

**EXPANDING THE CHEMICAL DIVERSITY OF PHAGE-DISPLAYED PEPTIDE**

**LIBRARIES**

A Dissertation

by

JEFFERY MICHEAL THARP II

Submitted to the Office of Graduate and Professional Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe R. Liu
Committee Members,	Frank M. Raushel
	Tadhg P. Begley
	Christian Hilty
Head of Department,	Simon W. North

May 2018

Major Subject: Chemistry

Copyright 2018 Jeffery M. Tharp II

## ABSTRACT

Phage display is a widely-used tool for the directed evolution of peptide ligands from large combinatorial peptide libraries. A limitation of this technique arises from the lack of chemical diversity of the phage-displayed peptides. Being that they are genetically encoded, phage-displayed peptides are limited to the functional groups found in the twenty canonical amino acids. However, using orthogonal aminoacyl-tRNA synthetase/tRNA pairs, researchers have been able to greatly expand the genetic code to incorporate numerous non-canonical amino acids (ncAAs) bearing diverse chemical functionalities. In this thesis we utilize the technique of genetic code expansion to incorporate ncAAs into phage-displayed peptide libraries, thereby, greatly expanding their utility.

We begin by investigating the substrate spectrum of a rationally designed pyrrolysyl-tRNA synthetase (PylRS) mutant used for genetic code expansion. We demonstrate that, in addition to *meta*-, and *para*-substituted phenylalanine derivatives, this mutant can also facilitate the genetic incorporation of *ortho*-substituted phenylalanine in response to TAG codons in *E. coli* and in mammalian cells. We also report on the ability of one of the phenylalanine derivatives to serve as a fluorescent probe of protein local environment.

Next, we turn our attention to using PylRS to incorporate ncAAs in response to TAG codons in phage display libraries. To do this, we developed a method for constructing phage libraries in which every clone contains an in-frame TAG codon. Such

“TAG-obligate” libraries overcome obstacles associated with libraries that contain both TAG and non-TAG clones. Using our TAG-obligate library, in conjunction with PylRS mutants, we demonstrate the genetic incorporation of lysine and phenylalanine derivatives into phage-displayed peptides. One lysine derivative,  $N_{\epsilon}$ -butyryl-lysine, is a naturally occurring lysine posttranslational modification. We use the library containing this ncAA to select for peptide ligands of the lysine deacylase sirtuin 2. We demonstrate that peptides identified from this selection, when synthesized with a structural analog of  $N_{\epsilon}$ -butyryl-lysine,  $N_{\epsilon}$ -thiobutyryl-lysine, are potent inhibitors of sirtuin 2-catalyzed lysine deacylation.

In a final study, we report on the design and synthesis of a regiospecific, kinetically controlled, bis-electrophile for cyclizing phage-displayed peptides. We use this new compound to cyclize peptides on a phage-displayed library and screen the library against two model proteins. Future directions for this ongoing work are also discussed.

To my parents and grandparents

Without their continued love and support this work would not have been possible.

## ACKNOWLEDGEMENTS

First, I would like to thank my advisor Professor Wenshe Liu for his guidance and tremendous support throughout the course of my graduate research. I would also like to thank my committee members, Professors Christian Hilty, Frank Raushel, and Tadhg Begley for their challenging questions, insightful comments, and helpful discussions.

There are many members of the Liu Group, past and present, to whom I owe a debt of gratitude. A special thanks to Yane-Shih Wang for being a fantastic mentor and friend. I would also like to thank my friends and colleagues Yan-Jiun Lee, Alfred Tuley, Bo Wu, Yadagiri Kurra, Catrina Reed, Xiaoshan Wang, and Wesley Wang for their support and kindness during my time in the Liu Group. I should give a special thanks to Willie Hsu, Vangmayee Sharma, Zhipeng “Zeepam” Wang, and Erol Vatansever for creating a fun and exciting work environment.

My sincere thanks to all of the Chemistry Department staff, especially Sandy Horton, Amy Liu, Dr. Joanna Pellois, and Ron Carter. Their hard work has allowed me to focus my attention on research and for that I am grateful. Thanks also to the Organic Chemistry Teaching Laboratory staff, including Carrie Nichols, Janet Robinson, and Dr. Robert Hildreth. Thanks to Drs. Yohannes Rezenom and Bo Wang of the Department of Chemistry Mass Spectrometry Facility for all of their help in collecting data that were essential to my research.

I would like to extend a special thanks to all of the great teachers I have had throughout my academic studies. In particular I would like to thank my high school

chemistry teacher Mrs. Melanie Huber. I would also like to extend my gratitude to the faculty in the Department of Chemistry and Physics at Indiana State University.

I would like to thank my friend Dan Rodriguez for encouraging me to pursue my PhD studies at TAMU and for showing me that true friendship cannot be weakened with distance. A special thanks to Andreas Ehnbohm who has made the past three years immeasurably more enjoyable.

Finally, I would like to thank my family. I am extremely fortunate to have a large and loving family. So many of you have expressed your encouragement that I could not possibly list everyone here. Your support has made this challenging journey a little easier. A special thanks to my Aunt Kathy, Aunt Niki, Aunt Dar, and Great Grandma Sheese. Thanks to my Uncle Curt, Aunt Augusta, Collin, Logan, and Mason. Huge thanks to my sister, Kaylee, and brother, Kyle. Thanks to my grandma and grandpa for their “love and prayers”. Finally, thanks to my mother and father for their support, love, and encouragement throughout this time. I could not have done this without you.

## CONTRIBUTORS AND FUNDING SOURCES

### *Faculty Committee Recognition:*

This work was conducted under the advisement of Professor Wenshe Liu of the Department of Chemistry and supervised by a thesis committee consisting of Professors Christian Hilty and Tadhg Begely of the Department of Chemistry and Professor Frank Raushel who holds joint appointment in the Department of Chemistry and the Department of Biochemistry and Biophysics.

### *Student/Collaborator Contributions:*

For the work described in Chapter 3, Yane-Shih Wang contributed to experiments involving the expression of sfGFP-S2X mutants, Yan-Jiun Lee constructed the plasmid pBAD-MPP8-sfGFP and mutants thereof, Yanyan Yang constructed the plasmid pEGFP-N1-EGFR and advised in cell culturing and image collection. For the work described in Chapter 4, Catrina Reed constructed the heptapeptide phagemid library and performed the superinfection-selection for TAG enrichment. Xiaoshan Wang synthesized the compound Boc-(Ac)Lys-AMC which was used as a substrate for the discontinuous sirtuin assay described in Chapter 5. Compounds **28**, **29**, **35**, **41**, **47**, **53**, and **61** were synthesized by Yadagiri Kurra. All ESI and MALDI-TOF mass spectra were collected in collaboration with the Department of Chemistry Mass Spectrometry Facility. All other work conducted for this thesis was completed by the student independently.

*Funding Sources*

Graduate study was partially supported by a fellowship from Texas A&M University, by the National Institutes of Health (Grants R01CA161158 and R01GM121584), and by the Welch Foundation (Grant A-1715). Funding agencies had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.



## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	v
CONTRIBUTORS AND FUNDING SOURCES.....	vii
TABLE OF CONTENTS .....	ix
LIST OF FIGURES.....	xiii
LIST OF TABLES .....	xvii
CHAPTER I INTRODUCTION .....	1
1.1 Peptide Display .....	1
1.2 Bacteriophages .....	2
1.3 Filamentous Bacteriophages .....	3
1.3.1 The Structure of Ff Bacteriophages .....	5
1.3.2 Adsorption, Infection, and Replication .....	8
1.4 Phage Display.....	11
1.4.1 Principles of Phage Display .....	11
1.4.2 Display on Filamentous Phages .....	14
1.5 Concluding Remarks .....	18
CHAPTER II EXPANDING THE SCOPE OF FILAMENTOUS PHAGE DISPLAY ..	19
2.1 Expanding the Chemical Diversity of Phage-Displayed Peptides .....	19
2.1.1 Enzymatic Posttranslational Modification .....	20
2.1.2 Native Chemical Ligation .....	22
2.1.3 Chemical Posttranslational Modification .....	23
2.1.4 Phage Display with an Expanded Genetic Code .....	25
2.2 Phage Display of Cyclic Peptides .....	31
2.2.1 Genetically Encoded Cyclic Peptides.....	31
2.2.2 Chemically Cyclized Peptides.....	33
2.3 Concluding Remarks .....	37

	Page
CHAPTER III GENETIC INCORPORATION OF SEVEN <i>ORTHO</i> -SUBSTITUTED PHENYLALANINE DERIVATIVES .....	39
3.1 Introduction .....	39
3.2 Experimental Details .....	42
3.2.1 Miscellaneous Information .....	42
3.2.2 Primer List .....	42
3.2.3 Plasmid Construction .....	43
3.2.4 Protein Expression and Purification .....	44
3.2.5 Mammalian Cell Cultures and Transfection .....	46
3.2.6 Fluorescence Measurements .....	46
3.3 Results and Discussion .....	47
3.3.1 PylRS(N346A/C348A) Recognizes <i>Ortho</i> -Substituted Phenylalanine as a Substrate .....	47
3.3.2 The use of PylRS(N346A/C348A) in Mammalian Cells .....	54
3.3.3 <i>Ortho</i> -Cyano-Phenylalanine Is an Environmentally Sensitive Fluorescent Probe .....	54
3.4 Conclusion .....	61
CHAPTER IV CONSTRUCTION AND CHARACTERIZATION OF A TAG-OBLIGATE PHAGE DISPLAY LIBRARY .....	62
4.1 Introduction .....	62
4.2 Experimental Details .....	65
4.2.1 Miscellaneous Information .....	65
4.2.2 Primer List .....	66
4.2.3 Plasmid Construction .....	66
4.2.4 Protein Expression and Purification .....	67
4.2.5 Biotinylation of TEV Protease .....	69
4.2.6 Phage Quantification .....	69
4.2.7 Superinfection-Selection to Generate the TAG-Obligate Phagemid Library .....	70
4.2.8 Expression and Purification of Phages Containing ncAAs .....	71
4.2.9 Affinity Selection against TEV Protease .....	73
4.2.10 Affinity Selection against Streptavidin .....	75
4.2.11 Chemical Synthesis .....	76
4.3 Results and Discussion .....	79
4.3.1 Construction of a TAG-Obligate Phage Display Library .....	79
4.3.2 Genetic Incorporation of ncAAs into Phage-Displayed Peptides .....	86
4.3.3 Affinity Selection of Peptides Containing ncAAs .....	92
4.4 Conclusion .....	95

CHAPTER V PHAGE SELECTION OF PEPTIDES CONTAINING  
NON-CANONICAL AMINO ACIDS FOR THE INHIBITION OF SIRTUIN 2 ..... 96

5.1	Introduction .....	96
5.2	Experimental Details .....	99
5.2.1	Miscellaneous Information .....	99
5.2.2	Primer List .....	100
5.2.3	Plasmid Construction .....	100
5.2.4	Protein Expression and Purification .....	101
5.2.5	Biotinylation of Sirtuin 2 .....	106
5.2.6	BuKRS Library Screening .....	107
5.2.7	In-Cell Fluorescence Measurements .....	107
5.2.8	Phagemid Library Construction and Transformation .....	108
5.2.9	Phage Expression and Purification .....	108
5.2.10	Affinity Selection against Sirtuin 2 .....	109
5.2.11	Solid-Phase Peptide Synthesis .....	111
5.2.12	Sirtuin 2 Discontinuous Inhibition Assay .....	119
5.2.13	Sirtuin 1, 2, and 3 Inhibition Assay .....	119
5.2.14	Chemical Synthesis .....	120
5.3	Results and Discussion .....	130
5.3.1	Genetic Incorporation of thBuK .....	130
5.3.2	Affinity Selection against Sirtuin 2 .....	135
5.3.3	Synthetic Peptides Containing thBuK Inhibit Sirtuin 2 .....	139
5.3.4	Selected Peptides Containing thBuK Exhibit Isoform Selectivity Comparable to TB .....	142
5.4	Conclusion .....	144

CHAPTER VI EFFORTS TOWARDS THE SELECTION OF  
PHAGE-DISPLAYED CYCLIC PEPTIDES USING A REGIOSPECIFIC,  
KINETICALLY CONTROLLED, BIS-ELECTROPHILE..... 146

6.1	Introduction .....	146
6.2	Experimental Details .....	149
6.2.1	Miscellaneous Information .....	149
6.2.2	Primer List .....	149
6.2.3	Plasmid Construction .....	149
6.2.4	Protein Expression and Purification .....	150
6.2.5	Biotinylation of HRas .....	152
6.2.6	CA <sub>5</sub> C-sfGFP Cyclization with CAmCBT .....	153
6.2.7	Phage Expression and Purification .....	154
6.2.8	Labelling Phages with FITC-CBT .....	154
6.2.9	Phage Infectivity Assay .....	155

	Page
6.2.10 Peptide Library Cyclization with CAmCBT .....	155
6.2.11 Affinity Selection against HRas .....	155
6.2.12 Affinity Selection against Plasma Kallikrein .....	156
6.2.13 Solid-Phase Peptide Synthesis .....	156
6.2.14 Synthetic Peptide Cyclization .....	157
6.2.15 Chemical Synthesis .....	157
6.3 Results and Discussion.....	161
6.3.1 Cyclization of a Model Peptide with CAmCBT .....	161
6.3.2 Cyclization of Phage-Displayed Peptides .....	162
6.3.3 Affinity Selection against HRas and Plasma Kallikrein .....	165
6.3.4 Modification of Synthetic Peptides with CAmCBT and First Attempts towards Measuring Binding Affinity .....	167
6.3.5 Future Directions.....	168
6.4 Conclusion.....	169
 CHAPTER VII CONCLUDING REMARKS .....	 170
 REFERENCES .....	 174
 APPENDIX LIST OF ABBREVIATIONS .....	 209

## LIST OF FIGURES

FIGURE	Page
1.1 Bacteriophage lifecycles .....	4
1.2 The structure of Ff bacteriophages .....	6
1.3 The mechanism of Ff bacteriophage infection .....	9
1.4 Ff bacteriophage replication .....	11
1.5 Principles of phage display .....	13
1.6 Display systems for filamentous phage display.....	16
2.1 Enzymatic modifications of phage-displayed peptides .....	21
2.2 Chemical methods for introducing ncFGs into peptides displayed on Ff bacteriophages .....	23
2.3 Genetic incorporation of selenocysteine into peptides displayed on the N-terminus of pIII .....	26
2.4 The genetic incorporation of ncAAs via nonsense suppression .....	28
2.5 Genetic incorporation of ncAAs into peptides displayed on the N-terminus of pIII .....	30
2.6 Disulfide constrained phage-displayed peptide libraries .....	32
2.7 Chemical cyclization of phage-displayed peptides.....	35
3.1 SDS-PAGE analysis of sfGFP and MPP8.....	45
3.2 The superimposed structures of OmeRS in complex with <i>O</i> -methyl-tyrosine and PylRS in complex with pyrrolysyl-adenosyl monophosphate .....	48
3.3 Genetic incorporation of seven <i>ortho</i> -substituted phenylalanine derivatives .....	49
3.4 The genetic incorporation of <i>ortho</i> -trifluoromethyl-phenylalanine and <i>ortho</i> -fluoro-phenylalanine .....	51

FIGURE	Page
3.5 The genetic incorporation of <i>ortho</i> -substituted phenylalanine derivatives in LB media .....	53
3.6 Expression of EGFP-N128→X in HEK293T cells co-transfected with the plasmid containing genes encoding PylRS(N346A)/(C348A) and tRNA <sub>CUA</sub> <sup>Pyl</sup> .....	55
3.7 Fluorescence emission spectrum of 20 μM <i>ortho</i> -cyano-phenylalanine dissolved in water, ethanol, and 6 M guanidine hydrochloride .....	57
3.8 <i>Ortho</i> -cyano-phenylalanine fluorescence in sfGFP variants .....	58
3.9 Monitoring protein unfolding with <i>ortho</i> -cyano-phenylalanine .....	59
3.10 Fluorescence intensity at 297 nm of wildtype sfGFP as a function of urea concentration.....	60
4.1 SDS-PAGE analysis of purified TEV protease .....	67
4.2 The synthesis of 2-amino-8-oxodecanoic acid .....	76
4.3 <sup>1</sup> H-NMR spectrum of 2-amino-8-oxodecanoic acid.....	79
4.4 A general strategy for constructing TAG-obligate phage display libraries ....	82
4.5 The percentage of cells infected during rounds 1 and 2 of superinfection-selection .....	84
4.6 The sequence of clones isolated after the 2 <sup>nd</sup> round of superinfection-selection .....	85
4.7 The genetic incorporation of phenylalanine derivatives into phage-displayed peptides.....	87
4.8 The genetic incorporation of N <sub>ε</sub> -butyryl-lysine and N <sub>ε</sub> -crotonyl-lysine into phage-displayed peptides.....	90
4.9 The genetic incorporation of 2-amino-8-oxodecanoic acid.....	91
4.10 The amino acid sequence of peptides containing phenylalanine derivatives isolated from selections against TEV protease and streptavidin .....	93

FIGURE	Page
5.1 The mechanism of sirtuin-catalyzed lysine deacylation and inhibition by thioamides.....	97
5.2 SDS-PAGE analysis of purified sirtuin 2 .....	103
5.3 SDS-PAGE analysis of purified sirtuin 1 and GST-sirtuin 3 .....	105
5.4 Characterization data for the peptide S2P07 .....	114
5.5 Characterization data for the peptide S2P04 .....	115
5.6 Characterization data for the peptide S2P04-5 .....	116
5.7 Characterization data for the peptide Ac-S2P04-5 .....	117
5.8 Characterization data for the peptide S2P03 .....	118
5.9 The synthesis of $N_{\epsilon}$ -thiobutyryl-lysine and $N_{\alpha}$ -Fmoc- $N_{\epsilon}$ -thiobutyryl-lysine.....	121
5.10 Characterization data for $N_{\alpha}$ -Fmoc- $N_{\epsilon}$ -thiobutyryl-lysine.....	125
5.11 The synthesis of TB .....	126
5.12 Characterization data for TB .....	129
5.13 The genetic incorporation of $N_{\epsilon}$ -thiobutyryl-lysine.....	131
5.14 Phage expression with thBuK and BuK .....	132
5.15 The crystal structure of PylRS in complex with adenylated-BuK and expression of sfGFP with BuKRS(N346A).....	134
5.16 The amino acid sequence of peptides containing BuK identified from affinity selection against sirtuin 2.....	136
5.17 Inhibition curves for the inhibition of sirtuin 2 with TB and synthetic peptides containing thBuK .....	140
5.18 Inhibition curves for the inhibition of sirtuin 1, sirtuin 2, and sirtuin 3 with TB and S2P04.....	143

FIGURE	Page
6.1 The reaction of 6-hydroxy-2-cyanobenzothiazole with cysteine and the reaction of CAmCBT with phage-displayed peptides .....	148
6.2 SDS-PAGE analysis of purified HRas .....	153
6.3 The synthesis of CAmCBT .....	158
6.4 The synthesis of FITC-CBT .....	160
6.5 ESI mass spectra of CA <sub>5</sub> C-sfGFP before and after treatment with CAmCBT .....	162
6.6 Labelling phages with FITC-CBT and the effect of CAmCBT on phage viability.....	164
6.7 The amino acid sequence of peptides identified from affinity selection against plasma kallikrein and HRas.....	166
6.8 MALDI-TOF mass spectra of the peptide HRP24 before and after reaction with CAmCBT .....	168



## LIST OF TABLES

TABLE		Page
3.1	ESI-MS characterization of sfGFP variants incorporated with different ncAAs .....	50
5.1	IC <sub>50</sub> values measured for TB and synthetic peptides with sirtuin 2.....	141
5.2	IC <sub>50</sub> values measured for TB and S2P04 with sirtuin 1, sirtuin 2, and sirtuin 3 .....	143

# CHAPTER I

## INTRODUCTION

### 1.1 Peptide Display

There is a growing need for the discovery of novel therapeutics for the treatment of human diseases. Recent years have seen increased interest in the development of protein and peptide-based therapeutics.<sup>1-3</sup> Peptides offer several advantages over traditional small molecule drugs. First, due to their larger size, peptides provide higher target selectivity and binding affinity giving rise to lower toxicity and fewer off-target side effects compared to small molecules.<sup>4, 5</sup> Peptides also offer benefit in terms of their ease of access. Modern methods of peptide synthesis allow for the large-scale production of highly modified peptides that can be obtained at pharmaceutical grade purity.<sup>1, 6</sup> Finally, novel peptide ligands for a particular target can be readily identified using *in vitro* evolution techniques known as peptide display. Peptide display allows for the selection of peptide ligands from enormous combinatorial peptide libraries that are far larger than those that can be achieved with small molecules via high-throughput screening.<sup>7</sup> These techniques have the potential to greatly accelerate the rate at which new therapeutic peptide leads are discovered.

In contrast to synthetic peptide libraries, peptide display libraries are genetically encoded, that is, the peptides are translated from large pools of randomized DNA. A distinctive feature of peptide display is that the peptides remain physically linked to their encoding DNA. This connection between genotype and phenotype (peptide sequence)

provides two considerable advantages over synthetic libraries. First, it provides a means for replication of the peptide library and, thereby, a mechanism for the directed evolution of peptides with desired properties. Second, as peptides with desired properties are evolved, their sequences can be readily determined through sequencing of the corresponding DNA.

Various peptide display strategies have been reported.<sup>7</sup> The first, and most widely used among these is phage display.<sup>8</sup> To facilitate discussion in subsequent chapters, this chapter will provide a brief introduction to phage display with a particular emphasis on the biology of filamentous bacteriophages and their use as scaffolds for peptide display.

## **1.2 Bacteriophages**

Bacteriophages (also called phages) are viruses that infect bacteria and use the biosynthetic machinery of their host for reproduction. In Nature, bacteriophages are ubiquitous and are the most numerous types of organisms outnumbering bacteria by up to ten to one.<sup>9, 10</sup> Although highly diverse, most bacteriophages consist of a DNA or RNA genome encapsidated by a proteinaceous coat.<sup>11</sup> The viral coat contains structural proteins that serve to protect the genome from the environment as well as functional proteins that are directly involved in host recognition and infection.

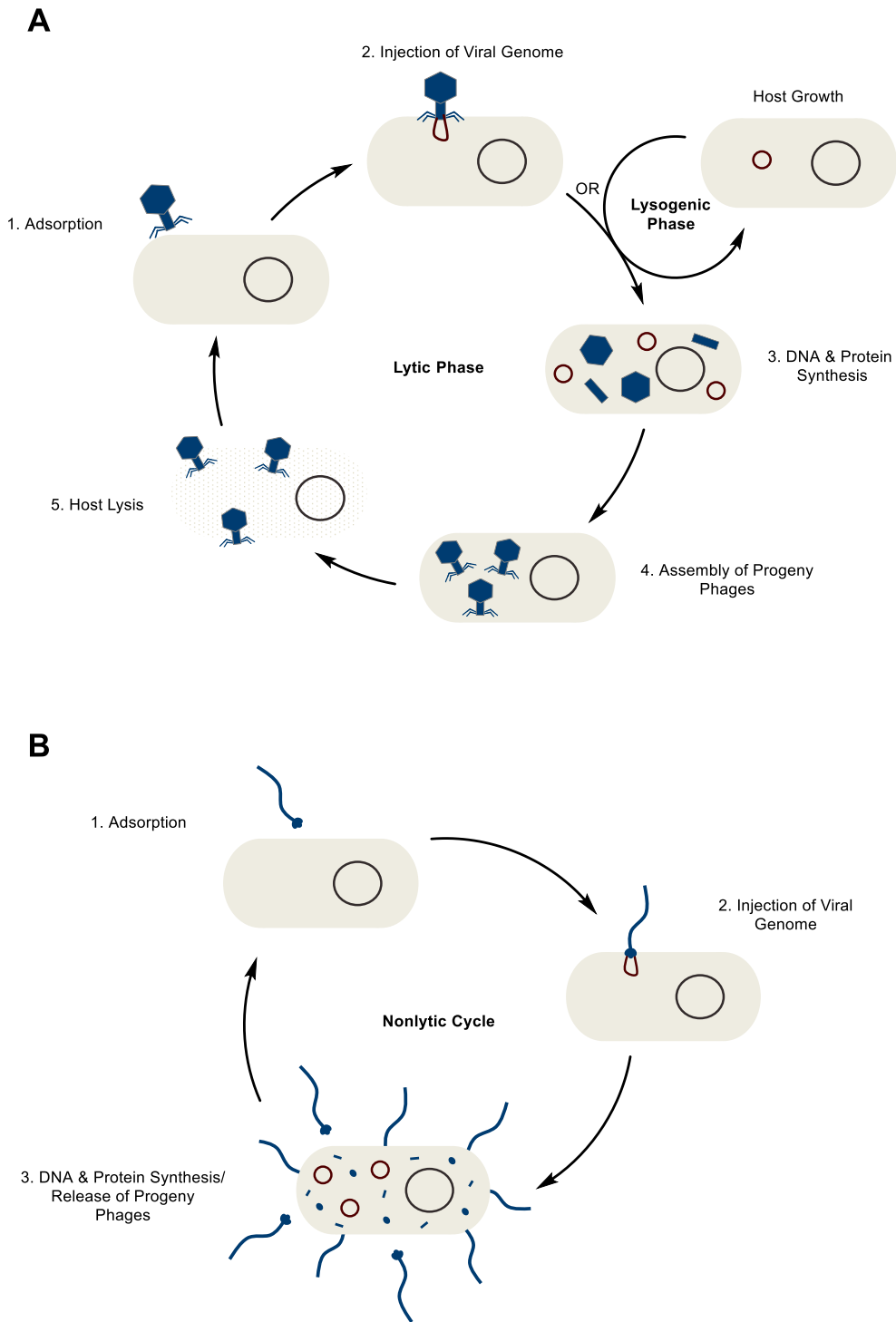
The process of infection is initiated by adsorption of the phage to its host. This involves highly specific interactions between phage coat proteins and receptors on the host's outer surface.<sup>12</sup> Many bacterial cell surface structures are known to be exploited by bacteriophages for this purpose. The presence or absence of a receptor determines the

narrow range of hosts that a phage is able to infect.<sup>12, 13</sup> Following adsorption, the phage completes infection by injecting its genome into the cytoplasm of the bacterial cell.

Phages can be broadly classified based on their reproductive lifecycle after infection. Three classes of bacteriophages are known including lytic (virulent), lysogenic (temperate), and nonlytic phages.<sup>12</sup> Following infection, lytic phages immediately begin the process of genome replication and protein expression. These newly synthesized molecules are assembled into progeny virions that are released upon phage-induced lysis of the host (Figure 1.1A).<sup>14</sup> Lysogenic phages, however, enter a dormant phase where, after infection, the phage genome is integrated into that of its host or is maintained as a separate entity within the cytoplasm. This lysogenic phase is highly stable and can be maintained for many generations. Certain environmental conditions trigger lysogenic phages to enter into a lytic phase in which progeny virions are produced and released following host lysis.<sup>15</sup> Unlike lytic and lysogenic phages, nonlytic phages do not lead to death of the host cell. Rather, nonlytic phages result in a state of chronic infection during which progeny virions are continually produced and released from the host (Figure 1.1B).<sup>9</sup> Bacteria infected with nonlytic phages continue to grow and divide albeit at a slower rate than uninfected cells.

### **1.3 Filamentous Bacteriophages**

Several types of bacteriophages have been used for phage display including the lytic T4<sup>16, 17</sup> and T7<sup>18</sup> phages and the lysogenic lambda phage<sup>19</sup>. By far the most widely used



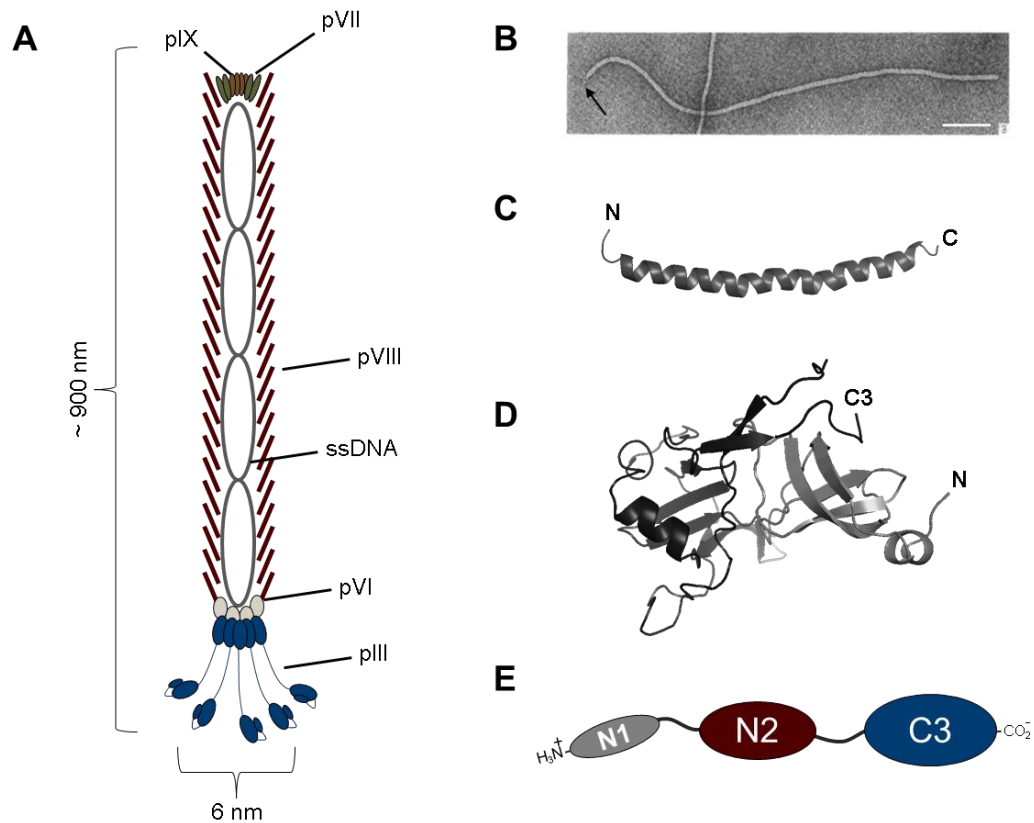
**Figure 1.1. Bacteriophage lifecycles.** (A) The lifecycle of lytic and lysogenic bacteriophages. (B) The lifecycle of nonlytic bacteriophages. Phages and phage proteins are represented as blue shapes. Viral DNA is represented as maroon circles whereas host DNA is represented as black circles. The host cell is represented by a gray oval.

phages for phage display, and the phages used in this work, are the nonlytic, F-pilus-specific, filamentous phages of *E. coli*. termed the Ff phages (F-pilus, filamentous). These include the closely related and highly similar f1, fd and M13 bacteriophages.<sup>8, 20</sup> Ff phages are easily manipulated, tolerate a wide range of structural modifications, and are stable at extremes of temperature and pH making them exceptional tools for peptide display.<sup>21</sup>

### *1.3.1 The Structure of Ff Bacteriophages*

As their name implies, filamentous phages are characterized as long flexible filaments of about 900 nm long and 6 nm wide (Figure 1.2A).<sup>22, 23</sup> The Ff viral particle consists of only five different proteins which encapsidate a single-stranded DNA (ssDNA) genome of 6.4 kb.<sup>24</sup> The genome has a noncoding intergenic region containing the f1 origin of replication as well as a packing signal that marks the genome for packaging into progeny virions during replication.<sup>25</sup> There are nine genes that, due to the presence of alternate start sequences, encode 11 proteins.<sup>21</sup> These proteins include the five capsid proteins, three proteins needed for genome replication, and three proteins required for assembly of the progeny phages.

The majority of the viral particle (87% by mass) consists of a short helical protein called pVIII, also known as the major coat protein (Figure 1.2C).<sup>8</sup> pVIII coats the genome in an overlapping pattern reminiscent of fish scales that are held together by hydrophobic interactions between neighboring subunits.<sup>26</sup> Positively charged lysine



**Figure 1.2. The structure of Ff bacteriophages.** (A) Diagram of an Ff bacteriophage. (B) An fd bacteriophage visualized via electron microscopy. The arrow points to the N1-N2 complex of pIII which can be seen as faint ‘knobs’. The white bar represents 100 nm. Reprinted from the Journal of Molecular Biology, Vol 146, C.W. Gray, R.S. Brown, and D.A. Marvin, Adsorption Complex of Filamentous fd Virus, 621-627, Copyright 1981, with permission from Elsevier. (C) The structure of the major coat protein pVIII. N- and C-termini are labeled. (PDB: 2C0W). (D) X-ray crystal structure of the N1-N2 Complex of pIII. N1 and N2 are shown in gray and black, respectively. The N-terminus of N1 is labeled whereas the end that is attached to the C-terminal domain in the full length protein is labelled C3. The glycine rich linker connecting the two domains is not shown. (PDB: 1G3P) (E) Cartoon representation of the three domains of pIII. The black lines between the domains represent glycine rich linkers.

residues at the C-terminus of pVIII interact with the negatively charged viral DNA. In wildtype phages there are, on average, 2,700 copies of pVIII; however, the number of pVIII monomers is highly flexible and large additions and deletions in the genome are accommodated by increasing or decreasing the number of pVIII molecules in the capsid.<sup>8</sup>

The two distal ends of the virus are capped by the four minor coat proteins that are each present in about five copies.<sup>27, 28</sup> pVII and pIX together occupy one end of the particle. These proteins, each about 30 residues, are the first to be extruded from the bacterial cell during replication.<sup>29</sup> The opposite end of the virus is occupied by pVI and pIII. These proteins are the last to be incorporated into the capsid as newly formed particles leaves the host.<sup>30</sup> Little structural information is available for the minor coat proteins and their exact arrangement within the viral particle is not known.

Along with its role in terminating phage assembly, pIII is also solely responsible for host recognition and adsorption. Slightly more is known about the structure of pIII than the other minor coat proteins. pIII is the largest of the viral coat proteins at 424 residues (including an 18 residue leader sequence). It is composed of three domains that are separated by glycine rich linkers (Figure 1.2E).<sup>31</sup> Deletion studies have revealed distinct roles for each of these domains. The 121 residue C-terminal domain (C3) is responsible for mediating the termination of viral assembly during replication as well as anchoring pIII into the viral capsid.<sup>32, 33</sup> The two N-terminal domains (N1 and N2, 67 and 131 residues, respectively) are responsible for host recognition and adsorption.<sup>34, 35</sup> The structure of the N-terminal domains has been solved by X-ray crystallography and solution NMR.<sup>36-38</sup> These structures have revealed that N1 and N2 share a highly similar fold with each domain composed of two antiparallel  $\beta$ -sheets and a short  $\alpha$ -helix (Figure 1.2D). Prior to adsorption, these two domains interact via a well-defined binding surface.<sup>36, 39</sup> The N1-N2 complex has been visualized via electron microscopy as ‘knobs’

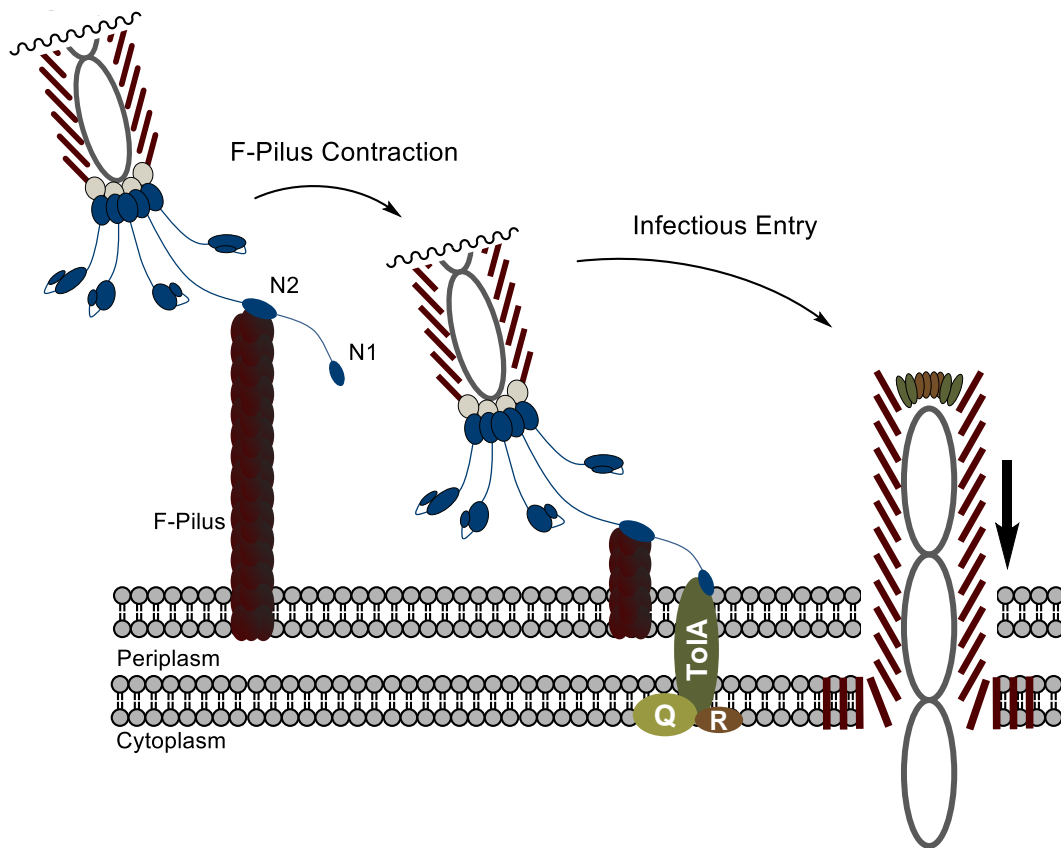


situated slightly apart from the tip of the virus that remain tethered by the interdomain glycine rich linker (Figure 1.2B).<sup>40</sup>

### *1.3.2 Adsorption, Infection, and Replication*

Ff phages make use of two receptors for adsorbing to their host. The primary receptor is a long hair-like appendage of *E. coli* known as the conjugative F sex pilus. In the first step of adsorption the N2 domain of pIII binds to the tip of the F-pilus (Figure 1.3).<sup>41</sup> The secondary receptor for adsorption is the TolA protein located in the TolQRA complex which spans the inner and outer membrane of *E. coli*.<sup>39, 42</sup> The N1 domain of pIII binds specifically to the C-terminus of TolA. Prior to attachment to the F-pilus, the TolA binding surface on N1 lies buried in the interface between the N1 and N2 domains. Upon binding of N2 to the F-pilus, a conformational change in pIII weakens the interaction between N1 and N2 exposing the TolA binding surface on N1.<sup>43-45</sup> The F-pilus of *E. coli* has been shown to undergo spontaneous cycles of extension and contraction.<sup>46</sup> After N2 binds, contraction of the F-pilus brings the phage in close proximity to the outer membrane allowing N1 to bind to TolA. Following adsorption, the viral coat proteins integrate into the host membrane as the ssDNA enters the cytoplasm.<sup>47</sup> The exact molecular mechanism of infectious entry of the viral genome is still poorly understood.

Once in the cytoplasm, the single stranded viral genome (the (+) strand) is converted to the double stranded 'replicative form' by host RNA and DNA polymerases

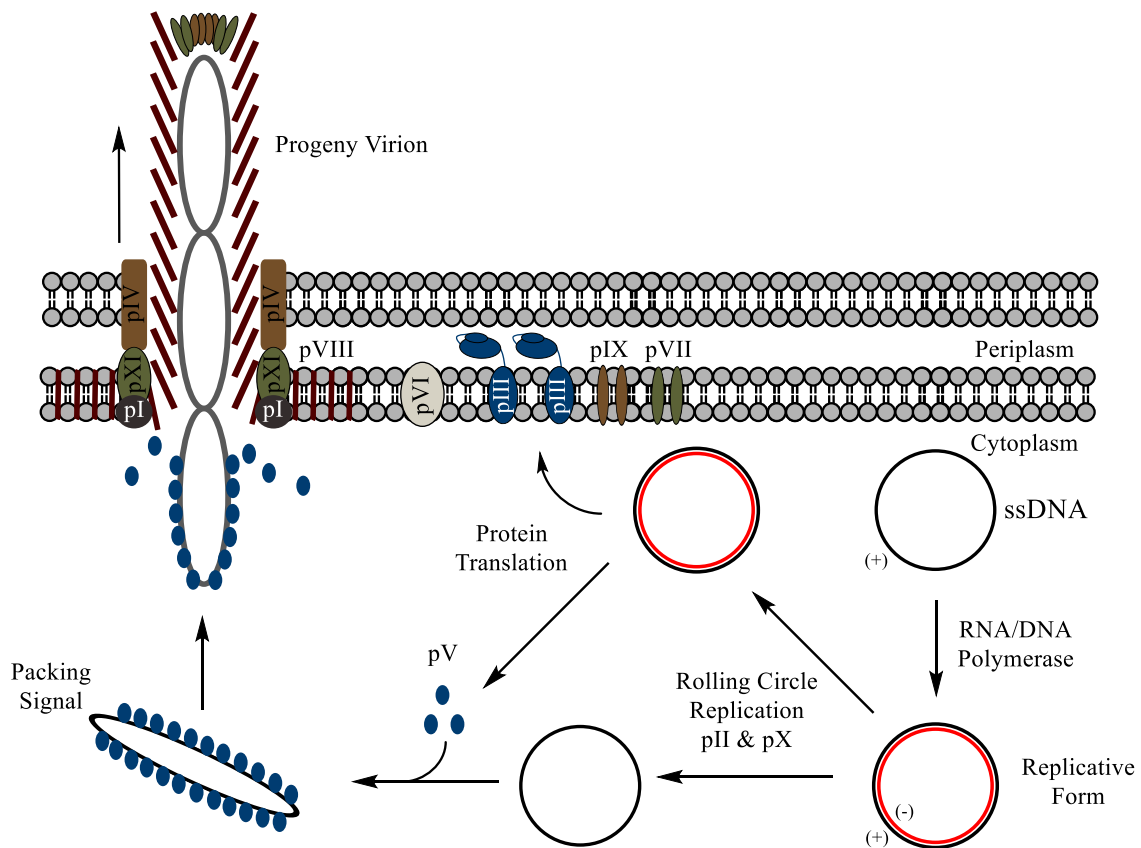


**Figure 1.3. The mechanism of Ff bacteriophage infection.** In the first step of infection the N2 domain of pIII binds to the tip of the F-pilus. F-pilus contraction brings the phage in proximity to the cell surface where the N1 domain of pIII binds to the C-terminus of TolA. After adsorption, the phage genome is injected into the cytoplasm as the coat proteins integrate into the inner membrane.

(Figure 1.4). The former acts as a primase for the synthesis of the new complimentary (-) strand by DNA polymerase.<sup>48</sup> The newly synthesized (-) strand serves as a template for gene transcription and for the synthesis of new (+) strands. Protein translation is entirely dependent on host machinery whereas three phage proteins (pII, pX, and pV) are involved in genome replication and regulating transcription. pII assists in replication of the genome via a rolling circle mechanism, specifically, cleavage of the template and ligation of the newly synthesized DNA.<sup>49</sup> The pX protein is identical to the C-terminal

one-third of pII and is likely involved in regulating DNA synthesis by inhibiting pII.<sup>50</sup> pV acts as a chaperone for the (+) strand DNA before it is packaged into the new virions. As the cell accumulates sufficient quantities of viral DNA and proteins, pV dimers begin to bind to and sequester newly synthesized (+) strands. The single stranded (+) strand is completely coated with pV with the exception of a hairpin sequence that serves as a signal for packaging into progeny virions. Along with this, pV has also been shown to be involved in regulating transcription of pVI and pII.<sup>51, 52</sup>

The remaining phage proteins are directly involved in assembly and secretion of progeny virions. Assembly of progeny phages is facilitated by an ATP-dependent assembly complex formed from three proteins: pI, pXI, and pIV. The assembly complex spans the inner and outer membrane of *E. coli*. pI and pXI are located in the inner membrane whereas pIV is located in the outer membrane within the complex. Together these proteins form a channel through which newly assembled phages pass out of the cell.<sup>53-55</sup> Following translation, all five coat proteins are localized to the *E. coli* inner membrane.<sup>28</sup> In a process that is not completely understood, progeny phages are simultaneously assembled and secreted from the host. Extrusion of the phage is initiated by binding of pIX and pVII to the packaging signal sequence on the phage genome. As the phage exits the cell, pV which coats the ssDNA in the cytoplasm, is replaced by pVIII. Finally, assembly and release of progeny phages is completed with the addition pVI and pIII to the viral capsid.<sup>56</sup>



**Figure 1.4. Ff bacteriophage replication.** Once in the cytoplasm, the single-stranded viral DNA ((+) strand) is converted to double-stranded ‘replicative form’ by host RNA and DNA polymerase. The newly synthesized (-) strand serves as a template for transcription and the synthesis of new (+) strands. As phage proteins accumulate in the cell they are assembled and secreted by the assembly complex formed from viral proteins pI, pXI, and pIV.

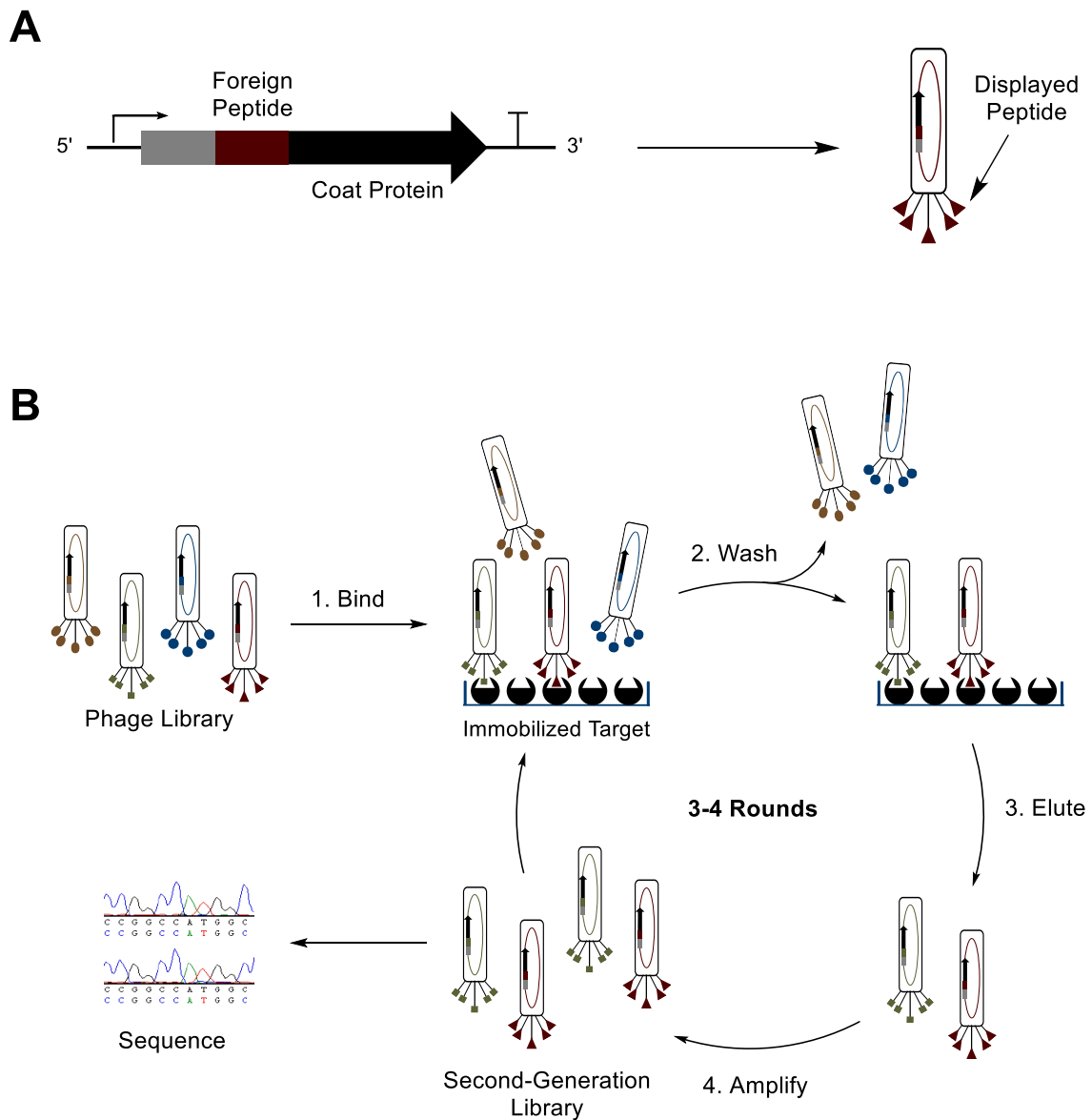
## 1.4 Phage Display

### 1.4.1 Principles of Phage Display

Bacteriophages were first described in 1915 as “ultra-microscopic viruses” that, when introduced to micrococci, result in destruction of the bacterial culture.<sup>57</sup> Over the past century much has been learned about the structure and function of bacteriophages. It was quickly realized that these organisms could be exploited as useful tools for chemical biology with applications ranging from drug delivery to nanowire assembly.<sup>14, 58</sup> Phages

can also be used as scaffolds for protein and peptide display. To display peptides on bacteriophages, DNA encoding the peptide is spliced into a gene that codes for one of the viral coat proteins. The gene product is then a fusion protein of the foreign peptide and the viral coat protein. When this fusion protein is integrated into the viral coat, it results in a phage particle that ‘displays’ the foreign polypeptide on its surface (Figure 1.5A).<sup>8</sup> This configuration creates a physical link between the displayed peptide and the genetic information encoding it—both are contained within the viral particle. Using standard techniques for DNA randomization, large libraries of phages can be constructed in which each phage is displaying a unique polypeptide sequence on its surface. Such phage-displayed peptide libraries can be used for the directed evolution of novel peptide ligands for various protein and non-protein targets.

Directed evolution through phage display requires several iterative rounds of selection followed by amplification (Figure 1.5B). Phage display libraries can contain upwards of  $10^{10}$  unique peptide sequences however, only a small fraction of these will bind to a particular target protein.<sup>59</sup> Peptides that bind to a desired target are selected through an affinity selection often called biopanning. During biopanning, the peptide library (in the form of purified virions) is incubated with the immobilized protein target. Phages displaying peptide sequences that bind to the immobilized target are retained whereas those displaying peptides that do not bind are removed in a series of washing steps. Next, captured phages are eluted from the target. Elution is achieved through competition with a known ligand, proteolysis, or through disruption of the protein-peptide interaction by altering the binding conditions, e.g. lowering the pH.



**Figure 1.5 Principles of phage display.** (A) To construct phage display libraries DNA corresponding to a foreign peptide is spliced into the gene encoding one of the viral coat proteins. The corresponding peptide-coat fusion protein is integrated into the phagemid particle and the peptide is displayed on the viral surface. (B) An overview of affinity selection for identifying novel peptide ligands to a protein target. 3-4 cycles of selection and amplification are performed to identify the peptide ligands.

Following selection, the eluted phages are used to infect *E. coli* wherein they will reproduce. Progeny phages inherit their DNA from the parent phage and, therefore, display the same peptide sequence on their surface. This amplification step produces a second-generation library that is enriched with phages displaying peptides that bind to the target. The second-generation library can be used for a further round of affinity selection, often with more stringent binding and washing conditions. After 3-4 rounds of selection and amplification the genomes of selected phages are isolated and sequenced to determine the sequence of peptides that bind to the target.

#### *1.4.2 Display on Filamentous Phages*

Peptides have been successfully displayed on all five coat proteins of the Ff phages, however, the most commonly used proteins for display are pIII or pVIII. In pIII display, peptides are typically fused to the N-terminus of pIII, though there are some reports of display on the C-terminus of the protein.<sup>60</sup> Since each virion contains five copies of pIII in the capsid, five copies of the peptide will be displayed on the phage. Careful consideration must be given to the size of the protein being displayed as large peptides and proteins fused to pIII can impair phage infectivity.<sup>61</sup> In pVIII display, peptides are also fused at the N-terminus. However, given that there are 2,700 copies of pVIII per virion, even short peptide additions to pVIII significantly increase the mass of the phage and can have a negative impact on capsid stability.<sup>61</sup> For this reason, display on pVIII is limited to short peptides of up to eight amino acids.

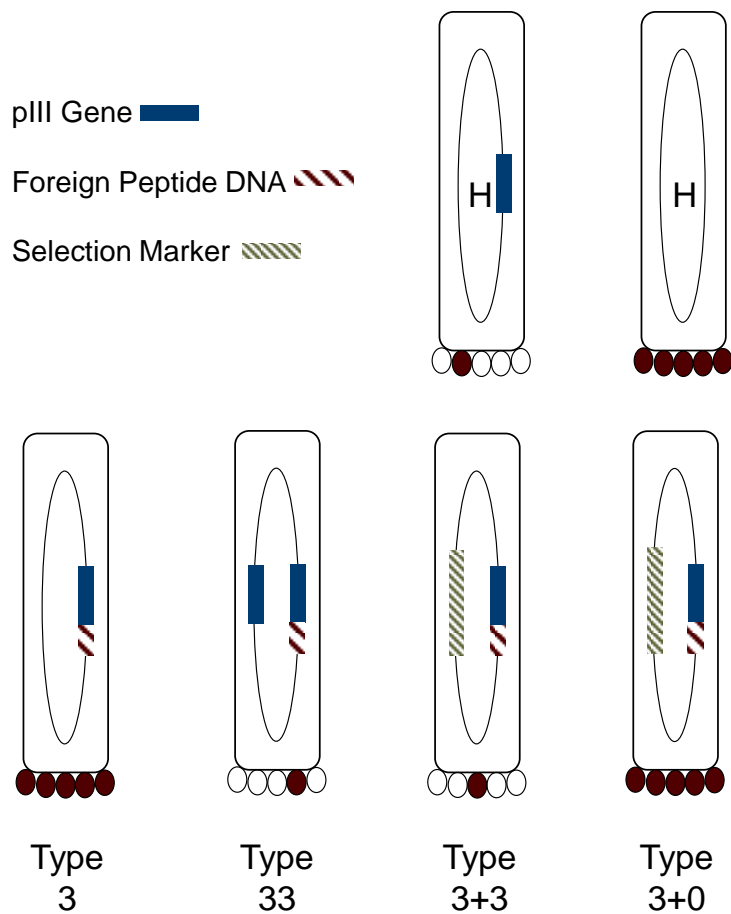
Various strategies have been used for expression of the peptide-coat fusion protein. Phage display in which the peptide-coat fusion is expressed directly from the phage genome, as the only source of the coat protein, is referred to as Type 3 display<sup>\*</sup>. In Type 3 display, a homogenous phage virion is produced in which every copy of pIII is displaying the foreign peptide (Figure 1.6). As mentioned previously, homogenous display of large peptides can negatively impact phage infectivity. To overcome this problem, expression systems have been developed that allow for display on heterogeneous phages containing a mixture of wildtype and fusion coat proteins. In one such system, a wildtype copy of pIII is inserted into the phage genome along with the peptide-coat fusion protein. During phage assembly, the wildtype protein competes with the fusion protein for incorporation into the phage capsid. This type of system is referred to as Type 33 display and is amenable to the display of larger proteins that would otherwise be detrimental to the phage infectivity and stability.

Another system that allows for heterogeneous phage production is Type 3+3 display (Figure 1.6). In this system, the peptide-coat fusion protein is not expressed from the phage genome; rather, it is expressed from a type of plasmid called a phagemid.<sup>8</sup> The phagemid contains the gene encoding the peptide-coat fusion protein as well as an f1 origin that allows for its synthesis as ssDNA and packaging into 'phagemid particles' (PhPs). Further, phagemids contain a gene encoding antibiotic resistance as well as a

---

<sup>\*</sup> Here we will focus on peptide display using pIII, although the same principles apply for display on pVIII.





