2019

Development Of Thioamides As Protein Probes

Daniel Miklos Szantai-Kis
University of Pennsylvania, miklos.szk@gmail.com

Development Of Thioamides As Protein Probes

Abstract

Thioamides have been used for various applications with small molecules and peptides, including as protease sensors, fluorescence quenching probes, folding probes, and handles for site-specific chemical modification. However, their use in proteins has been limited due to cumbersome incorporation through semi-synthesis, as well as their unknown effects on protein structure. In this work, I address these issues from multiple perspectives. I present improvements for the synthesis of thioamide-containing peptides, which are needed for semi-synthesis. I also investigate the effects that thioamide substitutions have on tertiary structure in a model protein and show that based on thermal denaturation experiments, the position of thioamide substitution can have a significant impact on the tertiary fold. Furthermore, I demonstrate a new semi-synthetic strategy to obtain mg quantities of a thioamide-containing protein that will enable structural investigation by Nuclear Magnetic Resonance spectroscopy and X-Ray crystallography. Finally, I describe my attempts at the genetic incorporation of thioamides using a combination of unnatural amino acid mutagenesis and amber suppression. While the last part was not successful, the efforts to improve semi-synthesis of thioamides will enable the use of thioamides as probes in proteins in a similar fashion as they have been used previously in peptides.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Biochemistry & Molecular Biophysics

First Advisor

Ernest J. Petersson

Keywords

Native Chemical Ligation, Protein Stability, Thioamide

Subject Categories

Biochemistry | Cell Biology | Organic Chemistry

DEVELOPMENT OF THIOAMIDES AS PROTEIN PROBES

Daniel Miklos Szantai-Kis

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

_____

E. James Petersson, PhD
Associate Professor of Chemistry, and of Biochemistry and Molecular Biophysics

Graduate Group Chairperson

_____

Kim A. Sharp, PhD
Associate Professor of Biochemistry and Molecular Biophysics

Dissertation Committee*:*

Yale E. Goldman, MD, PhD (chair), Professor of Physiology

Carol J. Deutsch, PhD, Professor of Physiology

David M. Chenoweth, PhD, Associate Professor of Chemistry

Kathleen P. Howard, PhD, Professor of Chemistry, Swarthmore College

DEVELOPMENT OF THIOAMIDES AS PROTEIN PROBES

Dedication

Dedicated to my family, old and new

ACKNOWLEDGMENT

First and foremost, I'd like to thank my PhD advisor, Prof. E. James Petersson. He helped me to become the well-rounded scientist that I am today. I am very grateful for that he believed in me, from allowing me to join the lab, to work on many exciting projects, and to travel the US and the world for conferences. I would also like to thank my thesis committee members Prof. Yale Goldman, Prof. Carol Deutsch, Prof. David Chenoweth for their kind support and guidance. I'd also like to acknowledge my external thesis reviewer, Prof. Kathleen Howard, for taking the time to read my thesis and to come to Philadelphia for my defense. I'd also like to thank my previous scientific advisors, Prof. Ralph Mazitschek and Prof. Katja Schmitz, without their motivation and support I wouldn't have made it this far.

Next I would like to acknowledge all of my fellow lab members, a group of talented young scientists that I had the pleasure of working throughout the years: Lee Speight, Solongo Ziraldo, Jerri Wang, Stella Chen, Chris Walters, Rebecca Wissner, Jack Ferrie, Ohm Sungwienwong, Christina Cleveland, Vicky Jun, Buyan Pan, Taylor Barrett, Kristen Fiore, Chloe Jones, Marshall Lougee, Sam Giannakoulias, Hoang-Anh Phan, Keith Keenan, Anand Muthusamy, Lily Owei, Eileen Hoang, Tiberiu Mihaila, Sumant Shringari, Jieliang Wang, Joo Hyung Park, Dr. Yun Huang, Dr. Naoya Ieda, and Dr. Conor Haney. We all shared our moments of happiness when things worked and moments of sadness when things didn't work. But no matter how the science went, it was always a very helpful and collaborative atmosphere. I also want to acknowledge my amazing cohort of the incoming BMB class of 2013: Abigail Cember, Korrie Mack, Jane Schulte, Matthew Thompson, Sarah Welch, Dr. Enrique Lin Shao, Dr. Jamie DeNizio and Dr. Mara Olenick. Together, we had a lot of fun exploring our new city together and getting through classes.

Through my life in two departments, I made a lot of friends at Penn in general. There are too many to mention, but I want to generally thank everyone who had a positive impact on my life here at Penn. Additionally, I would also like to acknowledge Prof. Phil Rae and Ruth Elliott for giving me the opportunity to teach and for showing me what it means to be a great educator. I learned a lot from both of you. I also would like to acknowledge my Co-TAs, Kelly Karch and Ananth Srinivasan who made teaching this huge class a really enjoyable experience.

Finally, I would like to acknowledge my families. First, I'd like to thank my parents, whom always supported and nurtured my curiosity for biochemistry. They supported me whenever and however they could and I would not be where I am today if it weren't for their never-ending love and support. I'd also like to acknowledge my mother-in-law, Shirley Robson, for not only putting up being the 'third wheel', but for taking me in as if were her son too and thereby helping making Philadelphia feel like home. Last, but certainly not least, I would like to thank my wife and soulmate Victoria. Her endless love and support helped me to get through the challenging parts of my PhD. Without her, I would have never experienced Philadelphia the way I did and thanks to her, this city will always hold a special place in my heart.

ABSTRACT

DEVELOPMENT OF THIOAMIDES AS PROTEIN PROBES

Daniel Miklos Szantai-Kis

E. James Petersson

Thioamides have been used for various applications with small molecules and peptides, including as protease sensors, fluorescence quenching probes, folding probes, and handles for site-specific chemical modification. However, their use in proteins has been limited due to cumbersome incorporation through semi-synthesis, as well as their unknown effects on protein structure. In this work, I address these issues from multiple perspectives. I present improvements for the synthesis of thioamide-containing peptides, which are needed for semi-synthesis. I also investigate the effects that thioamide substitutions have on tertiary structure in a model protein and show that based on thermal denaturation experiments, the position of thioamide substitution can have a significant impact on the tertiary fold. Furthermore, I demonstrate a new semi-synthetic strategy to obtain mg quantities of a thioamide-containing protein that will enable structural investigation by Nuclear Magnetic Resonance spectroscopy and X-Ray crystallography. Finally, I describe my attempts at the genetic incorporation of thioamides using a combination of unnatural amino acid mutagenesis and amber suppression. While the last part was not successful, the efforts to improve semi-synthesis of thioamides will enable the use of thioamides as probes in proteins in a similar fashion as they have been used previously in peptides.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1 : Introduction

Reprinted and adapted with permission from:

Wang, Y. J.; Szantai-Kis, D. M.; Petersson, E. J., Semi-synthesis of thioamide containing proteins. *Organic & Biomolecular Chemistry* **2015**, *13*, 5074-5081. Copyright 2015 The Royal Society of Chemistry[1]

and

Mahanta, N.*; Szantai-Kis, D. M.*; Petersson, E. J.; Mitchell, D. A.; Biosynthesis and Chemical Applications of Thioamides. *ACS Chemical Biology* **2019**, *14*, 142-163. Copyright 2019 The American Chemical Society[2]

*denotes equal contribution

Thioamide Properties

A thioamide is a single-atom deviation from the canonical amide functional group (hereafter also referred to as “oxoamide”), where the carbonyl oxygen is replaced with sulfur. In the literature the thioamide functional group has historically been referred to as a thionoamide, thioxoamide, endothiopeptide, mercaptoamide, thiopeptide unit, thiodepsipeptide, and thiocarboxamide. For the sake of brevity and to be consistent with current Organic Chemistry usage patterns, we prefer thioamide. Formally, according to IUPAC recommendations, thioamide backbone modifications in peptides and proteins should be annotated systematically as ψ[CS-NH]. For brevity, we have previously used a superscript S (^{S}) to indicate the presence of a thioamide between two residues. Additionally, it should be noted that we use the term thiopeptides to describe thioamide-containing peptides while this term is often also used to describe thiazole- and thiazolene-containing natural products such as thiostrepton[3].

The potential benefits of incorporation of thioamide into both natural and synthetic compounds result from the at once subtle and dramatic changes to amide interactions that can come from this single atom substitution. The physicochemical properties of thioamides are summarized in Table 1.1. Thioamides are more reactive with both nucleophiles and electrophiles than amides[4,5], with a weaker carbonyl bond (130 vs 170 kcal/ mol)[6] and therefore have been used as chemical synthesis intermediates. Thioamides also have greater affinity for certain metals compared to those of amides. For instance, the natural product methanobactin (Figure 1-1) exhibits extremely high affinity copper binding[7], and selective, silver and mercury-catalyzed transformations have been used to convert thioamides into other carbonyl derivatives[8,9]. The differences in amide and thioamide geometry govern many of the noncovalent interactions exhibited by thioamide-containing

peptides. The thioamide C=S bond (1.71 Å) is one-third longer than the amide C=O bond (1.23 Å)[10], and sulfur has a van der Waals radius (1.85 Å) one-third larger than that of oxygen (1.40 Å)[11]. Peptide conformational changes can result from the elongated C=S bond and the higher rotational barrier for the C−N bond (~5 kcal/mol)[12], which reduces conformational flexibility. Additional altered physiochemical properties include (i) thioamide N−H groups being more acidic ($\Delta pK_a = -6$) than the corresponding amide[13], (ii) thioamide N−H groups being better hydrogen bond donors[14], and (iii) the sulfur lone pairs of thioamides being weaker hydrogen bond acceptors relative to oxygen lone pairs in amides[15]. Therefore, thioamides are suitable for evaluating the contribution of single hydrogen bonds to protein folding and/or stability. Substitution of an amide with a thioamide also imparts significant spectroscopic and electrochemical changes. The thioamide C=S bond has an ultraviolet (UV) absorption maximum at 265 ± 5 nm and an infrared (IR) stretch at 1120 ± 20 cm^{-1} , compared to values of 220 ± 5 nm and 1660 ± 20 cm^{-1} , respectively, for the amide C=O bond[16]. Moreover, the 13C nuclear magnetic resonance (NMR) chemical shift of the thioamide carbonyl is found 30 ppm downfield (200−210 ppm) of the corresponding amide resonance[17]. The oxidation potential of a model thioamide (1.21 eV) is significantly lower than that of the amide (3.29 eV)[18], which has prompted speculation about a potential role in electron transfer in biological settings.

Table 1.1: Summary of the Physicochemical Properties of Thioamides

Property	Amide	Thioamide
van der Waals radius (Å)	1.40	1.85[11]
C=X BDE[a] (kcal/mol)	170	130[6]
C=X length (Å)	1.23	1.71[10]
C–N rotational barrier (kcal/mol)	17	22[12]
Electronegativity of the heteroatom	3.44	2.58
C=X ···H–N hydrogen bond (kcal/mol)	6.1	4.8[14]
N–H pK_a	17	12[13]
$\pi \rightarrow \pi^*$ absorption (nm)	200	270[16]
E_{Ox}[b] (V vs SHE)	3.29	1.21[18]

[a]BDE is the bond dissociation energy. [b]E_{Ox} is the oxidation potential vs a standard hydrogen electrode.

The origin of the higher C−N rotational barrier in thioamides has been examined by NMR[12] and ab initio calculations[19]. It was found that the amino group of thioformamide is more conformationally rigid than in formamide. Additionally, the change in charge density at sulfur upon rotation of the amino group in thioformamide is greater than that at oxygen in formamide due to a predominant bipolar amide resonance form. The small difference in electronegativity between carbon and sulfur (Pauling electronegativities for C of 2.55 and for S of 2.58) and the larger size of sulfur are the predominant factors that allow charge transfer from nitrogen to sulfur in thioamides. A special feature of the thioamide bond arises from the red-shifted absorption bands and the higher barrier for cis−trans isomerization. This property allows access to either the cis or trans isomer by irradiation with the appropriate UV wavelength. The photoinduced isomerization is efficient (30%) and fast (<600 ps), whereas thermal relaxation is comparatively slow (>10 min)[16].

