodies are small (15KDa) proteins derived from the variable heavy chain region of antibodies and are responsible for the binding of antigens. Antibodies can be created to bind virtually any disease relevant target and thus are the leading biologic on the drug market to date. However, their large (150KDa) size and need for post-translational modifications for biological activity have limited their ability to treat intracellular targets. Luckily, an array of recombinant nanobodies has been developed, including those that bind GFP, HER2 and β -lactamase. The McNaughton lab reconstructed the framework region of three different nanobodies net charge to +15, referred to as cationic resurfacing.⁸⁶ Unlike the +36GFP resurfacing, resurfaced nanobodies were shown to endow cellular penetration, maintain near wild-type binding affinity for its target, and demonstrated potential to access the cytosol (**Figure 1.7**).

1.4 Assays to detect internalization and protein-protein interactions

As has been portrayed, proteins play a vital role in the regulation of cellular processes such as growth, proliferation and apoptosis. Biologics that are successfully delivered to the cytosol must have ways of reporting back that this destination has been reached. Often, biological

therapeutics work by binding to a large surface of a disease-relevant protein and disrupting native activity and/or interactions with other biopolymers.¹⁹ Therefore, assays must be functional in the complex environment of a living cell, since they ensure that native PPIs identified are physiologically relevant, and new PPIs (involving a biological drug lead) operate in a cellular context. One of the biggest challenges are well-established assays capable of distinguishing the delivery of biologics trapped in the endosomes from the successful transfer to the cytosol.

1.4.1 Cre recombinase

Possibly the most commonly used mechanism implemented to detect the successful delivery of cytosolic proteins is the Cre recombinase-*loxP* reporter assay. Cre recombinase is an enzyme capable of catalyzing genetic modification, of genes between two Lox sites. The first use of the Cre/*loxP* reporter system demonstrated the transient activation of a luciferase gene in mammalian cells to create an inducible reporter assay for signal transduction pathways.⁸⁸ Exogenously administered Cre recombinase must enter the cell, escape the endosomes, and find its way in the nuclear envelope in order to catalyze the recombination and translation of a desired reporter gene. Therefore, Cre has been a commonly used cargo protein to analyze the efficiency of newly developed CPPs, miniature proteins, and other protein delivery methods. However, this means the system can require multiple fusions to a therapeutic cargo since Cre recombinase cannot enter mammalian cells on its own. One example of this was applied to discover new endosomalytic peptides using a +34GFP-cre-recombinase fusion. As mentioned previously, +34GFP penetrates cells at a much high potency than common CPPs, however, the vast majority of this protein and its fusions get trapped within the endosomes. Therefore, +34GFP was utilized to discover a new class of peptides specific for the endosomal release of exogenously delivered proteins.⁸⁹ By fusing +34GFP to cre recombinase, the successful delivery to the cytosol would

result in the translation of a tdTomato gene, a red fluorescent protein that can be imaged and quantified by flow cytometry.

1.4.2 GIGI/GIGT Assay

The GIGI/GIGT assay that was developed as an adaptation to a luciferase assay involving a steroid activator of the protein transduction across the nuclear envelope to turn on transcription and translation.⁹⁰ In short, the small molecule would passively diffuse across the cell membrane and access the cytosol. Here, the steroid would bind to a ligand binding domain (glucocorticoid receptor) of a protein that has been fused to DNA binding protein Gal4 and release chaperones that were preventing the fusion from crossing the nuclear envelope. The release of the chaperone proteins induced protein transduction into the nucleus where the Gal4 protein would bind the Gal4 recognition domain on the DNA and activate transcription and translation of the reporter protein luciferase.

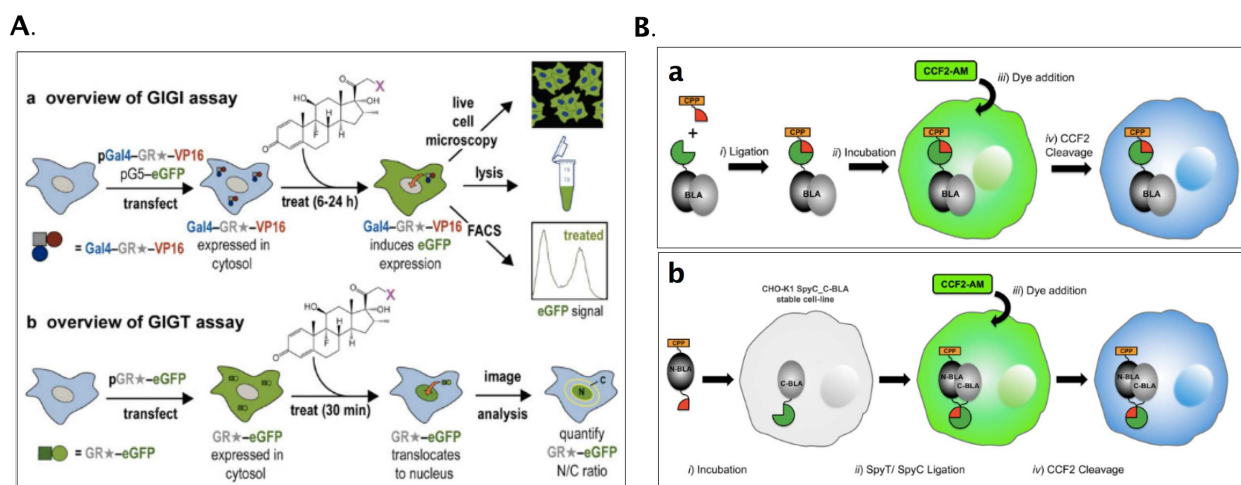


Figure 1.8: Reporter assays for cytosolic delivery of protein cargo. A. Overview of the GIGI/GIGT assay developed by the Schepartz lab⁹¹ B. Overview of the split b-Lactamase assay developed by the Milech lab.⁹²

The Schepartz lab widened the application of this assay by engineering the glucocorticoid receptor ligand binding domain to bind dex (the steroid) at a higher affinity such that dex tagged peptides and proteins would be recognized.⁹¹ The interaction between the dex-tagged peptides or proteins and the Gal4 protein fusion must occur within the cytosol of the cell in order to activate the production of the reporter gene (eGFP in this assay), thus fulfilling the first component of their assay. The GIGI component, glucocorticoid-induced eGFP induction, is an amplifiable assay that can inform of cytosolic residence of dex-tagged peptides and proteins without the need for imaging equipment for quantification. An amplified signal from the subsequent transcription and translation of a given gene is beneficial to detect proteins that have a low cytosolic delivery. The second component, GIGT, Glucocorticoid-induced eGFP Translocation, involves the transport of cytosolic eGFP glucocorticoid receptor fusion protein to the nucleus through interaction with Dex tagged molecules. In GIGI, the amplified translation of the eGFP would result in green fluorescence being observed in the cytosol of the cell (**Figure 1.8a**). However, in GIGT, fluorescence would be measured in the nucleus alone if translocation of the GFP fusion protein made it all the way to the nucleus. A ratio of nuclear fluorescence to cytosolic fluorescence would be measured by high content imaging for quantitation of delivery. GIGI/GIGT was used to analyze the difference of pent-arginine grafting upon α -helix peptides, proteins or peptide mimetics with relation to cytosolic access. Although highly quantitative, these assays require two complementing components, multiple plasmid transfections, synthetic conjugation of the protein or peptide of interest to Dex, and advanced fluorescence imaging technology is necessary to implement these methods.

1.4.3 Complimentary Split Reporter Assays for Cytosolic Detection

Split protein complementation assays have become a facile approach to observing protein-protein interactions *in vitro* and *in vivo*. A reporter protein such as GFP, luciferase, or in this case β -lactamase is split in to two stable halves and each half is fused to a protein of interest. The spontaneous recombination of the reporter protein is dependent upon their relative proximity, therefore, a signal is only detected when the two proteins of interest are interacting with one another. These split reporter assays have since been applied to internalization assays. Typically, one half of the reporter protein is transfected into the mammalian cells, while the smaller half is fused to the protein of interest and exogenously delivered. The meeting of the two halves within the cytosol catalyzes the complementation of the halves and produces the signal. The Dowdy lab used the split GFP reporter assay to discover a new class of endosomal escape domains, using the TAT peptide as the protein transduction domain.⁸⁹

β -lactamase is not a luminescent or fluorescent protein but rather a protein that provides antibiotic resistance to the cell in the presence of β -lactam antibiotics such as penicillin. In short, this two part assay first relies on the CPP-dependent internalization of the full β -lactamase protein and secondly the recombination of a split β -lactamase (**Figure 1.8b**).⁹² CCF2-AM is a compound that exhibits a native fluorescence through fluorescence resonance energy transfer (FRET) and is rapidly taken up by cells, de-esterified by cytoplasmic esterases and converted to CCF2. CCF2 then becomes trapped within the cytosol where β -lactamase is able to cause a cleavage reaction that is capable of shifting the FRET signal from a green to blue fluorescence. Thus, exogenously delivered β -lactamase can measure efficient CPP mediated delivery by the green to blue shift in CCF2 fluorescent signal. The second assay utilizes this technology, yet relies on the necessary internalization of one half of the β -lactamase before it is able to

recombine with the stably expressed second half in the cytosol to begin the cleavage reaction. This technology claims that the cleavage of the CCF2 molecule performs at a faster rate than that of split GFP reporter assay that require refolding of the protein before the fluorescence is detected. However, β -lactamase is a rather large protein by which adding more fusions to it increase the difficulty of expression. In addition, these bulky components only verify the successful delivery and does not inherently contain any additional protein of interest, therefore, it is unknown whether the bulk would interfere with the necessary interactions a biological therapeutic would need to perform.

1.5 Conclusion

Protein engineering methods continue to advance the ability to explore biology and all of its interactions. Through directed evolution, a native protein with a desired function can be subjected to over 10^6 mutational variants to either discover the most optimal activity or to completely alter its wildtype binder. Surprisingly, only the gene of interest and a function is necessary to drive this evolutionary pressure. If an experiment is fortunate enough to obtain a crystal structure, rational design of residues within a binding pocket, active site, on the surface and terminal ends of any protein of interest can be altered to induce new chemical abilities and function. Thanks to recombinant protein synthesis, these mutational libraries can be rapidly and easily produced in *E. Coli* for use in studies *in vitro*, or *in vivo* in yeast or mammalian cells,

Although mass mutational libraries of any protein can be made and analyzed, the natural 20 amino acids are limited in their chemical capabilities. Non-canonical amino acid incorporation mechanisms like amber suppression gave synthetic organic chemistry a biological purpose. Now, the biological activity of virtually any peptide, amino acid, protein mimetic, or ligations of any small molecule is possible in a site-specific manner. Non-canonical amino acids have been used

to analyze proteins in their native environment, observe protein folding, analyze new protein structures, and discover specific residues involved in PPIs.

One of the greatest efforts in recombinant design is the progression towards biological therapeutics. Small molecules have dominated the field due to their ability to cross the lipid bilayer of mammalian cells. Although the vast majority of disease and cancer are a result of protein-protein interactions that have gone awry, their size and generally hydrophobic character prevent them from accessing intracellular targets. However, the discovery of the TAT protein transduction domain spearheaded the development of other cell-penetrating peptides, protein mimetics and resurfaced proteins that efficiently deliver cargo to the cytosol of mammalian cells.

The advancement of protein therapeutics has brought another area of research in to question. The need for assays that function in a complex biological environment and sufficiently distinguish the delivery of exogenous protein to the cytosol over the endosomes. Macromolecules are almost always trafficked into the cell through micropinocytosis or receptor-mediated cytosol, both resulting in yet another membrane for the biologic to escape. The GIGI/GIGT and split reporter assays discussed are both able to detect the successful delivery of a biologic to the cytosol upon the translation or activation of a fluorescent signal. These signals would otherwise not exist if the cargo had been trapped in the endosomes or degraded within lysosomes. Yet there still remains a need for assays that do not require multiple fusion proteins or require high resolution fluorescent imaging equipment for quantitative analysis.

The combined aspects of protein engineering will be discussed in this work. First, the synthesis of a novel non-canonical amino acid and the library design of an aaRS will demonstrate the ability to alter a proteins function towards new biomaterials. Secondly, the

rational design of resurfaced nanobodies to endow cellular penetration of proteins that are capable of delivering therapeutic cargo, or embodying the therapeutic itself. And lastly, the design of a facile luminescent assay to detect successful cytosolic delivery of varying proteins.

REFERENCES

- (1) Khan, S.; Vihinen, M. Spectrum of Disease-Causing Mutations in Protein Secondary Structures. *BMC Struct. Biol.* **2007**, *7*, 1–18. <https://doi.org/10.1186/1472-6807-7-56>.
- (2) Sherman, M. Y.; Goldberg, A. L. Cellular Defenses against Unfolded Proteins. *Neuron* **2001**, *29* (1), 15–32. [https://doi.org/10.1016/s0896-6273\(01\)00177-5](https://doi.org/10.1016/s0896-6273(01)00177-5).
- (3) Omary, M. B.; Coulombe, P. A.; McLean, W. H. I. Intermediate Filament Proteins and Their Associated Diseases. *N. Engl. J. Med.* **2004**, *351* (20), 2087–2100. <https://doi.org/10.1056/NEJMra040319>.
- (4) Zügel, U.; Kaufmann, S. H. E. Role of Heat Shock Proteins in Protection from and Pathogenesis of Infectious Diseases. *Clin. Microbiol. Rev.* **1999**, *12* (1), 19–39.
- (5) Romero, Z.; Lomova, A.; Said, S.; Miggelbrink, A.; Kuo, C. Y.; Campo-Fernandez, B.; Hoban, M. D.; Masiuk, K. E.; Clark, D. N.; Long, J.; et al. Editing the Sickle Cell Disease Mutation in Human Hematopoietic Stem Cells: Comparison of Endonucleases and Homologous Donor Templates. *Mol. Ther.* **2019**, *125* (17), 2597–2605. <https://doi.org/10.1016/j.ymthe.2019.05.014>.
- (6) Wilkinson, A. J.; Fersht, A. R.; Blow, D. M.; Carter, P.; Winter, G. A Large Increase in Enzyme-Substrate Affinity by Protein Engineering. *Nature* **1984**, *307* (5947), 187–188. <https://doi.org/10.1038/307187a0>.
- (7) Harris, L. Enzyme Specificity.
- (8) Cella, F.; Wroblewska, L.; Weiss, R.; Siciliano, V. Engineering Protein-Protein Devices for Multilayered Regulation of mRNA Translation Using Orthogonal Proteases in Mammalian Cells. *Nat. Commun.* **2018**, *9* (1). <https://doi.org/10.1038/s41467-018-06825-7>.
- (9) Link, C. Engineering a Direct and Inducible Protein-RNA Interaction To Regulate RNA Biology Accessed. **2015**, *5* (9), 851–861.
- (10) Tobin, P. H.; Richards, D. H.; Callender, R. A.; Wilson, C. J.; Haven, N.; Haven, N.; Haven, N. Protein Engineering: A New Frontier for Biological Therapeutics. *HHS Public Access* **2016**, *15* (7), 743–756.
- (11) Cornish, V. W.; Mendel, D.; Schultz, P. G. Probing Protein Structure and Function with an Expanded Genetic Code. *Angew. Chemie Int. Ed. English* **1995**, *34* (6), 621–633. <https://doi.org/10.1002/anie.199506211>.
- (12) Mou, Y.; Yu, J. Y.; Wannier, T. M.; Guo, C. L.; Mayo, S. L. Computational Design of Co-Assembling Protein-DNA Nanowires. *Nature* **2015**, *525* (7568), 230–233. <https://doi.org/10.1038/nature14874>.
- (13) Nuñez, J. K.; Harrington, L. B.; Doudna, J. a. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chem. Biol.* **2016**, *acschembio.5b01019*. <https://doi.org/10.1021/acschembio.5b01019>.

- (14) Lutolf, M. P.; Hubbell, J. A. Synthetic Biomaterials as Instructive Extracellular Microenvironments for Morphogenesis in Tissue Engineering. *Nat. Biotechnol.* **2005**, *23* (1), 47–55. <https://doi.org/10.1038/nbt1055>.
- (15) Zhao, J.; Kardashliev, T.; Joëlle Ruff, A.; Bocola, M.; Schwaneberg, U. Lessons from Diversity of Directed Evolution Experiments by an Analysis of 3,000 Mutations. *Biotechnol. Bioeng.* **2014**, *111* (12), 2380–2389. <https://doi.org/10.1002/bit.25302>.
- (16) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16* (7), 379–394. <https://doi.org/10.1038/nrg3927>.
- (17) Lutz, S. Beyond Directed Evolution - Semi-Rational Protein Engineering and Design. **2010**, *21* (6), 734–743. <https://doi.org/10.1016/j.copbio.2010.08.011>.Beyond.
- (18) Lehmann, M.; Wyss, M. Engineering Proteins for Thermostability : The Use of Sequence Alignments versus Rational Design and Directed Evolution. 371–375.
- (19) Xing, S.; Wallmeroth, N.; Berendzen, K. W.; Grefen, C. Techniques for the Analysis of Protein-Protein Interactions in Vivo. *Plant Physiol.* **2016**, *171* (2), 727–758. <https://doi.org/10.1104/pp.16.00470>.
- (20) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V; Kopka, M. L. The Anatomy of A- , Bo- , and Z-DNA. **1960**, *0*.
- (21) Cobb, M. 60 Years Ago, Francis Crick Changed the Logic of Biology. *PLoS Biol.* **2017**, *15* (9), 1–8. <https://doi.org/10.1371/journal.pbio.2003243>.
- (22) Crick, F. H. C. On Protein Synthesis. *Med. Res. Counc. Unit Study Mol. Biol.* **1958**, *12*, 138–163.
- (23) Cookson, N. A.; Cookson, S. W.; Tsimring, L. S.; Hasty, J. Cell Cycle-Dependent Variations in Protein Concentration. *Nucleic Acids Res.* **2009**, *38* (8), 2676–2681. <https://doi.org/10.1093/nar/gkp1069>.
- (24) Crick, F. H. C.; Barnett, L.; BRENNER, S.; Watts-tobin, R. J. General Nature of the Genetic Code for Proteins. *Nature* **1961**, *192*, 1227–1232.
- (25) Agris, P. F. Decoding the Genome: A Modified View. *Nucleic Acids Res.* **2004**, *32* (1), 223–238. <https://doi.org/10.1093/nar/gkh185>.
- (26) Francois, B. Recombinant Protein Expression in Escherichia Coli. *Curr. Opin. Biotechnol.* **1999**, *60*, 411–421.
- (27) Khan, S.; Ullah, M. W.; Siddique, R.; Nabi, G.; Manan, S.; Yousaf, M.; Hou, H. Role of Recombinant DNA Technology to Improve Life. *Int. J. Genomics* **2016**, *2016*. <https://doi.org/10.1155/2016/2405954>.
- (28) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79* (March), 413–444. <https://doi.org/10.1146/annurev.biochem.052308.105824>.
- (29) Wang, L.; Xie, J.; Schultz, P. G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.

<https://doi.org/10.1146/annurev.biophys.35.101105.121507>.

- (30) Furter, R. Expansion of the Genetic Code: Site-Directed p-Fluoro-Phenylalanine Incorporation in Escherichia Coli. *Protein Sci.* **2008**, 7 (2), 419–426. <https://doi.org/10.1002/pro.5560070223>.
- (31) Koonin, E.; Novozhilov, A. Origin and Evolution of the Genetic Code: The Universal Enigma. *IUBMB Life* **2009**, 61 (2), 99–111. <https://doi.org/10.1002/iub.146.Origin>.
- (32) Kwon, I.; Kirshenbaum, K.; Tirrell, D. a. Breaking the Degeneracy of the Genetic Code. *J. Am. Chem. Soc.* **2003**, 125 (25), 7512–7513. <https://doi.org/10.1021/ja0350076>.
- (33) Cirino, P. C.; Tang, Y.; Takahashi, K.; Tirrell, D. A.; Arnold, F. H. Global Incorporation of Norleucine in Place of Methionine in Cytochrome P450 BM-3 Heme Domain Increases Peroxygenase Activity. *Biotechnol. Bioeng.* **2003**, 83 (6), 729–734. <https://doi.org/10.1002/bit.10718>.
- (34) Reay, P. A.; Kantor, R. M.; Davis, M. M. Use of Global Amino Acid Replacements to Define the Requirements for MHC Binding and T Cell Recognition of Moth Cytochrome c (93-103). **2019**.
- (35) Van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *J. Am. Chem. Soc.* **2000**, 122 (7), 1282–1288. <https://doi.org/10.1021/ja992749j>.
- (36) Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. Y. Identification of an Expanded Set of Translationally Active Methionine Analogues in Escherichia Coli. *FEBS Lett.* **2001**, 505, 25256.
- (37) Merkel, L.; Schauer, M.; Antranikian, G.; Budisa, N. Parallel Incorporation of Different Fluorinated Amino Acids: On the Way to “Teflon” Proteins. *ChemBioChem* **2010**, 11 (11), 1505–1507. <https://doi.org/10.1002/cbic.201000295>.
- (38) Athey, J.; Alexaki, A.; Osipova, E.; Rostovtsev, A.; Santana-Quintero, L. V.; Katneni, U.; Simonyan, V.; Kimchi-Sarfaty, C. A New and Updated Resource for Codon Usage Tables. *BMC Bioinformatics* **2017**, 18 (1), 1–10. <https://doi.org/10.1186/s12859-017-1793-7>.
- (39) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science (80-.)*. **2001**, 292 (5516), 498–500. <https://doi.org/10.1126/science.1060077>.
- (40) Jolla, L. Progress toward the Evolution of an Organism with an Expanded. **1999**, 96 (April), 4780–4785.
- (41) Chatterjee, A.; Xiao, H.; Schultz, P. G. Evolution of Multiple, Mutually Orthogonal Prolyl-TRNA Synthetase/TRNA Pairs for Unnatural Amino Acid Mutagenesis in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109 (37), 14841–14846. <https://doi.org/10.1073/pnas.1212454109>.
- (42) Neumann, H.; Slusarczyk, A. L.; Chin, J. W. De Novo Generation of Mutually Orthogonal Aminoacyl-TRNA Synthetase/ TRNA Pairs. *J. Am. Chem. Soc.* **2010**, 132 (7), 2142–2144.

<https://doi.org/10.1021/ja9068722>.

- (43) Pastrnak, M.; Magliery, T. J.; Schultz, P. G. A New Orthogonal Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. *Helv. Chim. Acta* **2000**, *83* (9), 2277–2286. [https://doi.org/10.1002/1522-2675\(20000906\)83:9<2277::AID-HLCA2277>3.0.CO;2-L](https://doi.org/10.1002/1522-2675(20000906)83:9<2277::AID-HLCA2277>3.0.CO;2-L).
- (44) Kast, P.; Hennecke, H. Amino Acid Substrate Specificity of Escherichia Coli Phenylalanyl-tRNA Synthetase Altered by Distinct Mutations. *J. Mol. Biol.* **1991**, *222* (1), 99–124. [https://doi.org/10.1016/0022-2836\(91\)90740-W](https://doi.org/10.1016/0022-2836(91)90740-W).
- (45) Cload, S. T.; Liu, D. R.; Froland, W. A.; Schultz, P. G. Development of Improved tRNAs for in Vitro Biosynthesis of Proteins Containing Unnatural Amino Acids. *Chem. Biol.* **1996**, *3* (12), 1033–1038. [https://doi.org/10.1016/S1074-5521\(96\)90169-6](https://doi.org/10.1016/S1074-5521(96)90169-6).
- (46) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. A New Functional Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins [16]. *J. Am. Chem. Soc.* **2000**, *122* (20), 5010–5011. <https://doi.org/10.1021/ja000595y>.
- (47) Wang, L.; Brock, A.; Schultz, P. G. Adding L-3-(2-Naphthyl)Alanine to the Genetic Code of E. Coli. *J. Am. Chem. Soc.* **2002**, *124* (9), 1836–1837. <https://doi.org/10.1021/ja012307j>.
- (48) Fechter, P.; Rudinger-Thirion, J.; Tukalo, M.; Giegé, R. Major Tyrosine Identity Determinants in Methanococcus Jannaschii and Saccharomyces Cerevisiae tRNA^{Tyr} Are Conserved but Expressed Differently. *Eur. J. Biochem.* **2001**, *268* (3), 761–767. <https://doi.org/10.1046/j.1432-1327.2001.01931.x>.
- (49) Biddle, W.; Schmitt, M. A.; Fisk, J. D. Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen. *Biochemistry* **2015**, *54* (50), 7355–7364. <https://doi.org/10.1021/acs.biochem.5b00870>.
- (50) Wang, L.; Schultz, P. G. A General Approach for the Generation of Orthogonal tRNAs. *Chem. Biol.* **2001**, *8* (9), 883–890. [https://doi.org/10.1016/S1074-5521\(01\)00063-1](https://doi.org/10.1016/S1074-5521(01)00063-1).
- (51) Guo, J.; Melançon, C.; Lee, H. S.; Groff, D.; Schultz, P. Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angew. Chemie* **2009**, *121* (48), 9312–9315. <https://doi.org/10.1002/ange.200904035>.
- (52) Schmitt, M. A.; Biddle, W.; Fisk, J. D. Mapping the Plasticity of the Escherichia Coli Genetic Code with Orthogonal Pair-Directed Sense Codon Reassignment. *Biochemistry* **2018**, *57* (19), 2762–2774. <https://doi.org/10.1021/acs.biochem.8b00177>.
- (53) Rozov, A.; Demeshkina, N.; Khusainov, I.; Westhof, E.; Yusupov, M.; Yusupova, G. Novel Base-Pairing Interactions at the tRNA Wobble Position Crucial for Accurate Reading of the Genetic Code. *Nat. Commun.* **2016**, *7*, 1–10. <https://doi.org/10.1038/ncomms10457>.
- (54) de la Torre, B. G.; Albericio, F. The Pharmaceutical Industry in 2018. An Analysis of

- FDA Drug Approvals from the Perspective of Molecules. *Molecules* **2019**, 2017. <https://doi.org/10.3390/molecules24040809>.
- (55) Guan, K.; Dixon, J. E. Eukaryotic Proteins Expressed in Escherichia Coli: An Improved Thrombin Cleavage and Purification Procedure of Fusion Proteins with Glutathione S-Transferase. *Anal. Biochem.* **1991**, 192 (2), 262–267. [https://doi.org/10.1016/0003-2697\(91\)90534-Z](https://doi.org/10.1016/0003-2697(91)90534-Z).
 - (56) Waugh, D. S. An Overview of Enzymatic Reagents for the Removal of Affinity Tags. *Protein Expr. Purif.* **2011**, 80 (2), 283–293. <https://doi.org/10.1016/j.pep.2011.08.005>.
 - (57) Strohl, W. R. Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters. *BioDrugs* **2015**, 29 (4), 215–239. <https://doi.org/10.1007/s40259-015-0133-6>.
 - (58) van den Berg, A.; Dowdy, S. F. Protein Transduction Domain Delivery of Therapeutic Macromolecules. *Curr. Opin. Biotechnol.* **2011**, 22 (6), 888–893. <https://doi.org/10.1016/j.copbio.2011.03.008>.
 - (59) Baeshen, N. A.; Baeshen, M. N.; Sheikh, A.; Bora, R. S.; Ahmed, M. M. M.; Ramadan, H. A. I.; Saini, K. S.; Redwan, E. M. Cell Factories for Insulin Production. *Microb. Cell Fact.* **2014**, 13 (1), 1–9. <https://doi.org/10.1186/s12934-014-0141-0>.
 - (60) Su, W. W. Fluorescent Protein as Tools to Aid Protein Production. *Microb. Cell Fact.* **2005**, 4, 1–6. <https://doi.org/10.1186/1475-2859-4-12>.
 - (61) Suzuki, T.; Ishii-Watabe, A.; Tada, M.; Kobayashi, T.; Kanayasu-Toyoda, T.; Kawanishi, T.; Yamaguchi, T. Importance of Neonatal FcR in Regulating the Serum Half-Life of Therapeutic Proteins Containing the Fc Domain of Human IgG1: A Comparative Study of the Affinity of Monoclonal Antibodies and Fc-Fusion Proteins to Human Neonatal FcR. *J. Immunol.* **2010**, 184 (4), 1968–1976. <https://doi.org/10.4049/jimmunol.0903296>.
 - (62) Morgenstern, J.; Gil Alvaradejo, G.; Bluthardt, N.; Beloqui, A.; Delaittre, G.; Hubbuch, J. Impact of Polymer Bioconjugation on Protein Stability and Activity Investigated with Discrete Conjugates: Alternatives to PEGylation. *Biomacromolecules* **2018**, 19 (11), 4250–4262. <https://doi.org/10.1021/acs.biomac.8b01020>.
 - (63) Doherty, G. J.; McMahon, H. T. Mechanisms of Endocytosis. *Annu. Rev. Biochem.* **2009**, 78 (March), 857–902. <https://doi.org/10.1146/annurev.biochem.78.081307.110540>.
 - (64) Elkin, S.; Lakoduk, A.; Schmid, S. Endocytic Pathways and Endosomal Trafficking: A Primer. *Wien Med Wochenschr* **2016**, 166 (Clic), 196–204. <https://doi.org/10.1016/j.physbeh.2017.03.040>.
 - (65) Rabideau, A. E.; Pentelute, B. L. Delivery of Non-Native Cargo into Mammalian Cells Using Anthrax Lethal Toxin. *ACS Chem. Biol.* **2016**, 11 (6), 1490–1501. <https://doi.org/10.1021/acschembio.6b00169>.
 - (66) Liao, X.; Rabideau, A. E.; Pentelute, B. L. Delivery of Antibody Mimics into Mammalian Cells via Anthrax Toxin Protective Antigen. *ChemBioChem* **2014**, 15 (16), 2458–2466. <https://doi.org/10.1002/cbic.201402290>.
 - (67) Bruce, V. J.; McNaughton, B. R. Inside Job: Methods for Delivering Proteins to the

- Interior of Mammalian Cells. *Cell Chem. Biol.* **2017**, *24* (8), 924–934. <https://doi.org/10.1016/j.chembiol.2017.06.014>.
- (68) Löfblom, J.; Feldwisch, J.; Tolmachev, V.; Carlsson, J.; Ståhl, S.; Frejd, F. Y. Affibody Molecules: Engineered Proteins for Therapeutic, Diagnostic and Biotechnological Applications. *FEBS Lett.* **2010**, *584* (12), 2670–2680. <https://doi.org/10.1016/j.febslet.2010.04.014>.
 - (69) Mann, D. A.; Frankel, A. D. Endocytosis and Targeting of Protein HIV-1 Tat. *EMBO J.* **1991**, *10* (7), 1733–1739.
 - (70) Gump, J. M.; Dowdy, S. F. TAT Transduction: The Molecular Mechanism and Therapeutic Prospects. *Trends Mol. Med.* **2007**, *13* (10), 443–448. <https://doi.org/10.1016/j.molmed.2007.08.002>.
 - (71) Eguchi, A.; Akuta, T.; Okuyama, H.; Senda, T.; Yokoi, H.; Inokuchi, H.; Fujita, S.; Hayakawa, T.; Takeda, K.; Hasegawa, M.; et al. Protein Transduction Domain of HIV-1 Tat Protein Promotes Efficient Delivery of DNA into Mammalian Cells. *J. Biol. Chem.* **2001**, *276* (28), 26204–26210. <https://doi.org/10.1074/jbc.M010625200>.
 - (72) Montrose, K.; Yang, Y.; Sun, X.; Wiles, S.; Krissansen, G. W. Xentry, a New Class of Cell-Penetrating Peptide Uniquely Equipped for Delivery of Drugs. *Sci. Rep.* **2013**, *3*, 1–7. <https://doi.org/10.1038/srep01661>.
 - (73) Ponnappan, N.; Chugh, A. Cell-Penetrating and Cargo-Delivery Ability of a Spider Toxin-Derived Peptide in Mammalian Cells. *Eur. J. Pharm. Biopharm.* **2017**, *114*, 145–153. <https://doi.org/10.1016/j.ejpb.2017.01.012>.
 - (74) Joliot, A.; Pernellet, C.; Deagostini-Bazint, H.; Prochiantz, A. Antennapedia Homeobox Peptide Regulates Neural Morphogenesis (Neuron/Development/Culture). *Proc. Natl. Acad. Sci. USA* **1991**, *88* (March), 1864–1868.
 - (75) Fuchs, S. M.; Raines, R. T. Arginine Grafting to Endow Cell Permeability. *ACS Chem. Biol.* **2007**, *2* (3), 167–170. <https://doi.org/10.1021/cb600429k>.
 - (76) Calnan, B. J.; Biancalana, S.; Hudson, D.; Frankel, A. D. Analysis of Arginine-Rich Peptides from the HIV Tat Protein Reveals Unusual Features of RNA-Protein Recognition. *Genes Dev.* **1991**, *5* (2), 201–210. <https://doi.org/10.1101/gad.5.2.201>.
 - (77) Kim, D. T.; Fathman, C. G.; Rothbard, J. B.; Steinman, L.; Mitchell, D. J. Polyarginine Enters Cells More Efficiently than Other Polycationic Homopolymers. *J. Pept. Res.* **2002**, *56* (5), 318–325. <https://doi.org/10.1034/j.1399-3011.2000.00723.x>.
 - (78) Appelbaum, J. S.; Larochelle, J. R.; Smith, B. A.; Balkin, D. M.; Holub, J. M.; Schepartz, A. Arginine Topology Controls Escape of Minimally Cationic Proteins from Early Endosomes to the Cytoplasm. *Chem. Biol.* **2012**, *19* (7), 819–830. <https://doi.org/10.1016/j.chembiol.2012.05.022>.
 - (79) Habault, J.; Poyet, J. L. Recent Advances in Cell Penetrating Peptide-Based Anticancer Therapies. *Molecules* **2019**, 1–17. <https://doi.org/10.3390/molecules24050927>.
 - (80) Aguilera, T. A.; Olson, E. S.; Timmers, M. M.; Jiang, T.; Tsien, R. Y. Systemic in Vivo

- Distribution of Activatable Cell Penetrating Peptides Is Superior to That of Cell Penetrating Peptides. *Integr. Biol.* **2009**, *1* (5–6), 371–381. <https://doi.org/10.1039/b904878b>.
- (81) Hyun, S.; Choi, Y.; Lee, H. N.; Lee, C.; Oh, D.; Lee, D. K.; Lee, C.; Lee, Y.; Yu, J. Construction of Histidine-Containing Hydrocarbon Stapled Cell Penetrating Peptides for: In Vitro and in Vivo Delivery of SiRNAs. *Chem. Sci.* **2018**, *9* (15), 3820–3827. <https://doi.org/10.1039/c8sc00074c>.
 - (82) Lawrence, M. S.; Phillips, K. J.; Liu, D. R. Supercharging Proteins Can Impart Unusual Resilience NIH Public Access. *J Am Chem Soc* **2007**, *129* (33), 10110. <https://doi.org/10.1021/ja071641y>.
 - (83) Liu, D.; Cronican, J. J.; Thompson, D. B. Engineering and Identifying Supercharged Proteins for Macromolecule Delivery into Mammalian Cells. *Methods Enzym.* **2012**, *503*, 293–319. <https://doi.org/10.1038/jid.2014.371>.
 - (84) Laber, J. R.; Dear, B. J.; Martins, M. L.; Jackson, D. E.; Divenere, A.; Gollihar, J. D.; Ellington, A. D.; Truskett, T. M.; Johnston, K. P.; Maynard, J. A. Charge Shielding Prevents Aggregation of Supercharged GFP Variants at High Protein Concentration. *Mol. Pharm.* **2017**, *14* (10), 3269–3280. <https://doi.org/10.1021/acs.molpharmaceut.7b00322>.
 - (85) Cronican, J. J.; Thompson, D. B.; Beier, K. T.; McNaughton, B. R.; Cepko, C. L.; Liu, D. R. Potent Delivery of Functional Proteins into Mammalian Cells in Vitro and in Vivo Using a Supercharged Protein. *ACS Chem. Biol.* **2010**, *5* (8), 747–752. <https://doi.org/10.1021/cb1001153>.
 - (86) Bruce, V. J.; Lopez-Islas, M.; McNaughton, B. R. Resurfaced Cell-Penetrating Nanobodies: A Potentially General Scaffold for Intracellularly Targeted Protein Discovery. *Protein Sci.* **2016**, *25* (6), 1129–1137. <https://doi.org/10.1002/pro.2926>.
 - (87) Wissner, R. F.; Steinauer, A.; Knox, S. L.; Thompson, A. D.; Schepartz, A. Fluorescence Correlation Spectroscopy Reveals Efficient Cytosolic Delivery of Protein Cargo by Cell-Permeant Miniature Proteins. *ACS Cent. Sci.* **2018**, *4* (10), 1379–1393. <https://doi.org/10.1021/acscentsci.8b00446>.
 - (88) Mattheakis, L. C.; Olivan, S. E.; Dias, J. M.; Northrop, J. P. Expression of Cre Recombinase as a Reporter of Signal Transduction in Mammalian Cells. *Chem. Biol.* **1999**, *6* (11), 835–844. [https://doi.org/10.1016/S1074-5521\(99\)80130-6](https://doi.org/10.1016/S1074-5521(99)80130-6).
 - (89) Lönn, P.; Kacsinta, A. D.; Cui, X. S.; Hamil, A. S.; Kaulich, M.; Gogoi, K.; Dowdy, S. F. Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **2016**, *6* (August), 1–9. <https://doi.org/10.1038/srep32301>.
 - (90) Yu, P.; Liu, B.; Kodadek, T. A High-Throughput Assay for Assessing the Cell Permeability of Combinatorial Libraries. *Nat. Biotechnol.* **2005**, *23* (6), 746–751. <https://doi.org/10.1038/nbt1099>.
 - (91) Holub, J. M.; LaRochelle, J. R.; Appelbaum, J. S.; Schepartz, A. Improved Assays for Determining the Cytosolic Access of Peptides, Proteins, and Their Mimetics. *Biochemistry* **2013**, *52* (1), 9036–9046. <https://doi.org/10.1038/jid.2014.371>.

- (92) Stone, S. R.; Heinrich, T.; Juraja, S. M.; Satiaputra, J. N.; Hall, C. M.; Anastasas, M.; Mills, A. D.; Chamberlain, C. A.; Winslow, S.; Priebatsch, K.; et al. B-Lactamase Tools for Establishing Cell Internalization and Cytosolic Delivery of Cell Penetrating Peptides. *Biomolecules* **2018**, 8 (3). <https://doi.org/10.3390/biom8030051>.

CHAPTER TWO

SYNTHESIS OF NON-CANONICAL AMINO ACID BITHIOPHENE TOWARDS PROTEINS WITH POLYMER LIKE QUALITIES

2.1 Overview

Historically, there has been a dual interest in the synthesis and application of non-canonical amino acids. Organic chemists used these quaternary structures as targets to develop asymmetric methodologies and synthetic routes that analyze and enhance stereoselective reaction mechanisms. This interest has since merged with protein engineering efforts to evolve and accelerate protein performance and chemical advancement in biological systems. The amber codon suppression system was the first to incorporate synthetically made amino acids *in vivo* into proteins in a site-selective manner.^{1,2} Since then, the Schultz lab discovered a tyrosyl tRNA and tRNA synthetase pair from the organism *Methanococcus Jannaschii* that can efficiently incorporate ncAAs with a variety of novel side chains into proteins using an *E. coli* host.³ Of particular interest to us was their published crystal structure of the tyrosine aaRS engineered to bind a bipyridylalanine amino acid.⁴⁻⁶ We sought to apply rational design to this enzyme in hopes of binding our novel bithiophene amino acid, a ncAA which bears a biaryl side chain with similar structural composer and chemistry to bipyridylalanine. Thiophene is a heterocycle known polymerize through the 2-position and has been widely used for its fluorescent, luminescent and organic semi-conductive properties towards biomaterials and other biological sensing applications.⁷⁻¹¹ Here, we will show the successful synthesis of a protected racemic bithiophene ncAA in a 15% total yield.

2.2 Introduction

Biology has evolved from the collaboration of interactive chemical compounds and macromolecules known as deoxyribonucleic acids (DNA), ribonucleic acids (RNA), and amino acids that create complex proteins.^{12,13} The central dogma describes how these three components work together to produce every cellular function and fate of life. The genome of an organism provides the DNA template for any particular gene encoding a protein. This gene is then transcribed into messenger RNA, which is translated into a polypeptide of amino acids that fold into a protein.¹⁴ The genetic code is a map created from the observed relationship between the triplicate combinations (known as codons) of the four DNA nucleotides, and the corresponding amino acids for the templated construction of every protein. The 20-22 natural amino acids possess the same quaternary structural motif, an α -carbon with a tertiary amine, a carboxylic acid, an α -hydrogen and the amino acid side chain (R group) as seen in **Figure 2.1**.^{15,16} The primary amine and carboxylic acid are necessary for ligation into a polypeptide chain, creating the backbone of every protein. The amino acid side chains are the single component that varies between the natural 20 building blocks of life. Folding of each protein, macromolecular interactions with other proteins, DNA or RNA molecules, and even the binding pockets of enzymes and tRNA synthetase are driven by the chemical interactions between these side chains.¹⁷⁻²⁰ Protein engineering methodologies have utilized the natural toolbox of amino acid side chains to evolve proteins with designer capabilities and functions that stray far from their wild type callings. However, this chemical variety is limited to acid-base chemistry between charged side chains, covalent bonds through disulfide bridges, polar amino acids to provide hydrogen-bonding, and hydrophobic side chains that provide van der Waals interactions. Perhaps

a more exciting realm of designer proteins would be the introduction of side chains with nearly unlimited chemical functionality beyond the natural 20.