**Figure 1.6. Display systems for filamentous phage display.** The products of various display systems are shown. Long rectangles represent Ff bacteriophage virions. Internal ovals represent single-stranded DNA. White circles represent wildtype pIII whereas maroon circles represent peptide-pIII fusion proteins. Type 3+3 and Type 3+0 are phagemid display systems that require helper phages which are labelled “H”.

bacterial origin of replication allowing for the added benefit of selection and propagation as a plasmid.<sup>62</sup> Since the phagemid encodes just one coat protein, cells harboring only the phagemid do not have the necessary machinery to package PhPs. To package the phagemid into PhPs the cell must first be infected with a special type of phage called a helper phage. The helper phage has a complete phage genome with all of the genes needed for replication and assembly of progeny virions. Proteins expressed

from the helper phage synthesize ssDNA of the phagemid and package it into PhPs that can be purified and used for affinity selection. Since the helper phage contains genes encoding all five coat proteins, the PhP will contain a mixture of wildtype and peptide-fusion coat proteins. Further, since the helper phage contains a packing sequence and f1 origin it will compete with the phagemid for packaging into progeny virions. For this reason helper phages with defects in their packing sequence or origin of replication have been constructed and these allow for the preferential packaging of the phagemid over the helper phage.<sup>63</sup>

Phagemid display offers several advantages over conventional phage display.<sup>62</sup> For one, phagemids are smaller than phages. This allows for larger DNA fragments to be inserted and greater transformation efficiency when introducing the cloned library into the host cell. The latter is a significant advantage since transformation efficiency is often the limiting factor for library size in phage display.<sup>64</sup> The selection marker, bacterial replication origin, and presence of multiple restriction sites allow for facile cloning of peptide display libraries on phagemids. Finally, the inducible promoter of phagemids allows for the expression level of the peptide-coat fusion to be tightly regulated and modulated.<sup>62</sup> For these reasons it may also be desirable to use phagemids for homogenous peptide display. In the fourth type of display, the peptide-coat fusion protein is expressed from a phagemid in conjunction with a helper phage that has a deletion of the pIII gene (Figure 1.6).<sup>65, 66</sup> Dubbed Type 3+0 display, this system produces phagemid particles in which every copy of the coat protein contains a peptide fusion, similar to Type 3 display, but with the added benefits of using phagemids.

## **1.5 Concluding Remarks**

Since its discovery in 1985, phage display has emerged as a powerful tool for the directed evolution of peptides with novel properties. The remarkable tolerance of Ff bacteriophages to structural modifications, coupled with their resistance to harsh chemical conditions, has made them the choice phage for peptide display. The focus of this thesis is to expand the chemical repertoire of peptides displayed on Ff bacteriophages—in particular N-terminal display on pIII (Type 3+0) using the M13 bacteriophage. We will review some methods that have been used for this purpose in the next chapter.

## CHAPTER II

### EXPANDING THE SCOPE OF FILAMENTOUS PHAGE DISPLAY

#### 2.1 Expanding the Chemical Diversity of Phage-Displayed Peptides

With few exceptions, Nature has limited the number of proteinogenic amino acids to 20. Although these include relatively diverse chemical functional groups, they are not sufficient to account for the myriad of structures and functions found within proteins in most organisms. For this reason many proteins undergo enzyme-catalyzed covalent modification following translation.<sup>67</sup> These posttranslational modifications (PTMs) have evolved as a means to expand the chemical diversity of the proteome beyond that which can be accomplished via direct genetic encoding, thereby imparting new function onto many proteins.

Being that they are ribosomally synthesized, phage display libraries are limited in terms of their chemical diversity to the functional groups found within the 20 canonical amino acids. However, just as PTMs have allowed for the evolution of proteins with expanded function in Nature, increasing the chemical diversity of phage display libraries would allow for the directed evolution of peptides with novel properties. As a general rule, increasing the chemical and structural diversity of a library increases the likelihood of identifying high affinity hits from selection.<sup>68</sup> On top of this, the ability to install non-canonical functional groups (ncFG) into peptide libraries could allow for the evolution of peptide ligands that are specific to the target protein. For example, metal chelating groups could be introduced into libraries targeted towards metalloenzymes. Reactive

functional groups could be installed to evolve covalent peptide inhibitors containing chemical warheads. Libraries containing naturally occurring posttranslational modifications could be generated and used to target enzymes that interact with these modifications. Numerous other target-specific applications could be easily envisaged.

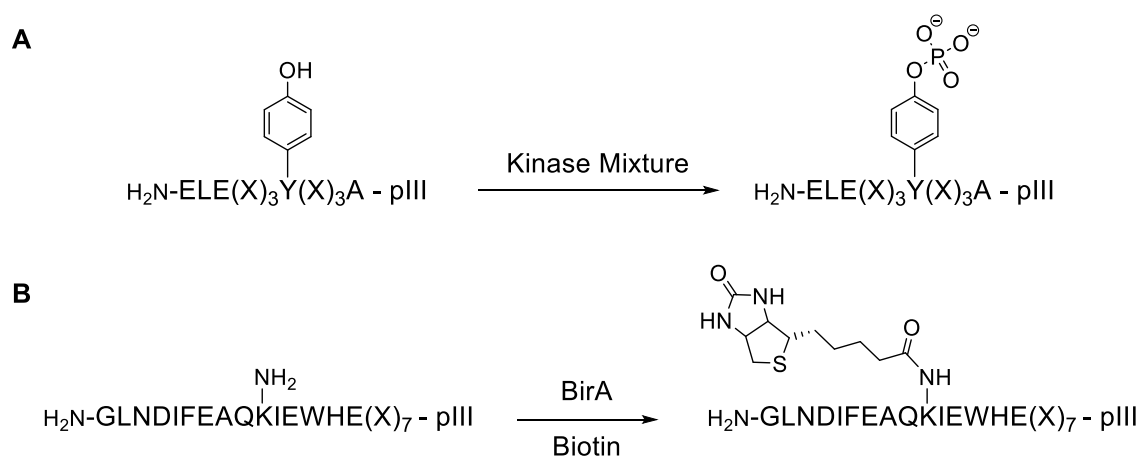
To this end, several methods for introducing ncFGs into phage display libraries have been developed. Since bacteriophages have recently risen in popularity for their use in materials science, new methods for the selective covalent modification of phage coat proteins are continually being explored.<sup>69</sup> Here we will review a few of these methods with a focus on those that have been applied for display on filamentous bacteriophages.

### *2.1.1 Enzymatic Posttranslational Modification*

One of the first methods developed for installing ncFGs employed recombinant enzymes to install natural PTMs on phage-displayed peptides. After enzymatic modification of the displayed peptides, the peptide library containing ncFGs can be used for *in vitro* selection. For example, Baumann et al. generated a peptide library containing phosphotyrosine (pTyr) in this way.<sup>70</sup> The library of peptides containing pTyr was used to identify ligands for the pTyr-binding protein src-homology 2 (SH2). They first constructed a peptide library containing octamers of the form (Xxx)<sub>3</sub>-Tyr-(Xxx)<sub>4</sub> (where Xxx is any amino acid). The library was cloned as a genetic fusion to the N-terminus of pIII on the M13 bacteriophage. After expression of the peptide library, purified phages were incubated with a mixture of promiscuous pTyr kinases to install phosphate onto the fixed tyrosine (Figure 2.1A). The modified libraries were then panned against SH2 to

reveal peptide ligands that bound to SH2 with an affinity three-fold greater than the natural ligand. More recently, a similar strategy has also been used to install pTyr onto peptides displayed on pVIII.<sup>71</sup>

Many enzymes that install PTMs, however, are not as promiscuous and only install modifications at residues located in highly specific peptide sequences. These enzymes are less useful for phage display where it is desirable to insert modifications into the random sequence of a peptide library.<sup>69</sup> Less promiscuous enzymes can, however, install PTMs on residues adjacent to the random sequence which is useful for some applications. For example, Kay et al. used the sequence-specific biotin ligase BirA to install biotin at a position directly upstream of a randomized peptide library on M13 (Figure 2.1B).<sup>72</sup> After enzymatic modification, the biotinylated library was panned



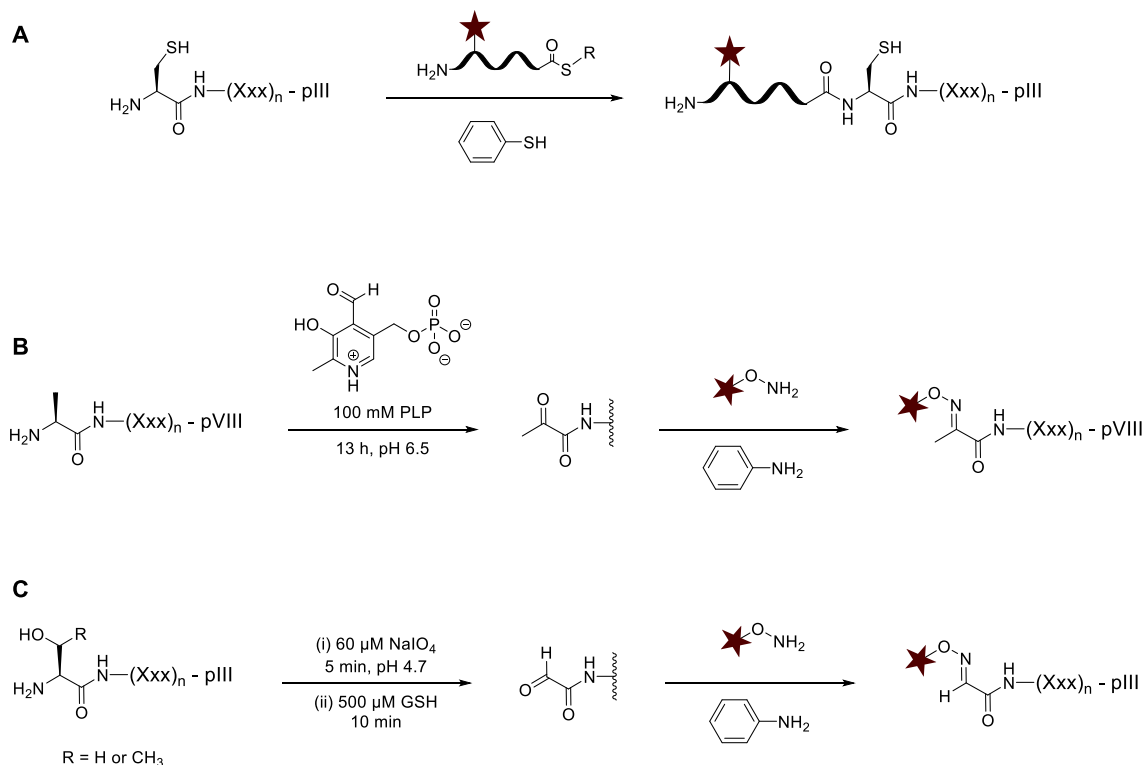
**Figure 2.1 Enzymatic modifications of phage-displayed peptides.** (A) Installation of pTyr into phage-displayed peptides using a kinase mixture. (B) Installation of biotin onto phage displayed peptides using the biotin ligase BirA. “X” represents any amino acid.

against several proteases to profile their substrate specificity. In their experiments, the biotinylated library was first immobilized via capture with streptavidin coated beads and then incubated with the various proteases. If the randomized sequence contained the protease recognition site, cleavage at this site released the phage from the solid support. Freed phages were then collected and sequenced to determine the peptide sequences cleaved by the protease.

Although these examples demonstrate the utility of enzymes for modifying phage display libraries, this technique is highly limited. As mentioned, most enzymes only install modifications at specific recognition sequences which limits their use for installing modifications in random libraries. The technique is limited further by the lack of available enzymes for installing many types of modifications. For these reasons, enzymatic modification has seen little use for phage display.

### *2.1.2 Native Chemical Ligation*

Native chemical ligation has been used to install protein fragments containing ncFGs onto both pIII and pVIII of the M13 bacteriophage. Kossiakoff et al. demonstrated that pIII and pVIII can both be efficiently expressed with N-terminal cysteine residues in *E. coli*.<sup>73</sup> Once incorporated into the viral capsid, these proteins can be modified with synthetic peptides containing C-terminal thioesters using standard native chemical ligation protocols (Figure 2.2A). While this method has been used to generate phage-displayed libraries containing ncFGs for selection, it only allows for the installation of ncFGs at positions adjacent to the genetically encoded randomized sequence.



**Figure 2.2. Chemical methods for introducing ncFGs into peptides displayed on Ff bacteriophages.** (A) Native chemical ligation can introduce synthetic peptide fragments containing ncFGs onto phage-displayed peptides with N-terminal cysteines. (B) Pyridoxal phosphate (PLP) can be used to introduce a ketone at the N-terminus of pVIII. (C) Sodium periodate can be used to introduce an aldehyde at the N-terminus of pIII. Ketone and aldehyde handles can be readily modified with aminoxy derivatives. “Xxx” represents any amino acid.

### 2.1.3 Chemical Posttranslational Modification

Many reactions have been developed, and are routinely used, to install ncFGs onto the sidechains of natural amino acids in proteins.<sup>74</sup> These reactions typically target and modify all of a particular functional group, e.g. *N*-hydroxysuccinimide esters react with the  $\epsilon$ -amine of Lys residues, iodoacetamide reacts with the thiol of Cys residues, etc. Such biocompatible reactions have been applied to the modification of Ff bacteriophages for various purposes including fluorophore conjugation for cell imaging and drug



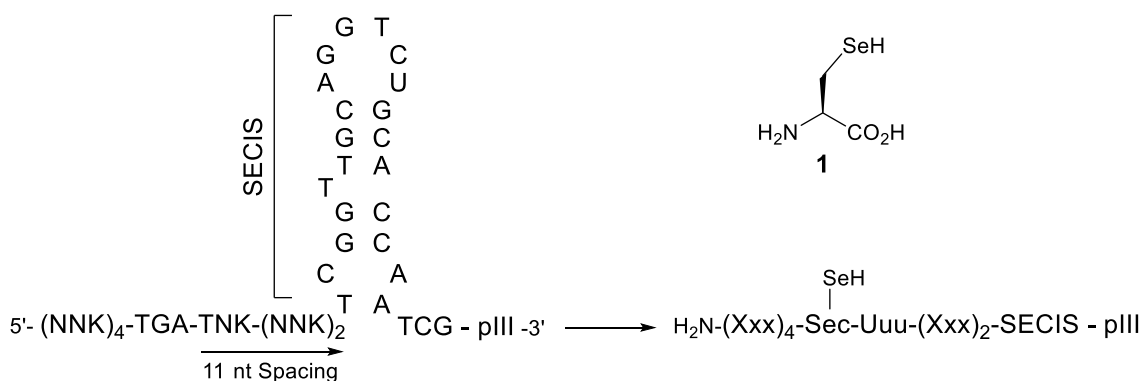
conjugation for cell-specific drug delivery.<sup>75, 76</sup> However, in terms of directed evolution via peptide display, it is desirable to install ncFGs at specific residues in a site-selective manner. Few reactions are available for mediating such transformations.<sup>77</sup>

Efforts to install ncFGs onto phage-displayed peptides in a site-selective manner have targeted their focus towards the N-terminal amine and N-terminal residues. For example, Francis et al. were able to selectively install a ketone at the N-terminus of peptides displayed on pVIII on a derivative of the M13 bacteriophage.<sup>78</sup> The installed ketone can undergo subsequent reaction with aminoxy-derivatives for the site-selective installation of various functional groups. The ketone was installed via transamination with pyridoxal 5'-phosphate. At pH 6.5 the reaction was specific for only N-terminal amines and went to 80% completion in 16 hours (Figure 2.2B). In a similar way, Derda and coworkers installed an aldehyde at the N-terminus of a heptapeptide library displayed on the N-terminus of pIII in M13.<sup>79</sup> The aldehyde was installed via oxidative cleavage of N-terminal serine and threonine residues using periodate (Figure 2.2C). Following installation of the aldehyde, treatment of the library with aminoxy-mannose allowed for the formation of a glycopeptide phage display library with  $1 \times 10^8$  unique glycopeptides. Notably, the aldehyde reaction was significantly faster than the ketone going to completion in under 1.5 hours. However, both methods required intermediate phage purification to remove high concentrations of the aniline catalyst.

#### *2.1.4 Phage Display with an Expanded Genetic Code*

As an alternative to enzymatic and chemical posttranslational modification, ncFGs can be introduced into phage coat proteins directly with the use of an expanded genetic code. In this technique amino acids bearing unique functional groups are incorporated co-translationally in response to a redefined codon during normal ribosomal peptide synthesis. A major benefit of this technique is that it allows for the site selective installation of modified residues without concern for nonspecific reaction byproducts or the need for subsequent purification to remove excess reagent. Further, it allows for the installation of amino acids with functional groups that are difficult or impossible to obtain with other techniques due to the lack of enzymes or selective chemical means for their installation.

The first example of phage display with an expanded genetic code came in 2000 when Sandman et al. demonstrated the genetic incorporation of selenocysteine (**1**, Sec) into a peptide library displayed on the N-terminus of pIII in M13.<sup>80</sup> Sec is the 21<sup>st</sup> naturally occurring amino acid that is incorporated in response to TGA codons in some organisms. Its incorporation requires a cognate tRNA ( $\text{tRNA}_{\text{UGA}}^{\text{Sec}}$ ), a specialized elongation factor, two enzymes for Sec biosynthesis, as well as a specific mRNA hairpin structure downstream of the TGA codon.<sup>81</sup> To incorporate Sec into M13, the authors introduced a fragment of DNA of the form  $(\text{NNK})_4\text{-TGA-TNK-(NNK)}_2$  (where N = A, T, G, or C, and K = G or T), encoding a partially randomized octamer, onto the N-terminus of pIII. They also introduced the hairpin element that is required for Sec



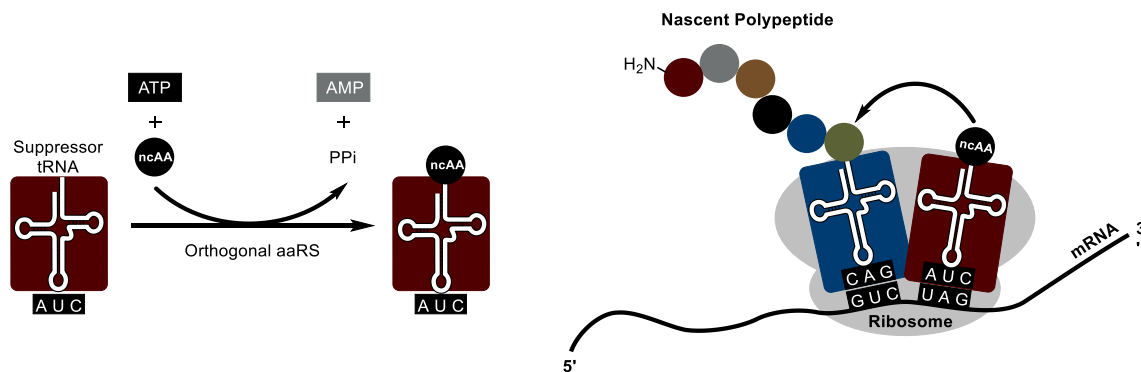
**Figure 2.3. Genetic incorporation of selenocysteine into peptides displayed on the N-terminus of pIII.** The required genomic elements for selenocysteine incorporation, including the 11 nucleotide spacing and selenocysteine incorporation sequence (SECIS), are shown. Selenocysteine (**1**, Sec) is incorporated at TGA codons in *E. coli* that are concurrently expressing 1) tRNA<sup>Sec</sup><sub>UGA</sub>, 2) two enzymes needed for Sec biosynthesis on tRNA<sup>Sec</sup><sub>UGA</sub>, and 3) a specialized elongation factor.

incorporation immediately after the last downstream codon (Figure 2.3). Though TGA codons usually signal translation termination, low level Trp incorporation at TGA codons has been observed in *E. coli*. It has been shown that the degree to which Trp suppresses TGA codons is dependent on the identity of the adjacent downstream nucleotide.<sup>82</sup> Therefore, to minimize Trp incorporation, the nucleotide immediately downstream was fixed to T. Expressing this phage library in *E. coli* that were concurrently expressing the Sec incorporation machinery in media supplemented with selenium resulted in the genetic incorporation of Sec at TGA codons in peptides displayed on pIII. Notably, the lower pKa of Sec (pKa = 5.47) compared to Cys (pKa = 8.18) allowed for the selective modification of Sec with biotin-conjugated iodoacetamide at low pH.<sup>80</sup>

Although useful as a means for the installation of a selective nucleophile in phages, in a subsequent study the same group also showed that phages containing the

TAGT sequence were prone to mutation at the 3' T upon propagation. It was further demonstrated that mutations at the 3' T resulted in the loss of selenium-dependent phage expression indicating an increase in TGA suppression by Trp.<sup>83</sup> Given that phage display requires iterative rounds of selection and amplification, such a susceptibility to mutation severely limits the utility of Sec incorporation. It is worth noting, however, that recent advances in next-generation sequencing have allowed for the identification of peptide ligands after only one round of biopanning.<sup>84</sup> These modern sequencing techniques may bring about renewed interest in Sec phage display. In any case, the method is still limited by the need for a downstream mRNA hairpin which constrains the Sec residue to a fixed position within the library.

Schultz and coworkers have developed a general strategy for the genetic incorporation of non-canonical amino acids (ncAAs) based on the suppression of nonsense codons.<sup>85</sup> In this technique an aminoacyl-tRNA synthetase (aaRS) and its cognate suppressor tRNA, (both from an exogenous source), are co-expressed along with the protein of interest containing an in-frame nonsense mutation. The ncAA is then incorporated at the nonsense mutation during translation by the suppressor tRNA which has been aminoacylated with the ncAA by the aaRS (Figure 2.4). This strategy for ncAA incorporation requires that the exogenous aaRS/tRNA pair be orthogonal to the host system. That is, the aaRS does not aminoacylate endogenous tRNAs and the suppressor tRNA does not serve as a substrate for endogenous aaRSs. One widely used aaRS for this purpose is the tyrosyl-tRNA synthetase from *Methanocaldococcus jannaschii* (Mj-



**Figure 2.4. The genetic incorporation of ncAAs via nonsense suppression.** (A) An orthogonal suppressor tRNA is selectively aminoacylated with the ncAA by its cognate orthogonal aaRS. (B) The suppressor tRNA incorporates the ncAA in response to nonsense codons during ribosomal peptide synthesis.

TyrRS) whose cognate tRNA ( $Mj$ -tRNA<sup>Tyr</sup><sub>CUA</sub>) has been engineered for amber (TAG) suppression.<sup>86</sup> Another commonly used aaRS is the pyrrolysyl-tRNA synthetase (PylRS) (from archaea of the genus *Methanocarcina*) whose cognate tRNA (tRNA<sup>Pyl</sup><sub>CUA</sub>) is a naturally occurring amber suppressor.<sup>87-89</sup> These archaeal aaRS/tRNA pairs are both orthogonal to the *E. coli* translation system and the PylRS/tRNA<sup>Pyl</sup><sub>CUA</sub> is also orthogonal in eukaryotes.<sup>90, 91</sup> Along with this method for ncAA incorporation, Schultz and coworkers have also developed powerful methods for the directed evolution of aaRSs with expanded or redefined substrate spectrums. Nonsense suppression coupled with aaRS evolution has allowed for the genetic incorporation of over 150 different ncAAs into proteins expressed in bacteria, eukaryotes, and even entire organisms.<sup>92, 93</sup>

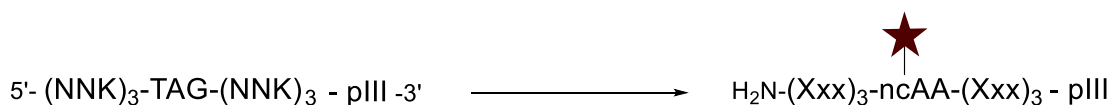
Nonsense suppression has been applied for the genetic incorporation of ncAAs into peptides and proteins displayed on the N-terminus of pIII in M13. In 2004, Schultz and colleagues used evolved *Mj*-TyrRSs in conjunction with *Mj*-tRNA<sup>Tyr</sup><sub>CUA</sub>, to introduce

five different ncAAs into peptides displayed on Ff bacteriophages<sup>94</sup> (Figure 2.5A). In their study, a short peptide containing a fixed TAG codon was introduced to the N-terminus of pIII on a derivative of the M13 bacteriophage. Expressing phages containing the TAG codon in *E. coli* that were concurrently expressing evolved *Mj*-TyrRS and *Mj*-tRNA<sub>CUA</sub><sup>Tyr</sup>, and in media supplemented with an ncAA, yielded 100- to 1000-fold more virions compared to when the media was not supplemented with an ncAA. This difference in expression level indicates successful suppression of the nonsense codon by the ncAA. Further, *para*-azido-phenylalanine (**3**) bears a reactive azide group that can undergo copper-catalyzed 1,3-dipolar cycloaddition with alkynes. Phages containing **3** were able to be selectively labelled with an alkyne-fluorescein conjugate providing further evidence for successful TAG suppression.

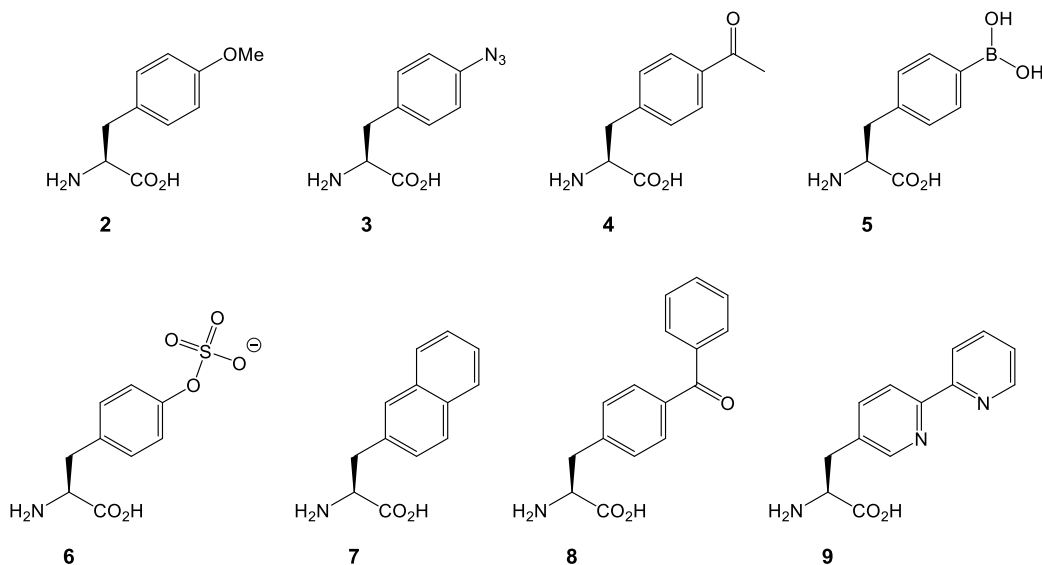
In later studies the same group utilized nonsense suppression to incorporate several different ncAAs into antibody fragments displayed on the N-terminus of pIII (Figure 2.5B). In one study, phage-displayed antibody libraries containing *para*-sulfo-tyrosine (**6**) were panned against the HIV coat protein gp120. This selection yielded antibodies containing **6** that bound to the antigen with significantly higher affinity than a naturally occurring anti-gp120 antibody.<sup>95</sup> In a subsequent study they demonstrated that two *para*-sulfo-tyrosine residues could be incorporated into phage-displayed antibodies, at fixed positions, and these libraries were also panned against gp120.<sup>96</sup> Nonsense suppression has been further applied for the selection of antibodies containing *para*-borono-phenylalanine (**5**) that react and bind with acyclic glycans<sup>97</sup> and for the selection of peptides containing the metal chelating amino acid bipyridyl alanine (**9**) that bind to

iron (II)<sup>98</sup> and nickel (II)<sup>99</sup>. In total, eight different ncAAs have been hitherto incorporated into phage display libraries (Figure 2.5B). In principle, this could readily be expanded to any of the ncAAs previously incorporated via nonsense suppression. A notable advantage of this technique is that it allows for site-selective installation anywhere in the peptide sequence without the need for enzyme recognition sequences or RNA elements.

**A**



**B**



**Figure 2.5. Genetic incorporation of ncAAs into peptides displayed on the N-terminus of pIII.** (A) ncAAs are incorporated at TAG codons in *E. coli* by an orthogonal aaRS in conjunction with an orthogonal suppressor tRNA. (B) The structures of all ncAAs incorporated into Ff bacteriophages to date.

Despite the apparent benefits, nonsense suppression has seen little use in phage display. A major drawback of this technique arises from the strong expression bias against clones that contain ncAAs in favor of those containing only canonical amino acids. Such systematic biases significantly impede selection efforts.<sup>95</sup> This problem will be addressed in detail in Chapter 4.

## **2.2 Phage Display of Cyclic Peptides**

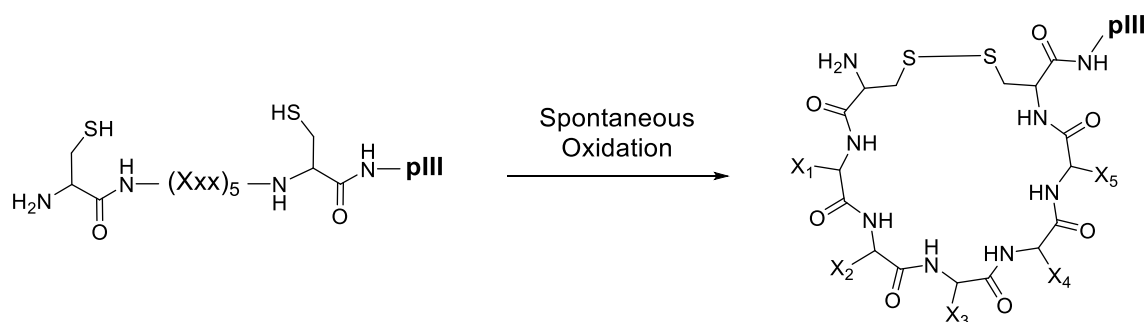
Conformationally constrained cyclic peptides offer several advantages over linear peptides in terms of the use for research or as therapeutic agents. For one, a major limitation of linear peptides is their high susceptibility to proteolytic degradation.<sup>100</sup> However cyclic peptides, due to their more rigid structure and protected terminal residues, are significantly more resistant to proteolysis<sup>101, 102</sup> Further, cyclization limits the number of conformations that peptides adopt in solution. This conformational constraint means that there is less of an entropic cost upon binding to the target protein which manifests as an increase in binding affinity.<sup>103</sup> Therefore, evolving peptides from cyclic peptide libraries will likely afford ligands with higher binding affinities and greater protease resistance. To this end, two general strategies have been developed for cyclizing peptides that are displayed on bacteriophages.

### *2.2.1 Genetically Encoded Cyclic Peptides*

The first examples of phage-displayed cyclic peptide libraries utilized genetically encoded cyclic peptides. In these types of libraries, the randomized peptide sequenced is



flanked on both sides by fixed cysteine residues. Upon spontaneous oxidation, the cysteines form a disulfide bond that cyclizes the peptide with a sidechain-to-sidechain linkage (Figure 2.6). DeGrado et al. first demonstrated the utility of disulfide constrained phage display libraries in 1992.<sup>104</sup> In their study, a randomized hexapeptide flanked by cysteine residues was introduced onto the N-terminus of pIII and panned against the platelet glycoprotein IIb/IIIa (gpIIb/IIIa). Peptides isolated from this selection were shown to be potent inhibitors of gpIIb/IIIa with IC<sub>50</sub> values in the low nanomolar. Using a reducing agent to break the disulfide bond and generate the linear peptide resulted in up to a 20-fold increase in IC<sub>50</sub> values indicating that the potency of the peptides was derived, in part, from their cyclic structure.



**Figure 2.6. Disulfide constrained phage-displayed peptide libraries.**

Shortly after the report by DeGrado et al. the crystal structures of two peptides isolated from disulfide-constrained phage display libraries were solved in complex with their target protein (streptavidin).<sup>105</sup> These structures revealed that the cysteines in both peptides were not in direct contact with streptavidin. This supported the hypothesis that

the observed increased binding affinity of the peptides was a result of entropic effects and not due to direct interactions between streptavidin and the Cys residues themselves. Further, these observations implied that disulfide constraint could be used as a general strategy for evolving peptide ligands with improved binding affinities.

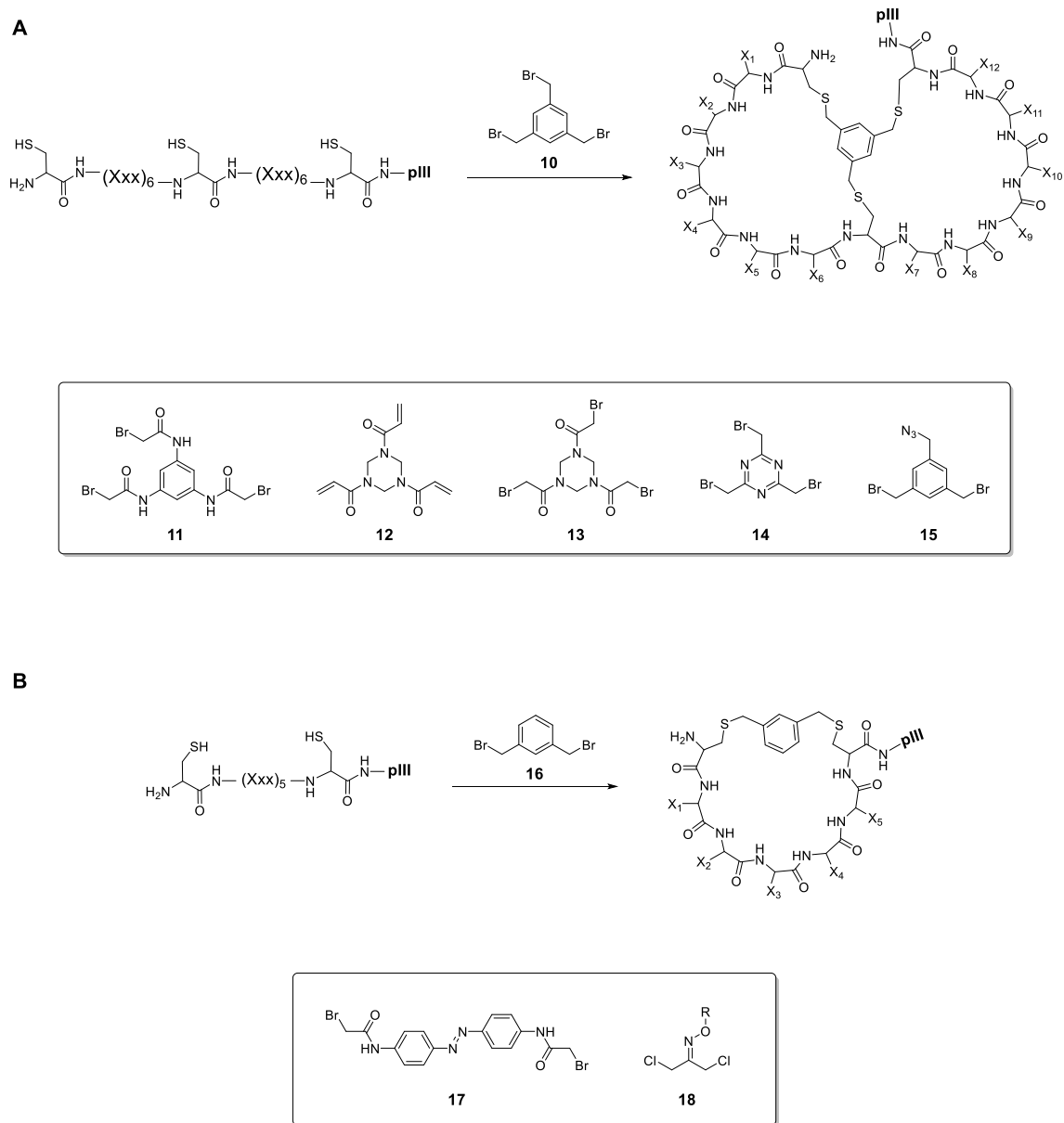
Since the first report by DeGrado et al. disulfide constraint has become a widely-used strategy for the identification of cyclic peptide ligands via phage display. Often selections with disulfide-constrained libraries afford peptides with affinities significantly higher than those obtained with linear peptide libraries.<sup>106</sup> However, a limitation of this technique lies within the labile disulfide bond on which it relies. Disulfide bonds are easily broken in the presence of a reducing agent. Therefore, disulfide-constrained peptides are not good ligands for protein targets found with the reducing cytosol of cells or for targets that require a reducing agent for stability or activity.

### *2.2.2 Chemically Cyclized Peptides*

As an alternative to disulfide-constrained peptide libraries, linear peptides displayed on bacteriophages can be chemically modified to afford cyclic peptides. In contrast to the labile disulfide bond of genetically-encoded cyclic libraries, posttranslational chemical modification affords cyclic peptides with linkages that are significantly more stable. The first example of chemically-cyclized phage display libraries was published in 2009.<sup>107</sup> In this study, Heinis et al. introduced a peptide with twelve randomized positions and three fixed cysteines of the form Cys-(Xxx)<sub>6</sub>-Cys-(Xxx)<sub>6</sub>-Cys onto the N-terminus of pIII of the fd bacteriophage. After reduction of spontaneously formed disulfide bonds, the

displayed peptides were treated with 1,3,5-tris(bromomethyl)benzene (**10**) to afford bicyclic peptides containing two macrocyclic rings. (Figure 2.7A). These phage-displayed bicyclic libraries were screened against several different human proteases to afford bicyclic peptides that inhibited the proteases with IC<sub>50</sub> values as low as 100 nM. Based on the consensus sequence identified from the selection against the protease human plasma kallikrein (PK), a second library was constructed and screened for affinity maturation. This second selection yielded potent inhibitors of PK with IC<sub>50</sub> values as low as 1.7 nM. Notably, the IC<sub>50</sub> of the peptide prior to modification with **10** was greater than 10 μM.

Following this initial report, several other compounds have been used for chemically cyclizing phage-displayed peptides. Heinis et al. have subsequently demonstrated that derivatives of **10** containing hydrogen bond acceptors and donors could be used to generate bicyclic peptide libraries. X-ray structures of these cyclic peptides in complex with their target protein have revealed that amino acids in the peptide can hydrogen bond with the cyclizing compound. These hydrogen bonds give rise to extremely compact and rigid peptides with exquisite binding affinity.<sup>108</sup> Other compounds have also been reported with additional applications. For example, compounds containing azobenzene cores, such as **17**, have been used to cyclize phage-displayed peptides bearing two fixed cysteines (Figure 2.7B).<sup>109-111</sup> These azobenzene compounds undergo photoinduced *trans-cis* isomerization which has allowed for the selection of photoswitchable peptide ligands. Displayed peptides containing Cys residues at the *i* and *i* + 4 positions can be cyclized with  $\alpha,\alpha'$ -dibromo-*m*-xylene (**16**)



**Figure 2.7. Chemical cyclization of phage-displayed peptides.** (A) The formation of bicyclic peptides by the reaction of a displayed peptide containing three cysteines with various tris-electrophiles. (B) The formation of monocyclic peptides by the reaction of a displayed peptide containing two cysteines with bis-electrophiles.

allowing for ligand evolution from libraries containing stapled  $\alpha$ -helical peptides.<sup>112</sup> Derda et al. have shown that peptide libraries bearing two Cys can be cyclized with various dichloro-oxime derivatives such as **18**. They have used this for the construction and selection of macrocyclic glycopeptide libraries.<sup>113</sup> The dichloro-oxime derivatives were formed *in situ* via oxime ligation between 1,3-dichloroacetone and various aminoxy derivatives with near quantitative conversion before being used to cyclize the phage peptides.

Recent advances in the ability to chemically cyclize peptides on bacteriophages have revolutionized peptide phage display. Peptide inhibitors with exquisite selectivity and dissociation constants in the picomolar range have been identified using chemically cyclized peptide libraries and several of these are nearing clinical trials.<sup>106</sup> However, a limitation of this technique arises from the lack of selectivity of the compounds used to cyclize peptides. All currently available methods utilize nucleophilic cysteine residues to react with a bis- or tris-electrophilic linker. Ideally, this reaction will take place only on the displayed peptide; however, wildtype Ff phages contain eight additional cysteine residues within pIII that can also react. Reduction or modification of these internal disulfides leads to loss of phage infectivity.<sup>114</sup> To circumvent this problem, several strategies have been employed including careful selective reduction of only the most sterically accessible cysteines<sup>109</sup> or the use of an engineered disulfide-free fd phage<sup>107</sup>. Selective reduction requires the use of expensive reagents and prolonged reaction times for efficient reduction of disulfide bonds, whereas engineered disulfide-free phages have a 100-fold reduced infectivity compared to the wildtype phage.<sup>114</sup> Further, the

concentration of the modifying compound must be carefully controlled as, for example, concentrations of **10** greater than 10  $\mu\text{M}$  lead to significant reduction in phage infectivity. This has been attributed to crosslinking of phage coat proteins when **10** nonspecifically reacts with lysine residues.<sup>107</sup> For these reasons, there is a strong need for compounds that can cyclize phage-displayed peptides with more regio- and chemoselectivity. This problem will be addressed further in Chapter 6.

### **2.3 Concluding Remarks**

The work presented in this thesis has two primary specific aims. In the first aim, we strive to expand the chemical diversity of phage-displayed peptides using the nonsense suppression technique. In Chapter 3, we profile the ncAA substrate specificity of a rationally designed pyrrolysyl-tRNA synthetase. In particular, we examine the ability of the enzyme to incorporate *ortho*-substituted phenylalanine derivatives. In Chapter 4, we develop and test a new method of constructing phage-displayed peptide libraries containing ncAAs. We further demonstrate that the enzyme described in Chapter 3 can be used to incorporate a variety of ncAAs into peptides displayed on pIII of the M13 bacteriophage. This work improves up previous studies involving phage display with an expanded genetic code by allowing for the selection from libraries containing ncAAs with varying incorporation efficiencies. In Chapter 5, we use the method developed in Chapter 4 to generate a library of peptides containing a specific posttranslational modification. This library is then used to identify potent inhibitors of an enzyme that removes this modification in Nature.

The second aim of this work is to develop a more regiospecific compound for cyclizing peptides displayed on the M13 bacteriophage. In Chapter 6, we describe the design and synthesis of a kinetically controlled bis-electrophile that can be used for this purpose. Further, we demonstrate that this compound reacts specifically with N-terminal cysteine residues and describe ongoing efforts to evolve cyclic peptide ligands using this compound.

**CHAPTER III**  
**GENETIC INCORPORATION OF SEVEN *ORTHO*-SUBSTITUTED**  
**PHENYLALANINE DERIVATIVES<sup>†</sup>**

### **3.1 Introduction**

The genetic incorporation of non-canonical amino acids (ncAAs) into recombinant proteins has provided researchers with powerful tools for the study of protein structure and function. The ability to genetically incorporate ncAAs with bioorthogonal functional groups, coupled with click chemistry, allows for the site-selective introduction of protein modifications that have been used for fluorescence labeling<sup>115-121</sup> and photo-cross-linking,<sup>122-126</sup> as well as for the study of posttranslational modifications that are traditionally difficult or impossible to obtain in bacterial expression systems.<sup>127-131</sup> Other analytic probes that have been site-selectively introduced using the genetic incorporation approach include infrared, Raman, NMR, and EPR spectroscopic probes and heavy atoms for X-ray crystallographic studies.<sup>132-134</sup> As described in the previous chapter, ncAAs are introduced into a protein through the suppression of an in frame nonsense or quadruplet codon by a suppressor tRNA coupled with an evolved aminoacyl-tRNA synthetase (aaRS). One widely used synthetase for this purpose is the tyrosyl-tRNA synthetase (*Mj*-TyrRS) derived from *Methanocaldococcus jannaschii*.<sup>86</sup> Indeed, directed evolution of *Mj*-TyrRS, coupled with a suppressor tRNA<sub>CUA</sub><sup>Tyr</sup>, for the suppression of TAG (amber) codons has been highly successful for the incorporation of a number of ncAAs

---

<sup>†</sup> Portions of this chapter are reprinted with permission from Tharp et al. *ACS Chem. Biol.*, **2014**, 9 (4), 884-890. Copyright 2014 American Chemical Society.



into proteins in *E. coli*. Despite the success of engineering the *Mj*-TyrRS-tRNA<sup>Tyr</sup><sub>CUA</sub> pair, its weakness lies in that it cannot be used to incorporate ncAAs into eukaryotic organisms owing to the fact that tRNA<sup>Tyr</sup><sub>CUA</sub> contains a recognition element that allows for its aminoacylation with canonical amino acids by the endogenous aaRSs in eukaryotic organisms.<sup>134, 135</sup> Alternatives to the *Mj*-TyrRS-tRNA<sup>Tyr</sup><sub>CUA</sub> pair include the *Ec*-TyrRS-tRNA<sup>Tyr</sup><sub>CUA</sub> pair, which was derived from *E. coli* tyrosyl-tRNA synthetase-tRNA<sup>Tyr</sup> pair, the *Ec*-LeuRS-tRNA<sup>Leu</sup><sub>CUA</sub> pair derived from *E. coli* leucyl-tRNA synthetase-tRNA<sup>Leu</sup> pair, and the pyrrolysyl-tRNA synthetase (PylRS)- tRNA<sup>Pyl</sup><sub>CUA</sub> pair, which naturally exists in certain methanogenic archaea and some bacteria. Both *Ec*-TyrRS-tRNA<sup>Tyr</sup><sub>CUA</sub> and *Ec*-LeuRS-tRNA<sup>Leu</sup><sub>CUA</sub> pairs can be applied in eukaryotic cells but not in bacterial systems.<sup>124, 136, 137</sup> However, tRNA<sup>Pyl</sup><sub>CUA</sub> is a naturally occurring amber suppressor tRNA that is specifically recognized by PylRS but does not cross-interact with endogenous aaRSs in both bacterial and eukaryotic cells.<sup>138-142</sup>

To date, the PylRS-tRNA<sup>Pyl</sup><sub>CUA</sub> pair has been successfully used for the genetic incorporation of ncAAs into recombinant proteins expressed in bacteria and a variety of eukaryotes including mammalian cells and multicellular organisms.<sup>143, 144</sup> While the ability of the PylRS-tRNA<sup>Pyl</sup><sub>CUA</sub> pair to be used in both bacterial and eukaryotic cells is advantageous in that it allows for the same ncAA to be used in different model systems, the method still has limitations. The incorporation of a new ncAA typically requires selection of large libraries of aaRS with randomized active site mutations. Schultz and co-workers have developed a powerful method for the directed evolution of aaRSs

involving both positive and negative selections based on the ability to suppress a nonsense mutation in the presence of the desired ncAA.<sup>145</sup> However, this process still requires multiple rounds of selection for each new substrate. Therefore, a single aaRS with high substrate promiscuity is ideal to facilitate rapid progress and widespread use of the genetic ncAA incorporation technique. We have previously reported the ability of a single, rationally designed PylRS mutant (PylRS(N346A/C348A)), coupled with tRNA<sub>CUA</sub><sup>Pyl</sup>, to genetically incorporate a wide variety of phenylalanine derivatives with large *para* and small *meta* substitutions at amber mutation sites in living cells.<sup>146-148</sup> More recent studies indicate that this mutant is also able to recognize more than 10 phenylalanine derivatives with large *meta* substitutions.<sup>149</sup> PylRS(N346A/C348A) displays high activity toward these ncAAs while remaining relatively orthogonal toward canonical amino acids. Here we reveal that the same enzyme is also able to mediate the genetic incorporation of seven *ortho*-substituted phenylalanine derivatives, which brings the total substrate availability of this enzyme close to 40 and leads to the conclusion that PylRS(N346A/C348A) is able to recognize phenylalanine derivatives with a substitution at any side-chain aromatic position as a substrate. To our knowledge, PylRS(N346A/C348A) is the only aaRS with such a high substrate promiscuity. We also demonstrate, for the first time, the ability of PylRS(N346A/C348A) to facilitate the genetic incorporation of phenylalanine derivatives in mammalian cells. Aside from *o*-nitro-phenylalanine,<sup>150</sup> which has been genetically encoded in *E. coli* using an evolved *Mj*-TyrRS-tRNA<sub>CUA</sub><sup>Tyr</sup> pair, this constitutes the first report of the genetic incorporation for all of these *ortho*-substituted phenylalanine derivatives.<sup>151</sup> One of the *ortho*-substituted

phenylalanine derivatives, *ortho*-cyano-phenylalanine is a small, environmentally sensitive, fluorescent probe that can be used as a sensor to probe the local environment of proteins, and we applied this to monitor protein folding/unfolding.

## 3.2 Experimental Details

### 3.2.1 Miscellaneous Information

All primers were purchased from Integrated DNA Technologies, Inc. (Coralville, PA). Enzymes and buffers for molecular cloning were purchased from New England Biolabs (Ipswich, MA). All ncAAs were purchased from Chem-Impex International, Inc. (Wood Dale, IL) and used without further purification. DNA sequencing was performed by Eton Biosciences, Inc. (San Diego, CA). Mass spectra were collected by the Department of Chemistry Mass Spectrometry Facility at Texas A&M University.

### 3.2.2 Primer List

MPP8-*ScaI*-F: 5'-AGTCGAGCTCGGCGAAGACGTGTTCTGAAGTAGAGAAGATC-3'

MPP8-*Sall*-R: 5'-AGTCGTCGACTTATTTAGCTTTGTTCTCGGCAATCTTCTTGCGAAATTC-3'

MPP8-TAG-F: 5'-AGTCGAGCTCGGCGAAGACGTGTAGGAAGTAGAGAAGATC-3'

MPP8-TAG-R: 5'-AGTCGTCGACTTATTTAGCTTTGTTCTCGGCAATCTTCTTGCGAAATTC-3'

EGFR-TAG-F: 5'-TAGAAAACCGGACTGAAGGAGCTG-3'

EGFR-TAG-R: 5'-TGCATCATAGTTAGATAAAGACTGC-3'

PylRS-Ampl-F: 5'-CTAGTCTAGAATGGATAAAAAACCACTAAACA-3'

PylRS-Ampl-R: 5'-CTAGCTAGCTCACAGGTTGGTAGAAATCCCGTTAT-3'

### 3.2.3 Plasmid Construction

The construction of pET-PylT-sfGFP-S2(TAG), pBAD-sfGFPwt and mutants F27(TAG) and N135(TAG), and pEVOL-PylT-PylRS(N346A/C348A) has been described by us previously.<sup>152, 153</sup>

pBAD-MPP8-sfGFP: The codon optimized MPP8 chromo domain gene was synthesized by Epoch Life Science Inc. (Houston, TX) and was amplified using a 5'-end primer (MPP8-*ScaI*-F) containing an *ScaI* site and a 3'-end primer (MPP8-*Sall*-R) containing an *Sall* site. The PCR product was cloned into the *SacI* and *Sall* sites of a previously constructed pBAD-sfGFP to afford pBAD-MPP8-sfGFP. The MPP8-F59TAG containing construct for *ortho*-cyano-phenylalanine incorporation was generated by Phusion high-fidelity DNA polymerase-based site-directed mutagenesis with primers MPP8-TAG-F and MPP8-TAG-R to afford pBAD-MPP8-F59TAG-sfGFP.

pEGFP-N1-EGFR: The pEGFP-N1-EGFR plasmid was purchased from Addgene (Plasmid 32751). The EGFR-N128(TAG) construct was generated by Phusion high-fidelity DNA polymerase-based site-directed mutagenesis with the primers EGFR-TAG-F and EGFR-TAG-R.

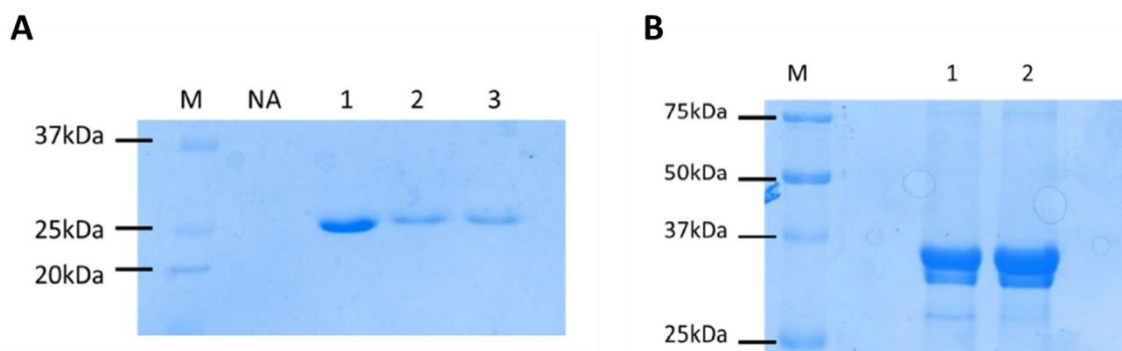
pCMV-U6-PylT-PylRS(N346A/C348A): PylRS(N346A/C348A) was amplified from the pEVOL construct using the primers PylRS-Ampl-F and PylRS-Ampl-R. The PCR product was cloned into *Xba*I and *Nhe*I site of pCMV- U6-PylT.

### 3.2.4 Protein Expression and Purification

sfGFP-S2X: *E. coli* BL21(DE3) containing the plasmids pET-PylT-sfGFP-S2(TAG) and pEVOL-PylT-PylRS(N346A/C348A) were grown in 500 mL LB media until the OD<sub>600</sub> reached 0.8 at which time the media was switched to a liquid glycerol minimal media (33.7 mM Na<sub>2</sub>HPO<sub>4</sub>, 22 mM KH<sub>2</sub>PO<sub>4</sub>, 8.6 mM NaCl, 9.4 mM NH<sub>4</sub>Cl, 1 mM MgSO<sub>4</sub>, 0.3 mM CaCl<sub>2</sub>, 1% glycerol, 0.3 mM leucine). To switch the media, cells were pelleted by centrifugation (4k rpm, 15 min), washed twice with phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5), and resuspended in the minimal media. All media was supplemented with 34 µg·mL<sup>-1</sup> chloramphenicol and 100 µg·mL<sup>-1</sup> ampicillin. Protein expression was induced by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and 0.2 % arabinose along with 2 mM of one of the ncAAs (**19-25**). After expression for 16 h, cells were harvested by centrifugation (4k rpm, 20 min), resuspended in lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated. The soluble fraction was recovered by centrifugation (10k rpm, 50 min) and the clear supernatant was incubated with 2 mL Ni<sup>2+</sup>-NTA resin with end-over-end rotation for 1 h at 4°C. The resin was then washed with 40 mL of the lysis buffer and the protein was eluted from the column with an elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH

8.0). The eluted protein was dialyzed against PBS and concentrated via ultrafiltration (10 kDa MWCO). The purified proteins were analyzed via 12% SDS-PAGE and electrospray ionization mass spectrometry.

*sfGFP-N135X*, *sfGFP-F27X*, and *MPP8-F59X-sfGFP*: *E. coli* Top10 were transformed with a pBAD plasmid containing the gene encoding sfGFP-N135X, sfGFP-F27X, or MPP8-F59X-sfGFP, along with pEVOL-PylT-PylRS(N346A/C348A). Transformed cells were grown in 200 mL of a synthetic auto-induction media (described elsewhere)<sup>154</sup> supplemented with 2 mM of ncAA **25**. To avoid background, when preparing the auto-induction media, phenylalanine was not included. All media was supplemented with 34  $\mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol and 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin. The culture was grown for 16 h at 37°C (250 rpm) and cells were harvested by centrifugation (4k rpm, 20 min). The proteins were then purified as described above. SDS-PAGE of the purified wildtype and mutant proteins is shown in Figure 3.1.



**Figure 3.1. SDS-PAGE analysis of sfGFP and MPP8.** (A) SDS-PAGE analysis of sfGFP-F27→**25** (1), sfGFP-N135→**25** (2), and wildtype sfGFP (3). NA indicates no ncAA was provided to the growth media. (B) SDS-PAGE analysis of MPP8-F59→**25** (1) and MPP8wt-sfGFP (2).