Hence, thioamide-containing peptides are good candidates for fast photoswitches in proteins, either to regulate activity or to initiate conformational transitions for time-resolved studies[20]. The properties mentioned above, allow thioamide modification to affect biomolecule function in valuable ways that have led to the evolution of several distinct mechanisms for thioamidation in natural products.

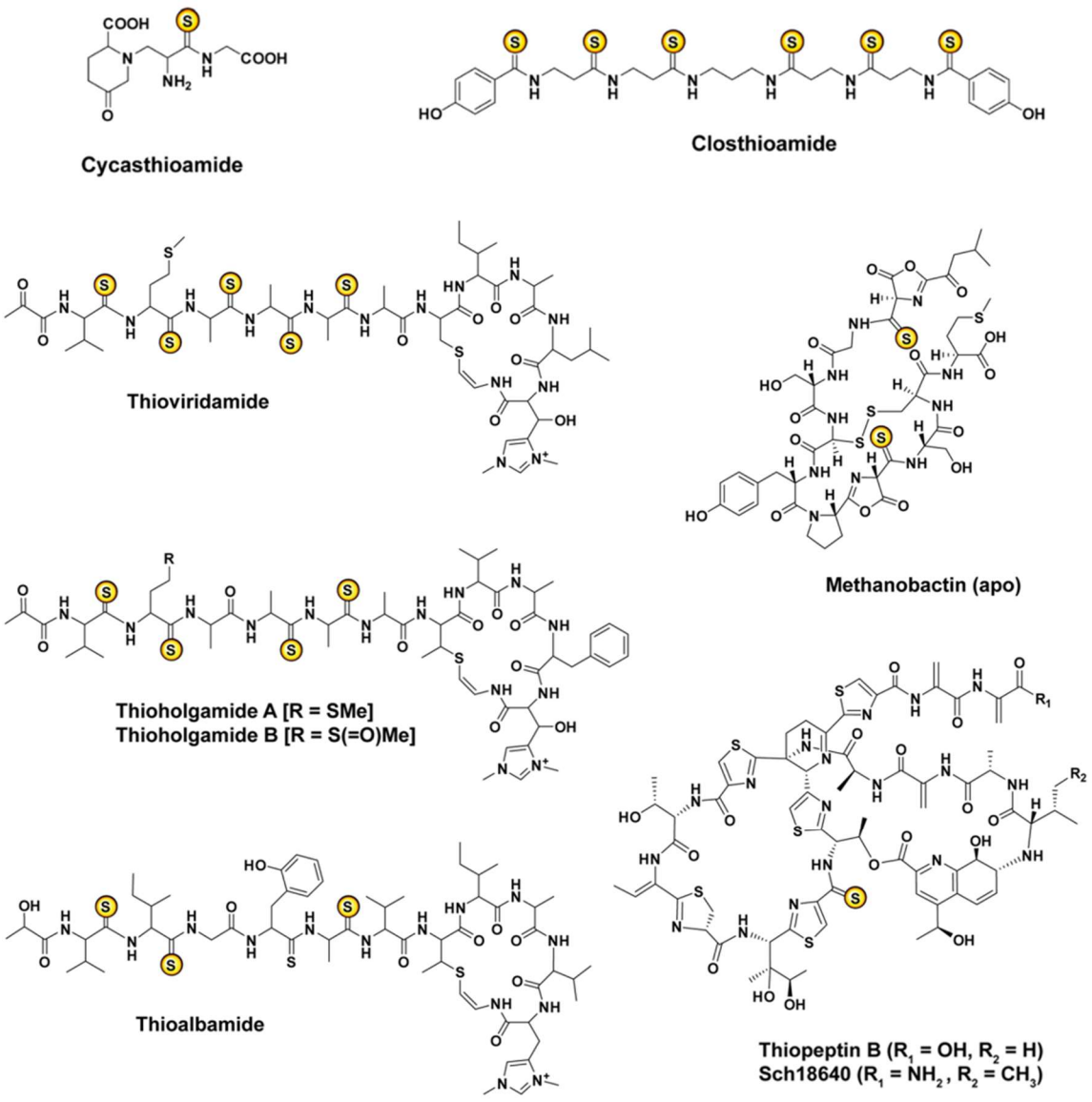


Figure 1-1: Structures of various thioamidated natural products. The stereo-chemistries in Thioviridamide, Thiohologamide and Thioalbamide have not yet been elucidated.

Thioamides in Nature: Biosynthesis, Structure, and Function

Thioamides are exceptionally rare in biology. Most reported natural thioamides are of bacterial origin, except for the plant-derived cycasthioamide (Figure 1-1). Among these bacterial thioamide products are the ribosomally synthesized and posttranslationally modified peptide (RiPP) natural products thioviridamide and its analogues, thiohologamide and thioalbamide, methanobactin, and thioamidated thiopeptides as well as closthioamide, which is a nonribosomal natural product (Figure 1-1)[3,21–23]. Known thioamide-containing nucleotides include thiouridine, thiocytidine, and thioguanine, which will not be discussed in the context of this thesis. Recent identification of the proteins responsible for thioglycine formation in the active site of methyl-coenzyme M reductase (MCR), the only known thioamidated protein, adds impetus to this area[24,25]. On the other hand, a distinct biosynthetic mechanism has been identified for closthioamide[26], which shares features of thionated nucleoside biosynthesis[27], as well as for methanobactin[28]. Collectively, these revelations suggest that natural thioamidation pathways arose through multiple different independent routes[24–26]. In the context of this thesis I will highlight two distinct pathways that lead to thioamidated natural products: The biosynthesis of Closthiamide and the YcaO-TfuA-dependent thioamidation as a posttranslational modification.

Closthioamide: Closthioamide, a polythioamide compound, produced by *Ruminiclostridium cellulolyticum* (previously *Clostridium cellulolyticum*), a strictly anaerobic, Gram-positive, soil-dwelling bacterium (Figure 1-1)[21]. Until recently, no R. cellulolyticum natural products had been isolated via standard laboratory cultivation. Hertweck and co-workers mimicked the natural environment by adding soil extracts to the culture, which led to the production of closthioamide[21]. Closthioamide is a symmetrical

hexathioamide and displays a central diaminopropyl group with four β-alanyl extender units and two terminal p-hydroxybenzoyl groups. Closthioamide is growth-suppressive toward several human pathogens, including Staphylococcus aureus, Enterococcus faecalis, and Neisseria gonorrheae[21,29]. Replacement of the thioamides with amides abolished the antibacterial activity of closthioamide, indicating the importance of thioamide moieties in its biological activity[21].

Studies using whole cell-based assays showed that closthioamide inhibits the ATPase activity of DNA gyrase and topoisomerase IV, thus effectively blocking DNA replication[30]. Closthioamide also inhibits the relaxation activity of DNA gyrase, which does not require ATP hydrolysis and thus may allosterically, rather than directly, interfere with the ATPase activity of gyrase. Notably, this mode of action differs from that of the other DNA gyrase inhibitors such the fluoroquinolones and aminocoumarins[30]. The discovery of closthioamide from an underexplored anaerobe with a new mode of action holds promise that additional, novel antibiotics might be discovered by parallel methods.

Manipulation of the regulatory elements involved in the global activation of secondary metabolism in *R. cellulolyticum* sustained the production of closthioamide and allowed isolation of seven congeners that showed varied levels of bioactivity[31]. Evaluation of the antibacterial activity of the congeners demonstrated the importance of all six thioamide moieties, terminal aromatic residues, the modular arrangement of the β-thioalanyl units, and the distinct length of spacer units for antibacterial activity[21,31]. Moreover, it was also found that closthioamide is a selective Cu(I) chelator akin to methanobactin and forms a compact symmetrical dinuclear copper complex[7,32].

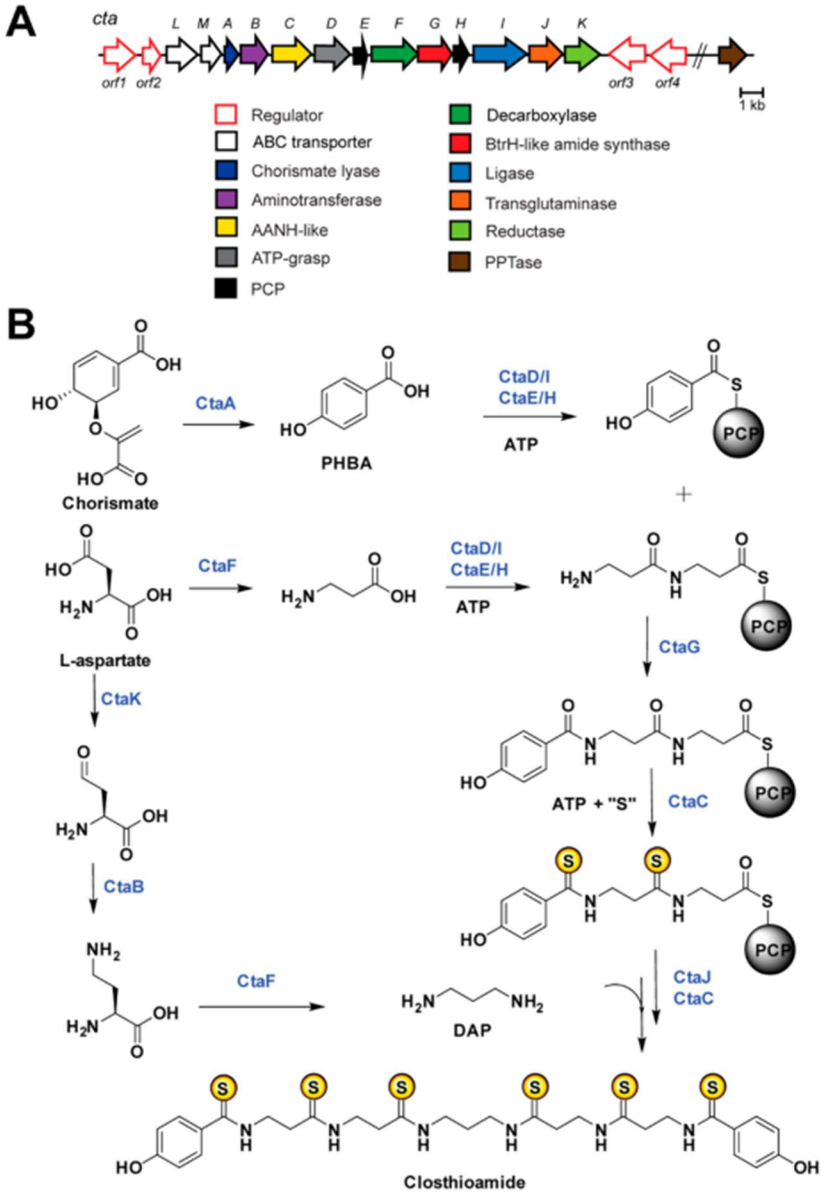


Figure 1-2: Biosynthesis of Closthioamide. (A) Biosynthetic gene cluster of closthioamide from *R. cellulolyticum* DSM5812. (B). Proposed NRPS biosynthetic pathway for closthioamide. (PCP = peptididyl carrier protein; PHBA = *p*-hydroxybenzoic acid; DAP = diamino propane; AANH = α-adenine nucleotide hydrolase)

Through synthesis and application of deuterium- and fluorine-labeled probes, initial insights into closthioamide biosynthesis were obtained, which predicted the involvement of an unusual nonribosomal peptide synthetase (NRPS)[17,31]. A very recent study reported

the closthioamide biosynthetic gene cluster (BGC) in *R. cellulolyticum* and demonstrated that closthioamide biosynthesis involves a novel thiotemplated peptide assembly line that differs from known NRPSs[26,33]. Extensive genome editing and isolation of intermediates from the knockout mutants revealed that the closthioamide BGC consisted of genes ctaA−M (Figure 1-2A), and a preliminary biosynthetic pathway has been proposed (Figure 1-2B)[26]. First, p-hydroxybenzoic acid (PHBA) is synthesized from chorismate by chorismate lyase CtaA and loaded onto a peptidyl carrier protein (PCP), either CtaE or CtaH[34]. This step is probably catalyzed by either CtaD or CtaI, as both enzymes are members of PCP-loading protein families (ATPgrasp and AMP-dependent ligase, respectively)[35]. The second PCP would be loaded with two molecules of β-alanine by the iterative action of CtaD/I, which itself could be biosynthesized from L-aspartate by the decarboxylase CtaF. Then, the PHBA thioester is proposed to be loaded onto this PCP-bound β-alanyl dipeptide. CtaG, whose N-terminus is homologous to an amide synthase, BtrH from butirosin biosynthesis, is the likely candidate[36]. The resultant PCP-bound product would be converted to the corresponding polythioamide intermediate by CtaC, which is homologous to the α-adenine nucleotide hydrolase (AANH) superfamily[22]. Finally, it would be coupled to diamino propane (DAP) by the amide synthase CtaJ and thionated by CtaC to yield mature closthioamide[26]. The biosynthetic intermediate, DAP, could be biosynthesized from L-aspartate by the sequential action of CtaK, CtaB, and CtaF[37]. Currently, the source of sulfur is not known[26]. This is the first example of a thioamidated compound that employs a novel NRPS biosynthetic assembly line[26].