Although the effort to synthesize non-proteinogenic amino acids is not a new one, dating back to the 1960's,²¹ the Schultz lab pioneered the *in vivo* incorporation of non-canonical amino acids into proteins and has since incorporated over 150 ncAA's.²² Of these, ncAAs have allowed for control of post-translational modifications, intracellular imaging and trafficking of proteins, and side chains that probe protein structure and functionality (**Figure 2.1**).^{2,23–26} Their wildly successful orthogonal *Mj*Tyr aaRS and tRNA system for *in vivo* incorporation has been proven to accept multiple aromatic residues and structural analogs. Recently, the synthesis and incorporation of bipyridylalanine caught our attention due to the co-crystallization of the non-canonical amino acid with the engineered bipyridylalanine aaRS.²⁷ Due to this generous amount of information; we sought to add to growing protein functionality by designing amino acids with side chains that possess polymerization characteristics. To do this, we attempted well known amino acid synthetic routes to create a bithiophene amino acid and subsequently evolve the orthogonal *Mj*Tyr aaRS and tRNA for incorporation into proteins of interest.

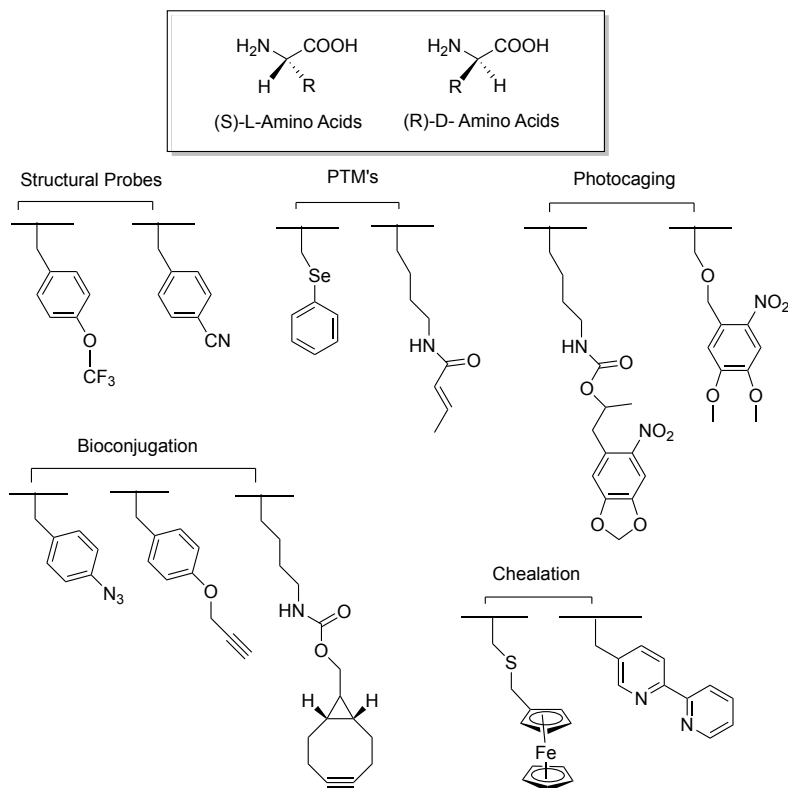


Figure 2.1: Various non-canonical amino acid side chains with complex functionalities that have been incorporated for the desired purpose stated above.

Synthetic routes that provide stereoselectivity and enantiomeric purity of amino acids are of utmost importance. (L)-amino acids are the natural configuration, whereas the (D)-configuration is rarely accepted into the binding pocket of wild type enzymes (such as tRNA synthetases.)²⁸ Cyclic and acyclic asymmetric systems have been developed to direct the optical purity of amino acid syntheses and have become suitable for production of ncAAs with challenging functionality of side chain structures.^{29–33} Here, we report the synthetic efforts towards the construction of bithiophene side chains utilizing two major acyclic synthetic routes; transition metal cross coupling of an established β -carbon, and use of a phase transfer catalyst.

2.3 Synthetic Plan

To build our non-canonical bithiophene amino acid we will break our synthesis into two pathways; transition metal cross coupling in order to utilize the pre-existing stereocenter of the serine amino acid, or by nucleophilic addition of the bithiophene unit to afford a racemic amino acid through phase transfer catalysis. Using natural amino acids as starting materials, cross-coupling reactions can attach the new side chain to the existing β -carbon of alanine or its derivatives. All transition-metal cross-coupling reactions proceed through the same general mechanism, regardless of which transition metal catalyst or organometallic is used.^{34,35} In our synthesis, we will utilize the Kumada-Corriu reaction to create our bithiophene unit (**R1**) that will be used in all succeeding synthetic routes.³⁶ Following the creation of our bithiophene unit, we will couple our zinc iodide protected serine to the side chain using Negishi coupling methodology (**Figure 2.2**).

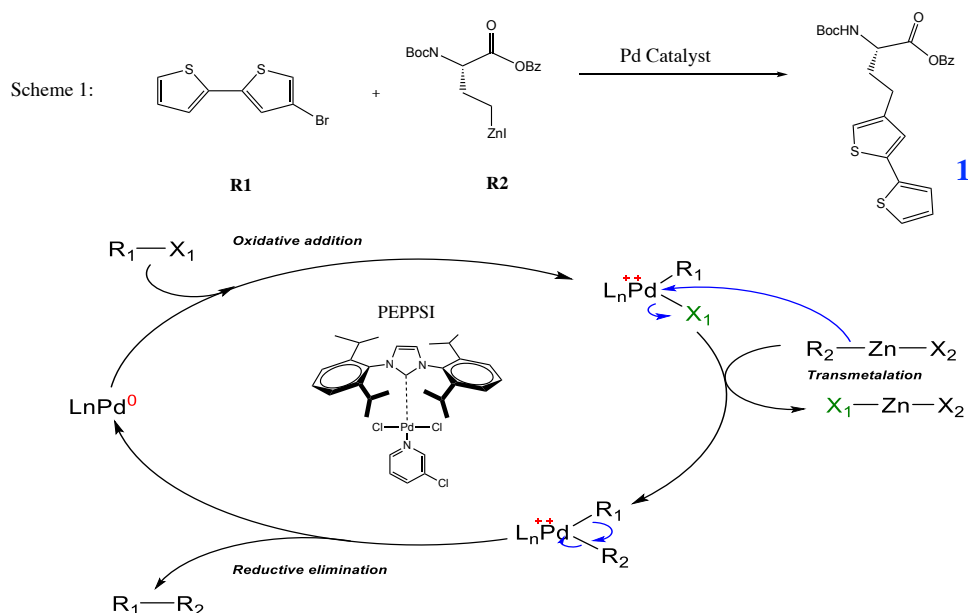


Figure 2.2: Transition metal cross coupling mechanism. PEPPSI will be utilized as the Pd catalyst in the coupling between the protected serine substrate and the bithiophene side chain motif

Simultaneously, we will explore phase-transfer catalysis pathways to utilize the inexpensive, readily available starting materials to create an acyclic Schiff base protected glycine. Phase-transfer catalysts are quaternary ammonium cations that transfer the reacting components across a biphasic barrier as depicted in **Figure 2.3**.³⁷ Therefore, these reactions do not need to proceed under an oxygen free environment, drastically increasing the ease of production; however, stereoselectivity is compromised and resulting racemic amino acids will need further enantiomeric purification. In 1978 O'Donnell began expanding this system for the synthesis of natural and non-canonical amino acids. His group condensed benzophenone and glycine derivatives to make the Schiff bases **2** and **6** that can be deprotonated by an aqueous base, alkylated by the desired side chain and undergo acid hydrolysis to expose the deprotected amino acid.^{37,38} Here, we will synthesize our bithiophene ncAA using benzyltriethylammonium chloride (BTEAC) as our phase transfer catalyst as shown in **Figure 2.3**.

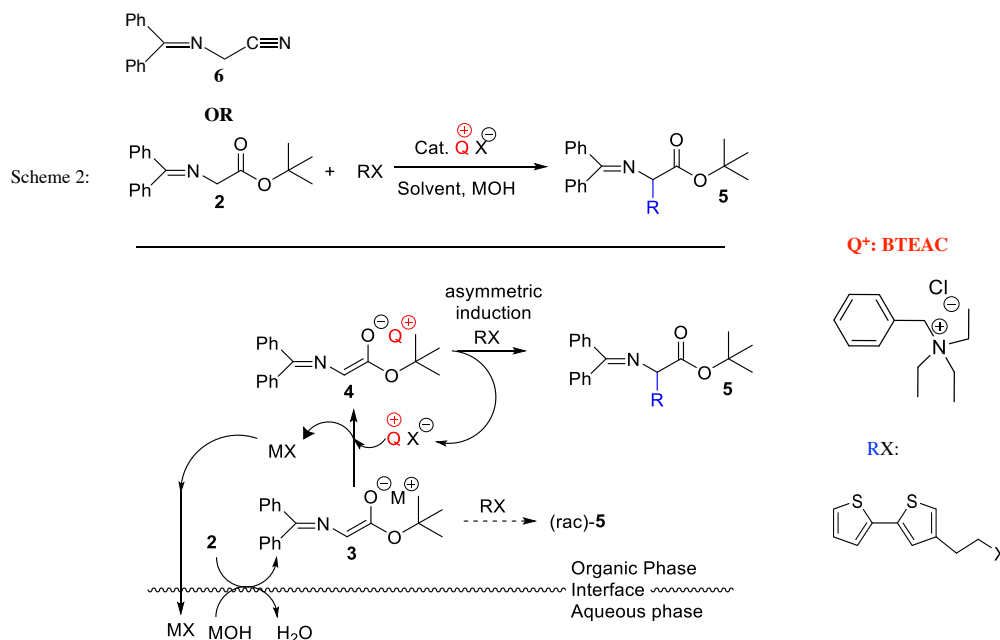


Figure 2.3: Phase Transfer Catalysis mechanism using our desired substrates and BTEAC as the catalyst to produce a racemic amino acid.

2.4 Results and Discussion

2.4.1 Transition Metal Cross Coupling

Nickel and palladium have been the most widely used for the construction of the catalyst with varying ligands to increase the rate of oxidative addition and transmetallation.^{39,40} Kumada and Corriu initially broadened the scope of aromatic couplings using nickel catalysts but shortly shifted to palladium catalysts due to their increased tolerance for a variety of organometallic coupling substrates, a more rapid transmetallation step, higher yields and better stereoselectivity compared to nickel.

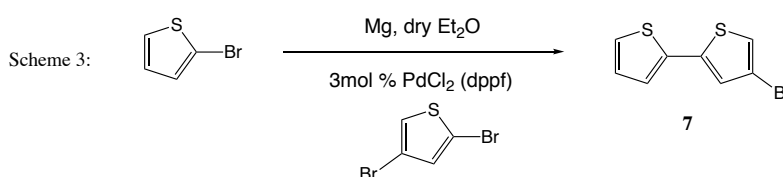


Figure 2.4: Kumada-Corriu coupling reaction to create the essential bithiophene unit with a halide handle for further functionalization.⁴¹

The Kumada-Corriu methodology was applied for the preparation of 4-Bromo-2,2'-bithiophene **7** as shown in **Figure 2.4**. A freshly prepared 2-bromothiophene Grignard reagent was produced in dry ether using dibromoethane to activate the magnesium and initiate a Grignard reaction. From here, the Grignard was slowly added to 2,4- dibromothiophene and underwent a palladium catalyzed cross-coupling reaction to provide the regioselective 4-bromo-2,2'-bithiophene **7** in moderate to good yield of 82%. The product 4-bromo-2,2'-bithiophene **7** is an intermediate in all subsequent synthetic routes to bithiophene containing ncAAs discussed in this paper. To couple intermediate **7** to a pre-existing stereocenter, we chose to construct a derivative of NBoc-serine-benzyl ester for our substrate as many ncAAs with aromatic side chains have been synthesized using this substrate via Negishi organozinc coupling reactions.^{42–44}

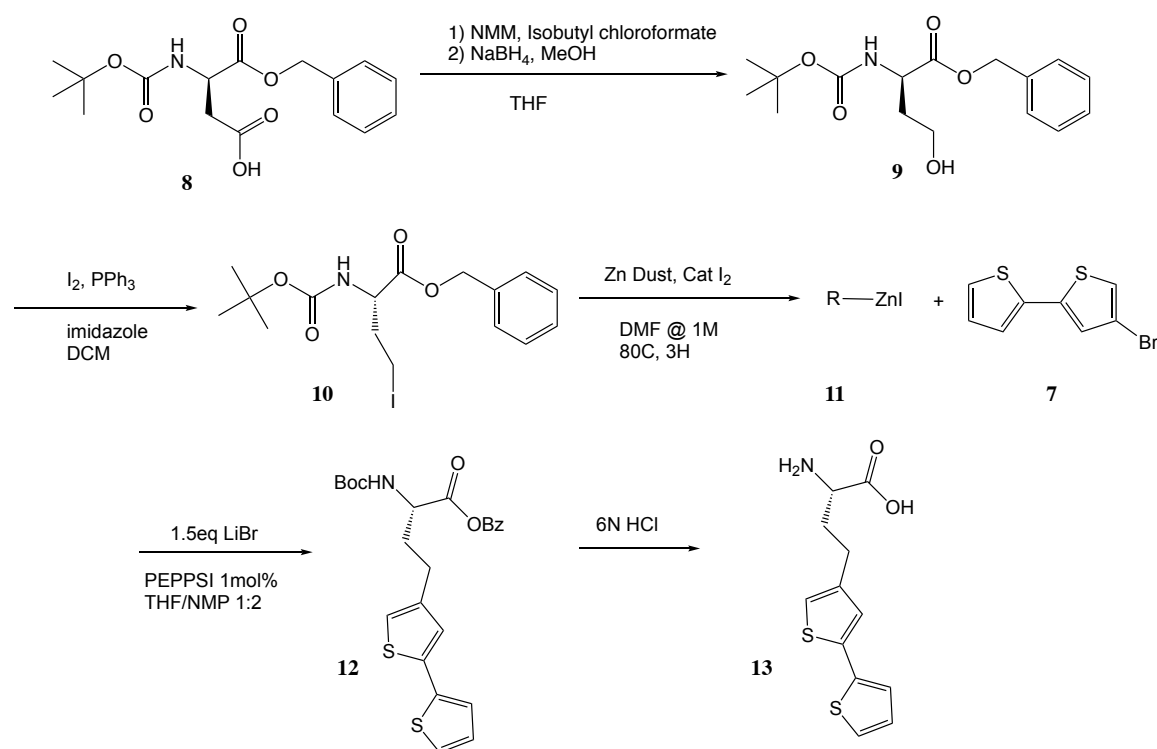
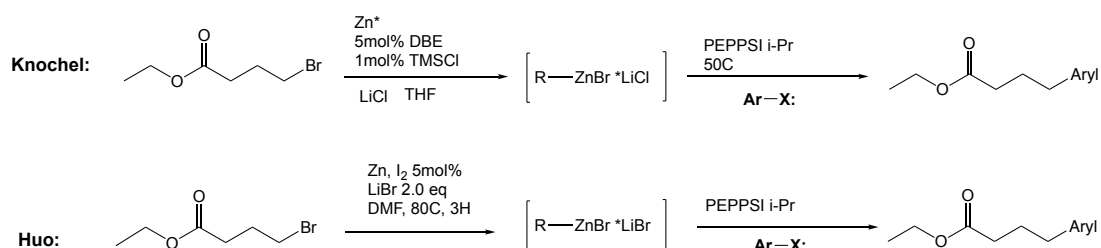


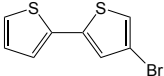
Figure 2.5: Reaction scheme for the synthesis of bithiophene ncAA using the transition metal cross coupling pathway. The Kumada Corriu reaction to create 7 is used in all synthetic pathways.^{42–44}

For our purposes, we wished to synthesize homoserine, as the additional carbon within the side chain allows for increased flexibility and functionality when the ncAA is incorporated into our proteins. To do this, **Figure 2.5** shows commercially available NBoc-(L) aspartic acid-benzyl ester **8** was initially converted to the acyl chloride, followed by NaBH₄ reduction to afford our desired homoserine **9**. Triphenylphosphine and iodide were then added to NBoc-(L) homoserine-benzyl ester to undergo iodination of the primary alkyl alcohol **10**. The final step in the synthesis is the Negishi palladium catalyzed cross coupling of substrate **7** with **10**. We selected PEPPSI iPr (pyridine- enhanced precatalyst preparation, stabilization and initiation) as our catalyst due to its outstanding air stability and previous application to a wide variety of cross coupling reactions using aromatic substrates (**Figure 2.2**).^{45,46} Our organozinc compound **11** was

created using the Knochel method for Zn activation, and the proceeding coupling reaction was performed.^{36,47,48} Unfortunately, no product formation was observed. Therefore, we wanted to solidify a method of zinc activation for proper insertion into the C-I bond. Model reactions using less reactive substrates such as 2-bromothiophene and 2-bromopyridine were tested in reactions **1** and **2**. Zinc dust activation was attempted using the Huo method utilizing catalytic iodide versus the more popular Knochel preparation using trimethylsilane chloride (TMSCl) and dibromoethane (DBE, **Table 2.1**).⁴⁹ 2-bromopyridine was successfully coupled with ethyl bromobutyrate in favor of the Huo activation, giving an 87% yield of **14**. 2-Bromothiophene and 4-bromo-2,2'-bithiophene test couplings were then carried forth with the Huo activation and both were successful, with 4-Bromo-2,2'-bithiophene producing a 42% yield and 2-bromothiophene producing a much lower 13% yield.

Table 2.1: Reaction series of Ethyl 4-bromobutyrate and aromatic halides



Rxn #	Ar-X	Knochel	Huo	Cat	Product #
1		NR	87%	PEPPSI @ 1mol%	14
2		NR	13%	PEPPSI @ 2mol%	15
3		NR	42%	PEPPSI @ 2mol%	16

Moving forward, we applied the Huo method to our coupling reaction between substrates **11** and **7**, yet continued to struggle to obtain any product yield. We hypothesized the addition of the β -carbon in iodo-homoserine may be decreasing the internal coordination of the inserted zinc by creating a 6-membered ring over the more favorable 5-membered ring. The higher energy of coordination for a 6-membered ring decreases the overall stability of the compound and subsequently increases a likelihood of β -elimination within the starting material.

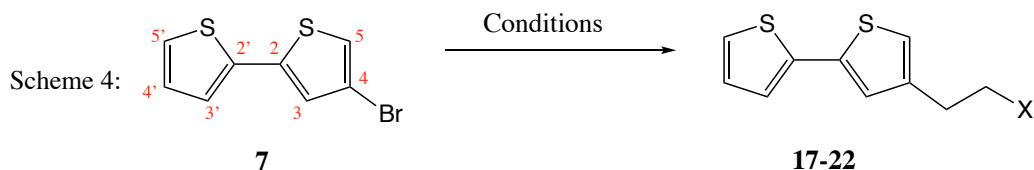
Due to the inherent struggles with efficient reduction of transition metals for successful cross coupling reactions, we focused our efforts on the phase transfer catalyst pathway. We sadly would lose the installation of a stereocenter during synthesis, but our hopes were to produce large enough quantities of racemic amino acid that we could then make enantiomerically pure.

2.4.2 Phase Transfer Catalysis

The Strecker synthesis in 1850 was the first to develop an adapted method for creating an α -carbon center through an aldehyde condensation with ammonia to form an imine intermediate. Nucleophilic addition of a cyanide ion affords an α -amino nitrile that can be hydrolyzed into the racemic amino acid.^{33,50} Modern reactions have utilized an imine intermediate, or Schiff base, for its two beneficial functionalities; protection of a free amine, and the increased acidity it provides the α -hydrogen. A Schiff base glycine derivative is the most commonly used amino acid precursor for alkylation reactions. Here, we will be utilizing Schiff bases with two different carboxylic acid protecting groups, a benzylic ester and a nitrile group as shown in **2** and **6** respectively. We will also construct the bithiophene unit with our desired two carbon chain, containing a leaving group for subsequent alkylation.

To do this, we start with our coupled thiophene moiety **7** and use ethylene oxide to install an ethyl halide handle shown in scheme 4 of **Table 2.2**. In reaction **4**, *n*-BuLi was added to perform a lithium-halogen exchange, which, upon addition of ethylene oxide, affords an ethyl oxide anion that is then trapped with *p*-toluenesulfonyl chloride to generate **17**.⁵¹ Many reactions of this character were performed with the highest yielding 41% of the tosylated product. It has been shown that there are competing deprotonation reactions that occur at the 5-position of this bithiophene unit. This competing deprotonation results in a mixture of substituted bithiophenes, some still containing the 4-bromo with additional substitution at the 5 position, therefore cutting down on the overall yield of our desired product. The lithium-halogen exchange results in an equilibrium between the two reacting species, with the lithium ion residing on its more stable substrate.⁵²⁻⁵⁴ In the case of *n*-BuLi, the lithium ion will reside on our thiophene moiety, potentially allowing for a more stable rearrangement to the 5 position, and allowing for the resulting bromobutane to react and give unwanted side products. We shifted our attention to *t*-butyl lithium in reaction **5** to help eliminate these issues and increase the yield of our desired variant. To our satisfaction, addition of *t*-butyl lithium followed by nucleophilic attack of the oxirane and anionic stabilization with *p*-toluenesulfonyl chloride afforded a 52% yield of **18**. The installation of the tosyl group was then substituted for an iodide leaving group using reaction **7** conditions resulting in a 68% yield.

Table 2: Synthetic routes and conditions to alkyl halide bithiophenes



Rxn #	Conditions	Yield	Product #
4	1) n-BuLi, THF, -78C 2) Ethylene Oxide -78C → RT 3) TsCl	41 %	17
5	1) t-BuLi, THF, -78C 2) Ethylene Oxide -78C → RT 3) TsCl	52 %	18
6	1) Mg, dibromoethane, Et ₂ O 2) Ethylene Oxide -78C → RT 3) NH ₄ Cl	50 %	19
7*	NaI Acetone	68 %	20*
8*	PPh ₃ , I ₂ Imidazole DCM	100 %	21*
9*	PBr ₃ Et ₂ O	44 %	22*

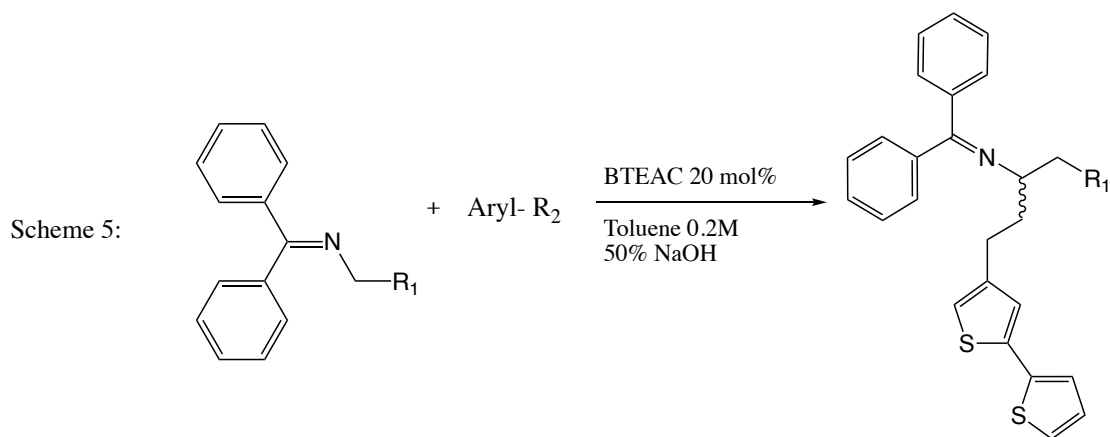
Although we were thankful to see an increase in using *t*-Buli yield, we investigated other options in hopes of reducing the strict reaction conditions and safety concerns associated with *t*-butyl lithium. In reaction **6**, we avoided the inherent complications with lithiation reactions and turned towards Grignard reagents.⁴¹ Dibromoethane was added to activate the magnesium turnings preparing our bithiophene Grignard to perform nucleophilic attack upon addition of

ethylene oxide. Rather than trapping the oxide with a tosyl leaving group, we added ammonium chloride to quench the reaction and produce a primary alcohol for consequent iodination. With a 50% yield of the alcohol **19**, we proceeded with reaction **8** and achieved 100% conversion to an iodide leaving group through the addition of triphenylphosphine and iodine to afford product **21**. Finally, we knew there could be potential for β -elimination with a 4-iodoethyl-2,2'-bithiophene that would result in a fully conjugated system with the inability to react with our Schiff bases, therefore, in reaction **9** we converted our primary alcohol **19** to possess a bromide leaving group in **22**.

Our next task was to explore the optimal conditions necessary for successful addition of an alkyl side chain into the Schiff bases **2** and **6** as seen in **Table 2.3**. Following the conditions presented by O'Donnell's group for our selected Schiff bases, bromoethyl benzene was added to a solution of toluene with 50% NaOH and BTEAC (Benzyl triethyl ammonium chloride) as the phase-transfer catalyst and stirred vigorously at room temperature. In reactions **10** and **11**, we see alkylation of the benzene with the α -cyano Schiff base at 22% yield, and t-butyl ester protected glycine at a crude yield of 53%. Control reactions were run in absence of the aryl halides to measure any possible decomposition of the α -cyano Schiff; however, little to no decomposition was observed. Varying concentrations of the organic layer were considered and it was observed that reaction **11** with the t-butyl ester Schiff base only proceeded in a more concentrated organic layer, while the cyano Schiff base was successful at lower concentrations in reaction **10**. Using bromoethyl benzene and the t-butyl ester Schiff base, we then tested a slightly bulkier phase transfer catalyst BTBAC (Benzyltributylammonium chloride), as well as a cinchona alkaloid derived catalyst that O'Donnells' group developed to help drive the synthesis of optically active

amino acids.^{55,56} Unfortunately, we were not able to observe any clean product from either of these reactions (data not shown).

Table 2.3: Synthesis of racemic bithiophene amino acid with various Schiff bases



Rxn #	R1	Aryl-R ₂	Yield	Product #
10	CN		22%*	23
11	tButyl Ester		52%*	24
12	CN		NR	25
13	tButyl Ester		NR	26
14	CN		38%	27

Although **23** and **24** did not result in high yields, we decided to proceed with our alkyl halide bithiophene's. In reaction **13**, the α -cyano Schiff base and 4-bromoethyl-2,2'-bithiophene were combined using BTEAC, a low concentration of toluene and 50% NaOH as the base. The reaction was stirred at room temperature for 24 hours, yet again, no clean product was obtained. We decided to continue our reactions solely with the 4-iodoethyl-2,2'-bithiophene for alkylation provided that the production of **21** is much higher yielding than for **22**, even if β -elimination is seen with **21**. From here, we were pleased to see the best performance from alkylations with **21** and the α -cyano Schiff base, yielding 38% of our protected racemic amino acid **27**.

Finally, we attempted acid hydrolysis of the racemic amino acid **27** in order to uncover the free amine and carboxylic acid of common amino acids. Multiple batches of products **26** and **27** were subject to a 1N hydrogen chloride stir at room temperature for 24 hours prior to a 6N acid reflux for 4 hours. However, only acid hydrolysis of **27** could really be observed due to the low resulting yields of **26**. In every case, acid hydrolysis seemed to result in a dark green solid product that became insoluble in most solvents besides basic water. We believe this was our ncAA, however; the solubility issue made the subsequent NMR and other forms of analysis a challenge, with our best result shown in **Appendix A, Figure A.17**. To observe any adverse effects of an acid reflux on the bithiophene side chain, we refluxed our starting material 4-bromo-2,2'-bithiophene under the same conditions; however, we did not observe any noticeable changes to the bithiophene structure. Therefore, further optimization of the work up and NMR analysis will need to take place in order to completely verify the existence of the free bithiophene amino acid.

2.5 Future Directions

Upon the completion of our first ncAA target, we began designing a directed evolution library in the *Mj* tyrosyl aaRS. The Schultz lab compared the mutations made from wild type *Mj*TyrRS to the second generation BpyAlaRS to see which residues were most important in providing the new fit. Mutations that eliminated residues which provided hydrogen bonding contacts to tyrosine, as well as mutations incorporating more glycine residues provided room within the binding pocket for new biaryl side chains. In the crystal structure of bipyridylalanine aaRS, the bound ncAA has a dihedral angle of 21° between the two rings to complete a face-to-face pi interaction with Tyr65 while the other pyridyl ring interacts with the carboxylate of Glu155 as seen in **Figure 2.6**.⁴

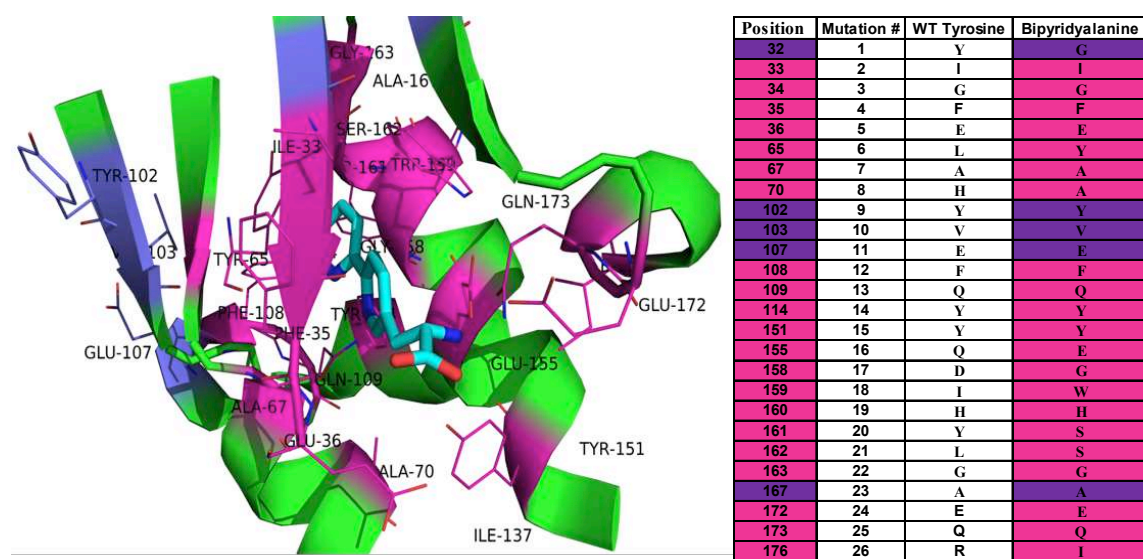


Figure 2.6: crystal structure of the mutant synthetase binding pocket for the bipyridylalanine ncAA. Pink depicts residues within $5\text{\AA}$ of the binding amino acid. Purple residues are outside $5\text{\AA}$, but were library members when creating the bipyridylalanine synthetase mutant from wild type tyrosyl tRNA synthetase.

The backbone sequence remained relatively conserved between the wild-type and bipyridylalanine aaRS. These observations were considered when designing an aaRS library for bithiophene. The positions of library members and their change from wild type tyrosyl aaRS to Bipyridyl aaRS and can be observed in the table in **Figure 2.6**. This library will be screened using the positive and negative selection systems developed by the Schultz lab.^{1,57} If difficulties arise from the specific incorporation of the bithiophene containing ncAA, a wide selection of structural variants can also be tested that would still serve the goal of creating conductive polymers.

2.6 Conclusion

Ultimately, we synthesized a racemic non-canonical amino acid with a side chain capable of polymerization using common techniques for ncAA synthesis. Transition metal cross coupling was used to create the 4-bromoethyl-2,2'-bithiophene unit in 82% yield; however, efforts to install a Negishi coupling towards the final ncAA product was unsuccessful. Concomitantly, phase transfer catalysis with an α -cyano Schiff base was successful in producing our racemic amino acid in a 15.5% overall yield. Efforts for a successful acid hydrolysis and enantiomeric resolution is necessary before utilization for *in vivo* incorporation. Building off of the bipyridylalanine aaRS, amber codon suppression can site selectively incorporate the bithiophene amino acid into essential genes to create a selection assay for tRNA synthetases with evolved binding capabilities towards the novel ncAA. Once tRNA synthetases are identified, the incorporation of this bithiophene moiety can be placed in any protein of interest for polymerization towards biologically based materials.

2.7 Materials and Methods

4-bromoethyl-2,2'-bithiophene, 7, 82% yield, (*Appendix A, Figure A.1 and A.2*)

Magnesium turnings were dried at 140°C for 2 hours prior. 2.5 equivalents of magnesium turnings were stirred with 50 mL (0.33 M with respect to 2-bromothiophene) of dry ether. 20 mmol of purchased 2-bromothiophene from Sigma was added to the mixture. Dibromoethane was added drop wise until reaction initiated, then added rapidly. The solution was refluxed for 1 hour. A second flask was prepared with 1 mol% PdCl₂(dppf) and 2,4-dibromothiophene in 50 mL dry ether. After the reflux was complete- the Grignard was added drop wise to the second flask at 0°C for 1 h. The reaction stirred at 0°C for 2 h and room temperature for 2 h before NH₄Cl was added to quench. Ether was used for extraction, the crude was concentrated and hexanes were added to crash out the excess magnesium salts and catalyst. The solution was vacuum filtered and filtrate was concentrated. The crude was columned with hexanes, giving an R_f of 0.73 in hexanes, being the first spot off the column to give a 82% yield. Product will be clear to a very slight yellow when initially columned. After sitting on a bench top for a day it will turn light blue and gradually get a darker teal. This does not change anything in the NMR. ¹H NMR (CDCl₃): 7.03 (1H, dd, J=5.1, 3.6 Hz), 7.07 (1H, d, J= 1.2 Hz), 7.09 (1H, d, J=1.5 Hz), 7.17 (1H, dd, J= 1.2, 3.6 Hz), 7.25 (1H, dd, J= 1.8, 5.1 Hz). ¹³C NMR: 110.25, 121.32, 124.36, 125.28, 126.01, 127.88, 135.92, 138.53.

N-*boc*-Homoserine-Benzyl ester, **9**, 77% yield, (*Appendix A, Figure A.3*)

NBoc-aspartic acid-benzyl ester was purchased from Alfa Aesar and 0.5mmol was dissolved in diethyl ether at 0.2M. 1.2 equivalents of Isobutyl Chloroformate (Sigma Aldrich) and 1.1 equivalents of N-methylmorpholine was added slowly at -15C to 0C over a 1 hour period with continuous stirring. The precipitated N-methylmorpholine hydrochloride was removed by filtration and washed with THF. Sodium borohydride and water were added dropwise at 0C and stirred for another 30 min. The reaction was quenched with ammonium chloride and the product was extracted with ethyl acetate to afford a 77% yield. ^1H NMR (CDCl_3): 1.43 (9H, s), 2.16 (2H, m), 3.61-3.67 (2H, m), 4.52 (1H, m), 5.17 (2H, s), 5.35-5.37 (1H, br s), 7.34 (5H, s)

N-*boc*-Iodohomoserine-benzyl Ester, **10**, 90% yield (*Appendix A, Figure A.4*)

Triphenylphosphine and imidazole were both purchased from Sigma and dissolved in DCM at 1.5 equivalents each and a final concentration of 0.4M. The solution was cooled to 0C and I_2 was added at 1.5 equivalents. The solution was warmed to room temperature and then cooled back to 0C before the slow addition of N-Boc Homoserine-Benzyl Ester **9** in 1M DCM. The reaction stirred for 8H before becoming diluted in Et_2O and purified via a silica column in 1:1 Hexane:Ethyl Acetate with an R_f of 0.82 : ^1H NMR (CDCl_3): 1.42 (9H, s), 2.16-2.19 (1H, M), 2.39 (1H, m), 3.1-3.15 (2H, t, $J = 8.4, 7.2$ Hz), 4.35 (1H, br s), 5.05 (1H, br s), 5.16 (2H, d, $J = 3$ Hz), 7.34 (5H, s)

Huo method prep for **14**, **15**, **16**: Zinc dust was added to a dry flask under argon in dry DMF at 1.5 equivalents to the aryl bromide. Catalytic I_2 was added to the flask at 5 mol% equivalents and heated to 80C. Ethyl-4-bromobutyrate was added at 1.6 equivalents and stirred for 3H at 80C until zinc insertion is completed. A separate flask was prepped containing flame dried LiBr at 2

equivalents and 1-2mol% of Pd catalyst PEPPSI i-Pr in 4mL of a 2:1 THF:NMP solution. The organozinc was then added to the second flask at room temperature and stirred to completion. The product was extracted with ethyl acetate and purified on a silica column with 2:1 hexane: ethyl acetate solution.

Ethyl-4-pyridin-2-yl) butanoate, 14, 87% yield (Appendix A, Figure A.5)

^1H NMR (CDCl_3): 1.26 (3H, t, $J = 6.9, 7.2$ Hz), 2.11 (m, 2H), 2.37 (t, 2H, $J = 7.2, 7.8$ Hz), 2.84 (2H, t, $J = 7.5, 7.8$ Hz), 4.1 (2H, q, $J = 6.9, 7.2$ Hz), 7.1 (2H, dt, $J = 7.5, 10.5, 3, 2.7, 4.8$ Hz), 7.6 (1H, td, $J = 1.8, 15.6$ Hz), 8.5 (1H, d, $J = 4.2$ Hz).

Ethyl-4-(thiophen-2-yl) butanoate, 15, 13% yield (Appendix A, Figure A.6)

^1H NMR (CDCl_3): 1.27 (3H, t, $J = 7.5, 6.9$ Hz), 2.04 (2H, m), 2.37 (2H, t, $J = 7.2, 7.5$ Hz), 2.89 (2H, t, $J = 7.5$ Hz), 4.16 (2H, q, $J = 6.9, 7.2$ Hz), 6.79 (1H, d, 3.3 Hz), 6.92 (1H, dd, 3.3, 4.8 Hz), 7.11 (1H, d, 4.5 Hz)

Ethyl-4-([2,2'-bithiophen]-4-yl) butanoate, 16, 42% yield (Appendix A, Figure A.7)

^1H NMR (CDCl_3): 1.27 (3H, t, 7.2, 6.9 Hz), 2.00 (m, 2H), 2.36 (2H, t, 7.5 Hz), 2.65 (2H, t, $J = 7.5, 7.8$ Hz), 4.16 (2H, q, 7.2 Hz), 6.81 (1H, s), 6.97 (1H, s), 7.00 (1H, dd, $J = 1.2, 3.6$ Hz), 7.14 (1H, dd, 1.2, 3.6 Hz), 7.19 (1H, dd, 1.2, 5.1 Hz)

2-([2,2'-bithiophen]-4-yl)ethyl-4-methylbenzenesulfonate, 17, 18, 41-52% yield

(Appendix A, Figure A.8) 2 equivalents of n-BuLi solution was added at room temperature to a dark dry flask under argon at -78C in 1.6M THF. 1mmol of 4-bromo-2,2'-bithiophene **7** was added to the n-BuLi dropwise over 15 min. Solution stirred for an additional 15 min before 2.5 equivalents of ethylene oxide in THF at 1.2M was added to the solution. The reaction slowly

warmed up to room temperature before the addition of 1 mmol of 4-toluene sulfonylchloride to quench the reaction. N-BuLi solution obtains a light to dark green color until addition of 4-toluene sulfonylchloride additions changes it to a bright orange. Purified product in 41% yield is a light green oil 0.36 Rf in 2:1 hexane: ethyl acetate. Reactions with t-Buli preceded similarly, with 2 equivalents of t-BuLi added to **7** in the dark and gave a 52% yield. ^1H NMR (CDCl_3): 2.36 (3H, s), 3.13 (2H, t, J= 6.6, 6.3 Hz), 4.23 (2H, t, J= 6.6 Hz) 6.69 (1H, s), 6.80 (1H, s), 6.93 (1H, d), 6.98 (1H, t), 7.05 (1H, s), 7.18 (1H, d, J=5.1 Hz), 7.28 (2H, d, J=8.7 Hz), 7.72 (2H, d, J= 7.8) 2-([2,2'-bithiophen]-4-yl)ethan-1-ol, **19**, 50% yield (*Appendix A, Figure A.9*)

5 equivalents of magnesium turnings were dried for 2 hours at 140C before being added to the flask. 20mL of dry ether was added along with 4mmol of the 4-bromo-2,2-bithiophene. 1.4mL of dibromoethane was added drop wise over 20min. The Grignard was refluxed for 1 hour before being cooled to 0C. 1.2 equivalents of ethylene oxide prepared in THF was added to the cooled reaction. The reaction is stirred for 5 min added warmed up to reflux for another hour. The reaction is then quenched with NH_4Cl and extracted with ethyl acetate. The crude is run through a silica column using 6:1 hexane to ethyl acetate solution being the last spot off the column for a 50% yield. ^1H NMR (CDCl_3): 2.85 (1H, br s), 3.06 (2H, t, J= 6, 3.9 Hz), 3.87 (2H, t, J= 6.3, 2.7 Hz), 6.76 (1H, s), 7.00 (2H, t, J= 1.8, 3.3 Hz), 7.1 (1H, s), 7.16 (1H, s), 7.25 (1H, t, J= 2.1, 3 Hz) ^{13}C NMR: 33.8, 62.7, 120.7, 123.6, 124.36, 124.96, 127.73, 137.36, 137.79, 139.58

4-(2-iodoethyl)-2,2'-bithiophene, **20, 21**, 68-100% yield (*Appendix A, Figure A.10, A.11*)

To begin with the synthesis of **20**, substrate **17** or **18** were dissolve in a 15% sodium iodide solution in acetone and reflux at 50C over night. The solution was then worked up in ethyl acetate to exact the product in 68% yield. To synthesis **21**, triphenylphosphene, I_2 and imidazole were added to DCM at 0.6M and 1.5 equivalents each and stirred at 0C until all contents had

dissolved. The substrate **19** was dissolved in DCM and added drop wise to the chilled flask. The reaction was quenched with aqueous sodium sulfite and worked up with DCM. Only one spot resulted on the TLC, therefor the crude product was carried on in following alkylation reactions. ^1H NMR (CDCl_3): 3.35 (4H, s), 6.77 (1H, d, $J=3.3$), 7.01 (2H, t, $J= 3.6, 5.4$ Hz), 7.12 (1H, dd, $J= 1.2, 3.3$ Hz), 7.20 (1H, dd, $J= 1.2, 5.1$ Hz)

*4-(2-bromoethyl)-2,2'-bithiophene, **22**, 44% yield (Appendix A, Figure A.12)*

3mmol of substrate **19** was dissolved in diethyl ether at 0.5M at -15C. 1 equivalent of phosphorous tribromide was added and the reaction was stirred with gradual warming to room temperature over one and a half hours. Reaction was worked up in ethyl acetate to afford a 44% yield of **22**. ^1H NMR (CDCl_3): 3.18 (2H, t, $J= 5.7, 11.4$ Hz) 3.59 (2H, t, 5.7, 11.4 Hz), 6.94 (1H, s), 7.01 (1H, q, $J=2.7, 1.2$ Hz), 7.04 (1H, d, $J= 1.2$), 7.16 (1H, dd, $J= 0.6, 2.7$ Hz), 7.21 (1H, dd, $J= 0.9, 3.9$ Hz)

Phase transfer catalysis reactions for products **23,24**, and **27**: 0.75mmol (1.2eq) of the respective Schiff base (CN or t-butyl) were added with 40mol% of the phase transfer catalyst, which in majority of successful reactions was BTEAC (Benzyltriethyl ammonium chloride.) 3.1ml of toluene and 20 equivalents (1.3mL) of 50% sodium hydroxide were added to the flask. 0.62mmol of substrate **20** and **21** were dissolved in 2 mL of toluene and added drop wise to the stirring reaction. The reaction was stirred at max speed and left overnight which resulted in a redish-orange color of solution. After stirring overnight, if product was not observed the reaction was heated to 35C. The final product was purified on a silica column in 6:1 hexane to ethyl acetate with an R_f of 0.41 for **27**.