### *3.2.5 Mammalian Cell Cultures and Transfection*

HEK293T cells purchased from American Type Culture Collection (ATCC, Manassas, VA) were grown in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) containing 10% fetal bovine serum (Invitrogen). Cells were seeded and grown on a Becton Dickinson 24-well plate. At approximately 70-80% confluence, cells were transfected with two plasmids, pCMV-U6-PylT-PylRS(N346A/C348A) and pEGFP-N1-EGFR-N128(TAG). Transient transfection was carried out using Lipofectamine® 2000 Transfection Reagent from Invitrogen, according to the manufacturer's protocol. After transfection, the growth medium was DMEM with 10% FBS which contained ncAAs **19-25** or no unnatural amino acids. The cell images were recorded on an Olympus FluoView FV1000 Confocal Microscope after 48 hours and analyzed using ImageJ (NIH).

### *3.2.6 Fluorescence Measurements*

For all samples the fluorescence was measured using a Quanta Master 40 fluorometer (Photon Technology International) equipped with a Xenon lamp. Spectra were collected at room temperature using a 1 cm quartz sample holder, 0.5 mm slit width, 1 nm resolution, and 1 s·nm<sup>-1</sup> integration time. The samples were excited at 240 nm and the fluorescence was measured from 250-450 nm. For the protein folding measurements, stock solutions of the wildtype and mutant proteins were diluted to an appropriate volume in PBS, pH 7.5 and varying concentrations of a 10 M urea solution were added to give a final urea concentration ranging from 0 to 8. After the addition of the

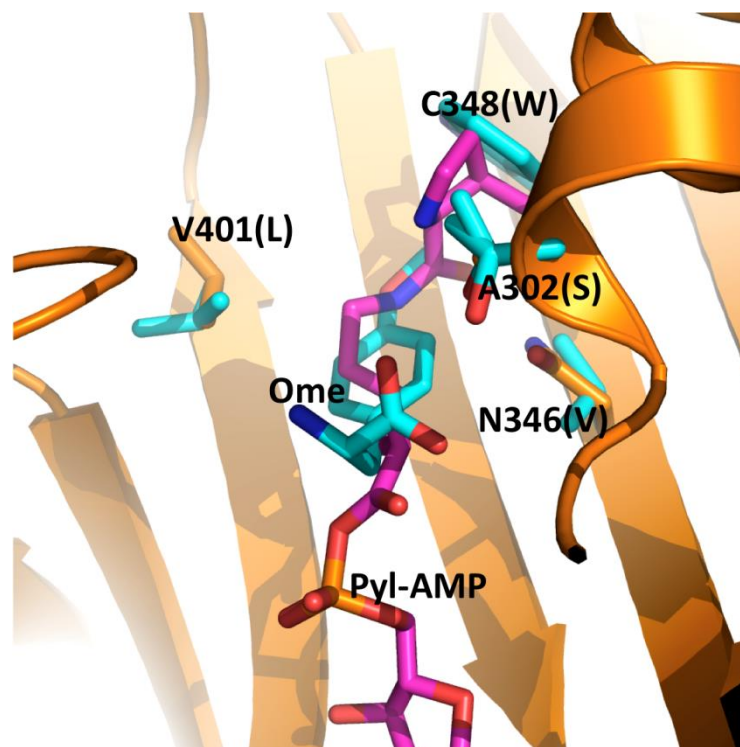
denaturant, 2 h were allotted for the transitions to occur before recording the fluorescence spectrum.

### 3.3 Results and Discussion

#### 3.3.1 PylRS(N346A/C348A) Recognizes Ortho-Substituted Phenylalanine as a Substrate

Wang and co-workers previously determined the crystal structure of a PylRS mutant (OmeRS) in complex with O-methyl-tyrosine.<sup>155</sup> Figure 3.2 presents the active site structure of this enzyme–substrate complex superimposed on the wild-type PylRS with pyrrolysyl-AMP bound. The two overlaid structures clearly show that the N346 side chain is adjacent to the *ortho* position of the O-methyl-tyrosine side chain phenyl group. Removing the N346 side chain would potentially leave space for adding an additional substitution to this position. Since PylRS(N346A/C348A) is expected to bind to a phenylalanine derivative in a similar fashion as OmeRS, we suspect that PylRS(N346A/C348A) may well recognize an *ortho*-substituted phenylalanine derivative and, together with tRNA<sub>CUA</sub><sup>Pyl</sup>, mediate its incorporation at an amber mutation site. We exploited an *E. coli* cell system harboring plasmids pEVOL-PylT-PylRS(N346A/C348A) and pET-PylT-sfGFP-S2(TAG) to assess the ability of the PylRS(N346A/C348A)-tRNA<sub>CUA</sub><sup>Pyl</sup> pair to incorporate *ortho*-substituted phenylalanine derivatives in response to an amber mutation. The plasmid pEVOL-PylT-PylRS(N346A/C348A) contains genes coding for PylRS(N346A/C348A) and tRNA<sub>CUA</sub><sup>Pyl</sup>, while pET-PylT-sfGFP-S2(TAG) contains genes coding for tRNA<sub>CUA</sub><sup>Pyl</sup> and superfolder

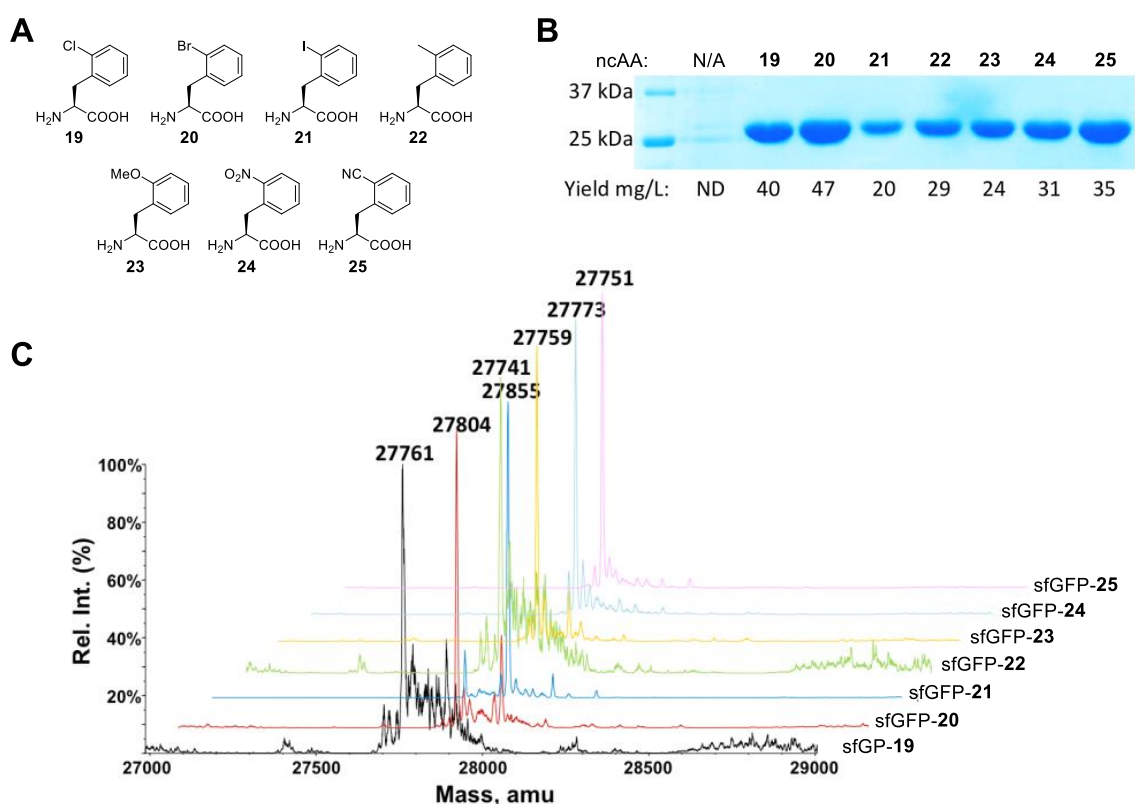




**Figure 3.2** The superimposed structures of OmeRS in complex with *O*-methyl-tyrosine and PylRS in complex with pyrrolysyl-adenosyl monophosphate (Pyl-AMP). The structures are based on PDB entries 2Q7H and 3QTC. The PylRS complex with Pyl-AMP is shown in orange for the protein carbon atoms and pink for the Pyl-AMP carbon atoms. Four mutated residues in OmeRS and the *para*-methoxy-phenylalanine ligand are shown in cyan for the carbon atoms. Letters in parenthesis indicate the four mutated residues in OmeRS.

green fluorescent protein (sfGFP) with an amber mutation at its S2 position and a 6 × His tag at the C-terminus. Growing cells in GMML medium (a minimal medium supplemented with 1% glycerol and 0.3 mM l-leucine) led to translation termination at the amber mutation site of sfGFP; therefore no full-length sfGFP was expressed. However, supplementing the media with 2 mM of one of seven *ortho*-substituted phenylalanine derivatives (**19–25**, shown in Figure 3.3A) promoted overexpression of full-length sfGFP (Figure 3.3B). Incorporation of the ncAAs was confirmed via electrospray ionization mass spectrometry (ESI-MS) (Figure 3.3C). For each of the

seven ncAAs, the detected molecular weight agreed well with the theoretical value (Table 3.1). Except for **24**, this is the first report of the genetic incorporation of these ncAAs in living cells, and for all of the ncAAs this is the first report of their incorporation using an aaRS derived from PylRS.

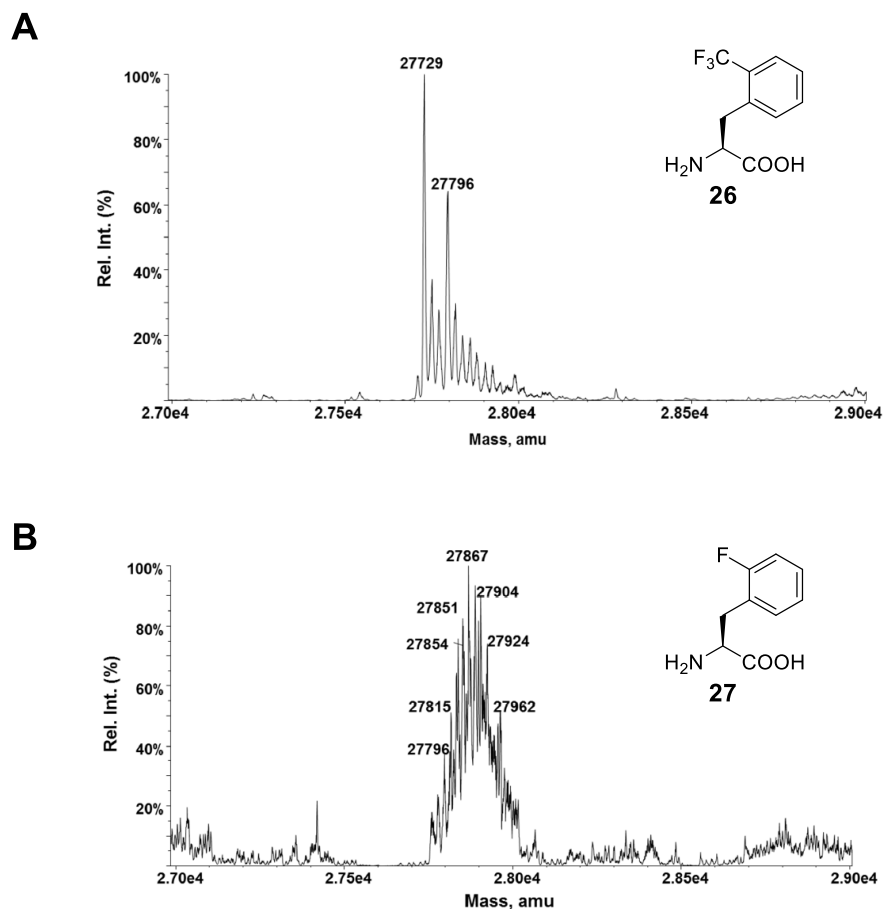


**Figure 3.3. Genetic incorporation of seven *ortho*-substituted phenylalanine derivatives.** (A) Structures of **19-25**. (B) Site-specific incorporation of **19-25** into sfGFP at its S2 site. N/A indicates no ncAA was available in the medium. ND represents non-detected. (C) Deconvoluted ESI mass spectra of sfGFP variants incorporated with **19-25**. sfGFP-X in which X is one of **19-25** and represents the specific sfGFP variant with X incorporated at its S2 site.

ncAA	Calculated Mass/Da	Detected Mass/Da <sup>a</sup>
<b>19</b>	<b>27764</b>	<b>27761</b>
<b>20</b>	<b>27808</b>	<b>27804</b>
<b>21</b>	<b>27855</b>	<b>27855</b>
<b>22</b>	<b>27743</b>	<b>27741</b>
<b>23</b>	<b>27759</b>	<b>27759</b>
<b>24</b>	<b>27774</b>	<b>27773</b>
<b>25</b>	<b>27754</b>	<b>27751</b>

**Table 3.1. ESI-MS characterization of sfGFP variants incorporated with different ncAAs.** Detected masses have an error of  $\pm 1$  Da.

Along with **19-25**, two additional amino acids were tested, namely, *ortho*-trifluoromethyl-phenylalanine (**26**) and *ortho*-fluoro-phenylalanine (**27**) (Figure 3.4). In GMMML supplemented with 2 mM ncAA **26**, we obtained minimal expression of full-length sfGFP. The mass spectra of the purified protein suggested the incorporation of both **26** and phenylalanine at the S2 site (Figure 3.4A), indicating the recognition of **26** by PylRS(N346A/C348A) is not significantly better than that of phenylalanine itself. Minimal incorporation of phenylalanine at an amber codons is a common observation for evolved *Mj*-TyrRS-tRNA<sub>CUA</sub><sup>Tyr</sup> and PylRS-tRNA<sub>CUA</sub><sup>Pyl</sup> pairs.<sup>156</sup> In a previous study, we found that while PylRS(N346A/C348A) was able to incorporate *meta*-fluoro-phenylalanine at an amber codon, ESI-MS analysis of the expressed protein indicated



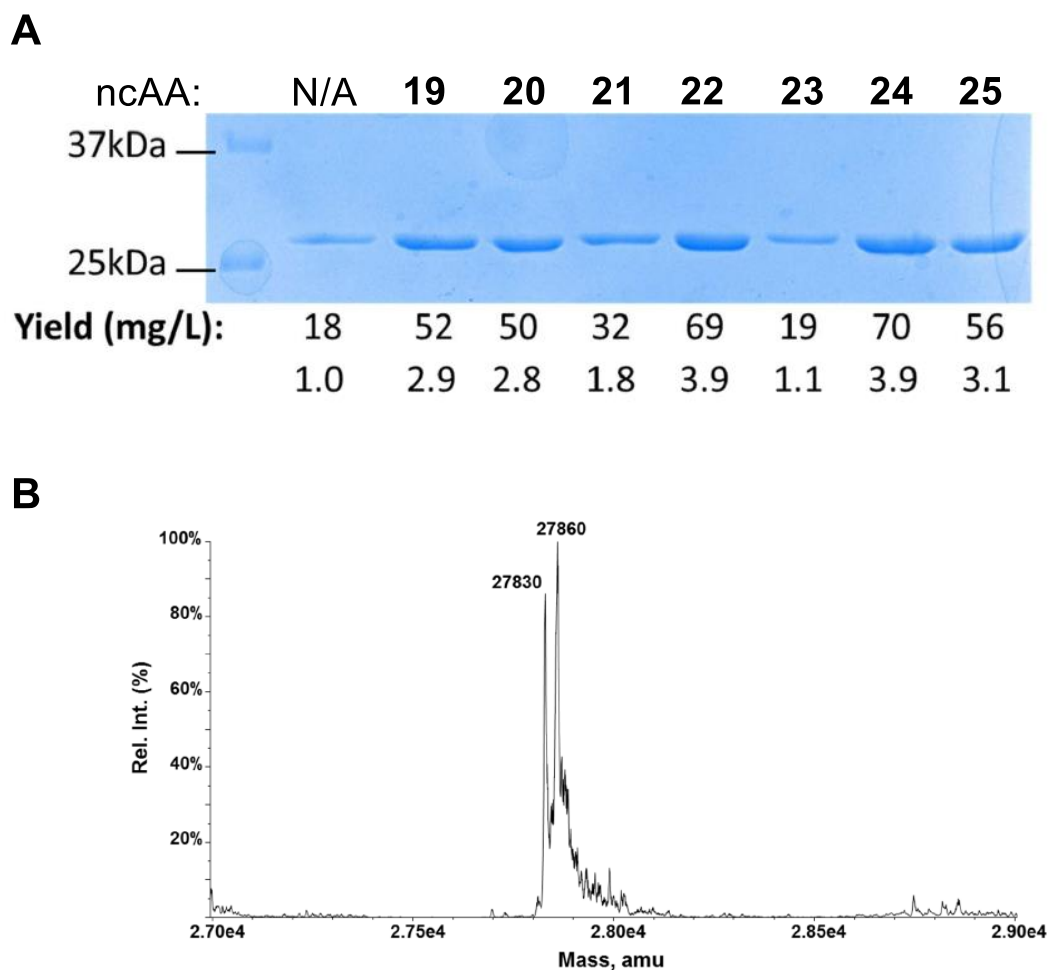
**Figure 3.4. The genetic incorporation of *ortho*-trifluoromethyl-phenylalanine and *ortho*-fluoro-phenylalanine. (A) Deconvoluted ESI mass spectra of sfGFP-**26**. (B) Deconvoluted ESI mass spectra of sfGFP-**27**.**

that this ncAA was also incorporated at phenylalanine codons.<sup>147</sup> Given the size similarity of fluorine and a proton, we attributed the miss-incorporation of *meta*-fluoro-phenylalanine at phenylalanine sites to an inability of the endogenous phenylalanyl-tRNA synthetase to distinguish between the two substrates. Similarly, when testing the incorporation of **27** at the S2 site of sfGFP, we found that the miss-incorporation rate was high. Indeed, major ESI-MS peaks of the purified protein suggest the incorporation

of **27** into sfGFP at multiple sites (Figure 3.4B). Taken together, these observations suggest that while the PylRS(N346A/C348A)- tRNA<sub>CUA</sub><sup>Pyl</sup> pair is able to direct the genetic incorporation of a monofluorinated phenylalanine at amber codons, *E. coli* cells with a more stringent phenylalanyl-tRNA synthetase that excludes monofluorinated phenylalanine as a substrate will need to be engineered for cleaner incorporation.

In a previous publication, we showed that PylRS(N346A/C348A) recognizes phenylalanine and mediates its incorporation at an amber mutation site in coordination with tRNA<sub>CUA</sub><sup>Pyl</sup>.<sup>146</sup> However, this background incorporation was suppressed when a ncAA that serves as a better substrate of PylRS(N346A/C348A) was provided. For instance, the PylRS(N346A/C348A)- tRNA<sub>CUA</sub><sup>Pyl</sup> pair induced significant background phenylalanine incorporation at an amber mutation at F27 of sfGFP in LB medium. However, providing *meta*-trifluoromethyl-phenylalanine obviated this background incorporation, and the expressed sfGFP had quantitative occupancy of *meta*-trifluoromethyl-phenylalanine at F27.<sup>147</sup> We used the same system to test the incorporation efficiencies of seven *ortho*-substituted phenylalanine derivatives (2 mM) in LB. Our data show that, except for **23**, all ncAAs led to sfGFP expression significantly higher (1.8–3.9-fold) than the background (Figure 3.5A). Further, although the sfGFP expression yield in the presence of **23** was comparable to the background, the major peak in the ESI-MS spectrum of purified sfGFP indicated occupancy of **23** at the designated amber mutation site (Figure 3.5B), indicating that **23** was efficiently incorporated. Presumably, further optimization studies that modify expression conditions, such as increasing the concentration of **23**, may lead to higher incorporation

efficiency. These data demonstrate that the practical use of the PylRS(N346A/C348A)- $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  pair in LB for the genetic incorporation of ncAAs is dependent on the ncAA identities and their concentrations.



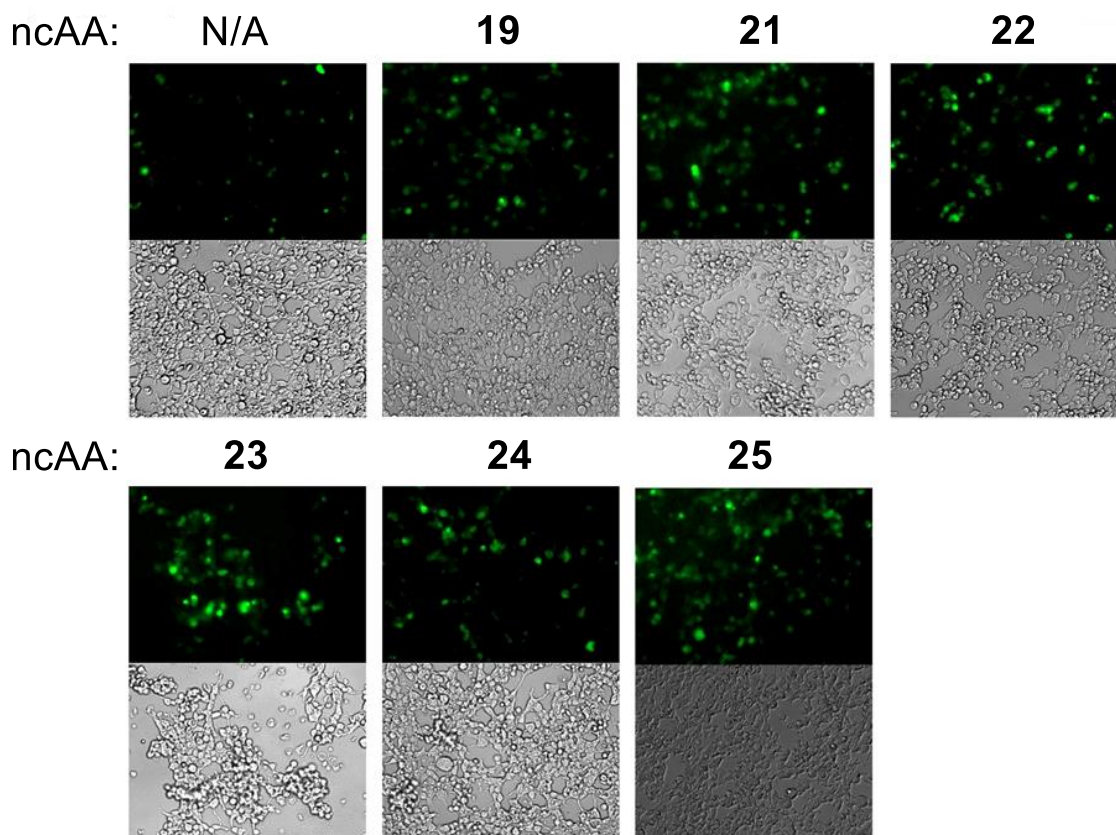
**Figure 3.5 The genetic incorporation of *ortho*-substituted phenylalanine derivatives in LB media.** (A) Expression of sfGFP-F27→**X** where **X** is one of the ncAAs **19-25**. N/A indicates no ncAA was provided in the growth media. The bottom numbers indicate the expression level relative to the control. (B) Deconvoluted ESI mass spectrum of sfGFP-F27→**23**. The major peak in the spectrum is from the incorporation of **23** (theoretical molecular weight = 27859 Da), while the minor peak is from the incorporation of phenylalanine (theoretical molecular weight = 27830 Da).

### 3.3.2 The use of PylRS(N346A/C348A) in Mammalian Cells

Next we asked whether the PylRS(N346A/C348A)-tRNA<sub>CUA</sub><sup>Pyl</sup> pair could be used to incorporate ncAAs into proteins in mammalian cells. To test the ability of the PylRS(N346A/C348A)-tRNA<sub>CUA</sub><sup>Pyl</sup> pair to facilitate the selective incorporation of ncAAs into proteins in mammalian cells, HEK293T cells were transfected with plasmids containing genes encoding epidermal growth factor receptor with a C-terminal GFP fusion (EGFR-GFP) and an amber mutation at the N128 position as well as PylRS(N346A/C348A) and tRNA<sub>CUA</sub><sup>Pyl</sup>.<sup>157</sup> Cells in cultures supplemented with 5 mM of ncAAs **19** or **21-25** displayed greater GFP fluorescence intensity compared to cells in the absence of an ncAA, suggesting enhanced suppression of the amber codon in the presence of ncAAs (Figure 3.6). Supplementing cultures with 5 mM **20** led to cell death. This is likely due to the low solubility of **20** in the growth medium. As expected, a low level of background protein expression was observed in the control culture, which can likely be attributed to background phenylalanine incorporation at the amber mutation.

### 3.3.3. Ortho-Cyano-Phenylalanine Is an Environmentally Sensitive Fluorescent Probe

Of the seven *ortho*-substituted phenylalanine substrates available to PylRS(N346A/C348A), **21** has the potential to be used for protein labeling via the palladium-catalyzed Suzuki–Miyaura cross-coupling reaction,<sup>148, 158</sup> and **24** can be used to cleave the polypeptide backbone upon irradiation at 365 nm, as a fluorescent distance



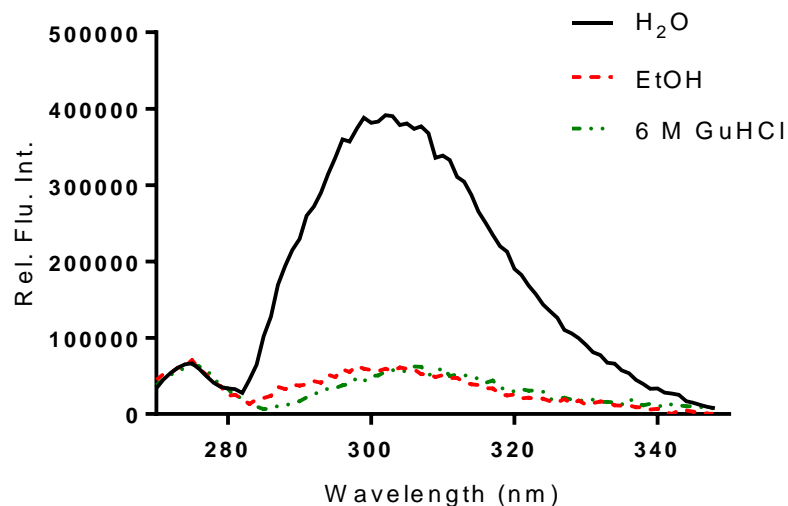
**Figure 3.6. Expression of EGFP-N128→X in HEK293T cells co-transfected with the plasmid containing genes encoding PylRS(N346A/C348A) and  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$ .** Cell cultures were supplemented with 5 mM of **19-25** or no ncAA (N/A). Supplementing the cultures with 5 mM of compound **20** resulted in nearly 100% cell mortality. The relative ratio of expressed EGFP proteins compared to the control, based on fluorescence intensity, is 2.6, 4.3, 3.2, 4.1, 2.2 and 4.9 for ncAAs **19**, **21**, **22**, **23**, **24**, and **25**, respectively.

probe due to its ability to quench the fluorescence of tryptophan, and as an IR probe due to the strong vibrational absorption of the nitro group.<sup>150, 159, 160</sup> **25** could also be used as an IR probe due to the strong vibrational absorption of the nitrile group<sup>132</sup> and as a sensor of protein local environment, protein folding/dynamics, and protein–ligand interactions due to the environmentally sensitive fluorescent features of the benzonitrile moiety. Indeed, an isomeric ncAA of **25**, *para*-cyano-phenylalanine (*p*CNF), has been



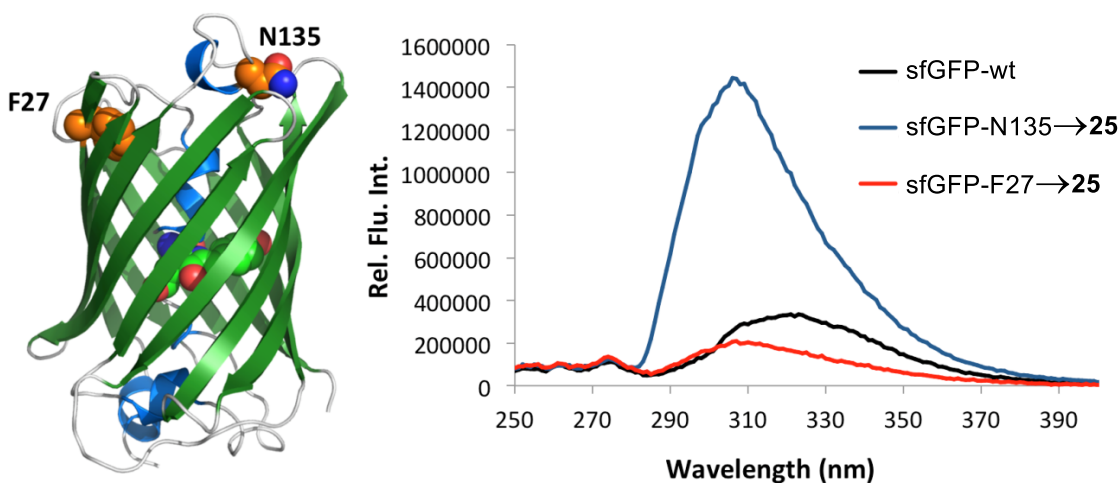
widely used as a reporter of protein local environment to study folding/dynamic analysis and protein–peptide interactions.<sup>161, 162</sup> Additionally, *p*CNF can be selectively excited in the presence of both tryptophan and tyrosine and has considerable spectral overlap with both fluorophores; therefore, its ability to form a Förster resonance energy transfer (FRET) pair with these residues was also recently explored.<sup>163, 164</sup> **25** could potentially serve as an alternative to *p*CNF and, because of its naturally occurring orthogonal tRNA<sub>CUA</sub><sup>Pyl</sup>, could be used in eukaryotic systems. Since the intermediate polarity of the nitrile group is well tolerated in both polar and nonpolar environments and the nitrile group is relatively small, replacing phenylalanine with **25** should minimally disturb the protein's structure and folding pattern.<sup>164</sup> Our preliminary measurements revealed that **25** has absorption and fluorescence spectra similar to those reported for *p*CNF and displays a nearly 10-fold increase in fluorescence when going from hydrophobic to hydrophilic solvents (Figure 3.7), suggesting that **25** could be used as a probe of protein local environment.

To test whether **25** could be used as a fluorescent probe of protein local environment, two sfGFP proteins were expressed: sfGFP-N135→**25** and sfGFP-F27→**25** with **25** incorporated at N135 and F27, respectively. N135 is solvent exposed and **25** is expected to be highly fluorescent at this site; however, F27 is buried within the hydrophobic interior of the protein, and therefore **25** incorporated at this position should display reduced fluorescence. Previous measurements indicate that mutations at these two positions do not significantly alter the folding of sfGFP.<sup>147</sup> The pEVOL-PylT-PylRS(N346A/C348A) plasmid was co-transformed into *E. coli* Top10 cells along with



**Figure 3.7. Fluorescence emission spectrum of 20 μM *ortho*-cyano-phenylalanine dissolved in water, ethanol, and 6 M guanidine hydrochloride.** Spectra were recorded in a 1 cm quartz cuvette at room temperature with a 1 nm resolution and 1 s·m<sup>-1</sup> integration time.  $\lambda_{\text{ex}} = 240$  nm.

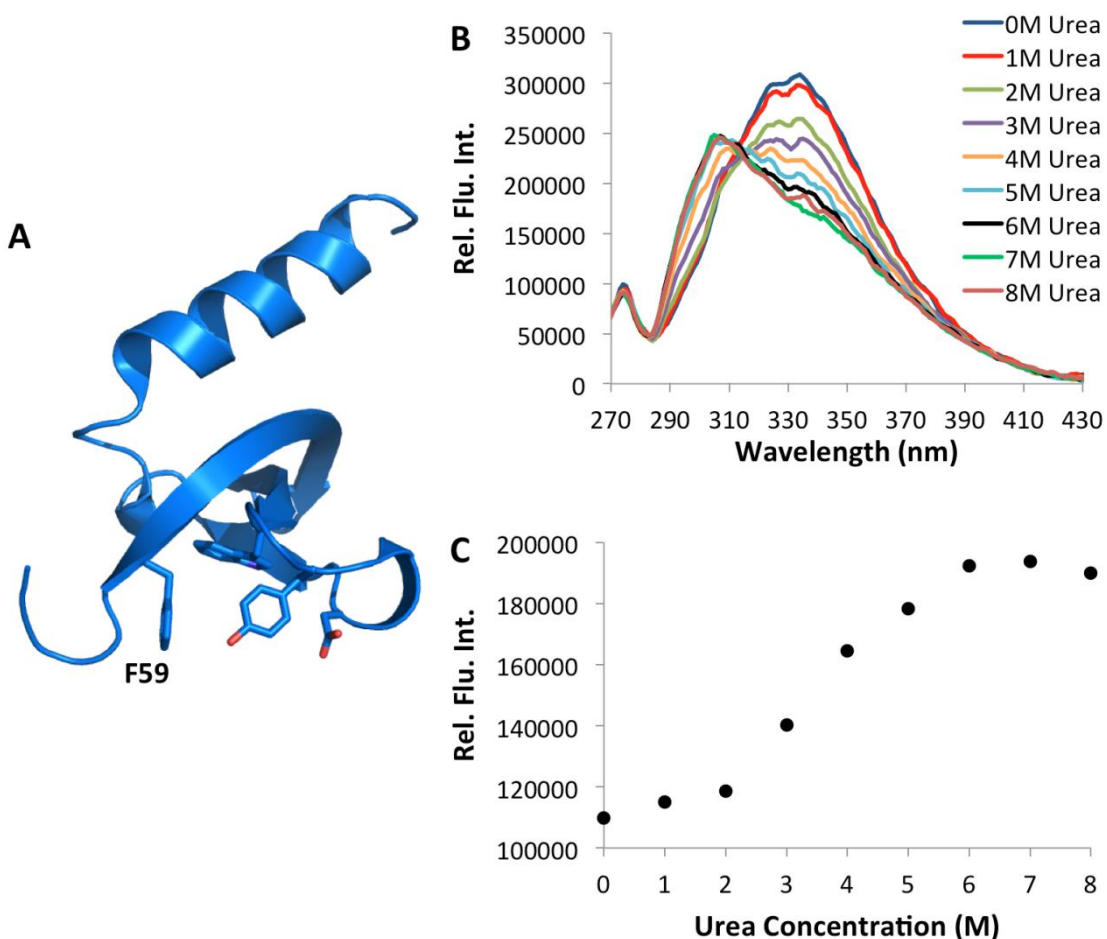
pBAD-sfGFP that contained sfGFP with an amber mutation at either the 27th or 135th position. Cultures were grown in a synthetic autoinduction medium supplemented with 2 mM ncAA **25** to afford full-length sfGFP-N135→**25** and sfGFP-F27→**25** in yields of 140 and 220 mg·L<sup>-1</sup>, respectively. No protein was detected in the absence of **25**. Figure 3.8 shows the fluorescence spectra of the folded proteins in phosphate buffered saline at pH 7.5 when excited at 240 nm. In the folded state, sfGFP-F27→**25** displays very little fluorescence, consistent with the positioning of **25** in the hydrophobic interior of the protein. However, sfGFP-N135→**25** displays nearly an 8-fold increase in fluorescence intensity consistent with **25** being solvent exposed in this protein. As a control, we also measured the fluorescence of the wildtype protein under the same



**Figure 3.8. Ortho-cyano phenylalanine fluorescence in sfGFP variants.** Fluorescence spectra of sfGFP-F27→**25** (red line), sfGFP-N135→**25** (blue line), and wild-type sfGFP (black line) excited at 240 nm. The structure of sfGFP is presented in the left side of the figure with F27 and N135 labeled.

conditions. The wildtype protein displayed fluorescence intensity similar to that of sfGFP-F27→**25** with a slightly red-shifted spectrum. These results indicate that **25** can be selectively excited in the presence of both tyrosine and tryptophan, as sfGFP contains one tryptophan residue and nine tyrosine residues, and demonstrate that **25** serves as an effective indicator of local environment within a protein.

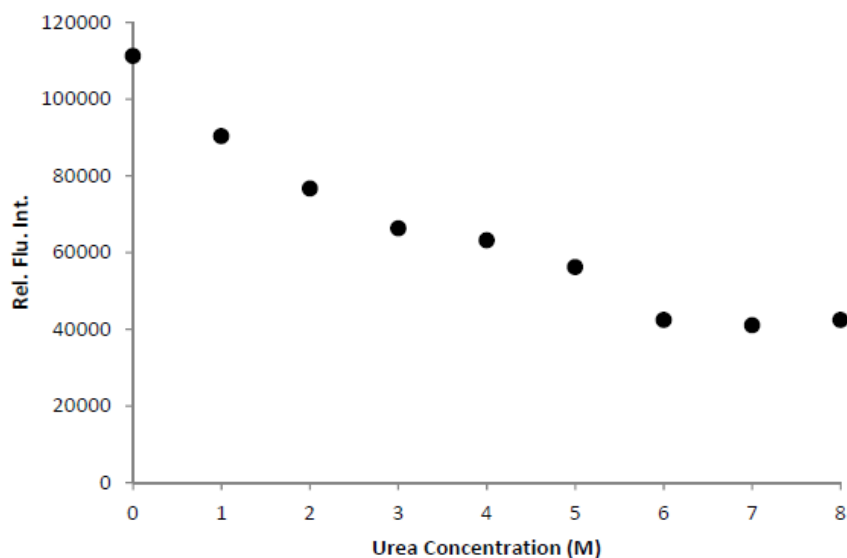
Next we demonstrated the use of **25** for the study of protein folding using the chromodomain of M-phase phosphoprotein 8 (MPP8). MPP8 has been shown to bind specifically to methylated lysine 9 of histone 3 and overexpression of MPP8 is linked to increase tumor metastasis.<sup>165</sup> The crystal structure of MPP8 revealed that F59 is located in the hydrophobic cage that forms a binding pocket for the methylated lysine (Figure 3.9A).<sup>166</sup> Therefore, in the folded protein, this residue is not exposed to the solvent and an increase in the fluorescence of **25** would be expected as the protein is



**Figure 3.9. Monitoring protein unfolding with *ortho*-cyano-phenylalanine** (A) The structure of MPP8 (based on the PDB entry: 3QO2). F59 is labeled. (B) Fluorescence spectra of the MPP8-F59→**25** at varied concentrations of urea. (C) Fluorescence intensity of MPP8-F59→**25** measured at 297 nm as a function of the urea concentration. The protein was measured at 2.5 $\mu$ M in PBS at pH 7.5. The excitation wavelength was 240 nm. A seven point, second order Savitsky-Golay filter was applied to Figure 3.9B to refine curves.

denatured. The chromodomain of MPP8, containing an amber mutation at F59, was expressed as an N-terminal fusion protein to sfGFP in *E. coli* Top10 cells that also expressed the PylRS(N346/C348A)-tRNA<sub>CUA</sub><sup>Pyl</sup> pair. Fusion to sfGFP allowed for high levels of the soluble MPP8 to be expressed.<sup>167</sup> The cells were grown in the synthetic

autoinduction medium supplemented with 2 mM ncAA **25** to afford the full length fusion protein. Figure 3.9B displays the fluorescence spectra of MPP8 in the presence of varying concentrations of urea. At low concentrations of the denaturant, a peak corresponding to the fluorescence of **25** was not detected, consistent with this residue being buried in the hydrophobic cage. As the concentration of urea is increased from 0 to 8 molar, an increase in **25** fluorescence at 297 nm was observed accompanied by a decrease in the tryptophan fluorescence of the protein, which is consistent with both **25** and tryptophan becoming solvent exposed. No increase in fluorescence at 297 nm was observed for the wild-type protein (Figure 3.10). Figure 3.9C shows the two-state unfolding curve for MPP8-F59→**25** obtained by exciting the protein at 240 nm and measuring the resulting fluorescence at 297 nm.



**Figure 3.10. Fluorescence intensity at 297 nm of wildtype sfGFP as a function of urea concentration.** The protein was measured at 2.5  $\mu$ M in PBS at pH 7.5. The excitation wavelength was 240 nm.

### 3.4 Conclusion

In conclusion, we have demonstrated an expanded ability of the PylRS(N346A/C348A) mutant to incorporate phenylalanine derivatives with substitutions at the *ortho* position. The enzyme demonstrates remarkable substrate promiscuity while retaining relative orthogonality toward canonical amino acids, making it a valuable tool for the genetic incorporation of ncAAs. Nearly 40 phenylalanine derivatives have been hitherto genetically encoded at amber codons using the PylRS(N346A/C348A)- tRNA<sub>CUA</sub><sup>Pyl</sup> pair. In addition, we have demonstrated the ability of PylRS(N346A/C348A) to facilitate the genetic incorporation of ncAAs in mammalian cell lines. Finally, we have demonstrated the ability of **25** to act as a sensitive fluorescent probe of protein local environment and protein folding. **25** shows significant differences in fluorescence intensity between hydrophobic and hydrophilic environments, and the nitrile substitution at the *ortho* position makes it an extremely small and noninvasive substitution in protein studies. We believe this advancement will greatly promote the adoption of the genetic ncAA incorporation approach in a broader biochemistry research field. In subsequent chapters, we will utilize the PylRS(N346A/C348A)- tRNA<sub>CUA</sub><sup>Pyl</sup> pair along with other PylRS mutants for the genetic incorporation of ncAAs into phage displayed peptides.

## CHAPTER IV

### CONSTRUCTION AND CHARACTERIZATION OF A TAG-OBLIGATE PHAGE DISPLAY LIBRARY

#### 4.1 Introduction

As discussed in Chapter 2, a straightforward method for increasing the chemical diversity of phage-displayed peptides is through the genetic incorporation of ncAAs via stop codon suppression. To date, over 150 ncAAs bearing diverse functional groups have been site-specifically incorporated into proteins using this technique.<sup>93</sup> This could theoretically be used to install any of these ncAAs at any position within a peptide library. However, direct application of nonsense suppression for phage display has proven to be extremely challenging, largely due to the inefficiency of ncAA incorporation via stop codon suppression.

Inefficient incorporation of ncAAs arises from several factors. For one, the orthogonal tRNA must compete with release factors for suppression of the nonsense codon during translation. Often, rather than the ncAA being incorporated at the nonsense codon, translation is terminated. These competing processes result in a reduction in the yield of proteins containing ncAAs compared to those that do not rely on nonsense suppression.<sup>168, 169</sup> Further, it has been demonstrated that engineered aaRSs are far less active (up to 1000-fold) than wildtype aaRSs.<sup>170</sup> This decreased activity is another contributing factor to inefficient ncAA incorporation.

In order for affinity selection of phage-displayed peptides to be successful, enrichment of a phage clone during selection should be a result of the functional advantage (binding affinity) provided to the clone by the displayed peptide. All other factors that could bias the selection and favor the enrichment of clones irrespective of the binding affinity of the peptide should be carefully minimized. The differential reproductive rate of bacteriophages within a library has been shown to be a significant source of bias in phage display affinity selections. In this type of bias, known as propagation-related (PR)-bias, certain clones are enriched over others due to their faster reproductive rate during the amplification steps between rounds of selection.<sup>59, 171</sup> Depending on the difference in propagation rate, PR-bias can make the selection of slowly reproducing clones difficult, if not impossible, even if the displayed peptide binds to the target. When the genetic ncAA incorporation technique is applied to phage display, inefficient stop codon suppression can severely limit the rate at which clones containing ncAAs propagate. This PR-bias against these clones has made the selection of phage-displayed peptides containing ncAAs extremely challenging.

Despite this hurdle, Schultz and colleagues have utilized phage display to evolve proteins and peptides containing ncAAs for several different targets. In these selections careful and extensive optimization of expression conditions (including time, temperature, plasmid system, and ncAA concentration) was used to maximize the rate of phage production.<sup>95</sup> Further, these studies utilized highly efficient ncAA/aaRS pairs to maximize the efficiency of stop codon suppression. However, even under the best conditions the rate of phage production for clones containing ncAAs was 3-fold lower



than clones that contain only canonical amino acids.<sup>95</sup> To overcome this expression bias, the ncAAs used for selection in these studies were carefully chosen such that they would provide a strong selective advantage. For example, phage-displayed antibody libraries containing *para*-sulfo-tyrsoine (*p*-STyr) were used to evolve antibodies that bind to the HIV coat protein gp120.<sup>95, 96</sup> *p*-STyr was chosen with the foreknowledge that naturally occurring anti-gp120 antibodies contain *p*-STyr residues, and that these residues contribute significantly to the binding interaction between the antibody and gp120.<sup>172</sup> In another example, a phage-displayed peptide library containing bipyridyl alanine (Bpy-Ala) was used for the evolution of peptides that bind to Ni<sup>2+</sup>.<sup>99</sup> However, it was known prior to selection that phages containing Bpy-Ala would be favored given the ability of the ncAA to chelate Ni<sup>2+</sup> in a bidentate fashion. While this work by Schultz et al. has pioneered the use of ncAAs in phage display, the method is currently limited to highly efficient ncAA/aaRS pairs and to ncAAs that will provide a clear advantage during selection. Only under these constraints can one overcome the PR-bias and select for proteins and peptides containing ncAAs.

One way to minimize PR-bias would be to increase the ncAA incorporation efficiency. Numerous studies have sought to address this problem through engineering of the orthogonal aaRS and tRNA,<sup>173, 174</sup> ribosome,<sup>175</sup> elongation factors,<sup>176</sup> or through the deletion of release factors<sup>169</sup>. While significant advances have been made, the efficiency of incorporation is still low compared to canonical amino acids. Another option would be to simply eliminate the fast propagating clones from the library pool. By removing sequences that contain only canonical amino acids, clones containing

ncAAs can be selected based on their functional performance without concern for their slower propagation. Herein, we report a novel strategy, based on a unique property of the Ff bacteriophage infection mechanism, for the construction of phage-displayed peptide libraries. These libraries are designed such that every clone contains an ncAA that is incorporated in response to amber (TAG) codons. We demonstrate that this “TAG-obligate” library can be used to display peptides containing ncAAs that have varying incorporation efficiencies. Finally, we show that this strategy for library construction generates robust libraries that can undergo iterative rounds of selection and amplification without apparent PR-bias.

## **4.2 Experimental Details**

### *4.2.1 Miscellaneous Information*

All primers were purchased from Integrated DNA Technologies, Inc. (Coralville, PA). Enzymes and buffers for molecular cloning were purchased from New England Biolabs (Ipswich, MA). ncAAs were purchased from Chem-Impex International, Inc. (Wood Dale, IL) and used without further purification or synthesized as described. DNA sequencing was performed by Eton Biosciences, Inc. (San Diego, CA). Mass spectra were collected by the Department of Chemistry Mass Spectrometry Facility at Texas A&M University.

#### 4.2.2 Primer List

CloDFori-AscI-F: 5'-TTGGCGCGCCCAATTAGCTAGCTCACTCGGTC-3'

CloDFori-AvrII-R: 5'-TG TTCCTAGGGATAAAATTGCACTGAAATCTAG-3'

pEVOLori-AvrII-F: 5'-TG TTCCTAGGTCTTCAAATGTAGCACCTGAAG-3'

pEVOLori-AscI-R: 5'-TTGGCGCGCCCCTTTTTTCTCCTGCCACATG-3'

pADL-NcoI-(NNK)<sub>7</sub>-g3-F: 5'-CATGCCATGGCC(NNK)<sub>7</sub>GCGGCGAAAGCGG-3'

pADL-NcoI-R: 5'-CATGCCATGGCCGGCTGGGCCGC-3'

#### 4.2.3 Plasmid Construction

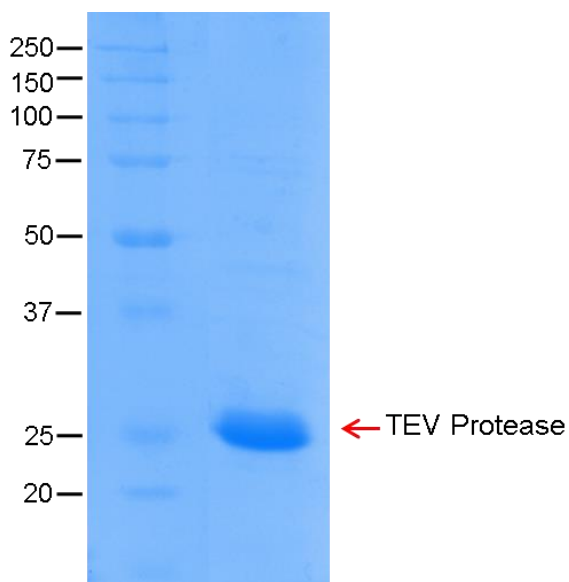
pEVOL-CDF: To exchange the origin of replication on pEVOL from p15A to CloDF, the CloDF origin was amplified from the plasmid pCDF-1 (EMD Millipore) using the primers CloDFori-AscI-F, containing an *AscI* site, and CloDFori-AvrII-R, containing an *AvrII* site. The PCR product was cloned into *AscI* and *AvrII* sites that had been introduced into previously reported pEVOL-PylT-PylRS constructs using the primer pair pEVOLori-AvrII-F and pEVOLori-AscI-R to afford pEVOL-CDF-PylT-PylRS.<sup>146, 177</sup>

pIII Phagemid Library: The phagemid pADL-10b was purchased from Antibody Design Labs (San Diego, CA). To introduce a seven site library at the N-terminus of pIII, the phagemid was amplified using the primers pADL-NcoI-(NNK)<sub>7</sub>-g3-F and pADL-NcoI-R. The PCR product was digested with *NcoI*, ligated with T4 DNA ligase, and used to transform electrocompetent *E. coli* Top10 F' yielding  $1.4 \times 10^9$  transformants.

#### 4.2.4 Protein Expression and Purification

TEV Protease: His<sub>6</sub>-TEV(S219V) protease was expressed and purified essentially as described previously<sup>178</sup> with minor modifications. Briefly, *E. coli* BL21(DE3) containing the plasmid encoding TEV protease were grown in 2xYT to an OD<sub>600</sub> = 0.6 at which point 1 mM IPTG was added and the protein was expressed at 30°C. 4 h after induction the cells were pelleted (6k rpm, 20 min) and the cell pellets were stored at -80°C until purification.

TEV protease was purified via nickel affinity chromatography exactly as described in Chapter 3. Following elution from the column, the protein was dialyzed overnight against TEV storage buffer (25 mM Tris pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM DTT), concentrated to 15 μM, flash frozen, and stored at -80°C. SDS-PAGE analysis of purified TEV protease is shown in Figure 4.1.



**Figure 4.1.** SDS-PAGE analysis of purified TEV protease.

*sfGFP-N134* → **35**: To express sfGFP with **35** at position 134, *E. coli* Top10 were transformed with the plasmid pBAD-sfGFP-N134(TAG), which contains the gene encoding sfGFP with a TAG mutation at position N134, and a previously reported pEVOL construct for the incorporation of  $N_{\epsilon}$ -propionyl-lysine (pEVOL-PylT-PrKRS).<sup>177</sup> Cells were grown at 37°C in 2xYT media supplemented with 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin and 34  $\mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol to an  $\text{OD}_{600} = 0.6$  at which point protein expression was induced with the addition of 0.2% arabinose, 5 mM **35** (racemic), and 5 mM nicotinamide. 16 h after induction cells were pelleted (4k rpm, 20 min) and cell pellets were stored at -80°C until purification. sfGFP-N134 → **35** was purified via nickel affinity chromatography exactly as described in Chapter 3.

*Ubiquitin-K48* → **35**: To express ubiquitin with **35**, *E. coli* BL21(DE3) were co-transformed with the previously mentioned pEVOL-PylT-PrKRS plasmid and a previously reported<sup>179</sup> pET-Ub-K48(TAG) plasmid which contains the gene encoding ubiquitin with a TAG mutation at position K48 and a 6 x His tag at the C-terminus. The transformed cells were grown at 37°C in 100 mL of 2xYT media supplemented with 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin and 34  $\mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol to an  $\text{OD}_{600} = 0.6$  at which point protein expression was induced with the addition of 5 mM **35** (racemic), 1 mM IPTG, and 0.2% arabinose to the growth media. 8 h after induction, the cells were pelleted (4k rpm, 20 min) and cell pellets were stored at -80°C until purification. Ubiquitin-K48 → **35** was purified by low pH fractionation as described previously<sup>179</sup> followed by purification by nickel affinity chromatography as described in Chapter 3.

#### 4.2.5 Biotinylation of TEV Protease

To generate biotin labelled TEV protease, aliquots of the purified protein, in storage buffer, were combined and dialyzed against PBS pH 7.6 overnight at 4°C. The following day the protein was dialyzed against PBS for an additional 1.5 h. The pH was adjusted to 8.0 using 0.1 M NaOH and the protein was concentrated to 10 µM. The concentrated protein was biotinylated with a 10-fold molar excess of Ez-Link™ NHS-Biotin (Thermo Fisher Scientific) for 1 h at room temperature with end-over-end rotation. After 1 h the excess biotin reagent was removed by ultrafiltration (10 kDa MWCO) by performing ten-fold dilutions in PBS and then concentrating to 1 mL (3x). Aliquots of the biotinylated protein were flash frozen and stored at -80°C.

#### 4.2.6 Phage Quantification

Phages were quantified via a colony forming unit (cfu) assay. In this assay, 10-fold serial dilutions of the phage preparation were performed in 2xYT media. 10 µL of each dilution was added to 90 µL of log phase *E. coli* Top10 F' (OD<sub>600</sub> = 0.45-0.55) in a 1.5 mL tube. The tubes were incubated at 37°C for 45 min. After incubation, 10 µL of the infected culture was spotted onto an LB agar plate containing the appropriate antibiotics. The liquid was allowed to evaporate, and the plate was incubated overnight at 37°C. The next day, the colonies in each spot were counted and this number was used to calculate the number of colony forming units in the phage preparation.

#### 4.2.7 Superinfection-Selection to Generate the TAG-Obligate Phagemid Library

First Round of Superinfection: The seven site phagemid library in *E. coli* Top10 F' was grown at 37°C in 2xYT media supplemented with 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin and 10  $\mu\text{g}\cdot\text{mL}^{-1}$  tetracycline. Upon reaching an OD<sub>600</sub> of 0.27, pIII expression was induced with the addition of 0.2 mM IPTG. 30 minutes post induction, CM13 helper phage (Antibody Design Labs,  $1 \times 10^{11}$  pfu, MOI  $\approx$  15) was added and the culture was incubated at 37°C for an additional 40 minutes. Following infection with CM13, cells were pelleted (6k rpm, 10 min), the supernate was removed, and the cell pellet was resuspended in 150 mL 2xYT containing 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, 25  $\mu\text{g}\cdot\text{mL}^{-1}$  kanamycin, and 1 mM IPTG. The culture was grown overnight at 37°C. The following day, the phagemid library was extracted with a commercial Midi prep kit (Epoch Life Sciences) according to the manufacturer's protocol and then purified by agarose gel electrophoresis.

Passage through DH5 $\alpha$ : The purified phagemid library was used to transform electrocompetent *E. coli* DH5 $\alpha$  containing M13KO7(gIII). Transformed cells were recovered for 1 h in 2xYT media and then grown overnight in 500 mL 2xYT containing 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, 25  $\mu\text{g}\cdot\text{mL}^{-1}$  kanamycin, and 1 mM IPTG. A small aliquot of the transformation was tittered to ensure the number of transformants exceeded the post-superinfection titer from the first round of superinfection-selection. The following day, the cells were pelleted (4k rpm, 30 min), the supernate was transferred to a new tube, and phagemids were precipitated by the addition of a 5x polyethylene glycol (PEG) precipitation solution (2.5 M NaCl, 20% PEG 8,000) with incubation on ice for 2 h.

Precipitated phages were collected by centrifugation (4k rpm, 1 h) and then resuspended in 3 mL PBS for 30 min at 4°C. The resuspended phages were centrifuged (10k rpm, 20 min) to remove residual cells and then incubated at 65°C for 15 min before titering.