YcaO-TfuA-dependent thioamidation. Thioviridamide (and its derivative JBIR-140) is a RiPP natural product obtained from *Streptomyces olivoviridis* NA05001 that exhibits activity in several cancer cell lines as well as antibiotic properties[38,39]. Thioviridamide features an N-terminal pyruvyl group (although it is hypothesized to be an artifact from extraction[40]), a *β*-hydroxy-N^1,N^3-dimethylhistidinium (hdmHis) residue, and an S-(2-aminovinyl)-cysteine (AviCys) residue that forms part of a macrocycle (see Figure 1-1)[38,40,41]. Thioviridamide also has five thioamide groups in place of backbone amide groups. The thioviridamide (tva) BGC was identified from *S. olivoviridis* and confirmed by heterologous production of thioviridamide in *Streptomyces lividans* TK23[42].

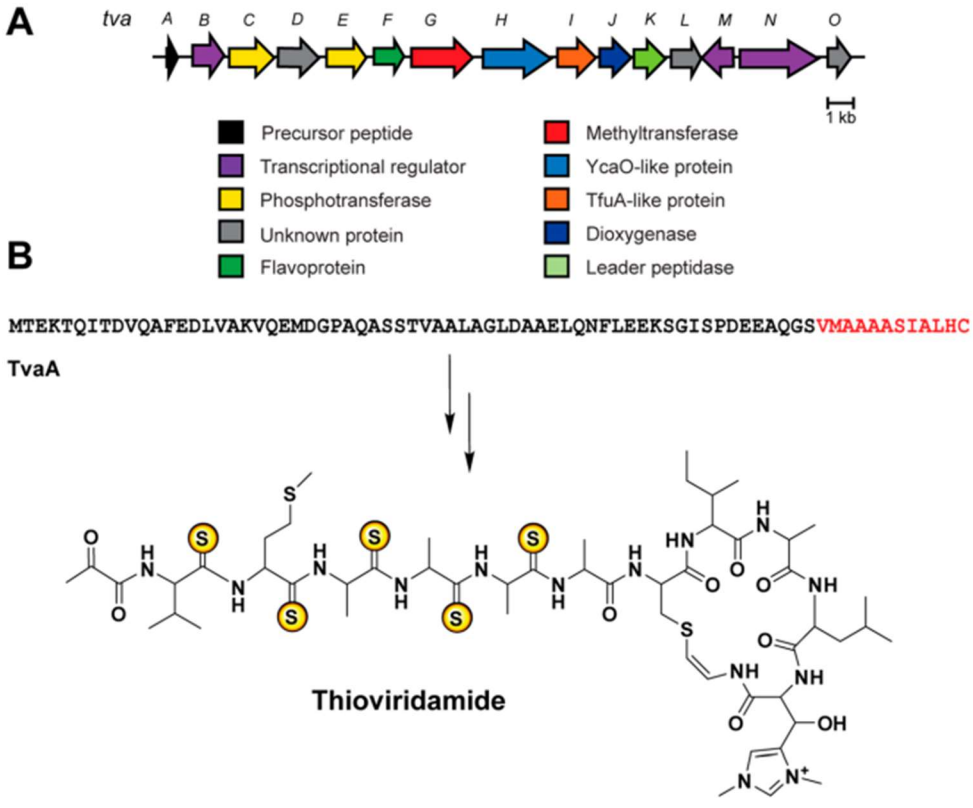


Figure 1-3: (A) Thioviridamide biosynthetic gene cluster from S. olivoviridis NA05001. (B) The core region (red) of the precursor peptide TvaA is converted to thioviridamide by the downstram biosynthetic enzymes. The stereochemistry has not yet been elucidated.

This demonstrated the ribosomal origin of this molecule, which is derived from a 12-amino acid core peptide at the C-terminus of the TvaA precursor peptide. An additional 11 proteins encoded by this gene cluster (TvaB−TvaL) are predicted to be involved in the maturation of the precursor peptide into thioviridamide, although virtually nothing has been reported regarding the individual steps (Figure 1-3)[42]. Two proteins encoded by the thioviridamide BGC have plausible roles in thioamide synthesis. TvaH is a member of the YcaO superfamily[43], while the second, TvaI, is annotated as a "TfuA-like" protein[42]. Biochemical characterization of YcaO family proteins illustrates that they catalyze ATP-dependent cyclodehydration of cysteine, serine, and threonine residues to the corresponding thiazoline, oxazoline, and methyloxazoline, respectively (see Figure 1-4A, cysteine shown), in the biosynthesis of various RiPP natural products[3,43].

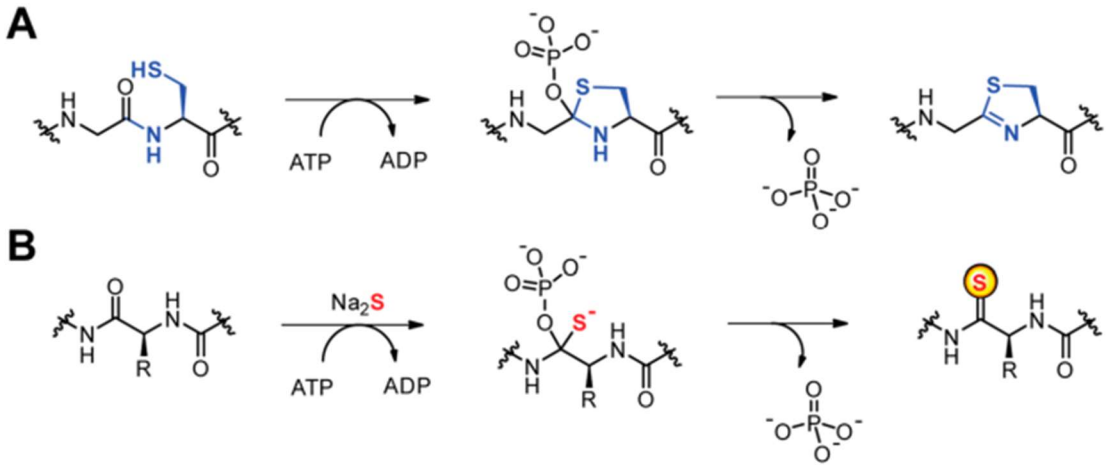


Figure 1-4: Comparison of reactions catalyzed by YcaO enzymes. (A) ATP-dependent cyclodehydration of cysteine, serine or threonine to thiazoline or (methyl)oxazoline. (B) Analogous proposed mechanism for biosynthesis of thioamides on peptidic backbones, with an exogenous sulfide source as nucleophile.

Characterized YcaOs often require a partner protein for efficient cyclodehydration[44,45]. Two types of such partner proteins have been reported so far, with one type resembling E1-ubiquitin activating-like enzymes and the other being termed an "ocin-ThiF"[44,46] protein.

These partner proteins harbor N-terminal RREs[47] that bind mostly the N-terminal leader peptide region. Once the substrate peptide is bound, the YcaO performs modifications in the C-terminal core region[43]. In contrast, two recently characterized YcaOs from bottromycin biosynthesis can catalyze the formation of thiazoline and macrolactamidine moieties in vitro in a manner independent of a partner protein[43,48,49]. On the basis of sequence similarity, the TfuA protein encoded adjacent to the YcaO in the thioviridamide BGC and in numerous other genomic contexts[24] is predicted to assist in the thioamidation reaction, although this proposal awaits biochemical validation in thioviridamide biosynthesis. However, an exogenous source of a sulfide equivalent will be required for thioamide formation (Figure 1-4B). Recently, genetic deletion studies in *Methanosarcina acetivorans* corroborated by in vitro studies of thioglycine formation, a universal post-translation modification (PTM) in MCR, provide the first evidence for this proposal[24,25]. Potential roles for TfuA include allosteric activation of the YcaO and/or delivery of sulfur equivalents, possibly in collaboration with sulfurtransferases. Recent genome-mining efforts led to the discovery of several thioviridamide-like compounds with improved bioactivities (Figure 1-1). Truman and co-workers reported three such analogues, thioalbamide (from *Amycolatopsis alba* DSM44262), thiostreptamide S4 (from *Streptomyces sp.* NRRL S-4), and thiostreptamide S87 (from *Streptomyces sp*. NRRL S-87)[40]. Of these, thioalbamide possesses nanomolar antiproliferative activity with ~6-fold selectivity for cancer cells[40]. Around the same time, the groups of Müller and Koehnke reported the discovery of thioholgamides A/B from *Streptomyces malaysiense* (Figure 1-1)[50]. Thioholgamide A showed markedly increased activity compared to that of thioviridamide with submicromolar activity against several cancer cell lines (30 nM against HCT-116 cells) and a potency 10-fold higher than that of thioholgamide B[50]. The minimal

set of responsible biosynthetic genes has been identified; however, the pathways have not yet been biochemically evaluated[50].

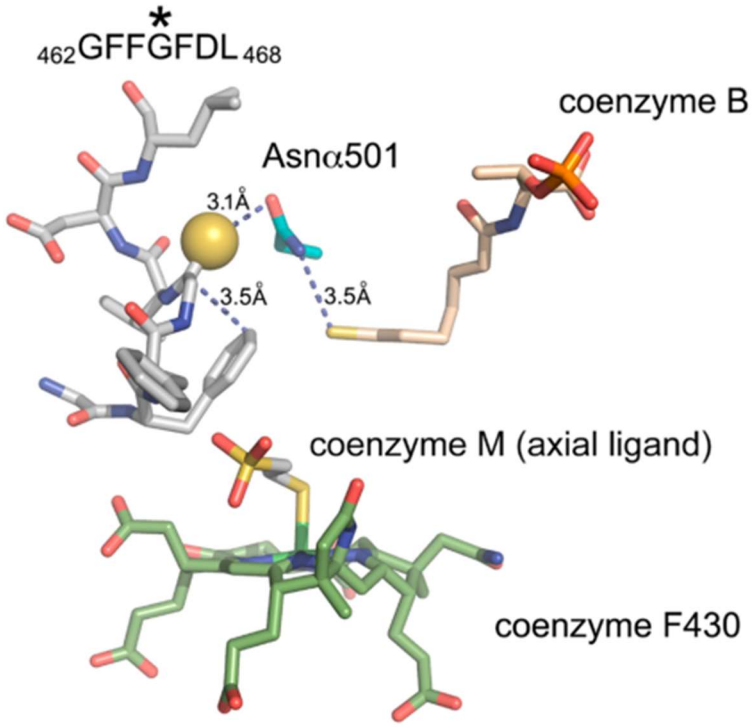


Figure 1-5: View of the MCR active site with the thioglycine involved in several stabilizing interactions using the crystal structure of *Methanosarcina barkeri* (Protein Data Bank entry 1E6Y)[51].