2-([diphenylmethylene]amino)-4-phenylbutanenitrile, 23, 22% yield

(Appendix A, Figure A.13) ^1H NMR (CDCl_3): 2.11-2.35 (2H, m), 2.82 (2H, q, $J = 7.2, 7.5, 6.9$ Hz), 4.23 (1H, t, $J = 6.9$ Hz), 7.10 (4H, m), 7.1-7.25 (3H, m), 7.32-7.37 (2H, td, $J = 7.2$ Hz), 7.44 (4H, m), 7.65 (2H, dd, $J = 7.2$ Hz)

Tert-butyl-2-([diphenylmethylene]amino)-4-phenylbutanoate, 24, 52% yield

(Appendix A, Figure A.14) ^1H NMR (CDCl_3): 1.46 (9H, s), 3.18 (2H, t, $J = 6$ Hz), 3.59 (2H, t, $J = 6$ Hz), 7.21 (4H, m), 7.34 (4H, m), 7.50 (6H, m), 7.57-7.67 (4H, m), 7.81 (4H, d, $J = 5.4$) – extra aromatic hydrogens

4-([2,2'-bithiophen]-4-yl)-2-((diphenylmethylene)amino)butanenitrile, 27, 38% yield

(Appendix A, Figure A.15, A.16) ^1H NMR (CDCl_3): 2.17-2.33 (2H, m), 3.00 (2H, m), 4.31 (1H, t, $J = 6.9$ Hz), 6.59 (1H, d, $J = 3.6$ Hz), 6.89 (1H, d, $J = 3.6$ Hz), 7.00 (1H, m), 7.06 (1H, dd, $J = 3.6$ Hz), 7.14 (2H, m), 7.21 (3H, m), 7.32-7.36 (3H, m), 7.4-7.49 (7H, m), 7.55-7.69 (4H, m), 7.80 (2H, $J = 9.3$ Hz) ^{13}C NMR: 26.00, 30.92, 51.88, 123.26, 123.39, 124.04, 125.61, 127.23, 127.70, 127.80, 128.23-128.36, 128.84, 128.95-128.99, 129.32, 129.68, 129.86, 130.03, 130.12, 130.80, 131.25, 137.57

REFERENCES

- (1) Wang, L.; Xie, J.; Schultz, P. G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249. <https://doi.org/10.1146/annurev.biophys.35.101105.121507>.
- (2) Cornish, V. W.; Mendel, D.; Schultz, P. G. Probing Protein Structure and Function with an Expanded Genetic Code. *Angew. Chemie Int. Ed. English* **1995**, *34* (6), 621–633. <https://doi.org/10.1002/anie.199506211>.
- (3) Wang, L.; Schultz, P. G. A General Approach for the Generation of Orthogonal TRNAs. *Chem. Biol.* **2001**, *8* (9), 883–890. [https://doi.org/10.1016/S1074-5521\(01\)00063-1](https://doi.org/10.1016/S1074-5521(01)00063-1).
- (4) Xie, J.; Liu, W.; Schultz, P. G. A Genetically Encoded Bidentate, Metal-Binding Amino Acid. *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (48), 9239–9242. <https://doi.org/10.1002/anie.200703397>.
- (5) Liu, C. C.; Mack, A. V.; Tsao, M.; Mills, J. H.; Soo, H.; Choe, H.; Farzan, M.; Schultz, P. G.; Smider, V. V. PNAS, 2008, 105, 17688.Pdf. **2008**.
- (6) Turner, J. M.; Graziano, J.; Spraggon, G.; Schultz, P. G. Structural Plasticity of an Aminoacyl-TRNA Synthetase Active Site. **2006**, *103* (17), 6–11.
- (7) McCullough, R. D. The Chemistry of Conducting Polythiophenes. *Adv. Mater.* **1998**, *10* (2), 93–116. [https://doi.org/10.1002/\(SICI\)1521-4095\(199801\)10:2<93::AID-ADMA93>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1521-4095(199801)10:2<93::AID-ADMA93>3.0.CO;2-F).
- (8) Barbarella, G.; Melucci, M.; Sotgiu, G. The Versatile Thiophene: An Overview of Recent Research on Thiophene-Based Materials. *Adv. Mater.* **2005**, *17* (13), 1581–1593. <https://doi.org/10.1002/adma.200402020>.
- (9) Åslund, A.; Nilsson, K. P. R.; Konradsson, P. Fluorescent Oligo and Poly-Thiophenes and Their Utilization for Recording Biological Events of Diverse Origin—When Organic Chemistry Meets Biology. *J. Chem. Biol.* **2009**, *2* (4), 161–175. <https://doi.org/10.1007/s12154-009-0024-8>.
- (10) Yagci, Y.; Jockusch, S.; Turro, N. J. Mechanism of Photoinduced Step Polymerization of Thiophene by Onium Salts: Reactions of Phenyliodonium and Diphenylsulfonium Radical Cations with Thiophene. *Macromolecules* **2007**, *40* (13), 4481–4485. <https://doi.org/10.1021/ma070586a>.
- (11) Poly, P. Á.; Polythiophene, Á. Progress in the Synthesis of Poly (3-Hexylthiophene). **2014**, No. July, 1–38. <https://doi.org/10.1007/12>.
- (12) Crick, F. H. C. On Protein Synthesis. *Med. Res. Counc. Unit Study Mol. Biol.* **1958**, *12*, 138–163.

- (13) Cobb, M. 60 Years Ago, Francis Crick Changed the Logic of Biology. *PLoS Biol.* **2017**, *15* (9), 1–8. <https://doi.org/10.1371/journal.pbio.2003243>.
- (14) The, O. N.; Of, R.; Ribonucleic, S.; In, A.; For, C.; Francois, B. Y.; Weisblum, B.; Ray, W. J. On the Role of Soluble Ribonucleic Acid in Coding for Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, No. 1, 1086–1092.
- (15) Vickery, H. B.; Schmidt, C. L. A. The History of the Discovery of the Amino Acids. *Chem. Rev.* **1931**, *9* (2), 169–318. <https://doi.org/10.1021/cr60033a001>.
- (16) Clarke, H. T.; Bean, H. J. α -Aminoisobutyric Acid. *Org. Synth.* **2003**, No. September, 4–4. <https://doi.org/10.1002/0471264180.os011.02>.
- (17) Berkholz, D. S.; Shapovalov, M. V.; Dunbrack, R. L.; Karplus, P. A. Conformation Dependence of Backbone Geometry in Proteins. *Structure* **2009**, *17* (10), 1316–1325. <https://doi.org/10.1016/j.str.2009.08.012>.
- (18) Chem, J. B. Prediction of Protein Conformation? **1973**, *13* (2), 222–245.
- (19) Dill, K. A.; Maccallum, J. L.; Folding, P. The Protein-Folding Problem , 50 Years On. **2012**, No. NOVEMBER, 1042–1047.
- (20) Baldwin, R. L.; Rose, G. D. How the Hydrophobic Factor Drives Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (44), 12462–12466. <https://doi.org/10.1073/pnas.1610541113>.
- (21) Noyori, R. Noyori-2002-Angewandte_Chemie_International_Edition. **2008**. [https://doi.org/10.1002/1521-3773\(20020617\)41:12<2008::AID-ANIE2008>3.0.CO;2-4](https://doi.org/10.1002/1521-3773(20020617)41:12<2008::AID-ANIE2008>3.0.CO;2-4).
- (22) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79* (March), 413–444. <https://doi.org/10.1146/annurev.biochem.052308.105824>.
- (23) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79* (1), 413–444. <https://doi.org/10.1146/annurev.biochem.052308.105824>.
- (24) Murakami, H.; Hohsaka, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M. Site-Directed Incorporation of Fluorescent Nonnatural Amino Acids into Streptavidin for Highly Sensitive Detection of Biotin. *Biomacromolecules* **2000**, *1* (1), 118–125. <https://doi.org/10.1021/bm990012g>.
- (25) Fahmi, N. E.; Dedkova, L.; Wang, B.; Golovine, S.; Hecht, S. M.; Charlottes, V. Site-Specific Incorporation of Glycosylated Serine and Tyrosine Derivatives into Proteins. **2007**, 3586–3597. <https://doi.org/10.1021/ja067466n>.

- (26) Vázquez, M. E.; Nitz, M.; Stehn, J.; Yaffe, M. B.; Imperiali, B. Fluorescent Caged Phosphoserine Peptides as Probes to Investigate Phosphorylation-Dependent Protein Associations. *J. Am. Chem. Soc.* **2003**, *125* (34), 10150–10151. <https://doi.org/10.1021/ja0351847>.
- (27) Imperiali, B.; Prins, T. J.; Fisher, S. L. Chemoenzymatic Synthesis of 2-Amino-3-(2,2'-Bipyridinyl)Propanoic Acids. *J. Org. Chem.* **1993**, *58* (6), 1613–1616. <https://doi.org/10.1021/jo00058a058>.
- (28) Zhang, G.; Sun, H. J. Racemization in Reverse: Evidence That D-Amino Acid Toxicity on Earth Is Controlled by Bacteria with Racemases. *PLoS One* **2014**, *9* (3), 1–6. <https://doi.org/10.1371/journal.pone.0092101>.
- (29) Yamada, S. I.; Oguri, T.; Shioiri, T. Asymmetric Synthesis of α -Amino-Acid Derivatives by Alkylation of a Chiral Schiff Base. *J. Chem. Soc. Chem. Commun.* **1976**, No. 4, 136–137. <https://doi.org/10.1039/C39760000136>.
- (30) Schöllkopf, B. U.; Groth, U.; Hiinig, S. Enantioselective Synthesis of (R)-Amino Acids Using L-Valine as Chiral Agent[**]. **1981**, *1187* (9), 798–799.
- (31) Schöllkopf, U. Enantioselective Synthesis of Non-Proteinogenic Amino Acids via Metallated Bis-Lactim Ethers of 2,5-Diketopiperazines. *Tetrahedron* **1983**, *39* (12), 2085–2091. [https://doi.org/10.1016/S0040-4020\(01\)91926-X](https://doi.org/10.1016/S0040-4020(01)91926-X).
- (32) Williams, R. M.; Sinclair, P. J.; Zhai, D.; Chen, D. Practical Asymmetric Syntheses of α -Amino Acids through Carbon-Carbon Bond Constructions on Electrophilic Glycine Templates. *J. Am. Chem. Soc.* **1988**, *110* (5), 1547–1557. <https://doi.org/10.1021/ja00213a031>.
- (33) Wang, J.; Liu, X.; Feng, X. Asymmetric Strecker Reactions. *Chem. Rev.* **2011**, *111* (11), 6947–6983. <https://doi.org/10.1021/cr200057t>.
- (34) Hassan, J.; Sévignon, M.; Gozzi, C.; Schulz, E.; Lemaire, M. Aryl-Aryl Bond Formation One Century after the Discovery of the Ullmann Reaction. *Chem. Rev.* **2002**, *102* (5), 1359–1469. <https://doi.org/10.1021/cr000664r>.
- (35) Negishi, E. I. A Genealogy of Pd-Catalyzed Cross-Coupling. *J. Organomet. Chem.* **2002**, *653* (1–2), 34–40. [https://doi.org/10.1016/S0022-328X\(02\)01273-1](https://doi.org/10.1016/S0022-328X(02)01273-1).
- (36) Manolikakes, G.; Knochel, P. Radical Catalysis of Kumada Cross-Coupling Reactions Using Functionalized Grignard Reagents. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (1), 205–209. <https://doi.org/10.1002/anie.200803730>.
- (37) Donnell, M. J. O.; Boniece, J. M.; Earp, S. E. The Synthesis of Amino Acids by Phase-Transfer Reactions. *Tetrahedron Lett.* **1978**, No. 30, 2641–2644.

- (38) O'DONNELL, M. J.; BENNETT, W. D.; BRUDER, W. A.; JACOBSEN, W. N.; KNUTH, K.; LECLEF, B.; POLT, R. L.; BORDWELL, F. G.; MROZACK, S. R.; CRIPE, T. A. ChemInform Abstract: Acidities of Glycine Schiff Bases and Alkylation of Their Conjugate Bases. *J. Am. Chem. soc.* **1988**, *110* (25), 8520–8525. <https://doi.org/10.1002/chin.198912054>.
- (39) Hartwig, J. F. Recent Advances in Palladium- and Nickel- Catalyzed Chemistry Provide New Ways to Construct C À N and C À O Bonds. *Angew. Chem. Int. Ed.* **1998**, *37*, 2046–2067. [https://doi.org/10.1002/\(SICI\)1521-3773\(19980817\)37:15<2046::AID-ANIE2046>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1521-3773(19980817)37:15<2046::AID-ANIE2046>3.0.CO;2-L).
- (40) Jana, R.; Pathak, T. P.; Sigman, M. S. Advances in Transition Metal (Pd,Ni,Fe)-Catalyzed Cross-Coupling Reactions Using Alkyl-Organometallics as Reaction Partners. *Chem. Rev.* **2011**, *111* (3), 1417–1492. <https://doi.org/10.1021/cr100327p>.
- (41) Rasmussen, S. C., Pickens, J. C., Hutchinson, J. E. A General Synthetic Route to 4-Substituted-2-2'-Bithiophenes. *J. Heterocycl. Cem.* **1996**, *34*, 285.
- (42) Jackson, R. F. W.; Moore, R. J.; Dexter, C. S.; Elliott, J.; Mowbray, C. E. Concise Synthesis of Enantiomerically Pure Phenylalanine, Homophenylalanine, and Bishomophenylalanine Derivatives Using Organozinc Chemistry: NMR Studies of Amino Acid-Derived Organozinc Reagents. *J. Org. Chem.* **1998**, *63* (22), 7875–7884. <https://doi.org/10.1021/jo981133u>.
- (43) Ross, A. J.; Lang, H. L.; Jackson, R. F. W. Much Improved Conditions for the Negishi Cross-Coupling of Iodoalanine Derived Zinc Reagents with Aryl Halides. *J. Org. Chem.* **2010**, *75* (1), 245–248. <https://doi.org/10.1021/jo902238n>.
- (44) Usuki, T.; Yanuma, H.; Hayashi, T.; Suzuki, N.; Masuyama, Y. Improved Negishi Cross-Coupling Reactions of an Organozinc Reagent Derived from L-Aspartic Acid with Monohalopyridines. *J. Heterocycl. Chem.* **2014**, *51* (January), 269. <https://doi.org/10.1002/jhet>.
- (45) Organ, M. G.; Abdel-Hadi, M.; Avola, S.; Hadei, N.; Nasielski, J.; O'Brien, C. J.; Valente, C. Biaryls Made Easy: PEPPSI and the Kumada-Tamao-Corriu Reaction. *Chem. - A Eur. J.* **2007**, *13* (1), 150–157. <https://doi.org/10.1002/chem.200601360>.
- (46) Valente, C.; Belowich, M. E.; Hadei, N.; Organ, M. G. Pd-PEPPSI Complexes and the Negishi Reaction. *European J. Org. Chem.* **2010**, No. 23, 4343–4354. <https://doi.org/10.1002/ejoc.201000359>.
- (47) Knochel, P.; Dohle, W.; Gommermann, N.; Kneisel, F. F.; Kopp, F.; Korn, T.; Sapountzis, I.; Vu, V. A. Highly Functionalized Organomagnesium Reagents Prepared through Halogen-Metal Exchange. *Angew. Chemie - Int. Ed.* **2003**, *42* (36), 4302–4320. <https://doi.org/10.1002/anie.200300579>.

- (48) Ila, H.; Baron, O.; Wagner, A. J.; Knochel, P. Functionalized Magnesium Organometallics as Versatile Intermediates for the Synthesis of Polyfunctional Heterocycles. *Chem. Commun.* **2006**, 6 (6), 583–593. <https://doi.org/10.1039/b510866g>.
- (49) Huo, S. Highly Efficient, General Procedure for the Preparation of Alkylzinc Reagents from Unactivated Alkyl Bromides and Chlorides. *Org. Lett.* **2003**, 5 (4), 423–425. <https://doi.org/10.1021/ol0272693>.
- (50) Yan, H.; Suk Oh, J.; Lee, J. W.; Eui Song, C. Scalable Organocatalytic Asymmetric Strecker Reactions Catalysed by a Chiral Cyanide Generator. *Nat. Commun.* **2012**, 3, 1212–1217. <https://doi.org/10.1038/ncomms2216>.
- (51) Ma, C.; Liu, X.; Li, X.; Flippen-anderson, J.; Yu, S.; Cook, J. M. Efficient Asymmetric Synthesis of Biologically Important Tryptophan Analogues via a Palladium-Mediated Heteroannulation Reaction. **2001**, No. 12, 4525–4542.
- (52) Parham, W. E.; Bradsher, C. K. Aromatic Organolithium Reagents Bearing Electrophilic Groups. Preparation by Halogen-Lithium Exchange. *Acc. Chem. Res.* **1982**, 15 (10), 300–305. <https://doi.org/10.1021/ar00082a001>.
- (53) Bailey, W. F.; Patricia, J. J. Mechanism of Lithium Halogen Interchange. *J. Organomet. Chem.* **1988**, 351, 1–46.
- (54) Tian, Q.; Shang, S.; Wang, H.; Shi, G.; Li, Z.; Yuan, J. Halogen–Metal Exchange on Bromoheterocyclics with Substituents Containing an Acidic Proton via Formation of a Magnesium Intermediate. *Molecules* **2017**, 22 (11). <https://doi.org/10.3390/molecules22111952>.
- (55) Cat, C. P. Synthesis of Novel Chiral Phase-Transfer Catalysts and Their Application to Asymmetric Synthesis of α -Amino Acid Derivatives. **2008**, No. Figure 1, 1311–1314. <https://doi.org/10.1055/s-0029-1216736>.
- (56) O'Donnell, M. J. *Aldrichimica Acta*; 2001; Vol. 34.
- (57) Schultz, P. G.; Melancon III, C. One Plasmid Selection System for the Rapid Evolution of Aminoacyl-TRNA Synthetases. *Bioorg Med Chem Lett* **2009**, 19 (14), 3845. <https://doi.org/10.1038/jid.2014.371>.

CHAPTER THREE

UPTAKE AND CELLULAR LOCALIZATION OF DIFFERENTIALLY RESURFACED CELL-PENETRATING NANOBODIES

3.1 Overview

By virtue of their size and complexity, proteins can bind and modulate therapeutically relevant receptors that challenge drug-like small molecules. However, in contrast to most small molecule drugs, proteins typically do not penetrate the lipid bilayer surface of mammalian cells. As a result, most proteins are limited to targeting cell surface receptors. We recently reported cationic-resurfaced nanobodies that penetrate mammalian cells. Expanding on initial findings, here, we present four new differentially resurfaced cell-penetrating nanobodies. These proteins differ in lysine/arginine composition, and in the surface area of their cationic feature. Using flow cytometry, we characterized their cell-penetrating potencies. Cytosolic fractionation through western blot analysis and confocal microscopy was used to characterize cellular residence (endosomes or cytosol) of each cell-penetrating nanobody. Collectively, we report optimized and minimized resurfacing strategies for cell-penetrating nanobodies. In addition, we report the ability of these scaffolds to undergo further mutagenesis through CDR swapping of existing relevant nanobodies and verify binding of their native targets. In doing so, we hope to provide valuable insight into the design of intracellularly targeted nanobodies, which will likely find use as research tools and drug leads that bind and modulate disease-relevant intracellular receptors.

3.2 Introduction

Proteins capable of modulating cell function and fate are a significant and expanding sector of the pharmaceutical industry, and also serve as valuable reagents in basic research. However, in contrast to their small molecule counterparts, the overwhelming majority of proteins do not appreciably penetrate mammalian cells. As a result, current protein therapeutics and basic research tools typically target disease-relevant cell surface receptors or secreted polypeptides. Fusion to polycationic sequences like polyarginine¹⁻³, Tat⁴⁻⁷, and penetratin⁸ peptides have been shown to facilitate intracellular delivery of proteins in mammalian cells. More recently, researchers have used protein engineering to graft or resurface a protein with a focused or extended polycationic surface. For example, Raines and co-workers have reported ‘arginine grafting’⁹ and Liu and co-workers have reported ‘protein supercharging’^{10, 11} (extensive mutagenesis of solvent exposed residues to lysine (K) or arginine (R)), as approaches to generate proteins that access the interior of mammalian cells.

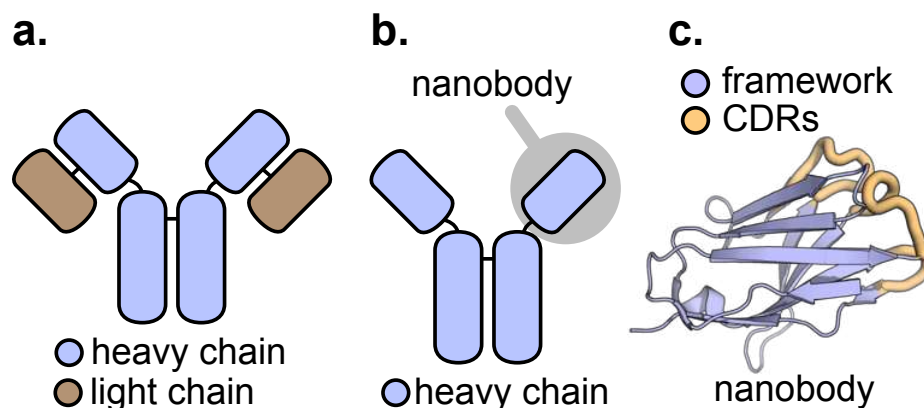


Figure 3.1. (a) Cartoon depiction of an IgG antibody consisting of heavy chain (blue) and light chain (brown) domains. (b) Cartoon depiction of a heavy chain IgG (hclgG); a monomer of the variable region (circled) is referred to as a nanobody. (c) Structure of a GFP-binding nanobody (PDB: 3OGO), with framework (blue) and Complementary Determining Regions (CDRs, light orange) highlighted.

While cationic resurfacing, in its various forms, can endow a protein with cell-penetrating properties, there are no universal ‘rules’ for this extensive mutagenesis. In our experience, most proteins are not amenable to the extensive mutagenesis associated with cationic resurfacing. An alternative approach is to identify a single protein scaffold that can be evolved to bind disease-relevant targets, but is also amenable to extensive cationic resurfacing and thus cell-penetrating. Antibodies are macromolecules that harbor the ability to be evolved to bind virtually any target of interest. However, antibodies are too large to cross the lipid bilayer of mammalian cells and cannot be recombinantly expressed in *E. coli*. In contrast to humans, and many other species, which have heavy and light chain domains in antibodies (IgG, **Figure 3.1a**), IgG antibodies in camelids (e.g. llamas and camels) and cartilaginous fishes (e.g. sharks) only contain heavy chains (**Figure 3.1b**). The monomeric variable domain (antigen-binding region) of these antibodies is referred to as a nanobody (**Figure 3.1c**).^{12, 13} A nanobody consists of a framework region and three loops referred to as the Complementary Determining Regions (CDRs, **Figure 3.1C**), which are responsible for target binding. A nanobody is a tenth of the full antibody size and some variants have been successfully expressed in *E. coli* with full binding capability.

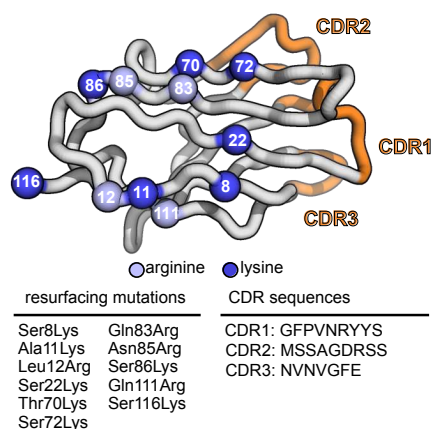


Figure 3.2. Ribbon structure of GFP-binding nanobody (PDB:3OGO) with residues mutated to achieve cationic resurfacing numbered and lighted with colored spheres. Complementary Determining Regions (CDRs) 1-3 are annotated and highlighted in orange. Resurfacing mutations and CDR sequences are provided.

We recently reported three cationic resurfaced cell-penetrating nanobodies, derived from previously reported GFP-, HER2-, β -lactamase-binding nanobodies.¹⁴ A high-resolution (2.8 Å) crystal structure of the GFP-binding nanobody has been reported (**Figure 3.1c** and **Figure 3.2**).¹⁵ Using the structure of the GFP-binding nanobody as a guide, we developed intracellularly targeting variants of all three nanobodies. Our resurfacing strategy was relatively simple. Based on previous findings^{10, 16}, we set out to generate resurfaced protein with a theoretical net charge of approximately +15. Mutation was focused within the framework region, since intracellularly targeting nanobodies derived from this scaffold would have altered CDRs that bind a therapeutically relevant target. We attempted to space out mutations over the entire framework surface in order to avoid cation/cation repulsion. Once identified, candidate residues were mutated to either lysine or arginine.

As shown in **Figure 3.2** (blue spheres), a total of eleven mutations were made throughout the framework region of the GFP-binding nanobody. The CDRs 1-3 are necessary for target binding, and were thus left untouched. Cationic resurfacing of the protein generated a nanobody (CatNB), with a theoretical net charge of +14 that retained affinity for GFP.¹⁴ CatNB is well suited to serve as a scaffold for further analysis and development, since a high-resolution structure of the parent protein is available, and mutational effects on protein folding or function can easily be assessed. We thus set out to design next-generation resurfaced nanobodies with two goals in mind: 1) optimizing the cell penetrating ability of cargo-conjugated scaffolds, 2) verifying the generality of the approach by demonstrating that resurfaced nanobodies can be directed to new targets by plug-and-play swapping of CDR domains.

3.3 Results and Discussion

3.3.1 Optimization CatNB Resurfaced Nanobody Scaffold For Delivery of Cargo proteins:

We have previously demonstrated that resurfaced nanobodies fused to GFP are capable of delivering their fluorescent cargo across cell membranes. Here we sought to optimize that performance by exploring the type and distribution of cationic mutation. We chose superfolder GFP (sfGFP) as our standard cargo. This 30 KDa protein cannot cross the cell membrane on its own, and successful internalization is conveniently probed by flow cytometry.

Before evaluating resurfacing patterns, we first sought to improve the performance of our tracking assay. Since the CDR domains of CatNB retain affinity for GFP, they provide a mechanism for intracellular self-association of the nanobody-cargo fusion, which in turn blends together the signal from multiple chromophores and confounds precise estimates of cellular location. To prevent aggregation of the CatNB-GFP fusion, we mutated three residues in CDR3 previously shown to turn off GFP recognition.¹⁵ Specifically, Asn100, Gly102 and Phe103 were all mutated to alanine. The triple mutant (CatNB_{OFF}) lacks appreciable affinity for sfGFP affinity in a simple co-purification assay (**Appendix B, Figure B.1**).

Disabling the potential of the CatNB_{OFF} to aggregate into larger complexes with other GFP molecules allowed for increased accuracy in imaging cellular localization, as seen in **Figure 3.3**. Live Hela cells were treated with 500nM of CatNB_{OFF}-GFP or CatNB-GFP, washed three times with 20 U/mL heparin sulfate, which has been demonstrated to remove cell surface bound proteins¹⁸⁻²¹, and stained with Hoechst 33342 and LysoTracker Red DND-99 for nucleus and lysosome staining, respectively. The cells were then imaged live and showed very little overlap of the GFP fluorescence with labeled lysosomes, indicating cytosolic localization. In addition, the GFP signal is uniformly sharper for the OFF mutant.

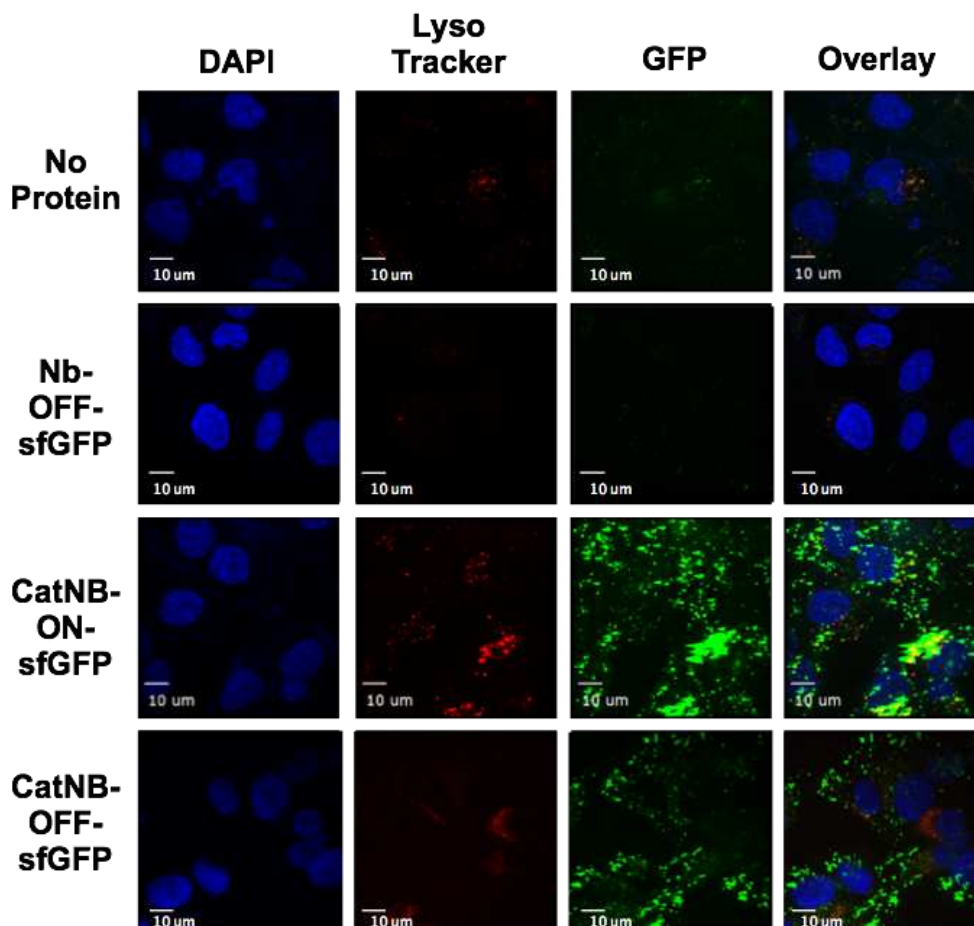


Figure 3.3: Confocal images of HeLa cells following exposure to 500nM CatNB_{ON}-GFP, or CatNB_{OFF}-GFP. Cells were incubated with protein for 1 hour, followed by staining with Hoescht 33342 and LysoTracker Red DND-99.

From here, we sought to optimize the original resurfacing strategy by dramatically increasing the number of arginine mutations, since this has been shown to be more effective in cell penetration than lysine residues. In our initial paper¹⁴, framework residues Ser22, Thr70, Ser72, Ser86, and Ser116 in the GFP-binding nanobody were all mutated to lysine, but some were converted to arginine in the resurfaced HER2- and/or β -lactamase-binding nanobodies. This suggests that those residues are amenable to arginine mutation in the context of the CatNB framework. We decided to not mutate lysine 8 to arginine since this residue was kept lysine in all

three previous nanobodies studied. Collectively, this extensive mutagenesis resulted in an arginine rich nanobody, referred to as PolyR-NB (**Figure 3.4a**). A PolyR_{OFF}-NB-GFP fusion (with the same CDR3 mutations as CatNB_{OFF}) expresses as a soluble protein in *E. coli* and is readily purified by Ni-NTA chromatography (**Appendix B, Figure B.3a**). PolyR-NB alone copurifies with GFP when expressed in *E. coli*, suggesting that despite extensive mutagenesis, the protein is folded and retains its binding function (**Appendix B, Figure B.3b,c**). This polyarginine resurfaced nanobody fused to GFP penetrates Hela cells more potently than CatNB_{OFF}-GFP over varying concentrations (250nM, 500nM, or 1uM **Figure 3.4f**).

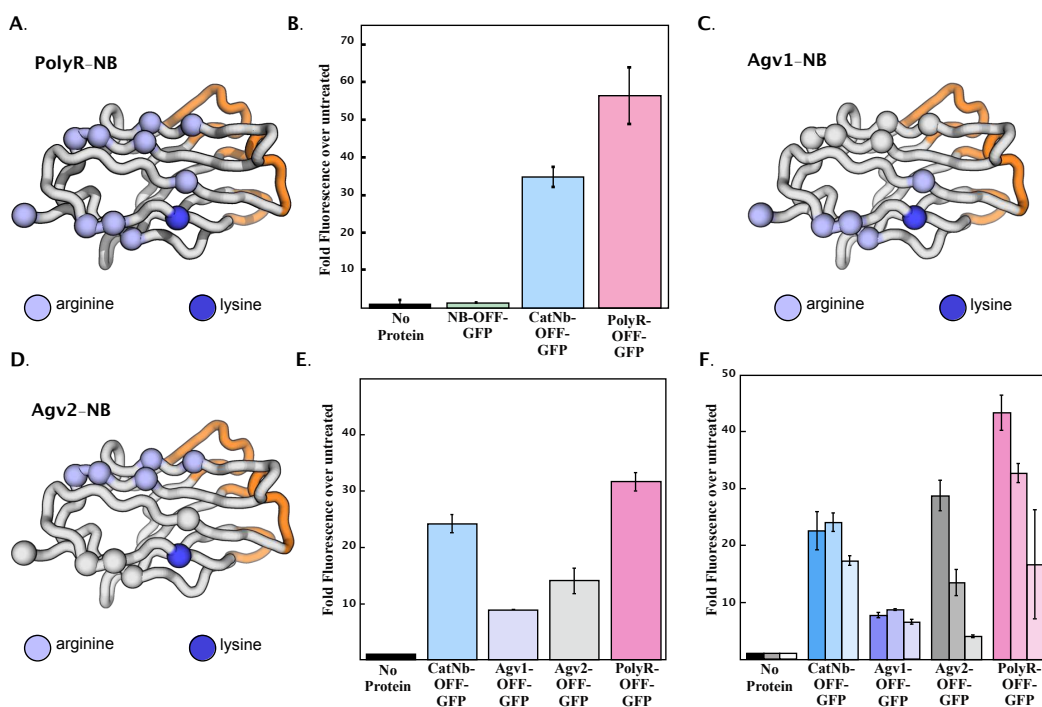


Figure 3.4. (a) Ribbon structure of GFP-binding nanobody (PDB: 3OGO) with residues mutated to achieve arginine rich cationic resurfacing (PolyR—NB). (b) Flow cytometry data of Hela cells following exposure to 500nM CatNB_{OFF}-GFP or PolyR_{OFF}-GFP. (c)-(d) Ribbon structure of GFP-binding nanobody (PDB:3OGO) with residues mutated to achieve arginine grafted variant 1 (Agv1-NB) and variant 2 (Agv2-NB). (e) Flow cytometry of Agv1_{OFF}-GFP and Agv2_{OFF}-GFP variants compared to PolyR_{OFF}-GFP at 500nM (f) Flow cytometry data of Hela cells following exposure to 1uM, 500nM or 250nM respectively (dark to light colors) of all variants.

Having demonstrated that PolyR_{OFF}-NB-GFP penetrates mammalian cells with improved potency compared to a first generation cationic nanobody CatNB_{OFF}-GFP, we next attempted to minimize resurfacing while retaining appreciable cell-penetration – analogous to previous arginine grafting studies⁹. Arginine resurfacing was essentially bisected into northern and southern hemispheres. To prepare an arginine grafted variant 1 (Agv1-NB, **Figure 3.4c**), five reversion mutations were performed, wherein arginines in the northern hemisphere were mutated to their composition in the wild-type GFP nanobody (R70T, R72S, R83Q, R85N, R86S).

To prepare arginine grafted variant 2 (Agv2-NB, **Figure 3.4d**), five reversion mutations were performed, wherein arginines in the southern hemisphere were mutated to their composition in the wild-type GFP nanobody (R11A, R12L, R22S, R111Q, R116S). Versions with the wtCDR3 that retains affinity for GFP co-purifies with GFP when expressed in *E. coli*, suggesting that both proteins are folded and retain function (**Appendix B, Figure B.3a-c**). HeLa cells were treated with 500nM of Agv1-NB_{OFF}-GFP or Agv2-NB_{OFF}-GFP, washed three times with PBS containing 20 U/mL heparin sulfate, trypsinized (which has also been shown to remove cell surface bound proteins²²) and pelleted. Cells were then analyzed for fluorescence by flow cytometry. As shown in **Figure 3.4e**, both Agv1_{OFF}-GFP and Agv2_{OFF}-GFP nanobodies are less capable of cellular penetration than PolyR-NB_{OFF}-GFP. Interestingly, PolyR-NB_{OFF}-GFP consistently shows a broader histogram than the CatNB_{OFF}-GFP (**Appendix B, Figure B.3d**), resulting in a mean fluorescence higher than that of CatNB_{OFF}-GFP even though the median peak fluorescence does not significantly improve. This may suggest PolyR_{OFF}-NB-GFP penetrates a larger population of mammalian cells at a range of concentrations compared to CatNB_{OFF}-GFP, therefore further tests were conducted to determine which scaffold would provide optimal performance.

3.3.2 Physical Characteristics of Newly Resurfaced NanoBody Scaffolds

In order to identify the best scaffold, we performed a series of tests that could help determine if the resurfacing variants differ in fold, function, and potential affect on the fused cargo protein. To start, it is important to verify that the applied charge to the outside of the proteins can resist protein aggregation. Non-native protein aggregation can lower protein yield, functionality and can cause unforeseen immunogenicity when the protein is administered.²⁴ To do so, the wtCDR3 variants with no GFP fusion were imaged immediately following a two-minute denaturation at 97°C. It can be seen that wild type GFPnb and Agv1-NB show the most aggregation with CatNB and PolyR-NB showing little to no aggregation (**Figure 3.5a**), increasing their likely-hood of becoming therapeutic scaffolds.

In our previous findings¹⁴ we tested whether the resurfacing had affected the nanobodies overall thermal stability and structural robustness and found that CatNB was able to maintain binding after full denaturation. The same test was applied to our new resurfaced variants. All five variants of the wtCDR3 nanobody scaffold, without any GFP fusion, were heated to 97°C for two minutes to ensure full denaturation. The proteins were allowed to cool to room temperature over a period of two hours, and incubated with sfGFP lysate lacking a His6 tag. The lysate and proteins incubated for 15 min before Ni-NTA beads were added to incubate for 1 hour. The Ni beads for each protein were then washed with phosphate buffer containing increasing amounts of imidazole until eluted and imaged under UV light. All protein variants were able to maintain binding potential to GFP after denaturation (**Figure 3.5b**).

Finally, we wanted to ensure that resurfacing did not have an adverse effect on the fused protein. These scaffolds can be used to deliver therapeutic proteins to mammalian cells that would otherwise not be internalized. With GFP as the stand-in “cargo protein,” we are able to

analyze the excitation/emission spectra of each fusion variant to determine any changes in fluorescence. Overall, the maximum emission wavelength remains the same for all variants, and the CatNB_{OFF}-GFP fluorescence intensity is comparable to PolyR-NB_{OFF}-GFP. The noticeable changes in fluorescence intensity between the original CatNB-GFP and CatNB_{OFF}-GFP, provides further evidence that it is the binding interaction between the GFPnb-GFP fusion molecules that affects fluorescence intensity and not the resurfacing.