Second Round of Superinfection: Purified phages ( $3.7 \times 10^9$  cfu) were used to infect actively growing *E. coli* Top10 F' at  $OD_{600} = 0.5$  in 100 mL 2xYT supplemented with  $10 \mu\text{g}\cdot\text{mL}^{-1}$  tetracycline (MOI  $\approx 0.15$ ). After 1 hour, the culture was diluted 1:1 with fresh 2xYT and then ampicillin and IPTG were added to a final concentration of  $100 \mu\text{g}\cdot\text{mL}^{-1}$  and 0.5 mM, respectively. The culture was grown at 37°C to  $OD_{600} = 0.42$  at which point 60 mL was removed and infected with CM13 ( $1 \times 10^{11}$  pfu, MOI  $\approx 10$ ). After incubating for an additional 45 min, the cells were pelleted (6k rpm, 10 min) and then resuspended in 500 mL 2xYT containing  $100 \mu\text{g}\cdot\text{mL}^{-1}$  ampicillin,  $25 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin, and 1 mM IPTG. The culture was grown overnight at 30°C and the phagemid was extracted and purified as describe above.

#### 4.2.8 Expression and Purification of Phages Containing ncAAs

Expression with Phenylalanine Derivatives: To express the TAG-obligate phagemid library with phenylalanine derivatives we used a previously reported pyrrolysyl-tRNA synthetase mutant containing the N346A/C348A mutations (PylRS(N346A/C348A)).<sup>146</sup> PylRS(N346A/C348A) and its cognate tRNA ( $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$ ) were expressed from a pEVOL construct containing the CloDF replication origin dubbed pEVOL-CDF-PylT-PylRS(N346A/C348A) which was constructed as described above. Electrocompetent *E.*



*coli* containing M13KO7(pIII<sup>-</sup>) and pEVOL-CDF-pyIT-PylRS(N346A/C348A) were transformed with the TAG-obligate phagemid library to yield  $6.92 \times 10^8$  transformants. Transformed cells were grown at 37°C in 200 mL of 2xYT media supplemented with  $100 \mu\text{g}\cdot\text{mL}^{-1}$  ampicillin,  $34 \mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol, and  $25 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin to  $\text{OD}_{600} = 0.4-0.6$  at which point 1 mM IPTG and 0.2% arabinose was added and the culture was split into two flasks. To one flask was added one of the ncAA **19** or **36-38** to a final concentration of 4 mM. Following induction, the phagemids were expressed at 30°C for 18-24 hours.

After expression, each of the 100 mL cultures was split into two 50 mL tubes and the cells were pelleted (6k rpm, 20 min). The upper 40 mL of the supernate was poured into new 50 mL tubes and to each tube was added 10 mL of a 5x phage precipitation solution (2.5 M NaCl, 20 % PEG 8,000). The tubes were mixed by inversion and the phages precipitated at 4°C overnight. The following day the phages were pelleted (11k rpm, 25 min), the supernate was decanted, and both pellets were resuspended in the same 10 mL of PBS pH 7.5. Residual cells were pelleted (6k rpm, 20 min) and the supernate was transferred to a new tube. The phages were precipitated a second time by the addition of 2.5 mL of phage precipitation solution and incubation at 4°C overnight. The following day the phages were pelleted (11k rpm, 25 min) and the pellet was resuspended in 1 mL PBS. Residual cells were pelleted (14k rpm, 10 min) and any remaining cells were killed by incubating the phage solution at 65°C for 15 min. To verify successful ncAA incorporation, the purified phages that were expressed in the presence and absence of ncAA were quantified using the cfu assay described above.

Expression with Lysine Derivatives: For expression of phages containing **28** and **29** electrocompetent *E. coli* Top10 containing M13KO7(pIII) and a pEVOL-CDF construct containing tRNA<sub>CUA</sub><sup>Pyl</sup> and a previously reported<sup>177</sup> PylRS mutant with a Y384W mutation (BuKRS) were transformed with the purified phagemid library yielding  $3.1 \times 10^9$  transformants. For the expression of phages containing **35**, electrocompetent *E. coli* Top10 containing containing M13KO7(pIII) and a pEVOL-CDF construct containing tRNA<sub>CUA</sub><sup>Pyl</sup> and a previously reported<sup>177</sup> PylRS mutant containing the mutations L301M/Y306L/L309A/C348F/Y384W (PrKRS) were transformed with the TAG-obligate phagemid library yielding  $2.1 \times 10^9$  transformants. Phages were expressed and purified exactly as described for the phenylalanine derivatives but with the addition of 5 mM nicotinamide to the growth media.

#### 4.2.9 Affinity Selection against TEV Protease

Streptavidin coated magnetic beads (100  $\mu$ L, 50% slurry, Genscript) were washed with PBS pH 7.4 (2x 1 mL) and then resuspended in 50  $\mu$ L of PBS. 5  $\mu$ g of biotin labelled TEV protease in 400  $\mu$ L PBS was added and the protein/bead mixture was incubated for 15 min at room temperature. After 15 min the beads were washed with PBS (3x 1 mL), resuspended in 300  $\mu$ L PBS and then 150  $\mu$ L of 2x blocking buffer was added to the tubes (PBS pH 7.4, 0.3% Tween 20, 3% BSA). At the same time 0.5 mL of 2x blocking buffer was also added to the purified phagemid library in 1 mL PBS. Both tubes were incubated at room temperature with end-over-end rotation for 30 min. After 30 min the blocking buffer was removed from the tube with the beads and the blocked phage library

(0.75 mL) was added. The beads/library mixture was incubated at room temperature with end-over-end rotation for 30 min. After 30 min the supernatant was removed and the beads were washed (8x 1 mL PBS containing 0.1% Tween-20; 2x 1 mL PBS). Bound phages were eluted from the beads by incubating with 100  $\mu$ L elution buffer (50 mM glycine pH 2.2) for 10 min. The elution was then immediately added to 50  $\mu$ L of neutralization buffer (1 M Tris pH 8.0).

A small portion (10  $\mu$ L) of the eluted phages was used for titering and the remainder was amplified by adding the elution to a 20 mL culture of *E. coli* Top10 F<sup>-</sup> (OD<sub>600</sub> = 0.45-0.55) in 2xYT media supplemented with 10  $\mu$ g·mL<sup>-1</sup> tetracycline, and incubated at 37°C, for 45 minutes. After 45 minutes, the cells were pelleted (4k rpm, 20 min), resuspended in 300 mL of fresh 2xYT media supplemented with 100  $\mu$ g·mL<sup>-1</sup> ampicillin, and grown overnight at 37°C. The following day the phagemids were extracted using a commercial plasmid Miniprep kit (Epoch) according to the manufacturer's protocol. The purified phagemids were used to transform electrocompetent *E. coli* Top10 containing pEVOL-PylT-PylRS(N346A/C348A) and M13KO7(pIII<sup>-</sup>). After electroporation the transformed cells were recovered in 1 mL 2xYT supplemented with 35  $\mu$ g·mL<sup>-1</sup> chloramphenicol and 25  $\mu$ g·mL<sup>-1</sup> kanamycin for 1 hour at 37°C. A small portion of the recovered cells were titered to ensure the number of transformed cells exceeded the number of eluted phages by at least 100-fold. After 1 hour, the recovered cells were added to 50 mL of 2xYT media supplemented with ampicillin, chloramphenicol, and kanamycin and amplified overnight. This amplified

culture was used to prepare glycerol stocks for expressing phages for further rounds of selection according to the methods described above.

#### *4.2.10 Affinity Selection against Streptavidin*

Selection was performed against streptavidin which had been nonspecifically adsorbed to a polystyrene petri dish. To immobilize the target protein, lyophilized streptavidin (Chem-Impex International, Woodale IL) was dissolved in 0.1 M NaHCO<sub>3</sub> pH 8.6 to a final concentration of 25 µg·mL<sup>-1</sup>. 1.5 mL of the protein solution was added to a sterile petri dish (60 x 15 mm, Falcon 35-3004) and incubated at 4°C overnight in a sealed, humidified, box. The following day, the protein coating solution was removed, the petri dish was filled with blocking buffer (0.1 M NaHCO<sub>3</sub>, 5 mg·mL<sup>-1</sup> BSA, 0.1 mg·mL<sup>-1</sup> streptavidin, pH 8.6), and the plate was incubated for 2 h at 4°C. Following blocking, the plate was washed 6 times with 2 mL PBST (PBS + 0.1 % Tween-20, pH 7.5) and the phage library, in 1 mL PBS, was added to the dish. The phages were incubated with the target at room temperature in a humidified box for 1 hour (rounds 1-3) or 30 minutes (round 4). After incubating the library with the target, unbound phages were removed and the dish was washed 10 times by completely filling the dish with PBST, swirling for 30 seconds, pouring off the PBST, and then slapping the dish face-down on a clean paper towel. The bound phages were then eluted with the addition of 1 mL of 0.1 mM D-biotin in PBS incubating at room temperature for 30 minutes. Eluted phages were quantified and amplified exactly as described for the selection against TEV protease.

#### 4.2.11 Chemical Synthesis

The Synthesis of  $N_\epsilon$ -butyryl-lysine (28) and  $N_\epsilon$ -crotonyl-lysine (29):  $N_\epsilon$ -butyryl-lysine and  $N_\epsilon$ -crotonyl-lysine were synthesized according to our previously reported procedure.<sup>177</sup>

The Synthesis of 2-amino-8-oxodecanoic acid (35): The synthetic scheme shown in Figure 4.2 was followed for the synthesis of 2-amino-8-oxodecanoic acid (35).

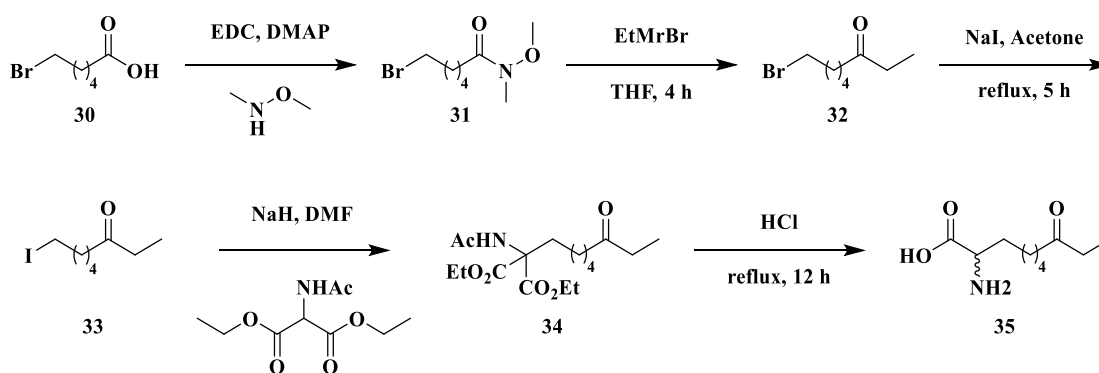


Figure 4.2. The synthesis of 2-amino-8-oxodecanoic acid (35).

A mixture of **30** (2.73 g, 14.0 mmol), 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (3.22 g, 16.8 mmol), 4-dimethylaminopyridine (0.17 g, 1.4 mmol), and *N,O*-dimethylhydroxylamine (1.5 g, 15.4 mmol) in dichloromethane (75 mL) was stirred at room temperature for 16 h. The mixture was diluted in dichloromethane (100 mL), washed with water (30 mL), hydrochloric acid (0.5 M, 30 mL), sodium hydroxide (0.5 M, 30 mL) and brine (15 mL),

dried ( $\text{Na}_2\text{SO}_4$ ), and concentrated under reduced pressure to give the crude **31** (3.39 g, quant.) as a yellow oil. The material was directly used without further purification.

To a stirred solution of **31** (3.39 g, ~14.0 mmol) in anhydrous THF (40 mL) cooled in an ice bath under argon protection was added a solution of ethylmagnesium bromide (3.0 M in diethyl ether, 6.6 mL, 20.0 mmol) dropwise over 20 min. After stirring for 3 h, hydrochloric acid (1.0 M, 21.9 mL) was cautiously added dropwise. The mixture was diluted in ethyl acetate (50 mL), washed with water (20 mL) and brine (20 mL), dried ( $\text{Na}_2\text{SO}_4$ ), and concentrated under reduced pressure to give crude **32** (2.52 g, 83%) as a yellow oil, which was directly used in the next step.

A mixture of **32** (2.5 g, 12.1 mmol) and sodium iodide (5.39 g, 36.2 mmol) in acetone (40 mL) was refluxed for 5 h. The solvent was then evaporated under reduced pressure and ether (50 mL) was added. The suspension was filtered and washed with ether (20 mL). The combined filtrate was washed with water (15 mL), sodium thiosulfate (1.0 M, 10 mL) and brine (10 mL), dried ( $\text{Na}_2\text{SO}_4$ ), and concentrated under reduced pressure to give **33** (2.90 g, 95%) as a yellow oil, which was directly used in the next step.

To a suspension of diethylacetamidomalonate (1.60 g, 7.2 mmol) in anhydrous DMF (5 mL) was added sodium hydride (60% dispersion in mineral oil, 0.30 g, 7.68 mmol), and the mixture was stirred at room temperature for 10 min. Compound **33** (2.00 g, 7.04 mmol) in DMF (4 mL) was then added dropwise over 5 min, and the resulting yellow solution was stirred at room temperature overnight. Water (5 mL) was added, and the mixture was diluted in ethyl ether (50 mL), washed with a mixture of brine (9 mL)

and hydrochloric acid (1 M, 1 mL), dried ( $\text{Na}_2\text{SO}_4$ ), concentrated under reduced pressure, and chromatographed (ethyl acetate/hexanes, 1:3 to 1:1) to give **34** (2.2 g, 82%) as a yellow oil.

$^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$  6.77 (br, s, 1H), 4.23-4.18 (m, 4H), 2.40-2.25 (m, 4H), 1.99 (s, 3H), 1.53-1.50 (m, 2H), 1.20-1.25 (m, 8H), 1.01-0.99 (m, 6H).

A suspension of **34** (3.0 g) in concentrated hydrochloric acid (12 M, 20 mL) was heated at reflux overnight. The solvent was evaporated under a high vacuum generated by an oil pump, and the residue was dissolved in minimal amount of water and loaded onto an ion-exchange column made from Dowex 50WX4-400 cation-exchange resin (wet volume ~10 mL) which had been prewashed with hydrochloric acid (6 M, 30 mL) and then copious amount of water until neutral. The column was desalted with excessive water washing (200 mL) and then eluted with hydrochloric acid (0.5 M to 3 M) to give **35** (2.17 g, quant.) as a white solid. The proton NMR of compound **35** is shown in Figure 4.3.

$^1\text{H}$  NMR (300 MHz,  $\text{D}_2\text{O}$ )  $\delta$  3.90 (t, 1H,  $J = 3.6$  Hz), 2.40-2.36 (m, 4H), 1.81-1.72 (m, 2H), 1.40-1.35 (m, 2H), 1.28-1.14 (m, 4H), 0.80 (t, 3H,  $J = 0.9$  Hz).

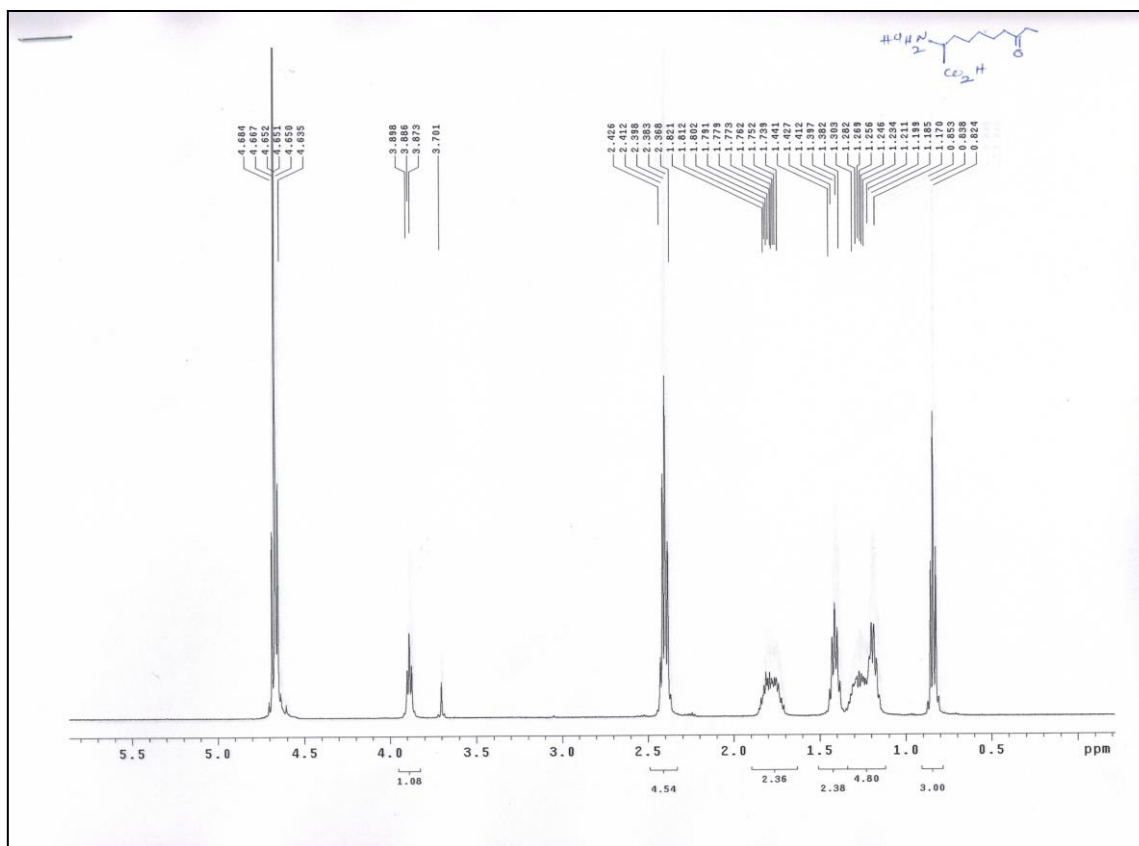


Figure 4.3. <sup>1</sup>H-NMR spectrum of 2-amino-8-oxodecanoic acid (35).

## 4.3 Results and Discussion

### 4.3.1 Construction of a TAG-Obligate Phage Display Library

Our strategy for constructing a TAG-obligate phage display library relies on a phenomenon of Ff bacteriophage biology known as superinfection immunity. The mechanism of Ff phage infection of *E. coli* was described in detail in Chapter 1. Briefly, adsorption of the virus to the host cell relies on two interactions between *E. coli* cell surface proteins and the phage coat protein pIII. During adsorption of the phage to the host, the N2 domain of pIII binds to the F-pilus of *E. coli*. This is followed by binding of



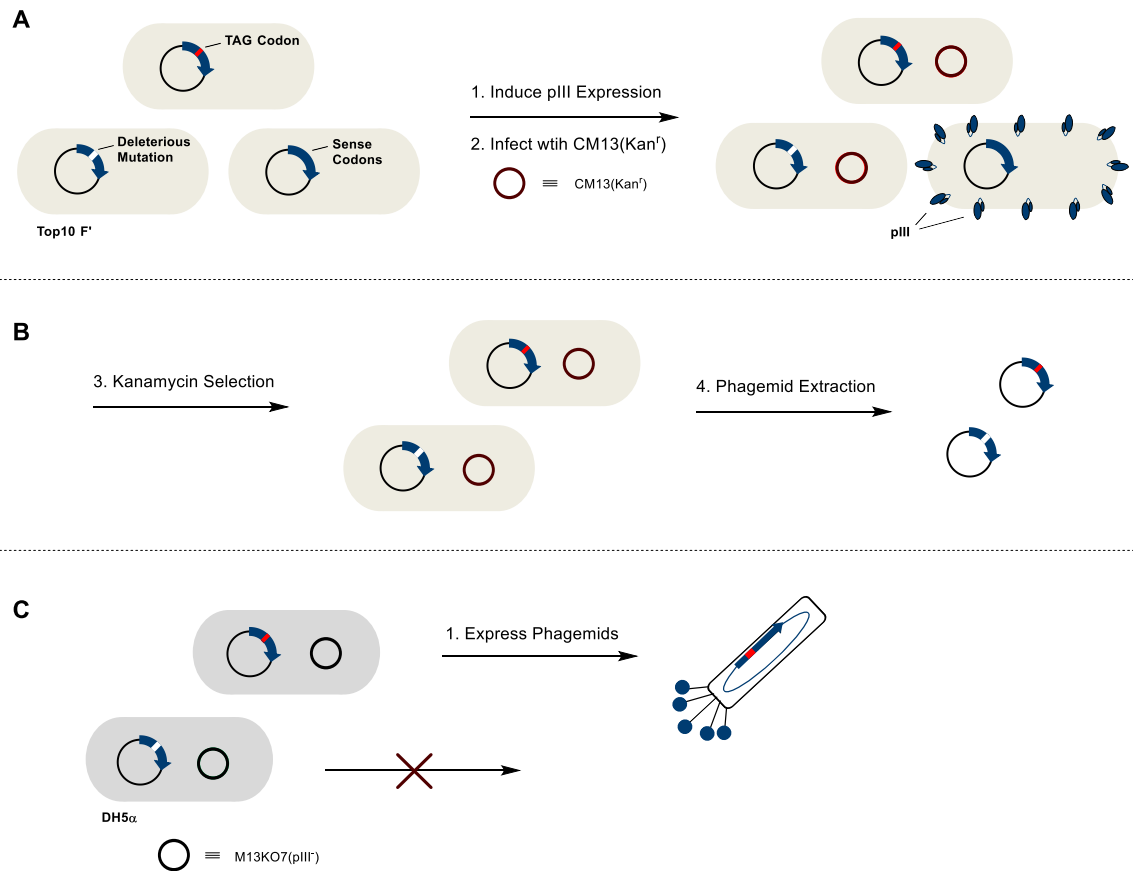
the N1 domain of pIII to the *E. coli* membrane bound protein TolA. However, if the bacterium is already expressing pIII, either from a plasmid or from previous phage infection, the endogenously-expressed pIII will bind to TolA preventing adsorption and, therefore blocking infection (superinfection) by the second phage.<sup>180, 181</sup> Such cells are said to have superinfection immunity. Our plan to construct the TAG-obligate library was to first build a standard library with both sequences containing TAG codons and those containing only sense codons. Then, we would perform a selection based on superinfection immunity (superinfection-selection) to remove sequences that contain only sense codons from the library pool.

We started by constructing a combinatorial heptapeptide (7-mer) library fused to the N-terminus of pIII using the pADL-10b phagemid system. In this phagemid vector, pIII is under control of a lac promoter which allows for tight regulation and high expression of the peptide-pIII fusions. When supplemented with a helper phage that is lacking a functional gene encoding pIII, this phagemid allows for pentavalent display of peptide-pIII fusions on phagemid particles. To construct the library seven randomized codons were introduced by site-saturation mutagenesis, using NNK codons, at the N-terminus of pIII immediately downstream of the PelB leader sequence. Fusion to the PelB leader peptide is required to direct pIII to the periplasm following translation.<sup>182</sup> Once in the periplasm, the leader sequence is removed by a signal peptidase to afford the 7-mer as a free N-terminal fusion to pIII.<sup>183</sup> Following cloning, transformation of electrocompetent *E. coli* Top10 F' with the phagemid library yielded  $1.4 \times 10^9$  unique

transformants (theoretical peptide diversity of a seven site library with 21 amino acids =  $1.80 \times 10^9$ ).

Next, we performed superinfection-selection to remove the peptide sequences that contain only canonical amino acids from the library pool. To do this, *E. coli* Top10 F' harboring the phagemid library were grown in media supplemented with ampicillin and tetracycline to  $OD_{600} = 0.24$  at which IPTG was added to induce the expression of pIII. Under these conditions if the cell harbors a copy of pIII that contains only sense codons, pIII will be expressed rendering the cell immune to superinfection. However, if pIII contains a TAG codon, translation termination at the nonsense codon will prevent the expression of pIII leaving the cell vulnerable to superinfection (Figure 4.4A). Therefore, these cells can be selected after infection with a second phage that bears an antibiotic resistance marker (Figure 4.4B). Shortly after the induction of pIII, the culture was infected with the CM13 helper phage which bears a kanamycin selection marker. Changing the growth media after infection to one containing  $25 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin allowed for the selective growth of cells that were infected with CM13, i.e., those that contain a TAG within the randomized peptide library.

Sequencing analysis of several clones after the first round of superinfection-selection revealed that a significant number of clones containing deleterious mutations in pIII were enriched along with those containing TAG codons. These mutations prevented the expression of full length pIII and thus allowed for superinfection by CM13. To cure the population of clones containing nonfunctional pIII, the phagemid library from the

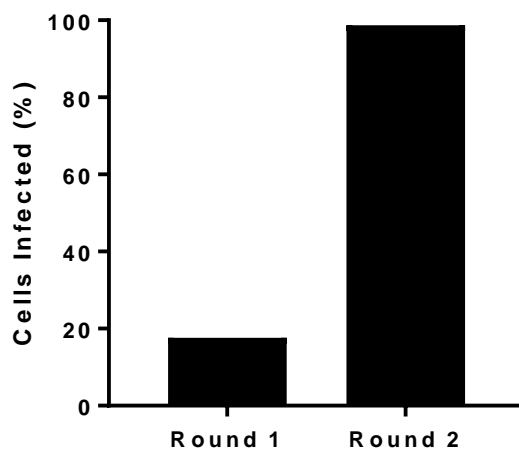


**Figure 4.4. A general strategy for constructing TAG-obligate phage display libraries.** (A) The cloned library is expressed in Top10 F' followed by superinfection with a helper phage bearing a selectable marker. Clones that express pIII are immune to superinfection. (B) Antibiotic selection allows for the selective growth of clones that were susceptible to superinfection. (C) To remove clones bearing a nonfunctional pIII gene the library is expressed in an amber suppressor strain with the help of a helper phage that lacks pIII.

first round of superinfection-selection was extracted, purified, and used to transform *E. coli* DH5 $\alpha$  containing the M13KO7 helper phage with a phenotypic knockout of pIII (M13KO7(pIII<sup>-</sup>)). DH5 $\alpha$  is a *supE* strain of *E. coli* that incorporates glutamine in response to TAG codons. Therefore, expression of the phagemid library in DH5 $\alpha$  containing M13KO7(pIII<sup>-</sup>) produces phagemid particles even if the peptide contains a

TAG codon. However, since the helperphage lacks functional pIII, the phagemid library is the only source of pIII protein. Therefore, phages are not produced when a deleterious mutation prevents the expression of pIII from the phagemid (Figure 4.4C). Expressing the library in this way allowed us to generate phagemid particles containing the 7-mer library free of deleterious mutations in pIII. Phagemid particles expressed from DH5 $\alpha$  were purified and used to infect *E. coli* Top10 F' for a second round of superinfection-selection.

Over the two rounds of superinfection-selection, TAG enrichment was monitored by quantifying the number of cells that were infected with CM13 (as determined by kanamycin resistance) as a percentage of the total number of cells in the culture. In the first round of selection, approximately 18% of the population was susceptible to superinfection. This is comparable to the number of clones expected to contain at least one TAG codon (19.9%) in a seven site library, based on a binomial distribution, using a probability of 1/32 for TAG codons. After the second round of selection, greater than 98% of the population was susceptible to superinfection indicating a significant enrichment of clones containing TAG codons (Figure 4.5). Sequencing after the second round of superinfection-selection confirmed successful TAG enrichment as all clones sequenced ( $n = 10$ ) contained a TAG codon within the random peptide sequence (Figure 4.6) In the subsequent time, dozens of clones have been sequenced from this 7-mer library and all have been found to contain a TAG codon with the random peptide sequence. Notably, all clones sequenced to date have been found to have only one TAG codon. This likely arises from the PR-bias against clones containing two ncAAs.



**Figure 4.5** The percentage of cells infected during rounds 1 and 2 of superinfection-selection. The increase in the number of cells infected from round 1 to 2 indicates enrichment of clones containing TAG.

We note that this strategy for library construction leads to a net reduction in library diversity. For example, a 7-mer library containing 21 amino acids will encode  $21^7 = 1.80 \times 10^9$  unique peptide sequences. Multiplying by the probability of finding at least one TAG codon within the peptide ( $p = 0.199$ ) we arrive at a theoretical library size of  $3.58 \times 10^8$  after superinfection-selection. Therefore, these libraries have a theoretical complexity between that of six and seven site libraries that contain only the 20 canonical amino acids. The drop in diversity decreases, however, as the peptide length increases since the probability of finding a TAG codon in longer peptides is significantly greater. We anticipate that the usefulness of libraries containing ncAAs will offset this reduction in diversity and note that libraries containing only canonical amino acids can easily be screened in parallel for final comparison after selection.

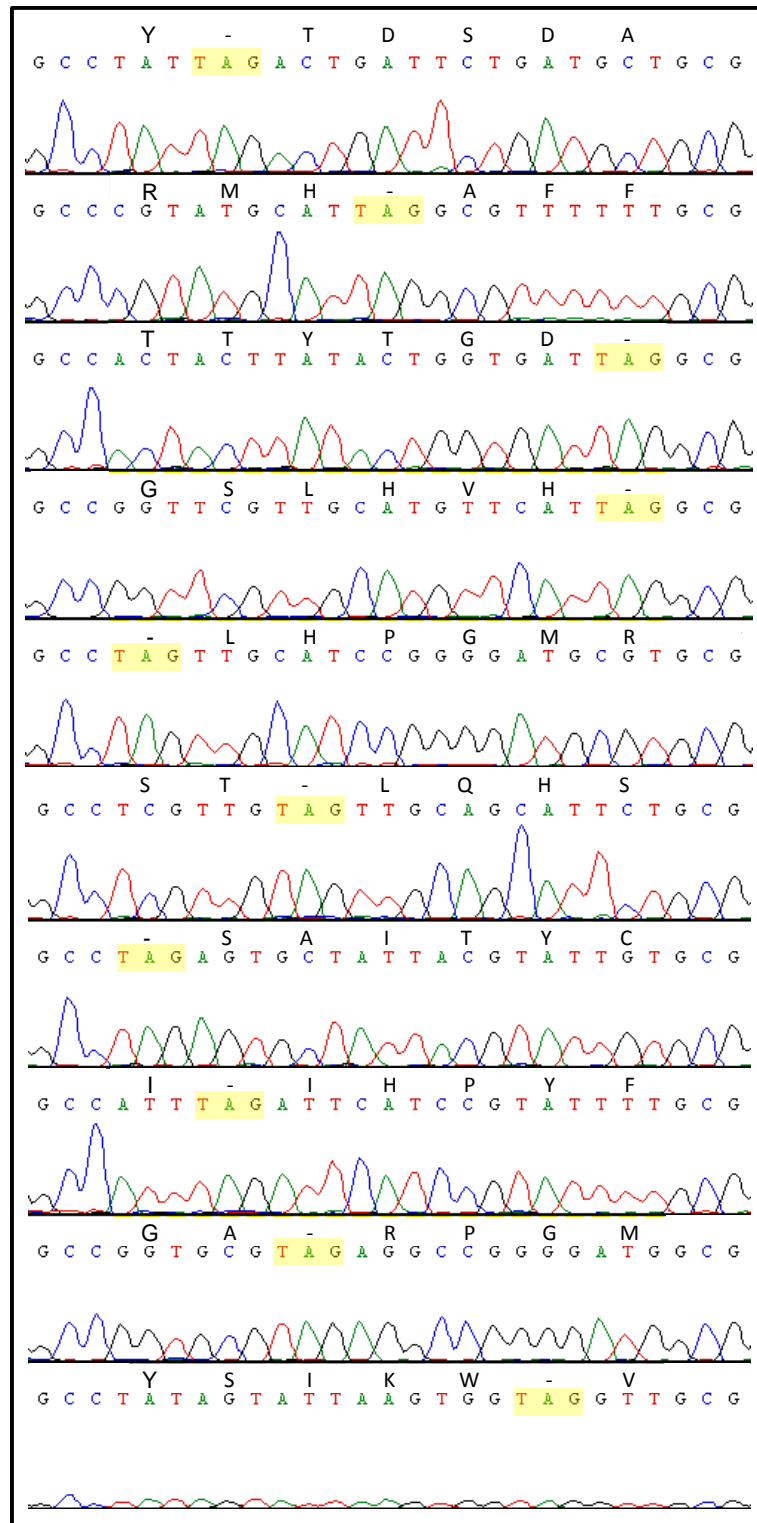
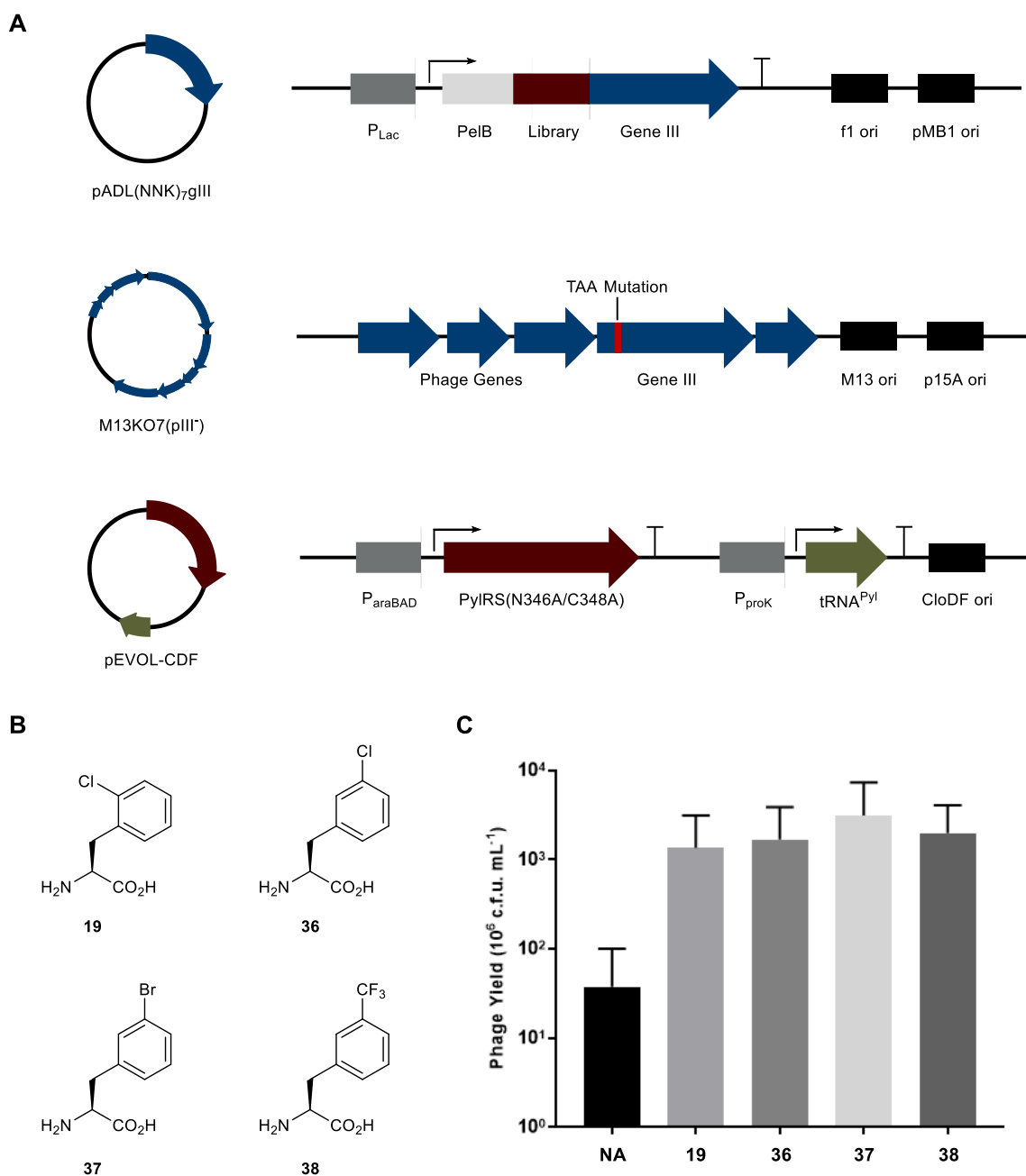


Figure 4.6 The sequences of clones isolated after the 2<sup>nd</sup> round of superinfection-selection

#### 4.3.2 Genetic Incorporation of ncAAs into Phage-Displayed Peptides

After constructing the TAG-obligate 7-mer library we next tested the expression of the library with various ncAAs. We first tested the incorporation of four phenylalanine derivatives (Figure 4.7B). To incorporate these ncAAs into phage-displayed peptides we utilized a previously reported polyspecific pyrrolysyl-tRNA synthetase (PylRS) mutant containing the N346A and C348A mutations (PylRS(N346A/C348A) in conjunction with its amber suppressor tRNA ( $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$ ) that were described in Chapter 3. These amino acids were chosen because they are all incorporated by the same PylRS mutant but with varying incorporation efficiencies. Using a sfGFP reporter protein with an amber mutation at position S2, **19** was found to be one of the best substrates for PylRS(N346A/C348A) yielding  $40 \text{ mg}\cdot\text{mL}^{-1}$  of purified protein.<sup>184</sup> By contrast, **36** represents a modest substrate for PylRS(N346A/C348A) with an incorporation efficiency 2.5-fold lower than **19**.<sup>147</sup> The commonly used plasmid for aaRS expression (pEVOL) utilizes the same p15A origin of replication as M13KO7(pIII<sup>-</sup>). To ensure compatibility of pEVOL with M13KO7(pIII<sup>-</sup>), we first constructed a derivative of pEVOL (pEVOL-CDF) in which the p15A origin was replaced with a CloDF origin. CloDF maintains approximately the same number of copies of the plasmid per cell, yet is compatible with the p15A origin of M13KO7(pIII<sup>-</sup>) and the pMB1 origin of the phagemid library.<sup>185</sup>

To express the phagemid library containing phenylalanine derivatives, *E. coli* Top10 were transformed with the 7-mer TAG-obligate phagemid library, M13KO7(pIII<sup>-</sup>



**Figure 4.7 The genetic incorporation of phenylalanine derivatives into phage-displayed peptides. (A)** A diagram of the three plasmid system used to express phenylalanine derivatives. **(B)** The structure of the phenylalanine derivatives incorporated in this study. **(C)** The yield of phages expressed in the presence of phenylalanine derivatives or in the absence of an nCAA (NA). The yield is displayed in millions of colony forming units per milliliter of culture. Error bars represent positive standard deviations of three independent expressions.



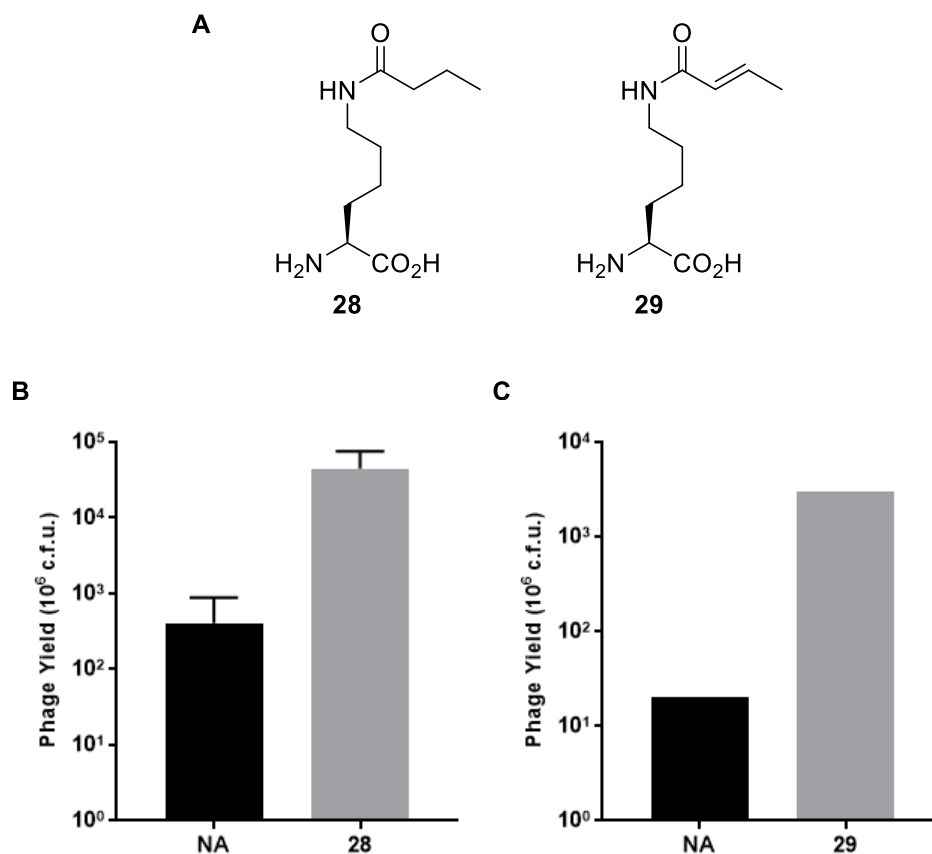
), and pEVOL-CDF-PylIT-PylRS(N346A/C348A) (Figure 4.7A). Growing cells in 2xYT media supplemented with 4 mM of one of the four phenylalanine derivatives resulted in a 36 to 83-fold increase in the number of phages produced as determined by a colony forming unit assay (Figure 4.7C). The increase in phage yield is consistent with suppression of the TAG codon by tRNA<sub>CUA</sub><sup>Pyl</sup> and indicates successful incorporation of the ncAAs. Background phage produced in the absence of an ncAA is likely due to phenylalanine incorporation as we have shown that PylRS(N346A/C348A) facilitates the incorporation of phenylalanine at low levels when used in rich media.<sup>147</sup> However, as discussed in Chapter 3, we have also shown that the level of phenylalanine incorporation by PylRS(N346A/C348A) decreases in the presence of a preferred ncAA substrate.<sup>184</sup> This implies that phenylalanine incorporation into peptides expressed in the presence of an ncAA will be minimal.

Interestingly, although the phage yield was significantly lower than that of wildtype M13, the yield did not correlate with the previously observed incorporation efficiency of the ncAAs. This suggests that, in this system, phage production may be limited by something other than stop codon suppression efficiency. For example, the negative impact of global TAG reassignment in the host cell could contribute to the low phage yield. If this is the case one would expect that increasing TAG suppression efficiency may actually be detrimental to viral fecundity and could actually contribute to PR-bias. This could be circumvented by using *E. coli* strains that have had TAG codons completely removed from their genome, such as those that have been recently reported.<sup>186</sup> However, with the low accuracy of the colony forming unit assay used to

quantify phages it is difficult to compare differences in expression yield of only 2 to 3-fold.

Next, we tested the incorporation of several lysine derivatives into the phagemid library. We have recently reported a PylRS mutant containing the Y384W mutation (BuKRS) that can facilitate the efficient incorporation of  $N_\epsilon$ -butyryl-lysine (**28**) and  $N_\epsilon$ -crotonyl-lysine (**29**) in response to amber codons in *E. coli*.<sup>177</sup> Both **28** and **29** are naturally occurring lysine posttranslational modification and their incorporation into phage-displayed peptides could prove useful for studying the enzymes that interact with these modifications.<sup>187, 188</sup> Expressing the phagemid library in *E. coli* Top10 containing the plasmids pEVOL-CDF-PylT-BuKRS and M13KO7(pIII<sup>r</sup>) and in the presence of 5 mM **28** or **29** led to greater than a 100-fold increase in phage yield compared to when no ncAA was added to the growth media indicating successful incorporation of both ncAAs (Figure 4.8).

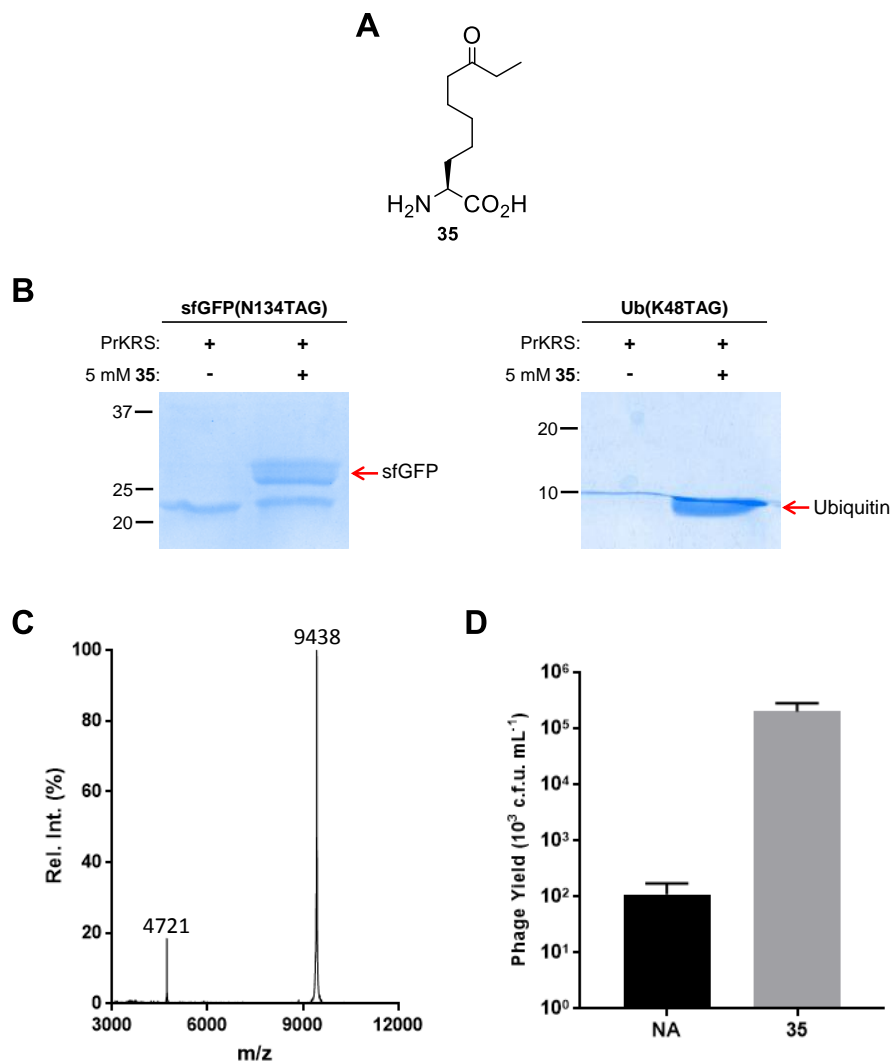
We also tested the incorporation of a new ketone-containing amino acid (**35**) into the phage library. We have recently reported a PylRS mutant (PrKRS) that can facilitate the incorporation of  $N_\epsilon$ -propionyl-lysine in response to TAG codons in *E. coli*. To test the ability of PrKRS to also recognize **35** as a substrate, we utilized two model proteins, namely, superfolder green fluorescent protein (sfGFP) containing a TAG mutation at position 134 (sfGFP-N134→X) and ubiquitin with a TAG mutation at position 48 (Ub-K48→X). Expression of full length sfGFP-N134→X and Ub-K48→X in *E. coli* that were concurrently expressing PrKRS and tRNA<sub>CUA</sub><sup>Pyl</sup> was dependent on the addition of



**Figure 4.8** The genetic incorporation of  $N_{\epsilon}$ -butyryl-lysine and  $N_{\epsilon}$ -crotonyl-lysine into phage-displayed peptides. (A) The structures of  $N_{\epsilon}$ -butyryl-lysine (**28**) and  $N_{\epsilon}$ -crotonyl-lysine (**29**). (B) The yield of phages expressed in the presence and absence (NA) of **28**. Error bars represent positive standard deviations of four independent expressions. (C) The yield of phages expressed in the presence and absence (NA) of **29**. Values on the Y-axis represent the total yield, obtained from 50 mL cultures, and are displayed in millions of colony forming units.

5 mM of **35** to the growth media (Figure 4.9B). MALDI-TOF mass spectrometry of purified Ub-K48→**35** displayed one major peak with a mass that agreed well with the theoretical value for ubiquitin with **35** at position K48 (Figure 4.9C). Together, these results demonstrate the ability of PrKRS to recognize **35** as a substrate and facilitate its incorporation in response to TAG codons in *E. coli*. When PrKRS and  $\text{tRNA}_{\text{CUA}}^{\text{Pyl}}$  were expressed in *E. coli* containing the 7-mer phagemid library and M13KO7(pIII), we

observed an increase in phage yield of approximately 100-fold that was dependent on the addition 5 mM of **35** to the growth media (Figure 4.9D). This result indicates successful incorporation of **35** into the phage-displayed peptide library.



**Figure 4.9. The genetic incorporation of 2-amino-8-oxodecanoic acid.** (A) The structure of 2-amino-8-oxodecanoic acid (**35**). (B) The incorporation of **35** at the N134 position of sfGFP and the K48 position of ubiquitin. (C) MALDI-TOF mass spectrum of Ub-48→**35**. Calculated mass:  $[M+H]^+ = 9442$  Da; Observed Mass = 9438 Da. (D) The yield of phages expressed in the presence and absence (NA) of **35**. The yield is displayed in thousands of colony forming units per milliliter of culture. Error bars represent positive standard deviations of four independent expressions.

### 4.3.3 Affinity Selection of Peptides Containing ncAAs

Next, we asked whether the TAG-obligate library could withstand iterative rounds of selection and amplification. Library integrity over iterative rounds is a fundamental prerequisite for evolution via phage display. If random mutations lead to the emergence of clones containing only canonical amino acids, the PR-bias in favor of these sequences would likely cause them to become the dominant clones during selection.

To test this we performed model selections against two commonly used laboratory proteins. We first performed selection against the Tobacco Etch Virus (TEV) protease using the phagemid library that had been expressed with either **37** or **38**. To select for peptides that bind to TEV protease, the purified phagemid library was incubated with the immobilized, biotinylated, protein for 30 minutes at room temperature. Following incubation and washing the bound phages were eluted non-specifically with a low pH buffer. Eluted phagemids were amplified and sequenced ( $n = 6$ ). Sequencing of the eluted phagemids revealed that in both selections the population had converged onto one or two sequences (Figure 4.10). The most abundant sequences in both selections consisted of highly hydrophobic peptides with four of the seven residues having aromatic side chains. Although the TEV protease recognition sequence (ENLYFQS) contains two neighboring aromatic residues, given the sequence specificity of this protease it is unlikely that the selected peptides bind to the substrate binding pocket of the enzyme.<sup>189, 190</sup> Peptides containing a high number of aromatic residues have been routinely isolated in phage display selections. Generally these peptides are unrelated to the target protein and many have been shown to bind to the plastic tubes or

plates used during selection.<sup>59</sup> Such target-unrelated peptides (TUPs) can often be avoided by including a round of negative selection as is described for the selections in Chapter 5.<sup>191</sup> However, notably, each of the peptides identified from both selections contained an in-frame TAG codon after three rounds of amplification and selection.

### TEV Protease Selection

#### ncAA 37

V	Y	F	F	X	D	E
V	Y	F	F	X	D	E
V	Y	F	F	X	D	E
V	Y	F	F	X	D	E
D	Y	F	F	X	D	S
D	Y	F	F	X	D	S

#### ncAA 38

V	Y	A	X	Y	I	F
V	Y	A	X	Y	I	F
V	Y	A	X	Y	I	F
V	Y	A	X	Y	I	F
X	V	G	L	F	S	I
S	Q	R	X	E	D	H

X ≡ ncAA

### Streptavidin Selection

#### Round 3

K	X	N	F	G	Y	Y
K	W	K	V	X	T	S
T	W	X	K	S	N	W
M	A	K	P	Q	R	X
R	E	H	K	P	X	N
K	L	X	K	H	Y	P

#### Round 4

K	L	X	K	H	Y	P
K	L	X	K	H	Y	P
K	L	X	K	H	Y	P
K	L	X	K	H	Y	P
K	H	X	K	A	I	P
K	M	X	P	Q	R	N
K	M	X	P	Q	R	N
K	M	X	P	Q	R	N
K	M	X	P	Q	R	N
K	M	X	P	Q	R	N
K	M	X	P	Q	R	N
K	M	X	P	Q	R	N
K	W	C	X	N	C	R
S	A	A	K	T	X	I

**Figure 4.10** The amino acid sequence of peptides containing phenylalanine derivatives isolated from selections against TEV protease and streptavidin. The position of the TAG codon within the sequence corresponds to the position of the ncAA and is shown with a blue X.

To rule out the possibility that the sequences identified from selection were obtained due to some PR-bias in favor of highly hydrophobic peptides, we performed an additional model selection against the common laboratory protein streptavidin. During this selection we altered the immobilization and elution strategy to rule out the possibility that hydrophobic peptides were enriched due to binding to the solid support. To select for peptides binding to streptavidin, the phagemid library expressed with **37** was incubated with the target protein that had been non-specifically adsorbed onto polystyrene plates. After incubation, the plate was washed and bound phages were competitively eluted with D-biotin (0.1 mM). Eluted phagemids were amplified and sequenced after the third round of selection ( $n = 6$ ). By the third round the population had not converged onto a clear consensus. Therefore, we performed a fourth round of amplification and selection. Sequencing of phagemids after the fourth round of selection ( $n = 12$ ) revealed that the majority of sequenced clones had converged onto two sequences (Figure 4.10). The most dominant sequence (KMXPQRN, where X = **37**) contained a fragment of a known streptavidin-binding sequence (HPQ).<sup>192</sup> A different clone that contained the PQR motif was also observed in round three which suggests specific enrichment of the PQR sequence. Notably, no peptides were isolated with greater than one aromatic residue (aside from the ncAA) further supporting our suspicion that the previously isolated peptides were TUPs that were enriched due to binding to the plastic support.

#### 4.4 Conclusion

In summary, we have developed a novel method for generating TAG-obligate phage display peptide libraries. By eliminating the more quickly replicating phage clones that contain only canonical amino acids, this type of library eliminates propagation-related bias that has previously interfered with selections involving ncAAs. Using this TAG-obligate library, we have vastly expanded the number of ncAAs that have been incorporated into phage-displayed peptides. With this work, the number of ncAAs that have been incorporated into phage-displayed peptides has increased from 8 to 15. These ncAAs include two naturally occurring lysine posttranslational modifications, and the brand new amino acid **35** which contains a ketone reactive handle. Finally, we demonstrated that libraries generated through superinfection-selection are robust and can withstand iterative rounds of amplification and selection to identify peptide sequences that bind to a given target.