Thioamide-Containing Protein: Methyl-Coenzyme M Reductase (MCR). MCR is found strictly in methanogenic (methane-producing) and methanotrophic (methane-consuming) archaea and carries out the reversible conversion of methyl-coenzyme M (CoM, 2-methylmercaptoethanesulfonate) and coenzyme B (CoB, 7-thioheptanoyl-threoninephosphate) to methane and a CoB−CoM heterodisulfide[52,53]. MCR plays an important role in the global carbon cycle by maintaining steady-state levels of atmospheric methane, a potent greenhouse gas[54]. The 300 kDa enzyme is a heterodimer of three subunits in an $\alpha_2\beta_2\gamma_2$ arrangement and uses a tightly bound, Ni-containing coenzyme F_{430}[52,55]. The Ni(I) oxidation state of this porphinoid cofactor is crucial for catalysis[56].

Notably, the α subunit of MCR (MCRα) has several unusual PTMs, including 3-methylhistidine, S-methylcysteine, 5-methylarginine, 2-methylglutamine, didehydro-aspartate, and thioglycine, with varying frequencies of occurrence[57–59]. Thioglycine is present in all of the methanogens analyzed thus far, and MCR is the only protein known to biology to bear a thioamide PTM[58]. Although there have been proposals for thioglycine in the MCR catalytic mechanism[60–62], recent work has shown that it might instead play a structural role in properly organizing the active site[24]. Thauer originally proposed that thioglycine formation in MCR could be similar to thioviridamide biosynthesis[38,58]. Upon elucidation of the thioviridamide BGC, it was then proposed that ycaO−tfuA would be involved in MCR thioamidation. Bioinformatics analysis of all available methanogen genomes revealed the universal occurrence of ycaO with ∼90% also encoding a tfuA gene[24]. To evaluate any role in thioglycine formation, ycaO and tfuA genes were deleted (individually and together) *in M. acetivorans*[24]. MCRα peptide fragments from the wild-type and deletion strains were analyzed by mass spectrometry, and it was found that each variant (i.e., ΔycaO, ΔtfuA, and ΔycaO−tfuA) lacked thioglycine (at Gly465 in MCRα), revealing their importance in thioamide installation[24]. While each variant was viable, they were incapable of growth at higher temperatures and displayed marked growth defects compared to wild-type *M. acetivorans*[24].This could be due to the increased flexibility of the unmodified glycine in the ΔycaO−tfuA strain, which would render the contorted conformation of the peptide backbone, otherwise stabilized by several interactions involving the thioglycine, in the Gly462−Leu469 region of MCRα considerably less stable (Figure 1-5)[24]. This genetic deletion study was further supported by in vitro thioamidation using purified YcaO, TfuA, ATP, and substrate peptides from *M. acetivorans* MCRα. The requisite bisulfide nucleophile was supplied chemically as sodium sulfide or produced enzymatically from L-cysteine and *M. acetivorans* IscS[25]. The residues encompassing

Gly465 that are directly engaged by YcaO were evaluated using biochemical and biophysical methods. Structural insights were obtained using thermophilic YcaO homologues from *Methanopyrus kandleri* and *Methanocaldococcus jannaschii*, which confirmed the ATP-binding pocket. Sequence- and structure-guided alanine scanning mutagenesis was performed to elucidate the role of residues involved in ATP/Mg^{2+} and possible peptide binding[25]. According to a mechanistic proposal supported by spectroscopic and isotopic labeling experiments (Figure 1-4), upon substrate binding, an external source of sulfide will attack the target amide bond (in this case, Gly465) generating a tetrahedral intermediate. The amide oxyanion will then attack the *γ*-phosphate of ATP, releasing ADP and a phosphorylated thiolate intermediate. This thermodynamically favorable step in which ATP cleavage is coupled with C−S bond formation could be concerted or stepwise. Finally, the tetrahedral intermediate will collapse by releasing phosphate and the thioamidated peptide[25]. Further studies are needed to elucidate the substrate orientation in the active site as well as the role of TfuA, which is currently proposed to act as a partner to YcaO that may regulate ATP usage or participate in sulfur delivery in collaboration with sulfurtransferases[25].

Potential for Other Thioamide Natural Products. The discoveries of thioamide biosynthesis pathways and the assignments of thioamide-specific functions in polypeptides reveal the potential that this backbone modification might be widely used in nature to confer evolutionarily valuable properties that presumably cannot be attained through side chain modification. However, one is then left to wonder why natural thioamides are not more prevalent. Studies of MCR, thioviridamide-like molecules, and thioamidated thiopeptides have identified biosynthetic signatures of such compounds, which surveys predict to be considerably more numerous than currently appreciated[23–25].

Indeed, given that a thioamide containing peptide is ~16 Da heavier than the corresponding amide, the mass change can easily be misinterpreted as oxidation. This may explain why some thioamide-containing natural products have gone undetected, although thioamidation may still represent a relatively rare PTM. Bioinformatics-driven searches based on the presence of ycaO and tfuA genes are already aiding in the discovery of thioamide-containing molecules and will undoubtedly help to resolve such misassignments. As more thioamide-containing compounds are discovered, the ability to synthetically introduce thioamides will be important to define the functional role of the thioamide. This knowledge can be applied by exogenously introducing thioamides to confer new functions on peptides and proteins.

Synthetic Thioamide in Peptides and Proteins

Studies of the roles of natural thioamides are enhanced by the ability to site-specifically incorporate thioamides to generate biosynthetic precursors, probe molecules, or natural product analogues. This is well-illustrated in the studies of closthioamide by Hertweck and co-workers noted above[30,32]. In addition to testing the activity of synthetic closthioamide[63], they prepared closthioamide variants with some or all of the thioamides replaced by amides, variants that had been observed in *R. cellulolyticum* extracts[32]. Using these compounds, amide variants were not converted to thioamide variants, but thioamide variants were hydrolyzed to form the amide variants. Thus, it appeared that thioamide incorporation occurred during closthioamide backbone synthesis, not through downstream modification of amides. In the second study, the motifs that are crucial to the antibiotic properties of closthioamide were investigated by modulating, among other things, the number and position of thioamides[30]. Removal of even one thioamide caused a >100-fold decrease in potency toward methicillin-resistant *S. aureus* and vancomycin-resistant

Enterococcus faecalis. As these studies make clear, the ability to insert thioamides at specific locations allows one to interrogate the biosynthesis and mechanism of action of thioamide containing natural products.

Chemical Synthesis of Thioamide-Containing Peptides. Site-specific installation of a thioamide into a peptide of interest is readily accomplished by incorporating a thioamide precursor during peptide synthesis rather than attempting a selective transformation of a specific amide bond. Various routes have been investigated, including activation of thioacids with phosphonium-based reagents like benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP)[30] or the synthesis of thioacyl-benzimidazolinones[64]. However, these reactions suffered from low yields and racemization of the thioamide residue. An improved route was devised based on thioacyl-benzotriazoles, which are sufficiently reactive that they can thioacylate amines in the presence of a base with no other activating reagent[65]. While tert-butyloxycarbonyl (Boc)-protected precursors were synthesized in the initial report, the methods have been shown to be compatible with fluorenylmethyloxycarbonyl (Fmoc)-protected amino acids[66], which are more commonly used. These precursors have become widely used; the route to their synthesis and incorporation is illustrated in Figure 1-6: Site-specific incorporation of thioamides using solid phase peptide synthesis (SPPS). (A) Thioacyl-benzotriazole monomers can be synthesized in three steps from Fmoc-protected amino acids. (B) The thioacyl-benztriazoles can be used to introduce the thioamide during SPPS.. The synthesis starts with Fmoc protected amino acids, which are commercially available in most cases, with typical yields of 40% over the three steps. So far, thioamide versions of 15 of the 20 common proteinogenic amino acids (Ala[67], Arg[68], Asp[69], Cys[68], Glu[70], Gly[68], Ile[71], Leu[72], Lys[68], Phe[73], Pro[73], Ser[68], Trp[68], Tyr[71], and Val[74]), as well as hydroxyproline[71],

have been reported. While the aforementioned benzotriazole synthesis can be used for Boc- or Fmoc-based precursors, thioamides are incompatible with the harsh HF cleavage used in Boc-based solid phase peptide synthesis (SPPS). However, thioamides are compatible with the conditions and reagents used in Fmoc based SPPS, including various activating reagents (for the other amino acids), bases, cleavage additives, capping solutions, and strong acids like trifluoroacetic acid (TFA) that are used in cleavage reactions. Fmoc-based SPPS using thioacyl-benzotriazole building blocks has been used to routinely prepare thioamide-containing peptides of 20−40 amino acids on a multimilligram scale[71,75].

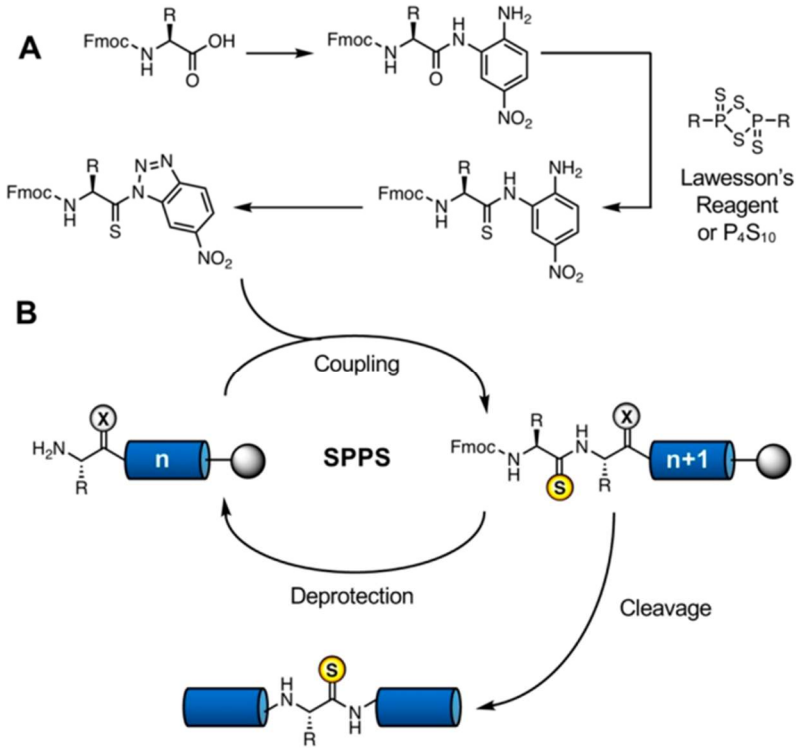


Figure 1-6: Site-specific incorporation of thioamides using solid phase peptide synthesis (SPPS). (A) Thioacyl-benzotriazole monomers can be synthesized in three steps from Fmoc-protected amino acids. (B) The thioacyl-benztriazoles can be used to introduce the thioamide during SPPS.

There are several side reactions that must be considered during SPPS and cleavage. First, prolonged exposure (>30 min) to strongly acidic conditions [95% TFA] typically used during peptide cleavage can lead to an Edman-degradation type reaction that results in backbone cleavage at the n + 1 site[76]. Therefore, one must limit TFA concentrations and deprotection times to achieve a delicate balance between full removal of all acid-labile side chain protecting groups while also maintaining the integrity of the thioamide. Second, thioamide precursors can react with residual amounts of water during coupling reactions, resulting in an S-to-O exchange reaction to yield the corresponding amide. Fortunately, the use of anhydrous methylene chloride during thioamide coupling can greatly reduce the level of this side reaction[68]. Last, thioamides can undergo epimerization during peptide synthesis. While there have been few studies in peptides, studies of model compounds consistently show that the pK_a of the thioamide C_α proton is ~5 units lower than that of the corresponding amide, equating to a pK_a of 12−13 in peptides. A more detailed analysis can be found in chapter 2 and elsewhere[77]. Therefore, prolonged exposure to base during Fmoc deprotection steps can lead to peptide epimerization. Although there have been various approaches to this problem[70,77], the use of a 2% (v/v) solution of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) seems to give the best combination of high deprotection efficiency and minimal epimerization. This is particularly important for the synthesis of peptides containing multiple adjacent thioamides[72,78], motifs found in the thioviridamide family of compounds. While alternate methods such as recently developed ynamide-based couplings may lead to further improvements[79], with these considerations, the thioacyl-benzotriazoles can be used to access many thioamide containing peptides.

Uses of Synthetic Thioamide Peptides as Fluorescence Quencher Probes.