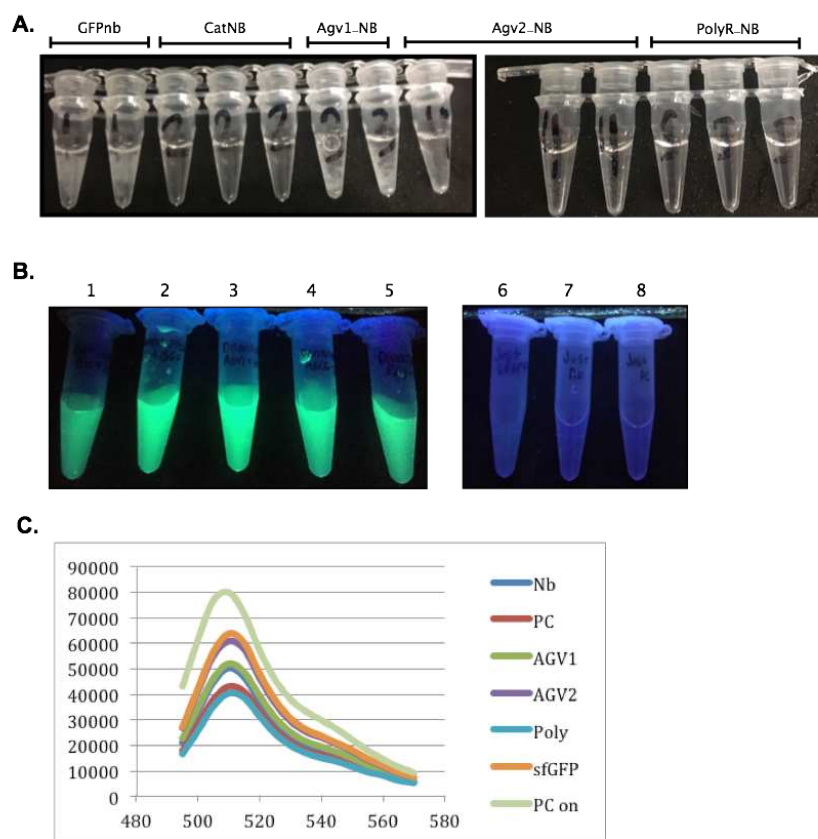


Figure 3.5: Physical characteristics of nanobody variants a) aggregation was seen in GFPnb and AGV1-NB after 1 minute at 98°C denaturation whereas no aggregation was seen for CatNB or PolyR-NB and only mild aggregation existed for AGV2-NB b) Samples were allowed to cool at room temperature and refold after denaturation. All refolded nanobodies were able to maintain binding to sfGFP (containing no his6) and co-purified in a Ni Pull down assay c) Excitation emission spectrum of a variant nanobodies fused to sfGFP.

3.3.3 Cytosolic Fractionation of PolyR-NB_{OFF}-GFP and CatNB_{OFF}-GFP

Having demonstrated that arginine resurfacing can improve NB-promoted cell penetration without compromising the stability or function of scaffold or cargo, we sought to further delineate the cellular residency of our best platforms when fused to GFP (CatNB_{OFF}-GFP and PolyR-NB_{OFF}-GFP). We used a combination of western blot and fluorescence measurements, with compartmental lysing buffers as previously described.¹⁴ Since flow cytometry provides only a measurement of total cellular fluorescence, independent of where the fluorescently-tagged molecule resides (e.g., cytosol or endosomes), we used selective lysing agents to obtain information on intracellular location. The two NB_{OFF}-GFP fusions were incubated with NIH3T3 cells for 16 hours, then washed three times with PBS containing heparin sulfate, trypsinized and pelleted. This pellet was then resuspended in a buffer containing digitonin and re-pelleted. The supernatant was removed (digitonin fraction) and the remaining pellet was resuspended in a solution containing RIPA for total cell lysis. The cells were re-pelleted to obtain the final supernatant (RIPA fraction). Since digitonin selectively lyses cell membranes but not endosomal ones, while RIPA disrupts endosomes as well, the digitonin fractions should be enriched in cytosolic proteins. As controls, the same procedures were repeated with added sfGFP or NB_{OFF} (the original GFP nanobody, but with the resurfacing mutations), as well as without added protein.

Western blot analysis was utilized to confirm the efficient separation of cytosolic and endosomal fractions by identifying separation of the endosomal marker protein Rab5. Rab5 is recruited to the early endosome membrane, and is not existent in late endosomes or lysosomes. By loading each sample lysate onto an SDS-PAGE gel and separating the proteins by size, subsequent western blot visualization of Rab5 in the RIPA fraction and not in the digitonin

fraction using Anti-Rab5 antibody would confirm a successful separation of cytosol and endosomal compartments. As observed in **Figure 3.6a**, none of the digitonin fractions contain the endosomal marker Rab5, which is observed in the RIPA fractions.

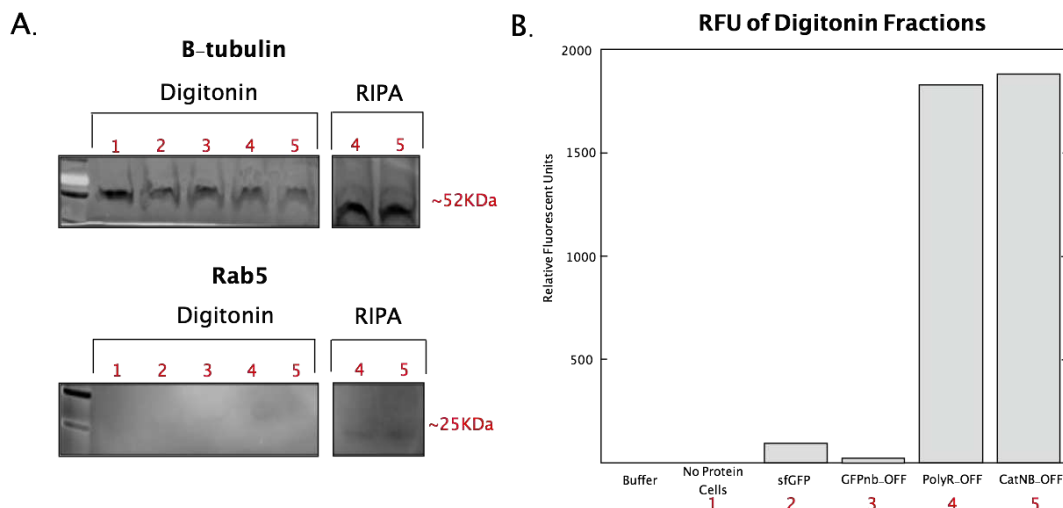


Figure 3.6: Cytosolic fractionation of NIH3T3 cells incubated with OFF GFP fusion proteins a) Western blot of the digitonin and RIPA lysate samples showing efficient separation of the cytosolic (digitonin) and endosomal (RIPA) fractions through the detection of Rab5. B-Tubulin is used as a “house-keeping” protein to verify the existence of cell lysate present in each lane b) Relative fluorescence of digitonin cell lysates show existence of PolyR-NB_{OFF} and CatNB_{OFF} in the cytosol.

Moving forward with our successfully separated fractions, the relative cytosolic concentration of each added protein was then determined by comparing fluorescence of the digitonin fraction to a standard curve established with the appropriate purified protein (**Figure 3.6b, Appendix B, Figure B.4**). It was satisfying to observe little to no exogenous sfGFP or GFPNB_{OFF}-GFP protein within the cytosol and total cell lysate, while both the CatNB and PolyR scaffolds deliver significant quantities of GFP cargo to the cytosol. CatNB_{OFF}-GFP and PolyR-NB_{OFF}-GFP were present in concentrations of 163nM and 67.7nM, respectively, corresponding to 16.3% and 6.77% of overall administered protein. Both scaffolds potently and specifically deliver exogenously administered cargo fusions to the cytosol of mammalian cells.

3.3.4 Cytosolic Fractionation with Endosomal Release Agent

Although we were pleased to observe our proteins existed within the cytosol of mammalian cells, a much larger fraction of what had been delivered existed within the total cell lysate. We wanted to take in to consideration that western blot analysis involves the lysis of cells, essentially halting the endosomal trafficking and release of our proteins at one point in time. We also wanted to partially verify the mechanism of internalization of our proteins, since some resurfaced proteins and CPPs have been proven to enter mammalian cells via multiple pathways. In effort to verify the mechanism of internalization, and to gain a better understanding of whether an extended time of administration may enhance of cytosolic delivery, we sought to utilize a small molecule known as chloroquine. In short, chloroquine inhibits lysosomal degradation, thereby extending the lifetime of endosomes and allowing time for protein cargo to escape. We applied the same compartmental lysis assay as above to verify the successful separation of cytosolic and endosomal fractions (**Figure 3.7a**).

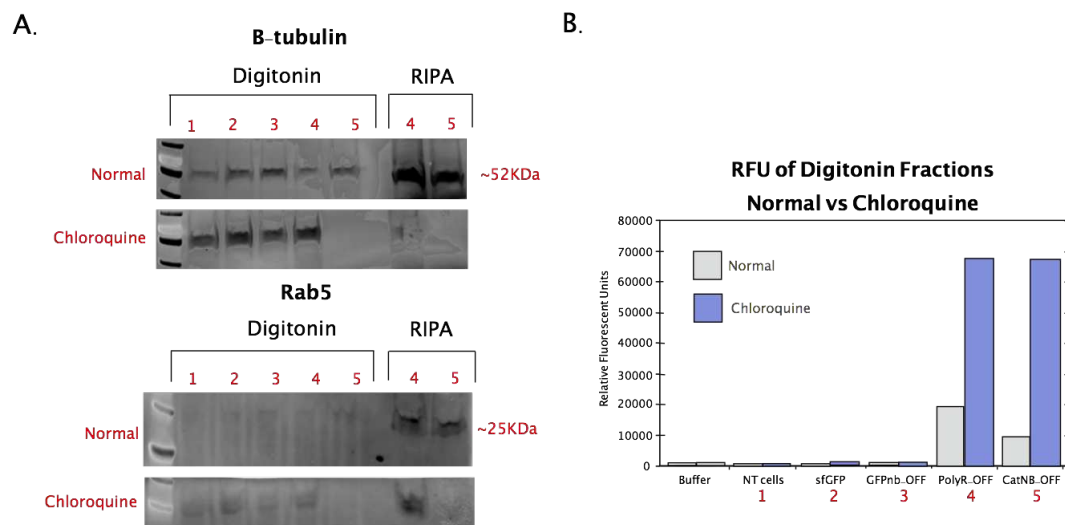


Figure 3.7: Comparison of normal internalization vs cells incubated with 100uM chloroquine to enhance endosomal release. A. Western blot verification of cell lysate using Anti-b-tubulin and successful separation verified by Anti-Rab5 antibody. B Relative fluorescence of digitonin fractions

Proteins were incubated at 1uM with NIH3T3 cells for 1 hour with and without the presence of 100uM chloroquine. Cells were pelleted after each lysis buffer and supernatants were collected as before. Fractions with chloroquine were found to have a 3.5 fold higher fluorescence within the digitonin fraction of PolyR_{OFF}-GFP and 7 fold higher fluorescence in CatNB_{OFF}-GFP as seen in **Figure 3.7b**. We then compared the digitonin fraction of each sample to a standard curve built from each corresponding protein with added NIH3T3 cell lysate to closer resemble the digitonin lysates background. To our great satisfaction, PolyR_{OFF}-GFP was measured with 333nM of cytosolic protein in cells treated with chloroquine, a 6.9 fold increase from the normal cells. Even more impressive, CatNB_{OFF}-GFP existed at a staggering 404nM within the cytosol of cells treated with chloroquine, a 4-fold increase over normal cells (**Appendix B, Figure B.5**). These concentrations are among some of the highest recorded in the cytosol of resurfaced proteins or CPPs known to date. The increased concentration of cytosolic protein in cells treated with chloroquine verifies that cells would benefit from a longer incubation time and methods that could measure concentration within live cells. These results also verify the internalization of these proteins via endocytosis mechanisms and could become extremely successful drug and drug delivery candidates with the addition of endosomal escape methodologies. However, the existence of the catNB_{OFF}-GFP protein above 100nM in the cytosol can challenge many common CPPs, including the TAT peptide, for cytosolic delivery

3.3.5 Inhibition of Endocytosis Pathways

In compliment to the experiment above, the use of small molecule endocytosis inhibitors of specific pathways of internalization will help understand the exact mechanism by which our proteins are brought into mammalian cells. Exogenous proteins can be delivered to mammalian cells through macropinocytosis, passive diffusion (if small enough) and through receptor-

mediated endocytosis dependent upon clathrin coated pits. It is well known that arginine grafted peptides and other cationic miniature proteins do not just enter the cell through a single pathway but multiple thermodynamic processes. To verify that our proteins follow the same mechanisms of endocytosis, we utilized two well known small molecule inhibitors, dynasore and amiloride.

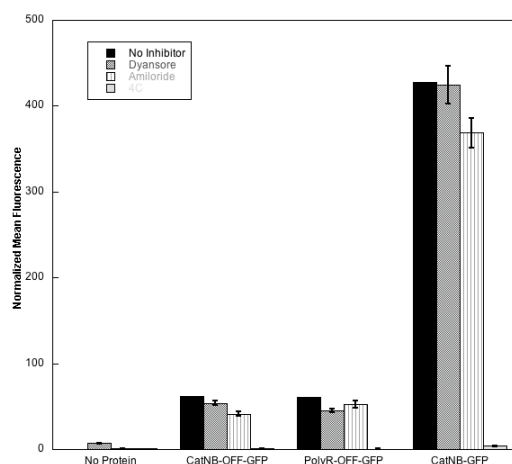


Figure 3.8: Inhibition of macropinocytosis and clathrin-dependant endocytosis of CatNB-GFP, CatNB-OFF-GFP, and PolyR-OFF-GFP

Dynasore is a small molecule inhibitor of clathrin-mediated endocytosis and Amiloride is a known small molecule inhibitor of macropinocytosis. Hela cells were plated in 24-well plates in triplicate for each set of tests. No inhibitor cells were incubated with respective protein at 500nM for 1 hour at 37C and 5%CO₂. Another set of Hela cells were incubated with 500nM of protein at 4C for 1 hour. Inhibitors were incubated with cells for 1 hour (dynasore at 50uM and amiloride at 5mM) prior to incubation with 500nM protein and remained in the presence of the protein during incubation for another hour at 37C and 5% CO₂. All cells were washed with PBS containing heparin sulfate, trypsinized and analyzed by flow cytometry. This data suggests that the inhibition of a single pathway may push the internalization through the unmonitored pathway, resulting in equal fluorescence to uninhibited cells (**Figure 3.8**). Overall, the 4C treated proteins show complete inhibition of internalization, proving that these resurfaced nanobodies do

enter the cells through thermodynamic processes. It was encouraging to continue to visualize equal amounts of internalization between CatNB_{OFF}-GFP and PolyR_{OFF}-GFP, and it was expected that the CatNB protein would exhibit a 4-fold high fluorescence than our OFF binding versions.

3.3.6 CDR Grafting of BC2 Nanobody on to Resurfaced Scaffolds:

The use of nanobodies has become extremely prevalent in biology since they essentially can be made to bind any target of interest with tight binding affinities. Immunization of camelids and phage display libraries are two of the extremely laborious processes that are mainly responsible for the evolution of new nanobodies.²⁵ Luckily, the framework regions of all nanobodies are relatively conserved while the CDRs are variable (**Figure 3.9a**), therefore grafting CDRs from recombinant nanobodies onto stable nanobody scaffolds has become a more common practice. In order for our nanobody scaffolds to become therapeutically relevant, it would be beneficial if grafting could be applied to the PolyR and CatNB scaffolds. The BC2 nanobody was the first nanobody evolved to bind a 15 amino acid peptide rather than a globular protein.²⁶ This unique binding event lead to applications that allow for super resolution imaging within mammalian cells without disrupting native biological targets.²⁷ The CDR3 of the BC2 nanobody is much longer than that of the GFP nanobody and mostly responsible for the “headlock” binding of the peptide; therefor, would serve as a good model to determine if this general scaffold could be grafted upon and still retain function (**Figure 3.9a**).

We started with our cargo protein scaffold PolyR. All three CDRs of the BC2 nanobody were grafted onto the scaffold and was expressed in *E. Coli*. We then transferred our original lysine and arginine mixed resurfacing strategy to the wild type BC2 nanobody and compared internalization of our grafted scaffold to a CatBC2nb by flow cytometry (**Figure 3.9b, Appendix**

B, Figure B.7). We were happy to discover that our PolyR scaffold not only maintained its internalization ability, but that the PolyR resurfacing with grafted CDRs was internalized more than the CatBC2nb.

We then wanted to graft upon the nanobodies without any fusion cargo to see if these scaffolds could become a therapeutic on their own. We compared the PolyR-NB as a FITC conjugate to the CatNB-FITC to determine the scaffold that would be stable enough to endure CDR grafting. These conjugated nanobodies were analyzed by flow cytometry and to our surprise; the PolyR-NB-FITC scaffold was internalized less than the original CatNB-FITC (**Figure 3.9c, Figure B.6a,b**). This could be suggesting that the stability of the nanobody was decreased when all lysines were replaced with arginines, and only the conjugation to a “cargo” protein aided in the folding process. However, it is exciting to note that these studies produced two separate useful scaffolds, PolyR-NB for passage of “cargo” proteins, and CatNB-FITC for the internalization of a nanobody possessing CDRs capable of binding new targets.

Finally, we wanted to make sure that this grafted nanobody could still maintain binding to the BC2 tag. Using a Ni-coated 96-well plate for ELISA, the BC2 CDR grafted CatNB was incubated initially to bind to the well through its His6 purification tag. sfGFP-BC2T was cloned without a his6 tag, expressed in *E. coli*, lysed and added to each well. Since the sfGFP-BC2T lysate does not have a his6 tag, the only way to identify GFP would be through the binding of the grafted nanobody and the BC2 tag on the GFP. Anti-GFP antibody was used to identify the successful binding of the two proteins as shown in **Figure 3.9d**. The successful CDR grafting proves that other therapeutically relevant CDRs could potentially be built upon this CatNB scaffold and internalized in mammalian cells to target disease relevant proteins and not just cell surface receptors.

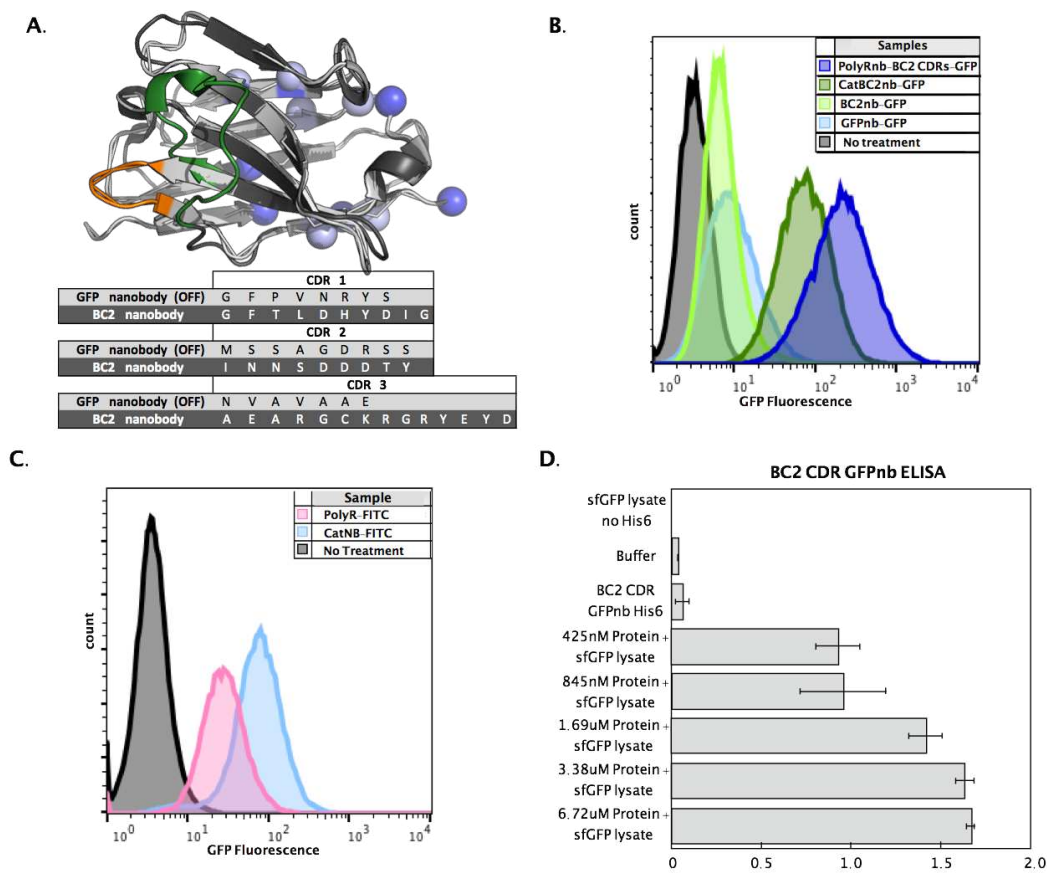


Figure 3.9. a) Ribbon structure of the CatNB in light gray overlaid with CatBC2-NB in dark gray. CDR3 is highlighted in orange for CatNB and dark green for CatBC2-NB (GFPnb: PDB 3OGO BC2nb: PDB 5IVO) b) Flow cytometry of CatBC2-NB-GFP, CatNB-GFP and PolyR-NB with BC2 CDRs fused to GFP. c) flow cytometry of CatNB-FITC and PolyRNB-FITC d) Ni coated plate ELISA showing retained binding of the CatNB-BC2 CDR (His6) with BC2 tagged sfGFP (no His6)

3.4 Conclusion

In summary, proteins are generally incapable of penetrating the lipid bilayer of mammalian cells, limiting their utility as basic research tools and therapeutic leads to targeting cell surface receptors. Researchers have used polycationic sequences—first fusions, then surface engineered features^{9, 10, 20, 21}—to dramatically enhance protein uptake in mammalian cells. However, many proteins are not receptive to the extensive mutagenesis required for surface engineering required by ‘arginine grafting’ and ‘supercharging’. An alternative approach is to

identify a single scaffold that can be resurfaced with a cationic feature-endowing cell-penetration—and can be evolved to target virtually any disease-relevant receptor.

We recently reported cationic resurfaced nanobodies (CatNB).¹⁴ These proteins potently penetrate mammalian cells, express well in *E. coli*, retain function (target recognition – at least in the case of GFP), and can, in principal, be evolved to recognize a multitude of targets. Using a cation resurfaced GFP-binding nanobody that has been mutated to turn off GFP affinity (CatNB_{OFF}), we evaluated various resurfacing strategies, improved uptake by extensive mutagenesis, and report minimalist resurfaced variants that penetrate mammalian cells. We demonstrated that CatNB_{OFF} is amenable to extensive mutagenesis of resurfaced lysine residues to arginine. When fused to a cargo protein such as sfGFP, the arginine rich nanobody (PolyR-NB_{OFF}-GFP) penetrates mammalian cells more potently than CatNB_{OFF}-GFP as seen by flow cytometry. Polycationic resurfacing was then minimized in the form of two arginine-grafted variants (agv1-NB and agv2-NB), which retain cell-penetrating properties. Using versions that retain affinity for GFP, we demonstrated that each resurfaced protein retains fold and function (GFP recognition) after denaturation, as well as resisting aggregation that WT GFP nanobody displays. We ensured that the new optimal variant PolyR_{OFF}-GFP and CatNB_{OFF}-GFP were able to access the cytosol of mammalian cells and provided quantification of their total cellular internalization from 67.7nM to 163nM respectively. Upon addition of chloroquine, a small molecule known to enhance endosomal release, we observed a cytosolic concentration up to 404nM of our CatNB_{OFF}-GFP protein. We showed that the CatNB scaffold is amenable to CDR grafting from additional recombinant nanobodies and maintains binding to a target peptide, proving that this technique can be used to develop novel grafted nanobodies with the capability of targeting intracellular disease relevant proteins. Overall, we demonstrated multiple

functionalities for these nanobodies, from the ability to drag in “cargo” proteins of therapeutic relevance using the PolyR scaffold, to the ability of the nanobody to become the therapeutic itself through CDR grafting upon the CatNB scaffold. We anticipate that these nanobodies may serve as scaffolds for the development of new research tools and drug leads that target and modulate disease-relevant intracellular receptors.

3.5 Materials and Methods

3.5.1 Materials

Hoechst 33342, LysoTracker Red DND-99 were purchased from Thermo Fischer Scientific. Heparin Sodium Sulfate from porcine intestinal mucosa, Grade I-A was purchased from Sigma Aldrich. 35mm Dish, No 1.5 coverslip 14mm glass diameter cell culture dishes were purchased from MatTek. GBlocks of nanobodies were purchased from IDT. Mouse monoclonal Anti-B-Tubulin was purchased from Abcam, Mouse monoclonal Anti-Rab5 was purchased from Cell Signaling Technology (E6N8S). Secondary antibody AlexaFluor 790 Donkey Anti-Mouse Antibody was purchased from Abcam. ELISA's were performed using 96-well Ni-coated plates purchased from Pierce, HRP antibodies from Abcam and TMB one solution from Promega. Vent DNA Polymerase, various endonucleases, and Quick Ligase were purchase from New England Biolabs

3.5.2 Cloning

All plasmids were constructed on a pETDuet-1 back- bone. All proteins were assembled from a set of over-lapping oligonucleotides for the construction of sfGFP fusions, C-terminal cysteines, and the OFF mutations in CDR3. Entire nanobody sequences for resurfaced nanobodies were purchased as GBlocks from IDT. Proteins were amplified using vent and the constructs were

ligated into either NcoI and NotI/KPN1 or BamHI and KPN1 restriction enzyme cleavage sites in the pETDuet-1 plasmid. All plasmids were transformed using 5-alpha cells and carbenicillin selection markers.

3.5.3 Protein Purification

Plasmids were transformed into BL21s (DE3). Cells were grown in either 1000 or 500 mL LB or 2XYT cultures containing carbenicillin at 30C to OD₆₀₀ of 0.7 and induced with 1 mM IPTG at 20 or 25C overnight. Cells were then collected by centrifugation and resuspended in either phosphate buffer with 150mM NaCl for NBs (20mM Sodium Phosphate, pH 7.4) or resuspended in phosphate buffer with 2M NaCl for CatNBs (20mM Sodium Phosphate, pH 7.4) and stored at -20C. Frozen pellets were thawed and incubated with complete ULTRA protease inhibitors tablets then sonicated for 4 min. The lysate was cleared by centrifugation (8000 rpm, 10 min) and the supernatant was mixed with 1 mL Ni-NTA resin for 1 hour. The resin was collected by centrifugation (4750 rpm, 10 min). The resin was washed with 20mL buffer and 20mM imidazole then 30mL buffer and 50mM imidazole and finally 20mL 75mM imidazole. The protein was then eluted with 4 mL buffer containing 400 mM imidazole. The proteins were dialyzed against 1M NaCl phosphate buffer containing 20mM HEPES and analyzed for purity by SDS- PAGE. Purified proteins were quantified using absorbance at 280 nm on a Nanodrop.

3.5.4 Mammalian Cell Culture

NIH3T3 and Hela cells were incubated at 37C and 5%CO₂ using Dulbecco's High Glucose Modified Eagle Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) for growth

3.5.5 Live Cell Confocal Microscopy

Cells were plated in 35mm glass coverslip dishes tissue coated plate and grown to 80% confluency overnight. Cells were then incubated with purified protein for 1 hour in DMEM with no FBS and washed three times with 20 U/l Heparin sulfate in warm PBS. Cells were then stained with 5mM LysoTracker Red DND-99 and allowed to incubate for 1 hour at 37C and %5 CO₂. The cells were washed and stained with Hoechst 33342 for 10 min hidden from light. Cells were then immediately imaged on a Zeiss 800 Laser Scanning Confocal Microscope.

3.5.6 Flow Cytometry

Mammalian cells were grown to 80% confluency in a 12-well plate. Cells were then washed once with PBS and 1 mL of 250nM, 500nM or 1uM protein fused with GFP or conjugated to maleimide FITC was added. The cells were incubated with the protein solution for 1h in DMEM with no FBS at 37C, 5% CO₂ environment. After the incubation period, cells were washed three times with PBS- HS (heparin sulfate 20 U/mL) for approximately 5 min at 37C, 5% CO₂. The cells were then removed from dish with 0.25% trypsin-EDTA and collected by centrifugation. The cells were then suspended in PBS and taken for flow cytometry analysis

3.5.7 Cytosolic Protein Extraction and Whole Cell Lysate Preparation for Fluorescent Microplate Analysis

3T3 cells were plated in 6-well plate and grown to 80% confluency. The cells were treated with 1uM protein for 16h in DMEM with no FBS at 37C, 5% CO₂. After treatment, cells were washed once with PBS and twice with PBS-HS (heparin sulfate 20 U/mL) for 5min then lifted with 0.25% trypsin-EDTA and pelleted. For cytosolic protein extraction, cell pellets were resuspended

in 90uL of 50 ug mL⁻¹ digitonin in 75 mM NaCl, 1 mM NaH₂PO₄, 8 mM Na₂HPO₄, 250 mM sucrose supplemented with Roche protease inhibitor cocktail for 10 min on ice. Cells were then centrifuged for 5 min at 3,000 xG. Supernatant was then used as cytosolic protein extraction. Left over pellets were then re-suspended in 90 uL RIPA buffer supplemented with Roche protease inhibitor cocktail and incubated on ice for 5 min then further lysed through 10 second vortex. 100uL supernatant was then used as whole cell lysate extraction. Both supernatants were collected and analyzed separately for their fluorescence emission and compared to a standard curve for each respective protein to determine intracellular concentration. Aliquots from the same samples were loaded onto an SDS-Page gel and analyzed through western blot analysis to identify Rab5, an early endosome marker protein to show successful compartmental separation. For detection of b-tubulin, a 1/333 dilution of anti-b-tubulin antibody was made in 3% BSA, 0.1% Tween20, 1X TBS buffer and incubated over night at 4^oC. For Rab5 detection, the same antibody dilution was made with Anti-Rab5 (E6NBS Cell Signaling Technology in 5% milk, 0.1% Tween20, 1X TBS buffer and incubated over night at 4^oC. The membranes were washed three times with 0.1% Tween20 in 1X TBS buffer and incubated with a 1/5000 dilution of Donkey Anti-Mouse AlexaFluor790 Antibody for one hour at room temperature. Membranes were washed three times and imaged on an OdesseyCLx Infrared Imager.

3.5.8 CDR grafted ELISA

100uL of BC2 grafted CatNB-His6 was incubated at concentrations varying from 3.36uM to 420nM in a Ni-coated plate for 1 hr with all respective controls. Wells were washed three times with 200uL of a wash buffer consisting of .1% Tween 20, 1% BSA in 1X TBS. Wells were then incubated with 100uL sfGFP-BC2T lysate containing no his6 tag for 1hr. Wells were washed

four times with 200uL wash buffer. Wells were incubated with 100uL of Anti-GFP-HRP diluted 1/5000 in Odessey blocking buffer for 1hr. Wells were washed three times with 200uL of wash buffer. Wells were finally incubated with 100uL of room temperature TMB one solution for 20min before being analyzed for HRP absorbance on a BioTek microplate reader.

REFERENCES

- (1) Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K., and Sugiura, Y. (2001) Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery, *J Biol Chem* 276, 5836-5840.
- (2) Mitsui, H., Inozume, T., Kitamura, R., Shibagaki, N., and Shimada, S. (2006) Polyarginine-mediated protein delivery to dendritic cells presents antigen more efficiently onto MHC class I and class II and elicits superior antitumor immunity, *J Invest Dermatol* 126, 1804-1812.
- (3) Stanzl, E. G., Trantow, B. M., Vargas, J. R., and Wender, P. A. (2013) Fifteen years of cell-penetrating, guanidinium-rich molecular transporters: basic science, research tools, and clinical applications, *Acc Chem Res* 46, 2944-2954.
- (4) Vives, E., Brodin, P., and Lebleu, B. (1997) A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus, *J Biol Chem* 272, 16010-16017.
- (5) Schwarze, S. R., Ho, A., Vocero-Akbani, A., and Dowdy, S. F. (1999) In vivo protein transduction: delivery of a biologically active protein into the mouse, *Science* 285, 1569-1572.
- (6) Frankel, A. D., and Pabo, C. O. (1988) Cellular uptake of the tat protein from human immunodeficiency virus, *Cell* 55, 1189-1193.
- (7) Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L. L., Pepinsky, B., and Barsoum, J. (1994) Tat-mediated delivery of heterologous proteins into cells, *Proc Natl Acad Sci U S A* 91, 664-668.
- (8) Derossi, D., Joliot, A. H., Chassaing, G., and Prochiantz, A. (1994) The third helix of the Antennapedia homeodomain translocates through biological membranes, *J Biol Chem* 269, 10444-10450.
- (9) Fuchs, S. M., and Raines, R. T. (2007) Arginine grafting to endow cell permeability, *ACS Chem Biol* 2, 167-170.
- (10) Cronican, J. J., Thompson, D. B., Beier, K. T., McNaughton, B. R., Cepko, C. L., and Liu, D. R. (2010) Potent delivery of functional proteins into Mammalian cells in vitro and in vivo using a supercharged protein, *ACS Chem Biol* 5, 747-752.

- (11) Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007) Supercharging proteins can impart unusual resilience, *J Am Chem Soc* 129, 10110-10112.
- (12) Muyldermans, S. (2013) Nanobodies: natural single-domain antibodies, *Annu Rev Biochem* 82, 775-797.
- (13) Bruce, V. J., Ta, A. N., and McNaughton, B. R. (2016) Minimalist Antibodies and Mimetics: An Update and Recent Applications, *Chembiochem* 17, 1892-1899.
- (14) Bruce, V. J., Lopez-Islas, M., and McNaughton, B. R. (2016) Resurfaced cell-penetrating nanobodies: A potentially general scaffold for intracellularly targeted protein discovery, *Protein Sci* 25, 1129-1137. **CatNB referred to as GFPnbPC
- (15) Kubala, M. H., Kovtun, O., Alexandrov, K., and Collins, B. M. (2010) Structural and thermodynamic analysis of the GFP:GFP-nanobody complex, *Protein Sci* 19, 2389-2401.
- (16) McNaughton, B. R., Cronican, J. J., Thompson, D. B., and Liu, D. R. (2009) Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins, *Proc Natl Acad Sci U S A* 106, 6111-6116.
- (17) Wender, P. A., Mitchell, D. J., Pattabiraman, K., Pelkey, E. T., Steinman, L., and Rothbard, J. B. (2000) The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters, *Proc Natl Acad Sci U S A* 97, 13003-13008.
- (18) DePorter, S. M., Lui, I., Bruce, V. J., Gray, M. A., Lopez-Islas, M., and McNaughton, B. R. (2014) Mutagenesis modulates the uptake efficiency, cell-selectivity, and functional enzyme delivery of a protein transduction domain, *Mol Biosyst* 10, 18-23.
- (19) DePorter, S. M., Lui, I., Mohan, U., and McNaughton, B. R. (2013) A protein transduction domain with cell uptake and selectivity profiles that are controlled by multivalency effects, *Chem Biol* 20, 434-444.
- (20) Thompson, D. B., Cronican, J. J., and Liu, D. R. (2012) Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells, *Methods Enzymol* 503, 293-319.

- (21) Cronican, J. J., Beier, K. T., Davis, T. N., Tseng, J. C., Li, W., Thompson, D. B., Shih, A. F., May, E. M., Cepko, C. L., Kung, A. L., Zhou, Q., and Liu, D. R. (2011) A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo, *Chem Biol* 18, 833-838.
- (22) Appelbaum, J. S., LaRochelle, J. R., Smith, B. A., Balkin, D. M., Holub, J. M., and Schepartz, A. (2012) Arginine topology controls escape of minimally cationic proteins from early endosomes to the cytoplasm, *Chem Biol* 19, 819-830.
- (23) LaRochelle, J. R., Cobb, G. B., Steinauer, A., Rhoades, E., and Schepartz, A. (2015) Fluorescence correlation spectroscopy reveals highly efficient cytosolic delivery of certain penta-arg proteins and stapled peptides, *J Am Chem Soc* 137, 2536-2541.
- (24) Chapman, A. M. and McNaughton, B. (2016) Scratching the Surface: Resurfacing Proteins to Endow New Properties and Function, *Cell Chem Biol*. 23, 543-553.
- (25) Schumacher, D., Helma, J., Schneider, A. F. L., Leonhardt, H., and Hackenberger, C. P. R., (2018) Nanobodies: Chemical Functionalization Strategies and intracellular Applications, *Angew. Chem. Int.* 57, 2314-2333.
- (26) Braun, M. B., Traenkle, B., Koch, P. A., Emele, F., Weiss, F., Poetz, O., Stehle, T., and Rothbauer, U. (2016) Peptides in a Headlock- a Novel high-Affinity and Versatile Peptide Binding Nanobody for Proteomics and Microscopy, *Sci Rep* 6, 19211.
- (27) Virant, D., Traenkle, B., Mauer, J., Kaiser, P., Bodenhofer, M., Schmees, C., Vojnovic, I., Pisak-Lukats, B., Endesfelder, U., and Rothbauer, U. (2018) A Peptide-tag Specific Nanobody Enables High-Quality Labeling for dSTORM Imaging, *Nature Commun.* 9, 930.

CHAPTER FOUR

SPLIT NANOLUCIFERASE AS A TOOL FOR IDENTIFYING PROTEIN

LOCALIZATION IN MAMMALIAN CELLS

4.1 Overview

Proteins are increasingly used as basic research tools and therapeutics. However, the inability of most proteins to cross the lipid bilayer of mammalian cells dramatically limits their broader use. Unsurprisingly, methods for intracellular delivery of exogenous proteins have received significant attention over the last decade cell-penetrating peptides, penta-arginine tags, miniature proteins and resurfacing have been utilized to improve cytosolic delivery but the detection of such success has been limited by the availability of an assay that can distinguish between endosomal entrapment or escape and cytoplasmic delivery. Therefore, simple and high-throughput methods are needed, which measure cytosolic residence of exogenously delivered proteins. Ideally, such methods would rely on easy to access and relatively inexpensive reagents and/or equipment. Here, we utilize the split nanoluciferase live cell mammalian assay to serve as a general tool for characterizing and investigating intracellular delivery of exogenous proteins. By transfecting mammalian cells with the larger half of nanoluciferase and adding the complimentary 13-residue tag to exogenous proteins, a quantifiable luminescent signal can only be produced if proteins successfully reached the cytosol. This assay can be performed in high-throughput using commercially available reagents and plate readers. We anticipate this method will prove useful to researchers interested in developing protein delivery methods, and measuring cytosolic residence of proteins in mammalian cells.

4.2 Introduction

Historically, small-molecule (<800 Da) modulators of biological processes have been used as basic research tools and/or therapeutics.^{1,2} While small-molecules will continue to play a significant—if not prominent—role in basic research and pharmaceutical settings, the use of protein modulators of cell function and fate have become increasingly popular.^{3–5} The large size and molecular complexity of proteins often generates surfaces that potently and selectively bind to disease-relevant surfaces that challenge—or evade—their small-molecule counterparts. Additionally, laboratory or in vivo evolution of proteins capable of binding a receptor is often higher-throughput and simpler than the analogous small-molecule centered discovery process.^{6–9} However, in contrast to many small-molecule reagents and drugs, the majority of proteins do not appreciably penetrate the lipid-bilayer of mammalian cells. As a result, exogenously administered proteins are typically limited to targeting cell surface or circulating receptors, limiting their broader use. Unsurprisingly, researchers have dedicated enormous effort to identifying methods for intracellular delivery of exogenous proteins.

Multiple approaches for intracellular delivery of exogenous proteins now exist, but the struggle to improve cytosolic delivery efficiencies continues. As a result, functional analysis of intracellular protein delivery technologies must be paired with assays that report on cytosolic residence. Common assays such as cre-recombinase, split β -lactamase and GIGI/ GIGT give quantitative cytosolic measurements but rely on fluorescent signal which is hard to do without high resolution imaging software.^{10–13} Also, Cre-recombinase and other split reporters assays require many large bulky fusions to the protein or peptide of interest, potentially disrupting the native interaction you wish to study. In addition, bulky fusions to the protein of interest most likely force internalization through macropinocytosis and increase the likelihood of endosomal

sequestration. Endosomal entrapment could lead to eventual degradation of the delivered proteins without ever reaching the cytosol (or nucleus) in order to engage their target. Here, we utilize Promega's 19.3KDa split nanoluciferase (NanoBit) seen in **Figure 4.1**, in order to identify success cytosol residence.¹⁴

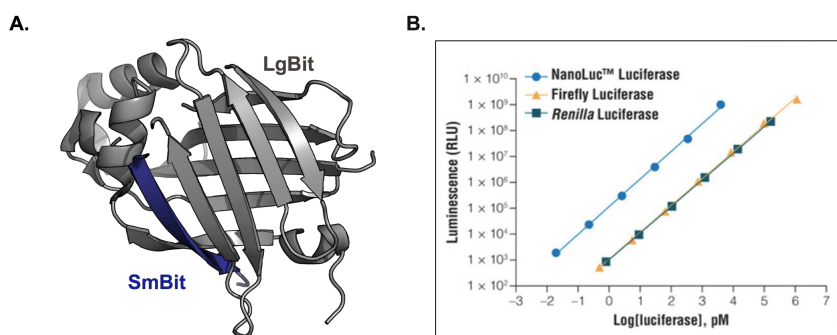


Figure 4.1) A. Crystal structure of NanoLuc with the depiction of LgBit and SmBit components. Dark Blue represents the 13 amino acid 86-peptide and dark gray represents the 18KDa LgBit portion (PDB: 5IBO) B. Logarithmic graph of the difference between commonly used luciferase reporter proteins.¹⁵

NanoBit is a small (~19.6KDa) protein that was successfully split into a LgBit (18KDa) and a SmBit (1.3KDa), that when reassemble produce a luminescence 100 fold brighter than commonly used North American Firefly and *Renilla reniformis* Luciferase.¹⁶⁻¹⁹ The creation of the NanoBit was originally to study native protein-protein interactions (PPIs), and the small fusions would be less of an interference from the native interactions. However, for PPIs to be studied, the two fragments of the reporter protein must have a significantly lower binding affinity for each other than the two interacting proteins. The team truncated the nanoluciferase in multiple positions and tested each binding affinity of the SmBit for the LgBit, and settled on the 114 with a binding affinity of 190uM for the LgBit.¹⁴ For the purposes of PPIs, this invention is extraordinary, yet we envisioned using this tool in a slightly different manner.

The purpose of this work was to redesign the use of this system towards the successful detection of protein internalization and cytosolic residence. By selecting the 86-peptide (VSGWRLFKKIS), with a binding affinity of 700pM, we sought out to determine cytosolic delivery by the spontaneous recombination of NanoBit within the cytosol. We choose to analyze four main proteins, a nanobody with β -lactamase as the wild type binder (BlacNB), a resurfaced version of this β -lactamase nanobody (Cat-BlacNB), super-folder GFP (sfGFP), and super positive GFP (34GFP).^{20–22} We choose these four proteins to help us distinguish between those proteins that do not exhibit internalization (BlacNB and sfGFP), those proven to end their journey in the endosomes (34GFP), and a protein that we have previously shown to access the cytosol (Cat-BlacNB).²³ In summary, we hope to mold the split nanoluciferase system to create an assay that is highly amenable to high-throughput experiments, can distinguish exogenous protein delivery from cytosolic delivery to endosomal entrapment, relies on relatively inexpensive and commercially available reagents, and commercially available luminescence plate readers.