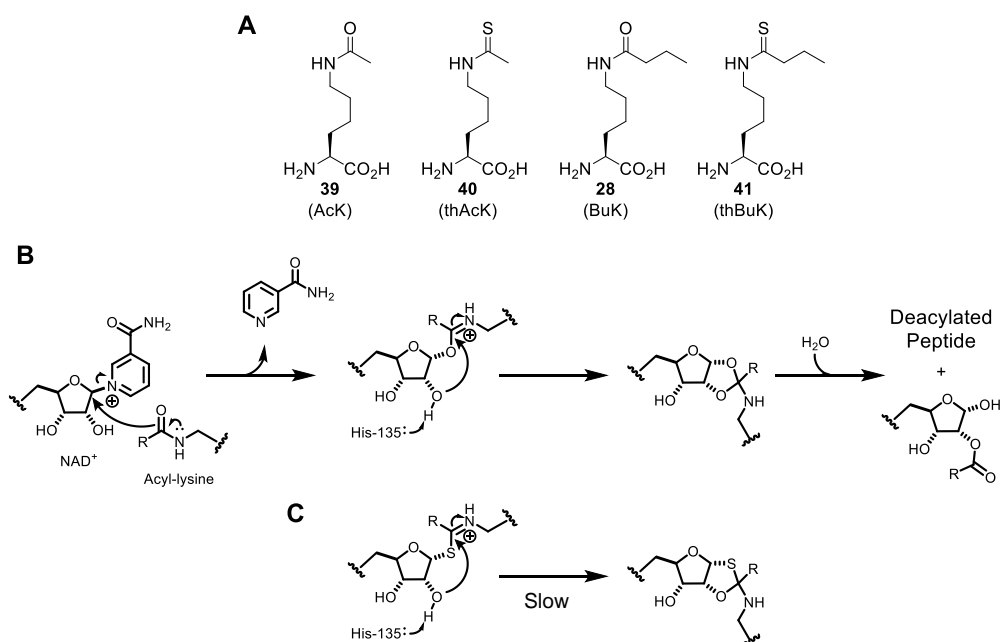


## CHAPTER V

### PHAGE SELECTION OF PEPTIDES CONTAINING NON-CANONICAL AMINO ACIDS FOR THE INHIBITION OF SIRTUIN 2

#### 5.1 Introduction

Sirtuin 2 (SIRT2) is one of the seven mammalian lysine deacylases, collectively referred to as the sirtuins, that are homologs of the yeast gene silencing protein, silent information regulator 2.<sup>193</sup> SIRT2, like all sirtuins, is an NAD<sup>+</sup>-dependent deacylase that catalyzes the conversion of N<sub>ε</sub>-acyl-lysine and NAD<sup>+</sup> to lysine and 2'-O-acyl-ADP-ribose (Figure 5.1).<sup>194, 195</sup> Predominantly localized in the cytoplasm, SIRT2 was originally categorized as a tubulin deacetylase.<sup>196</sup> However, numerous studies have shown that SIRT2 displays robust deacetylase activity towards several other protein substrates, both *in vitro* and *in vivo*, including histones<sup>197</sup> and nucleosomes<sup>198</sup> and several transcription factors including p53<sup>199, 200</sup> and Foxo1<sup>201</sup>. Further, SIRT2's deacylase activity is not limited to acetylation. Indeed, SIRT2 catalyzes the removal of numerous lysine acylations including propionyl-, butyryl-, crotonyl-, hexanoyl-, octanoyl-, decanoyl-, dodecanoyl-, and myristoylation.<sup>202</sup> These lysine modifications serve as important regulators of both the function and subcellular localization of proteins.<sup>203, 204</sup> Therefore, in its capacity as a lysine deacylase, SIRT2 has been implicated as a key regulator of numerous cellular processes from metabolism to DNA repair and aging.<sup>205, 206</sup>



**Figure 5.1. The mechanism of sirtuin-catalyzed lysine deacylation and inhibition by thioamides.** (A) The structures of  $N_{\epsilon}$ -acetyl-lysine (**39**),  $N_{\epsilon}$ -butryl-lysine (**28**), and their thioamide analogs, **40** and **41**, respectively. (B) The proposed catalytic mechanism of sirtuin-catalyzed lysine deacylation. (C) lysine thioamides undergo the same deacetylation mechanism as lysine amides, but form stalled thioimidate intermediates.

Given the important role of SIRT2 in regulating diverse cellular processes, this enzyme has emerged as a target for therapeutic intervention in human diseases. In terms of cancer, SIRT2 has been classified as a tumor suppressor, largely due to the observation that aging SIRT2 null mice have a higher incidence of aggressive tumor formation.<sup>207, 208</sup> However, SIRT2 has been observed to be upregulated in certain cancer types including hepatocellular carcinoma,<sup>209, 210</sup> non-small cell lung cancer,<sup>211</sup> and breast tumors. In the latter, high levels SIRT2 were associated with reduced time until disease-related death.<sup>212</sup> In addition, several SIRT2 inhibitors have been shown to have broad anticancer activity, validating SIRT2 as a target for therapeutic intervention in cancer.<sup>199, 213, 214</sup> SIRT2 inhibitors also show promise in the treatment of age-related

neurodegenerative diseases. As early as 2007 it was demonstrated that SIRT2 inhibitors protect against neuronal death in response to  $\alpha$ -synuclein in cellular and *Drosophila* models of Parkinson's Disease.<sup>215</sup> Recently it was shown that this neural protection is a result of inhibition of SIRT2-catalyzed deacetylation of K6 and K10 on  $\alpha$ -synuclein.<sup>216</sup> Further, it has been shown that inhibition of SIRT2 displays neuroprotective effects in a *Drosophila* model of Huntington's disease.<sup>217</sup>

Given the possibility of SIRT2 inhibitors to be used in the treatment of these diseases, there has been widespread interest in the development of potent, SIRT2-selective inhibitors. Through this search it has been shown that lysine thioamides, such as *N* $\epsilon$ -thioacetyl-lysine (**40**, thAcK; Figure 5.1A), which mimics the natural acetyl-lysine substrate, are potent deacetylase inhibitors.<sup>218</sup> Indeed, peptides based on the sequence of p53, histone H3, or  $\alpha$ -tubulin containing thAcK inhibit SIRT1-3 with IC<sub>50</sub> values in the low micromolar to high nanomolar range.<sup>218-220</sup> Pseudopeptide inhibitors containing thAcK have been shown to have antiproliferative effects on cancer cells and in animal models of human cancer.<sup>221, 222</sup> Kinetic and biochemical studies have revealed that thAcK is a mechanism-based sirtuin inhibitor that undergoes the same deacetylation mechanism as acetyl-lysine but forms a stalled 1'-*S*-alkylimidate intermediate after reacting with NAD<sup>+</sup> (Figure 5.1C).<sup>219</sup> This makes thAcK inhibition selective for sirtuins which are the only deacetylases known to use NAD<sup>+</sup>.<sup>223</sup>

Recently, it has been shown that altering the carbon chain length of the thioamide modification on lysine can lead to increased potency and sirtuin isoform selectivity of thioamide-based inhibitors. For example, lysine thioamides containing four, six, and

twelve carbon chains were significantly more potent than thAcK and displayed greater selectivity for SIRT2 over SIRT1 and SIRT3.<sup>214</sup> This selectivity arises from the fact that the sirtuins show differing activity towards lysine acyl groups based on the carbon chain length.<sup>202</sup> While sirtuins show a strong preference for acyl group identity, their preference for peptide sequences is less pronounced.<sup>224</sup> However, some studies have indicated that the identity of residues adjacent to thAcK can have a considerable impact on sirtuin inhibition.<sup>220, 221</sup> Herein, we utilize phage display to screen a library of greater than  $10^9$  peptides containing  $N_\epsilon$ -butyryl-lysine (**28**, BuK) to select for peptide sequences that bind to SIRT2. Following three rounds of selection, the phage population converged on clones containing conserved residues immediately adjacent to the butyryl-lysine which may point to subtle substrate preferences of SIRT2. We further demonstrate that these peptides, when synthesized with  $N_\epsilon$ -thiobutyryl-lysine (**40**, thBuK), a thioamide analog of BuK, are potent inhibitors of SIRT2-catalyzed lysine deacetylation.

## 5.2 Experimental Details

### 5.2.1 Miscellaneous Information

All primers were purchased from Integrated DNA Technologies, Inc. (Coralville, PA). Enzymes and buffers for molecular cloning were purchased from New England Biolabs (Ipswich, MA). DNA sequencing was performed by Eton Biosciences, Inc. (San Diego, CA). Mass spectra were collected by the Department of Chemistry Mass Spectrometry Facility at Texas A&M University.

### 5.2.2 Primer List

BuKRS-N346A-F: 5'-GCGTTTTGTCAGATGGGTTCTG-3'

BuKRS-N346A-R: 5'-CAGCATAGTGAACTCTCCAGGTG-3'

BuKRS-N346/C348 Lib-F: 5'-NNKTTTTGTCAGATGGGTTCTGGCT-3'

BuKRS-N346/C348 Lib-R: 5'-NNKTTTNNKCAGATGGGTTCTGGCTG-3'

### 5.2.3 Plasmid Construction

pEVOL-PylT-BuKRS(N346A): To generate the N346A mutation in BuKRS the previously reported<sup>177</sup> plasmid pEVOL-PylT-BuKRS was PCR amplified with the primer pair BuKRS-N346A-F and BuKRS-N346A-R. The PCR product was phosphorylated with T4 PNK, ligated with T4 DNA ligase, and used to transform *E. coli* Top10.

BuKRS-N346/C348 Library: To generate the two site library in BuKRS, the plasmid pEVOL-PylT-BuKRS was PCR amplified with the primer pair BuKRS-N346/C348 Lib-F and BuKRS-N346/C348 Lib-R. The PCR product was phosphorylated with T4 PNK, ligated with T4 DNA ligase, and used to transform *E. coli* Top10. Transformed cells were plated on LB agar plates containing 34  $\mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol and grown at 37°C overnight. The following day, colonies were harvested from the plate, combined, and the plasmids extracted using a commercial plasmid extraction kit (Epoch).

#### 5.2.4 Protein Expression and Purification

*sfGFP-N134* → X: *E. coli* Top10 containing the plasmid pBAD-*sfGFP-N134*(TAG) and pEVOL-PylT-BuKRS were grown at 37°C in 2xY media supplemented with 100 µg·mL<sup>-1</sup> ampicillin, and 34 µg·mL<sup>-1</sup> chloramphenicol to an OD<sub>600</sub> of 0.5-0.6 at which point protein expression was induced with the addition of 2 mM BuK or thBuK. 16 h after induction, cells were harvested (4k rpm, 20 min) and cell pellets were stored at -80°C until purification. Proteins were purified via nickel affinity chromatography exactly as described in Chapter 3. Purified proteins were analyzed via SDS-PAGE and Western blot using a pan anti-butyryl-lysine antibody (PTM Biolabs Inc.).

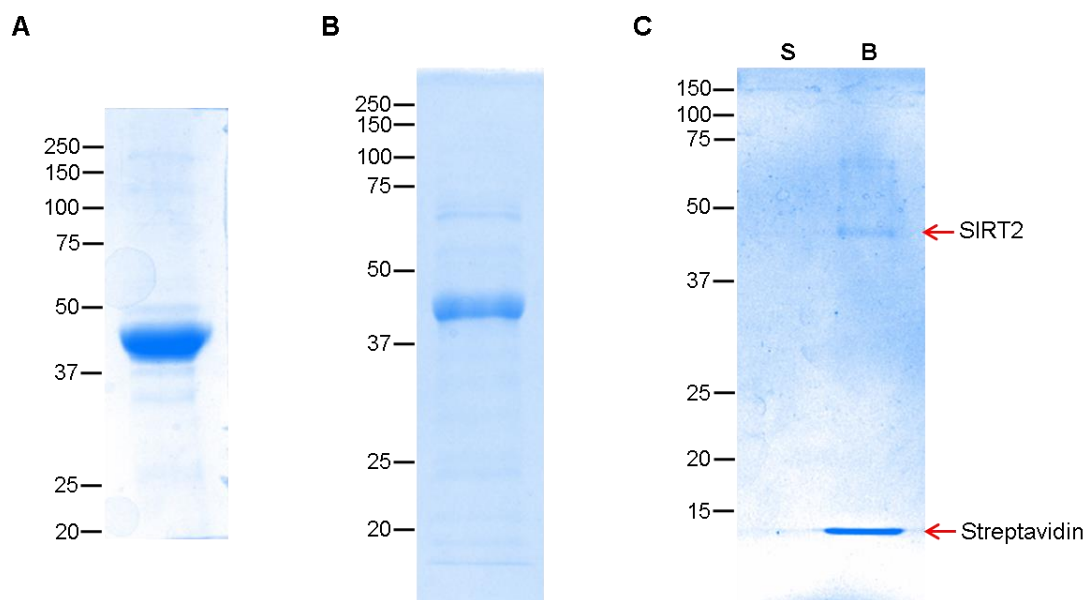
*Sirtuin 2*: The plasmid pHEX-His<sub>6</sub>-SIRT2 was a kind gift from Professor John Denu at the University of Wisconsin School of Medicine and Public Health. *E. coli* Top10 containing the plasmid pHEX-His<sub>6</sub>-SIRT2 were grown at 37°C in 2 L of 2xYT containing 100 µg·mL<sup>-1</sup> ampicillin to OD<sub>600</sub> of 0.6 at which point SIRT2 expression was induced by the addition of 0.1 mM IPTG. 2 h after induction the cells were harvested by centrifugation (6k rpm, 20 min) and the cell pellets were stored at -80°C until purification.

Cell pellets were thawed and resuspended in 40 mL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.1 mM phenylmethanesulfonyl fluoride (PMSF)). The resuspended pellet was incubated in the presence of lysozyme (1 mg·mL<sup>-1</sup>, chicken egg white) for 30 min on ice. Following incubation, the resuspension was sonicated twice (1 s on, 1 s off, 1 min total, 60% output) and clarified by

centrifugation (10k rpm, 45 min). To the clarified supernate was added 4 mL of a 50% slurry of high affinity Ni<sup>2+</sup>-charged resin (Genscript) and the mixture was incubated at 4°C with end-over-end rotation for 30 min. After 30 min, the mixture was filtered through a disposable column, the resin was washed with 40 mL of lysis buffer, and the bound protein was eluted with 7 mL of elution buffer (lysis buffer containing 250 mM imidazole). Subsequent purification steps vary depending on whether the protein was used for biotinylation and phage panning or SIRT2 inhibition assays.

For phage panning, the eluted protein was concentrated to 1 mL by ultrafiltration (10 kDa MWCO) diluted 1:9 with Q Sepharose buffer A (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 50 mM NaCl, 0.1 mM DTT) and concentrated again to 1 mL. The concentrated sample was applied to a 25 mL Q Sepharose column (GE Healthcare) that was pre-equilibrated with Q Sepharose buffer A. The column was washed with 2 column volumes (CV, 2 mL·min<sup>-1</sup>) of buffer A and the proteins were eluted with a linear gradient from 0-100% buffer B (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 1 M NaCl, 0.1 mM DTT) over 3 CV (2 mL·min<sup>-1</sup>). Fractions from the main peak were combined and concentrated to 1 mL for biotinylation. SDS-PAGE of SIRT2 purified using this protocol is shown in Figure 5.2B.

For inhibition assays, the eluted protein was loaded onto a HiPrep 26/10 desalting column (GE Healthcare) that had been pre-equilibrated with Source 15Q buffer A (20 mM Tris pH 8.0, 50 mM NaCl, 0.2 mM DTT, 10% glycerol). The desalted protein was concentrated by ultrafiltration to 1 mL (10 kDa MWCO) and applied to a 10 mL Source 15Q column (GE Healthcare) pre-equilibrated with Source 15Q buffer A. The



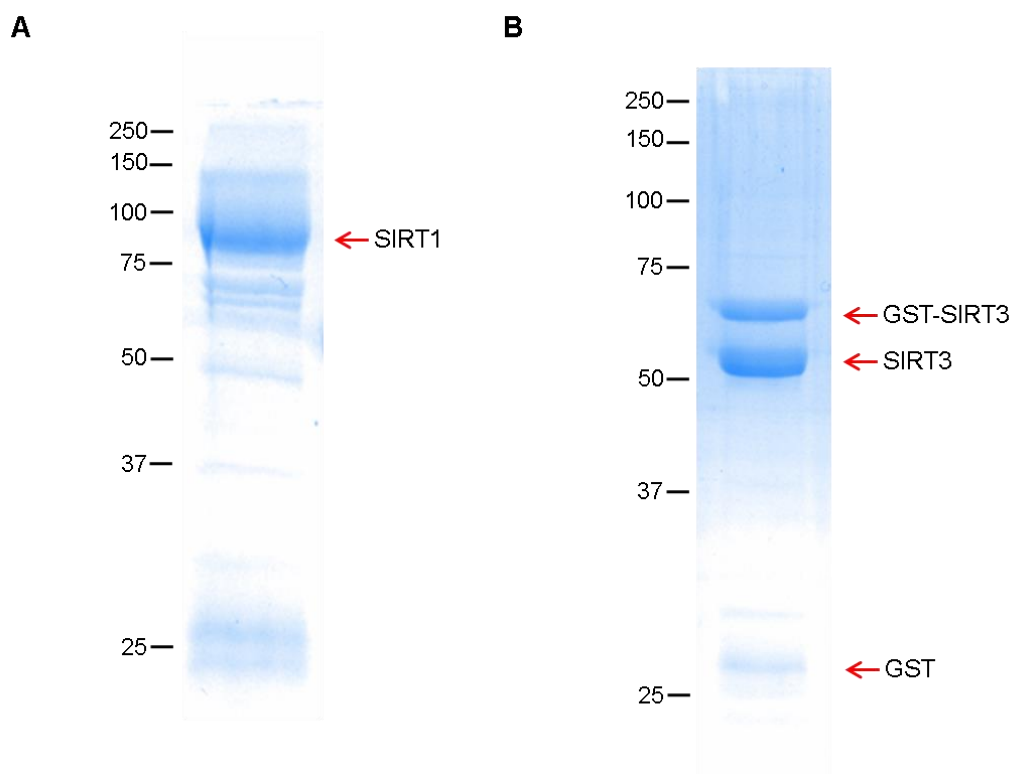
**Figure 5.2. SDS-PAGE analysis of purified sirtuin 2.** (A) SIRT2 purified for activity assays. (B) Biotinylated SIRT2 purified for phage panning. (C) Pull-down assay to confirm biotinylation of SIRT2. The lane labelled “S” denotes supernate whereas “B” denotes beads. Masses of a standard ladder are labelled.

column was washed with 3 CV ( $2 \text{ mL} \cdot \text{min}^{-1}$ ) of Source 15Q buffer A and then eluted with a linear gradient from 0-60% Source 15Q buffer B ( $20 \text{ mM Tris pH } 8.0, 1 \text{ M NaCl}, 0.2 \text{ mM DTT}, 10\% \text{ glycerol}; 1 \text{ mL} \cdot \text{min}^{-1}$ ) over 5 CV. The fractions containing pure SIRT2 were concentrated to  $40\text{-}70 \mu\text{M}$ , flash frozen in  $10 \mu\text{L}$  aliquots, and stored at  $-80^\circ\text{C}$ . Enzyme concentration was determined from the absorbance at  $280 \text{ nm}$  using the calculated extinction coefficient of  $32,470 \text{ M}^{-1} \cdot \text{cm}^{-1}$ . SDS-PAGE of SIRT2 purified using this protocol is shown in Figure 5.2A.

*Sirtuin 1:* The plasmid pQE80-His<sub>6</sub>-SIRT1 was a kind gift from Professor John Denu at the University of Wisconsin School of Medicine and Public Health. *E. coli* BL21(DE3)



containing the plasmid pQE80-His<sub>6</sub>-SIRT1 were grown at 37°C in 3 L of 2xYT media supplemented with 100 µg·mL<sup>-1</sup> ampicillin to an OD<sub>600</sub> of 0.7 at which point protein expression was induced with the addition of 0.5 mM IPTG. 4 hours post induction the cells were harvested by centrifugation (4k rpm, 20 min) and cell pellets were stored at -80°C until purification. Cell pellets were resuspended in 50 mL of lysis buffer (50 mM Tris pH 8.0, 250 mM NaCl, 10 mM imidazole, 0.1 mM PMSF, 1% Triton X-100) and incubated with 1 mg·mL<sup>-1</sup> lysozyme for 30 min on ice. Following incubation, the resuspension was sonicated twice (1 s on, 1 s off, 1 min total, 60% output) and clarified by centrifugation (10k rpm, 45 min). To the clarified supernate was added 4 mL of a 50% slurry of high affinity Ni-charged resin (Genscript) and the mixture was incubated at 4°C with end-over-end rotation for 30 min. After 30 min, the mixture was filtered through a disposable column, the resin was washed with 40 mL of lysis buffer containing 0.1% Triton X-100, followed by 40 mL of lysis buffer containing 0% Triton X-100. The bound protein was eluted with 7 mL of elution buffer (lysis buffer containing 250 mM imidazole). The eluted protein was then desalted and purified on a Source 15Q column exactly as described for SIRT2. The fractions containing SIRT1 were concentrated to 19 µM, flash frozen in 25 µL aliquots, and stored at -80°C. Enzyme concentration was determined from the absorbance at 280 nm using the calculated extinction coefficient of 40,340 M<sup>-1</sup>·cm<sup>-1</sup>. SDS-PAGE of SIRT1 purified using this protocol is shown in Figure 5.3A.



**Figure 5.3. SDS-PAGE analysis of purified sirtuin 1 and GST-sirtuin 3.** (A) SDS-PAGE analysis of SIRT1. (B) SDS-PAGE analysis of GST-SIRT3. Masses of a standard ladder are labelled.

*GST-Sirtuin 3*: *E. coli* Top10 containing the plasmid pGEX4T-GST-SIRT3 were grown at 37°C in 2 L of 2xYT media supplemented with 100 µg·mL<sup>-1</sup> ampicillin to an OD<sub>600</sub> of 0.6 at which point protein expression was induced with the addition of 0.2 mM IPTG. 2 hours post induction the cells were harvested by centrifugation (4k rpm, 20 min) and cell pellets were stored at -80°C until purification. Cell pellets were resuspended in 50 mL of lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.1 mM PMSF, 1% Triton X-100) and incubated in the presence of 1 mg·mL<sup>-1</sup> of lysozyme for 30 minutes on ice. Following incubation, the resuspension was sonicated twice (1 s on 1 s off, 1 min total, 60% output) and clarified by centrifugation (10k rpm, 45 min). To the clarified supernatant was added

3 mL of a 50% slurry of immobilized glutathione resin (Genscript) and the mixture was incubated at 4°C for 30 min with end-over-end rotation. After 30 min the mixture was filtered through a disposable column, and the resin was washed with 50 mL of lysis buffer containing 0.1% Triton X-100. After washing, the bound protein was eluted with 10 mL of lysis buffer containing 0% Triton X-100 and 10 mM of reduced glutathione. The elution was directly applied to a HiPrep 26/10 desalting column (GE Healthcare) that had been pre-equilibrated with 50 mM Tris pH 7.5, 150 mM NaCl, 0.2 mM DTT, and 10% glycerol. The desalted protein was concentrated to 16  $\mu$ M via ultrafiltration (10 kDa MWCO), flash frozen, and stored in 25  $\mu$ L aliquots at -80°C. Enzyme concentration was determined from the absorbance at 280 nm using the calculated extinction coefficient of 80,790  $M^{-1}\cdot cm^{-1}$ . SDS-PAGE analysis of GST-SIRT3 purified using this protocol is shown in Figure 5.3B.

#### *5.2.5 Biotinylation of Sirtuin 2*

SIRT2 (53  $\mu$ M) was biotinylated with a 10-fold molar excess of EZ-Link™ NHS-Biotin (Thermo Fisher Scientific) for 2 h on ice with gentle rocking. After 2 h the solution was directly loaded onto a Superdex 75 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated with 50 mM Tris pH 8.0, 250 mM NaCl, 0.2 mM DTT, 10% glycerol. The eluted protein was concentrated by ultrafiltration (10 kDa MWCO) to 10  $\mu$ M, flash frozen in 10  $\mu$ L aliquots, and stored at -80°C. Biotinylation of the protein was confirmed in two ways. In method one biotin was detected via Western blot using a streptactin-HRP conjugate (Bio-Rad). In method two biotinylation was confirmed by capture on

magnetic streptavidin beads (Genscript) using 2  $\mu\text{g}$  of biotinylated SIRT2 as described previously.<sup>225</sup> SDS-PAGE of the streptavidin pulldown is shown in Figure 5.2C. Enzymatic activity of the biotinylated SIRT2 was confirmed using the Boc-Lys(Ac)-AMC substrate as describe below.

#### *5.2.6 BuKRS Library Screening*

*E. coli* Top10 were co-transformed with the BuKRS library and the plasmid pBAD-sfGFP-N134(TAG) which contains the gene encoding superfolder green fluorescent protein with a TAG mutation at position 134. Transformed cells were plated on LB agar plates containing 0.1% arabinose, 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, and 34  $\mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol and grown at 37°C overnight. The following day, fluorescent colonies were picked and grown in 200  $\mu\text{L}$  2xYT for 2 h. Individual cultures were then spotted in duplicate on LB and GMMML agar plates with and without 2 mM thBuK and 0.1% arabinose. The plates were examined for fluorescent spots after 1 (LB agar) or 3 (GMMML agar) days at 37°C.

#### *5.2.7 In-Cell Fluorescence Measurements*

*E. coli* Top10 containing the plasmids pBAD-sfGFP-N134(TAG) and pEVOL-PylT-BuKRS were used to express sfGFP-134 $\rightarrow$ X as described above with varying concentrations of thBuK and BuK. 8 h after induction, cell cultures were diluted to  $\text{OD}_{600} = 0.9\text{-}1.1$  and the fluorescence of the GFP chromophore was measured in black

96-well plates using a BioTek Synergy H1 microplate reader at  $\lambda_{\text{Ex}} = 485 \text{ nm}$ ,  $\lambda_{\text{Em}} = 525 \text{ nm}$ .

#### *5.2.8 Phagemid Library Construction and Transformation*

Construction and transformation of the TAG-obligate phagemid library for the expression of phage-displayed peptides containing thBuK and BuK was described in Chapter 4.

#### *5.2.9 Phage Expression and Purification*

The phage library containing BuK and thBuK was expressed and purified as described in Chapter 4 with minor modifications. Briefly, library glycerol stocks were used to inoculate 100 mL of 2xYT media containing  $100 \mu\text{g}\cdot\text{mL}^{-1}$  ampicillin,  $34 \mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol, and  $25 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin to an  $\text{OD}_{600} = 0.1$ . Cells were grown at  $37^\circ\text{C}$  to  $\text{OD}_{600} = 0.6$  at which time the temperature was changed to  $30^\circ\text{C}$  and phage expression was induced by the addition of 1 mM IPTG, 0.2% arabinose, 5 mM nicotinamide, and 5 mM of the ncAA. 18-24 h post induction the cells were pelleted (6k rpm, 20 min), the upper 40 mL was removed, and phages were precipitated twice with polyethylene glycol, overnight at  $4^\circ\text{C}$ . Following the second precipitation, phages were resuspended in 1 mL PBS for affinity selection.

#### 5.2.10 Affinity Selection against Sirtuin 2

Rounds 1 and 3: Streptavidin coated magnetic beads (100  $\mu$ L, 50% slurry, Genscript) were transferred to a 1.5 mL tube, washed twice with 1 mL phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5), resuspended in 100  $\mu$ L PBS, and split into two tubes. Biotinylated SIRT2 (10  $\mu$ g) in 400  $\mu$ L PBS was added to one of the tubes and an equal volume of PBS was added to the other tube (hereafter referred to as +SIRT2 and -SIRT2, respectively). The beads/protein mixture was incubated with rocking for 20 min at room temperature. After 20 min the supernate was removed, 1 mL of blocking buffer (0.1 M NaHCO<sub>3</sub> pH 8.6, 5 mg·mL<sup>-1</sup> BSA) was added, and the tubes were incubated for 30 min at room temperature with end-over-end rotation. After 30 min the blocking buffer was removed from the -SIRT2 tube, the resin was washed with PBST (3x 1 mL; PBS containing 0.1% tween-20) followed by PBS (1x 1 mL) and the purified phage library in PBS (Round 1: 6.2 x 10<sup>10</sup> c.f.u., Round 3: 2.9 x 10<sup>10</sup> c.f.u.) was added to the tube for negative selection. The phage bead mixture was incubated for 30 min at room temperature with end-over-end rotation. After 30 min, the +SIRT2 resin was washed (3x 1 mL PBST, 1x 1 mL PBS) and the phage library was transferred to the +SIRT2 tube and incubated for 30 min at room temperature with end-over-end rotation. After 30 min the phage library was removed and the resin was washed (10x 1 mL PBST) to remove nonspecifically bound phages. During each washing step the resin was completely resuspended by pipetting up and down. To remove phages binding to the polypropylene tube, the resin was transferred to a new tube after the 3<sup>rd</sup>, 6<sup>th</sup>, and 9<sup>th</sup> washes. After the last wash, phages were eluted by incubating for 10 min

with 100  $\mu\text{L}$  of a low pH elution buffer (50 mM glycine pH 2.2). After 10 min, the supernate was removed and immediately added to 50  $\mu\text{L}$  of neutralization buffer (1 M Tris pH 8). The neutralized elution was immediately used for amplification.

Round 2: Eight wells of a pre-blocked netutravidin coated polystyrene plate (Pierce) were washed with PBS (3x 250  $\mu\text{L}$ ). Biotinylated SIRT2 (1.25  $\mu\text{g}$  in 100  $\mu\text{L}$  PBS) was added to each of the wells and the plates were incubated in a sealed, humidified box at room temperature for 30 min. After 30 min the SIRT2 solution was removed and the wells were washed (3x 250  $\mu\text{L}$  PBST, 1x 250  $\mu\text{L}$  PBS). The phage library in PBS (800  $\mu\text{L}$ ,  $5.4 \times 10^9$  c.f.u.) was divided evenly into each of the wells and incubated for 30 min at room temperature in a sealed humidified box. After 30 min, the phages were removed and the wells were washed (10x 250  $\mu\text{L}$  PBST). The bound phages were eluted with low pH elution buffer (100  $\mu\text{L}$  per well) for 10 min. The pooled elutions were added to 400  $\mu\text{L}$  neutralization buffer and immediately used for amplification.

Phage Amplification: A small aliquot (10  $\mu\text{L}$ ) of the phage elution was removed for titering. The remaining phages were added to an actively growing culture of *E. coli* Top10 F' (OD<sub>600</sub> = 0.5-0.6) in 20 mL 2xYT containing 10  $\mu\text{g}\cdot\text{mL}^{-1}$  tetracycline for 45 min at 37°C with rotation. After 30 min the cells were pelleted (4k rpm, 15 min), resuspended in 100 mL 2xYT containing 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, and amplified overnight at 37°C. The following day, phagemids were extracted from the amplified culture using a commercial plasmid purification kit (Epoch). The purified phagemids were used to transform electrocompetent *E. coli* Top10 containing M13KO7(pIII) and

pEVOL-CDF-PylT-BuKRS for expressing phages for the next round. For each round the number of transformants was at least 1000-fold greater than the number of eluted phages.

#### *5.2.11 Solid-Phase Peptide Synthesis*

The peptides S2P03, S2P04, S2P04-5, Ac-S2P04-5, and S2P07 were synthesized via manual Fmoc-based solid-phase peptide synthesis on Rink-amide resin using *N,N,N',N'*-tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uranium hexafluorophosphate (HBTU) as a coupling agent. Rink-amide MBHA resin (Novabiochem, 100 mg, substitution: 0.78 mmol g<sup>-1</sup>) was swelled for 1 h in 10 mL of DMF. Deprotection, coupling, and cleavage reactions were carried out at room temperature in a glass reaction vessel agitated with end-over-end rotation under the following conditions:

*Deprotection:* A solution of 20% Piperidine in DMF (v:v, 5 mL) was added to the resin for 5 min. The resin was washed (3x 5 mL DMF) and the deprotection reaction was repeated for 15 min. After the second deprotection reaction the resin was washed again (5x 5 mL DMF, 2x 5 mL DCM).

*Coupling:* Each coupling reaction contained the Fmoc-protected amino acid (312 μmol, 4 eq.), HBTU (304.2 μmol, 3.9 eq.), and *N,N*-diisopropylethylamine (DIPEA, 780 μmol, 10 eq.) dissolved in 5 mL DMF. The Fmoc-protected amino acid was preactivated with



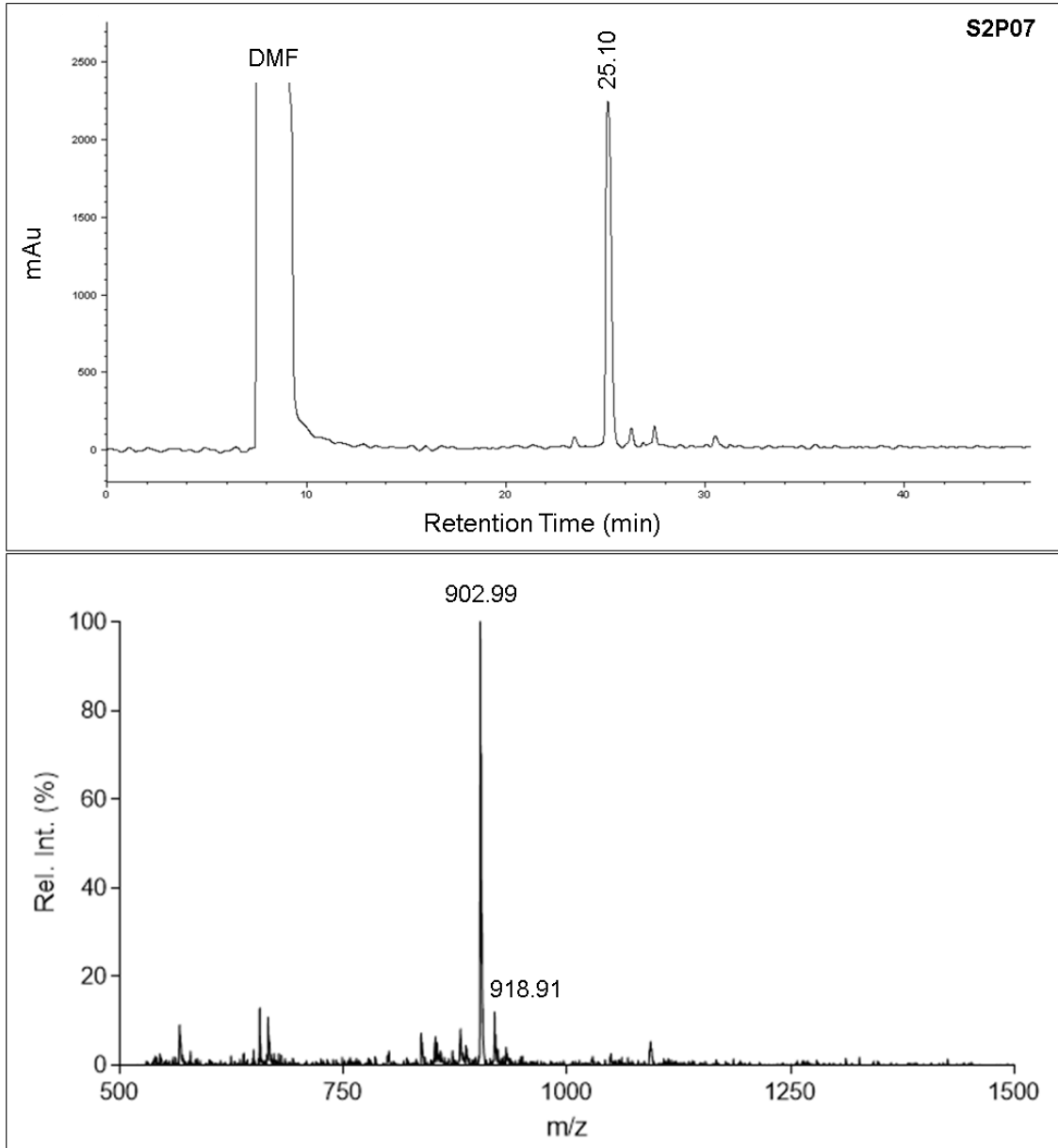
HBTU and DIPEA for 5 min before adding to the resin for 40 min. After 40 min the resin was washed (5x 5 mL DMF, 2x 5 mL DCM).

*N-Terminal Acetylation:* For synthesizing Ac-S2P04-5 with an N-terminal acetylation, the resin was washed (5x 5 mL DMF, 3x 5 mL DCM) following the last Fmoc deprotection and then incubated with a solution of acetic anhydride, pyridine, and DMF (1:2:3, v:v:v) for 1 hour at room temperature with agitation. The peptide was then cleaved and simultaneously sidechain deprotected as described.

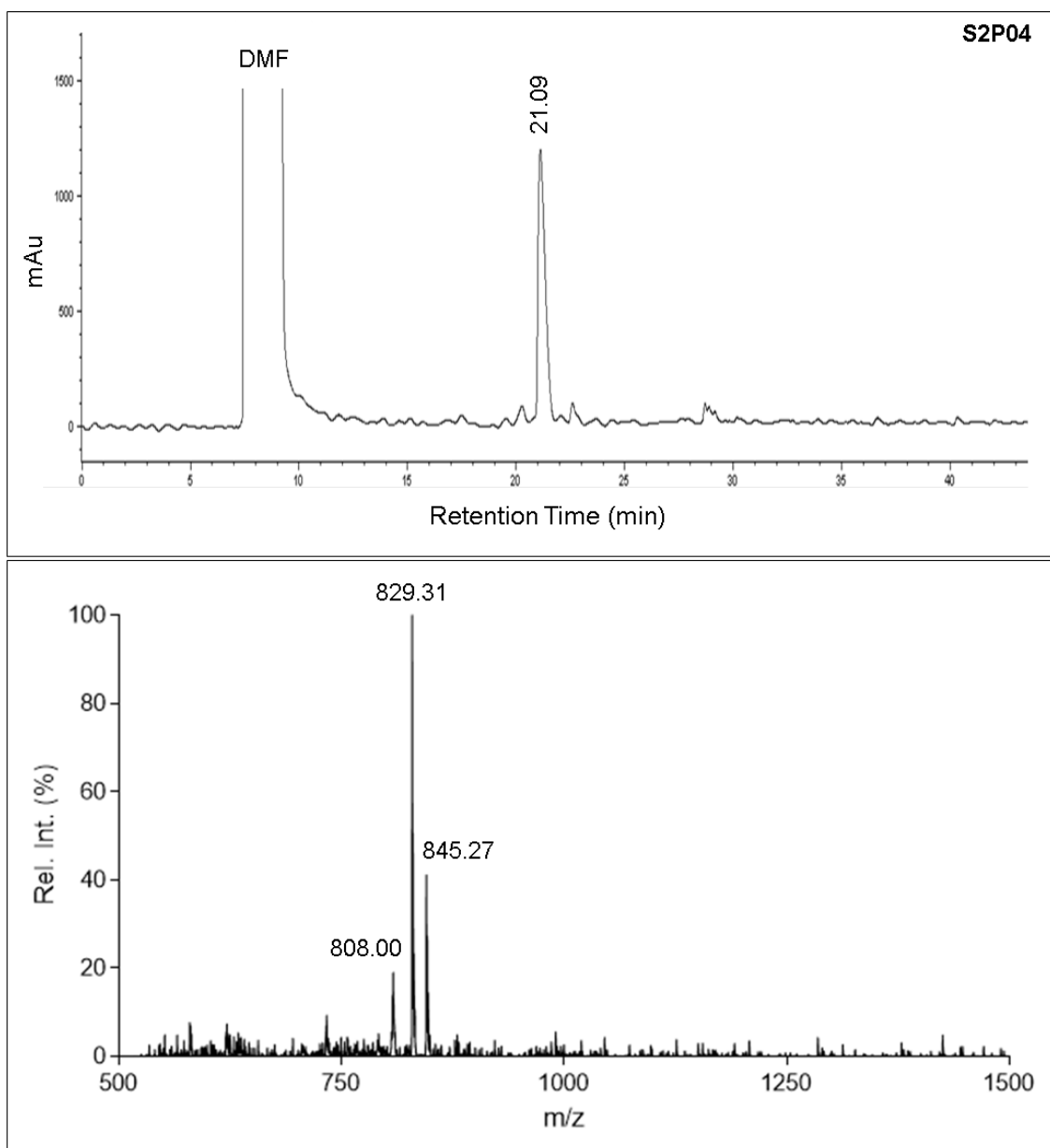
*Cleavage:* After the final Fmoc deprotection, the resin was washed (5x 5 mL DMF, 5x 5 mL DCM) and the peptide was cleaved from the resin and simultaneously sidechain deprotected using 7 mL of a cleavage cocktail containing 88% trifluoroacetic acid (TFA), 5% phenol, 5% H<sub>2</sub>O, and 2% triisopropylsilane (v:w:v:v) for two hours under a blanket of nitrogen. After two hours, the resin was filtered and washed (2x 1 mL cleavage cocktail) and the combined filtrate was reduced to ~4 mL using a stream of nitrogen. The concentrated solution was added dropwise to cold diethyl ether (40 mL) and incubated on ice for 10 min. The precipitated peptide was collected by centrifugation (6k rpm, 10 min), washed twice with cold diethyl ether (20 mL), dried under a stream of nitrogen, and stored at -20°C.

*Purification:* Crude peptides were dissolved in DMF and purified by semi-preparative RP-HPLC on a C18 column (Alltech, 10 x 250 mm, 10 μm, 100 Å pore or Supelco, 10 x

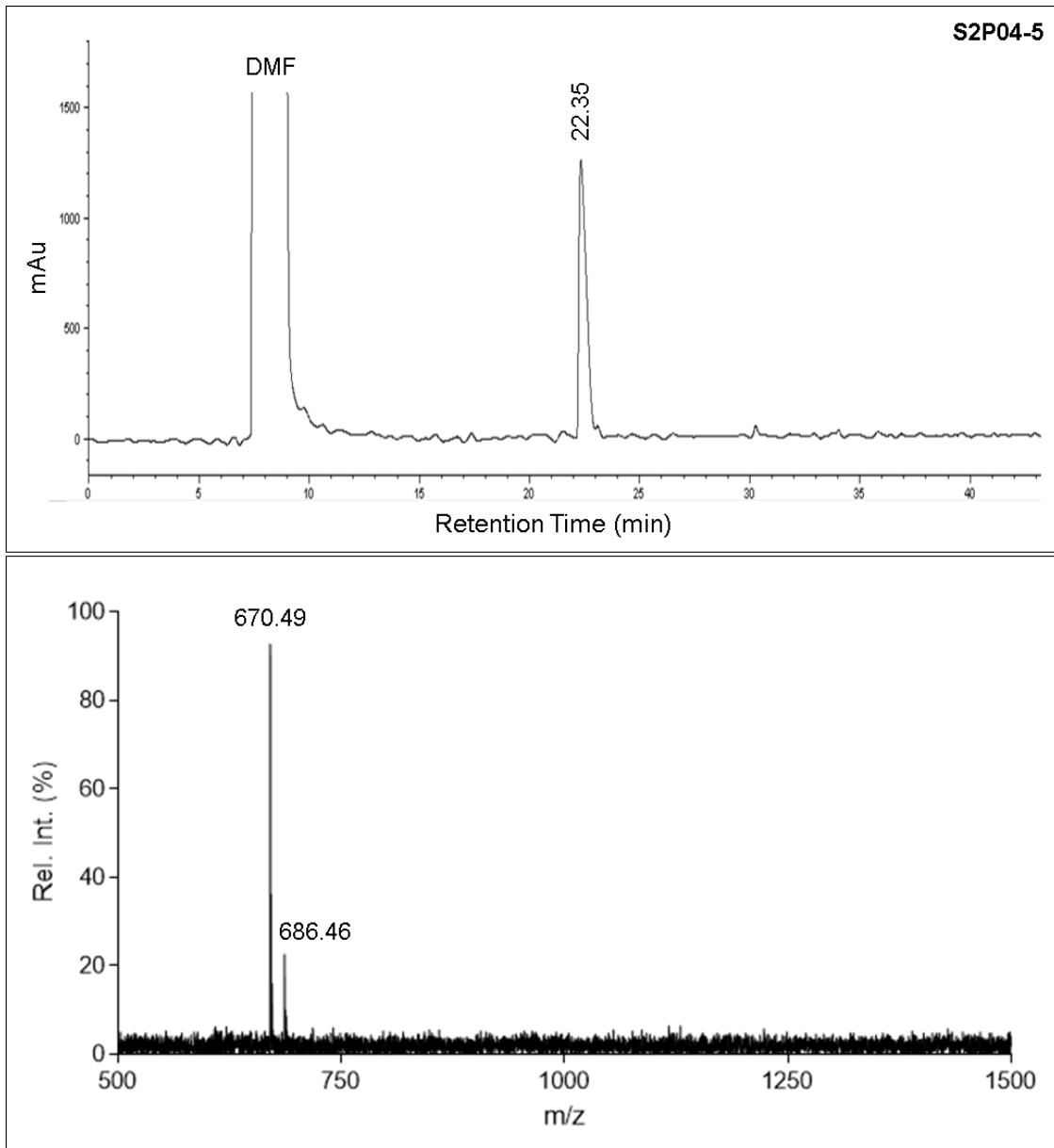
250 mm, 5  $\mu\text{m}$ , 300  $\text{\AA}$  pore). Solvent A: water with 0.1% TFA, Solvent B: acetonitrile with 0.1% TFA. The peptides were eluted with a linear gradient from 20-50% Solvent B over 30 min. Fractions containing the purified peptides were combined and lyophilized to afford a fluffy white powder. Masses of the peptides were confirmed by MALDI-TOF mass spectrometry. Peptide purity was determined according to the HPLC traces at 215 nm and was greater than 90% for all peptides used. MALDI-TOF mass spectra and HPLC traces for the purified peptides are shown in Figures 5.4-5.8.



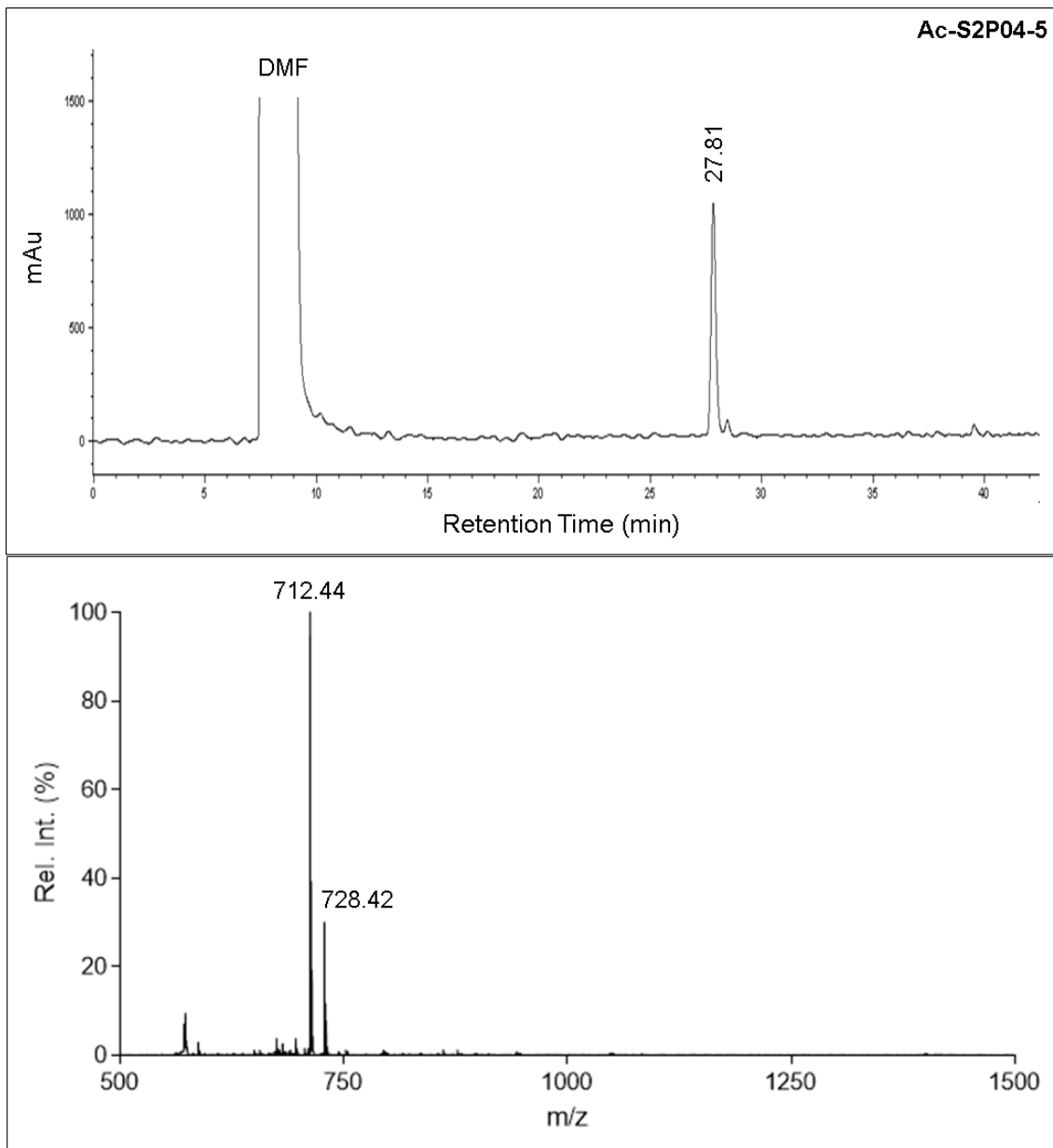
**Figure 5.4. Characterization data for the peptide S2P07.** (Top) HPLC trace at 215 nm. The peptide was eluted with a linear gradient from 20-50% acetonitrile in water (0.1% TFA) from 10-40 min. (Bottom) MALDI-TOF mass spectra. Calculated mass:  $[M+H]^+ = 880.44$  Da,  $[M+Na]^+ = 902.41$  Da,  $[M+K]^+ = 918.52$  Da; Observed Mass: 902.88 Da, 918.91 Da.



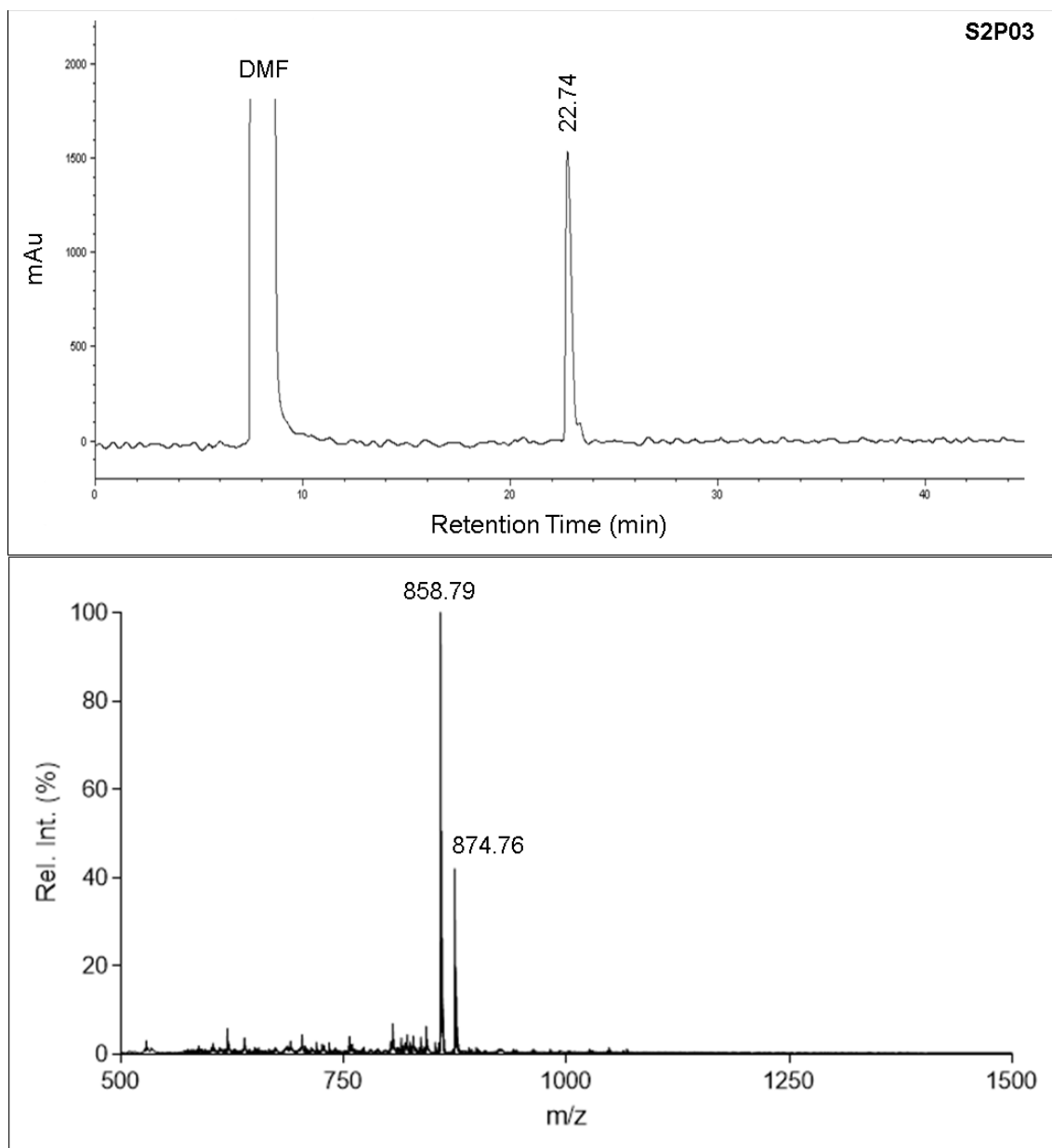
**Figure 5.5. Characterization data for the peptide S2P04.** (Top) HPLC trace at 215 nm. The peptide was eluted with a linear gradient from 20-50% acetonitrile in water (0.1% TFA) from 10-40 min. (Bottom) MALDI-TOF mass spectra. Calculated mass:  $[M+H]^+ = 806.43$  Da,  $[M+Na]^+ = 828.41$  Da,  $[M+K]^+ = 844.52$  Da; Observed Mass: 808.00 Da, 829.31 Da, 845.27 Da.



**Figure 5.6. Characterization data for the peptide S2P04-5.** (Top) HPLC trace at 215 nm. The peptide was eluted with a linear gradient from 20-50% acetonitrile in water (0.1% TFA) from 10-40 min. (Bottom) MALDI-TOF mass spectra. Calculated mass:  $[M+H]^+ = 648.36$  Da,  $[M+Na]^+ = 670.34$  Da,  $[M+K]^+ = 686.45$  Da; Observed Mass: 670.49 Da, 686.46 Da.



**Figure 5.7. Characterization data for the peptide Ac-S2P04-5.** (Top) HPLC trace at 215 nm. The peptide was eluted with a linear gradient from 20-50% acetonitrile in water (0.1% TFA) from 10-40 min. (Bottom) MALDI-TOF mass spectra. Calculated mass:  $[M+H]^+ = 690.37$  Da,  $[M+Na]^+ = 712.35$  Da,  $[M+K]^+ = 728.46$  Da; Observed Mass: 712.44 Da, 728.42 Da.



**Figure 5.8. Characterization data for the peptide S2P03.** (Top) HPLC trace at 215 nm. The peptide was eluted with a linear gradient from 20-50% acetonitrile in water (0.1% TFA) from 10-40 min. (Bottom) MALDI-TOF mass spectra. Calculated mass:  $[M+H]^+ = 836.44$  Da,  $[M+Na]^+ = 858.41$  Da,  $[M+K]^+ = 874.52$  Da; Observed Mass: 858.79 Da, 874.76.

### 5.2.12 Sirtuin 2 Discontinuous Inhibition Assay

The Sirtuin 2 discontinuous inhibition assay was modified from the procedure described by Jing et al.<sup>214</sup> The inhibition reactions were performed in 100  $\mu\text{L}$  volumes at 37°C in black 96-well plates (Nunc). Different concentrations of the thBuK-peptide inhibitors or TB (0.0064, 0.032, 0.16, 0.4, 0.8, 2, 4, and 100  $\mu\text{M}$ ) were incubated with SIRT2 (1  $\mu\text{M}$ ) and  $\text{NAD}^+$  (2 mM) in assay buffer (50 mM Tris pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM  $\text{MgCl}_2$ , 1 mM DTT, 5% DMF) for 15 min. After incubation with the inhibitor, deacetylase reactions were initiated by the addition of 50  $\mu\text{M}$  Boc-Lys(Ac)-AMC substrate. The reaction was allowed to proceed for 20 min before being stopped by the addition of 100  $\mu\text{L}$  of a 2x stop solution (1  $\text{mg}\cdot\text{mL}^{-1}$  trypsin, 50 mM nicotinamide in assay buffer). After addition of the stop solution reactions were incubated for an additional 30 min and then the AMC fluorescence was measured in a BioTek Synergy H1 microplate reader at  $\lambda_{\text{Ex}} = 390 \text{ nm}$ ,  $\lambda_{\text{Em}} = 460 \text{ nm}$  (Gain = 75, Read Height = 7 mm). Fluorescence data were normalized relative to 100% activity (no inhibitor) and 0% activity (no enzyme) controls.  $\text{IC}_{50}$  values were determined by nonlinear regression of the plot of normalized fluorescence vs. inhibitor concentration using GraphPad Prism.