The use of thioamides as spectroscopic labels, particularly as circular dichroism (CD) probes or fluorescence quenchers, has limited utility in complex biological environments but has been successfully been used in *in vitro* experiments. Thioamides have been employed as CD probes due to their unique signature at 270 nm[80], which is based on their π→π* transition[81]. Furthermore, thioamides have been successfully used as fluorescence quencher probes by our lab and others. Their versatility comes from the ability to quench a wide range of fluorophores through two commonly employed ways of energy transfer: The first one is Förster Resonance Energy Transfer (FRET), a non-radiative energy transfer mechanism that relies on spectral overlap of the fluorophore emission with the thioamide absorption spectrum. FRET is a distance dependent energy transfer mechanism and FRET pairs are in general described by the Förster radius (R_0), which is the distance at which energy transfer is half-maximal. R_0 can be calculated using the so called Förster equation shown in Equation 1, where κ^2 is a factor to describe the orientation of the interacting transition dipoles of the donor and acceptor, Φ is the donor quantum yield, J is the spectral overlap integral, n is the refractive index and N_A is Avogadro's number[82].

$$R_0{}^6 = \frac{(9000)\ln(10)\kappa^2\Phi J}{128\pi^5 n^4 N_A}$$

Equation 1: Equation to calculate Förster distance.

The spectral overlap integral is defined by Equation 2, where $f_D(\lambda)$ is the normalized donor emission spectrum and $\varepsilon_A(\lambda)$ is the normalized acceptor emission spectrum. As evident from Equation 1 and Equation 2, there is a strong wavelength dependence on the

Förster distance R_0, so the choice of acceptor/donor and their excitation/emission wavelengths influence the range in which distances can accurately be determined.

$$J = \int_0^\infty f_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda$$

Equation 2: Calculation of the spectral overlap integral *J*.

The calculated R_0 can be used to calculate the distance between two fluorophores using Equation 3.

$$E_Q(FRET) = \frac{1}{1 + (R/R_0)^6}$$

Equation 3: Linking Quenching efficiency to distance.

For thioamides, the distance that can accurately be measured is usually in the range of approximately 10-30 Å[83,84]. For shorter distances, a different quenching mechanism can be exploited. Due to their lowered oxidation potential (see Table 1.1), thioamides are able to quench fluorescence through photoinduced electron transfer (PET). Compared to FRET, the distance dependence of PET quenching is more complicated, so PET is only able to provide rough distance information and in most cases van der Waals contact between the fluorophore/quencher pair is required[84]. Whether PET is able to occur or not can be rationalized by calculating the free energy transfer of electron transfer using the Rehm-Weller Equation show in Equation 4, where E_{ox}(D) and E_{red}(A) are the oxidation and reduction potentials of donor and acceptor respectively, ΔG_{00} is the zero-vibrational electronic excitation energy[85]. The last term describes the coulombic attraction energy experienced by the pair post electron transfer with ε being the dielectric constant of the solvent and *d* being the distance. However, the last term is usually small enough to be neglected in aqueous buffers[84].

$$\Delta G = E_{ox}(D) - E_{red(A)} - \Delta G_{00} - \frac{e^2}{\varepsilon d}$$

Equation 4: Rehm-Weller Equation to calculate the free energy of electron transfer.

While ΔG is able to predict if an electron transfer will occur or not, the magnitude of ΔG is not correlated with the quenching efficiency EQ. While Marcus Theory could be applied to calculate the rate of electron transfer, detailed knowledge of all photophysical rates would be required. Therefore, currently experimental determination of EQ is the most straightforward way. In most cases where fluorescence quenching by thioamides was observed, thioamides were assumed to act as donor while the fluorophores were assumed to act as electron acceptor[86]. However, it is possible that in some cases a reversed role could be possible. However, electrochemical data to calculate this order of reaction is not yet available.

Our group and others have exploited the use of thioamides as fluorescence quenchers in various ways. In an early work from our lab, thioamides have been used to monitor protease activity[87]. Substrate peptides were designed to where one terminus had a thioamide, while 7-methoxycoumarinylalanine was attached to the other. In the initial state the fluorescence is quenched. Upon cleavage of the substrate the two fragments are able to diffuse away from one another, resulting in an increase of fluorescence. This increase in fluorescence is directly correlated to protease activity and kinetic parameters derived from the assays agreed well with reported literature values. Furthermore, our lab demonstrated that this method works for a variety of different proteases, including serine-, cysteine-, carboxyl- and metallo-protases such as trypsin, chymotrypsin, pepsin, thermolysin, papain and calpain.

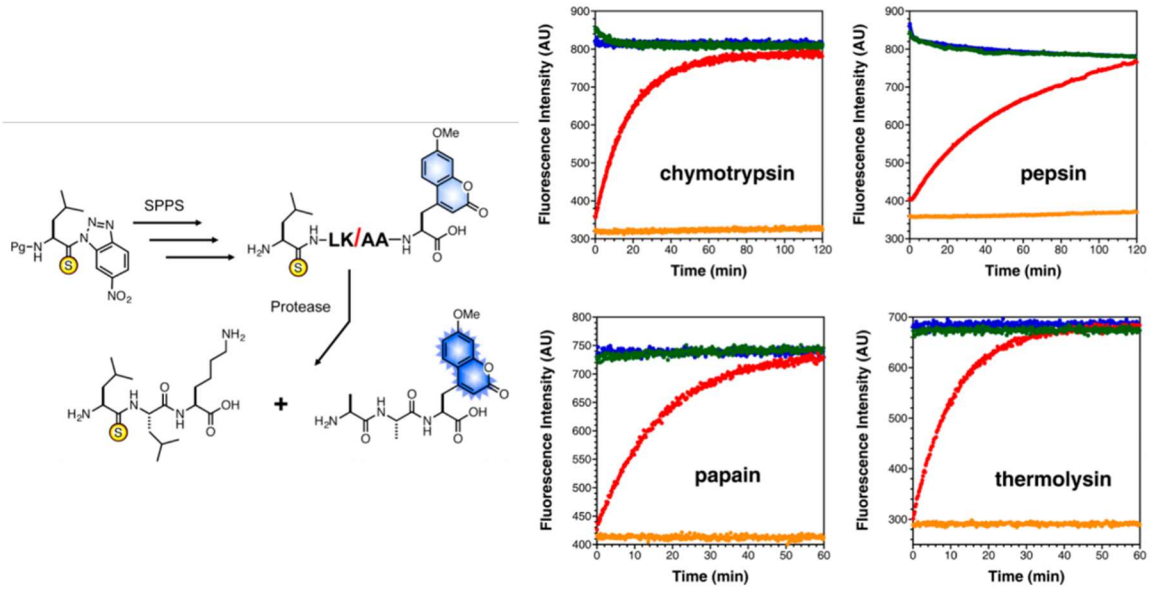


Figure 1-7: Using thioamides to monitor protease activity. Left: Reaction scheme; Right: Fluorescence time courses from activity assays testing various proteases. Green/blue trace: oxopeptide in the presence/absence of protease respectively. Red/Orange: thiopeptide in the presence/absence of protease respectively.

While these results demonstrate that thioamides can be used for this application, questions about the influence on activity with regards to positioning the thioamide in reference to the scissile bond remained. Our lab has conducted follow-up experiments for serine- and cysteine proteases in this regard and the results will be published soon.

Another similar utility of thioamides was demonstrated by the Vocadlo lab in their study of *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) modifying enzymes. Specifically, they are interested in studying activity of human *O*-GlcNAcase (hOGA). For this purpose they synthesized probes where Resorufin or 4-Methylcoumarin was *O*-linked at the anomeric position of GlcNAc and a thioamide was appended to the 2' or 4' position[88,89]. The quenching efficiency was determined by comparing fluorescence signals of equimolar solutions of the probes with or without the thioamide at the 2' position. The average quenching efficiency was >90% for the 4-Methylcoumarin modified GlcNAc and ~50% for

Resorufin modified one. Quenching efficiencies for probes with thioamides attached at the 4' position varied based on linked chemistry but were in the same range as for the 2' attachment. All of the probes worked with hOGA and were specific. Unfortunately, all of the thioamide modified probes reduced the enzyme efficiency k_{cat}/K_m. Depending on the linker and point of attachment there was a 20-1000 -fold decrease in enzyme efficiency[88,89].

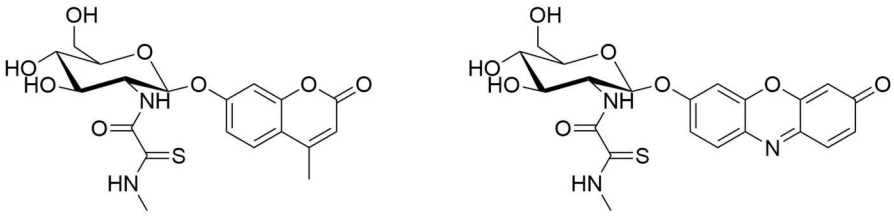


Figure 1-8: Thioamide containing probes to study *O*-GlcNAcase.

While these are only two examples, they demonstrate how thioamides can be used to develop enzymatic turn-on probes as a mean to study enzyme kinetics. However, they also demonstrate that careful consideration of thioamide placement is crucial to not impact enzyme efficiency.

Synthetic thioamide peptides for other uses. The prospects of using the nearly isosteric thioamide to intentionally modify the physical properties of the peptide backbone have long intrigued biochemists. As noted above, these properties can be roughly categorized into four categories: increased reactivity with nucleophiles, electrophiles, and soft metals; altered conformational properties; altered hydrogen bonding propensity; and altered photophysical or electrochemical properties. All of these properties have been exploited for various purposes. The first biochemical study of a synthetic thioamide containing peptide was published by Nobel laureate Vincent du Vigneaud in 1973[90]. This study focused on the C-terminal amide of the cyclic 9-mer peptide hormone oxytocin. It

was known that the C-terminal amide was crucial for biological activity, but none of the previously studied analogues altered the carbonyl moiety. Thioamide substitution at this site reduced the biological activity to ≤6% of oxytocin in all assays tested. This study demonstrated that although it is nearly isosteric with an amide and retains functional groups capable of donating and accepting hydrogen bonds, a thioamide modification can have significant effects on biological signaling. Since this pioneering work, thioamides have been used in biophysical studies of protein folding, as spectroscopic labels, as probes of hydrogen bonding in protein interactions, and as synthetic intermediates in the synthesis of other carbonyl modifications. For example, Boger's laboratory used thioamides both as synthetic tools and as probes of hydrogen bonding interactions with a backbone carbonyl in vancomycin[8]. Through the introduction of thioamides as useful intermediates, synthetic chemists have been able to synthesize analogues that have different hydrogen bonding patterns and test the derivatives with model target substrates to confirm the assumed mechanism of action[8] and to develop more potent variants of vancomycin[91,92].

Another very recent example is the Ag(I) promoted macrocyclization of peptides using N-terminal thioamides[93]. The Hutton laboratory was able to demonstrate that protected peptides with N-terminal thioamides can be cyclized as demonstrated in Figure 1-9. After subsequent deprotection the resulting macrocyclic peptide contains a canonical oxoamide. Compared to other methods, their method is fast (1 hr), high yielding (80-90% with >90% purity) and without epimerization. They successfully synthesized various sized macrocycles with 5-10 amino acids.

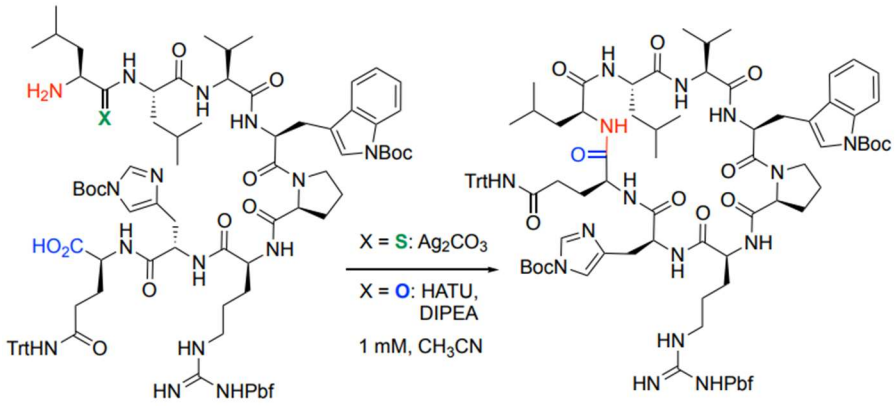


Figure 1-9: Reaction scheme conventional macrocyclization and silver(I) promoted thioamide macrocyclization.