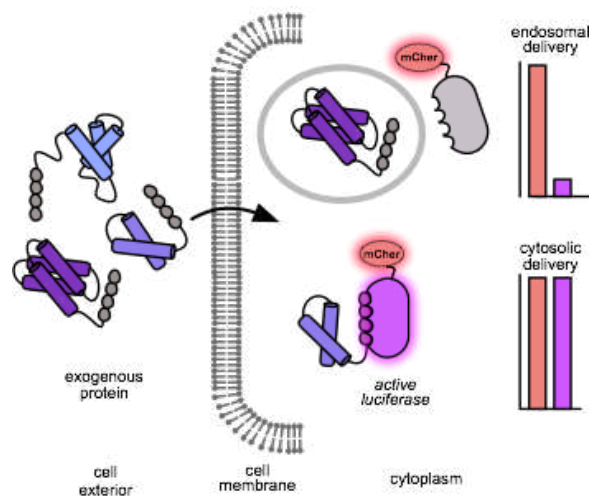


Figure 4.2) NanoBit intracellular assay scheme. This assay is designed to be used with any protein of interest that is to be delivered to the cytosol. By simply fusing the 86-peptide to the POI, and transfecting cells with the LgBit component, a luminescent signal will only be observed if the exogenous protein reaches the cytosol upon addition of Furimazine.

4.3 Assay Design and Methodology

Endocytosis of exogenously delivered proteins is the largest bottleneck in the advancement of biologics.^{21,24–26} Therefore, a facile method to identify endosomal escape is necessary. In this assay, a stronger interacting SmBit peptide was chosen such to induce rapid recombination as soon as it reaches the cytosol and interacts with the LgBit component. This is opposed to the original purpose of split nanoluciferase, where a weaker binding affinity is preferable to observe PPIs without interfering with the PPIs native interaction (**Appendix C, Figure C.1**). To start, the LgBit component is fused to mCherry and transfected into Hek293T mammalian cells. The fusion to mCherry, and thus translation of the mCherry-LgBit component in the cytosol can be observed through mCherry fluorescence, confirming successful transfection. The 13-residue amino acid peptide (SmBit-86) is then fused to any protein of interest. If the protein successfully reaches the cytosol, NanoBit should recombine, and through the addition of a small molecule known as furimazine, a detectable luminescent signal can be read using a common plate reader (**Figure 4.2**).¹⁴

4.3.1 In Vitro Assay to Identify Successful Recombination

As with any protein-protein interaction studies, specific orientation of the interacting components is necessary to achieve the best-fit and binding affinity. Since the spontaneous recombination of the SmBit and LgBit components are crucial to the identification of successful cytosolic delivery, the fusion of the SmBit to a POI, and the fusion of LgBit to mCherry, must be analyzed for a working orientation. To start, we designed our fusions to match the orientation of those that were successful when analyzing the PPI of FRB-FKBP in the original development of Nanoluciferase.¹⁴ The mCherry fusion was placed on the N-terminal end of the LgBit protein,

while the SmBit was placed C-terminus of our POI's. The fusions were ligated into a pETDuet plasmid and expressed within *E. Coli*. We discovered that our proteins were successfully expressed and purified via a Ni- column (**Appendix C, Figure C.2**). We then analyzed the ability of the mCherry-LgBit and POI-86 to recombine and produce a substantial luminescence. We began by analyzing our four 86-tagged proteins, BlacNB-86, Cat-BlacNB-86, sfGFP-86 and 34GFP-86 as purified proteins in a PBS buffer with purified mCherry-LgBit to measure luminescence. To our satisfaction, luminescence was observed in a concentration dependent manner across all purified proteins, verifying a working orientation of the fusions (**Figure 4.3a-b**). It was interesting to us that a higher concentration of protein within the well almost abolishes any luminescent signal, and that lower concentrations were thus most successful. Due to this, we sought to lower the analyzed concentrations, and also add Hek293T cell lysate to model the cellular environment the proteins will be exposed to (**Figure 4.3c**). Again, we observed a concentration dependent luminescence, verifying that our fusions were able to reassemble. Finally, we wanted to compare the luminescence of the proteins in PBS vs the assay with added cell lysate (**Figure 4.3d**) at 50nM. We observed a higher luminescence in the cell lysate assay over the proteins diluted in PBS. This not only suggests that our proteins have a functional orientation, but that, in the context of lysate proteins, our 86-tagged POIs are functional in a complex environment.

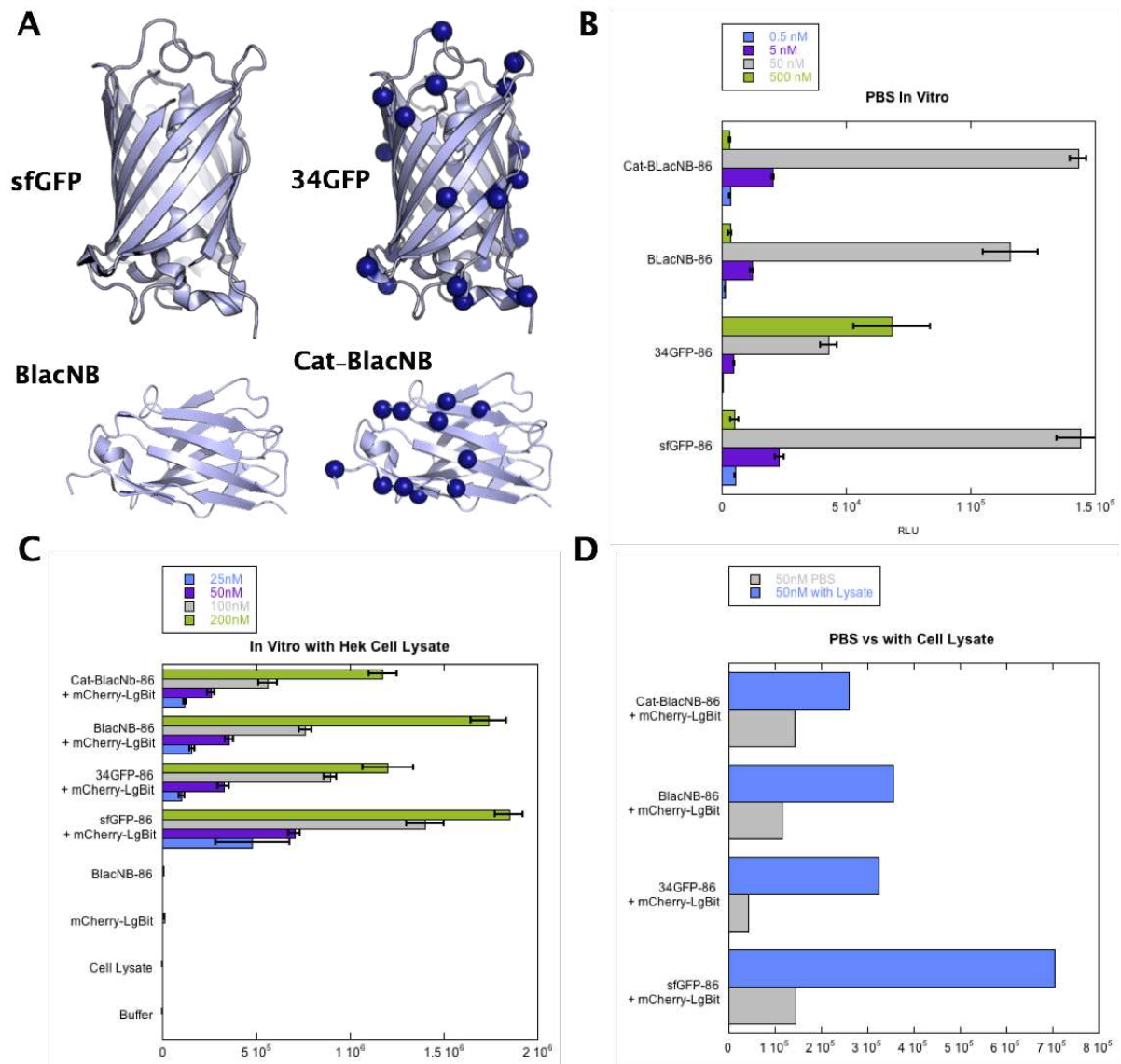


Figure 4.3) *in vitro* analysis of protein fusion orientations A) representative crystal structures of the 86-tagged proteins, comparing wild type and resurfaced structures. There is no crystal structure for BlacNB and Cat-BlacNB, therefore; crystal structure for GFP binding nanobody is used. PDB: 3OGO

4.3.2 Transfection conditions

Next, we wanted to find the proper transfection conditions needed for a functional assay. To start, we cloned the mCherry-Lgbit gene into the mammalian expression plasmid pIRES and transfected into 5a cells. The plasmid was extracted, ethanol precipitated and lyophilized before transfections. Hek293T cells were plated in a 24-well plate and transfected with four different combinations of plasmid DNA and Lipofectamine 2000, based off of previously successful transfections (data not shown). The cells were transfected with the mCherry-LgBit pIRES and allowed to incubate for 24 hours before flow cytometry analysis of mCherry fluorescence. Although the increase in fluorescence was not incredibly significant, it was determined that 150ng DNA and 1uL of Lipofectamine 2000 was observed to provide the most mCherry fluorescence (**Figure 4.4a**).

Next, we wanted to observe transfection and protein incubation timelines. Using our previously determined DNA transfection amounts, 150ngDNA and 1uL Lipofectamine, we transfected three new batches of Hek293T cells, and let them incubate for four hours prior to protein incubation. After four hours, fluorescence images were obtained on an Evos Fluorescence microscope, and minor translation of mCherry-LgBit was observed (data now shown). We then added our four 86-tagged proteins, Blacnb-86, Cat-BlacNB-86, sfGFP-86 and 34GFP-86 at three different concentrations (250nM, 500nM, and 1uM) and observed the fluorescence via Evos imaging and relative luminescence via our BioTek microplate reader. As expected, an increase in the amount of mCherry fluorescence is observed at each increasing time point. At 6 hours of protein incubation (10 hours total transfection), the cells were washed three times with PBS containing 20 U heparin sulfate and then placed in 100uL OptiMem media, according to the Promega protocol for NanoGlow Luciferase Live Cell Assay.¹⁴ The cells were

imaged and then double the amount of NanoGlow reagent was added to each well, and luminescence was immediately measured. Luminescence was measured for all four proteins, however, because BlacNB-86 is our negative control and Cat-BlacNB-86 is our positive control, only these proteins are represented in **Figure 4.4c** at all three time points. It was interesting to us that the BlacNB-86 showed just as high, if not higher luminescence than Cat-BlacNB-86. In fact, the only time point in which Cat-BlacNB-86 was higher in luminescence was at 24 hours of protein incubation. It can be seen in the in vitro data that BlacNB-86 does give a higher luminescence than Cat-BlacNB-86, suggesting that the orientation might be slightly more favorable. However, from previously published data involving BlacNB and Cat BlacNB, it has been proven that exogenous BlacNB protein is sufficiently washed away and does not internalize within mammalian cells. From here, we sought to optimize our conditions in hopes of eliminating the BlacNB-86 signal.

Originally, we had started with a 24-well plate to maintain as many cells as possible for flow cytometry analysis and washes after protein incubation, however, the NanoGlow protocol is optimized based off a 96-well plate. Due to this, we hypothesized that we were possibly not adding the correct amount of NanoGlow reagent for our modified application. We plated Hek293T cells into a 96-well plate and tested different amounts of reagent added to observe any change between our positive and negative control proteins. We tested a 5X, 10X and 20X dilution of NanoGlow, with 20X being following the Promega protocol (**Appendix C, Figure C.3**). We transfected the wells following the Sigma-Aldrich recommended amounts for a 96-well plate (75ng DNA and 0.25uL Lipofectamine) and incubated the cells four hours prior to a 24-hour incubation with our proteins at 250nM, 500nM and 1uM. Sadly, following the Promega protocol did not seem to eliminate signal for the BlacNB-86 and sfGFP-86 (**Figure 4.4d**).

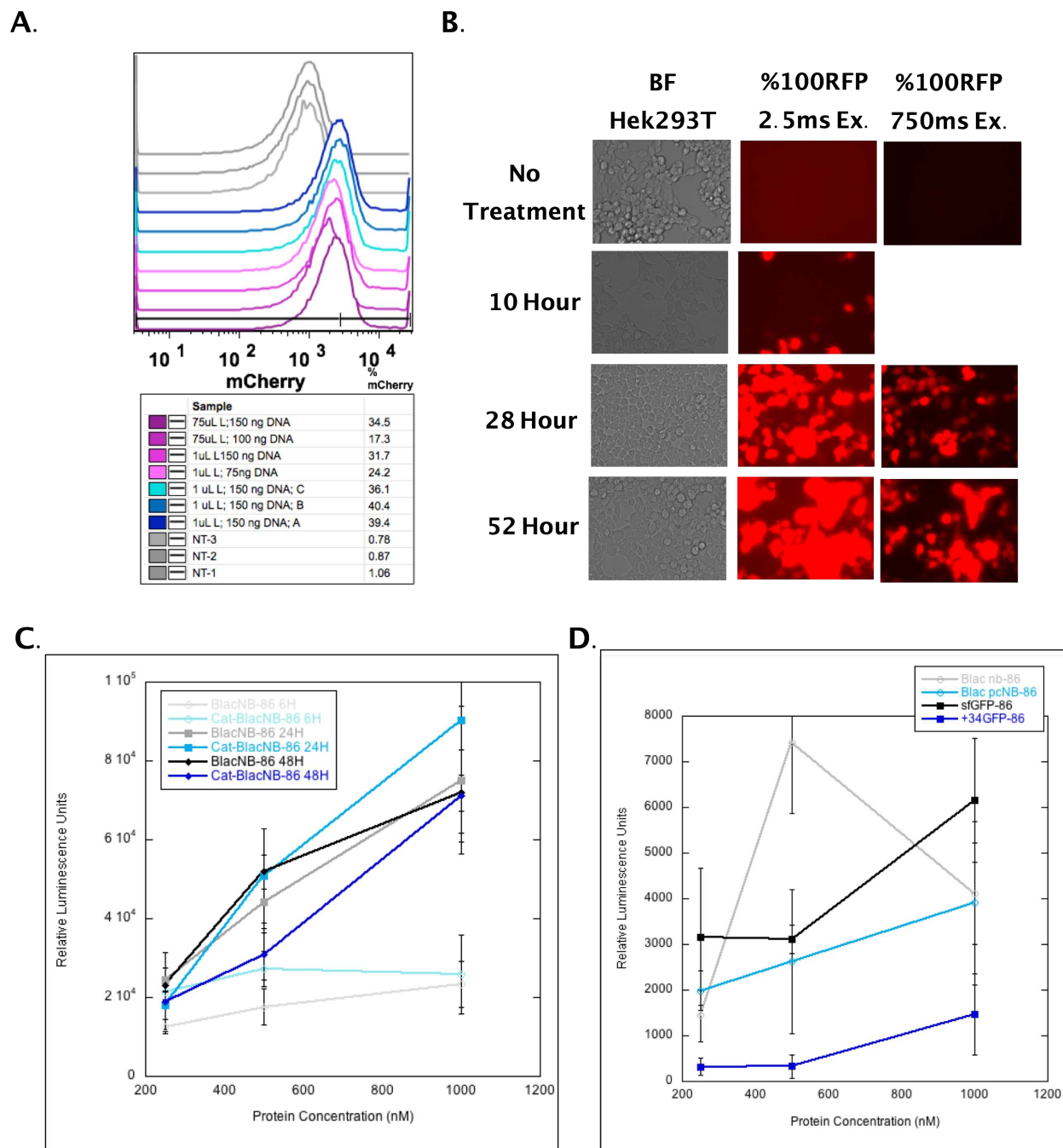


Figure 4.4) a. Flow cytometry of transfection conditions visualizing mCherry-LgBit in Hek293T cells. b. Evos Fluorescence images of transfections at three different time points. c. Using previously determined transfection conditions, we analyzed three time points of protein incubation and relative fluorescence using a 24-well plate. d Following Promega protocol for measuring relative fluorescence in a 96-well plate of our four proteins.

4.3.3 High copy plasmid conditions

Since there seemed to be no change between a 96-well plate and 24-well plate as far as BlacNB-86 and sfGFP-86 giving high luminescence readings, we decided to move back to the 24-well plate in order to maintain a healthy amount of cells in the wells for the final reading. After washing the 96-well plates, we noticed the amount of cells that remained adherent may have been causing some issues, and a 24-well plate may be more accurate for our application. Previously, we had been using a low copy plasmid known as pIRES, and this could possibly be giving us a lower translation of mCherry-LgBit then could be needed. We decided to switch to a new plasmid, H2B, a high copy plasmid that would hopefully increase the levels of mCherry-LgBit inside the cell. We decided to re-examine the transfection conditions using our new plasmid, comparing our previous conditions against conditions that are recommended by Sigma Aldrich for a 24-well plate (**Figure 4.5a**). We discovered that, with the H2B plasmid, 500ng DNA appeared to produce more mCherry fluorescence when visualized with an Evos Fluorescence Imager. There seemed to be minimal change between 1uL and 2uL Lipofectamine, therefore we decided to continue with 500ng of the H2B plasmid with 1uL Lipofectamine per well.

Another hypothesis to double check was that of the change in transfection timelines. We had been transfecting our cells for four hours prior to any protein incubation. This could potentially not be allowing enough mCherry-LgBit to be produced before the protein is added. We imaged the cells at four hours with the H2B plasmid and saw a very low rate of translation. We decided to continue incubating the transfection mixture with the Hek293T cells for a total of 12 hours prior to the addition of protein. Proteins were incubated for 24 hours at 500nM and 1uM, cells were washed with PBS containing heparin sulfate and imaged on the EVOS

fluorescence imager. As previously reported, sfGFP does not enter mammalian cells and 34GFP enters cells, but resides mostly in the endosomes. Our fluorescence images would fight to confirm this observation (**Figure 4.5b**).

There is also a chance that each well may be containing various levels of mCherry-LgBit protein production and thus altering some results. We tested our transfected 24-well plate for mCherry fluorescence using a micro-plate reader prior to the addition of NanoGlow and subsequent luminescence reading. It was positive to see that each well contained almost the same level of mCherry fluorescence (**Figure 4.5c**). Therefore, we proceeded with adding the NanoGlow reagent to our cells, and it remained surprising to us the sfGFP-86 fluorescence was higher than 34GFP-86, as well as BlacNB-86 remained higher than Cat-BlacNb-86 (**Figure 4.5d**).

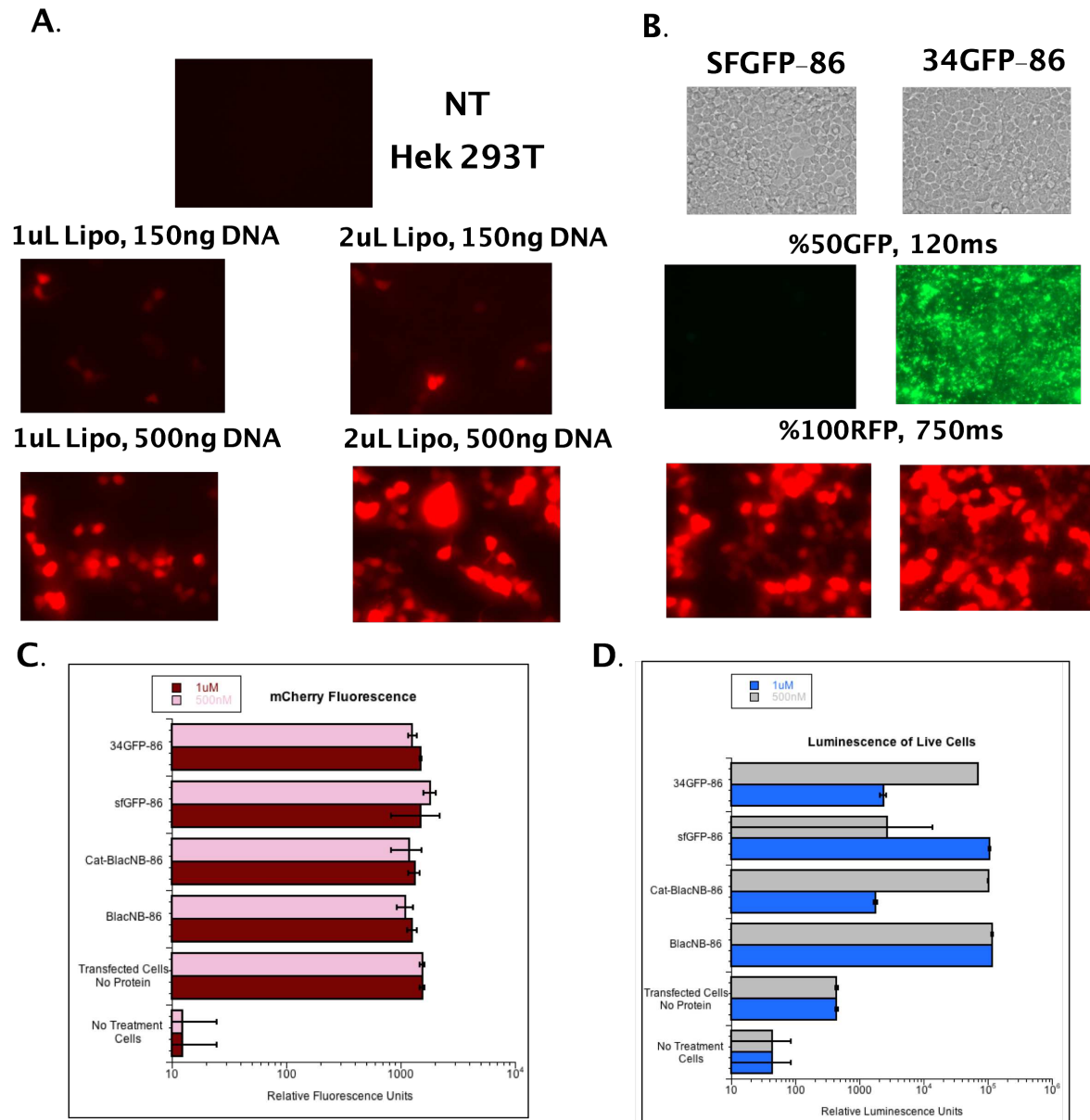


Figure 4.5) High copy plasmid H2B was tested in a 24-well plate to help increase levels of mCherry-LgBit within the cell A. Transfection amounts of H2B DNA and Lipofectamine 2000 on an Evos Fluorescence Imager B. Evos images of wells after incubation of sfGFP-86 and 34GFP-86 with PBS-HS washing. C. Micro-plate reader analysis of mCherry fluorescence in each well. D. Relative luminescence of each well at 500nM and 1uM D. Relative Luminescence of cells with protein incubation at two different concentrations.

4.3.4 Secretion tests

Originating back to the *in vitro* data where we discovered that BlacNB-86 gave a higher luminescence reading than Cat-BlacNB-86, we wanted to test if a new linker length may work better in our *in vivo* setting for the two proteins. The previously tested proteins contained a 10 amino acid linker (GGGS)₂, which were adopted from the original NanoBit paper when measuring their protein-protein interactions. We cloned two more proteins (BlacNB-86 and Cat-BlacNB-86) containing a shorter linker of 5 amino acids and ligated these into the H2B high copy plasmid.

From here, our next few assays hoped to answer two different hypothesis, one was the possibility that mCherry-LgBit was secreting from the cell, therefore reaching the minimal amount of sfGFP-86 and BlacNB-86 protein that may be left on the outside of the cell. The second was that the 86-tagged could be cleavage by proteases and therefore internalized, reaching the mCherry-LgBit when it shouldn't be. We transfected our H2B plasmid into Hek293T cells and incubated for 12 hours. We then collected the media, added exogenous purified sfGFP-86 protein to each sample, and added the NanGlow reagent to the media and measured for luminescence. Remaining plasmid was not a concern to us, since only translation of the mCherry-LgBit protein would result in luminescence. We observed a small level of luminescence in the transfected cells well, indicating that some mCherry-LgBit must be secreting from the cells (**Figure 4.6a**).

We had a simultaneous plate incubating such that we could continue on with a 24-hour assay. We switched out the media after 12 hours of transfection and incubated our proteins at 500nM for 24 hours. This time, protease inhibitor was added to our proteins to hopefully cut down on the possibility of cleavage during their 24-hour incubation with the cells. We also tested

our four proteins, as well as BlacNB-86 and Cat-BlacNB-86 with our shorter linker. After another 24 hours, media was collected after the incubation and placed in a separate plate for secretion analysis (**Figure 4.6b**). Cells were washed with PBS contain heparin sulfate, imaged on the evos (data not shown) and ran the NanoBit assay on both cells and media. As expected, secretion of mCherry-LgBit did increase after 24 hours; however, when compared to the luminescence produced from the cells, it was encouraging to see that the luminescence produced by the cells were consistently high. These results suggest that washes of the cells after the transfection and shorter protein incubation times may help decrease the background of secretion luminescence.

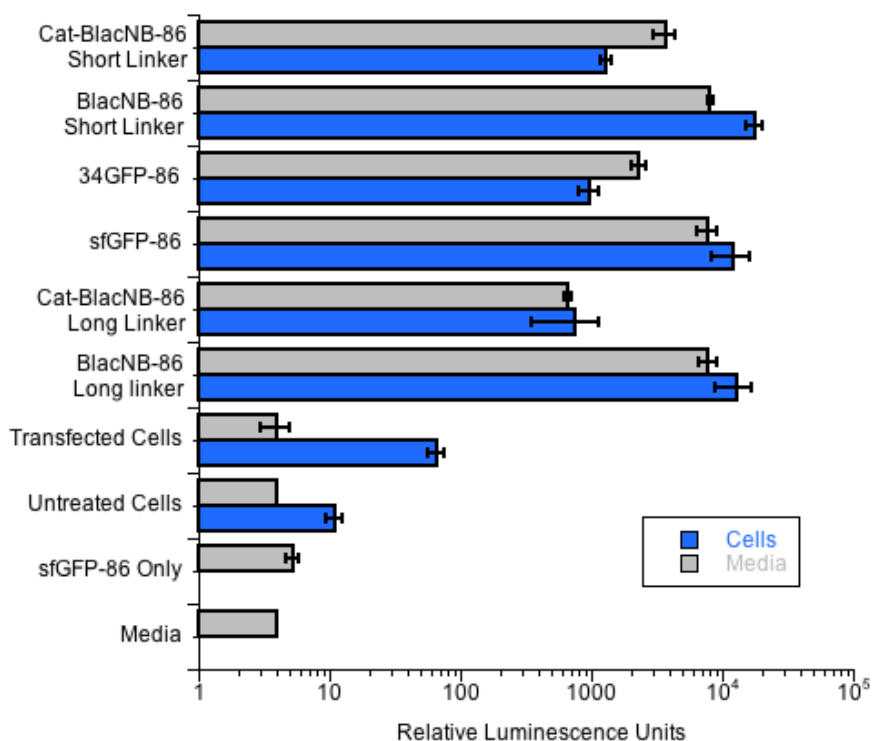


Figure 4.6: Relative luminescence of cells and media after exogenous protein incubation. Using the H2B plasmid for transfection, secretion appeared to be relatively high, but intracellular luminescence from the cells remained a stronger signal.

4.3.5 Luminescence of Cells in Suspension

Another hypothesis extends from our constant question of whether the number of cells in each well is sufficient to produce the necessary signal. In the 24-well plates, we had been noticing pockets of lost adherent cells in each well and thus the variation between cells could be causing these unexpected results. To answer this question, we wanted to see if measuring luminescence of live cells in suspension would produce a more accurate reading when measured in a 96-well plate. As before, Hek293T cells were plated in a 24-well plate and transfected with the H2B high copy plasmid for 12 hours. After 12 hours the media was removed and proteins were added at 500nM to incubate for another 12 hours. This shorter incubation would hopefully cut back on the level of mCherry-LgBit that is secreted from the cells. After 12 hours, the media was removed and trypsin was added to each well in order to resuspend the cells and remove exogenous proteins. Cells were pelleted and resuspended in PBS containing 20U Heparin sulfate twice, pelleted a third time and resuspended in PBS. Triplicate samples of 100uL each were placed in a 96-well and RFP and GFP were measured on a microplate reader. Surprisingly, mCherry fluorescence was not measured in any of the wells; however, green fluorescence was efficiently measured, and only 34GFP-86 wells resulted in a GFP signal, as expected (**Appendix C, Figure C.4**). It was unclear why there seemed to be a discrepancy between plated cells and resuspended cells for mCherry fluorescence measurements, but we proceeded forth with the luminescence assay. NanoGlow live cell reagent was added to each of the resuspended wells and measured on the microplate reader. It was unfortunate to observe that in a live cell suspension, absolutely no luminescence was measured (**Appendix C, Figure C.4**).

4.3.6 Comparison of Mammalian Expression Plasmids

Next, we explored the possibility that the H2B high copy plasmid was contributing an excess amount of mCherry-LgBit translation, and thus increasing the level of secretion. We decided to compare our two plasmids, the previous low copy pIRES, and the high copy H2B, over a variety of timelines. Subsequently, we wanted to consider the possibility that the luminescence we were getting could potentially be a baseline reading, and a complete positive result would produce a significantly higher luminescence. Therefore, we cloned a full version of the NanoBit protein (with no cleavage of the 86 peptide) fused to mCherry in the pIRES plasmid in order to visualize the maximum possible luminescence of our assay. We transfected our Hek293T cells in 24-well plates for 12 hours with all three plasmids. Our BlacNB-86 and Cat-BlacNB-86 proteins with short linkers were incubated at 500nM for 6, 18 and 24 hours (**Figure 4.7a,b**). Media was collected at each time point to measure the secretion levels, and cells were washed with PBS containing Heparin Sulfate to remove exogenous protein. It appeared that 18 hours of protein incubation produced the highest luminescence; however, amplitude of signal was not significantly different between the H2B and pIRES plasmids. Particularly in regards to BlacNB-86, the cells luminescence is similar between plasmids, and the media luminescence is within error of one another. Cat-BlacNB-86 cells and media appear to give a higher signal using the pIRES plasmid, but these results are within error of the H2B plasmid. Unfortunately, these changes continued to result in the BlacNB-86 protein producing a more positive signal than our Cat-BlacNb-86.

It was encouraging to see that the pIRES control, the transfected mCherry-NanoBit protein, only produced luminescence within the cells, and absolutely no secretion was measured in the media. This suggested to us that the NanoBit protein is not being secreted after LgBit and

SmBit recombine within the cell, but that minute amounts of the BlacNB-86 protein could be residing outside of the cell after all washes, and thus producing a positive signal. In our previous studies, this extremely small existence of wild-type nanobodies “within” the cells could be seen in flow cytometry by less than 2-fold over untreated cells. However, the binding affinity of the 86-peptide with LgBit is so low that picomolar amounts of exogenous protein left behind could result in a significant luminescent signal, such as the one we are seeing. It was now our hypothesis that a lower luminescence seen in 34GFP-86 and Cat-BlacNB-86 proteins could be the result of internalization of the protein, leaving less remaining on the outside of the cell, but most likely ending their journey trapped within the endosomes.

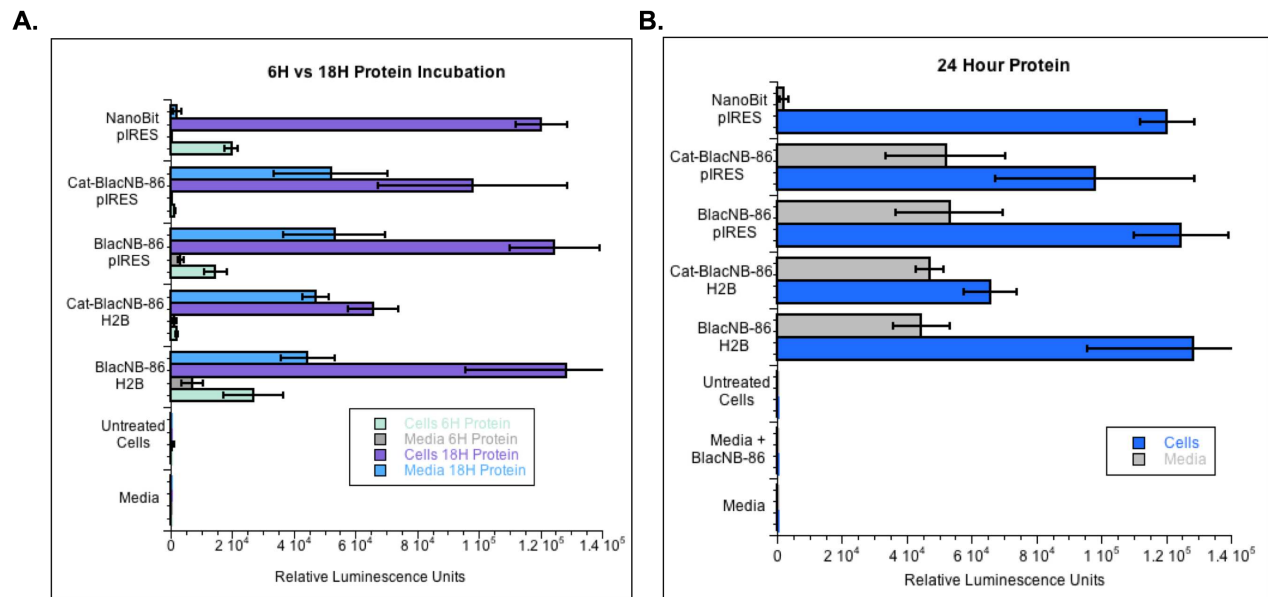


Figure 4.7: Relative luminescence of cells incubated with exogenous protein at various timelines A. Proteins were incubated for 6 hours and 18 hours and then luminescence of cells vs the media were measured. B. proteins were incubated for 24 hours prior to washes with PBS containing heparin sulfate and subsequent measurement of luminescence. For both graphs, luminescence in the media would only occur if there was secretion of the mCherry-LgBit to meet the 86-tagged proteins on the exterior of the cell.

4.3.7 Protein Interactions within Co-Transfected Cells

Moving forward, we wanted to make sure that if our resurfaced Cat BlacNB-86 and 34GFP-86 were successfully reaching the cytosol, that the NanoBit protein was able to recombine within the complex environment of the cell. It was encouraging that only cellular luminescence was measured by the mCherry-NanoBit protein, and verified to us that the NanoGlow reagent was able to access the full protein within the cell and produce a signal. To prove our assay could work in theory, we created four new plasmids using the pIRES vector due to its two multiple cloning sites. We cloned in BlacNB-86 and Cat-BlacNB-86 with the previously used short linkers, and BlacNB-86 and Cat-BlacNB-86 with longer linkers into one cloning site (individually), while the other contained the mCherry-LgBit. This way, both halves of our split NanoBit would be co-translated within the cytosol of the cell, letting us observe if their recombination is functional within the cell. We plated Hek293T cells in 24-well plates and transfected them with 500ng DNA of each individual plasmid and 1uL Lipofectamine per well and incubated for 24 hours. A longer transfection period was chosen over the previous 12 hours since there was no subsequent addition of exogenous protein for an extended period of time. After 24 hours, media was collected to measure for secretion of full protein, cells were washed, and the NanoGlow reagent was added to each well. To our satisfaction, little to no luminescence was measured in the media, verifying that the recombined NanoBit protein is not secreted. We were also happy to see luminescence from all plasmids, concluding that the overall vision for the assay could work with successful cytosolic binding (**Figure 4.8**). However, we still observed a higher signal for the BlacNB-86 protein, whether long or short linker, over our Cat-BlacNB-86 protein. We now believe that there is something inherent about the cationic resurfacing that is disturbing the recombination of the split NanoBit protein.

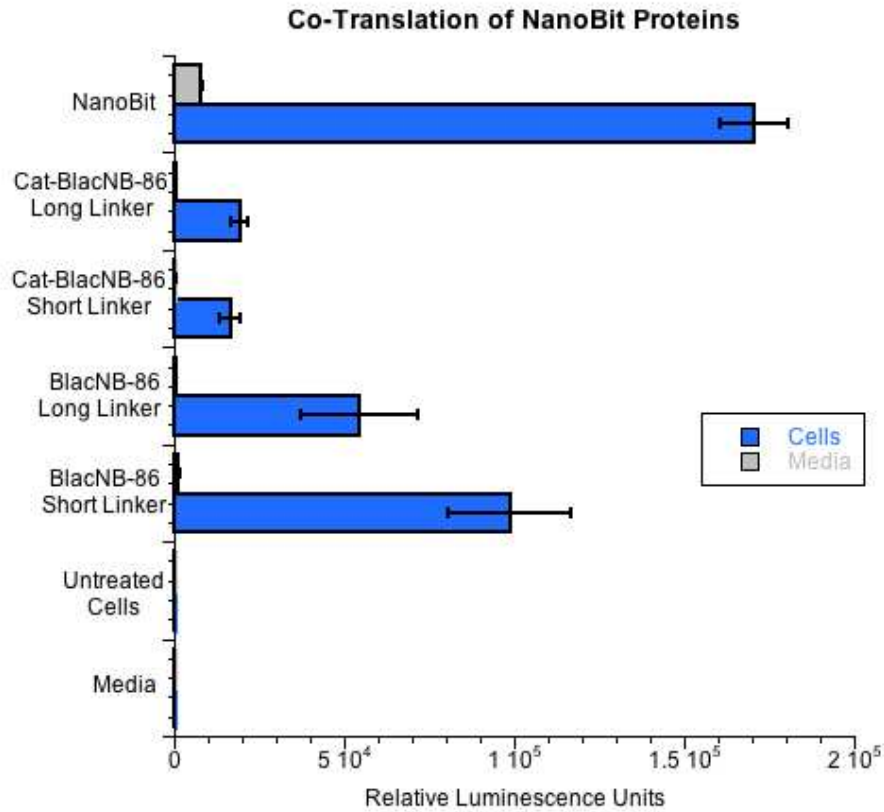


Figure 4.8: Transfection of bicistronic plasmid pIRES containing mCherry-LgBit in multiple cloning site 1 and POI-86 proteins in multiple cloning site 2. Full length NanoBit was also used as a control for maximum luminescence measurements. As cells were transfected with each plasmid for 24 hours prior to luminescence readings. Only successful translation of the SmBit and LgBit components and intracellular recombination would produce a luminescent signal

4.3.8 Luminescence of Total Cell Lysate

We have now verified successful detection of intracellular binding of our Split NanoBit fusion proteins within in cytosol of mammalian cells. We now want ed to understand why there is consistently a higher luminescence for our proteins that should not be entering the cell, versus ones that have been previously reported to enter cells. Our hypothesis was that residual BlacNB-86 and sfGFP-86 protein could remain on the outside of our cells, and our internalized proteins

are existing within the endosomes. We sought to perform a total cell lysate in order to break open endosomal contents and see if this would discover our “hidden” luminescence signal for our cationic proteins.

Now that we were sure that the assay could work in practice, we wanted to return the high-throughput set up a 96-well plate would offer. Throughout our experiments, we have not observed a significant difference between the 96-well plates and 24-well plates, therefore we were comfortable moving forward with optimizing a 96-well format. Hek293T cells were transfected with 200ng DNA each of our control pIRES plasmids including the short linker BlacNB-86 and Cat-BlacNB-86, as well as the full length mCherry-Nanobit protein and wells with just the mCherry-LgBit protein. Remaining cells were transfected with mCherry-LgBit and incubated for 24 hours. All four of our 86-tagged proteins were incubated at 1uM and 2uM for 1 hour (**Figure 4.9b**). The media was removed and trypsin was added to each well to remove and pellet the cells. Cells were resuspended in PBS containing heparin sulfate and pelleted two consecutive times to remove exogenous protein. The cell pellets were then resuspended in 300uL RIPA buffer, a strong lysis buffer, vortexed and incubated on ice for 30 minutes. Cells were then pelleted at 14,000 RPM and the supernatant was collected to analyze luminescence of the soluble protein content. Each sample was separated into triplicates in a white 96-well plate and the NanoGlow reagent was added (**Figure 4.9**). Interestingly enough, the overall relative luminescence was orders of magnitude lower than any previously run assay. We suspected these samples would act similarly to our *in vitro* assays and were surprised to see such low signals. However, it was encouraging that this was the first time we had gotten a higher luminescence for our cationic proteins than their wild type counterparts.

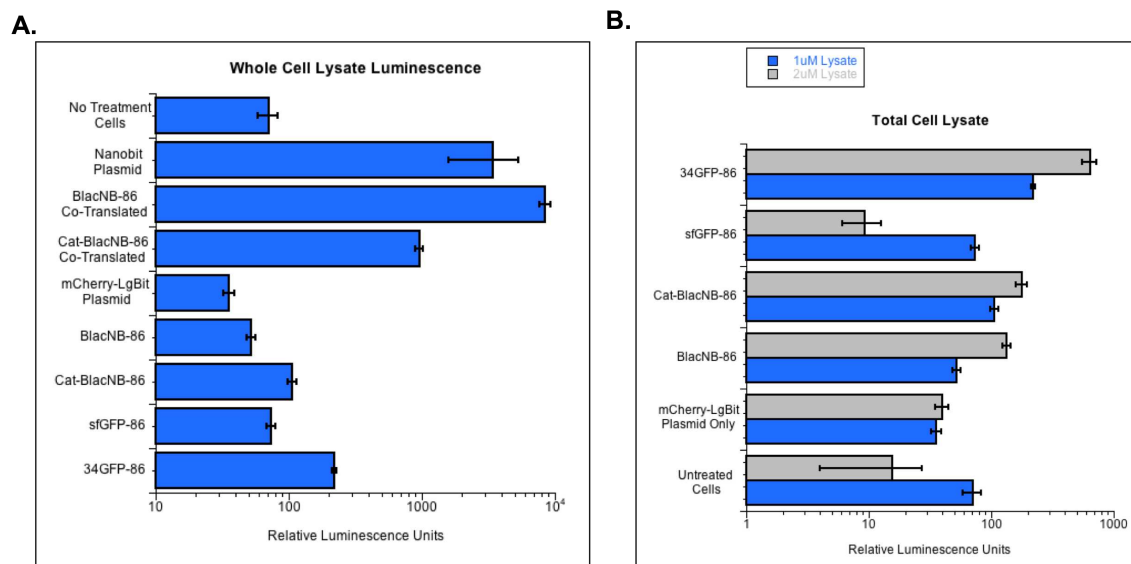


Figure 4.9: Relative luminescence of whole cell lysates to discover endosomally trapped proteins. A. Luminescence of all proteins and control plasmids. For the first time, the cationic resurfaced proteins give a higher luminescence than wild-type counterparts. B. Luminescence of lysates after cells were incubated with two different protein concentrations.