### 5.2.13 Sirtuin 1, 2, and 3 Inhibition Assay

The continuous assay for comparing isoform specificity of S2P04 was performed using the universal fluorogenic sirtuin substrate described by Schuster et al.<sup>226</sup> The inhibition reactions were performed in 50  $\mu\text{L}$  volumes at 37°C in black, half-area, 96-well plates. The sirtuin enzyme (0.25  $\mu\text{M}$ ) was pre-incubated with different concentrations of the



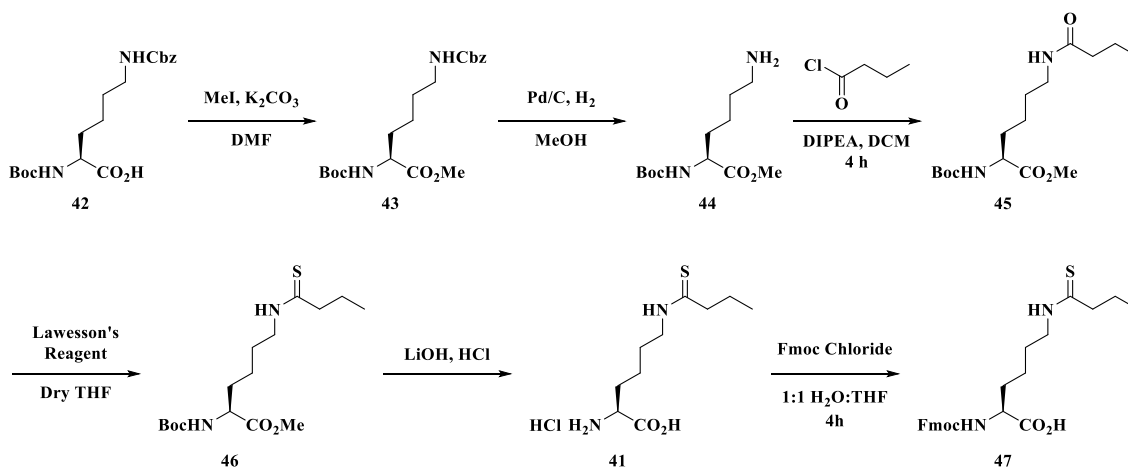
peptide S2P04 (0.0064 – 50  $\mu\text{M}$  for SIRT1 and SIRT2, 0.5 – 100  $\mu\text{M}$  for SIRT3) or TB (0.16 – 150  $\mu\text{M}$  for SIRT1, SIRT2, and SIRT3 ) and 1 mM  $\text{NAD}^+$  in assay buffer (20 mM Tris pH 7.8, 150 mM NaCl, 5 mM  $\text{MgCl}_2$ , 1 mM DTT) for 15 min. Following incubation, the reaction was started with the addition of the fluorogenic substrate (5  $\mu\text{M}$ ) and product formation was monitored by measuring the fluorescence intensity every 20 s for 5 min (SIRT2) or 30 min (SIRT1 and SIRT3) in a BioTek Synergy H1 microplate reader at  $\lambda_{\text{Ex}} = 320$  nm,  $\lambda_{\text{Em}} = 408$  nm (Gain = 130, Read Height = 7 mm). Initial rates of product formation were determined by a linear regression of the plot of fluorescence intensity vs. time and normalized relative to 0% (max inhibitor) and 100% (no inhibitor) controls.  $\text{IC}_{50}$  values were determined by nonlinear regression of the plot of normalized initial rate vs. inhibitor concentration using GraphPad Prism.

#### 5.2.14 Chemical Synthesis

##### Synthesis of $N_{\epsilon}$ -thiobutyryl-lysine (**41**) and $N_{\alpha}$ -Fmoc- $N_{\epsilon}$ -thiobutyryl-lysine (**47**)

The synthetic scheme shown in Figure 5.9 was followed for the synthesis of  $N_{\epsilon}$ -thiobutyryl-lysine (**41**) and  $N_{\alpha}$ -Fmoc- $N_{\epsilon}$ -thiobutyryl-lysine (**47**).

To a suspension of Boc-Lys(Z)-OH (**42**, 40.3 g, 0.106 mol) and potassium carbonate (27.6 g, 0.200 mol) in DMF (200 mL) was added iodomethane (9.90 mL, 0.159 mol), and the mixture was stirred at room temperature for 30 h. The mixture was filtered, and the filter cake was washed with ethyl acetate (50 mL), dissolved in water (100 mL), and extracted with ethyl acetate (100 mL 2x). All the ethyl acetate solutions



**Figure 5.9.** The synthesis of  $N_{\epsilon}$ -thiobutyryl-lysine and  $N_{\alpha}$ -Fmoc- $N_{\epsilon}$ -thiobutyryl-lysine.

were combined with the filtrate, and the solution was evaporated under vacuum until most of the DMF had been removed. The residue was dissolved in ether (250 mL), washed with water (100 mL) and brine (50 mL), dried ( $\text{Na}_2\text{SO}_4$ ), and evaporated to afford **43** (41.8 g, quant.) as a yellow oil. The material was pure enough for the next reaction without further treatment.

A solution of **43** (4.20 g, 10.6 mmol) in methanol (100 mL) was hydrogenated under an  $\text{H}_2$  balloon in the presence of palladium on carbon (10 wt.% Pd, 0.71 g, 0.67 mmol) at room temperature for 3 h, and TLC analysis showed complete conversion. The mixture was then filtered over a pad of Celite and the solution was directly used for the next reaction. The material should be immediately used without purification since prolonged storage at room temperature or flash chromatography would contribute to lactam formation.

To a solution of the above amine (**44**, ~9.15 mmol) in anhydrous dichloromethane (90 mL), cooled in an ice bath, was added *N,N*-diisopropylethylamine (2.80 mL, 16.07 mmol) dropwise, followed by a solution of *n*-butyryl chloride (1.10 mL, 11.07 mmol) in dichloromethane (10 mL) dropwise over 20 min. The mixture was then stirred at room temperature for 12 h, and it was washed with sodium hydroxide solution (0.5 M, 20 mL) and brine (20 mL 2x), dried ( $\text{Na}_2\text{SO}_4$ ), evaporated, and flash chromatographed (EtOAc/hexanes, 1:3) to give **45** (2.77 g, 79% for two steps) as a colorless oil.  $R_f = 0.28$  (EtOAc/hexanes, 1:2).

To a solution of compound **45** (2.5 g, 7.5 mmol) in THF (20 mL) was added Lawesson's reagent (3.0 g, 7.5 mmol) at room temperature. The reaction mixture was stirred overnight under nitrogen (monitored by TLC). After removal of THF using a rotary evaporator, the residue was purified by silica gel column (EtOAc/hexanes = 3:7) to give the product **46** as light yellow oil (1.96 g, 75% yield).

To a solution of **46** (1.90 g, 5.49 mmol) in THF (30 mL) was added lithiumhydroxide solution (1.0 M, 5.5 mL, 5.49 mmol), and the mixture was stirred at room temperature for 2 h. The mixture was diluted in water (20 mL) and extracted with ether (30 mL, 2x). The ether extracts were discarded, and the remaining aqueous solution was adjusted to pH 3 with hydrochloric acid (3 M), with the concomitant formation of white precipitate. The suspension was extracted with ethyl acetate (30 mL, 2x), and the combined organic phases were washed once with brine (20 mL), dried ( $\text{Na}_2\text{SO}_4$ ), and evaporated to give the crude carboxylic acid as a colorless oil, which was used without further purification.

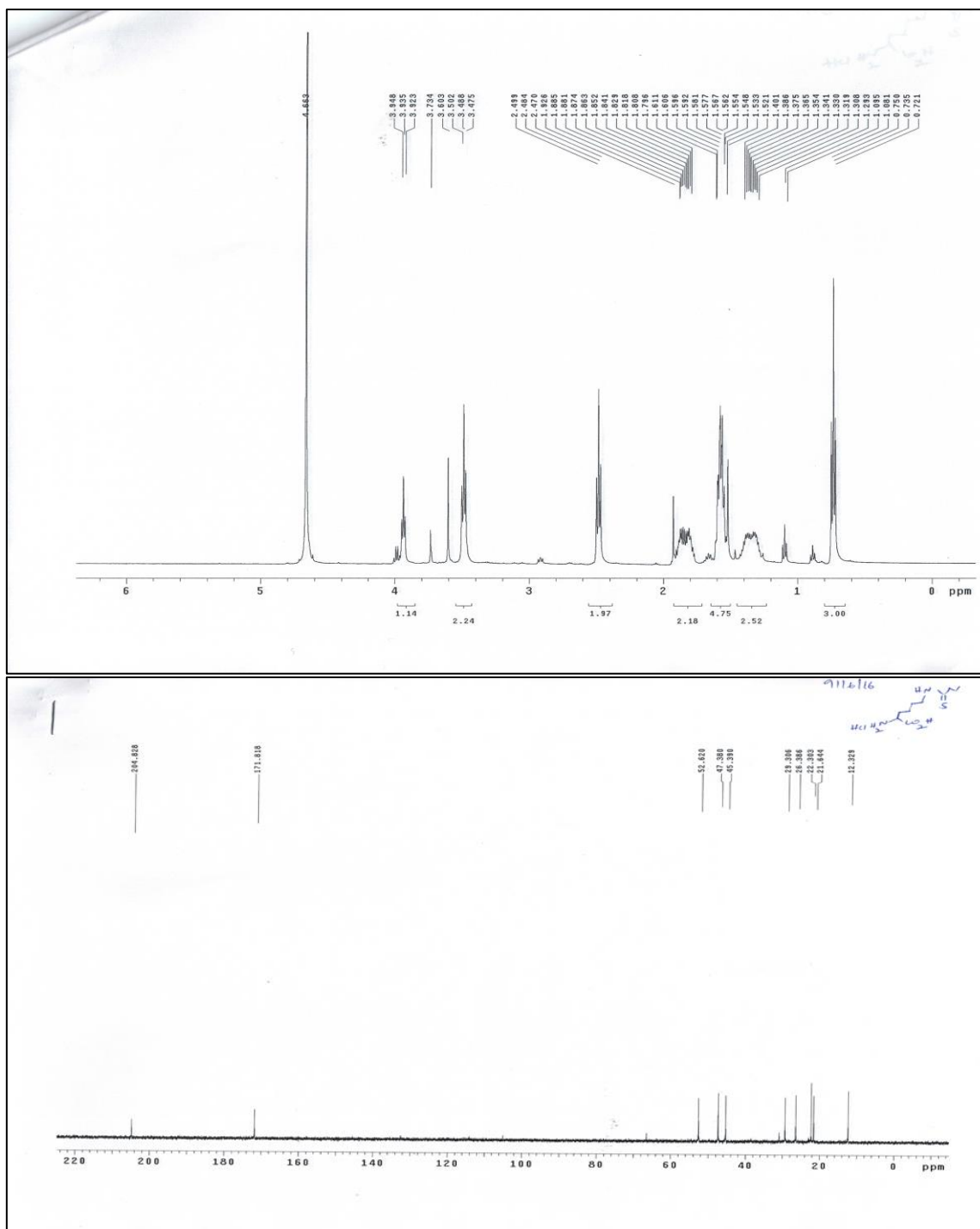
The above crude acid (~4.8 mmol) was dissolved in 1,4-dioxane (10 mL), and hydrogen chloride in 1,4-dioxane (4.0 M, 2.4 mL, 9.63 mmol) was added. The resulting white suspension was stirred at room temperature for 20 h, filtered, washed with dichloromethane, and dried to give **41** (1.03 g, 70% for two steps) as a white solid.

$^1\text{H}$  NMR ( $\text{D}_2\text{O}$ , 500 MHz)  $\delta$  3.93 (t, 1H,  $J = 3.6$  Hz), 3.48 (t, 2H,  $J = 3.6$  Hz), 2.49 (t, 2H,  $J = 4.2$  Hz), 1.92-1.79 (m, 2 H), 1.60-1.52 (m, 4 H), 1.38-1.30 (m, 2H), 0.73 (t, 3H,  $J = 4.2$  Hz).  $^{13}\text{C}$  NMR ( $\text{D}_2\text{O}$ , 125 MHz)  $\delta$  204.8, 171.8, 52.6, 47.3, 45.3, 29.3, 26.3, 22.3, 21.6, 12.3.

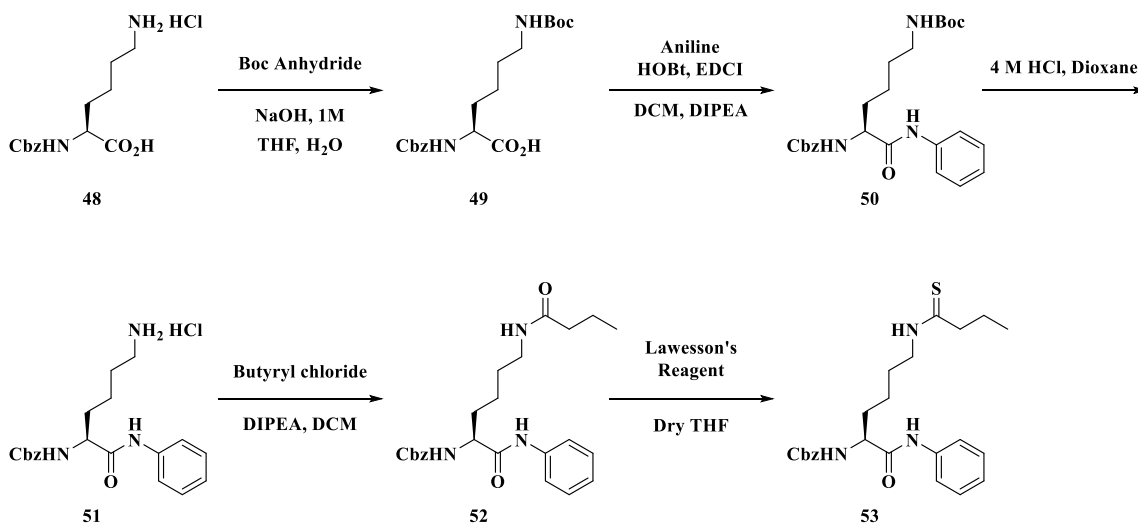
Compound **41** (1.00 g, 3.73 mmol) and  $\text{NaHCO}_3$  (1.17 g, 11.1 mmol) were dissolved in a mixture of 20 mL  $\text{H}_2\text{O}$  and 20 mL THF. The solution was cooled to 0 °C in an ice bath, and 9-fluorenylmethyl chloroformate (1.17 g, 4.47 mmol) dissolved in dry THF was added dropwise over 10 min. The reaction mixture was stirred for 4 h. THF was evaporated and the aqueous phase was acidified with 3 M HCl (aq.) to pH 2-3. The product was extracted 3 times with EtOAc and the organic phase was dried and evaporated. The product was purified by column chromatography (MeOH gradient 0-10% in DCM) to give product **47** (1.00 g, 60%) as a solid. Proton and carbon NMR for **47** are shown in Figure 5.10.

$^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ , 300 MHz):  $\delta$  7.80 (d, 2H,  $J = 4.5$  Hz), 7.69 (t, 2H, 4.8 Hz), 7.40 (t, 2H,  $J = 4.5$  Hz), 7.32 (t, 2H,  $J = 4.8$  Hz), 4.36 (d, 2H,  $J = 3.9\text{Hz}$ ), 4.23 (t, 1H,  $J = 4.2$  Hz), 4.16-4.13 (m, 1H), 3.60 (t, 2H,  $J = 3.9$  Hz), 2.57 (t, 2H,  $J = 4.5$  Hz), 1.90-1.85 (m, 1H), 1.78-1.65 (m, 5H), 1.47-1.44 (m, 2H), 0.93 (t, 3H,  $J = 4.5$  Hz).  $^{13}\text{C}$  NMR ( $\text{CD}_3\text{OD}$ ,

125 MHz);  $\delta$  204.7, 175.1, 157.2, 143.9, 143.7, 141.9, 127.4, 126.8, 124.8, 119.5, 66.8, 54.1, 45.2, 31.1, 26.9, 23.0, 22.6, 12.2.



Synthesis of TB (53): The synthetic scheme shown in Figure 5.11 was followed for the synthesis of TB (53).



**Figure 5.11. The synthesis of TB.**

To a 500 mL of round-bottom-flask was added Z-Lys-OH (**48**, 5.0 g, 15.82 mmol), followed by THF (100 mL) and aqueous NaOH (1 M, 17.40 mL, 17.4 mmol). The reaction mixture was cooled to 0°C. Di-*tert*-butyl dicarbonate (3.75g, 17.4 mmol) was dissolved in THF (100 mL) and added to the above mixture dropwise over 40 min. Then the resulting mixture was stirred at room temperature for 20 h. THF was removed under reduced pressure. The residue was diluted with water (100 mL), and extracted with EtOAc (100 mL). The aqueous layer was collected and acidified to pH = 3 with HCl (1 M in water), then extracted with EtOAc (100 mL × 3). The combined organic layers were dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to

afford compound **49** (4.81 g, 80% yield). Compound **49** was pure enough for the next step without further purification.

A mixture of **49** (2.20 g, 5.5 mmol), aniline (1.15 g, 11.0 mmol), EDCI (1.63 g, 8.5 mmol) and HOBt·H<sub>2</sub>O (1.34 g, 8.5 mmol) in DCM (100 mL) was stirred for 24 h at room temperature. The reaction mixture was poured into water and the whole was extracted with AcOEt. The AcOEt layer was separated, washed with saturated aqueous NaHCO<sub>3</sub>, brine, and dried over Na<sub>2</sub>SO<sub>4</sub>. Filtration and concentration *in vacuo* gave a colorless solid. The solid was suspended in n-hexane and collected by filtration to give 2.53 g (90%) of **50** as a colorless solid.

To a solution of the fully protected compound **50** (2.52 g, 5.53 mmol) dissolved in 15 mL dioxane was added HCl (2.76 mL, 11.07 mmol, 4 M in dioxane) and the mixture was stirred and monitored by TLC for 8 h. After complete removal of the solvent, compound **51** was obtained as a white powder (2 g, 92%). **51** was pure enough to use for the next step without further purification.

To a solution of the above amine (**51**, 2 g, 5.10 mmol), in anhydrous dichloromethane (30 mL), cooled in an ice bath, was added *N, N*-diisopropylethylamine (1.10 mL, 7.92 mmol) dropwise, followed by a solution of n-butyryl chloride (1.10 mL, 7.91 mmol) in dichloromethane (10 mL) dropwise over 5 min. The mixture was then stirred at room temperature for 12 h, and it was washed with sodium hydroxide solution (0.5 N, 20 mL) and brine (20 mL x 2), dried (Na<sub>2</sub>SO<sub>4</sub>), evaporated, and flash chromatographed (EtOAc/hexanes, 1:3) to give **52** (1.85 g, 85%) as a colorless oil. R<sub>f</sub> = 0.28 (EtOAc/hexanes, 1:2).



To a solution of compound **52** (1.5 g, 3.52 mmol) in THF (20 mL) was added Lawesson's reagent (1.42 g, 3.52 mmol) at room temperature. The reaction mixture was stirred overnight under nitrogen (monitored by TLC). After removal of THF using a rotary evaporator, the residue was purified by silica gel column (EA/Hexane = 3:7) to give the product **53** as light yellow oil (1.1 g, 75% yield). Proton and carbon NMR of **53** are shown in Figure 5.12.

$^1\text{H}$  NMR ( $\text{CDCl}_3$ , 300 MHz):  $\delta$  8.43 (br, 1H), 7.70 (br, 1H), 7.51(d, 2H,  $J = 4.5$  Hz), 7.34-7.28 (m, 7H), 7.12-7.11(m, 1H), 5.72 (d, 2H,  $J = 4.8$  Hz), 5.15-5.09 (m, 2H), 4.38 (br, s, 1H), 3.70-3.60 (m, 2H), 2.59 (t, 2H,  $J = 4.8$  Hz), 1.97-1.90 (m, 1H), 1.83-1.72 (m, 6H), 1.49-1.48 (m, 2H), 0.93 (t, 3H, 4.5 Hz).  $^{13}\text{C}$ NMR ( $\text{CDCl}_3$ , 125MHz);  $\delta$  205.4, 170.7, 156.8, 137.4, 135.9, 128.9, 128.5, 128.2, 127.8, 124.7, 120.3, 6.2, 60.5, 55.3, 48.6, 45.4, 32.2, 27.1, 22.8, 22.7, 21.1, 14.2, 13.9.

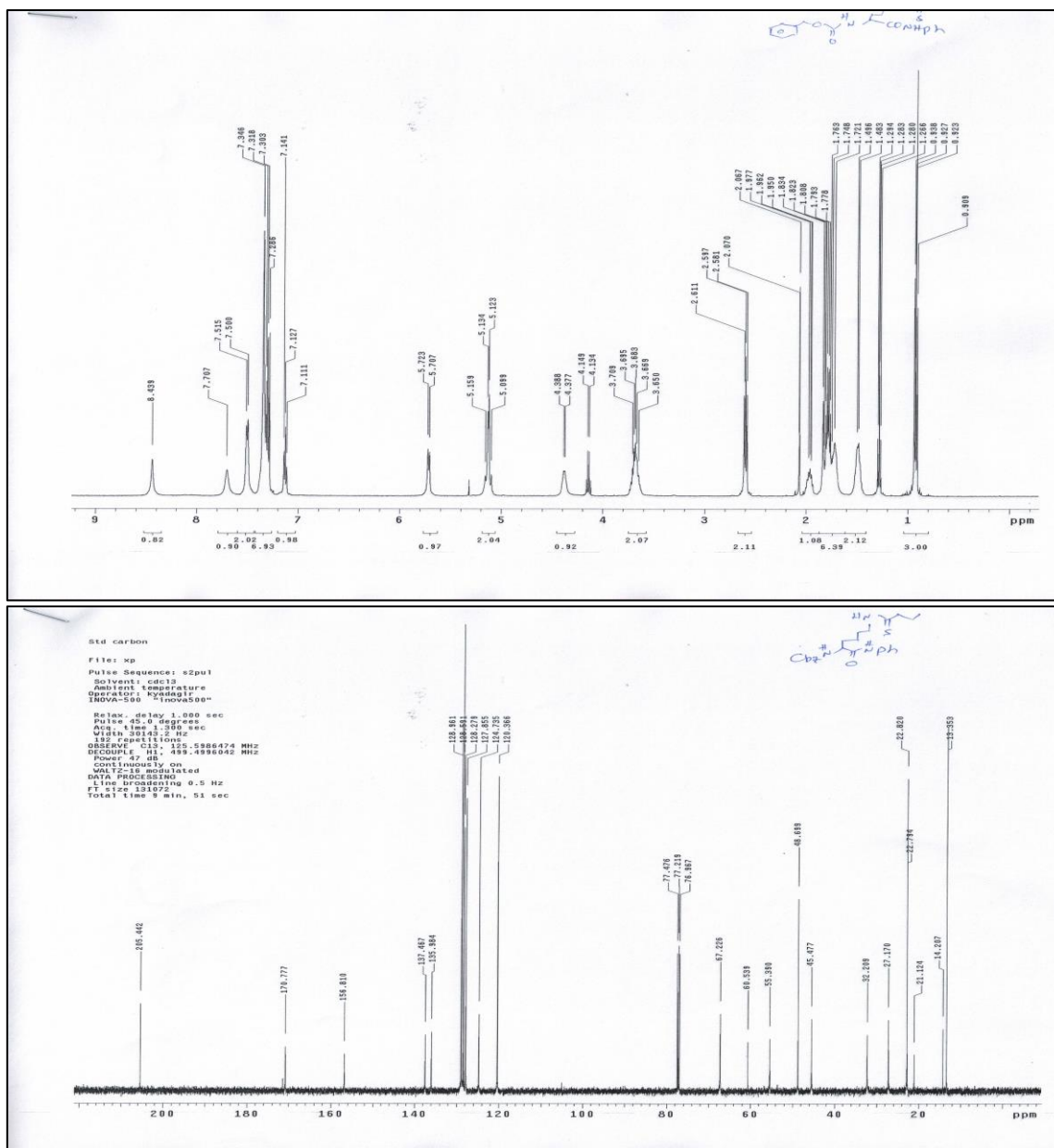


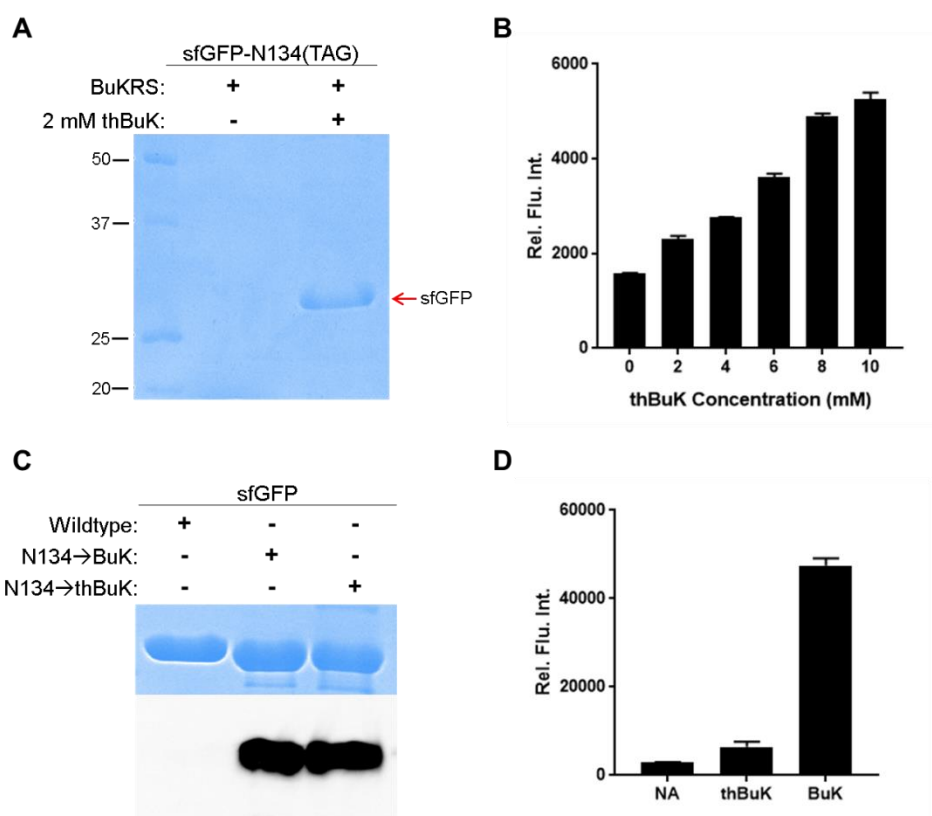
Figure 5.12. Characterization data for TB. (Top) <sup>1</sup>H-NMR spectrum for TB (53). (Bottom) <sup>13</sup>C-NMR spectrum for compound 53.

## 5.3 Results and Discussion

### 5.3.1 Genetic Incorporation of thBuK

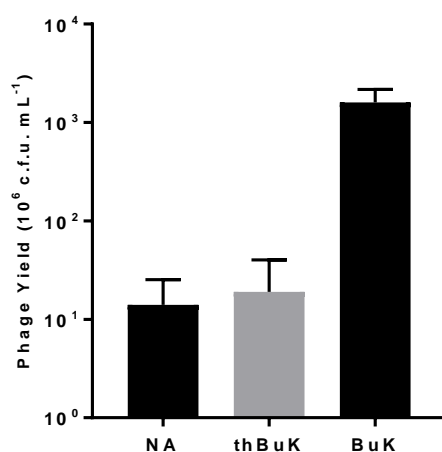
In Chapter 4 we described the construction and characterization of a phagemid system capable of displaying peptide libraries containing various ncAAs. In this system, ncAAs were incorporated in the peptides in response to TAG codons by the pyrrolysyl-tRNA synthetase (PylRS) in conjunction with its cognate tRNA (tRNA<sub>CUA</sub><sup>Pyl</sup>). In our efforts to identify peptide inhibitors of SIRT2, we initially aimed to directly screen this library containing thBuK. To express the peptide library with thBuK, we first investigated the ability of an available PylRS mutant to facilitate the genetic incorporation of thBuK. We have previously reported a PylRS mutant containing a Y384W mutation (BuKRS) that can efficiently incorporate *N*<sub>ε</sub>-butyryl-lysine (BuK) in response to amber codons in *E. coli*.<sup>177</sup> Given the structural similarity of BuK and thBuK, we tested whether BuKRS could also recognize thBuK as a substrate. To do this, *E. coli* Top10 were co-transformed with two plasmids: pEVOL-PylT-BuKRS, which contains genes encoding BuKRS and tRNA<sub>CUA</sub><sup>Pyl</sup>, and pBAD-sfGFP-N134(TAG), which contains the gene encoding a sfGFP reporter protein with a TAG mutation at position 134. Expression of full length sfGFP from these cells was dependent on addition of thBuK to the growth media, consistent with suppression of the TAG codon and incorporation of thBuK at position 134 (Figure 5.13A). In-cell fluorescence measurements showed that the intensity of sfGFP fluorescence increased with increasing concentrations of thBuK (Figure 5.13B). To further investigate the incorporation of thBuK into sfGFP, we performed Western blot with an antibody specific for butyryl-lysine. Previous studies

have shown that, due to the structural similarity of AcK and thAcK, both can be detected by antibodies generated to detect AcK.<sup>227, 228</sup> We utilized wildtype sfGFP and sfGFP-N134(TAG) that had been expressed with BuK as negative and positive controls, respectively. As shown in Figure 5.13C, the pan anti-butyryl-lysine antibody detected sfGFP when expressed with BuK and thBuK but not the wildtype protein, further supporting the genetic incorporation of thBuK.



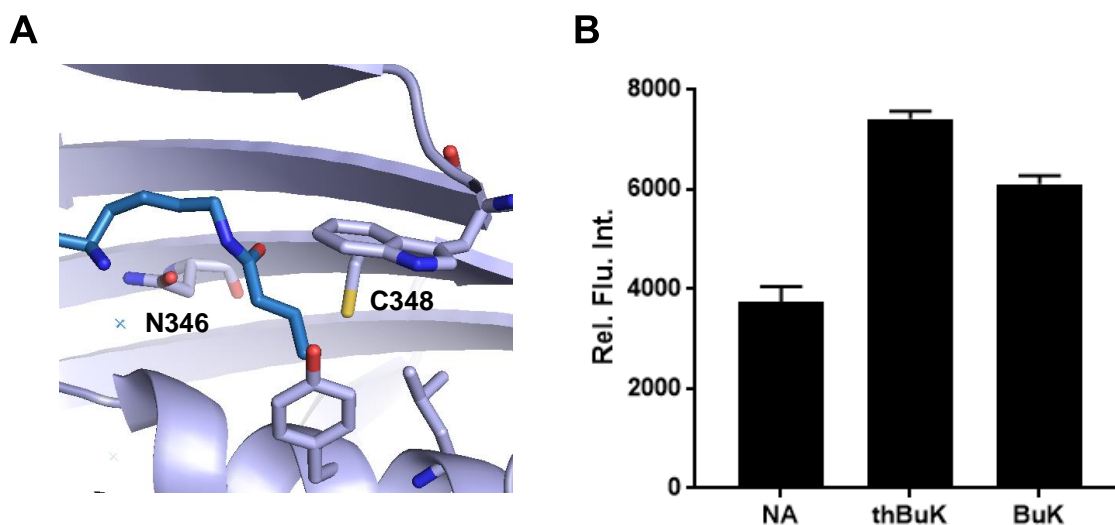
**Figure 5.13. The genetic incorporation of  $N_{\epsilon}$ -thiobutyryl-lysine.** (A) The genetic incorporation of  $N_{\epsilon}$ -thiobutyryl-lysine (thBuK) at the N134 position in sfGFP. (B) In-cell fluorescence measurements of sfGFP. sfGFP expression level is dependent on the concentration of thBuK provided in the growth media. Error bars represent one standard deviation of the mean of two measurements. (C) SDS-PAGE and Western blot of sfGFP-wt, sfGFP-N134→BuK, and sfGFP-N134→thBuK. Proteins were detected with a pan anti-butyryl-lysine antibody. (D) In-cell fluorescence of cells expressing sfGFP-N134(TAG) in the presence of BuK, thBuK, or no ncAA (NA). Error bars represent one standard deviation of the mean of two measurements.

Although these data demonstrate that BuKRS can facilitate the incorporation of thBuK, the efficiency with which the ncAA is incorporated is very low. This is illustrated by the fluorescence intensity signal of cells expressing sfGFP-N134(TAG) in the presence of thBuK which is 7.4-fold lower than when the same protein is expressed in the presence of BuK (Figure 5.13D). We anticipated that this low incorporation efficiency would make detection of the incorporation of thBuK into the phage library challenging. To test the incorporation of thBuK in the phage library, *E. coli* Top10 were transformed with a phagemid encoding the heptapeptide TAG-obligate library, as well as pEVOL-CDF-PylT-BuKRS, and M13KO7(pIII<sup>+</sup>). Phages were expressed from these cells in 2xYT media supplemented with BuK, thBuK, or no ncAA. Adding BuK to the growth media resulted in nearly a 100-fold increase in the number of infective phages that were produced compared to the no ncAA control. Whereas adding 5 or 10 mM of thBuK led to only insignificant increase in the number of phages produced (Figure 5.14).



**Figure 5.14. Phage expression with thBuK and BuK.** Yield is displayed in millions of colony forming units per milliliter of culture. Error bars represent one standard deviation of the mean of two independent expressions.

Next, we made several attempts to improve the incorporation efficiency of thBuK by modifying the BuKRS enzyme. From the crystal structure of PylRS in complex with adenylated-BuK, we identified two residues in the enzyme active site that were in close proximity to the side chain amide oxygen of BuK, namely, N346 and C348 (Figure 5.15A). Assuming that thBuK binds in a similar manner as BuK, we reasoned that mutating the closest residue (N346) to a smaller, more hydrophobic side chain may help to accommodate the sulfur of thBuK which is larger and more hydrophobic than oxygen. Mutating N346 to A has been used previously to expand the binding pocket of PylRS.<sup>146</sup> Therefore, we introduced an N346A mutation into BuKRS via site-directed mutagenesis. We again utilized the expression of sfGFP-N134(TAG) to gauge the incorporation efficiency of thBuK by BuKRS(N346A). Contrary to our prediction, fluorescence measurements of cells expression sfGFP-N134(TAG) indicated that there was no significant increase in the incorporation efficiency of thBuK with BuKRS(N346A) (Figure 5.15B). Interestingly, this mutation significantly reduced the ability of the enzyme to recognize BuK as a substrate. This likely arises from the fact that BuKRS(N346A) is missing both N346 and Y384. In wildtype PylRS these two residues are instrumental in substrate recognition. N346 makes two interactions with the Pyl substrate including a water-mediated hydrogen bond with the alpha amine and a direct hydrogen bond with the sidechain amide oxygen. Whereas, Y384 makes direct hydrogen bonding interactions with the alpha amine and the nitrogen of the pyrrole



**Figure 5.15. The crystal structure of PylRS in complex with adenylated-BuK and expression of sfGFP with BuKRS(N346A).** (A) The crystal structure of PylRS in complex with adenylated-BuK (PDB: 4CH3). The PylRS enzyme is shown as a purple ribbon. Residues lining the amino acid binding pocket are shown as sticks. The adenylated-BuK is shown in blue. Residues N346 and C348 which are in proximity to the sidechain amide oxygen of BuK are labelled. (B) In-cell fluorescence measurements of cells expressing sfGFP-N134(TAG) and BuKRS(N346A) in the presence of BuK, thBuK, or no ncAA (NA). Error bars represent one standard deviation of the mean of two measurements.

ring.<sup>229</sup> Since these two residues are mutated in BuKRS(N346A), three out of five potential hydrogen bond interactions between the enzyme and substrate are removed explaining the decreased incorporation of BuK.

In a further attempt to identify a PylRS mutant that could more efficiently incorporate thBuK we screened a PylRS mutant library in which N346 and C348 were randomized to any of the 20 canonical amino acids. A plasmid encoding the enzyme library was used to transform *E. coli* Top10 along with a second plasmid encoding sfGFP-N134(TAG). Transformed cells were grown on agar plates supplemented with thBuK and fluorescent colonies were screened for thBuK incorporation. Unfortunately, none of the mutants in the library were able to more efficiently incorporate thBuK while

excluding canonical amino acids. These results indicate that efficient thBuK incorporation will require more comprehensive restructuring of the PylRS substrate binding pocket to accommodate the thioamide sidechain.

### 5.3.2 Affinity Selection against Sirtuin 2

Since our attempts to improve the incorporation efficiency of thBuK were unsuccessful, we could not directly use the peptide library containing thBuK to identify inhibitors of SIRT2. Therefore, we devised an indirect selection strategy dubbed “bait-and-switch” selection. Under this strategy, we planned to perform selection against SIRT2 using the peptide library that had been expressed with BuK, instead of thBuK. This selection allows us to identify peptide sequences containing BuK that bind to SIRT2. Following selection, these peptides can be synthesized with thBuK and used for SIRT2 inhibition. Since BuK is a substrate for SIRT2, this strategy requires that selection be performed against the apoenzyme, lest the butyryl group be removed and the deacylated peptide released from the enzyme. Therefore, we performed selection against SIRT2 in the absence of the NAD<sup>+</sup> cofactor. We reasoned that withholding NAD<sup>+</sup> will not interfere with selection as it has been well documented that SIRT2 follows an ordered sequential mechanism in which the first substrate to bind is the acylated peptide, followed by NAD<sup>+</sup>.<sup>194</sup> Further, binding of the peptide to SIRT2 is not affected by the presence or absence of NAD<sup>+</sup>.<sup>230</sup>

We performed selection against human SIRT2 that had been recombinantly expressed in *E. coli* and subsequently biotinylated with biotin *N*-hydroxysuccinimide



ester. Briefly, biotinylated SIRT2 was immobilized by capture on streptavidin coated magnetic beads (rounds 1 and 3) or neutravidin coated polystyrene plates (round 2). The immobilized enzyme was then incubated with the phage library ( $5 \times 10^9 - 6 \times 10^{10}$  cfu) for 30 minutes followed by washing away of nonspecifically bound phages and elution of specific binders with a low pH buffer. Eluted phages were isolated after the third round of selection and their DNA sequenced ( $n = 20$ ). The peptide sequences of isolated clones are shown in Figure 5.16.

Clone	Amino Acid Sequence												
	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	
S2P01						L	X	M	T	S	I	L	
S2P02					H	L	X	T	F	F	Y		
S2P03				C	T	V	X	T	S	L			
S2P04			T	C	T	V	X	I	G				
S2P05			T	C	T	V	X	I	G				
S2P06			T	C	T	V	X	I	G				
S2P07				C	T	F	X	V	P	T			
S2P08			W	S	G	F	X	A	P				
S2P09			W	S	G	F	X	A	P				
S2P10						F	X	L	E	S	F	L	
S2P11			S	N	V	F	X	V	I				
S2P12					A	F	X	H	M	T	V		
S2P13				Q	M	R	F	X	P				
S2P14				Q	M	R	F	X	P				
S2P15							X	V	C	S	C	Y	A
S2P16							X	V	C	S	C	Y	A
S2P17			C	W	W	C	X	V	S				
S2P18		T	E	S	N	H	X	G					
S2P19	L	F	L	W	M	P	X						
S2P20				S	P	M	X	N	K	V			

**Figure 5.16. The amino acid sequence of peptides containing BuK identified from affinity selection against sirtuin 2.** The amino acid sequences of peptides isolated from the third round of selection against SIRT2 is shown. The position of BuK is indicated by X.

Sequencing of phages after the third round of selection revealed that the population had converged on consensus sequences primarily involving the residues immediately adjacent to the BuK residue. At the -1 position 40% of the sequenced clones contained a phenylalanine residue while 30% of clones contained a leucine or valine residue.<sup>‡</sup> At the +1 position, there was a strong preference for short aliphatic side chains with 55% of the sequenced clones containing isoleucine, valine, alanine, or leucine at this position. There was a slight preference for serine or threonine residues at position +3 with 30% of clones containing one of these two residues. However, it should be noted that not all clones contained a +3 position within the randomized sequence. Of those that did contain a +3, 66% contained serine or threonine. Finally, 25% of the sequenced clones converged on the unique consensus Cys – Thr – <sup>Val</sup>/<sub>Phc</sub> – BuK – <sup>Val</sup>/<sub>Ile</sub>.

Two recent studies have also utilized peptide libraries in selections against SIRT2. In one study, Morimoto et al. utilized a macrocyclic peptide library armed with a trifluoroacetyl-lysine warhead to search for inhibitors of SIRT2.<sup>231</sup> In a second study, Rauh et al. used an acetylated peptide microarray to investigate the deacetylases substrate specificity of SIRT2.<sup>232</sup> Consistent with our findings, both of these studies found that SIRT2 shows only a preference for the residues immediately adjacent to the acylated lysine. In screening their cyclic peptide library Morimoto et al. found that greater than 70% of peptides contained either a V or L at position -1. Further, Rauh et al. found that V, L, and I were three of the four most preferred residues at position -1 in

---

<sup>‡</sup> In discussing the peptide sequences we number the residues relative to their position with respect to BuK where -1 is immediately N-terminal to BuK and +1 is immediately C-terminal to BuK etc. (see Figure 5.16).

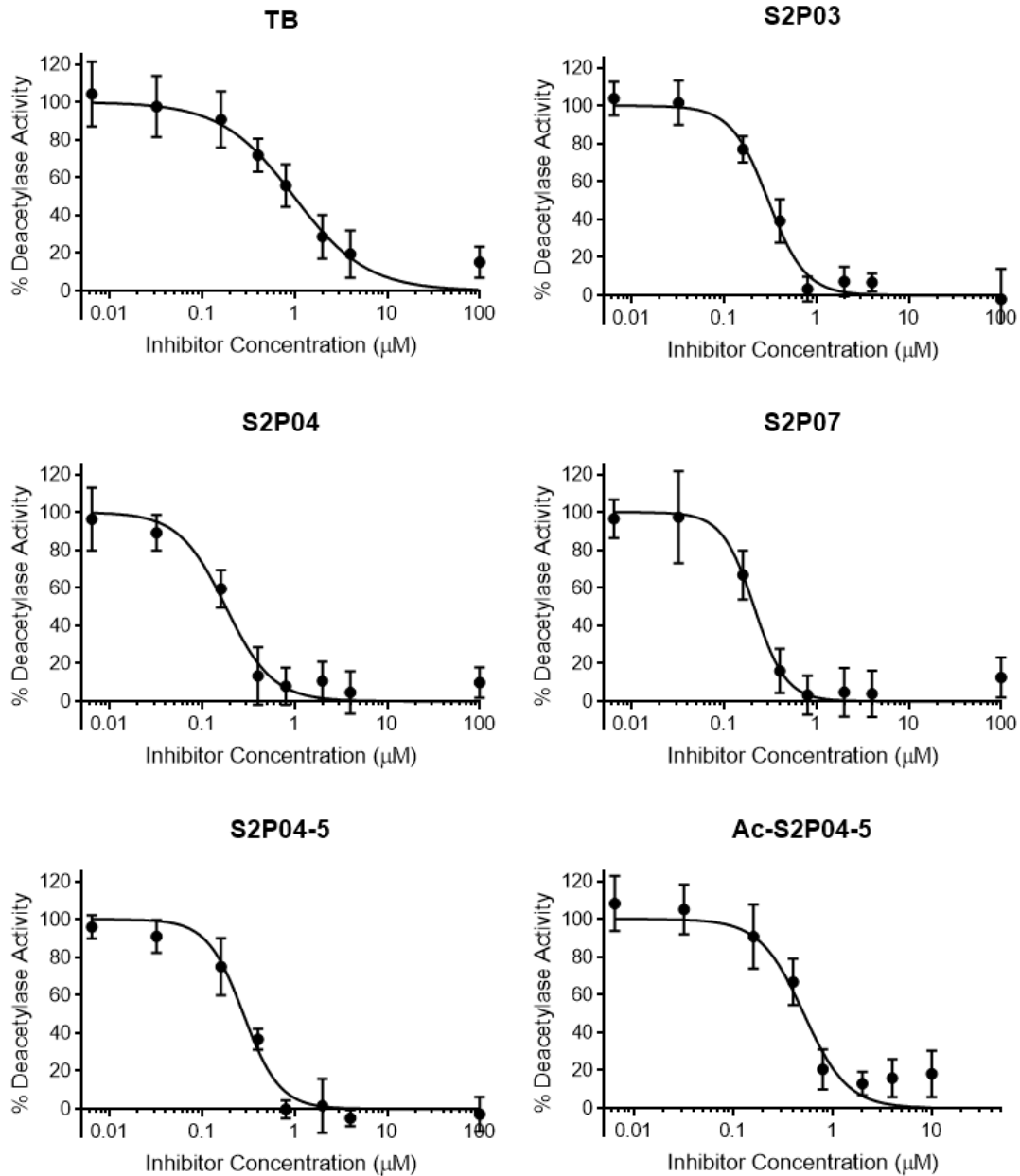
their screen of a peptide microarray. These findings are consistent with our results in which 30% of the clones contained an L or V at the -1 position. However, neither study reported an F at position -1 which constituted 40% of clones in our study. Interestingly, all three studies give conflicting results for which residue is preferred at position +1. The cyclic peptide library and peptide microarray report a strong preference for either R or Y/S at this position, respectively; whereas, we observed a strong enrichment of amino acids with aliphatic side chains at position +1. The fact that these studies have shown inconsistencies in the preferred residue at the +1 position may be attributed to an inability of SIRT2 to discriminate between residues at this position. Indeed, the crystal structure of a macrocyclic peptide identified by Morimoto et al. in complex with SIRT2 showed that the R at position +1 is pointing away from the surface of the enzyme and not in visible contact with any SIRT2 residues.<sup>233</sup> Further, crystal structures and isothermal titration calorimetry data of Sir2Tm, a bacterial ortholog of SIRT2, suggest that this residue is not important for substrate recognition.<sup>234</sup> In our study, 66% of clones containing a +3 position in the randomized sequence had either an S or T at this position. This is consistent with both the peptide array study, which found S to be one of the two most preferred residues at this position, and the cyclic peptide library screen which found an S at +3 in the most potent cyclic peptide inhibitor. The crystal structure of this cyclic peptide in complex with SIRT2 revealed that the S at position +3 makes hydrogen bonding interactions with Q265 of SIRT2 providing a possible explanation for why this residue is preferred at this position. Efforts are currently underway to obtain the crystal

structure of SIRT2 in complex with peptides identified in this study which will provide more detailed structural insight into the enzyme-substrate interactions.

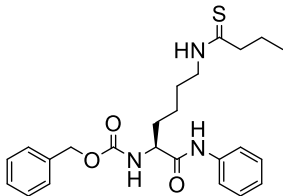
### 5.3.3 Synthetic Peptides Containing thBuK Inhibit Sirtuin 2

Next, we asked whether these peptides, when synthesized with thBuK in place of BuK could serve as inhibitors of SIRT2. To investigate this, we synthesized three of the peptides identified from our selection that had converged on the Cys – Thr – Val<sup>al</sup>/Phe – BuK – Val<sup>al</sup>/Ile consensus. Peptides were synthesized via manual Fmoc-based solid-phase peptide synthesis and their inhibitory potency measured against recombinant SIRT2 in a discontinuous inhibition assay as described in the experimental details. As a reference we used compound **53** (TB) which was recently demonstrated to be a potent, isoform selective inhibitor of SIRT2.<sup>214</sup> Figure 5.17 shows the inhibition curves for the inhibition of SIRT2 with TB and the synthetic peptides containing thBuK. The data show that all three of the synthetic peptides inhibited SIRT2 significantly more potently than TB. The most potent inhibitor, S2P04, inhibited SIRT2 with an IC<sub>50</sub> of 184 ± 19 nM which is 3.7-fold lower than the IC<sub>50</sub> of TB (Table 5.1).

Since the consensus sequence identified from our selection involved only five of the seven residues in the library, we synthesized a shortened version of S2P04 and measured its ability to inhibit SIRT2. The shortened peptide, S2P04-5, contained only the five residues of the consensus sequence, Cys – Thr – Val – thBuK – I. Interestingly, S2P04-5 inhibited SIRT2 with an IC<sub>50</sub> nearly double that of S2P04 (Table 5.1). This



**Figure 5.17. Inhibition curves for the inhibition of sirtuin 2 with TB and synthetic peptides containing thBuK.** Reaction mixtures contained 1 μM SIRT2, 2 mM NAD<sup>+</sup>, 50 μM Boc-(Ac)Lys-AMC, and 6.4 nM – 100 μM of the synthetic peptide or TB. Error bars represent one standard deviation of the mean of three independent experiments.

Compound	Structure	IC <sub>50</sub> , nM
TB		697 ± 59
S2P03	NH <sub>2</sub> -C-T-V-X-T-S-L-NH <sub>2</sub>	290 ± 58
S2P04	NH <sub>2</sub> -T-C-T-V-X-I-G-NH <sub>2</sub>	184 ± 19
S2P07	NH <sub>2</sub> -C-T-F-X-V-P-T-NH <sub>2</sub>	198 ± 30
S2P04-5	NH <sub>2</sub> -C-T-V-X-I-NH <sub>2</sub>	339 ± 53
Ac-S2P04-5	Ac-C-T-V-X-I-NH <sub>2</sub>	400 ± 48

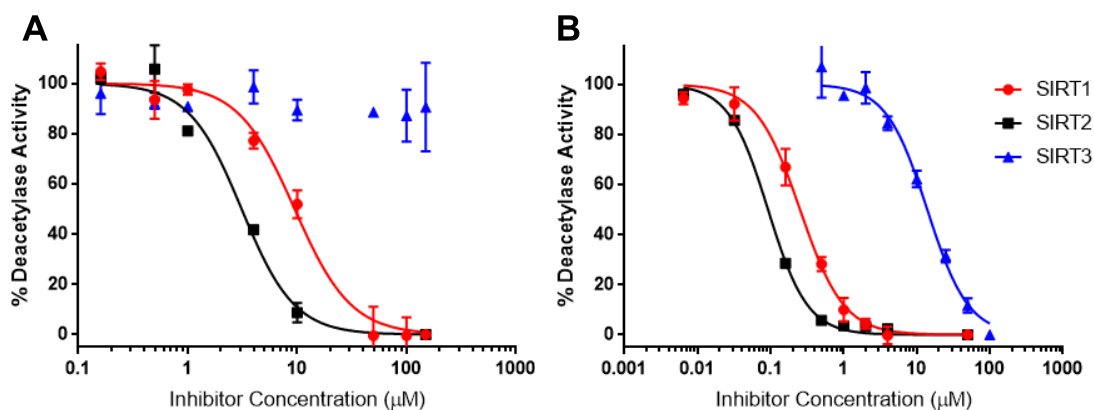
**Table 5.1.** IC<sub>50</sub> values measured for TB and synthetic peptides with sirtuin 2. X = thBuK

observation suggests that while SIRT2 shows no clear preference for specific residues, amino acids at these positions can still contribute to the substrate-enzyme interaction. This likely arises from interactions not involving side chains, such as backbone hydrogen bonds. In addition, we synthesized a version of S2P04-5 (Ac-S2P04-5) in which the N-terminal amine was capped with an acetyl group thereby removing the positive charge on the N-terminal amine. Ac-S2P04-5 inhibited SIRT2 with an IC<sub>50</sub> similar to S2P04-5 indicating that the charge of the N-terminus is likely not involved in the substrate-enzyme interaction.

#### 5.3.4 Selected Peptides Containing thBuK Exhibit Isoform Selectivity Comparable to TB

Next, we measured the potency of S2P04 against two other sirtuin isoforms that also recognize lysine butylation, namely sirtuin 1 (SIRT1), which is a nuclear sirtuin, and sirtuin 3 (SIRT3), which is located in the mitochondria.<sup>202</sup> TB has been shown to inhibit both SIRT1 and SIRT3 but is most potent against SIRT2.<sup>222</sup> The IC<sub>50</sub> of S2P04 and TB was measured against recombinant SIRT1, SIRT2, and GST-SIRT3 in a continuous inhibition assay as described in the experimental details. As shown in Figure 5.18 both S2P04 and TB displayed selectivity for SIRT2 over SIRT1 and SIRT3. S2P04 inhibited all three enzymes significantly more potently than TB, however, the selectivity of S2P04 for SIRT2 over other sirtuin isoforms was comparable to that of TB. For example, the IC<sub>50</sub> of TB was 3-fold lower for SIRT2 compared to SIRT1 while the IC<sub>50</sub> of S2P04 measured for SIRT2 was 2.7-fold lower than that measured for SIRT1 (Table 5.2). The low solubility of TB at concentrations greater than 150  $\mu$ M prevented an accurate determination of the IC<sub>50</sub> of TB with SIRT3. Collectively, these results suggest that while the amino acids adjacent to thBuK can lead to increased potency, acyl group identity is the driving force governing isoform selectivity for our synthetic peptides.

Given that these peptides were selected for binding to SIRT2, one might ask: why do they show no improvement in isoform selectivity over TB? In fact the observation that amino acid identity does not contribute to isoform selectivity is not entirely surprising. Current studies do not agree on the degree to which peptide amino acid sequence is a determinant of substrate specificity, however; many agree that



**Figure 5.18. Inhibition curves for the inhibition of sirtuin 1, sirtuin 2, and sirtuin 3 with TB and S2P04.** (A) Inhibition by TB. (B) Inhibition by S2P04. Reaction mixtures contained 0.25  $\mu\text{M}$  of the sirtuin enzyme, 1 mM  $\text{NAD}^+$ , 5  $\mu\text{M}$  fluorogenic substrate, and 0.16 – 150  $\mu\text{M}$  TB or 6.4 nM – 100  $\mu\text{M}$  S2P04. Error bars represent one standard deviation of the mean of two independent experiments.

Compound	$\text{IC}_{50}$ , $\mu\text{M}$		
-	SIRT1	SIRT2	SIRT3
TB	$9.5 \pm 0.9$	$3.1 \pm 0.3$	>150
S2P04	$0.28 \pm 0.01$	$0.096 \pm 0.006$	$16 \pm 7$

**Table 5.2.  $\text{IC}_{50}$  values measured for TB and S2P04 with sirtuin 1, sirtuin 2, and sirtuin 3.** S2P04 inhibited SIRT1 and SIRT2 ~33-fold more potently than TB.

the role of amino acid sequence is likely minimal.<sup>224, 235</sup> Indeed, some studies have reported robust deacetylase activity *in vitro* for SIRT1-2 across a range of substrates irrespective of amino acid sequence.<sup>198, 236</sup> Several hypotheses have been proposed to explain the apparent lack of sequence specificity of sirtuin-catalyzed deacetylation. One hypothesis supported by recent evidence suggests that protein fold, not amino acid



sequence, may be the primary determinant governing substrate specificity for SIRT1-3.<sup>237</sup> Another intriguing hypothesis suggests that the sirtuins may have evolved as group of protein quality-control enzymes to combat non-enzymatic lysine acylation by acyl-CoA.<sup>238, 239</sup> In any case, the fact that sirtuins display little sequence specificity explains why our selected peptides do not display greater isoform specificity—though they were selected for binding to SIRT2.

#### **5.4 Conclusion**

In this study, we have demonstrated the use of phage display with an expanded genetic code to generate peptide libraries containing naturally occurring lysine posttranslational modifications. We screened a peptide library containing  $N_\epsilon$ -butyryl-lysine to identify peptide sequences containing the ncAA that bind to the lysine deacylase SIRT2. We show that these peptides, when synthesized with a slowly-reacting analog of  $N_\epsilon$ -butyryl-lysine, are potent inhibitors of sirtuin 2-catalyzed lysine deacylation. The results of this work may help to elucidate subtle substrate preferences of SIRT2 and will aid in the rational design of SIRT2-selective inhibitors. Efforts are currently underway to solve the crystal structure of SIRT2 in complex with peptides identified in this study. These results may provide a rational explanation for why these peptides were specifically enriched in our selection. In this work we also demonstrated the ability of a PylRS mutant to facilitate the genetic incorporation of thBuK in response to TAG codons in *E. coli*. While the efficiency with which thBuK was incorporated was not sufficient to

support phage expression, the genetic incorporation of thBuK into other proteins, such as histones, could provide a useful tool for studying sirtuin activity.