The use of thioamides as spectroscopic labels, particularly as circular dichroism probes or fluorescence quenchers, has limited utility in complex biological environments, but they have found other uses in cell-based or *in vivo* experiments. One photophysical property that has a demonstrated utility in biological systems is the ability to use thioamides as photoswitches. Excitation at 270 or 340 nm ($\pi \rightarrow \pi^*$ or n $\rightarrow \pi^*$ transition, respectively)[16], allows for selective photoswitching of the thioamide between the trans and cis isomers. Additionally, the metastable photoinduced cis state has a slow thermal relaxation rate ($k < 1 \times 10^{-3}$ s^{-1})[94] and can populate up to 50% of the photostationary state[95]. Two studies illustrate the use of thioamide photoswitching in biological contexts. Kiefhaber and Fischer investigated the enzymatic activity of ribonuclease S[96]. This cleaved enzyme requires the S-protein and complementary S-peptide for activity. Incorporation of a thioamide moiety into the helical S-peptide at central positions had no significant effect on enzyme activity as measured by hydrolysis of cytidine 2′,3′- cyclic monophosphate (cCMP). However, UV irradiation induced 30% isomerization of the peptide to the *cis* state, correlating with a 30% decrease in enzyme activity without dissociation of the protein−peptide complex. This indicated that switching the thioamide

into the *cis* state led to a conformational change that abolished enzyme activity. In another study, the conformation−activity relationship of a C-terminal pentapeptide activator of insect kinin was investigated[95]. NMR and molecular modeling studies were inconclusive as to whether a 1−4 or 2−5 *β*-turn was the active conformation of the peptide[97]. In a cockroach hindgut myotropic activity assay, the photoswitched thioamide peptide in the *cis* conformation had a 4-fold lower EC_{50} compared to that of the native peptide or trans thioamide peptide, confirming that a 1−4 turn was necessary for bioactivity (Figure 1-10)[95]. These examples show that thioamide photoswitching can regulate peptide activity in cell lysates and similar systems, providing a backbone switch with much less structural perturbation than an azobenzene replacement[98].

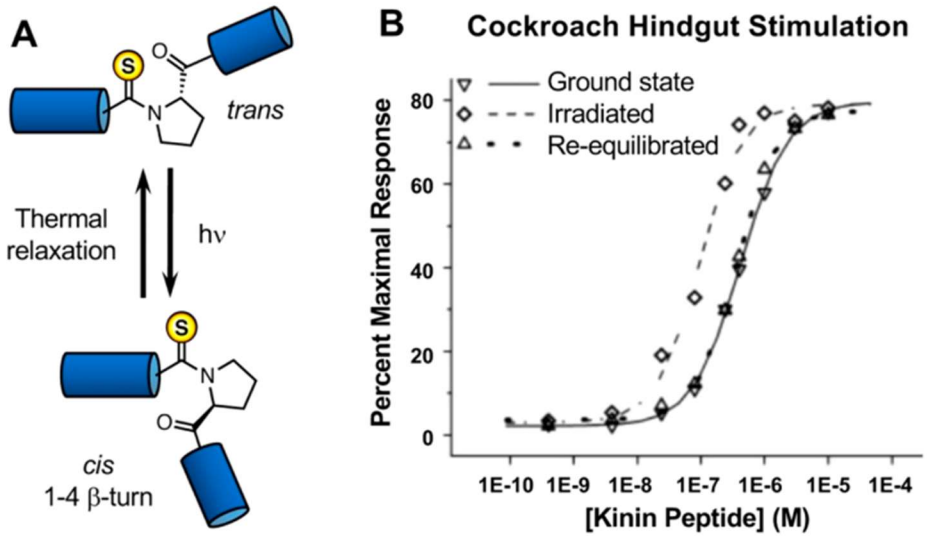


Figure 1-10: Controlling bioactivity through thioamide photoisomerization. (A) The insect kinin-derived thioamide pentapeptide can undergo UV-induced photoisomerization to adopt a 1-4 β-turn structure. (B) Titration of isolated *cis* or *trans* peptide in an *ex vivo* myotropic contraction assay shows a 4-fold lower EC_{50} for the *cis* peptide.

The distinct properties of thioamides have also been exploited in studies involving proteases. Most investigations have centered on the effects of thioamides in short peptide

substrates of various proteases, with the intention of developing inhibitors and investigating the protease mechanism (e.g., favored interactions with soft metals substituted in the active sites of metalloproteases)[99]. For example, Fischer and co-workers investigated thioalanine-proline-*p*-nitroaniline (Ala^S-Pro-pNA) as a substrate for dipeptidyl peptidase 4 (DPP-4), where the enzyme hydrolyzes the bond between Pro and pNA[100]. They found a 1100-fold decrease in k_{cat}/K_m compared to that of the amide, which they attributed to the decrease in k_{cat} caused by the increased rotational barrier of the thioamide. Recently, these results were extended by introducing thioamide substitutions in glucagonlike peptide 1 (GLP-1)[75], a natural substrate of DPP-4 that stimulates the release of insulin while suppressing glucagon release[101]. However, GLP-1 is rapidly degraded ($t_{1/2}$ = 2 min) *in vivo* by DPP-4, which cleaves the two N-terminal amino acids and renders GLP-1 inactive[102] (see Figure 1-11A). Injection of DPP-4- resistant GLP-1 variants has become a common treatment for type II diabetes[102]. The single-atom thioamide substitution at either of the two terminal positions increased the peptide halflife in an *in vitro* proteolysis assay from 2 min to 12 h (Figure 1-11C). Competition experiments with an alternate DPP-4 substrate revealed that thioamide GLP-1 was not a competitive inhibitor, seemingly in conflict with the finding of a primary k_{cat} effect. This may be due to the fact that the 36-residue GLP1 peptide cannot be repositioned in the active site to accommodate the thioamide, whereas the Ala^S-Pro-pNA can but in a way that is not optimal for catalysis. Examination of the crystal structure of DPP-4 with a peptide substrate[103] reveals bifurcated hydrogen bonds with the carbonyls of the two N-terminal amino acids (Figure 1-11B). Thioamide incorporation at either site leads to DPP-4 resistance, a finding that can be explained on the basis of weakened hydrogen bond acceptance by the thioamide (although the rotational barrier may still play a role).

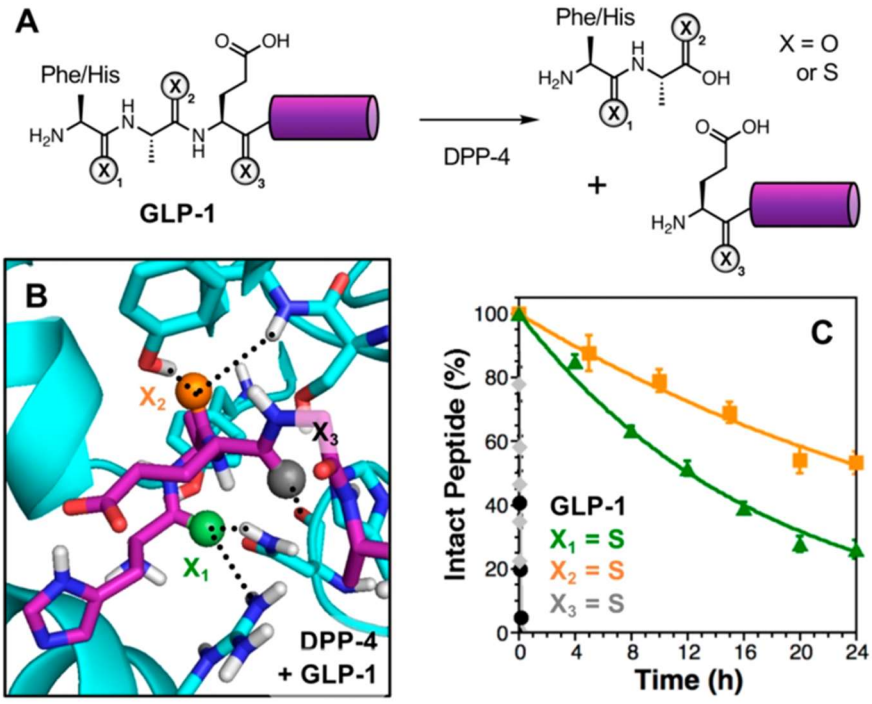


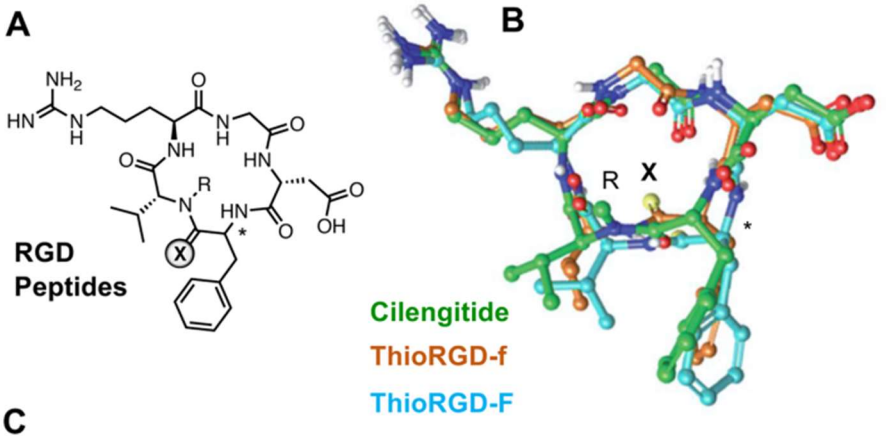
Figure 1-11: (A) Incorporation of a thioamide into Glucagon-like peptide 1 (GLP-1) at the X_1 or X_2 position prevents dipeptidyl peptidase 4 (DPP-4) degradation. (B) Image of the DPP-4 (cyan) active site with a modeled GLP-1 fragment (based on PDB entry 1R9N)[103]. X_1 – X_3 carbonyls and their interactions are highlighted. (C) *In vitro* degradation assay shows that thioamide substitution at X_1 or X_2 but not X_3 can effectively prevent proteolysis by DPP-4.

Rats injected with the thioamide-containing version of GLP1 had a significantly reduced blood glucose spike in an oral glucose tolerance test compared to that of rats injected with GLP-1 or the vehicle control, showing that DPP-4 resistance was retained *in vivo*. In cell-based receptor activation assays, thioamide GLP-1 had a comparable potency for cyclic AMP activation (which controls insulin release) but a substantially lower potency for activation of *β*-arrestin 1 and 2 (which has been correlated with effects of GLP-1 on appetite)[104,105]. While a detailed analysis of this bias effect was not possible at the time of publication of the thioamide GLP-1 study, subsequent disclosure of a cryo-electron microscopy structure of the GLP-1 receptor (PDB entry 6B3J) could be used as a basis

for modeling to explain both the signal bias and the thioamide positional effects[106]. These studies illustrate how thioamide substitution has the potential to influence interactions of peptides with other proteins, preventing proteolytic cleavage and eliciting subtle changes in receptor activation. While thioamide substitution has been used to confer protease resistance in several other *in vivo* studies[107,108], we highlight the DPP-4 work because of the level of mechanistic understanding available.