4.3.9 Endosomal Release of Cationic Resurfaced Proteins

Finally, we set out to clone on endosomal escape peptides discovered by the Schepartz and Lui lab to enhance the signal of our internalized proteins.²⁷ We cloned on the peptide A8 (GLFDIHKRIAESF) to the N-terminal of all our proteins of interest. Unfortunately, the addition of this second tag, to create A8-POI-86, drastically decreased the successful expression and thereby purification of the majority of our proteins. What purified proteins we did have we tested in our in vitro assay with Hek293T cell lysate to confirm recombination of Nanobit and subsequent luminescence (**Appendix C, Figure C.5**). The results from the in vitro test were not successful enough for us to wish to continue down this path to accomplish endosomal escape of our proteins.

Next, we thought to utilize the chloroquine, a small molecule known to enhance escape of proteins and macromolecules trapped within endosomes.²¹ We hoped to see an increase of

luminescence in our cationic proteins if they are in fact getting trapped in these intracellular vesicles. To do this, Hek293T cells were plated in a 96-well plate and transfected with our four control plasmids at 200ng DNA and 1uL Lipofectamine per well. The cells incubated with the transfection mixture for 24 hours before the media was removed and two simultaneous tests were run. Our first plate of cells was incubated with 1uM of our four 86-tagged proteins, while our second plate of cells incubated with 1uM of proteins along with 100uM of chloroquine. The proteins incubated for 1 hour, the media was removed and the cells were washed three times with PBS containing heparin sulfate. From here, each plate was examined for luminescence using the NanoGlow Live cell reagent (**Figure 4.10**). Chloroquine has potential to increase the levels of translation of a transfected plasmid, but we were happy to see that the level of luminescence remained relatively equal in our control plasmids. This gave us confidence that the differences between the exogenously delivered proteins in the normal cells vs chloroquine cells were due to endosomal release of protein into the cytosol. We see a significant increase in luminescence in the Cat-BlacNB-86 chloroquine wells than the normal run assay, but we did not see much of a change in the supercharged GFP-86 wells. Regular sfGFP-86 remained higher than supercharged GFP, and BlacNB-86 in the normal assay remained higher in luminescence than Cat-BlacNB-86 in the normal cells. To obtain a more solid conclusion on the existence of Cat-blacNB-86 within the endosomes, more tests would need to be run. However, this did get us a step closer to understanding the exact mechanism of action in developing our internalization assay.

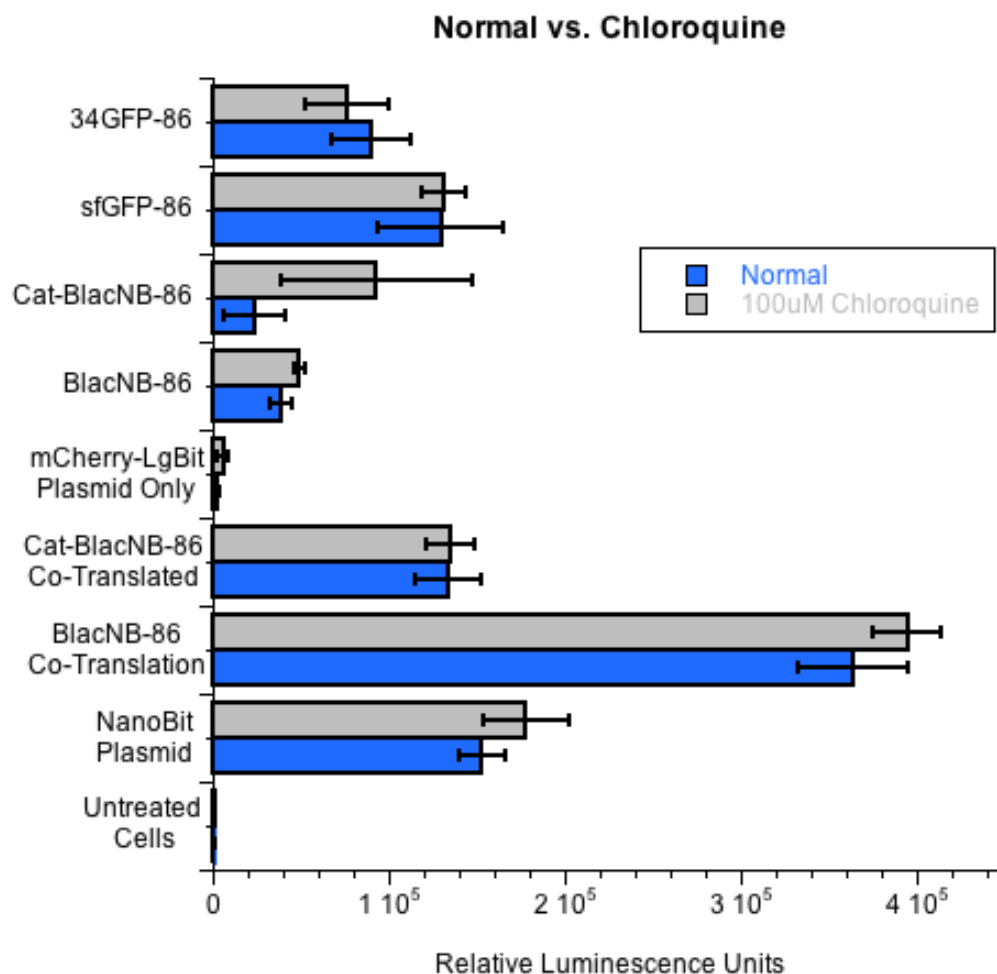


Figure 4.10: Addition of chloroquine, a small molecule known to enhance endosomal escape of trapped proteins. Again, it was observed that a higher luminescence is produced for the Cat-BlcNB-86 when chloroquine is added to the cells, suggesting entrapment of the protein. However, sfGFP-86 remains higher in luminescence over 34GFP-6, and there is little to no change between normal cells and those with addition of chloroquine

4.4 Conclusion and Future Directions

We initially set out to develop a facile assay that produces a high-throughput, stable luminescent signal for identification of exogenous proteins that have been delivered to the cytosol of cells. Previous assays developed for these purposes, such as cre-recombinase and split GFP or β -lactamase, require multiple large bulky fusions and require multiple domains for protein transduction, endosomal escape, and subsequent DNA translation to produce a

fluorescent signal of successful delivery. The binding orientation of proteins of interest to their target may be drastically altered by the addition of some many delivery and reporter proteins. Thus, it would be much simpler if a 13-amino acid tag (the 86-peptide) could be fused to any POI, or in our case a resurfaced nanobody, and delivered to the cytosol to recombine with LgBit and produce a stable luminescent signal. This assay can be easy adapted to a 96-well plate format, and is adaptable to any desired protein for study with little interference of native binding interactions.

We had analyzed multiple transfection conditions for our mCherry-Lgbit reporter, and found that 200ng of pIRES DNA with 1uL Lipofectamine per well supplied a significant level of mCherry fluorescence. We also established a 12 or 24-hour transfection time period, depending of the extent of protein incubation that is desired. A 24-hour transfection is sufficient to use when proteins are incubated for 1 hour, whereas a 12-hour transfection is useful for proteins with incubation of 6 hours or more. Addition of the NanoGlow live cell reagent is functional in both a 24-well or 96-well assay format and produces a substantial luminescent signal over untreated cells.

Most importantly, we have proven that the assay is functional when the genes of the SmBit and LgBit fusions are co-translated within Hek293T cells. This verified that the NanoBit components were able to reassemble and produce a strong luminescent signal from within live mammalian cells. We investigated the possibility that our exogenously delivered proteins were becoming trapped within endosomes, thus producing less of a signal then their wild-type counter parts, by analyzing total cell lysates and inducing endosomal escape. In both cases, we did observe a higher luminescent signal of our Cat-BlacNB-86 over the BlacNB-86, however, both were relatively small differences. In fact, we saw successful tests of both positive proteins when

the cells were completely lysed open, but the luminescent signal was orders of magnitude lower than the live cell assays.

If these tests were to continue, there are a few hypotheses that would be tested. First, a new peptide with a lower binding affinity would be selected to re-clone all examined proteins. A lower binding affinity could produce less of a background signal of the residual protein that is not completely washed away by heparin sulfate. The potential down side is that the SmBit and LgBit would have a harder time interacting and binding with one another in the cytosol, and therefore produce a lower positive signal overall. Second, linker engineering and new orientations should be made for the positively charged since there seems to be something inherently restricting or limiting with the versions used in this paper. Cationic resurfacing has been known to reduce aggregation, while sfGFP has been proven to aggregate. An aggregated fluorescent protein can intensify the overall signal, and therefore aggregation could have a potential amplifying affect on luminescence. Perhaps other protein transduction domains could be utilized to test this theory. This also leads into a third area of experimentation, the addition of proteins and protein transduction domains proven to access the cytosol to provide more accurate positive controls. For a short period, TAT-GFP-86 was cloned to help examine this theory. Unfortunately, there were too many difficulties in the expression and purification of the protein that these tests were never conclusive enough to comment on (data not shown). Overall, we believe every aspect of minor troubleshooting experiments were performed, and really only large reconstruction of all interacting components would be able to create a fully functional assay. However, the proof of concept and working theory has been shown to be a positive start to the future development of a facile intracellular detection assay.

4.5 Materials and Methods

Plasmids were either obtained from the Stasevich lab (H2B) or through Addgene for pIRES and pETDuet. All primers for overlap PCR of the 86-peptide and Gblocks for 34GFP and Cat-BlacNB were purchased from IDT. Restrictions enzymes Nco1 and KPN1, Vent DNA polymerase and ligation enzymes and buffers were purchased from New England Biolabs. LB and 2XYT broth were obtained from Sigma. 96-well and 24-well tissue cultured plates were purchased from Corning. Lipofectamine 200 was purchase from Life Technologies. Heparin Sodium Sulfate from porcine intestinal mucosa, Grade I-A was purchased from Sigma Aldrich. NanoGlow Live Cell Reagent was purchase from Promega and luminescence was measure using a Biotek microplate reader. Images were collected using an Evos EpiFluorescence Microscope.

4.5.1 Cloning and Protein Sequences

Proteins were cloned using Nco1 and KPN1 restriction cut sites into pETDuet for protein expression, NHe1 and BglII for H2B, and . The 86-peptide was cloned with two linkers (GGGGS) and (GGGGS)₂ through overlap PCRs. For the Aurein 8 endosomal escape tag (A8) a (GGG)₃ linker was cloned on the N-terminus of the proteins. Protein sequences are as follows:

Mcherry-LgBit:

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWD
ILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDEGGVVTVTQDSSLQDGEFIYKVKLR
GTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQ
LPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKSGLRSGGSGGSGGSGGVFTLEDF
VGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQI

EEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIID
ERLITPDGSMLFRVTINS

mCherry-NanoBit:

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWD
ILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLR
GTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQ
LPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYKSGLRSGGSGGSGGSGGVFTLEDF
VGDWEQTAAYNLDQVLEQGGVSSLLQNLAVSVTPIQRIVRSGENALKIDIHVIIPYEGLSADQMAQI
EEVFKVVYPVDDHHFKVILPYGTLVIDGVTPNMLNYFGRPYEGIAVFDGKKITVTGTLWNGNKIID
ERLITPDGSMLFRVTINSVSGWRLFKKIS

BlacNB-86 short linker:

MAQVQLVESGGGSVQAGGSLRLSCTASGGSEYSYSTFSLGWFRQAPGQEREAVAAIASMGGLTYYA
DSVKGRFTISRDNKNTVTLMNNLKPEDTAIYYCAAVRGYFMRLPSSHNFYWGQGTQVTVSS
HHHHHHGGGGSVSGWRLFKKIS

** Longer linker adds another GGGGS after the his6

Cat-BlacNB-86 short linker:

MAQVQLVEKGGGKVRAGGKLRLRCTASGGSEYSYSTFSLGWFRQAPGREREAVAAIASMGGLTYY
ADSVKGRFKIKRDNKNTVTLMNNLKPEDTAIYYCAAVRGYFMRLPSSHNFYWGQGTQVTVSS
RHHHHHHGGGGSVSGWRLFKKIS

** Longer linker adds another GGGGS after the his6

sfGFP-86:

SKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTCLKFICTTGKLPVPWPTLVTTLTYG
VQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNFNHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALS KDPNEKRDHMLLEFVTAAGITHGMDELYKHHHHHHGGGGSGGGGSVSGWRL
FKKIS

34GFP-86:

MGGASKGERLFRGKVPILVELKGDVNGHKFSVRGEGKGDATRGLTLCLKFICTTGKLPVPWPTLV
TLTYGVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKYKTRAEVKFEGRTLVRNRIKLKG
RDFKEKGNILGHKLRYNFNHSHKVYITADKRKNGIKAKFKIRHNVKDGSVQLADHYQQNTPIGRGP
VLLPRNHYLSTRSKLSKDPKEKRDHMLLEFVTAAGIKHGRDERYKHHHHHHGGGGSGGGGSVS
GWRLFKKIS

4.5.2 Protein Expression and Purification

Plasmids were transformed into BL21s (DE3). Cells were grown in either 1000 or 500 mL LB or 2XYT cultures containing carbenicillin at 30C to OD₆₀₀ of 0.7 and induced with 1 mM IPTG at 20C overnight. Cells were then collected by centrifugation and resuspended in either phosphate buffer with 150mM NaCl for wild type protein BlacNB and sfGFP (20mM Sodium Phosphate, pH 7.4) or resuspended in phosphate buffer with 2M NaCl for Cationic proteins (20mM Sodium Phosphate, pH 7.4) and stored at -20C. Frozen pellets were thawed and incubated with complete ULTRA protease inhibitors tablets then sonicated for 4 min. The lysate was cleared by centrifugation (8000 rpm, 10 min) and the supernatant was mixed with 1 mL Ni-NTA resin for 1 hour. The resin was collected by centrifugation (4750 rpm, 10 min). The resin

was washed with 20mL buffer and 20mM imidazole then 30mL buffer and 50mM imidazole and finally 20mL 75mM imidazole. The protein was then eluted with 4 mL buffer containing 400 mM imidazole. The proteins were dialyzed against 1M NaCl phosphate buffer for cationic proteins and 150mM phosphate buffer for wild type proteins and analyzed for purity by SDS-PAGE. Purified proteins were quantified using absorbance at 280 nm on a NanoDrop.

4.5.3 In Vitro Assays

For the PBS assay, proteins were diluted at four different concentrations in PBS in 100uL. The NanoGlow Reagent Buffer was made at 1:50 dilution ratio of reagent to buffer following Promegas protocol. 22.1uL of each protein concentration was added to wells in triplicate in a 96-well plate. 2.8uL of purified mCherry-LgBit was added to each well to make a final sample volume of 25uL. Then, 25uL of the NanoGlow reagent mix was added to each well and luminescence was measured.

For the cell lysate assay, 25uL of purified protein at 2X the desired concentration was added to each well in triplicate in a 96-well plate. 25uL of Hek293T cell lysate was added to each well for a total volume of 50uL. The NanoGlow Live Cell Reagent was used in this assay and the reagent mix is created at a 1:19 ratio reagent to buffer. Promega's protocol calls for the addition of 25uL of this mix to each well of 100uL, since each well contained 50uL, 12.5uL was added in this scenario and luminescence was measured.

4.5.4 Mammalian Cell Culture

Hek293T cells were grown in High Glucose DMEM containing 10% Fetal Bovine Serum (FBS) and incubated at 37C with 5% CO₂. When cells were plated in 24 or 96-well plates, cells

were allowed 24 hours to adhere before transfection. Transfections were carried out by mixing ethanol precipitated DNA and Lipofectamine 2000 in Optimem and adding the appropriate amount of mixture according to Invitrogen protocol for the respective plate size. Proteins were incubated with cells in High Glucose DMEM with no FBS at desired concentrations and incubated at 37C and %5 CO₂. Cells were washed with PBS containing 20U heparin sulfate and placed in Optimem for the incubation of the NanoGlow Reagent.

4.5.5 NanoBit Luminescence Assays

Depending on the assay type determined which NanoGlow reagent was used. In vitro assays, cell lysates, and media collection for secretion assays used the NanoGlow Luciferase Assay Reagent. This reagent was diluted 1:50 reagent to buffer to create the reagent mixture and added to wells in :1 volumes. For live cell assays and cells in resuspension, NanoGlow Live Cell Assay Reagent was made in a :19 dilution ratio of reagent to buffer. 25uL of reagent mix was added to 100uL samples, and adjusted according for varying samples volumes. Luminescence for all were read immediately after addition of reagent in a BioTek microplate reader.

REFERENCES

- (1) Sheng, C.; Georg, G. I. Targeting Protein-Protein Interactions by Small Molecules. *Target. Protein-Protein Interact. by Small Mol.* **2018**, 1–329. <https://doi.org/10.1007/978-981-13-0773-7>.
- (2) McFedries, A.; Schwaib, A.; Saghatelian, A. Methods for the Elucidation of Protein-Small Molecule Interactions. *Chem. Biol.* **2013**, 20 (5), 667–673. <https://doi.org/10.1016/j.chembiol.2013.04.008>.
- (3) de la Torre, B. G.; Albericio, F. The Pharmaceutical Industry in 2018. An Analysis of FDA Drug Approvals from the Perspective of Molecules. *Molecules* **2019**, 2017. <https://doi.org/10.3390/molecules24040809>.
- (4) Lagassé, H. A. D.; Alexaki, A.; Simhadri, V. L.; Katagiri, N. H.; Jankowski, W.; Sauna, Z. E.; Kimchi-Sarfaty, C. Recent Advances in (Therapeutic Protein) Drug Development. *F1000Research* **2017**, 6, 113. <https://doi.org/10.12688/f1000research.9970.1>.
- (5) Pelay-Gimeno, M.; Glas, A.; Koch, O.; Grossmann, T. N. Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew. Chemie - Int. Ed.* **2015**, 54 (31), 8896–8927. <https://doi.org/10.1002/anie.201412070>.
- (6) Wigglesworth, M. J.; Murray, D. C.; Blackett, C. J.; Kossenjans, M.; Nissink, J. W. M. Increasing the Delivery of next Generation Therapeutics from High Throughput Screening Libraries. *Curr. Opin. Chem. Biol.* **2015**, 26, 104–110. <https://doi.org/10.1016/j.cbpa.2015.04.006>.
- (7) O’Neil, K. T.; Hoess, R. H. Phage Display: Protein Engineering by Directed Evolution. *Curr. Opin. Struct. Biol.* **1995**, 5 (4), 443–449. [https://doi.org/10.1016/0959-440X\(95\)80027-1](https://doi.org/10.1016/0959-440X(95)80027-1).
- (8) Binz, H. K.; Amstutz, P.; Plückthun, A. Engineering Novel Binding Proteins from Nonimmunoglobulin Domains. *Nat. Biotechnol.* **2005**, 23 (10), 1257–1268. <https://doi.org/10.1038/nbt1127>.
- (9) Chapman, A. M.; McNaughton, B. R. Resurfaced Shape Complementary Proteins That Selectively Bind the Oncoprotein Gankyrin. *ACS Chem. Biol.* **2014**, 9 (10), 2223–2228. <https://doi.org/10.1021/cb5003834>.
- (10) Holub, J. M.; LaRochelle, J. R.; Appelbaum, J. S.; Schepartz, A. Improved Assays for Determining the Cytosolic Access of Peptides, Proteins, and Their Mimetics. *Biochemistry* **2013**, 52 (1), 9036–9046. <https://doi.org/10.1038/jid.2014.371>.
- (11) Larochelle, J. R.; Cobb, G. B.; Steinauer, A.; Rhoades, E.; Schepartz, A. Fluorescence Correlation Spectroscopy Reveals Highly Efficient Cytosolic Delivery of Certain Penta-Arg Proteins and Stapled Peptides. *J. Am. Chem. Soc.* **2015**, 137 (7), 2536–2541.

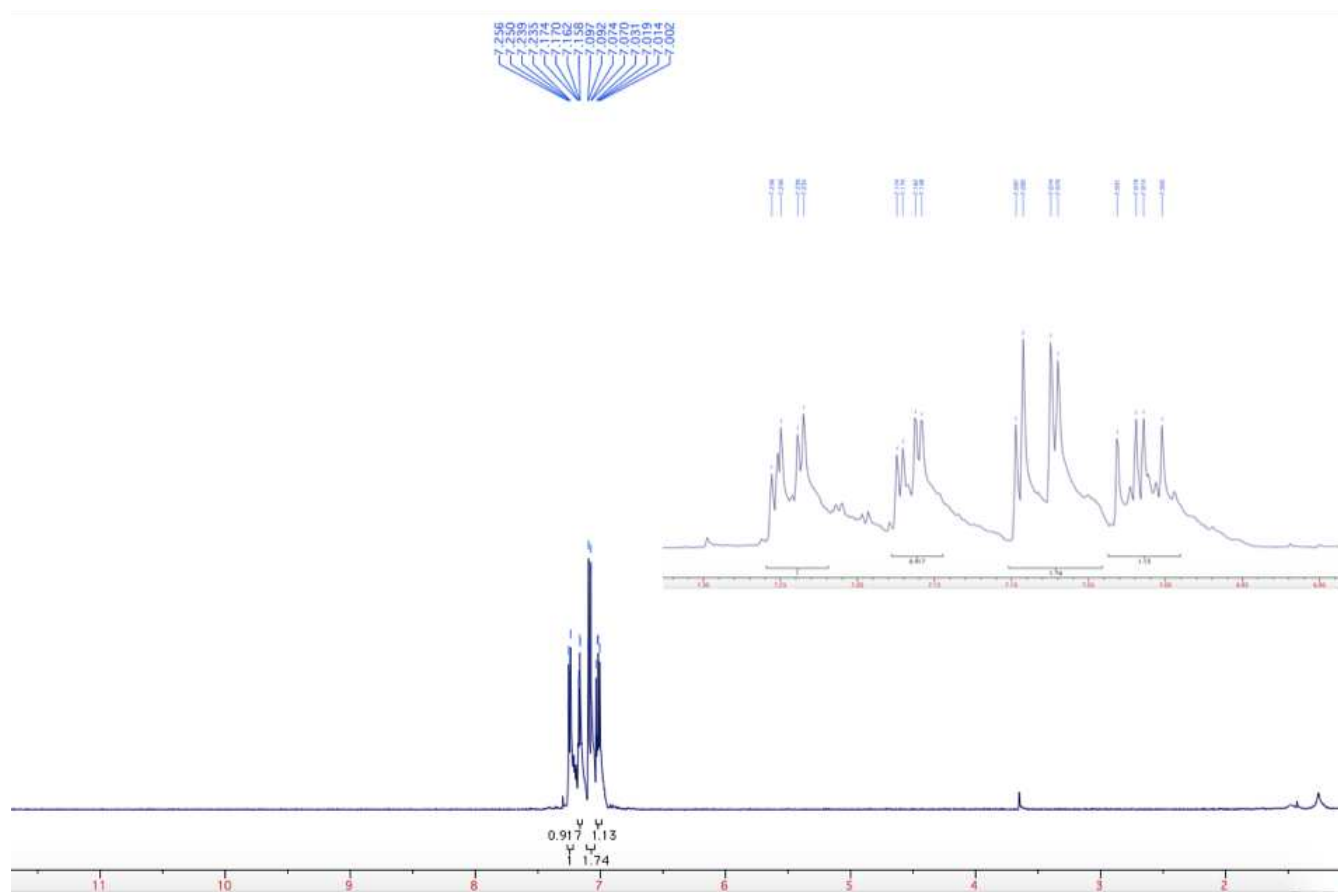
<https://doi.org/10.1021/ja510391n>.

- (12) Stone, S. R.; Heinrich, T.; Juraja, S. M.; Satiaputra, J. N.; Hall, C. M.; Anastasas, M.; Mills, A. D.; Chamberlain, C. A.; Winslow, S.; Priebatsch, K.; et al. B-Lactamase Tools for Establishing Cell Internalization and Cytosolic Delivery of Cell Penetrating Peptides. *Biomolecules* **2018**, 8 (3). <https://doi.org/10.3390/biom8030051>.
- (13) Wissner, R. F.; Steinauer, A.; Knox, S. L.; Thompson, A. D.; Schepartz, A. Fluorescence Correlation Spectroscopy Reveals Efficient Cytosolic Delivery of Protein Cargo by Cell-Permeant Miniature Proteins. *ACS Cent. Sci.* **2018**, 4 (10), 1379–1393. <https://doi.org/10.1021/acscentsci.8b00446>.
- (14) Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P.; Lubben, T. H.; Butler, B. L.; Binkowski, B. F.; MacHleidt, T.; Kirkland, T. A.; et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **2016**, 11 (2), 400–408. <https://doi.org/10.1021/acschembio.5b00753>.
- (15) England, C.; Ehlerding, E.; Cai, W. NanoLuc: A Small Luciferase Is Brightening up the Field of Bioluminescence. *Bioconj Chem* **2016**, 27 (5), 1175–1187. <https://doi.org/10.1021/acs.bioconjchem.6b00112>. NanoLuc.
- (16) Solomon, E. I.; Augustine, A. J.; Yoon, J. Firefly Luciferase Enzyme Fragment Complementation for Imaging in Cells and Living Animals. **2010**, 77 (30), 3921–3932. <https://doi.org/10.1039/b800799c>. O.
- (17) Auld, D. S.; Thorne, N.; Nguyen, D. T.; Inglese, J. A Specific Mechanism for Nonspecific Activation in Reporter-Gene Assays. *ACS Chem. Biol.* **2008**, 3 (8), 463–470. <https://doi.org/10.1021/cb8000793>.
- (18) Paulmurugant, R.; Gambhir, S. S. Monitoring Protein-Protein Interactions Using Split Synthetic Renilla Luciferase Protein-Fragment-Assisted Complementation. *Anal. Chem.* **2003**, 75 (7), 1584–1589. <https://doi.org/10.1021/ac020731c>.
- (19) Remy, I.; Michnick, S. W. A Highly Sensitive Protein-Protein Interaction Assay Based on Gaussia Luciferase. *Nat. Methods* **2006**, 3 (12), 977–979. <https://doi.org/10.1038/nmeth979>.
- (20) Liu, D.; Cronican, J. J.; Thompson, D. B. Engineering and Identifying Supercharged Proteins for Macromolecule Delivery into Mammalian Cells. *Methods Enzym.* **2012**, 503, 293–319. <https://doi.org/10.1038/jid.2014.371>.
- (21) Thompson, D. B.; Villaseñor, R.; Dorr, B. M.; Zerial, M.; Liu, D. R. Cellular Uptake Mechanisms and Endosomal Trafficking of Supercharged Proteins. *Chem. Biol.* **2012**, 19 (7), 831–843. <https://doi.org/10.1016/j.chembiol.2012.06.014>.

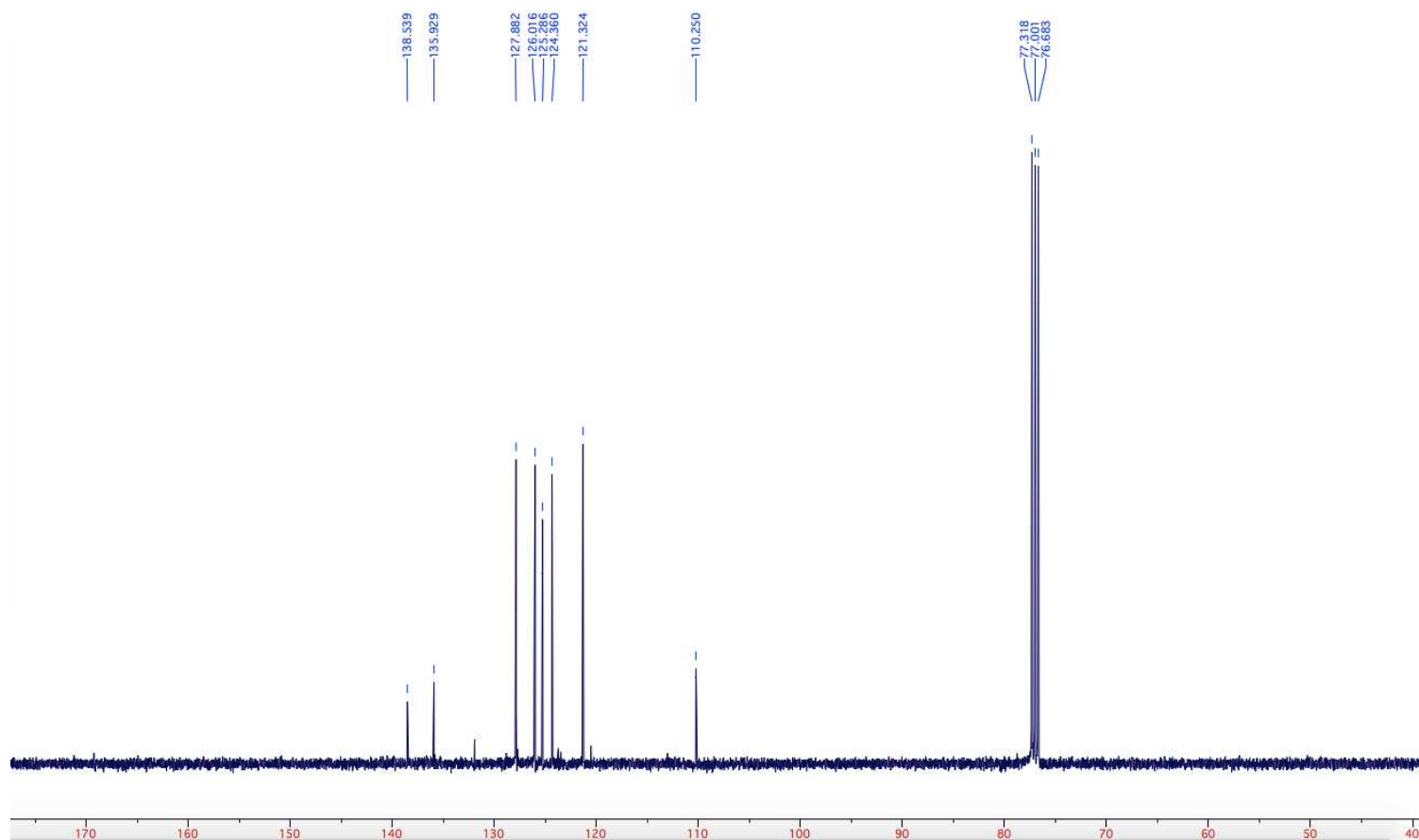
- (22) Bruce, V. J.; McNaughton, B. R. Inside Job: Methods for Delivering Proteins to the Interior of Mammalian Cells. *Cell Chem. Biol.* **2017**, *24* (8), 924–934. <https://doi.org/10.1016/j.chembiol.2017.06.014>.
- (23) Bruce, V. J.; Lopez-Islas, M.; McNaughton, B. R. Resurfaced Cell-Penetrating Nanobodies: A Potentially General Scaffold for Intracellularly Targeted Protein Discovery. *Protein Sci.* **2016**, *25* (6), 1129–1137. <https://doi.org/10.1002/pro.2926>.
- (24) Appelbaum, J. S.; Larochelle, J. R.; Smith, B. A.; Balkin, D. M.; Holub, J. M.; Schepartz, A. Arginine Topology Controls Escape of Minimally Cationic Proteins from Early Endosomes to the Cytoplasm. *Chem. Biol.* **2012**, *19* (7), 819–830. <https://doi.org/10.1016/j.chembiol.2012.05.022>.
- (25) Lönn, P.; Kacsinta, A. D.; Cui, X. S.; Hamil, A. S.; Kaulich, M.; Gogoi, K.; Dowdy, S. F. Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **2016**, *6* (August), 1–9. <https://doi.org/10.1038/srep32301>.
- (26) Steinauer, A.; LaRochelle, J. R.; Knox, S. L.; Wissner, R. F.; Berry, S.; Schepartz, A. *HOPS-Dependent Endosomal Fusion Required for Efficient Cytosolic Delivery of Therapeutic Peptides and Small Proteins.*; 2019; Vol. 116. <https://doi.org/10.1073/pnas.1812044116>.
- (27) Li, M.; Tao, Y.; Shu, Y.; LaRochelle, J. R.; Steinauer, A.; Thompson, D.; Schepartz, A.; Chen, Z. Y.; Liu, D. R. Discovery and Characterization of a Peptide That Enhances Endosomal Escape of Delivered Proteins in Vitro and in Vivo. *J. Am. Chem. Soc.* **2015**, *137* (44), 14084–14093. <https://doi.org/10.1021/jacs.5b05694>.

APPENDIX A: SUPPLEMENTAL DATA FOR CHAPTER TWO

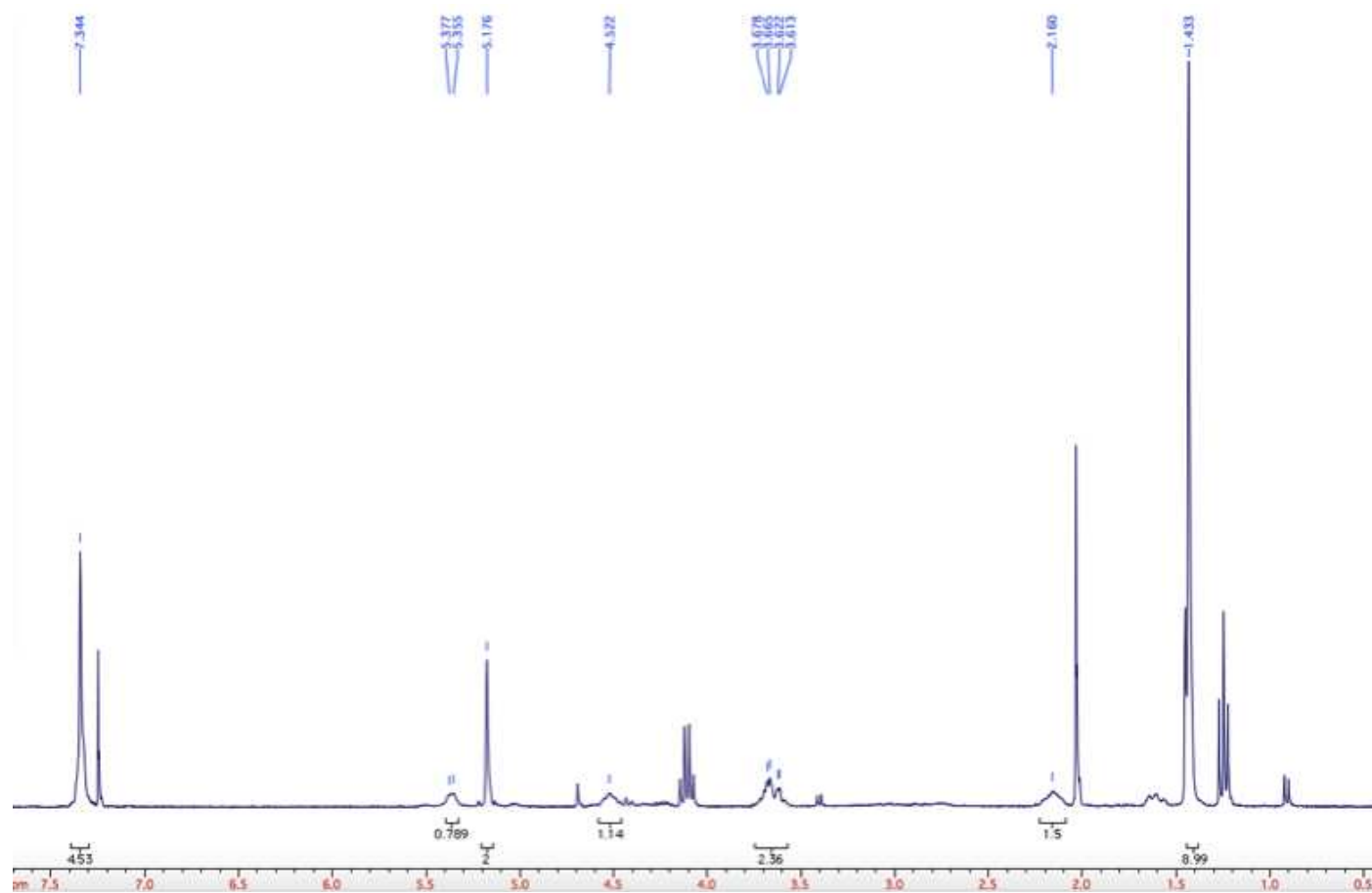
NMR's for Transition Metal Cross Coupling



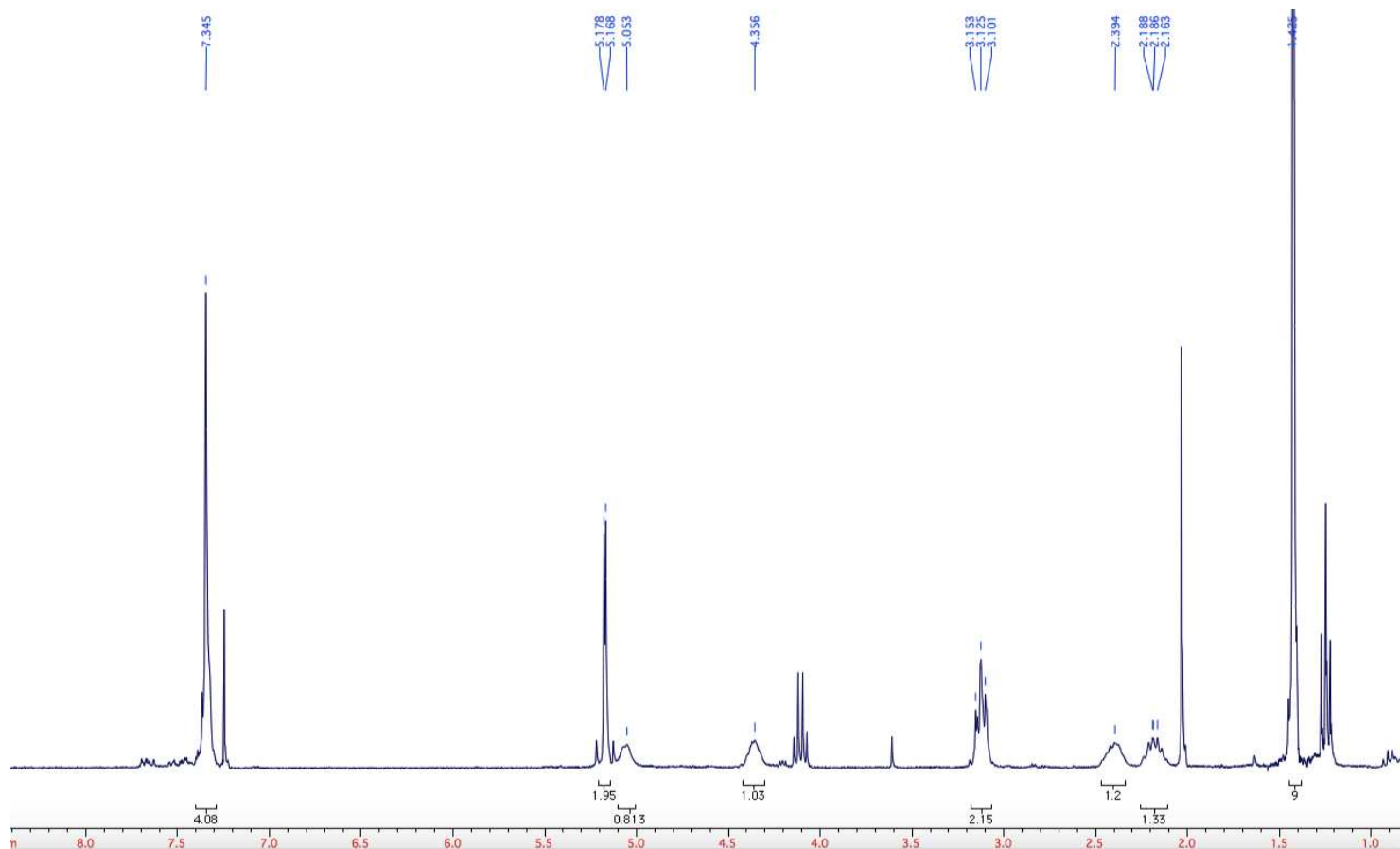
Supplemental Figure A.1: ^1H NMR of 4-bromo-2,2'-bithiophene (7)