**CHAPTER VI**

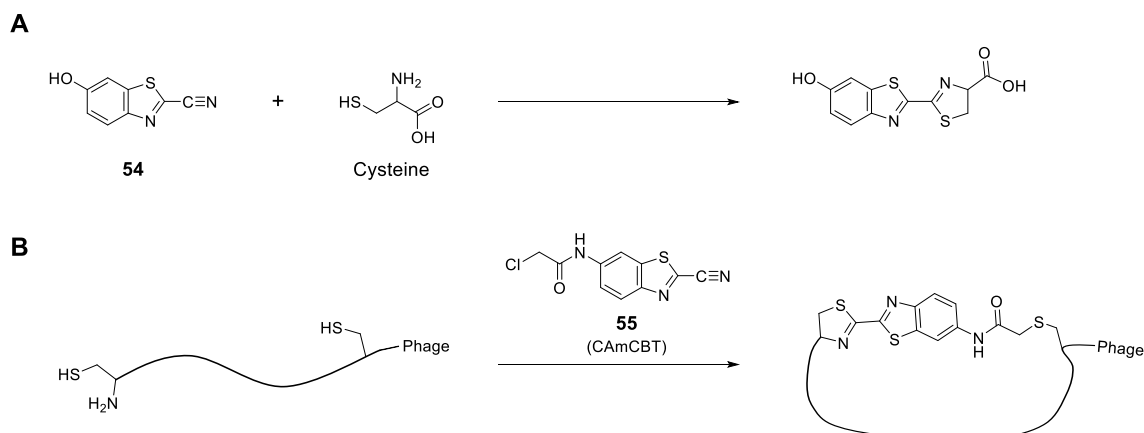
**EFFORTS TOWARDS THE SELECTION OF PHAGE-DISPLAYED CYCLIC  
PEPTIDES USING A REGIOSPECIFIC, KINETICALLY CONTROLLED,  
BIS-ELECTROPHILE**

**6.1 Introduction**

As described in Chapter 2, cyclization can be used as a general strategy to increase the binding affinity and proteolytic resistance of peptides. Phage display has been used as a tool to screen libraries of cyclic peptides for the identification of novel ligands for numerous targets. Current methods for cyclizing phage-displayed peptides utilize fixed cysteine residues that flank the randomized peptide sequence. In one such method, peptides are cyclized via an intrapeptide disulfide bond that forms upon oxidation of the cysteines during phage expression.<sup>8</sup> Phage-displayed cyclic peptide libraries based on disulfide bonds have been widely used for the identification of cyclic peptide ligands; however, this method is limited by the fact that disulfide bonds are labile, being easily broken by common reducing agents used to maintain protein stability and activity. More recent methods for cyclizing phage-displayed peptides have utilized organic bis-electrophiles which react with each of the nucleophilic cysteines to form the cyclic peptide via an organic linker.<sup>106</sup> This method affords cyclic peptides with more stable linkages that cannot be broken by commonly used buffer components. However, current compounds that are used for cyclizing phage displayed peptides consist of electrophilic alkyl halides or Michael acceptors that are nonspecific and will react with cysteines

present elsewhere in the phage.<sup>69</sup> This is problematic as wildtype Ff phages contain eight additional cysteine residues in pIII, and reaction of these cysteines with organic electrophiles leads to a significant reduction in phage viability.<sup>107, 114</sup>

With the goal of screening phage-displayed cyclic peptide libraries, we sought to identify more regiospecific electrophiles that will react with cysteines flanking the library, but spare cysteines elsewhere in the phage. To this end, we turned to the biocompatible condensation reaction between cysteine and 6-hydroxy-2-cyanobenzothiazole (**54**, Figure 6.1A). This reaction occurs rapidly ( $9.19 \text{ M}^{-1} \cdot \text{s}^{-1}$ ) and irreversibly in aqueous solvents at neutral pH. Rao and colleagues have demonstrated that **54** reacts specifically with 1,2- and 1,3-aminothiols.<sup>240</sup> This specificity allowed them to develop derivatives of **54** that can be used to label proteins at N-terminal cysteine residues while sparing internal cysteines.<sup>241</sup> We reasoned that the specificity of **54** for N-terminal cysteines could be used to cyclize peptides displayed at the N-terminus of pIII on Ff bacteriophages. Therefore, we designed the bis-electrophile derivative of **54**, **55** (CAmCBT), which contains a chloroacetamide moiety. We predicted that CAmCBT could be used to cyclize peptides bearing N-terminal and internal cysteines that are displayed on the surface of bacteriophages. The CBT moiety of CAmCBT is predicted to react with the N-terminal cysteine, while the chloroacetamide will react with an internal cysteine to give the cyclic product (Figure 6.1B). Although the chloroacetamide moiety can also react with the cysteine residue at the N-



**Figure 6.1. The reaction of 6-hydroxy-2-cyanobenzothiazole with cysteine and the reaction of CAMcBT with phage-displayed peptides.** (A) Cysteine and 6-hydroxy-2-cyanobenzothiazole (**54**) react to give luciferin. The reaction proceeds in aqueous solvents at neutral pH. (B) Phage-displayed peptides containing N-terminal and internal cysteines are predicted to react with **55** to give a cyclic peptide product.

terminus, the reaction of CBT with this cysteine is ~400 times faster than the reaction of chloroacetamide with this cysteine.<sup>240, 242</sup> Therefore, the desired cyclic product is kinetically favored.

In this chapter, we report the synthesis of CAMcBT and demonstrate its ability to form the desired cyclic peptide product when reacted with peptides containing N-terminal and internal cysteine residues. The reaction proceeds with high yields in only three hours, at room temperature, in aqueous solvents, and at neutral pH. We use CAMcBT to cyclize peptides on a phage-displayed dodecapeptide library and screen this library against two model proteins. Future directions for this ongoing work are also discussed.

## 6.2 Experimental Details

### 6.2.1 Miscellaneous Information

All primers were purchased from Integrated DNA Technologies, Inc. (Coralville, PA). Enzymes and buffers for molecular cloning were purchased from New England Biolabs (Ipswich, MA). DNA sequencing was performed by Eton Biosciences, Inc. (San Diego, CA). Mass spectra were collected by the Department of Chemistry Mass Spectrometry Facility at Texas A&M University.

### 6.2.2 Primer List

TEV-CAC-sfGFP-F: 5'-CGGCCGCGGCCTGCGTTAGCAAAGGTGAAGAACTG-3'

TEV-CAC-sfGFP-R: 5'-CGGCGCACTGAAAATACAGGTTTTCCATGGTTAATTCCTCCTG-3'

12mer-pIII-NcoI-F: 5'-CGGCCATGGCCTGC(NNK)<sub>12</sub>TGCGCGGCGAAAGCGGCCGGCCC-3'

12mer-pIII-NcoI-R: 5'-CATGCCATGGCCGGCTGGGCCGC-3'

### 6.2.3 Plasmid Construction

pBAD-TEV-CA<sub>5</sub>C-sfGFP: The plasmid pBAD-sfGFP has been described by us previously.<sup>146</sup> To introduce the polyalanine model peptide and TEV protease recognition sequence to the N-terminus of sfGFP, the plasmid pBAD-sfGFP was PCR amplified with the 5'-end primer, TEV-CA<sub>5</sub>C-sfGFP-F, and the 3'-end primer, TEV-CA<sub>5</sub>C-sfGFP-R using Phusion high-fidelity DNA polymerase. The PCR product was phosphorylated with T4 polynucleotide kinase, followed by ligation with T4 DNA ligase, and used to transform *E. coli* Top10.

12mer Peptide-pIII Library: The plasmid pADLg3 (Antibody Design Labs, San Diego, CA) was amplified with Phusion high-fidelity DNA polymerase using the primers 12mer-pIII-*NcoI*-F and 12mer-pIII-*NcoI*-R. The product was digested with *NcoI*, ligated overnight with T4 DNA ligase, and used to transform electrocompetent *E. coli* Top10 (1 ng DNA per  $\mu\text{L}$  of cells). Transformed cells were diluted 1:9 in 2xYT and incubated at 37°C. After 1 hour, 100  $\mu\text{L}$  was removed for to quantify transformation efficiency and the remaining culture was diluted to  $\text{OD}_{600} < 0.2$  in fresh 2xYT media, supplemented with 100  $\mu\text{g}\cdot\mu\text{L}^{-1}$  ampicillin, and grown at 37°C overnight. In total  $1.1 \times 10^9$  transformants were obtained. The following day, the plasmid was extracted using a commercial extraction kit (Qiagen) and used to transform electrocompetent *E. coli* ER2738 exactly as described above yielding  $1.8 \times 10^9$  transformants.

#### 6.2.4 Protein Expression and Purification

TEV-CA<sub>5</sub>C-sfGFP: *E. coli* Top10 containing the plasmid pBAD-TEV-CA<sub>5</sub>C-sfGFP-His<sub>6</sub> were grown at 37°C in 1 L of 2xYT media, containing 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, to  $\text{OD}_{600} = 0.6$  at which point protein expression was induced with the addition of 0.2% arabinose to the growth media. 18 h post induction, cell pellets were harvested by centrifugation (6k rpm, 20 min). The cells were lysed and the protein purified by Ni-NTA exactly as described in Chapter 2. Following Ni-NTA purification, the buffer was changed to TEV cutting buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT) by diluting 1:9 in TEV cutting buffer and concentrating by ultrafiltration (10 kDa MWCO; 2x). The protein was then incubated with TEV protease at a molar ratio of 1:500

(enzyme:substrate) for 48 h at 4°C. Following digestion with TEV protease, the protein was loaded onto a 10 mL Mono S column pre-equilibrated with Mono S buffer A (50 mM NH<sub>4</sub>PO<sub>4</sub> pH 4.4, 50 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol) the protein was eluted with a linear gradient from 0-60% Mono S buffer B (50 mM NH<sub>4</sub>PO<sub>4</sub> pH 4.4, 1 M NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol) over 20 min, followed by 60-100% over 1 h. The fractions containing pure protein were combined, concentrated to 60 μM, flash frozen, and stored at -80°C. Protein concentration was determined by measuring the absorbance of the GFP chromophore at 485 nm using the reported extinction coefficient of 83,300 M<sup>-1</sup>·cm<sup>-1</sup>.<sup>243</sup>

*His<sub>7</sub>-TEV-HRas(1-169)*: The plasmid pQLink-His<sub>7</sub>-TEV-HRas(1-169) was kind gift from Professor Sharon Campbell at the University of North Carolina, Chapel Hill. *E. coli* BL21(DE3) containing the plasmid pQLink-His<sub>7</sub>-TEV-HRas(1-169) were grown at 37°C in 2 L of 2xYT media, containing 100 μg·mL<sup>-1</sup> ampicillin, to OD<sub>600</sub> = 0.8 at which point the temperature was changed to 30°C and HRas expression was induced with the addition of 1 mM IPTG. 8 h post induction, cell pellets were harvested by centrifugation (6k rpm, 20 min) and stored at -80°C until purification.

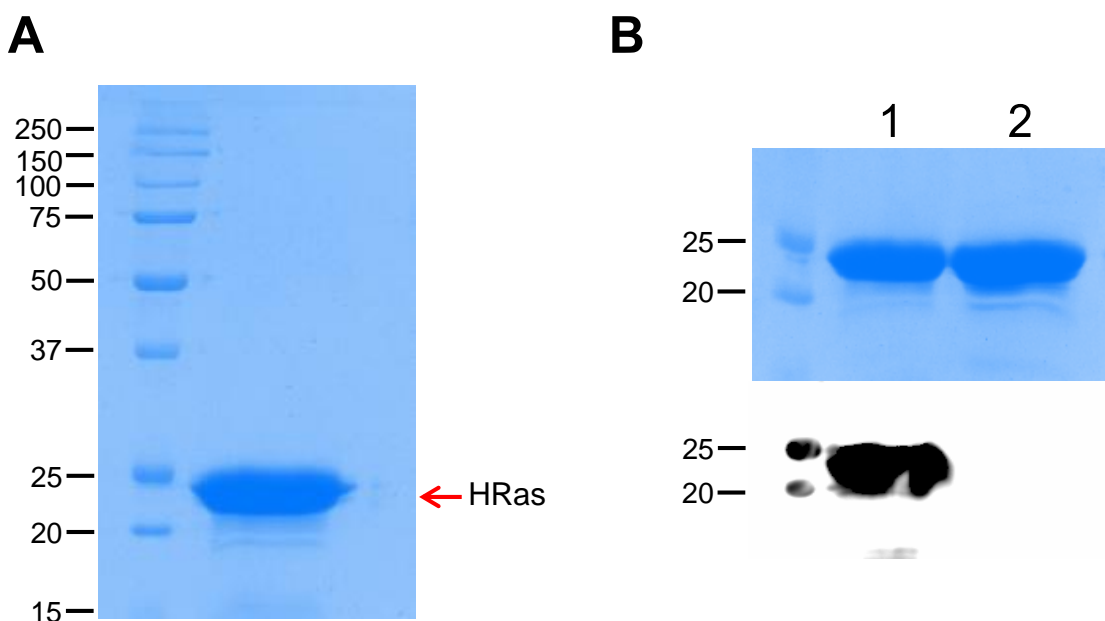
Cell pellets were resuspended in 40 mL of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub> pH 8.0, 300 mM NaCl, 10 mM imidazole) and lysed by sonication. The sonicated resuspension was clarified by centrifugation (10k rpm, 45 min) and the clarified supernatant was passed through a disposable column containing 2 mL of high affinity Ni<sup>2+</sup>-charged resin (Genscript). The resin was washed with 40 mL of lysis buffer and the bound protein



eluted with 7 mL of elution buffer (lysis buffer containing 250 mM imidazole). The eluted protein was concentrated by ultrafiltration (10 kDa MWCO) to 4 mL and then dialyzed overnight at 4°C against 2x 4 L Q Sepharose buffer A (64 mM Tris pH 7.5, 50 mM NaCl, 10 mM MgCl<sub>2</sub>; 3 kDa MWCO). The following day, the protein solution was concentrated to 1 mL and applied to a 31 mL Q Sepharose column (GE Healthcare) that was pre-equilibrated with Q Sepharose buffer A. The column was washed with 2 CV of buffer A (2 mL·min<sup>-1</sup>) and the bound protein eluted with a linear gradient of 0-100% buffer B (64 mM Tris pH 7.5, 1 M NaCl, 10 mM MgCl<sub>2</sub>) over 3 CV. The fractions containing pure HRas were concentrated and dialyzed overnight at 4°C against 3x 4 L of PBS pH 7.4. The purified protein was quantified using a Pierce<sup>TM</sup> BCA Protein Assay Kit, with BSA as a standard, according to the manufacturer's protocol. SDS-PAGE analysis of HRas purified using this protocol is shown in Figure 6.2A.

#### *6.2.5 Biotinylation of HRas*

The purified HRas was concentrated to 89 µM and biotinylated with a 5-fold molar excess of EZ-Link<sup>TM</sup> NHS-Biotin (Thermo Fisher Scientific) for 2 h at room temperature. After 2 h the solution was directly loaded onto a Superdex 75 10/300 GL size exclusion column (GE Healthcare) pre-equilibrated with 64 mM Tris pH 7.5, 50 mM NaCl, 30 mM MgCl<sub>2</sub>. The eluted protein was concentrated by ultrafiltration (10 kDa MWCO) to 1 mg·mL<sup>-1</sup>, flash frozen in 10 µL aliquots, and stored at -80°C. Biotinylation of the protein was confirmed via Western blot using a streptactin-HRP conjugate (Bio-Rad) which is shown in Figure 6.2B.



**Figure 6.2. SDS-PAGE analysis of purified HRas.** (A) SDS-PAGE of HRas purified using the above protocol. (B) Western blot to confirm biotinylation of HRas. The protein before (Lane 2) and after (Lane 1) treatment with NHS-biotin was blotted onto a nitrocellulose membrane and then visualized with a streptactin-HRP conjugate (Bio-Rad).

#### 6.2.6 CA<sub>5</sub>C-sfGFP Cyclization with CAMCBT

Purified CA<sub>5</sub>C-sfGFP was diluted to 10  $\mu$ M in PBS pH 7.4. To this was added a freshly prepared solution of tris(2-carboxyethyl)phosphine (TCEP) to a final concentration of 1 mM and the solution was incubated at 37°C. After 30 min, CAMCBT in DMSO was added to a final concentration of 50  $\mu$ M (5% DMSO final) and the solution was incubated at room temperature for 3 hours. The reaction was quenched with the addition of 10 mM L-cysteine and the buffer was changed to 10 mM NH<sub>4</sub>CO<sub>3</sub> by dilution followed by concentration via ultrafiltration (3 kDa MWCO) three times. The concentrated solution was lyophilized and then analyzed by electrospray ionization mass spectrometry.

### 6.2.7 Phage Expression and Purification

*E. coli* ER2738 containing the 12-mer peptide-pIII phagemid library were grown at 37°C in 250 mL of 2xYT media supplemented with 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, 12  $\mu\text{g}\cdot\text{mL}^{-1}$  tetracycline, and 1% glucose. Upon reaching  $\text{OD}_{600} = 0.5-0.6$ , 50 mL of the culture was transferred to a small flask and infected with M13KO7d3 helper phage (Antibody Design Labs, San Diego, CA;  $\text{MOI} = 25$ ) at 37°C with shaking. After 45 min the cells were pelleted (4k rpm, 15 min) and resuspended in 100 mL of 2xYT media containing 100  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin, 25  $\mu\text{g}\cdot\text{mL}^{-1}$  kanamycin, and 1 mM IPTG. 12-18 h post induction, the culture was transferred to two 50 mL tubes, cells were pelleted (6k rpm, 20 min), and the upper 40 mL was decanted into new tubes. Phages were precipitated twice with the addition of sodium chloride and polyethylene glycol solution as described in Chapter 4. Phages were resuspended in 3-4 mL of PBS pH 7.4 for panning against HRas or 2-3 mL of kallikrein binding buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 10 mM  $\text{MgCl}_2$ , 1 mM CaCl) for panning against plasma kallikrein.

### 6.2.8 Labelling Phages with FITC-CBT

The purified phage library or M13KO7 (Antibody Design Labs) in PBS (pH 7.4) was incubated with 1 mM TCEP for 30 min at 37°C followed by the addition of 50  $\mu\text{M}$  FITC-CBT. The reaction was incubated at room temperature for 2.5 h followed by precipitation of the proteins with the addition of trichloroacetic acid. The precipitated protein pellet was washed twice with 500  $\mu\text{L}$  of ice-cold acetone and the fluorescence was visualized under blue light.

### *6.2.9 Phage Infectivity Assay*

For determining the effect of treatment with CA<sub>m</sub>CBT on phage infectivity, the purified phagemid library in PBS (pH 7.4) was incubated under various reaction conditions including: (1) PBS only, (2) 1 mM TCEP for 30 min at 37°C followed by 3 h at room temperature, (3) 30 min at 37°C followed by the addition of 50 μM CA<sub>m</sub>CBT and incubation for 3 h at room temperature, and (4) 1 mM TCEP for 30 min at 37°C followed by the addition of 50 or 100 μM CA<sub>m</sub>CBT and incubation for 3 h at room temperature. Following incubation under the various conditions, the phages were quantified via the colony forming unit assay described in Chapter 4.

### *6.2.10 Peptide Library Cyclization with CA<sub>m</sub>CBT*

The purified phage library ( $10^{11}$ - $10^{12}$  cfu) in PBS or kallikrein binding buffer (pH 7.4) was incubated with 1 mM TCEP at 37°C for 30 min. After 30 min, CA<sub>m</sub>CBT in DMSO (5% DMSO final) was added to a final concentration of 100 μM. The solution was incubated for 3 h at room temperature with slow rotation. Following the reaction, the library was directly used for affinity selection without further purification.

### *6.2.11 Affinity Selection against HRas*

The cyclized phage library ( $10^{11}$ - $10^{12}$  cfu) in PBS was panned against 10 μg of biotinylated HRas immobilized on streptavidin coated magnetic beads (rounds 1, 3, and 4) or neutravidin coated polystyrene plates (round 2) exactly as described in Chapter 5 with the following modifications. The HRas ligand, guanosine triphosphate (30 μM),

was added to the phage library after cyclization and to all blocking and binding buffers for each round of selection. For round 4, the library was incubated with the immobilized target for 20 rather than 30 min and the number of washes was increased from 10 to 15.

#### *6.2.12 Affinity Selection against Plasma Kallikrein*

Biotin labelled human plasma kallikrein was purchased from Molecular Innovations (Novi, MI). The cyclized phage library ( $10^{11}$  cfu) in kallikrein binding buffer was panned against 10  $\mu$ g of plasma kallikrein immobilized on streptavidin coated magnetic beads (rounds 1 and 3) or neutravidin coated polystyrene plates (round 2) exactly as described in Chapter 5 with the following modifications. The blocking buffer consisted of 1% (w/v) BSA and 0.1% Tween-20 in kallikrein binding buffer, and the washing buffer was kallikrein binding buffer containing 0.1% Tween-20. Following library cyclization with CA<sub>m</sub>CBT, 0.25 mL of 5x blocking buffer was added to 1 mL of the cyclized library. The library was incubated for 30 min in the 1x blocking buffer before incubating with the immobilized kallikrein.

#### *6.2.13 Solid-Phase Peptide Synthesis*

The peptide HRP24 was synthesized via manual Fmoc-based solid-phase peptide synthesis on Rink-amide resin, using *N,N,N',N'*-tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uranium hexafluorophosphate (HBTU) as a coupling agent, essentially as described in Chapter 5 with the following modifications. Fmoc-Lys(5/6-FAM)-OH was coupled for 2.5 h with 3 rather than 4 eq. of the Fmoc-protected amino acid. The next 10 residues

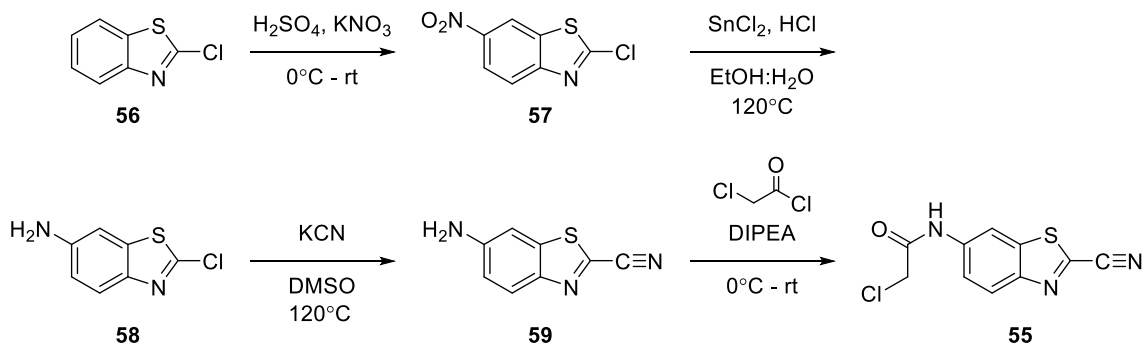
were coupled with 4 eq. for 40 min each. The final 6 residues were coupled with 4 eq. twice, once for 15 min followed by a second coupling reaction for 30 min. The crude peptide was dissolved in 20% acetonitrile in water containing 0.1% TFA and purified by semi-preparative RP-HPLC on a C18 column (Alltech, 10 x 250 mm, 10  $\mu$ m, 100 Å pore). Solvent A: water with 0.1% TFA, Solvent B: acetonitrile with 0.1% TFA. The peptide was eluted with a linear gradient from 20-100% solvent B over 15 min. Fractions containing the purified peptide were combined and lyophilized to afford a fluffy orange powder. The mass of the peptide was confirmed by MALDI-TOF mass spectrometry.

#### *6.2.14 Synthetic Peptide Cyclization*

The peptide HRP24 was dissolved in PBS (pH 7.4) to a concentration of 10  $\mu$ M. To this was added a freshly prepared solution of TCEP to a final concentration of 1 mM. The solution was incubated for 30 min at 37°C and then CAmCBT in DMSO was added to a final concentration of 50  $\mu$ M (5% DMSO final). The solution was incubated at room temperature for 2 h and then purified and concentrated by C<sub>18</sub> Zip-Tip® (Millipore Sigma) according to the manufacturer's protocol. The mass of the purified sample was analyzed by MALDI-TOF mass spectrometry.

#### *6.2.15 Chemical Synthesis*

*The Synthesis of CAmCBT*: The synthetic scheme shown in Figure 6.3 was followed for the synthesis of CAmCBT (**55**).



**Figure 6.3. The synthesis of CAMcBT.**

To a 250 mL round-bottom-flask containing  $\text{H}_2\text{SO}_4$  (81.85 mL) at  $0^\circ\text{C}$  was added **56** (11.51 mL, 88.4 mmol), dropwise, followed by  $\text{KNO}_3$  (9.87 g, 97.3 mmol, 1.1 eq.). The solution was stirred for 30 min at  $0^\circ\text{C}$  followed by 1 h at room temperature and then poured onto ice giving rise to a yellow precipitate (**57**). The precipitate was collected by vacuum filtration, washed with cold water (2x 5 mL), and then recrystallized from ethanol. The crystals were collected by vacuum filtration, washed with cold ethanol (3x, 5 mL), and dried. The recrystallized product (11.50 g, 60.6% yield) was used for the next step without further purification.

In a 500 mL round-bottom-flask, **57** (3.00 g, 13.9 mmol) was dissolved, with light heating, in a solution of 234 mL ethanol, 156 mL water, and 31.18 mL  $\text{HCl}$  (4.8 M). To this was added  $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$  (31.52 g, 139.7 mmol, 10 eq.) and the resulting solution was refluxed at  $120^\circ\text{C}$  until complete consumption of **57** as monitored by TLC. After approximately 3.5 h, the solution was cooled and made basic by adding aqueous  $\text{NaOH}$  (5 M) affording a yellow precipitate. The precipitate was removed by filtration

and the supernate was concentrated under reduced pressure. The concentrated solution was extracted with ethyl acetate (3x 33 mL). The organic layer was washed with brine, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, and evaporated. The crude solid was purified by flash chromatography (ethyl acetate/hexane = 1:1) to give **58** as an off white solid (3.03 g, 37% yield).

In a 250 mL round-bottom-flask, KCN (0.845 g, 17.04 mmol) was dissolved in 130 mL DMSO and the solution was refluxed at 135°C overnight under an atmosphere of argon. The next day **58** (0.900 g, 4.87 mmol) in 18.1 mL DMSO was added to the solution dropwise while stirring. The resulting solution was refluxed at 120°C until complete consumption of **58** as determined by TLC at which point the solution was cooled to room temperature and poured into a solution containing 87 mL of 1 M KH<sub>2</sub>PO<sub>4</sub> and 130.5 mL of diethyl ether. The organic layer was removed and the aqueous layer was extracted with ethyl acetate (3x 20 mL). The combined organic layers were washed with brine, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, and evaporated. The crude solid was purified by flash chromatography (ethyl acetate/hexane = 1:4 to 1:2) to afford **59** (0.131 g, 15%) as a yellow solid.

To a solution of the above amine **59** (0.120 g, 0.685 mmol) in dry DCM (5 mL), cooled in an ice bath, was added diisopropylethylamine (0.252 mL, 2 eq.) and a solution of chloroacetyl chloride (0.06 mL, 1 eq.). The reaction was stirred at room temperature overnight and the volatiles removed under reduced pressure. The residue was purified by silica gel column chromatography eluting with ethyl acetate in hexane to afford the compound **55** (0.115 g, 59%) as a yellow solid.



$^1\text{H}$  NMR ( $\text{CDCl}_3$  300 MHz):  $\delta$  8.71 (d, 1H,  $J = 4.5$  Hz), 8.35 (br, s, 1H), 8.23 (d, 1H,  $J = 7.3$  Hz), 7.53 (dd, 7H,  $J = 4.5, 8.6$  Hz) 4.21 (s, 2H).

The Synthesis of FITC-CAmCBT:

The synthetic scheme shown in Figure 6.4 was followed for the synthesis of FITC-CBT (**61**).

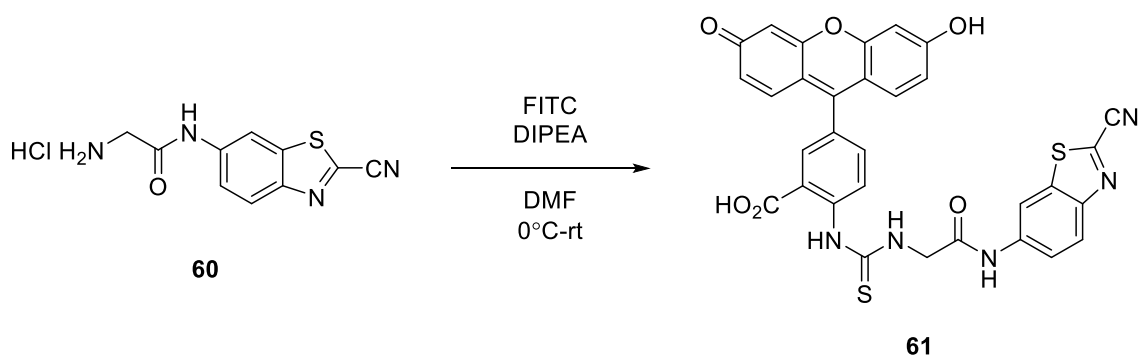


Figure 6.4. The synthesis of FITC-CBT.

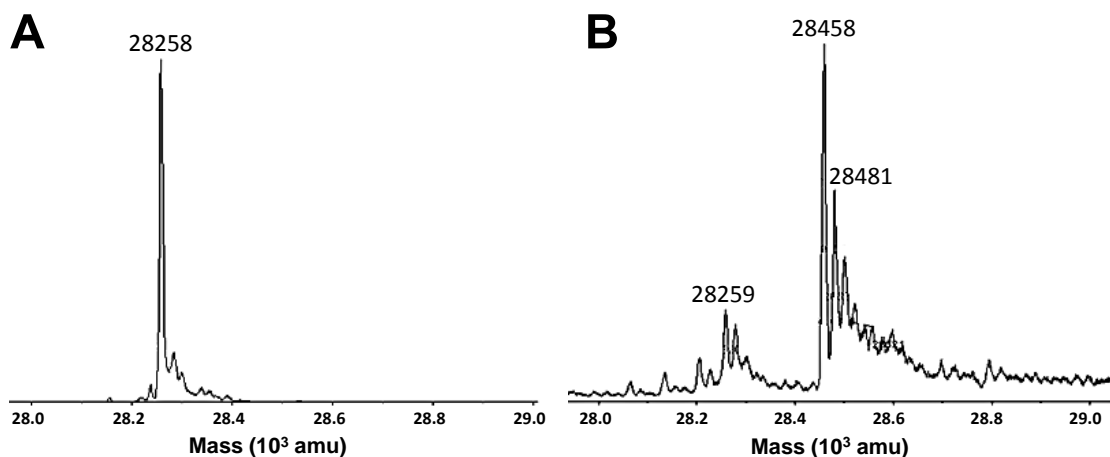
The amine **60** (50 mg, 0.18 mmol) was dissolved in a solution of diisopropylethyl amine (41  $\mu\text{L}$ , 0.22 mmol) in dry methanol and THF (2:1, 1.2 mL). The solution was cooled to  $0^\circ\text{C}$  before adding fluorescein-5-isothiocyanate (FITC, 72 mg, 0.18 mmol). The reaction mixture was warmed to room temperature and stirred overnight before concentrating and purifying by column chromatography on silica gel using a gradient of DCM:acetone:MeOH (80:10:10 to 0:90:10) to afford compound **61** (71 mg, 62%) as an orange solid.

$^1\text{H}$  NMR (DMSO- $d_6$ ), 300 MHz)  $\delta$  8.59 (d,  $J = 2.1$  Hz, 1H), 8.43 (s, 1H), 8.37 (d,  $J = 1.8$  Hz, 1H), 8.11-8.08 (m, 1H), 8.02 (2, 1H), 7.82-7.70 (m, 2H), 7.28 (d,  $J = 2.1$  Hz, 1H), 6.70-6.56 (m, 6H), 4.49 (d,  $J = 2.1$  Hz, 2H).

## 6.3 Results and Discussion

### 6.3.1 Cyclization of a Model Peptide with CA<sub>m</sub>CBT

In order to determine whether CA<sub>m</sub>CBT could react with peptides containing two cysteines to give the predicted cyclic product, we constructed a peptide-GFP fusion protein on which to test the reaction. A short sequence of five alanine residues, flanked on both sides by cysteine residues, was introduced as an N-terminal fusion to superfolder green fluorescent protein (CA<sub>5</sub>C-sfGFP) via PCR cloning as described in the experimental details. To ensure quantitative removal of the N-terminal methionine, we introduced a TEV protease recognition sequence immediately upstream of the first cysteine. Treatment of the protein with TEV protease following purification removes all upstream residues to afford a free N-terminal cysteine. The peptide-GFP fusion was expressed in *E. coli* Top10 followed by purification via nickel affinity chromatography and subsequent cation exchange chromatography after treatment with TEV protease. Analysis of the purified CA<sub>5</sub>C-sfGFP by electrospray ionization mass spectrometry (ESI-MS) revealed predominantly one peak with a mass that agreed with the theoretical value (Figure 6.5A). Next, the purified protein (10  $\mu\text{M}$ ) was incubated with tris(2-carboxyethyl)phosphine (TCEP, 1 mM) for 30 minutes at 37°C to ensure that the



**Figure 6.5. ESI mass spectra of CA<sub>5</sub>C-sfGFP before and after treatment with CAmbBT.** (A) Mass spectrum of CA<sub>5</sub>C-sfGFP before treatment with CAmbBT (Theoretical mass = 28259 Da). (B) Mass spectrum of CA<sub>5</sub>C-sfGFP after treatment with 50  $\mu$ M CAmbBT (Theoretical mass = 28458 Da). Other peaks in the spectrum correspond to unreacted CA<sub>5</sub>C-sfGFP and the sodium adduct of the cyclic product.

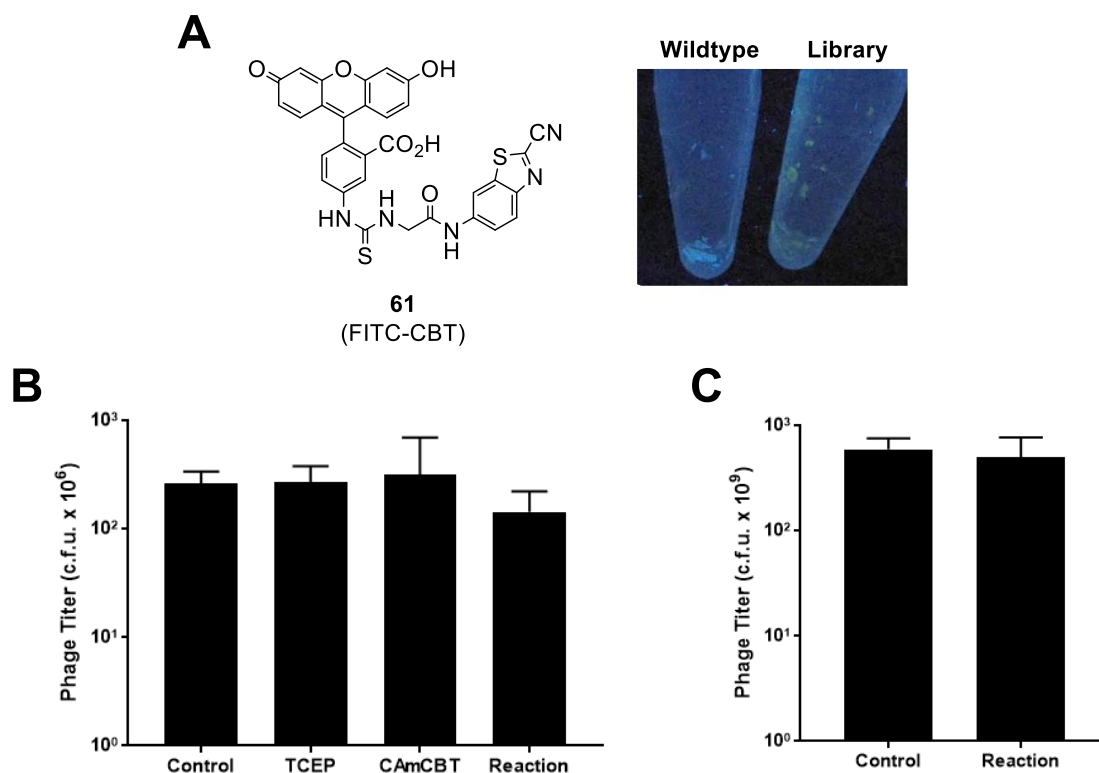
cysteines were fully reduced. Treating the reduced protein with five equivalents of CAmbBT, for three hours at room temperature, in phosphate buffered saline (PBS) resulted in a shift of the major peak on the mass spectrum ( $\Delta m/z = 199.0$ ) corresponding to formation of the cyclic product (Figure 6.5B). The relative intensity of the mass peaks indicated greater than 80% conversion to the cyclic product in just three hours. Incubating the protein with CAmbBT in the absence of TCEP led to only minimal conversion to the cyclic product (data not shown).

### 6.3.2 Cyclization of Phage-Displayed Peptides

Encouraged by our results with the model peptide, we moved to test the reaction on a phage-displayed peptide library. We started by constructing a peptide library flanked by N-terminal and internal cysteine residues. To do this, we utilized site-saturation

mutagenesis to introduce a library of twelve randomized residues, and two fixed cysteines, as a genetic fusion to the phage coat protein pIII. The library was constructed using the pADL-10b phagemid system that, when complimented with a pIII-defective helper phage, allows for peptide display on all five copies of pIII. Transformation of *E. coli* ER2738 following cloning yielded  $1.8 \times 10^9$  unique transformants. In this expression system, secretion of pIII into the periplasm for phage packaging is driven by the pelB leader sequence, a 22 residue peptide fused to the N-terminus of pIII.<sup>244</sup> Once in the periplasm, the leader sequence is removed by an endogenous *E. coli* leader peptidase.<sup>183, 245</sup> Previous studies have shown that the pelB leader sequence can be used for the *in vivo* generation of N-terminal cysteine residues and this has been applied for the expression of proteins for expressed protein ligation.<sup>246</sup> We designed our library such that cleavage of the leader sequences *in vivo* affords the free N-terminal cysteine required for reaction with CAmCBT. Incubating phages expressed from this library with a fluorescein isothiocyanate derivative of CBT, **61** (FITC-CBT) led to a precipitated phage pellet that was visibly fluorescent, indicative of the reaction of CBT with the N-terminal cysteine (Figure 6.6A). In contrast, no fluorescence was detected when FITC-CBT was incubated with wildtype M13. This suggests that the CBT moiety does not react with cysteines present elsewhere in the phage.

Next, we tested the impact of reaction with CAmCBT on phage viability. As discussed in the introduction, nonspecific reaction of CAmCBT with endogenous phage cysteines can lead to a significant reduction in the ability of phages to infect their host.



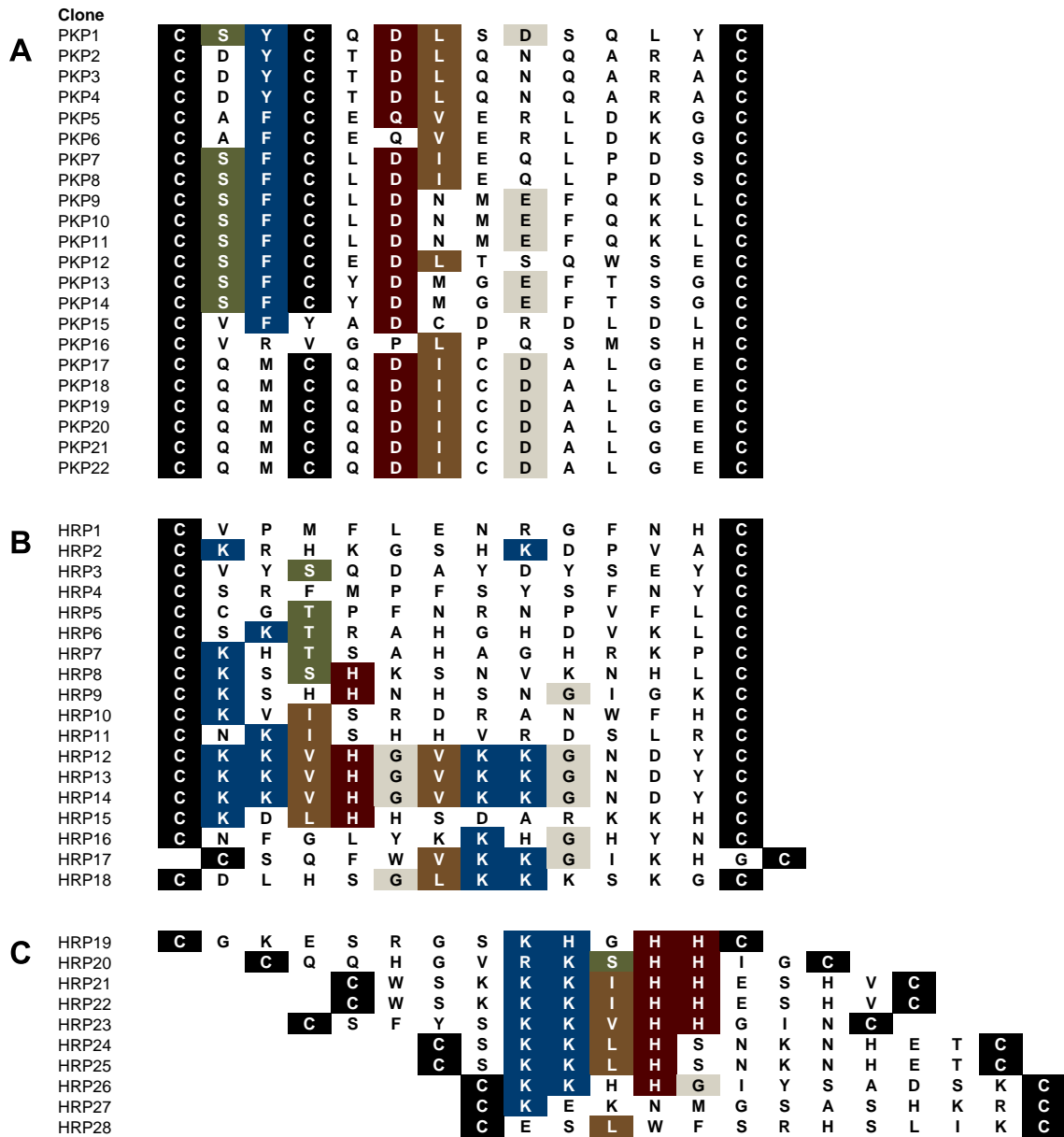
**Figure 6.6. Labelling phages with FITC-CBT and the effect of CAmCBT on phage viability.** (A) Labelling of phages with FITC-CBT. Phages from the library contain an N-terminal cysteine and are labeled with FITC-CBT, whereas wildtype phages are not. (B) The effect of 1 mM TCEP, 50  $\mu$ M CAmCBT (CAmCBT), or both (Reaction) on the number of infective phages. (C) The effect of 1 mM TCEP and 100  $\mu$ M CAmCBT (Reaction) on the number of infective phages. Control indicates the titer of phages incubated in PBS only. Error bars represent one standard deviation of the mean of three experiments.

To test the effect of cyclization with CAmCBT on phage viability, we utilized a colony forming unit assay in which the number of infective phages was quantified before and after cyclization with CAmCBT, as well as after incubation with the individual reaction components. As shown in Figure 6.6B incubating the phages under the reaction conditions led to a slight (1.8-fold) reduction in the number of infective phages when 50  $\mu$ M of CAmCBT was used. Higher concentrations of CAmCBT resulted in a similar

reduction in the number of infective phages of about 1.2-fold (Figure 6.6C). This low reduction in phage viability can likely be attributed to the greater selectivity of CAmCBT for N-terminal cysteines compared to other reported compounds.<sup>107, 108</sup>

### 6.3.3 Affinity Selection against HRas and Plasma Kallikrein

Next, we performed affinity selection using the cyclic peptide library against two model proteins namely, the GDP-bound form of the GTPase HRas and the human plasma protease kallikrein. Briefly, the biotinylated targets were first immobilized by capture on streptavidin coated magnetic beads or neutravidin coated magnetic plates. Next, the phage library that had been modified with CAmCBT was incubated with the immobilized target for 30 minutes before washing to remove nonspecifically bound phages and elution of specifically bound clones with a low pH buffer. Sequencing of clones after the third round of selection against plasma kallikrein ( $n = 22$ ) revealed that the population had converged on a strong consensus sequence of the type Ser – <sup>Phe</sup>/<sub>Tyr</sub> – Cys – Xxx – Asp – <sup>Leu</sup>/<sub>Ile</sub> (Figure 6.7A). Sequencing of clones from the selection against HRas ( $n = 18$ ) revealed a less convincing consensus (Figure 6.7B). Therefore, we performed an additional round of affinity selection against HRas in which more stringent binding and washing conditions were used. Eight out of the ten clones sequenced after the fourth round of selection contained a consensus sequence of the type Lys – Lys – Xxx – His – His. In total, 16 of the 28 clones sequenced from rounds three and four contained the consensus Lys – Lys – <sup>Val</sup>/<sub>Leu</sub> – His – His (Figure 6.7C).



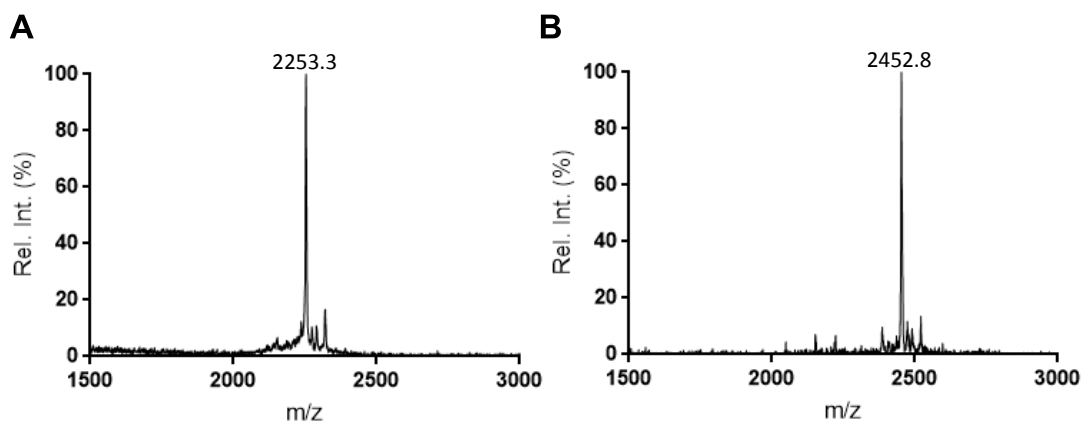
**Figure 6.7. The amino acid sequence of peptides identified from affinity selection against plasma kallikrein and HRAs. (A)** The sequence of peptides identified after the third round of affinity selection against plasma kallikrein. **(B)** The sequence of peptides identified after the third round of affinity selection against HRAs. **(C)** Peptides identified after the fourth round of affinity selection against HRAs.

#### *6.3.4 Modification of Synthetic Peptides with CAmCBT and First Attempts towards Measuring Binding Affinity*

After identifying consensus sequences from affinity selection, we next turned to measure the binding affinity of the selected peptides with HRas. To do this, we designed an assay based on fluorescence polarization utilizing synthetic, fluorescently-labeled, peptides. We synthesized the peptide HRP24 containing an Ala – Ala – Lys(5/6-FAM) linker at the amidated C-terminus via manual Fmoc-based solid-phase peptide synthesis. In, this peptide, the Lys residue at the C-terminus is conjugated at the sidechain amine to 5/6-carboxyfluorescein. The crude peptide was purified by reverse phase HPLC and then, following reduction with TCEP, incubated with five equivalents of CAmCBT, in PBS, for 2 hours at room temperature. MALDI-TOF mass spectrometry of the peptide before and after treatment with CAmCBT revealed a shift in the major peak corresponding to expected mass of the cyclic product (Figure 6.8). Notably, no unconjugated peptide was detected in the mass spectrum after only two hours.

Unfortunately, we noticed at this time that cyclization with CAmCBT led to a significant reduction in the solubility of HRP24. At high concentrations of the peptide this was observed as the formation of an orange precipitate shortly after the addition of CAmCBT. Although no precipitate was observed at lower concentrations of the peptide, a color change of the solution from green to orange was indicative of the formation of insoluble aggregates. The low solubility of the synthetic peptides, we feared, would interfere with the accurate measurements of binding affinity in our assay. Indeed,





**Figure 6.8. MALDI-TOF mass spectra of the peptide HRP24 before and after reaction with CAmCBT. (A)** Mass spectrum of HRP24 before reaction with CAmCBT (Theoretical mass = 2254.2 Da). **(B)** Mass spectrum of HRP24 after reaction with CAmCBT (Theoretical mass = 2453.2).

contrary to theory, the fluorescence polarization of the synthetic peptide decreased with the addition of HRAs. We are in the process of synthesizing peptides identified from selection against plasma kallikrein to see whether this observation is sequence dependent.

### 6.3.5 Future Directions

The low solubility of HRP24 when conjugated with CAmCBT interferes with our ability to measure the binding of the peptide to HRAs via fluorescence polarization. The high solubility of sfGFP may explain why no precipitation was observed in our preliminary trials in which we used CAmCBT to cyclize CA<sub>5</sub>C-sfGFP. Indeed, genetic fusion to sfGFP has been used as a general strategy to enhance the solubility of poorly soluble proteins.<sup>167, 247</sup> Based on these studies and our own observations, we plan to utilize peptide-sfGFP fusion proteins to measure the potential binding and inhibitory

properties of peptide sequences identified from selection against HRas and plasma kallikrein. Several of the peptides identified in selections against both targets have already been cloned and expressed as fusions to the N-terminus of sfGFP. Our preliminary results have shown that, when expressed as fusions to sfGFP, these peptides, including HRP24, remain soluble following conjugation with CAmCBT. Efforts are currently underway in our laboratory to measure the inhibition of plasma kallikrein using the aforementioned peptide-sfGFP fusion proteins.

#### **6.4 Conclusion**

In this study we have demonstrated the ability of a newly synthesized bis-electrophile (CAmCBT) to react with peptides bearing N-terminal and internal cysteines and yield cyclic peptide products. The reaction occurs in aqueous solvents, at neutral pH, to afford the cyclic product in high yield. Mass spectrometry revealed greater than 80% conversion to the cyclic product in 3 h when CAmCBT was used to cyclize a peptide fused to the N-terminus of sfGFP. Further, synthetic peptides underwent quantitative conversion to the cyclic product in just 2 h. The high selectivity of the CBT moiety for N-terminal cysteine residues resulted in a lower reduction in phage viability compared to other commonly used electrophiles for peptide cyclization. We used CAmCBT to cyclize a peptide library, displayed on the N-terminus of pIII on the M13 bacteriophage, and identified strong consensus sequences when the cyclic peptide library was panned against HRas and plasma kallikrein. Efforts are currently underway to measure the binding and potential inhibitory properties of the identified peptide sequences.

## CHAPTER VII

### CONCLUDING REMARKS

Over the past 30 years, phage display has emerged as a widely-used and powerful tool for the directed evolution of peptide ligands. In this thesis we sought to improve upon the utility of phage display by expanding the chemical diversity of phage-displayed peptides. In particular, we sought to utilize the technique of genetic code expansion, through nonsense suppression, to genetically incorporate non-canonical amino acids (ncAAs) in response to redefined TAG codons.

Throughout this thesis we have demonstrated a remarkable ability of the pyrrolysyl-tRNA synthetase (PylRS) to facilitate the genetic incorporation of a wide variety of ncAAs with diverse structural features. We began our studies by exploring the expanded substrate spectrum of a rationally designed PylRS mutant containing the N346A/C348A mutations (PylRS(N346A/C348A)). This mutant had been shown to facilitate the genetic incorporation of phenylalanine derivatives with *meta* and *para* substituents while remaining orthogonal to canonical amino acids.<sup>146, 147, 149</sup> The data presented in this thesis demonstrate that PylRS(N346A/C348A) can also recognize phenylalanine derivatives with substitutions at the *ortho* position as a substrate. A recent publication by our laboratory also demonstrated the ability of PylRS(N346A/C348A) to facilitate the genetic incorporation of histidine derivatives.<sup>248</sup> Although polyspecificity appears to be a general feature of evolved aaRSs<sup>170</sup>, the fact that this one enzyme mutant can facilitate the genetic incorporation of nearly 50 ncAAs with high efficiency is

striking. Polyspecificity of aaRSs is ideal for enzymes that will be used to incorporate ncAAs for phage display as this allows for the library to be expressed with different ncAAs using the same cell strain.

We have also demonstrated, in this thesis, a new method for constructing phage display libraries that results in a library in which every clone contains an in-frame TAG codon. This type of library eliminates propagation-related bias that has previously interfered with selections involving ncAAs. Using TAG-obligate libraries, in conjunction with PylRS mutants, we have vastly expanded the number of ncAAs that have been incorporated into phage-displayed peptides. These ncAAs include those that contain reactive handles, such as the ketone-containing ncAA (**35**) whose genetic incorporation was first reported in this work. These reactive functional groups can be used to selectively conjugate the phage-displayed peptides to further expand their chemical diversity.

The ncAAs we have incorporated into phage-displayed peptides also include two naturally occurring lysine posttranslational modifications, namely,  $N_\epsilon$ -butyryl-lysine and  $N_\epsilon$ -crotonyl-lysine. Phage-displayed peptide libraries containing posttranslational modifications will, no doubt, be useful tools for studying the proteins that interact with these modifications. For example, these libraries could be used to profile the substrate specificity of an enzyme that binds to a particular modification. They will also be useful to identify sequence-optimized peptides, containing the modification, which can be used as inhibitors of enzymes that interact with the modification. As a demonstration of this, we utilized our TAG-obligate heptapeptide library containing  $N_\epsilon$ -butyryl-lysine to select

for peptides that bind to the lysine deacylase sirtuin 2. Peptide sequences identified from this selection were synthesized with the structural analog of  $N_{\epsilon}$ -butyryl-lysine,  $N_{\epsilon}$ -thiobutyryl-lysine, and were shown to be potent inhibitors of sirtuin 2-catalyzed lysine deacylation. Various methods have been reported for the genetic incorporation of other posttranslational modifications and we anticipate that peptide libraries containing these modifications will be useful for the field of epigenetics.<sup>127</sup>

Herein, we also reported the design and synthesis of a regiospecific, kinetically-controlled, bis-electrophile for cyclizing phage-displayed peptides. In designing this compound, we coupled cyanobenzothiazole with chloroacetamide giving the compound CAMCBT (**55**), which reacts with peptides containing N-terminal and internal cysteines yielding a cyclic product. Kinetics ensure that CAMCBT reacts with the peptide in a particular order giving only one product. Yields of the reaction were high (>80%) when CAMCBT was used to cyclize a peptide-sfGFP fusion protein, and synthetic peptides reacted with CAMCBT quantitatively giving only the desired product. Little impact on phage viability was observed when phage-displayed peptides were cyclized with CAMCBT, owing to the specific reactivity of the compound. Efforts to select phage-displayed peptides that were cyclized with CAMCBT against two model proteins yielded strong consensus sequences. Though we are still working to measure the binding affinity and potential inhibitory properties of these peptides, this work demonstrates that the cyanobenzothiazole moiety can be used to design compounds that react with phage-displayed peptides in a regiospecific manner. We anticipate that tris-electrophiles based

on cyanobenzothiazole could also be used to generate phage-displayed bicyclic peptide libraries.