A recent publication by Chatterjee and co-workers nicely illustrates how the altered conformational properties of thioamides can be exploited in medicinal chemistry applications for macrocyclic peptides[109]. Because of their large surface area per molecular weight, easy variability of functional groups, and improved bioavailability compared to that of linear congeners, peptide macrocycles are primed to be used as protein−protein interaction inhibitors[110]. However, in spite of their cyclic constraint, macrocycles can exhibit significant conformational flexibility, often making them low-affinity or nonspecific binders *in vivo*. NMR experiments, including ^{13}C spin−lattice relaxation time and temperature coefficient measurements, were used to show that introduction of a thioamide into a peptide macrocycle narrowed the breadth of the conformational ensemble, limiting the macrocycle to a single observable conformation. After initial investigation of model peptide macrocycles, the authors turned to bioactive RGD peptides. RGD peptides are known antagonists of proangiogenic integrins and represent attractive drug targets[110] for cancer and other indications. Cilengitide, an *N*-methylated cyclic peptide that reached Phase III clinical trials against glioblastoma, was used as a reference for *in vitro* and *in cellulo* binding assays[111]. Some of the thioamide-containing RGD peptides bound more tightly to integrins than cilengitide did, and computational docking of their solution NMR structures aligned well to a bound cilengitide molecule in an integrin receptor cocrystal structure

(Figure 15)[112]. These docking studies allowed them to identify the basis for the increased affinity in the optimal thioamide RGD macrocycle as stabilization of a certain ring conformation. Interestingly, ex vivo metabolic stability assays in human serum over 72 h revealed that all tested thioamide RGD macrocycles were more stable than cilengitide, independent of the position of the thioamide. Like the thioamide GLP-1 molecules, these RGD analogues show strong prospects as injectable therapeutics or vehicles for imaging probes.



RGD Peptide	$t_{1/2}$	Breast Cancer	Glioblastoma
Cilengitide X = O, R = Me, *D-Phe	12 h	1.2 nM	3.0 nM
RGD-f X = O, R = H, *D-Phe	8 h	$>10^5$ nM	$>10^5$ nM
ThioRGD-f X = S, R = H, *D-Phe	9 h	4 .0 nM	326 nM
RGD-F X = O, R = H, *L-Phe	7 h	$>10^5$ nM	$>10^5$ nM
ThioRGD-F X = S, R = H, *L-Phe	36 h	1.0 nM	1.6 nM

Figure 1-12: (A) Chemical structures of RGD peptide macrocycles, including cilengitide and two matched amide/thioamide pairs. (B) NMR structures of the f^S (organge), F^S (cyan) peptides overlaid with structure of cilengitide bound to *αvβ3* integrin (PDB entry 1L5G)[112]. (C) Summary of serum stability ($t_{1/2}$) and activity (IC_{50} for binding to breast cancer and glioblastoma cells)

Together, these studies show that thioamides can profoundly modulate the biological activity of peptides. Their unique combination of easy installation and amide mimicry with distinct physical properties allows targeted manipulation of various systems

to tune protein interactions. These provide valuable tools for basic science or translational research and help to shed light on the roles of thioamides in natural peptides.

Thioprotein Semisynthesis. SPPS is largely sufficient to synthesize thioamide analogues of peptide hormones or natural product derivatives. However, to investigate the role of thioamides in proteins like MCR or to use them as probes in full-sized proteins, other methods are necessary. To the best of our knowledge, the longest directly synthesized thioamide peptide/small protein was a thioleucine-modified version of the 56-amino acid B1 domain from protein G (GB1)[71]. However, the isolated yield of this GB1 variant was very low (1%), showing the size limitations of SPPS for generating thioamide proteins. This problem has been rectified using native chemical ligation (NCL). NCL allows the ligation of two peptides or protein fragments together at low millimolar concentrations under neutral pH, aqueous conditions, resulting in a native amide bond[113]. The main requirement and/or limitation of this method is that proteins must be able to refold into their native structure after ligation in a denatured state. For this reaction to occur, the C-terminal fragment requires an N-terminal Cys or a β-, γ-, or δ-thiol-containing amino acid derivative, and the N-terminal fragment requires a C-terminal thioester. While the generation of C-terminal peptide fragments containing thioamides is straightforward, the generation of N-terminal fragments is complicated by the need to generate a C-terminal thioester. It was once speculated that ligation of thioamide-containing peptides was "not suitable because the presence of [thioamide] bonds is not compatible with the subsequent synthesis of the thioester moiety."[96]. Since then, thioamides have been shown to be compatible with various peptide activation methods to generate C-terminal thioesters. Such methods include conventional solution phase PyBOP activation[114] and more recent in situ thioesterification strategies making use of latent thioesters, such as C^bPG_o (N-to-S acyl

shift) [114,115], ChB (O-to-S acyl shift)[116,117], and Nbz, which promotes thioester formation by activating the C-terminal carbonyl[114,118]. Many of these older methods suffer from low yields, slow reactions, or racemization. Currently, the fastest and most effective way of generating C-terminal thioesters in thioamide-containing peptides makes use of acyl hydrazide methods, for which detailed procedures are available[119]. The acyl hydrazide modification is stable to most purification conditions and latent until activation using sodium nitrite at pH 4.0, which converts it to an acyl azide. Readjustment of the pH to 7.0 and addition of an aromatic thiol generates a thioester that can be immediately used in NCL reactions. Thioamide-containing peptides have been shown to be compatible with this activation reaction with no observed side reactions and have been used to generate proteins with thioamides near the N-terminus[120].

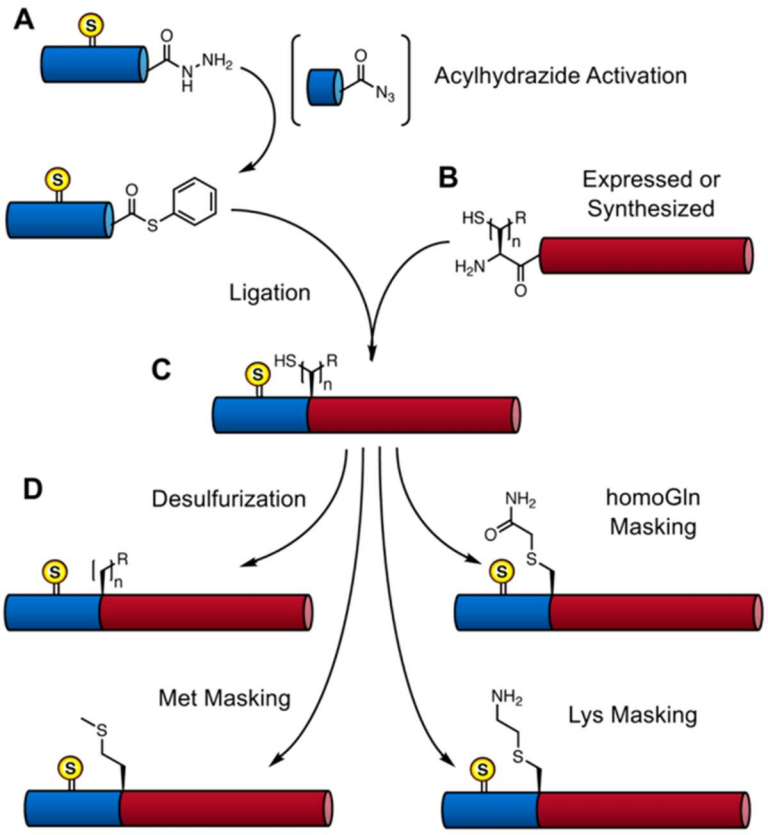


Figure 1-13: Traceless ligation strategy to synthesize thioamide-containing proteins. (A) Thioamide-containing thioester peptides (blue) are synthesized from acylhydrazide precursors to generate a thioester. (B) N-terminal cysteine fragments (red) are synthesized or recombinantly expressed. (C) Ligation can be performed without special precautions for the thioamide. (D) Masking of the ligation site can occur through desulfurization, Met masking, homoGln masking and Lys masking.

Another concern in the field of NCL is that the requirement for a Cys residue at the ligation site forces one to ligate long fragment sequences or results in semisynthetic proteins with non-native Cys residues included. There are two general approaches to circumventing this issue, and both have been demonstrated to be compatible with thioamides (see Figure 1-13). The first strategy involves masking the cysteine or analogue thereof, resulting in a mimic of another amino acid. For example, homocysteine can be used as a ligation handle and treated with methyl iodide after ligation to convert it to

methionine[69]. Alternatively, cysteine can be treated with alkyl halide reagents like iodoacetamide or 2-bromoethylamine, which converts it to thioether mimics of (homo)glutamine or lysine, respectively. The second strategy, radical desulfurization, converts cysteine to alanine and converts β-, γ-, or δ-thiol analogs of various amino acids to their respective canonical amino acids[121]. Desulfurization of thioamide-containing peptides using classical Raney nickel procedures results in undesired backbone cleavage at the thioamide position[120]. The more recently developed VA-044, a water-soluble diazo compound that acts as a radical initiator, has been shown to be compatible with backbone thioamides, provided that thioacetamide is used as a scavenger to suppress an S-to-O exchange side reaction[120]. With this combination of methods, thioamides can be incorporated into essentially any protein that is amenable to synthesis by NCL. At the millimolar concentrations typically used in NCL, reactions proceed in a few hours for ligations of short peptides and in 1−2 days for ligations with at least one large protein fragment (>60 residues). The limiting reagent in this process is the amount of purified thioamide-containing peptide, which usually is obtained in a 20−50% yield relative to SPPS of the corresponding all-amide peptide.

Thioamide Effects in Full Length Proteins. With the possibility of incorporating thioamides as spectroscopic probes into full length proteins[73], the question of their influence on protein stability has emerged. Because of their increased size and altered hydrogen bonding properties, they can stabilize or destabilize protein folds. While several older studies had focused on the effects of thioamide on the stability of peptide models of *α*-helical and *β*-sheet secondary structure[67,122], two recent publications have investigated their effects in full-sized proteins[71,123]. First, Raines and co-workers investigated the influence of thioamides on the thermal stability of collagen model peptides and identified

thioamides as the first backbone modification that does not compromise the thermal stability of these peptides. A separate team of investigators conducted a more comprehensive investigation, examining thioamide substitutions in three different protein model systems each representing a different class of secondary structure[71]. Taken together, the two studies indicate that thioamides can have a slight stabilizing effect in *α*-helices and polyproline type II (PPII) helices, if the additional steric bulk of the thiocarbonyl is accommodated and the thioamide acts primarily as a hydrogen bond donor. However, in a sterically crowded environment, thioamides can be destabilizing. All tested (internal) *β*-sheet positions, as well as some positions in a PPII helix, had a significant negative impact on protein stability (see Chapter 5). Destabilizing effects in *α*-helices, however, were comparatively mild, and some of the destabilizing effects in αhelices can be "rescued" with the incorporation of a second thioamide[78]. While Raines and others have shown that thioamides can exert stabilizing effects in model peptide systems by increasing the strength of n $\rightarrow \pi^*$ interactions between *i* and *i* + 1 carbonyls, these have not obviously contributed to the stabilizing effects observed to date in protein systems[72,124,125]. However, many of the observed effects cannot fully be rationalized using crystal structures of the native proteins. Determining high-resolution structures of the thioamide proteins themselves will benefit the field substantially. Combining structural information with biophysical studies should help to provide guidelines for thioamide placement beyond simple rules of avoiding steric clashes of the longer thiocarbonyl and taking advantage of stronger thioamide hydrogen bond donation.

Synopsis

Thioamides are a unique modification of canonical amide bonds. Although they only differ by a single Oxygen to Sulfur substitution, their physicochemical properties are

dramatically different from regular amides. Thioamides have been observed in various natural products with potential therapeutic properties. Over the last few decades, synthetic chemists and peptide chemists have exploited the unique properties of thioamides for various different purposes. However, their use in proteins is still very limited. The Petersson laboratory has created some of the first (semi-)synthetic thioamide labeled proteins (thioproteins). In this thesis, I will describe my efforts towards improving various aspects of thioprotein semi-synthesis and characterization. First, I will demonstrate how optimization of thiopeptide synthesis can increase isolated yield. Next, I will show that moving the thioamide motif to the side-chain of certain amino acids is synthetically feasible and that side-chain thioamides retain their fluorescence quenching ability. Furthermore, I will demonstrate how traceless ligation methods can be modified to be compatible with thiopeptides and thioproteins. This will be followed by an analysis of thioamides in different secondary protein structures and the influence of thioamides on protein stability. I will then demonstrate a novel semi-synthetic scheme for thioproteins that allows us to obtain mg quantities of protein, needed for high-resolution structural analysis. I will finish by describing my attempt of genetically incorporating thioamides into proteins. Together, these studies will aid the effort of developing thioamides as a probe that cannot only be used in peptides, but also in proteins.