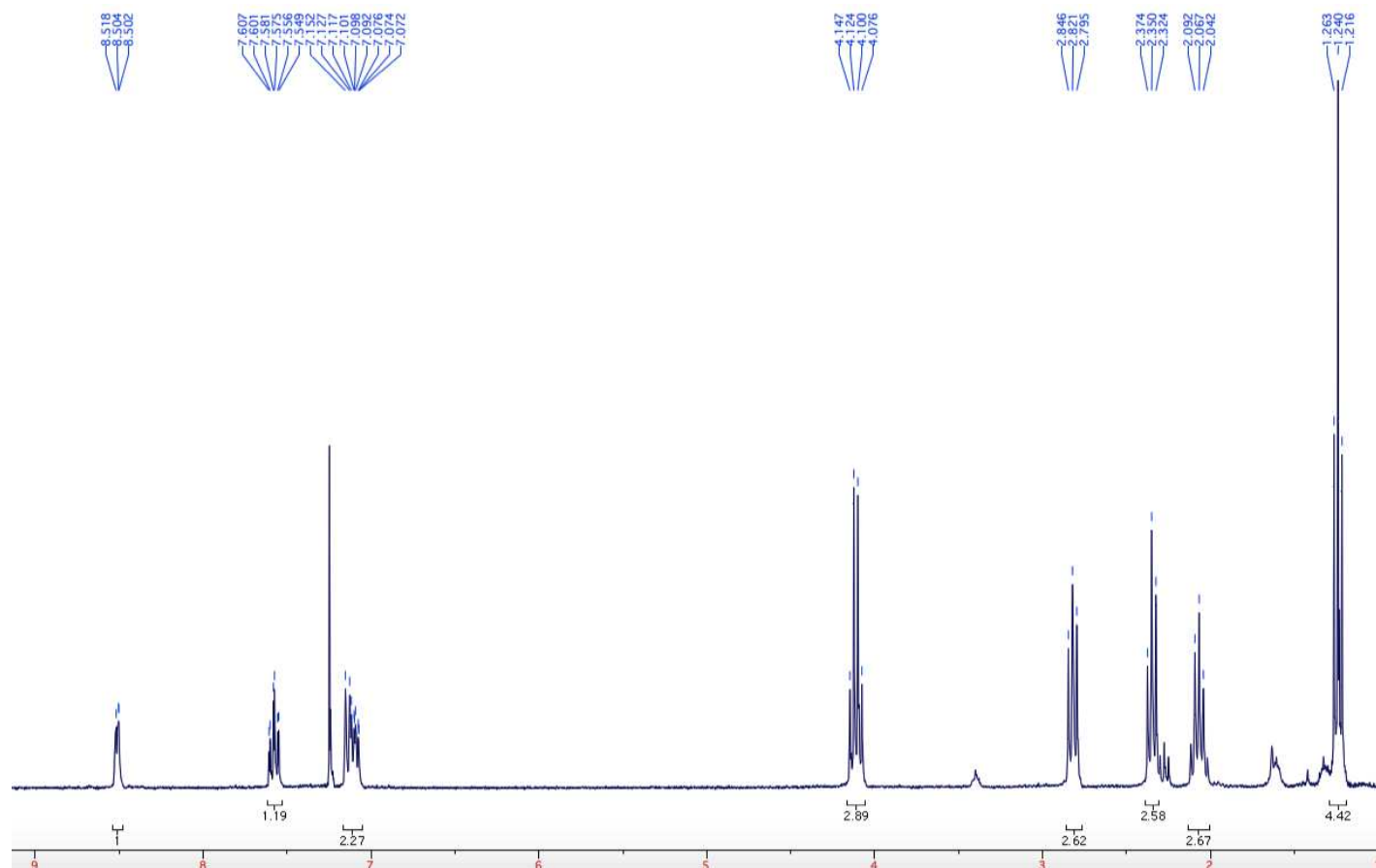
Supplemental Figure A.2: C^{13}NMR for 4-bromo-2,2'-bithiophene (7)



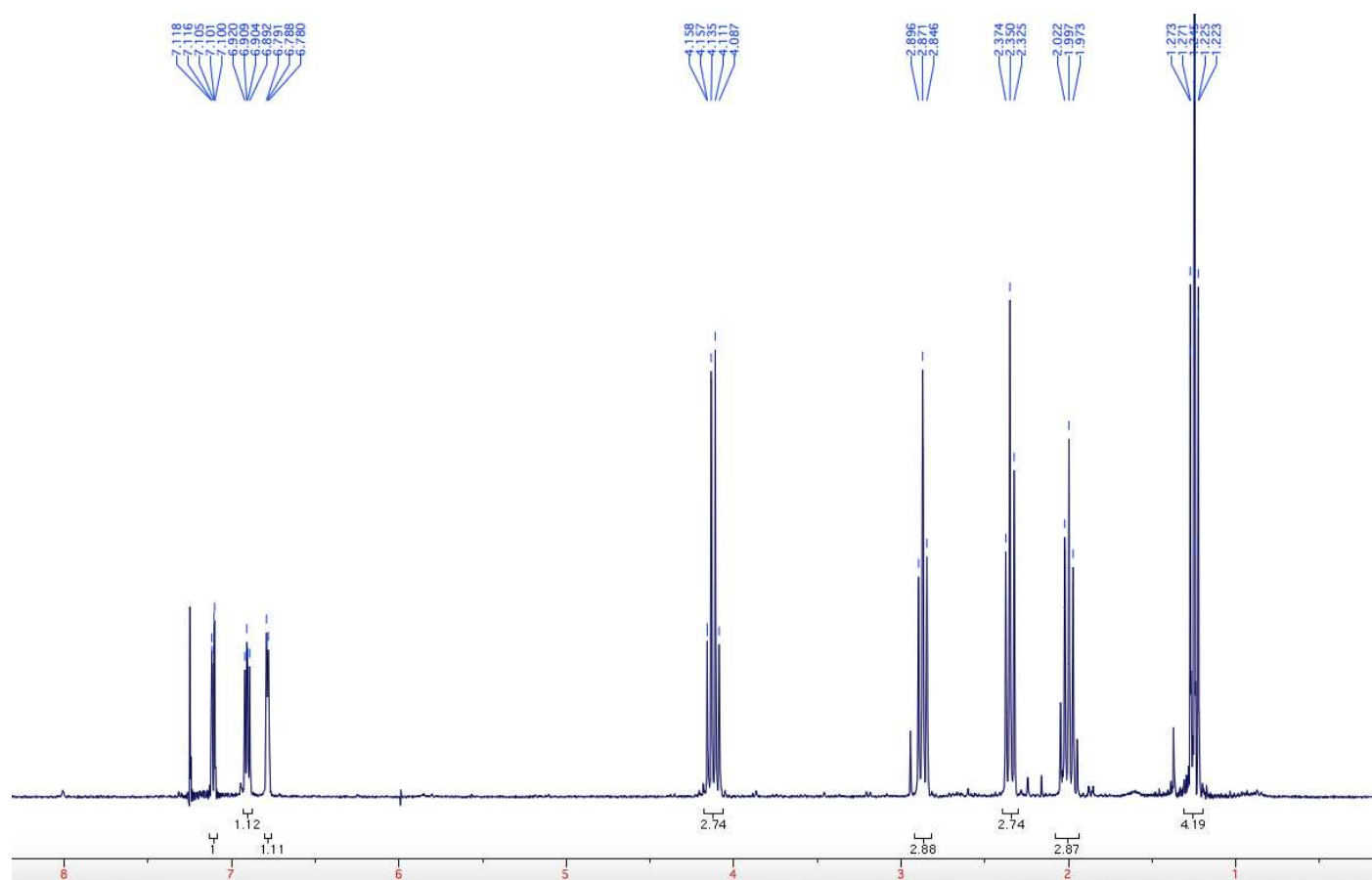
Supplemental Figure A.3: ^1H NMR of N-Boc-Homoserine-Benzyl Ester (**9**)



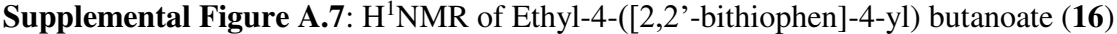
Supplemental Figure A.4: ¹H NMR of N-Boc-Iodohomoserine-Benzyl Ester (10)



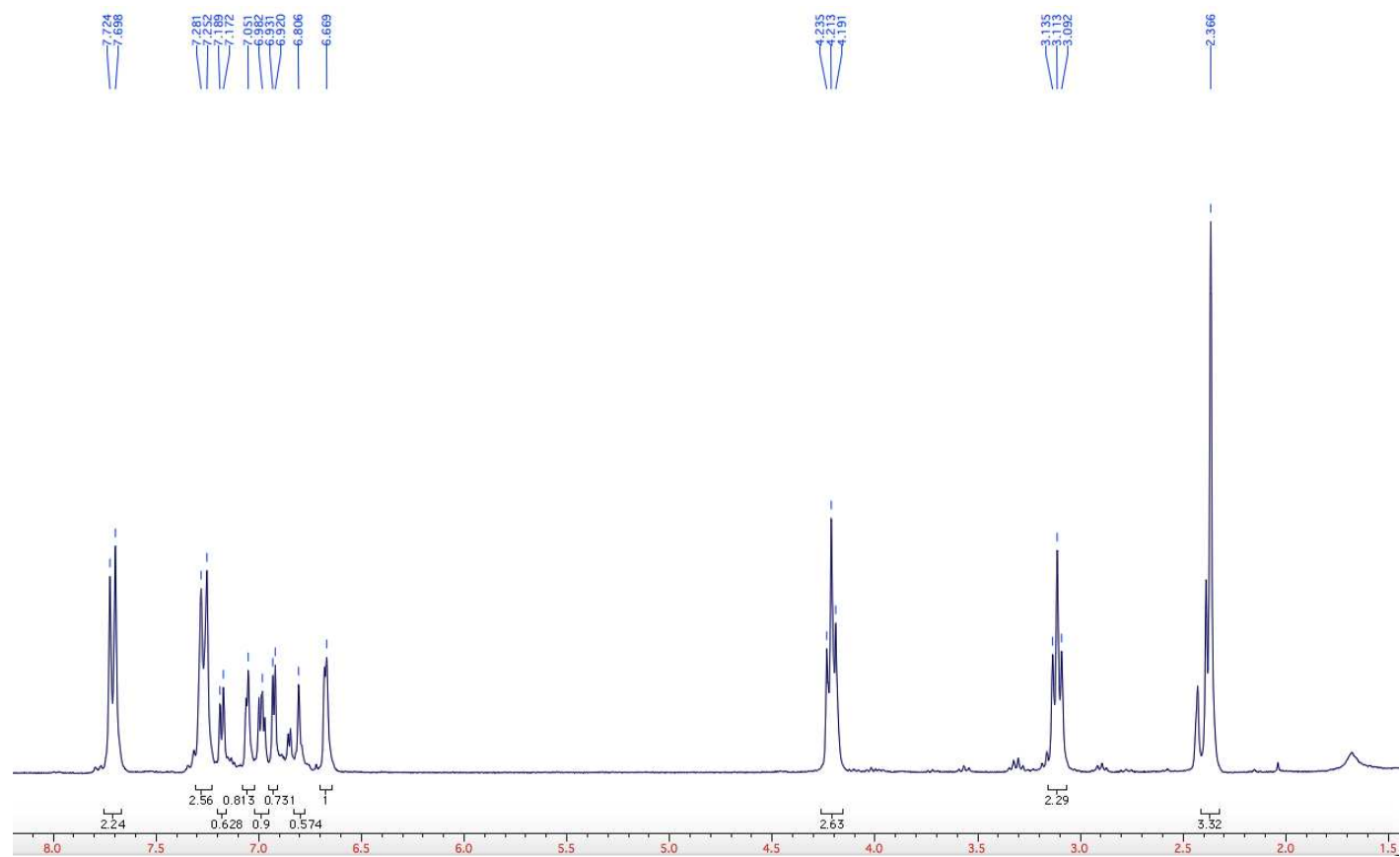
Supplemental Figure A.5: ¹H NMR of Ethyl-4-pyridin-2-yl) butanoate (14**)**



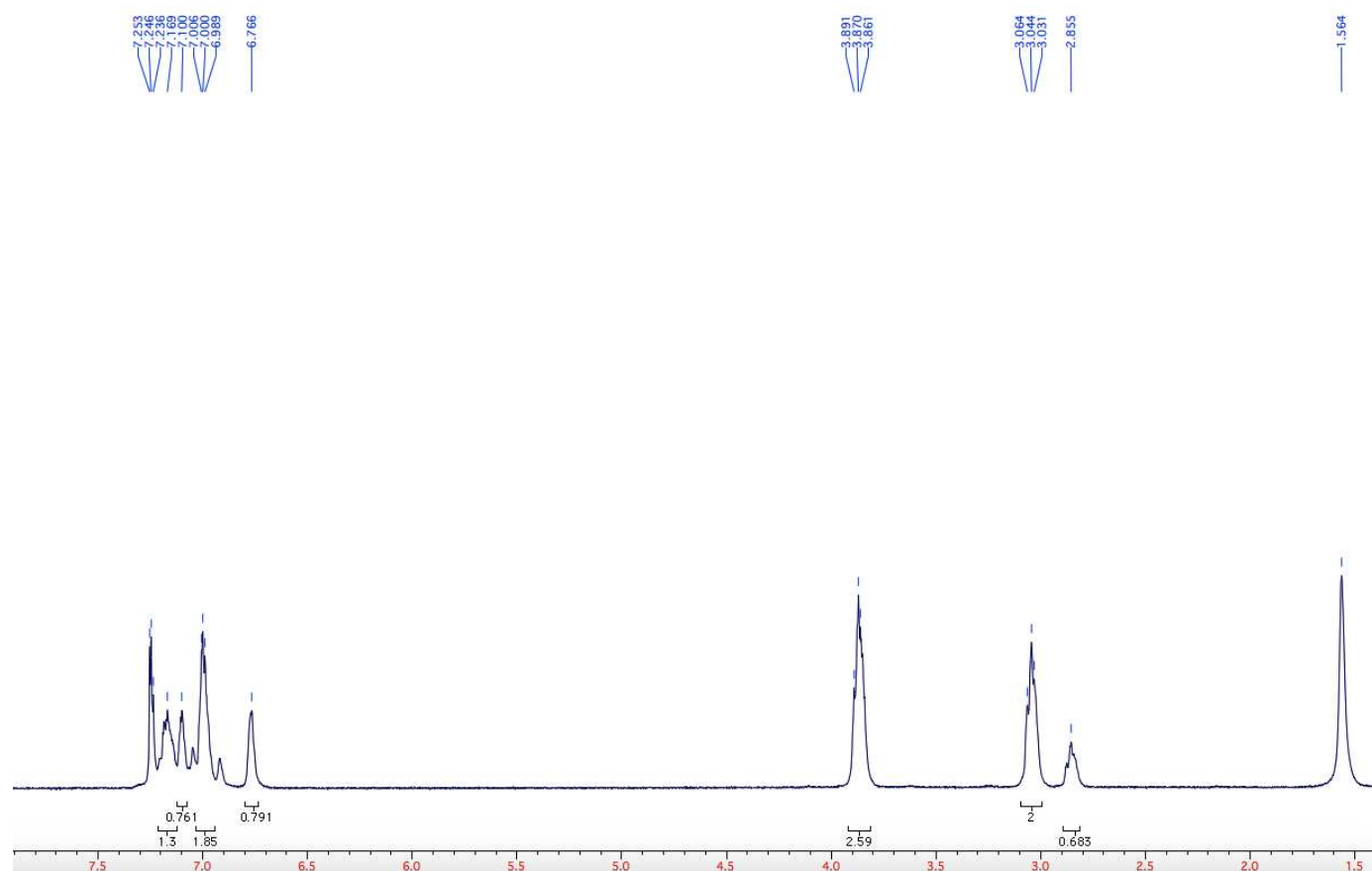
Supplemental Figure A.6: ¹H NMR of Ethyl-4-(thiophen-2-yl) butanoate (**15**)



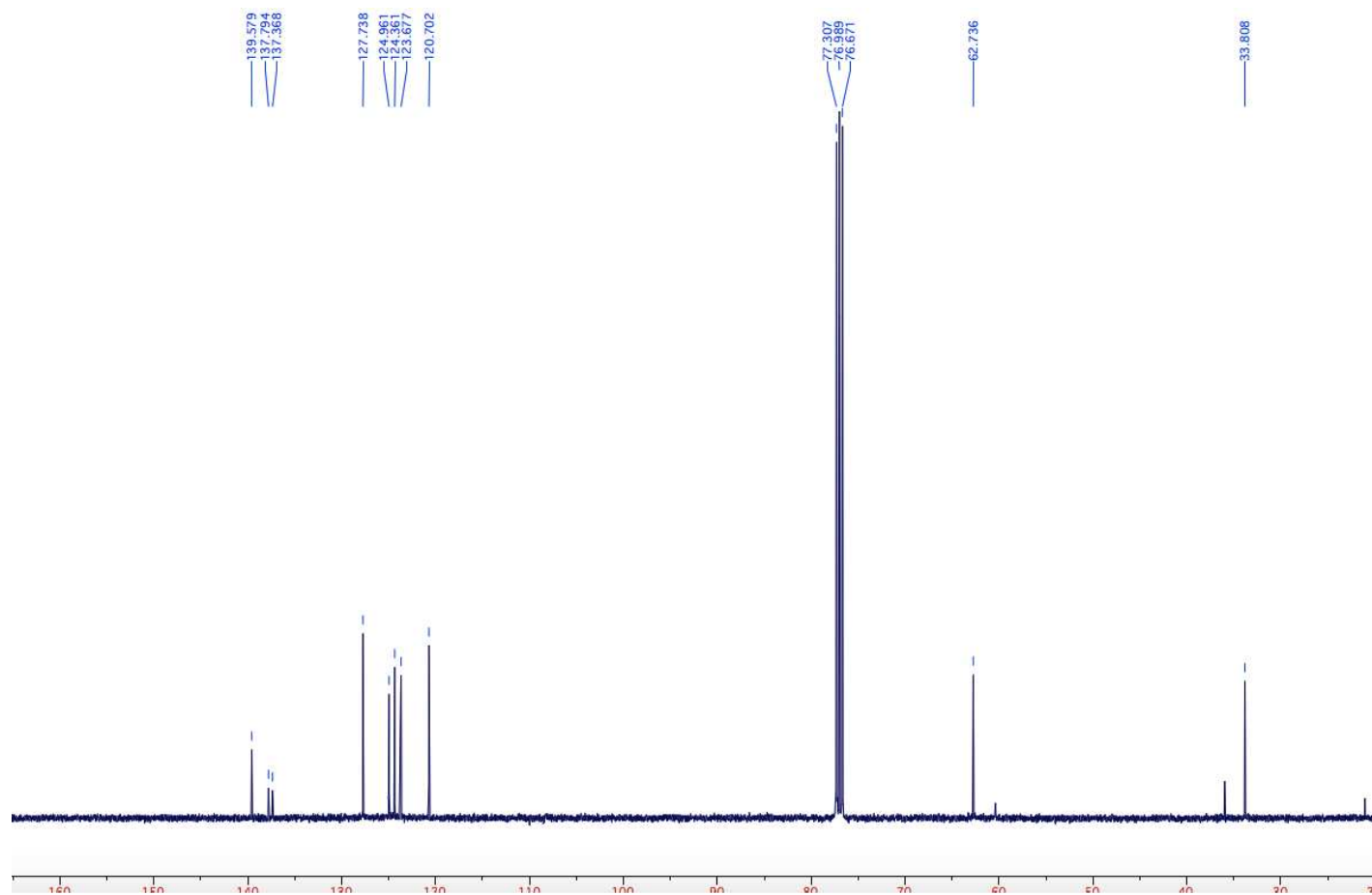
NMR's for Phase Transfer Catalysis



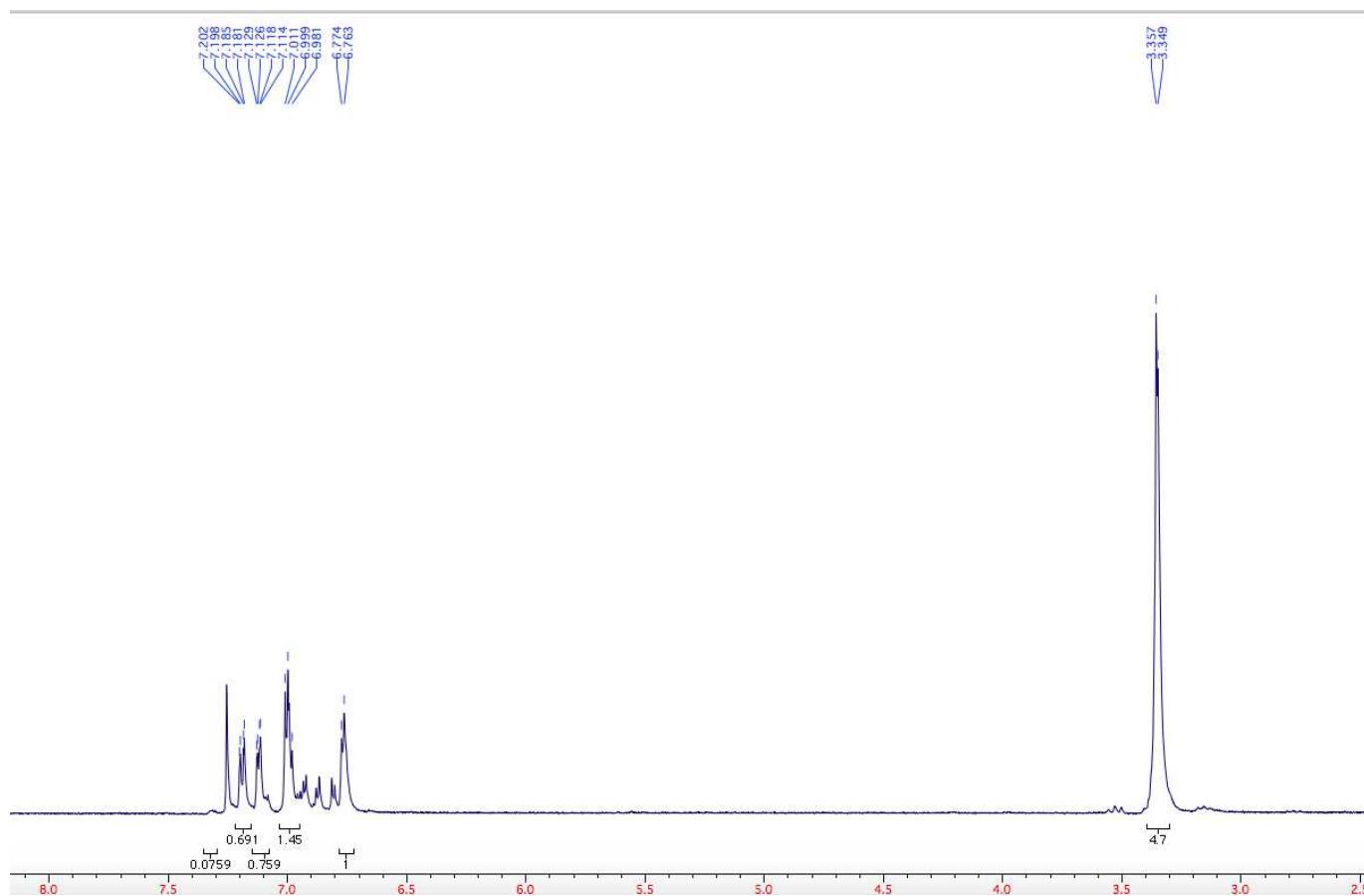
Supplemental Figure A.8: ^1H NMR for 2-([2,2'-bithiophen]-4-yl)ethyl-4-methylbenzenesulfonate (17,18)



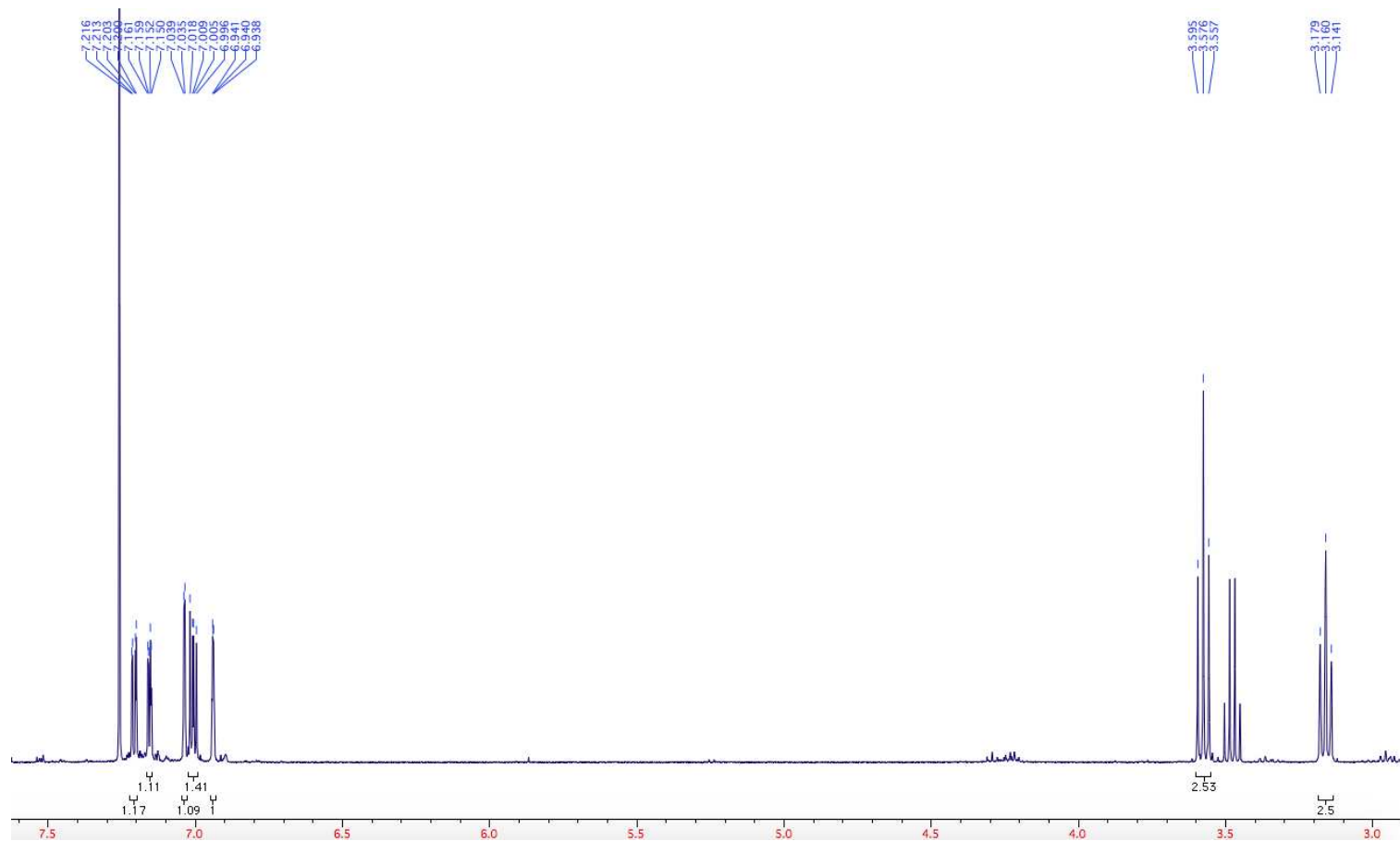
Supplemental Figure A.9: ¹H NMR of 2-([2,2'-bithiophen]-4-yl)ethan-1-ol (**19**)



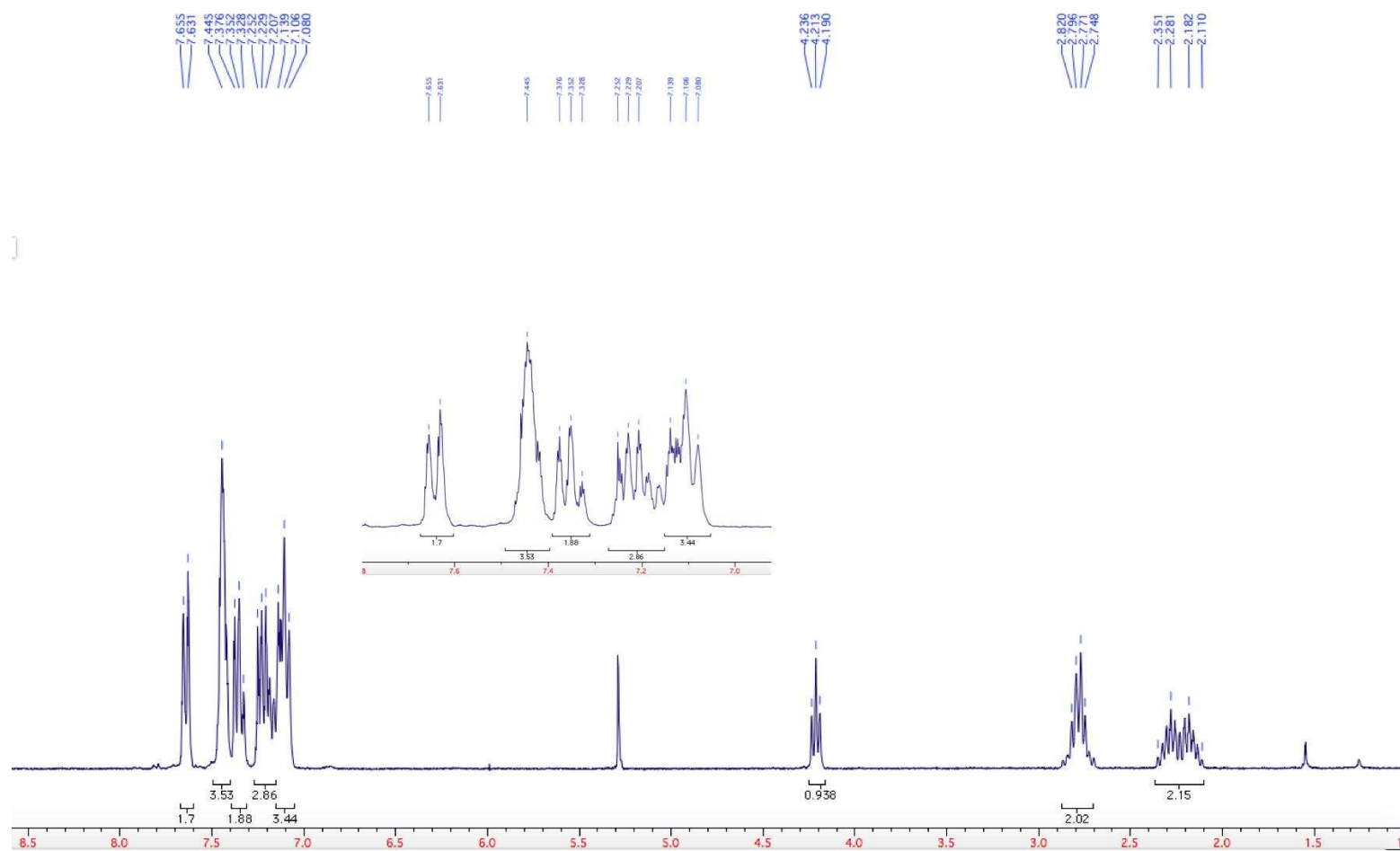
Supplemental Figure A.10: C^{13}NMR of 2-([2,2'-bithiophen]-4-yl)ethan-1-ol (**19**)



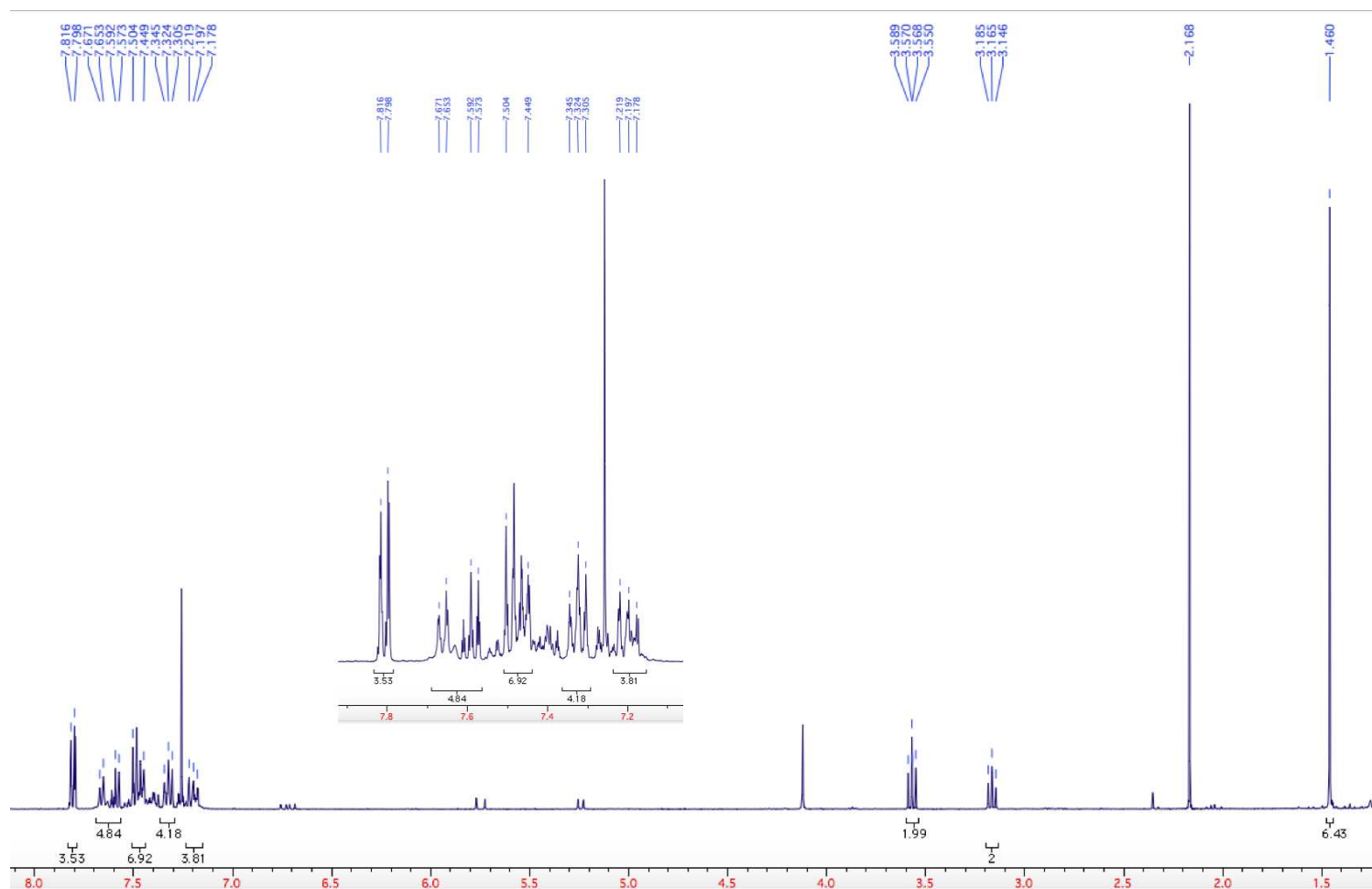
Supplemental Figure A.11: ^1H NMR of 4-(2-iodoethyl)-2,2'-bithiophene (20,21)



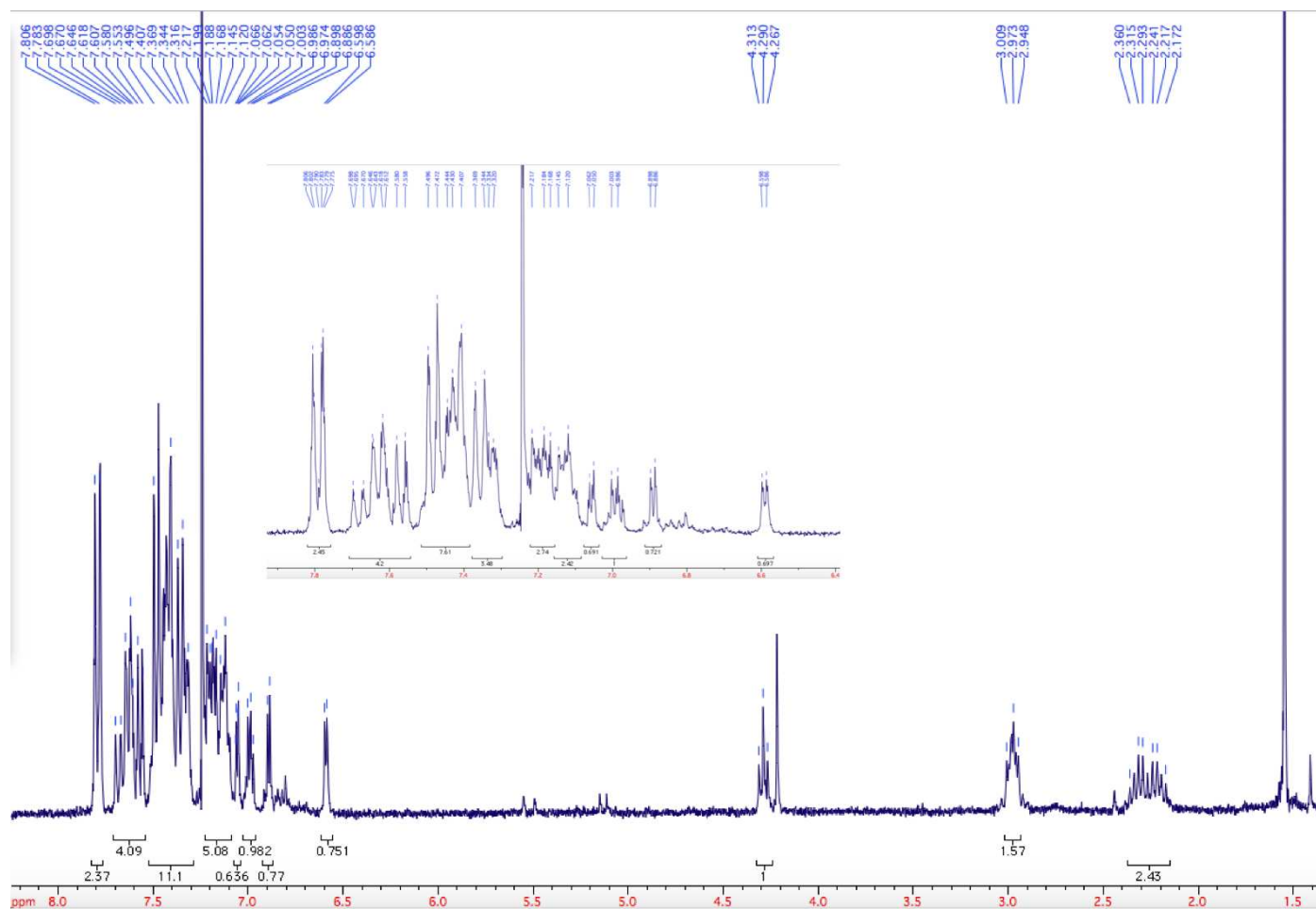
Supplemental Figure A.12: ¹H NMR of 4-(2-bromoethyl)-2,2'-bithiophene (**22**)



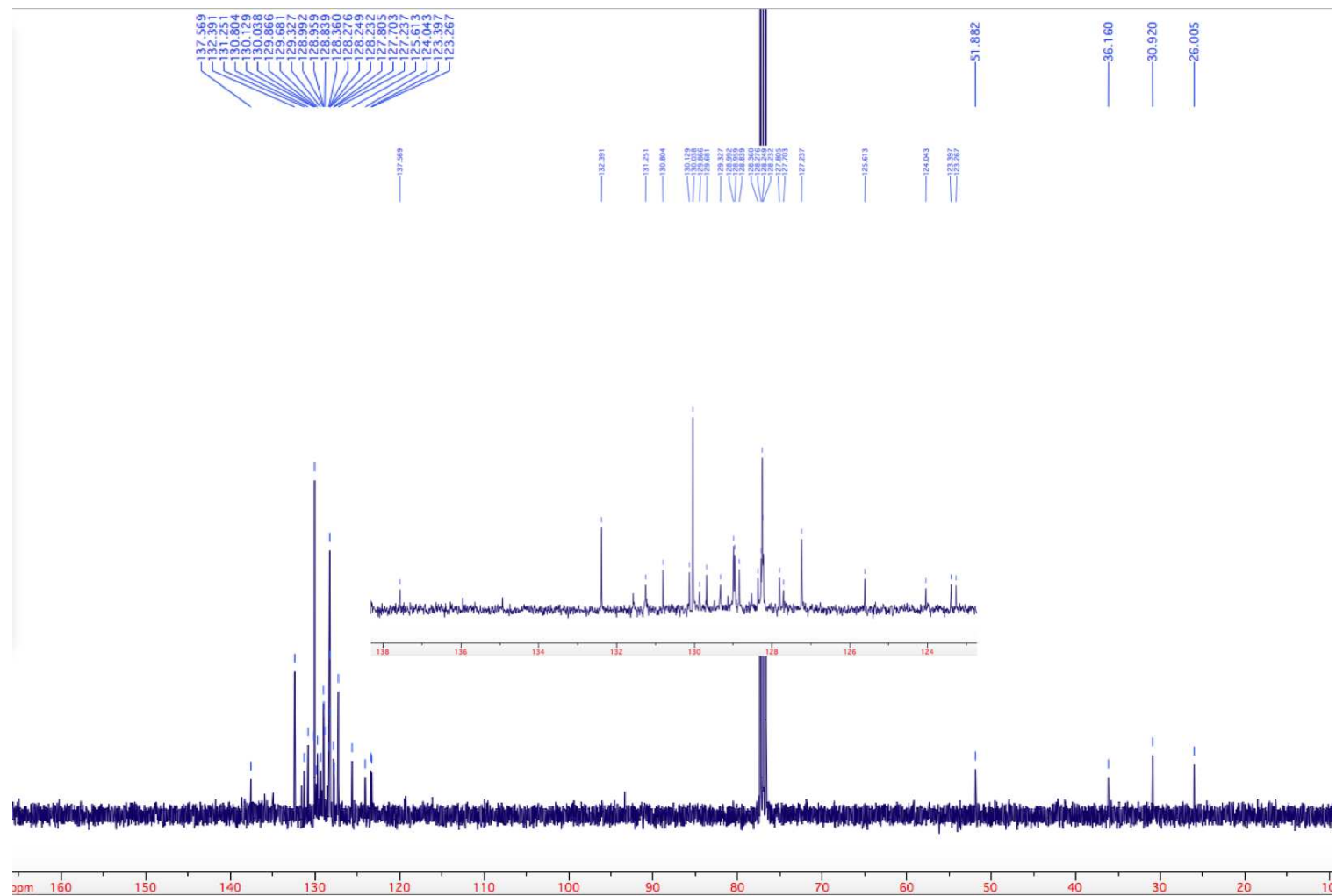
Supplemental Figure A.13: ^1H NMR of 2-([diphenylmethylene]amino)-4-phenylbutanenitrile (**23**)