## REFERENCES

- [1] Uhlig, T., Kyprianou, T., Martinelli, F. G., Oppici, C. A., Heiligers, D., Hills, D., Calvo, X. R., and Verhaert, P. (2014) The emergence of peptides in the pharmaceutical business: from exploration to exploitation, *EuPA Open Proteom.* 4, 58-69.
- [2] Fosgerau, K., and Hoffmann, T. (2015) Peptide therapeutics: current status and future directions, *Drug Discov. Today* 20, 122-128.
- [3] Sachdeva, S. (2017) Peptides as 'drugs': the journey so far, *Int. J. Pept. Res. Ther.* 23, 49-60.
- [4] Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. (2013) The future of peptide-based drugs, *Chem. Biol. Drug. Des.* 81, 136-147.
- [5] Dietrich, U., Durr, R., and Koch, J. (2013) Peptides as drugs: from Screening to application, *Curr. Pharm. Biotechno.* 14, 501-512.
- [6] Vlieghe, P., Lisowski, V., Martinez, J., and Khrestchatisky, M. (2010) Synthetic therapeutic peptides: science and market, *Drug Discov. Today* 15, 40-56.
- [7] Ullman, C. G., Frigotto, L., and Cooley, R. N. (2011) In vitro methods for peptide display and their applications, *Brief Funct. Genomics* 10, 125-134.
- [8] Smith, G. P., and Petrenko, V. A. (1997) Phage display, *Chem. Rev.* 97, 391-410.
- [9] Clokie, M. R. J., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011) Phages in nature, *Bacteriophage* 1, 31-45.

- [10] Bergh, O., Borsheim, K. Y., Bratbak, G., and Haldal, M. (1989) High abundance of viruses found in aquatic environments, *Nature* 340, 467-468.
- [11] Marks, T., and Sharp, R. (2000) Bacteriophages and biotechnology: a review, *J. Chem. Technol. Biot.* 75, 6-17.
- [12] Christie, G. E. (1999) Propagation of viruses | Bacteria A2 - Granoff, Allan, In *Encyclopedia of Virology (Second Edition)* (Webster, R. G., Ed.), pp 1413-1418, Elsevier, Oxford.
- [13] Rakhuba, D. V., Kolomiets, E. I., Dey, E. S., and Novik, G. I. (2010) Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell, *Pol. J. Microbiol.* 59, 145-155.
- [14] Salmond, G. P. C., and Fineran, P. C. (2015) A century of the phage: past, present and future, *Nat. Rev. Microbiol.* 13, 777-786.
- [15] Little, J. W. (2005) Lysogeny, prophage induction, and lysogenic conversion, In *Phages*, American Society of Microbiology.
- [16] Jiang, J., AbuShilbayeh, L., and Rao, V. B. (1997) Display of a porA peptide from *Neisseria meningitidis* on the bacteriophage T4 capsid surface, *Infect. Immun.* 65, 4770-4777.
- [17] Gamkrelidze, M., and Dąbrowska, K. (2014) T4 bacteriophage as a phage display platform, *Arch. Microbiol.* 196, 473-479.
- [18] Rosenberg, A., Griffin, K., Studier, F. W., McCormick, M., Berg, J., Novy, R., and Mierendorf, R. (1996) T7Select® phage display system: a powerful new protein display system based on bacteriophage T7, (Novagen, Ed.).



- [19] Sternberg, N., and Hoess, R. H. (1995) Display of peptides and proteins on the surface of bacteriophage lambda, *Proc. Natl. Acad. Sci. U.S.A.* 92, 1609-1613.
- [20] Bazan, J., Calkosinski, I., and Gamian, A. (2012) Phage display-A powerful technique for immunotherapy 2. Vaccine delivery, *Hum. Vacc. Immunother.* 8, 1829-1835.
- [21] Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D., and Russel, M. (2011) Filamentous bacteriophage: biology, phage display and nanotechnology applications, *Curr. Issues. Mol. Biol.* 13, 51-75.
- [22] Marvin, D. A., and Hoffmann-Berling, H. (1963) A fibrous DNA phage (fd) and a spherical RNA phage (Fr) specific for male strains of *E. coli*. II. Physical characteristics, *Z. Naturforsch. B* 18, 884-893.
- [23] Newman, J., Swinney, H. L., and Day, L. A. (1977) Hydrodynamic properties and structure of fd virus, *J. Mol. Biol.* 116, 593-603.
- [24] Day, L. A., and Berkowitz, S. A. (1977) The number of nucleotides and the density and refractive index increments of fd virus DNA, *J. Mol. Biol.* 116, 603-606.
- [25] Dotto, G. P., Enea, V., and Zinder, N. D. (1981) Functional-analysis of bacteriophage-F1 intergenic region, *Virology* 114, 463-473.
- [26] Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J., and Marvin, F. J. (1974) Filamentous bacterial viruses: XII. Molecular architecture of class I (fd, If1, IKE) Virion, *J. Mol. Biol.* 88, 581-582.

- [27] Grant, R. A., Lin, T. C., Konigsberg, W., and Webster, R. E. (1981) Structure of the filamentous bacteriophage-FI - location of the minor coat protein-a, protein-C and protein-D, *J. Biol. Chem.* 256, 539-546.
- [28] Endemann, H., and Model, P. (1995) Location of filamentous phage minor coat proteins in phage and in infected cells, *J. Mol. Biol.* 250, 496-506.
- [29] Marvin, D. A. (1998) Filamentous phage structure, infection and assembly, *Curr. Opin. Struc. Biol.* 8, 150-158.
- [30] Gailus, V., Ramsperger, U., Johner, C., Kramer, H., and Rasched, I. (1994) The role of the adsorption complex in the termination of filamentous phage assembly, *Res. Microbiol.* 145, 699-709.
- [31] Beck, E., and Zink, B. (1981) Nucleotide-sequence and genome organization of filamentous bacteriophages FI and fd, *Gene* 16, 35-58.
- [32] Nelson, F. K., Friedman, S. M., and Smith, G. P. (1981) Filamentous phage DNA cloning vectors: a non-Infective mutant with a non-polar deletion in gene-III, *Virology* 108, 338-350.
- [33] Crissman, J. W., and Smith, G. P. (1984) Gene-III protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis, *Virology* 132, 445-455.
- [34] Armstrong, J., Perham, R. N., and Walker, J. E. (1981) Domain-structure of bacteriophage fd adsorption protein, *FEBS Lett.* 135, 167-172.

- [35] Stengele, I., Bross, P., Garces, X., Giray, J., and Rasched, I. (1990) Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites, *J. Mol. Biol.* 212, 143-149.
- [36] Lubkowski, J., Hennecke, F., Pluckthun, A., and Wlodawer, A. (1998) The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p, *Nat. Struct. Biol.* 5, 140-147.
- [37] Holliger, P., and Riechmann, L. (1997) A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd, *Structure* 5, 265-275.
- [38] Lubkowski, J., Hennecke, F., Pluckthun, A., and Wlodawer, A. (1999) Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA, *Structure* 7, 711-722.
- [39] Riechmann, L., and Holliger, P. (1997) The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*, *Cell* 90, 351-360.
- [40] Gray, C. W., Brown, R. S., and Marvin, D. A. (1981) Adsorption complex of filamentous fd virus, *J. Mol. Biol.* 146, 621-627.
- [41] Deng, L. W., and Perham, R. N. (2002) Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*, *J. Mol. Biol.* 319, 603-614.
- [42] Click, E. M., and Webster, R. E. (1997) Filamentous phage infection: required interactions with the TolA protein, *J. Bacteriol.* 179, 6464-6471.

- [43] Holliger, P., Riechmann, L., and Williams, R. L. (1999) Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Å: evidence for conformational lability, *J. Mol. Biol.* 288, 649-657.
- [44] Chatellier, J., Hartley, O., Griffiths, A. D., Fersht, A. R., Winter, G., and Riechmann, L. (1999) Interdomain interactions within the gene 3 protein of filamentous phage, *FEBS Lett.* 463, 371-374.
- [45] Eckert, B., Martin, A., Balbach, J., and Schmid, F. X. (2005) Prolyl isomerization as a molecular timer in phage infection, *Nat. Struct. Mol. Biol.* 12, 619-623.
- [46] Clarke, M., Maddera, L., Harris, R. L., and Silverman, P. M. (2008) F-pili dynamics by live-cell imaging, *Proc. Natl. Acad. Sci. U.S.A.* 105, 17978-17981.
- [47] Armstrong, J., Hewitt, J. A., and Perham, R. N. (1983) Chemical modification of the coat protein in bacteriophage-fd and orientation of the virion during assembly and disassembly, *EMBO J.* 2, 1641-1646.
- [48] Zenkin, N., Naryshkina, T., Kuznedelov, K., and Severinov, K. (2006) The mechanism of DNA replication primer synthesis by RNA polymerase, *Nature* 439, 617-620.
- [49] Meyer, T. F., and Geider, K. (1982) Enzymatic synthesis of bacteriophage fd viral DNA, *Nature* 296, 828-832.
- [50] Fulford, W., and Model, P. (1988) Regulation of bacteriophage f1 DNA replication. I. New functions for genes II and X, *J. Mol. Biol.* 203, 49-62.
- [51] Yen, T. S. B., and Webster, R. E. (1982) Translational control of bacteriophage f1 gene II and gene X proteins by gene V protein, *Cell* 29, 337-345.

- [52] Grant, R. A., and Webster, R. E. (1984) The bacteriophage-f1 morphogenetic signal and the gene V protein/phage single-stranded DNA complex, *Virology* 133, 329-340.
- [53] Rapoza, M. P., and Webster, R. E. (1995) The products of gene I and the overlapping in-Frame gene XI are required for filamentous phage assembly, *J. Mol. Biol.* 248, 627-638.
- [54] Brissette, J. L., and Russel, M. (1990) Secretion and membrane integration of a filamentous phage-encoded morphogenetic protein, *J. Mol. Biol.* 211, 565-580.
- [55] Marciano, D. K., Russel, M., and Simon, S. M. (1999) An aqueous channel for filamentous phage export, *Science* 284, 1516-1519.
- [56] Russel, M., Linderoth, N. A., and Sali, A. (1997) Filamentous phage assembly: variation on a protein export theme, *Gene* 192, 23-32.
- [57] Twort, F. W. (1915) An investigation on the nature of ultra-microscopic viruses, *Lancet* 2, 1241-1243.
- [58] Pires, D. P., Cleto, S., Sillankorva, S., Azeredo, J., and Lu, T. K. (2016) Genetically engineered phages: a review of advances over the last decade, *Microbiol. Mol. Biol. Rev.* 80, 523-543.
- [59] Vodnik, M., Zager, U., Strukelj, B., and Lunder, M. (2011) Phage display: selecting straws instead of a needle from a haystack, *Molecules* 16, 790-817.
- [60] Fuh, G., and Sidhu, S. S. (2000) efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein, *FEBS Lett.* 480, 231-234.

- [61] Jacobsson, K., Rosander, A., Bjerketorp, J., and Frykberg, L. (2003) Shotgun phage display - selection for bacterial receptors or other exported proteins, *Biol. Proced. Online* 5, 123-135.
- [62] Qi, H., Lu, H. Q., Qiu, H. J., Petrenko, V., and Liu, A. H. (2012) Phagemid vectors for phage display: properties, characteristics and construction, *J. Mol. Biol.* 417, 129-143.
- [63] Russel, M., Kidd, S., and Kelley, M. R. (1986) An improved filamentous helper phage for generating single-stranded plasmid DNA, *Gene* 45, 333-338.
- [64] Chen, W. Z., Zhu, Z. G., Feng, Y., Xiao, X. D., and Dimitrov, D. S. (2008) Construction of a large phage-displayed human antibody domain library with a scaffold based on a newly identified highly soluble, stable heavy chain variable domain, *J. Mol. Biol.* 382, 779-789.
- [65] Rakonjac, J., Jovanovic, G., and Model, P. (1997) Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3, *Gene* 198, 99-103.
- [66] Rondot, S., Koch, J., Breitling, F., and Dubel, S. (2001) A helper phage to improve single-chain antibody presentation in phage display, *Nat. Biotechnol.* 19, 75-78.
- [67] Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications, *Angew. Chem. Int. Edit.* 44, 7342-7372.
- [68] Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J.,

- Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zacco, M., Gherardi, E., and Winter, G. (1994) Isolation of high-affinity human-antibodies directly from large synthetic repertoires, *EMBO J.* 13, 3245-3260.
- [69] Ng, S., Jafari, M. R., and Derda, R. (2012) Bacteriophages and viruses as a support for organic synthesis and combinatorial chemistry, *ACS Chem. Biol.* 7, 123-138.
- [70] Gram, H., Schmitz, R., Zuber, J. F., and Baumann, G. (1997) Identification of phosphopeptide ligands for the Src-homology 2 (SH2) domain of Grb2 by phage display, *Eur. J. Biochem.* 246, 633-637.
- [71] Walchli, S., Espanel, X., Harrenga, A., Rossi, M., Cesareni, G., and van Huijsduijnen, R. H. (2004) Probing protein-tyrosine phosphatase substrate specificity using a phosphotyrosine-containing phage library, *J. Biol. Chem.* 279, 311-318.
- [72] Scholle, M. D., Kriplani, U., Pabon, A., Sishtla, K., Glucksman, M. J., and Kay, B. K. (2006) Mapping protease substrates by using a biotinylated phage substrate library, *ChemBioChem* 7, 834-838.
- [73] Dwyer, M. A., Lu, W. Y., Dwyer, J. J., and Kossiakoff, A. (2000) Biosynthetic phage display: a novel protein engineering tool combining chemical and genetic diversity, *Chem. Biol.* 7, 263-274.
- [74] Basle, E., Joubert, N., and Pucheault, M. (2010) Protein chemical modification on endogenous amino acids, *Chem. Biol.* 17, 213-227.

- [75] Li, K., Chen, Y., Li, S. Q., Huong, G. N., Niu, Z. W., You, S. J., Mello, C. M., Lu, X. B., and Wang, Q. A. (2010) Chemical modification of M13 bacteriophage and its application in cancer cell imaging, *Bioconjugate Chem.* 21, 1369-1377.
- [76] Yacoby, I., Shamis, M., Bar, H., Shabat, D., and Benhar, I. (2006) Targeting antibacterial agents by using drug-carrying filamentous bacteriophages, *Antimicrob. Agents Chemother.* 50, 2087-2097.
- [77] Bernard, J. M. L., and Francis, M. B. (2014) Chemical strategies for the covalent modification of filamentous phage, *Front. Microbiol.* 5.
- [78] Carrico, Z. M., Farkas, M. E., Zhou, Y., Hsiao, S. C., Marks, J. D., Chokhawala, H., Clark, D. S., and Francis, M. B. (2012) N-Terminal labeling of filamentous phage to create cancer marker imaging agents, *ACS Nano* 6, 6675-6680.
- [79] Ng, S., Jafari, M. R., Matochko, W. L., and Derda, R. (2012) Quantitative synthesis of genetically encoded glycopeptide libraries displayed on M13 phage, *ACS Chem. Biol.* 7, 1482-1487.
- [80] Sandman, K. E., Benner, J. S., and Noren, C. J. (2000) Phage display of selenopeptides, *J. Am. Chem. Soc.* 122, 960-961.
- [81] Gonzalez-Flores, J. N., Shetty, S. P., Dubey, A., and Copeland, P. R. (2013) The molecular biology of selenocysteine, *Biomol. Concepts* 4, 349-365.
- [82] Miller, J. H., and Albertini, A. M. (1983) Effects of surrounding sequence on the suppression of nonsense codons, *J. Mol. Biol.* 164, 59-71.



- [83] Sandman, K. E., and Noren, C. J. (2000) The efficiency of *Escherichia coli* selenocysteine insertion is influenced by the immediate downstream nucleotide, *Nucleic Acids Res.* 28, 755-761.
- [84] 't Hoen, P. A. C., Jirka, S. M. G., ten Broeke, B. R., Schultes, E. A., Aguilera, B., Pang, K. H., Heemskerk, H., Aartsma-Rus, A., van Ommen, G. J., and den Dunnen, J. T. (2012) Phage display screening without repetitious selection rounds, *Anal. Biochem.* 421, 622-631.
- [85] Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code, *Annu. Rev. Bioph. Biom.* 35, 225-249.
- [86] Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*, *Science* 292, 498-500.
- [87] Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine encoded by UAG in archaea: charging of a UAG-decoding specialized tRNA, *Science* 296, 1459-1462.
- [88] Krzycki, J. A. (2005) The direct genetic encoding of pyrrolysine, *Curr. Opin. Microbiol.* 8, 706-712.
- [89] Tharp, J. M., Ehnbohm, A., and Liu, W. R. (2017) tRNA<sup>Pyl</sup>: structure, function, and applications, *RNA Biology*, 1-12.
- [90] Javahishvili, T., Manibusan, A., Srinagesh, S., Lee, D., Ensari, S., Shimazu, M., and Schultz, P. G. (2014) Role of tRNA orthogonality in an expanded genetic code, *ACS Chem. Biol.* 9, 874-879.

- [91] Crnkovic, A., Suzuki, T., Söll, D., and Reynolds, N. M. (2016) Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion, *Croat. Chem. Acta* 89.
- [92] Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool, *BBA-Proteins Proteom* 1844, 1059-1070.
- [93] Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - expanding the chemistry in biology, *Chem. Sci.* 6, 50-69.
- [94] Feng, T., Tsao, M. L., and Schultz, P. G. (2004) A phage display system with unnatural amino acids, *J. Am. Chem. Soc.* 126, 15962-15963.
- [95] Liu, C. C., Mack, A. V., Tsao, M. L., Mills, J. H., Lee, H. S., Choe, H., Farzan, M., Schultz, P. G., and Smider, V. V. (2008) Protein evolution with an expanded genetic code, *Proc. Natl. Acad. Sci. U.S.A.* 105, 17688-17693.
- [96] Liu, C. C., Choe, H., Farzan, M., Smider, V. V., and Schultz, P. G. (2009) Mutagenesis and evolution of sulfated antibodies using an expanded genetic code, *Biochemistry* 48, 8891-8898.
- [97] Liu, C. C., Mack, A. V., Brustad, E. M., Mills, J. H., Groff, D., Smider, V. V., and Schultz, P. G. (2009) Evolution of proteins with genetically encoded "chemical warheads", *J. Am. Chem. Soc.* 131, 9616-9617.
- [98] Kang, M. C., Light, K., Ai, H. W., Shen, W. J., Kim, C. H., Chen, P. R., Lee, H. S., Solomon, E. I., and Schultz, P. G. (2014) Evolution of iron(II)-finger peptides by using a bipyridyl amino acid, *ChemBioChem* 15, 822-825.

- [99] Day, J. W., Kim, C. H., Smider, V. V., and Schultz, P. G. (2013) Identification of metal ion binding peptides containing unnatural amino acids by phage display, *Bioorg. Med. Chem. Lett.* *23*, 2598-2600.
- [100] Tapeinou, A., Matsoukas, M. T., Simal, C., and Tselios, T. (2015) Cyclic peptides on a merry-go-round; towards drug design, *Biopolymers* *104*, 453-461.
- [101] Ferrie, J. J., Gruskos, J. J., Goldwaser, A. L., Decker, M. E., and Guarracino, D. A. (2013) A comparative protease stability study of synthetic macrocyclic peptides that mimic two endocrine hormones, *Bioorg. Med. Chem. Lett.* *23*, 989-995.
- [102] Roxin, A., and Zheng, G. (2012) Flexible or fixed: a comparative review of linear and cyclic cancer-targeting peptides, *Future Med. Chem.* *4*, 1601-1618.
- [103] Joo, S. H. (2012) Cyclic Peptides as therapeutic agents and biochemical tools, *Biomol. Ther.* *20*, 19-26.
- [104] Oneil, K. T., Hoess, R. H., Jackson, S. A., Ramachandran, N. S., Mousa, S. A., and Degrado, W. F. (1992) Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library, *Proteins* *14*, 509-515.
- [105] Katz, B. A. (1995) Binding to protein targets of peptidic leads discovered by phage display: crystal-structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence, *Biochemistry* *34*, 15421-15429.
- [106] Deyle, K., Kong, X. D., and Heinis, C. (2017) Phage selection of cyclic peptides for application in research and drug development, *Acc. Chem. Res.* *50*, 1866-1874.

- [107] Heinis, C., Rutherford, T., Freund, S., and Winter, G. (2009) Phage-encoded combinatorial chemical libraries based on bicyclic peptides, *Nat. Chem. Biol.* 5, 502-507.
- [108] Chen, S., Bertoldo, D., Angelini, A., Pojer, F., and Heinis, C. (2014) Peptide ligands stabilized by small molecules, *Angew. Chem. Int. Edit.* 53, 1602-1606.
- [109] Jafari, M. R., Deng, L., Kitov, P. I., Ng, S., Matochko, W. L., Tjhung, K. F., Zeberoff, A., Elias, A., Klassen, J. S., and Derda, R. (2014) Discovery of light-responsive ligands through screening of a light-responsive genetically encoded library, *ACS Chem. Biol.* 9, 443-450.
- [110] Bellotto, S., Chen, S. Y., Rebollo, I. R., Wegner, H. A., and Heinis, C. (2014) Phage selection of photoswitchable peptide ligands, *J. Am. Chem. Soc.* 136, 5880-5883.
- [111] Jafari, M. R., Lakusta, J., Lundgren, R. J., and Derda, R. (2016) Allene functionalized azobenzene linker enables rapid and light-responsive peptide macrocyclization, *Bioconjugate Chem.* 27, 509-514.
- [112] Diderich, P., Bertoldo, D., Dessen, P., Khan, M. M., Pizzitola, I., Held, W., Huelsken, J., and Heinis, C. (2016) Phage selection of chemically stabilized alpha-helical peptide ligands, *ACS Chem. Biol.* 11, 1422-1427.
- [113] Ng, S., and Derda, R. (2016) Phage-displayed macrocyclic glycopeptide libraries, *Org. Biomol. Chem.* 14, 5539-5545.
- [114] Kather, I., Bippes, C. A., and Schmid, F. X. (2005) A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution, *J. Mol. Biol.* 354, 666-678.

- [115] Wang, L., Zhang, Z. W., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* *100*, 56-61.
- [116] Deiters, A., and Schultz, P. G. (2005) In vivo incorporation of an alkyne into proteins in *Escherichia coli*, *Bioorg. Med. Chem. Lett.* *15*, 1521-1524.
- [117] Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J. W., and Schultz, P. G. (2006) A genetically encoded fluorescent amino acid, *Proc. Natl. Acad. Sci. U.S.A.* *103*, 9785-9789.
- [118] Wang, J. Y., Xie, J. M., and Schultz, P. G. (2006) A genetically encoded fluorescent amino acid, *J. Am. Chem. Soc.* *128*, 8738-8739.
- [119] Lang, K., Davis, L., Torres-Kolbus, J., Chou, C. J., Deiters, A., and Chin, J. W. (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction, *Nat. Chem.* *4*, 298-304.
- [120] Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., Fox, J. M., and Chin, J. W. (2012) Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions, *J. Am. Chem. Soc.* *134*, 10317-10320.
- [121] Plass, T., Milles, S., Koehler, C., Szymanski, J., Mueller, R., Wiessler, M., Schultz, C., and Lemke, E. A. (2012) Amino acids for Diels-Alder reactions in living cells, *Angew. Chem. Int. Edit.* *51*, 4166-4170.

- [122] Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* 99, 11020-11024.
- [123] Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*, *J. Am. Chem. Soc.* 124, 9026-9027.
- [124] Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z. W., and Schultz, P. G. (2003) An expanded eukaryotic genetic code, *Science* 301, 964-967.
- [125] Krishnamurthy, M., Dugan, A., Nwokoye, A., Fung, Y. H., Lancia, J. K., Majmudar, C. Y., and Mapp, A. K. (2011) Caught in the act: covalent cross-linking captures activator-coactivator interactions in vivo, *ACS Chem. Biol.* 6, 1321-1326.
- [126] Zhang, M., Lin, S. X., Song, X. W., Liu, J., Fu, Y., Ge, X., Fu, X. M., Chang, Z. Y., and Chen, P. R. (2011) A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance, *Nat. Chem. Biol.* 7, 671-677.
- [127] Liu, W. S. R., Wang, Y. S., and Wan, W. (2011) Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach, *Mol. Biosyst.* 7, 38-47.
- [128] Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N-epsilon-acetyllysine in recombinant proteins, *Nat. Chem. Biol.* 4, 232-234.

- [129] Nguyen, D. P., Alai, M. M. G., Kapadnis, P. B., Neumann, H., and Chin, J. W. (2009) Genetically encoding N-epsilon-methyl-L-lysine in recombinant histones, *J. Am. Chem. Soc.* *131*, 14194-14195.
- [130] Virdee, S., Kapadnis, P. B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D. P., Riechmann, L., and Chin, J. W. (2011) Traceless and site-specific ubiquitination of recombinant proteins, *J. Am. Chem. Soc.* *133*, 10708-10711.
- [131] Wang, Y. S., Wu, B., Wang, Z. Y., Huang, Y., Wan, W., Russell, W. K., Pai, P. J., Moe, Y. N., Russell, D. H., and Liu, W. R. (2010) A genetically encoded photocaged N-epsilon-methyl-L-lysine, *Mol. Biosyst.* *6*, 1575-1578.
- [132] Schultz, K. C., Supekova, L., Ryu, Y. H., Xie, J. M., Perera, R., and Schultz, P. G. (2006) A genetically encoded infrared probe, *J. Am. Chem. Soc.* *128*, 13984-13985.
- [133] Xie, J. M., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P. G. (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination, *Nat. Biotechnol.* *22*, 1297-1301.
- [134] Wang, L., and Schultz, P. G. (2005) Expanding the genetic code, *Angew. Chem. Int. Ed.* *44*, 34-66.
- [135] Thibodeaux, G. N., Liang, X. A., Moncivais, K., Umeda, A., Singer, O., Alfonta, L., and Zhang, Z. W. J. (2010) Transforming a pair of orthogonal tRNA-aminoacyl-tRNA synthetase from archaea to function in mammalian cells, *Plos One* *5*.

- [136] Liu, W. S., Brock, A., Chen, S., Chen, S. B., and Schultz, P. G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells, *Nat. Methods* 4, 239-244.
- [137] Wu, N., Deiters, A., Cropp, T. A., King, D., and Schultz, P. G. (2004) A genetically encoded photocaged amino acid, *J. Am. Chem. Soc.* 126, 14306-14307.
- [138] Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Church-Church, K. B., Chan, M. K., and Krzycki, J. A. (2004) Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo, *Nature* 431, 333-335.
- [139] Polycarpo, C., Ambrogelly, A., Berube, A., Winbush, S. A. M., McCloskey, J. A., Crain, P. F., Wood, J. L., and Söll, D. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine, *Proc. Natl. Acad. Sci. U.S.A.* 101, 12450-12454.
- [140] Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine encoded by UAG in archaea: charging of a UAG-decoding specialized tRNA, *Science* 296, 1459-1462.
- [141] Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008) Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases, *Biochem. Biophys. Res. Commun.* 371, 818-822.



- [142] Hancock, S. M., Uprety, R., Deiters, A., and Chin, J. W. (2010) Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair, *J. Am. Chem. Soc.* *132*, 14819-14824.
- [143] Greiss, S., and Chin, J. W. (2011) Expanding the genetic code of an animal, *J. Am. Chem. Soc.* *133*, 14196-14199.
- [144] Parrish, A. R., She, X. Y., Xiang, Z., Coin, I., Shen, Z. X., Briggs, S. P., Dillin, A., and Wang, L. (2012) Expanding the genetic code of *Caenorhabditis elegans* using bacterial aminoacyl-tRNA synthetase/tRNA Pairs, *ACS Chem. Biol.* *7*, 1292-1302.
- [145] Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity, *Nat. Biotechnol.* *20*, 1044-1048.
- [146] Wang, Y. S., Fang, X. Q., Wallace, A. L., Wu, B., and Liu, W. S. R. (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum, *J. Am. Chem. Soc.* *134*, 2950-2953.
- [147] Wang, Y. S., Fang, X. Q., Chen, H. Y., Wu, B., Wang, Z. Y. U., Hilty, C., and Liu, W. S. R. (2013) Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant, *ACS Chem. Biol.* *8*, 405-415.
- [148] Wang, Y. S., Russell, W. K., Wang, Z. Y., Wan, W., Dodd, L. E., Pai, P. J., Russell, D. H., and Liu, W. S. R. (2011) The de novo engineering of pyrrolysyl-

- tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives, *Mol. Biosyst.* 7, 714-717.
- [149] Tuley, A., Wang, Y. S., Fang, X. Q., Kurra, Y., Rezenom, Y. H., and Liu, W. R. (2014) The genetic incorporation of thirteen novel non-canonical amino acids, *ChemComm* 50, 2673-2675.
- [150] Peters, F. B., Brock, A., Wang, J. Y., and Schultz, P. G. (2009) Photocleavage of the polypeptide backbone by 2-nitrophenylalanine, *Chem. Biol.* 16, 148-152.
- [151] Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code, *Annu. Rev. Biochem.* 79, 413-444.
- [152] Wang, Y. S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum, *J. Am. Chem. Soc.* 134, 2950-2953.
- [153] Wang, Y. S., Fang, X., Chen, H. Y., Wu, B., Wang, Z. U., Hilty, C., and Liu, W. R. (2013) Genetic incorporation of twelve *meta*-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant, *ACS Chem. Biol.* 8, 405-415.
- [154] Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C., and Mehl, R. A. (2007) Preparation of site-specifically labeled fluorinated proteins for <sup>19</sup>F-NMR structural characterization, *Nat. Protocols* 2, 2601-2607.
- [155] Takimoto, J. K., Dellas, N., Noel, J. P., and Wang, L. (2011) Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient *in vivo*

- incorporation of structurally divergent non-native amino acids, *ACS Chem. Biol.* *6*, 733-743.
- [156] Odoi, K. A., Huang, Y., Rezenom, Y. H., and Liu, W. R. (2013) Nonsense and sense suppression abilities of original and derivative *Methanosarcina mazei* pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the *Escherichia coli* BL21(DE3) cell strain, *Plos One* *8*.
- [157] Carter, R. E., and Sorkin, A. (1998) Endocytosis of functional epidermal growth factor receptor-green fluorescent protein chimera, *J. Biol. Chem.* *273*, 35000-35007.
- [158] Chalker, J. M., Wood, C. S. C., and Davis, B. G. (2009) A convenient catalyst for aqueous and protein Suzuki-Miyaura cross-coupling, *J. Am. Chem. Soc.* *131*, 16346-16347.
- [159] Tsao, M. L., Summerer, D., Ryu, Y. H., and Schultz, P. G. (2006) The genetic incorporation of a distance probe into proteins in *Escherichia coli*, *J. Am. Chem. Soc.* *128*, 4572-4573.
- [160] Smith, E. E., Linderman, B. Y., Luskin, A. C., and Brewer, S. H. (2011) Probing local environments with the infrared probe: L-4-nitrophenylalanine, *J. Phys. Chem. B* *115*, 2380-2385.
- [161] Aprilakis, K. N., Taskent, H., and Raleigh, D. P. (2007) Use of the novel fluorescent amino acid *p*-cyanophenylalanine offers a direct probe of hydrophobic core formation during the folding of the N-terminal domain of the

- ribosomal protein L9 and provides evidence for two-state folding, *Biochemistry* 46, 12308-12313.
- [162] Tucker, M. J., Oyola, R., and Gai, F. (2006) A novel fluorescent probe for protein binding and folding studies: *p*-cyano-phenylalanine, *Biopolymers* 83, 571-576.
- [163] Tucker, M. J., Oyola, R., and Gai, F. (2005) Conformational distribution of a 14-residue peptide in solution: a fluorescence resonance energy transfer study, *J. Phys. Chem. B* 109, 4788-4795.
- [164] Taskent-Sezgin, H., Chung, J., Patsalo, V., Miyake-Stoner, S. J., Miller, A. M., Brewer, S. H., Mehl, R. A., Green, D. F., Raleigh, D. P., and Carrico, I. (2009) Interpretation of *p*-cyanophenylalanine fluorescence in proteins in terms of solvent exposure and contribution of side-chain quenchers: a combined fluorescence, IR and molecular dynamics study, *Biochemistry* 48, 9040-9046.
- [165] Kokura, K., Sun, L. D., Bedford, M. T., and Fang, J. (2010) Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion, *EMBO J.* 29, 3673-3687.
- [166] Chang, Y. Q., Horton, J. R., Bedford, M. T., Zhang, X., and Cheng, X. D. (2011) Structural insights for MPP8 chromodomain interaction with histone H3 lysine 9: potential effect of phosphorylation on methyl-lysine binding, *J. Mol. Biol.* 408, 807-814.
- [167] Wu, X. D., Wu, D., Lu, Z. S., Chen, W. T., Hu, X. J., and Ding, Y. (2009) A novel method for high-level production of TEV protease by superfolder GFP tag, *J. Biomed. Biotechnol.*

- [168] Short, G. F., Golovine, S. Y., and Hecht, S. M. (1999) Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids, *Biochemistry* 38, 8808-8819.
- [169] Johnson, D. B. F., Xu, J. F., Shen, Z. X., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z., Ecker, J. R., Briggs, S. P., and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites, *Nat. Chem. Biol.* 7, 779-786.
- [170] Guo, L. T., Wang, Y. S., Nakamura, A., Eiler, D., Kavran, J. M., Wong, M., Kiessling, L. L., Steitz, T. A., O'Donoghue, P., and Söll, D. (2014) Polyspecific pyrrolysyl-tRNA synthetases from directed evolution, *Proc. Natl. Acad. Sci. U.S.A.* 111, 15724-16729.
- [171] Zade, H. M., Keshavarz, R., Shekarabi, H. S. Z., and Bakhshinejad, B. (2017) Biased selection of propagation-related TUPs from phage display peptide libraries, *Amino Acids* 49, 1293-1308.
- [172] Huang, C. C., Venturi, M., Majeed, S., Moore, M. J., Phogat, S., Zhang, M. Y., Dimitrov, D. S., Hendrickson, W. A., Robinson, J., Sodroski, J., Wyatt, R., Choe, H., Farzan, M., and Kwong, P. D. (2004) Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120, *Proc. Natl. Acad. Sci. U.S.A.* 101, 2706-2711.
- [173] Wang, N. X., Ju, T., Niu, W., and Guo, J. T. (2015) Fine-tuning Interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids, *ACS Synth. Biol.* 4, 207-212.

- [174] Fan, C. G., Xiong, H., Reynolds, N. M., and Söll, D. (2015) Rationally evolving tRNA(Pyl) for efficient incorporation of noncanonical amino acids, *Nucleic Acids Res.* 43.
- [175] Wang, K. H., Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion, *Nat. Biotechnol.* 25, 770-777.
- [176] Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. (2007) Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system, *J. Am. Chem. Soc.* 129, 14458-14462.
- [177] Lee, Y. J., Wu, B., Raymond, J. E., Zeng, Y., Fang, X. Q., Wooley, K. L., and Liu, W. R. S. (2013) A genetically encoded acrylamide functionality, *ACS Chem. Biol.* 8, 1664-1670.
- [178] Tropea, J. E., Cherry, S., and Waugh, D. S. (2009) Expression and purification of soluble His6-tagged TEV protease, In *High Throughput Protein Expression and Purification: Methods and Protocols* (Doyle, S. A., Ed.), pp 297-307, Humana Press, Totowa, NJ.
- [179] Wang, Z. A., Kurra, Y., Wang, X., Zeng, Y., Lee, Y. J., Sharma, V., Lin, H., Dai, S. Y., and Liu, W. R. (2017) A versatile approach for site-specific lysine acylation in proteins, *Angew. Chem. Int. Ed.* 56, 1643-1647.
- [180] Boeke, J. D., Model, P., and Zinder, N. D. (1982) Effects of bacteriophage f1 gene III protein on the host-cell membrane, *Mol. Gen. Genet.* 186, 185-192.

- [181] Nilsson, N., Malmberg, A. C., and Borrebaeck, C. A. K. (2000) The phage infection process: a functional role for the distal linker region of bacteriophage protein 3, *J. Virol.* *74*, 4229-4235.
- [182] Jestin, J. L., Volioti, G., and Winter, G. (2001) Improving the display of proteins on filamentous phage, *Res. Microbiol.* *152*, 187-191.
- [183] Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J., and Wilcox, G. (1987) Characterization of the *Erwinia carotovora* pelB gene and its product pectate lyase, *J. Bacteriol.* *169*, 4379-4383.
- [184] Tharp, J. M., Wang, Y. S., Lee, Y. J., Yang, Y. Y., and Liu, W. S. R. (2014) Genetic incorporation of seven *ortho*-substituted phenylalanine derivatives, *ACS Chem. Biol.* *9*, 884-890.
- [185] Nijkamp, H. J. J., Delang, R., Stuitje, A. R., Vandanelzen, P. J. M., Veltkamp, E., and Vanputten, A. J. (1986) The complete nucleotide sequence of the bacteriocinogenic plasmid Clodf13, *Plasmid* *16*, 135-160.
- [186] Lajoie, M. J., Tovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., and Isaacs, F. J. (2013) Genomically recoded organisms expand biological functions, *Science* *342*, 357-360.
- [187] Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J. M., Gu, W., and Zhao, Y. M. (2007) Lysine propionylation and

butyrylation are novel post-translational modifications in histones, *Mol. Cell. Proteomics* 6, 812-819.

- [188] Tan, M. J., Luo, H., Lee, S., Jin, F. L., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z. Y., Rousseaux, S., Rajagopal, N., Lu, Z. K., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., and Zhao, Y. M. (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification, *Cell* 146, 1015-1027.
- [189] Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. (1994) Release of proteins and peptides from fusion proteins using a recombinant plant-virus proteinase, *Anal. Biochem.* 216, 413-417.
- [190] Nunn, C. M., Jeeves, M., Cliff, M. J., Urquhart, G. T., George, R. R., Chao, L. H., Tsuchia, Y., and Djordjevic, S. (2005) Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site, *J. Mol. Biol.* 350, 145-155.
- [191] Bakhshinejad, B., Zade, H. M., Shekarabi, H. S. Z., and Neman, S. (2016) Phage display biopanning and isolation of target-unrelated peptides: in search of nonspecific binders hidden in a combinatorial library, *Amino Acids* 48, 2699-2716.
- [192] Giebel, L. B., Cass, R. T., Milligan, D. L., Young, D. C., Arze, R., and Johnson, C. R. (1995) Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities, *Biochemistry* 34, 15430-15435.



- [193] Finkel, T., Deng, C. X., and Mostoslavsky, R. (2009) Recent progress in the biology and physiology of sirtuins, *Nature* 460, 587-591.
- [194] Borra, M. T., Langer, M. R., Slama, J. T., and Denu, J. M. (2004) Substrate specificity and kinetic mechanism of the Sir2 family of NAD(+)-dependent histone/protein deacetylases, *Biochemistry* 43, 9877-9887.
- [195] Smith, B. C., Hallows, W. C., and Denu, J. M. (2008) Mechanisms and molecular probes of sirtuins, *Chem. Biol.* 15, 1002-1013.
- [196] North, B. J., Marshall, B. L., Borra, M. T., Denu, J. M., and Verdin, E. (2003) The human Sir2 ortholog, SIRT2, is an NAD(+)-dependent tubulin deacetylase, *Mol. Cell* 11, 437-444.
- [197] Vaquero, A., Scher, M. B., Lee, D. H., Sutton, A., Cheng, H. L., Alt, F. W., Serrano, L., Sternglanz, R., and Reinberg, D. (2006) SirT2 is a histone deacetylase with preference for histone H4 Lys 16 during mitosis, *Gene. Dev.* 20, 1256-1261.
- [198] Hsu, W. W., Wu, B., and Liu, W. S. R. (2016) Sirtuins 1 and 2 are universal histone deacetylases, *ACS Chem. Biol.* 11, 792-799.
- [199] Hoffmann, G., Breitenbucher, F., Schuler, M., and Ehrenhofer-Murray, A. E. (2014) A novel sirtuin 2 (SIRT2) inhibitor with p53-dependent pro-apoptotic activity in non-small cell lung cancer, *J. Biol. Chem.* 289, 5208-5216.
- [200] Peck, B., Chen, C. Y., Ho, K. K., Di Fruscia, P., Myatt, S. S., Coombes, R. C., Fuchter, M. J., Hsiao, C. D., and Lam, E. W. F. (2010) SIRT inhibitors induce

- cell death and p53 acetylation through targeting both SIRT1 and SIRT2, *Mol. Cancer Ther.* 9, 844-855.
- [201] Jing, E. X., Gesta, S., and Kahn, C. R. (2007) SIRT2 regulates adipocyte differentiation through FoxO1 acetylation/deacetylation, *Cell Metab.* 6, 105-114.
- [202] Feldman, J. L., Baeza, J., and Denu, J. M. (2013) Activation of the protein deacetylase SIRT6 by long-chain fatty acids and widespread deacylation by mammalian sirtuins, *J. Biol. Chem.* 288, 31350-31356.
- [203] Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., and Mann, M. (2014) The growing landscape of lysine acetylation links metabolism and cell signalling, *Nat. Rev. Mol. Cell Bio.* 15, 536-550.
- [204] Drazic, A., Myklebust, L. M., Ree, R., and Arnesen, T. (2016) The world of protein acetylation, *BBA-Proteins Proteom.* 1864, 1372-1401.
- [205] Houtkooper, R. H., Pirinen, E., and Auwerx, J. (2012) Sirtuins as regulators of metabolism and healthspan, *Nat. Rev. Mol. Cell Bio.* 13, 225-238.
- [206] Mei, Z., Zhang, X., Yi, J. R., Huang, J. J., He, J., and Tao, Y. G. (2016) Sirtuins in metabolism, DNA repair and cancer, *J. Exp. Clin. Canc. Res.* 35, 182.
- [207] Kim, H. S., Vassilopoulos, A., Wang, R. H., Lahusen, T., Xiao, Z., Xu, X. L., Li, C. L., Veenstra, T. D., Li, B., Yu, H. T., Ji, J. F., Wang, X. W., Park, S. H., Cha, Y. I., Gius, D., and Deng, C. X. (2011) SIRT2 maintains genome integrity and suppresses tumorigenesis through regulating APC/C activity, *Cancer Cell* 20, 487-499.

- [208] Chalkiadaki, A., and Guarente, L. (2015) The multifaceted functions of sirtuins in cancer, *Nat. Rev. Cancer* 15, 608-624.
- [209] Chen, J., Chan, A. W. H., To, K. F., Chen, W. X., Zhang, Z. Z., Ren, J. H., Song, C. L., Cheung, Y. S., Lai, P. B. S., Cheng, S. H., Ng, M. H. L., Huang, A. L., and Ko, B. C. B. (2013) SIRT2 overexpression in hepatocellular carcinoma mediates epithelial to mesenchymal transition by protein kinase B/glycogen synthase kinase-3/-catenin signaling, *Hepatology* 57, 2287-2298.
- [210] Xie, H. J., Jung, K. H., and Nam, S. W. (2011) Overexpression of SIRT2 contributes tumor cell growth in hepatocellular carcinomas, *Mol. Cell Toxicol.* 7, 367-374.
- [211] Luo, J., Bao, Y. C., Ji, X. X., Chen, B., Deng, Q. F., and Zhou, S. W. (2017) SPOP promotes SIRT2 degradation and suppresses non-small cell lung cancer cell growth, *Biochem. Bioph. Res. Co.* 483, 880-884.
- [212] McGlynn, L. M., Zino, S., MacDonald, A. I., Curle, J., Reilly, J. E., Mohammed, Z. M. A., McMillan, D. C., Mallon, E., Payne, A. P., Edwards, J., and Shiels, P. G. (2014) SIRT2: tumour suppressor or tumour promoter in operable breast cancer?, *Eur. J. Cancer* 50, 290-301.
- [213] Cheon, M. G., and Kim, J. E. (2015) AK-1, a specific SIRT2 inhibitor, induces cell cycle arrest by downregulating Snail in HCT116 human colon carcinoma cells, *Cancer Res.* 75, 637-645.
- [214] Jing, H., Hu, J., He, B., Abril, Y. L. N., Stupinski, J., Weiser, K., Carbonaro, M., Chiang, Y. L., Southard, T., Giannakakou, P., Weiss, R. S., and Lin, H. N. (2016)

- A SIRT2-selective inhibitor promotes c-Myc oncoprotein degradation and exhibits broad anticancer activity, *Cancer Cell* 29, 297-310.
- [215] Outeiro, T. F., Kontopoulos, E., Altmann, S. M., Kufareva, I., Strathearn, K. E., Amore, A. M., Volk, C. B., Maxwell, M. M., Rochet, J. C., McLean, P. J., Young, A. B., Abagyan, R., Feany, M. B., Hyman, B. T., and Kazantsev, A. G. (2007) Sirtuin 2 inhibitors rescue alpha-synuclein-mediated toxicity in models of Parkinson's disease, *Science* 317, 516-519.
- [216] de Oliveira, R. M., Miranda, H. V., Francelle, L., Pinho, R., Szego, E. M., Martinho, R., Munari, F., Lazaro, D. F., Moniot, S., Guerreiro, P., Fonseca, L., Marijanovic, Z., Antas, P., Gerhardt, E., Enguita, F. J., Fauvet, B., Penque, D., Pais, T. F., Tong, Q., Becker, S., Kugler, S., Lashuel, H. A., Steegborn, C., Zweckstetter, M., and Outeiro, T. F. (2017) The mechanism of sirtuin 2-mediated exacerbation of alpha-synuclein toxicity in models of Parkinson disease, *Plos Biol* 15.
- [217] Taylor, D., Pallos, J., Lambert, E., Amore, A., Parker, A., Moffitt, H., Smith, D., Runne, H., Gokce, O., Kuhn, A., Xiang, Z., Maxwell, M., Reeves, S., Bates, G., Neri, C., Thompson, L., Marsh, L., Kazantsev, A., and Luthi-Carter, R. (2010) Sirt2 inhibition achieves neuroprotection by decreasing sterol biosynthesis, *J. Neurol. Neurosurg. Psychiatry* 81, A3-A4.
- [218] Fatkins, D. G., Monnot, A. D., and Zheng, W. P. (2006) N-epsilon-thioacetyl-lysine: a multi-facet functional probe for enzymatic protein lysine N-epsilon-deacetylation, *Bioorg. Med. Chem. Lett.* 16, 3651-3656.

- [219] Smith, B. C., and Denu, J. M. (2007) Mechanism-based inhibition of Sir2 deacetylases by thioacetyl-lysine peptide, *Biochemistry* 46, 14478-14486.
- [220] Kiviranta, P. H., Suuronen, T., Wallen, E. A. A., Leppanen, J., Tervonen, J., Kyrylenko, S., Salminen, A., Poso, A., and Jarho, E. M. (2009) N-epsilon-thioacetyl-lysine-containing tri-, tetra-, and pentapeptides as SIRT1 and SIRT2 inhibitors, *J. Med. Chem.* 52, 2153-2156.
- [221] Mellini, P., Kokkola, T., Suuronen, T., Salo, H. S., Tolvanen, L., Mai, A., Lahtela-Kakkonen, M., and Jarho, E. M. (2013) Screen of pseudopeptidic inhibitors of human sirtuins 1-3: two lead compounds with antiproliferative effects in cancer cells, *J. Med. Chem.* 56, 6681-6695.
- [222] Jing, H., Hu, J., He, B., Abril, Y. L. N., Stupinski, J., Weiser, K., Carbonaro, M., Chiang, Y. L., Southard, T., Giannakakou, P., Weiss, R. S., and Lin, H. N. (2016) A SIRT2-selective inhibitor promotes c-Myc oncoprotein degradation and exhibits broad anticancer activity, *Cancer Cell* 29, 767-768.
- [223] Schiedel, M., Robaa, D., Rumpf, T., Sippl, W., and Jung, M. (2018) The current state of NAD<sup>+</sup>-dependent histone deacetylases (sirtuins) as novel therapeutic targets, *Med. Res. Rev.* 38, 147-200.
- [224] Bheda, P., Jing, H., Wolberger, C., and Lin, H. N. (2016) The substrate specificity of sirtuins, *Annu. Rev. Biochem.* 85, 405-429.
- [225] Heinis, C. (2013) Phage selection of bicyclic peptides, *Biopolymers* 100, 230-230.

- [226] Schuster, S., Roessler, C., Meleshin, M., Zimmermann, P., Simic, Z., Kambach, C., Schiene-Fischer, C., Steegborn, C., Hottiger, M. O., and Schutkowski, M. (2016) A continuous sirtuin activity assay without any coupling to enzymatic or chemical reactions, *Sci. Rep.* 6.
- [227] Xiong, H., Reynolds, N. M., Fan, C. G., Englert, M., Hoyer, D., Miller, S. J., and Söll, D. (2016) Dual genetic encoding of acetyl-lysine and non-deacetylable thioacetyl-lysine mediated by flexizyme, *Angew. Chem. Int. Edit.* 55, 4083-4086.
- [228] Venkat, S., Nannapaneni, D. T., Gregory, C., Gan, Q. L., McIntosh, M., and Fan, C. G. (2017) Genetically encoding thioacetyl-lysine as a non-deacetylable analog of lysine acetylation in *Escherichia coli*, *FEBS Open Bio.* 7, 1805-1814.
- [229] Kavran, J. M., Gundliapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation, *Proc. Natl. Acad. Sci. U.S.A.* 104, 11268-11273.
- [230] Smith, B. C., and Denu, J. M. (2007) Sir2 deacetylases exhibit nucleophilic participation of acetyl-lysine in NAD(+) cleavage, *J. Am. Chem. Soc.* 129, 5802-5803.
- [231] Morimoto, J., Hayashi, Y., and Suga, H. (2012) Discovery of macrocyclic peptides armed with a mechanism-based warhead: isoform-selective inhibition of human deacetylase SIRT2, *Angew. Chem. Int. Edit.* 51, 3423-3427.
- [232] Rauh, D., Fischer, F., Gertz, M., Lakshminarasimhan, M., Bergbrede, T., Aladini, F., Kambach, C., Becker, C. F. W., Zerweck, J., Schutkowski, M., and

- Steegborn, C. (2013) An acetylome peptide microarray reveals specificities and deacetylation substrates for all human sirtuin isoforms, *Nat. Commun.* 4, 2327.
- [233] Yamagata, K., Goto, Y., Nishimasu, H., Morimoto, J., Ishitani, R., Dohmae, N., Takeda, N., Nagai, R., Komuro, I., Suga, H., and Nureki, O. (2014) Structural basis for potent inhibition of SIRT2 deacetylase by a macrocyclic peptide inducing dynamic structural change, *Structure* 22, 345-352.
- [234] Cosgrove, M. S., Bever, K., Avalos, J. L., Muhammad, S., Zhang, X. B., and Wolberger, C. (2006) The structural basis of sirtuin substrate affinity, *Biochemistry* 45, 7511-7521.
- [235] Feldman, J. L., Dittenhafer-Reed, K. E., and Denu, J. M. (2012) Sirtuin catalysis and regulation, *J. Biol. Chem.* 287, 42419-42427.
- [236] Blander, G., Olejnik, J., Olejnik, E. K., Mcdonagh, T., Haigis, M., Yaffe, M. B., and Guarente, L. (2005) SIRT1 shows no substrate specificity in vitro, *J. Biol. Chem.* 280, 9780-9785.
- [237] Knyphausen, P., de Boor, S., Kuhlmann, N., Scislawski, L., Extra, A., Baldus, L., Schacherl, M., Baumann, U., Neundorff, I., and Lammers, M. (2016) Insights into lysine deacetylation of natively folded substrate proteins by sirtuins, *J. Biol. Chem.* 291, 14677-14694.
- [238] Simic, Z., Weiwad, M., Schierhorn, A., Steegborn, C., and Schutkowski, M. (2015) The amino group of protein lysine residues is highly susceptible to nonenzymatic acylation by several physiological acyl-CoA thioesters, *ChemBioChem* 16, 2337-2347.

- [239] Wagner, G. R., and Hirschey, M. D. (2014) Nonenzymatic protein acylation as a carbon stress regulated by sirtuin deacylases, *Mol. Cell* 54, 5-16.
- [240] Ren, H., Xiao, F., Zhan, K., Kim, Y. P., Xie, H., Xia, Z., and Rao, J. (2009) A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins *Angew. Chem. Int. Ed.* 48, 9658-9662.
- [241] Yuan, Y., and Liang, G. (2014) A biocompatible, highly efficient click reaction and its applications, *Org. Biomol. Chem.* 12, 865-871.
- [242] Lindley, H. (1962) The reaction of thiol compounds and chloroacetamide. 2. The reaction between chloroacetamide and cysteine peptides, *Biochem. J.* 82, 418-425.
- [243] Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat. Biotechnol.* 24, 79-88.
- [244] Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P., and Winter, G. (1991) Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains, *Nucleic Acids Res.* 19, 4133-4137.
- [245] Steiner, D., Forrer, P., Stumpp, M. T., and Pluckthun, A. (2006) Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display, *Nat. Biotechnol.* 24, 823-831.



- [246] Hauser, P. S., and Ryan, R. O. (2007) Expressed protein ligation using an N-terminal cysteine containing fragment generated in vivo from a pelB fusion protein, *Protein Expres. Purif.* 54, 227-233.
- [247] Al-Homsi, L., Al-Okla, S., and Abbady, A. Q. (2015) Preparation of specific polyclonal antibody against the recombinant mutacin produced by sfGFP fusion protein technology, *Open Microbiol. J.* 9, 70-80.
- [248] Sharma, V., Wang, Y. S., and Liu, W. S. R. (2016) Probing the catalytic charge-relay system in alanine racemase with genetically encoded histidine mimetics, *ACS Chem. Biol.* 11, 3305-3309.

## APPENDIX

### LIST OF ABBREVIATIONS

aaRS	aminoacyl-tRNA synthetase
Bpy-Ala	bipyridyl alanine
BSA	bovine serum albumin
BuK	<i>N</i> <sub>ε</sub> -butyryl-lysine
cfu or c.f.u.	colony forming unit
CV	column volume
DMEM	Dulbecco's modified eagle medium
EGFR	epidermal growth factor receptor
EPR	electron paramagnetic resonance
ESI-MS	electrospray ionization mass spectrometry
FITC	fluorescein isothiocyanate
FRET	Förster resonance energy transfer
GST	glutathione <i>S</i> -transferase
IC <sub>50</sub>	half maximal inhibitory concentration
IPTG	isopropyl β-D-thiogalactoside
M13KO7(pIII)	A pIII phenotypic knockout of the M13KO7 helper phage containing a K10(TAA) mutation
MALDI-TOF MS	matrix-assisted laser desorption/ionization mass spectrometry
<i>Mj</i> -TyrRS	the tyrosyl-tRNA synthetase of <i>Methanocaldococcus jannaschii</i>
MPP8	M-phase phosphoprotein 8

MWCO	molecular weight cut-off
ncAA	non-canonical amino acid
ncFG	non-canonical functional group
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PBST	PBS supplemented with Tween 20
<i>p</i> CNF	<i>para</i> -cyano-phenylalanine
pfu or p.f.u.	plaque forming unit
Phage	bacteriophage
PhP	phagemid particle
PK	plasma kallikrein
PLP	pyridoxal 5'-phosphate
PMSF	phenylmethylsulfonyl fluoride
PR-bias	propagation-related bias
<i>p</i> -STyr	<i>para</i> -sulfo-tyrosine
PTM	posttranslational modification
pTyr	phospho-tyrosine
PylRS	pyrrolysyl-tRNA Synthetase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sec	selenocysteine
SECIS	selenocysteine incorporation sequence
sfGFP	superfolder green fluorescent protein

SH2	src-homology 2
SIRT2	sirtuin 2
ssDNA	single-stranded DNA
TCEP	tris(2-carboxyethyl)phosphine
TEV	tobacco etch virus
TFA	trifluoroacetic acid
thAcK	<i>N</i> <sub>ε</sub> -thioacetyl-lysine
thBuK	<i>N</i> <sub>ε</sub> -thiobutyryl-lysine
TUP	target-unrelated peptide
Ub	ubiquitin
wt	wildtype