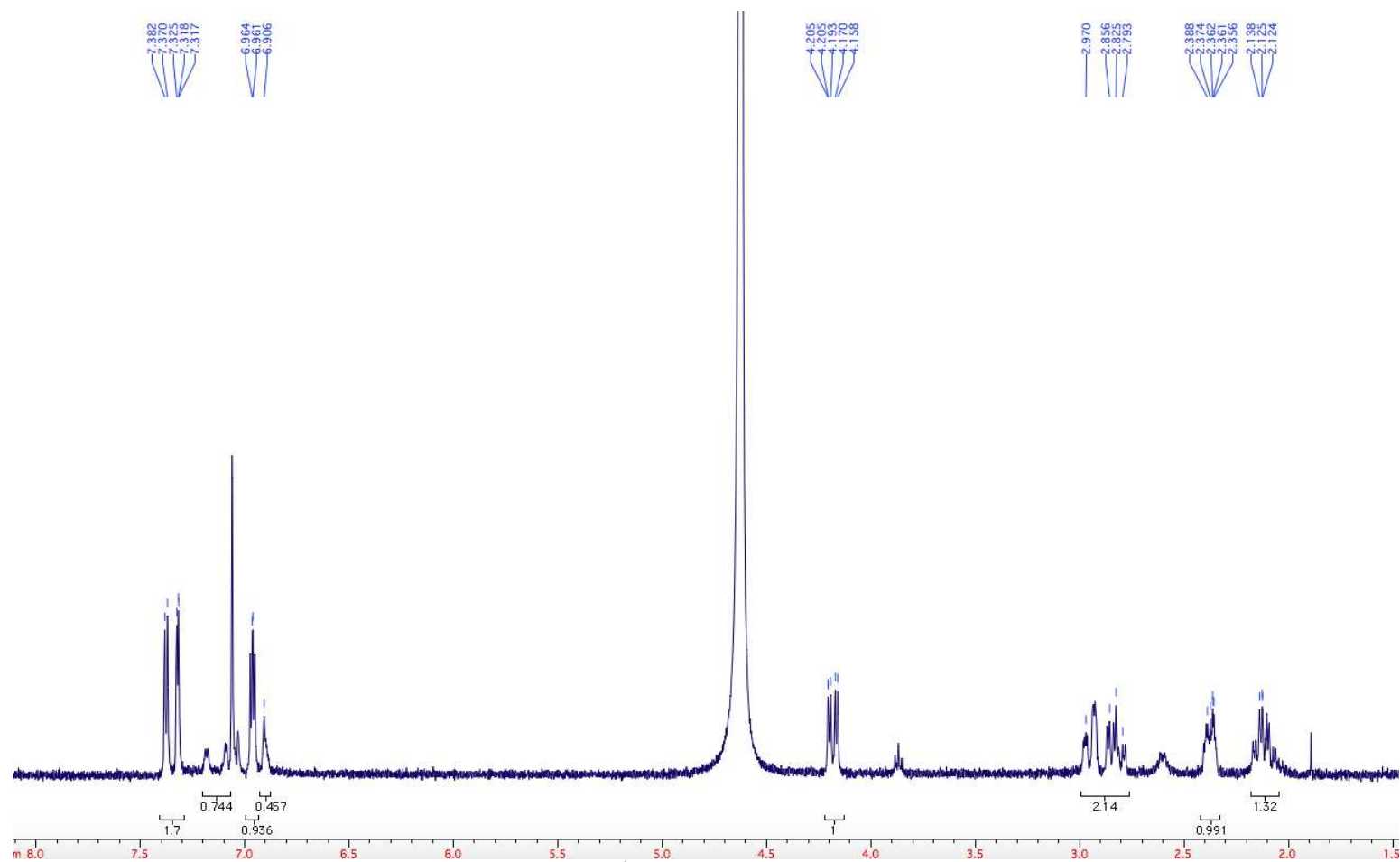
Supplemental Figure A.14: ¹H NMR of Tert-butyl 2-([diphenylmethylene]amino)-4-phenylbutanoate (**24**)



Supplemental Figure A.15: ^1H NMR of 4-([2,2'-bithiophen]-4-yl)-2-((diphenylmethylene)amino)butanenitrile (**27**)



Supplemental Figure A.16: C^{13} NMR of 4-([2,2'-bithiophen]-4-yl)-2-((diphenylmethylene)amino)butanenitrile (**27**)



Supplemental Figure A.17: ^1H NMR of deprotected Bithiophene amino acid

APPENDIX B: SUPPLEMENTAL DATA FOR CHAPTER THREE

Protein Sequences:

GFPnb OFF:

MGQVQLVESGGALVQPGGSLRLSCAASGFPVNRYSMRWYRQAPGKEREWVAG
MSSAGDRSSYEDSVKGRFTISRDDARNTVYQLQMNSLKPEDTAVYYCNVAVGAA
YWGQGTQVTVSSKHHHHHHH

CatNB OFF:

MGMQVQLVEKGGKRVQPGGSLRLKCAASGFPVNRYSMRWYRQAPGKEREWV
AGMSSAGDRSSYEDSVKGRFKIKRDDARNTVYLMRKLPEDTAVYYCNVAV
GAAYWGQGTRTVTSKKHHHHHHH

AGV1 OFF:

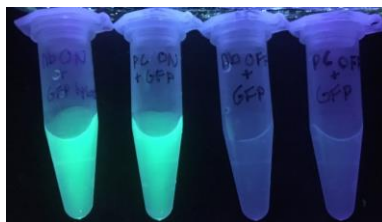
MQDPQVQLVEKGGRRVQPGGSLRLCAASGFPVNRYSMRWYRQAPGKEREWV
AGMSSAGDRSSYEDSVKGRFTISRDDARNTVYQLQMNSLKPEDTAVYYCNVAVG
AAYWGQGTRVTVSRKHHHHHHH

AGV2 OFF:

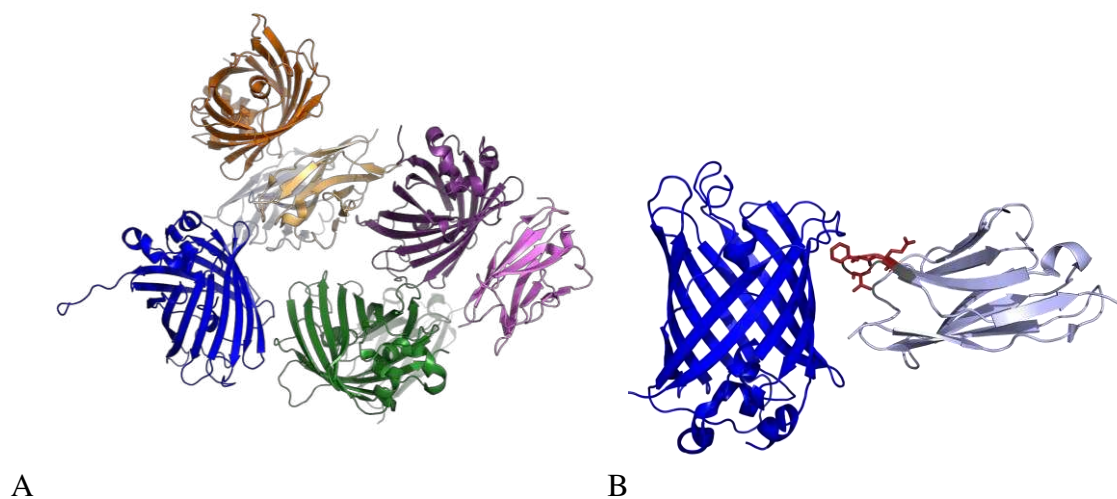
MQVQLVEKGGALVQPGGSLRLSCAASGFPVNRYSMRWYRQAPGKEREWVAG
MSSAGDRSSYEDSVKGRFRIRDDARNTVYLMRRLKPEDTAVYYCNVAVGAA
YWGQGTQVTVSSKHHHHHHH

PolyR OFF:

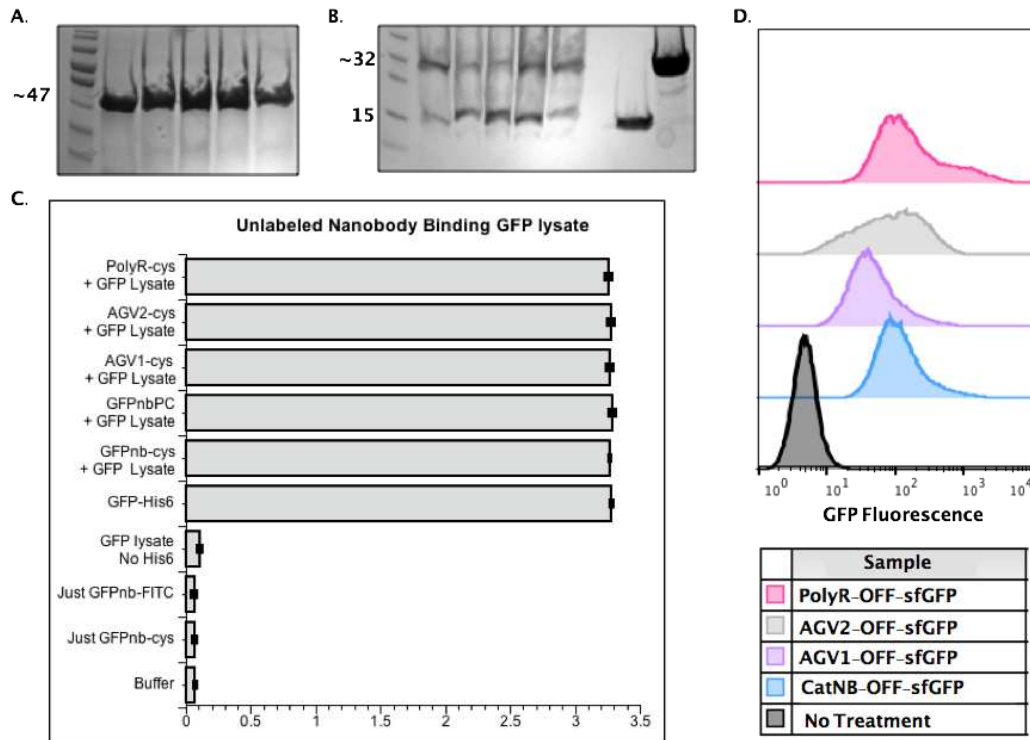
MGQVQLVEKGGRRVQPGGSLRLCAASGFPVNRYSMRWYRQAPGKEREWVA
GMSSAGDRSSYEDSVKGRFRIRDDARNTVYLMRRLKPEDTAVYYCNVAVGA
AYWGQGTRVTVSRKHHHHHHH



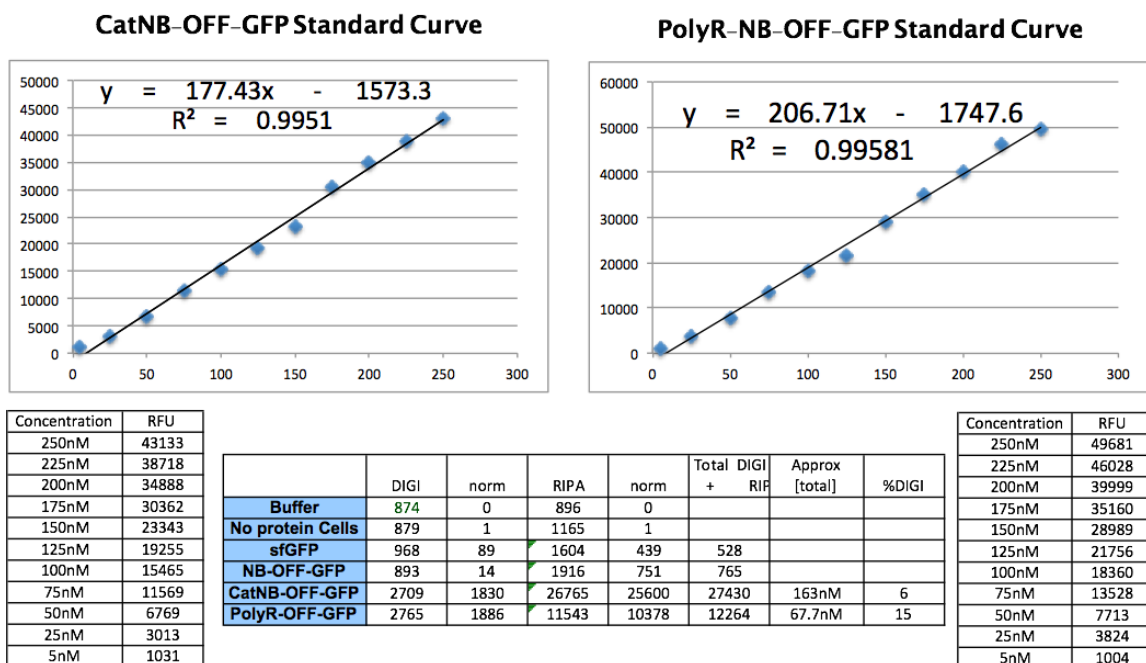
Supplemental Figure B.1: Co-purification of GFP binding nanobody and resurfaced GFP binding nanobody compared to OFF mutants. CDR3 mutations AVGAA were made in order to turn off GFP binding. 0.3mg of protein were incubated with 100uL of sfGFP lysate (no His6 tag) and incubated for 15 minutes at 4C. 100uL of Ni-NTA beads were added to each tube and incubated for 1 hour at 4C. Ni for each protein was then washed with 20mL 2mM imidazole phosphate buffer, 25mL 50mM imidazole phosphate buffer and 20mL 75mM imidazole phosphate buffer before being eluted with 500uL of 400mM Imidazole and imaged under UV light. Tubes left to right: GFPnb ON, CatNB ON, GFPnb OFF, CatNB OFF



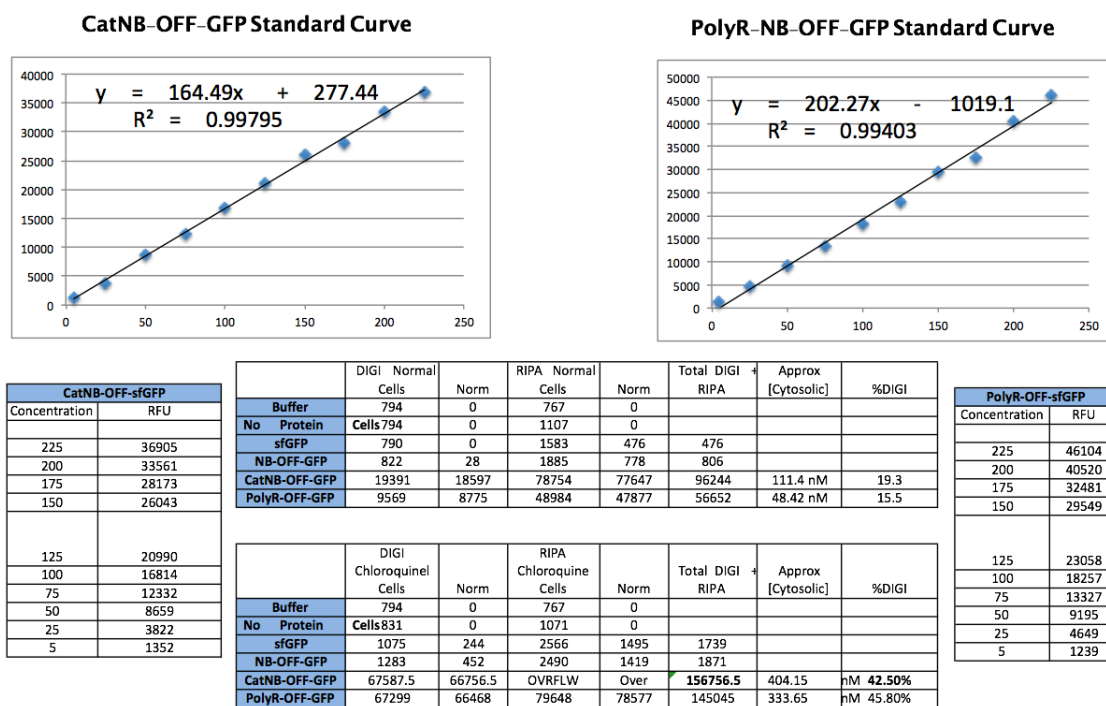
Supplemental Figure B.2: Crystal structures showing native binding of GFP nanobody and GFP. PDB: 3OGO **A.** Suggested formation multimeric complex of GFP nanobody fused to GFP without OFF mutations. Dark colors represent GFP and corresponding binding nanobody partners in lighter colors. **B.** A single GFP molecule binding a single GFP nanobody as a suggested representation of the OFF mutants with the CDR3 mutations important for binding highlighted in red.



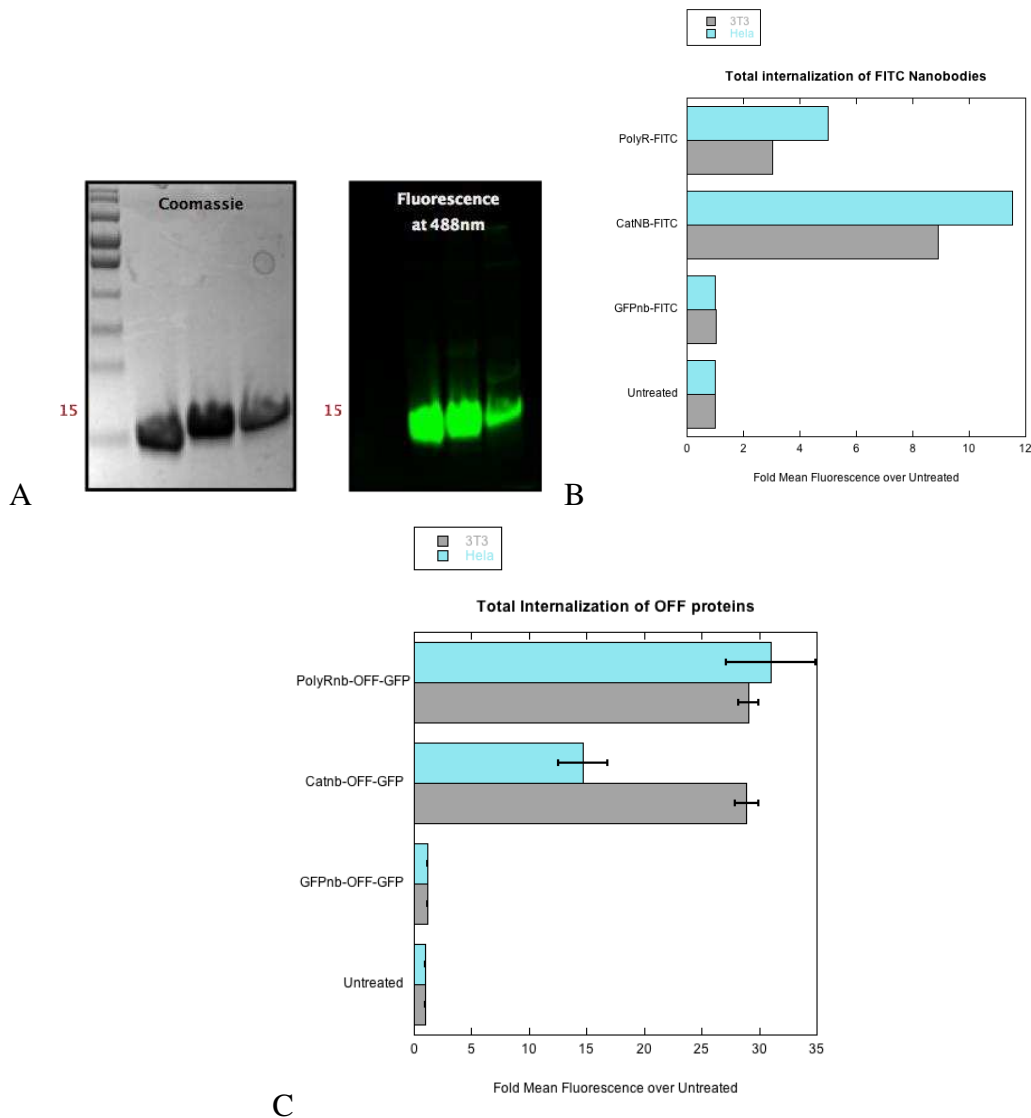
Supplemental Figure B.3) E coli expression of OFF mutants and Ni-Pull down in E coli of ON variants. A) E coli expression and purification of nanobody-OFF-sfGFP fusions. B) Co-purification using Ni-pull down assay of nanobody mutants with a his6 tag and sfGFP without a his6 tag. C) ELISA using Ni-coated plate of nanobody ON resurfaced mutants and sfGFP lysate with no His6. D) Representative histograms of nanobody total cell internalization



Supplemental Figure B.4: Fluorescent standard curves of cytosolic and total cell fractions for both CatNB-OFF and PolyR-OFF proteins. Proteins were incubated with NIH 3T3 cells for 1 hour, washed with PBS containing heparin sulfate, trypsinized and pelleted. The Pellet was resuspended in 80uL Digitonin buffer and incubated on ice for 10 min and pelleted. The supernatant was collected for the digitonin fractions and 100uL of each sample was loaded into a black 96-well plate to be read for fluorescence. The remaining pellet was then resuspended in 90ul RIPA buffer and incubated for 30min on ice and pelleted. 100uL of this supernatant was loaded on to the same black 96-well plate. The fractions were analyzed for their relative fluorescence, normalized to cell lysate that was not incubated with any protein, and compared to a standard curve built from their respective purified protein to estimate an intracellular concentration of delivered protein. These fractions were then used in a western blot analysis to show efficient separation of fractions (shown in paper.)



Supplemental Figure B.5: Fluorescent standard curves of cytosolic and total cell fractions for both CatNB-OFF and PolyR-OFF proteins. Proteins were incubated with NIH 3T3 cells for 1 hour at 1uM with and without 100uM Chloroquine. Each set of cells were washed with PBS containing heparin sulfate, trypsinized and pelleted. The pellet was resuspended in 80uL Digitonin buffer and incubated on ice for 10 min and pelleted. The supernatant was collected for the digitonin fractions and 100uL of each sample was loaded into a black 96-well plate to be read for fluorescence. The remaining pellet was then resuspended in 90ul RIPA buffer and incubated for 30min on ice and pelleted. 100uL of this supernatant was loaded on to the same black 96-well plate. The fractions were analyzed for their relative fluorescence, normalized to cell lysate that was not incubated with any protein, and compared to a standard curve built from their respective purified protein to estimate an intracellular concentration of delivered protein. It is obvious that both nanobodies exist in the cytosol at relatively high concentrations in the cytosol as seen in the normal cells, however; a lot may still be trapped in the endosomes since the concentrations of both proteins in the cytosol dramatically increased upon addition of Chloroquine.



Supplemental Figure B.6: a) Coomassie and fluorescence gel of FITC labeled nanobodies. Left to right: GFPnb-FITC, CatNB-FITC, PolyNB-FITC. b) Comparison of FITC nanobodies in HeLa and 3T3 cells. Cells were treated with 1uM of each protein and incubated for 1H at 37C and 5% CO₂. Cells were washed three times with PBS containing 20 U Heparin sulfate and analyzed using flow cytometry. c) Comparison of OFF proteins in HeLa and 3T3 cells. Cells were treated with 1uM of each protein and incubated for 1H at 37C and 5% CO₂. Cells were washed three times with PBS containing 20 U Heparin sulfate and analyzed using flow cytometry.

Residue	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
CatNB ₁ (GFP)	M	Q	V	Q	L	V	E	K	G	G	K	R	V	Q	P	G	G	S	L	R	L	K	C	A	A
CatNB ₁ (BC2)	M	Q	V	Q	L	V	E	K	G	G	K	R	V	Q	P	G	G	S	L	R	L	K	C	T	A
CatNB(GFP)-BC2 ₁ CDRsHis6	M	Q	V	Q	L	V	E	K	G	G	K	R	V	Q	P	G	G	S	L	R	L	K	C	A	A
CatNB(GFP)-BC2 ₁ CDRsHis6GFP	M	Q	V	Q	L	V	E	R	G	G	R	R	V	Q	P	G	G	S	L	R	L	R	C	A	A

Residue	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
CatNB ₁ (GFP)	S	-	-	G	F	P	V	N	R	Y	S	-	M	R	W	Y	R	Q	A	P	G	K	E	R	E
CatNB ₁ (BC2)	S	-	G	F	T	L	D	H	Y	D	I	G	-	-	W	F	R	Q	A	P	G	K	E	R	E
CatNB(GFP)-BC2 ₁ CDRsHis6	S	-	G	F	T	L	D	H	Y	D	I	G	M	R	W	Y	R	Q	A	P	G	K	E	R	E
CatNB(GFP)-BC2 ₁ CDRsHis6GFP	S	-	G	F	T	L	D	H	Y	D	I	G	M	R	W	Y	R	Q	A	P	G	K	E	R	E

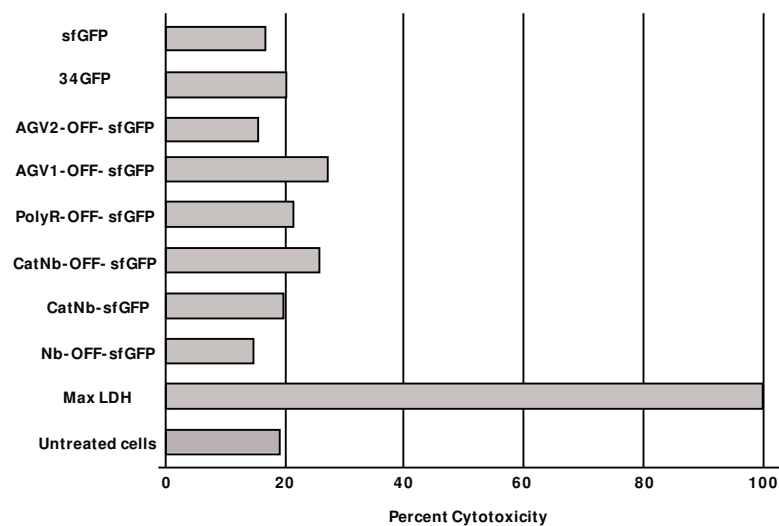
Residue	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
CatNB ₁ (GFP)	W	V	A	G	M	S	S	A	G	D	R	S	S	Y	E	D	S	V	K	G	R	F	K	I	K
CatNB ₁ (BC2)	G	V	S	C	I	N	N	S	D	D	D	T	Y	Y	A	D	S	V	K	G	R	F	K	I	K
CatNB(GFP)-BC2 ₁ CDRsHis6	W	V	A	G	I	N	N	S	D	D	D	T	Y	Y	E	D	S	V	K	G	R	F	K	I	K
CatNB(GFP)-BC2 ₁ CDRsHis6GFP	W	V	A	G	I	N	N	S	D	D	D	T	Y	Y	E	D	S	V	K	G	R	F	R	I	R

Residue	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
CatNB ₁ (GFP)	R	D	D	A	R	N	T	V	Y	L	R	M	R	K	L	K	P	E	D	T	A	V	Y	Y	C
CatNB ₁ (BC2)	R	N	N	A	K	D	T	V	Y	L	R	M	R	K	L	K	P	E	D	T	A	I	Y	Y	C
CatNB(GFP)-BC2 ₁ CDRsHis6	R	D	D	A	R	N	T	V	Y	L	R	M	R	K	L	K	P	E	D	T	A	V	Y	Y	C
CatNB(GFP)-BC2 ₁ CDRsHis6GFP	R	D	D	A	R	N	T	V	Y	L	R	M	R	R	L	K	P	E	D	T	A	V	Y	Y	C

Residue	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125
CatNB ₁ (GFP)	N	V	N	V	G	F	E	-	-	-	-	-	-	-	-	-	-	Y	W	G	Q	G	T	R	V
CatNB ₁ (BC2)	A	E	A	R	G	C	K	R	G	R	Y	E	Y	D	-	-	-	F	W	G	Q	G	T	R	V
CatNB(GFP)-BC2 ₁ CDRsHis6	A	E	A	R	G	C	K	R	G	R	Y	E	Y	D	-	-	-	Y	W	G	Q	G	T	R	V
CatNB(GFP)-BC2 ₁ CDRsHis6GFP	A	E	A	R	G	C	K	R	G	R	Y	E	Y	D	-	-	-	Y	W	G	Q	G	T	R	V

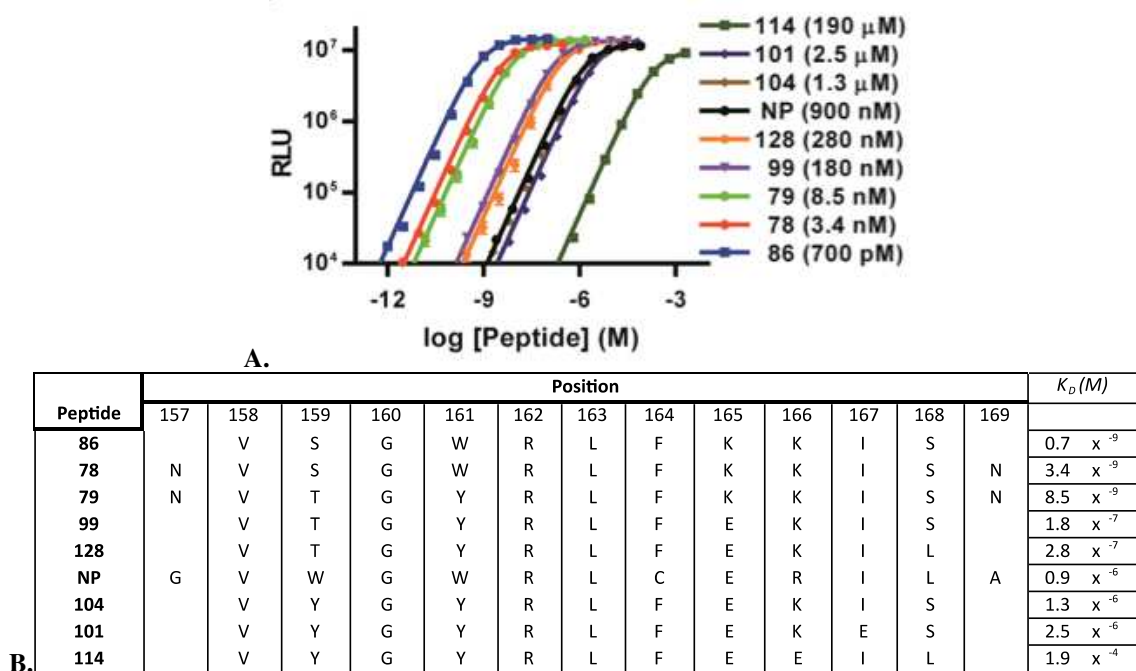
Residue	126	127	128	129	130	131	132	133	134	135	136	137	138	139
CatNB ₁ (GFP)	V	T	V	S	K	K	H	H	H	H	H	H	H	
CatNB ₁ (BC2)	V	T	V	S	S	K	K	K	H	H	H	H	H	H
CatNB(GFP)-BC2 ₁ CDRsHis6	V	T	V	S	K	K	H	H	H	H	H	H	H	
CatNB(GFP)-BC2 ₁ CDRsHis6GFP	V	T	V	S	R	K	H	H	H	H	H	H	H	

Supplemental Figure B.7: Sequences of wild type GFP binding nanobody, CatNB_{GFP}, wild type BC2 nanobody, CatNB_{BC2}, and CatNB_{GFP} with BC2 CDRs without any GFP fusions. For CatNB with BC2 CDRs fused to sfGFP, an all arginine version was used. Purple represents CDRs and blue represents the resurfaced residues.



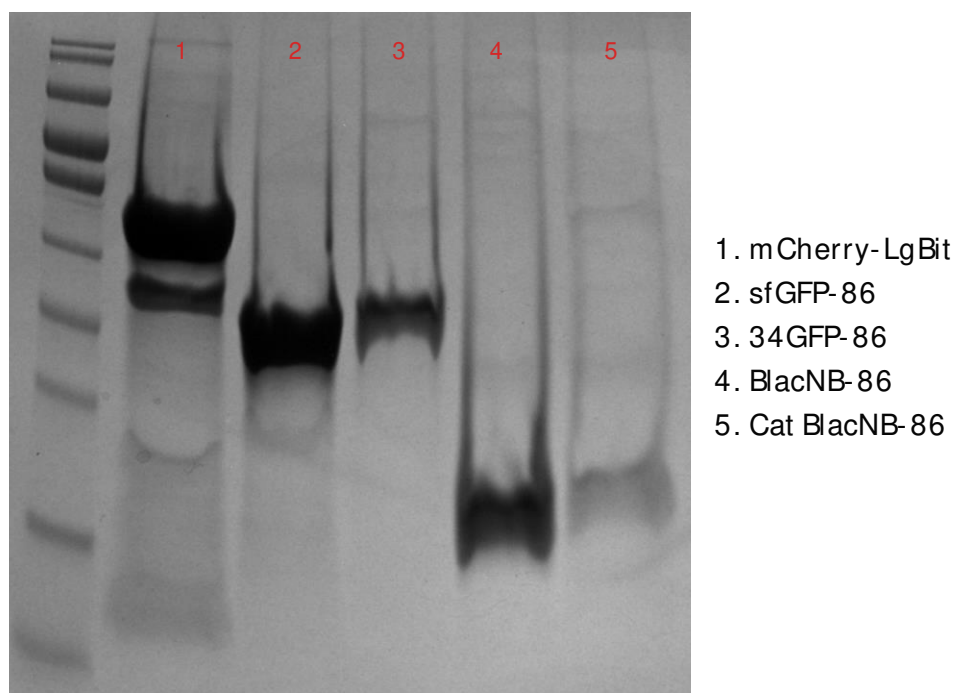
Supplemental Figure B.8: MTT assay of OFF proteins. Proteins were incubated with Hela cells for 1 hour prior to following the protocol for CytoTox 96 ® NonRadioactive Cytotoxicity Assay from Promega.

APPENDIX C: SUPPLEMENTAL DATA FOR CHAPTER FOUR

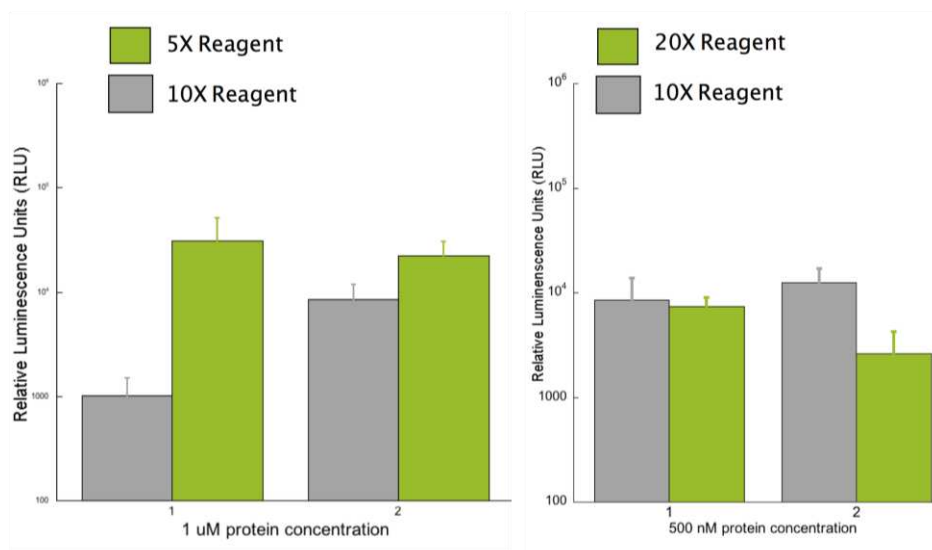


Supplemental Figure C.1: Sequences and binding affinities of the engineered by Promega for SmBit peptides of NanoBit¹. The 86 peptide was used for our studies while the 114 peptide is used by Promega for observing protein-protein interactions.

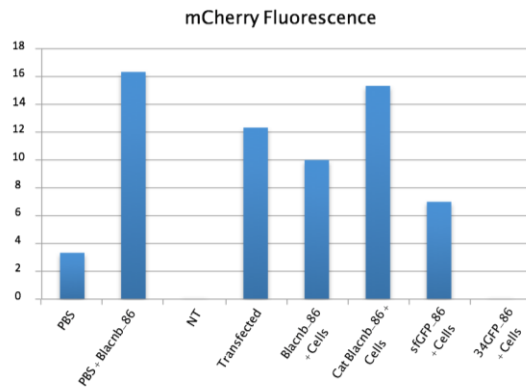
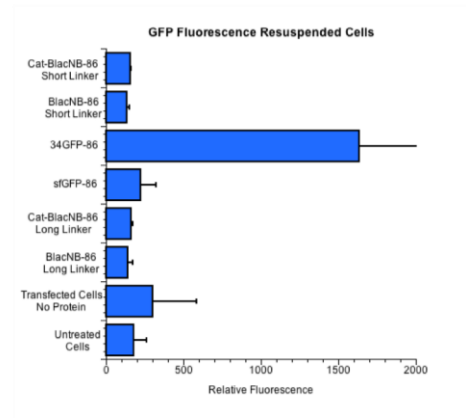
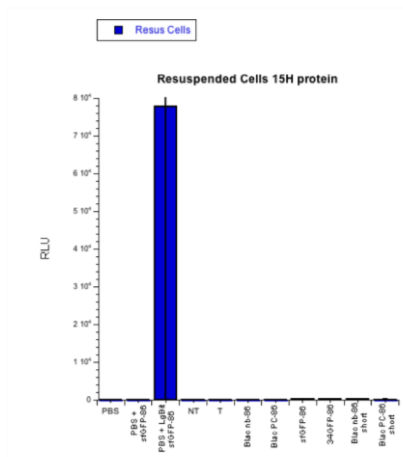
- (1) Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P.; Lubben, T. H.; Butler, B. L.; Binkowski, B. F.; MacHleidt, T.; Kirkland, T. A.; et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **2016**, *11* (2), 400–408. <https://doi.org/10.1021/acscchembio.5b00753>.



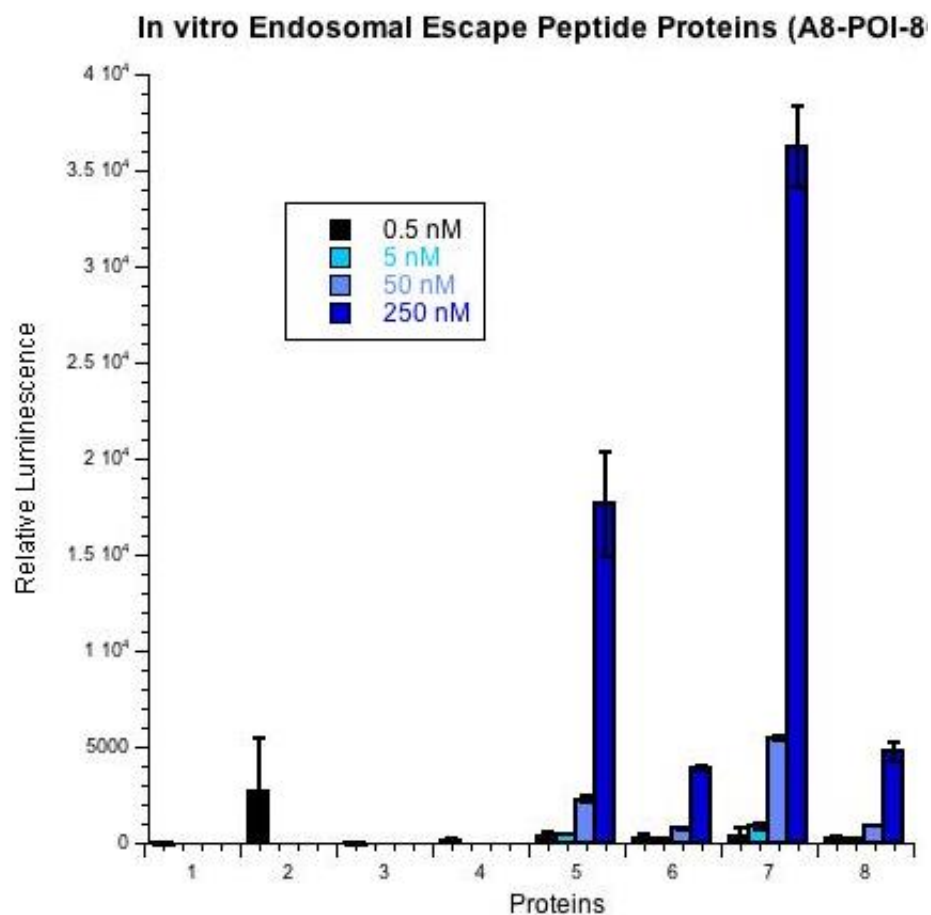
Supplemental Figure C.2: PAGE-SDS gel of Ni-purified SmBit and LgBit fusion proteins



Supplemental Figure C.3: Varying protein concentration incubation in a 96-well plate with 5X, 10X or 20X added Live Cell NanoGlow Reagent from Promega.

A**B****C**

Supplemental Figure C.4: Live Hek293T Cells in suspension analyzed with NanoGlow Live Cell Reagent and a BioTek microplate reader. Mcherry fluorescence and Luminescence were not measured in any sample. GFP Fluorescence was measured only in the wells where cells were incubated with 34GFP-86, which was expected.



Supplemental Figure C.5: In vitro concentration dependent assay using N-terminal endosomal escape peptide fused to our Smbit proteins. Lanes from left to right: PBS, cell lysate, A8-BlacNB-86 only, mCherry-LgBit Only, Cell lysate + A8-BlacNB-86, Cell lysate + A8-Cat BlacNB-86, Cell Lysate + A8-sfGFP-86, Cell Lysate + A8-34GFP-86. Non-resurfaced proteins still display a higher luminescence than resurfaced proteins