Chapter 2 : Improved Fmoc Deprotection Methods for the Synthesis of Thioamide-Containing Peptides

Reprinted and adapted with permission from: Szantai-Kis, D. M.; Walters, C. R; Barrett, T. M.; Hoang, E. M.; Petersson, E. J., Thieme Chemistry Journals Awardees – Where Are They Now? Improved Fmoc Deprotection Methods for the Synthesis of Thioamide-Containing Peptides and Proteins. *Synlett* **2017**, *28*, 1789-1794. Copyright 2017 Georg Thieme Verlag Stuttgart New York[77].

2.1 Introduction

Thioamides have traditionally been incorporated using adaptations of standard fluorenylmethyloxycarbonyl (Fmoc)-based solid phase peptide synthesis (SPPS) conditions, wherein monomers are introduced as pre-activated thioacyl compounds and deprotections are accomplished using 20% (v/v) piperidine in dimethylformamide (DMF) for Fmoc deprotections. However, we observed two different side-reactions while using this strategy for Fmoc deprotections. One of the side-reactions is epimerization at the α-carbon of the thioamino acid. Prolonged exposure to base during Fmoc deprotection can catalyze this epimerization reaction. The other undesired reaction observed during SPPS is the formation of a covalent adduct when incorporating N_ε-thioacetyl lysine, or Lys(Ac^S). While we only observed this adduct during syntheses involving Lys(Ac^S), we hypothesize that transient formation of this adduct with backbone thioamides may also lead to S-to-O exchange of the thioamide in the presence of water. Both of these side-reactions are detrimental to the yields of our thioamide peptides, and we therefore sought to understand their origins and eliminate them.

We primarily attribute epimerization at the α-carbon to the lowered pK_a of the α-C-H bond. While there have been only a limited number of measurements of the pK_as of protons in thioamide bonds and peptides, a survey of small molecule and peptide literature values for N-H and α-C-H pK_a values allows us to make reasonable estimates of the corresponding values in thioamide containing peptides (see Figure 2-1; all values in DMSO unless otherwise notes)[126–129]. In general, the α-C-H of thioamides is about 5 pH units lower than for the corresponding oxoamides and therefore the α-C-H pK_a of thioamides in peptides can be estimated to be around 12-13. The pK_a for the acidic form

of the bases used in SPPS is usually around the same value or higher, making base-catalyzed epimerization a reasonable pathway for the observed racemization.

While the formation of piperidine adducts during thioamide SPPS had not previously been reported, studies from Boger and coworkers had established the susceptibility of thioamides to reactions with amine nucleophiles[8]. Even in the absence of the silver catalysts used by Boger, it is not surprising that amidine adducts can form when using high concentrations of piperidine.

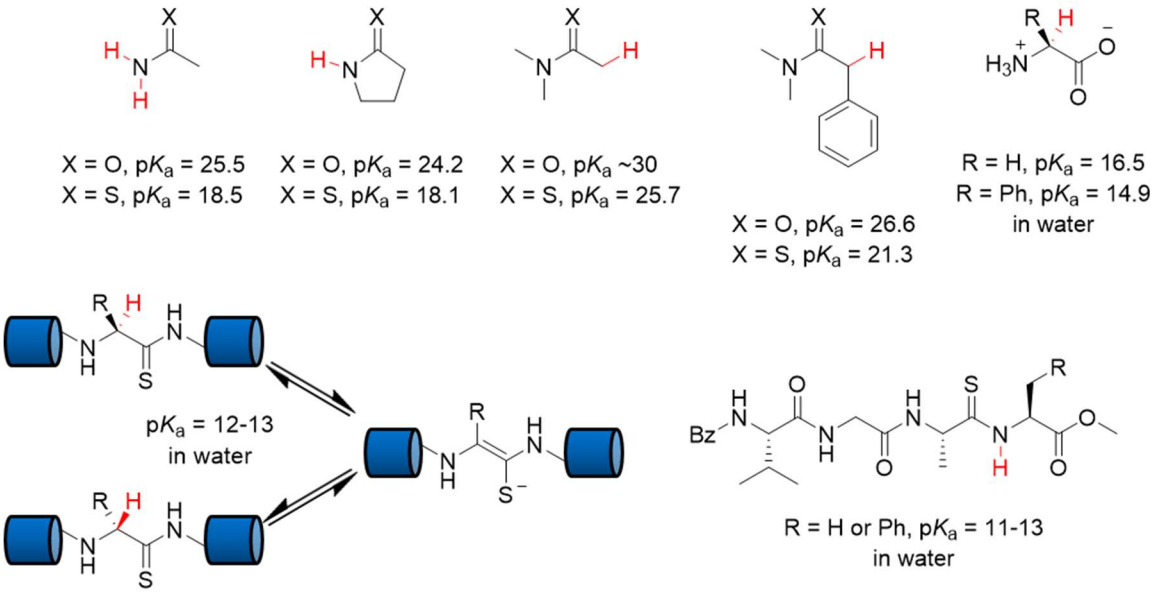


Figure 2-1: Reported pK_a values that allows estimation of α-C-H pK_a in thioamides

Previously, it was reported that lowering the concentration of piperidine and shortening Fmoc deprotection times to one minute resulted in less epimerization of thioamides[70]. However, since the half-life of the Fmoc group under these conditions is 20–45 seconds, this approach significantly decreases the yield of isolated protein[130]. We sought to test alternative deprotection solutions to reduce both observed side reactions. Other commonly used Fmoc deprotections solutions are 50% (v/v) Morpholine in DMF

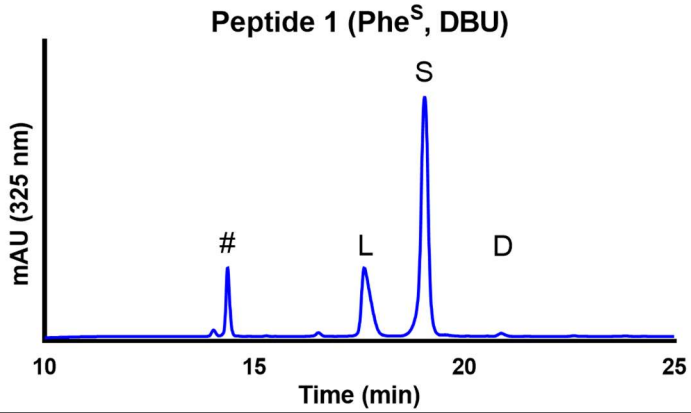
and 2% (v/v) 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU). Unlike piperidine and morpholine, DBU is a non-nucleophilic base and is not able to scavenge the reactive dibenzofulvene group resulting from the Fmoc deprotection. Therefore, either a nucleophilie needs to be added as a scavenger, or reaction times must be kept short and the resin washed extensively after deprotection[131]. The non-nucleophilic nature of DBU also reduces its likelihood of forming amidines by reaction with thioamides. Therefore, we evaluated the use of DBU for deprotection of Fmoc groups in thioamide-containing peptides to address both epimerization and adduct formation issues observed with piperidine.

2.2 Results and Discussion

We designed a set of model peptides containing a mixture of nonpolar (Phe and Ala) and polar (Lys and Glu) amino acids. For each sequence, 7-methoxycoumarin-4-ylalanine (Mcm or µ) was installed as the first amino acid on 2-chlorotrityl resin. This chromophore has an absorption maximum at 325 nm and allows for facile tracking of the peptides by analytical HPLC. Thioamides also have a shifted absorption maximum (272 nm) relative to oxoamides (215 nm), which can be used to characterize the thiopeptides. To discriminate between genuine L-thiaomino acid peptides and their D-epimers, we synthesized both L and D-thionitrobenzotriazolide monomers (Fmoc-ala^{S}-NBt and Fmoc-phe^{S}-NBt, using the lower-case letter convention for D-amino acids) for inclusion into the model peptides. The retention times for the authentic D-thioamino acid peptides were used as standards for each of the first four sets of thiopeptides. Additionally, a separate assay was performed to assess how each condition affected yield. In these experiments, 7-methoxycoumarin-4-acetic acid (Mca) was added to each peptide sample to a final concentration of 500 µM, representing the maximum theoretical yield of peptide in each

HPLC sample. The integration value of this standard peak was compared to the integration of the product peak in each chromatogram (at 325 nm) to calculate the yield. These experiments, including synthesis and characterization of peptides and small molecules were carried out by C. R. Walters and T. M. Barrett and a detailed description can be found elsewhere[77]. The results are summarized in Table 2.1 and show that the use of DBU reduced epimerization in all cases and therefore increased overall purity and yield of the corresponding peptide.

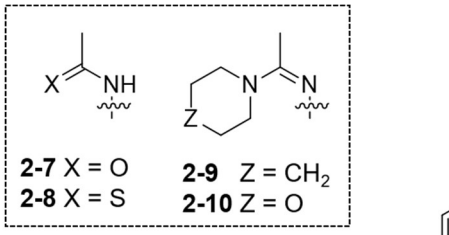
Table 2.1: Comparison of Yield and Epimerization between Peptides Synthesized with either DBU or Piperidine[a].

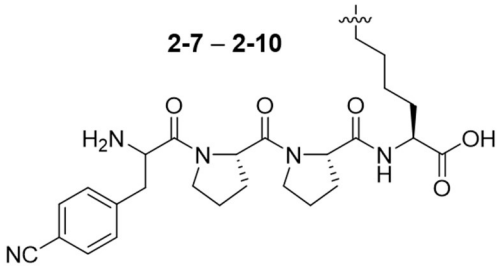


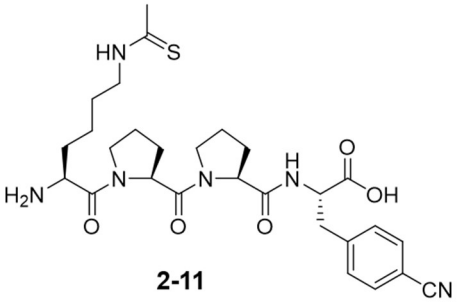
Peptide[b]	Base used	Epimerization (%)	Purity (%)[c]	Yield (%)
KAF^{S}AKμ (**2-1**)	Piperidine	16.1	56.2	29.9
	DBU	5.0	61.1	35.9
F^{S}AKAKμ (**2-2**)	Piperidine	1.0	89.0	44.3
	DBU	0.0	88.5	51.3
FA^{S}KAKμ (**2-3**)	Piperidine	3.2	68.5	44.0
	DBU	1.0	67.3	44.7
FAE^{S}AKμ (**2-4**)	Piperidine	5.6	71.5	18.2
	DBU	1.0	71.6	22.7
CVNY^{S}EEFVQMMTAK (**2-5**)	Piperidine	49.9	8.6	n/d[d]
	DBU	9.3	12.3	n/d[d]
CVNYEEF^{S}VQMMTAK (**2-6**)	Piperidine	28.1	25.0	n/d[d]
	DBU	4.5[e]	22.0[e]	n/d[d]

[a] Graph: Example chromatogram for epimer quantification and yield determination. L and D denote epimers respectively, # denotes KAAKμ (product of failed F^S coupling), S denotes Mca standard used for yield quantification
[b] Peptides in one-letter code. Superscript S denotes thioamino acids; μ = 7-methoxycoumarin-4-yl-alanine
[c] Purity calculations based on HPLC peak are of desired product.
[d] No external standard added for yield quantification.
[e] Average of two experiments.

We are also interested in incorporating sidechain thioamides as probes for fluorescence quenching studies or protein interactions and as mimics of post-translational modifications. Thus, we attempted to synthesize peptide **2-8** (Figure 2-2) containing the sidechain thioamide Lys(Ac^S). However, during this synthesis, we encountered a +51 mass product. This product was not observed while synthesizing the corresponding oxopeptide **2-7** (Figure 2-2), which contained N_ε-acetyl-lysine, or Lys(Ac). This implied that the modification must have been part of the thioamide motif. To investigate this, we tested several different cleavage solutions to cleave the peptide from the resin (See Table 2.2). However, all tested cleavage conditions still resulted in the +51 mass adduct, leading us to believe that this side product was not formed during peptide cleavage.



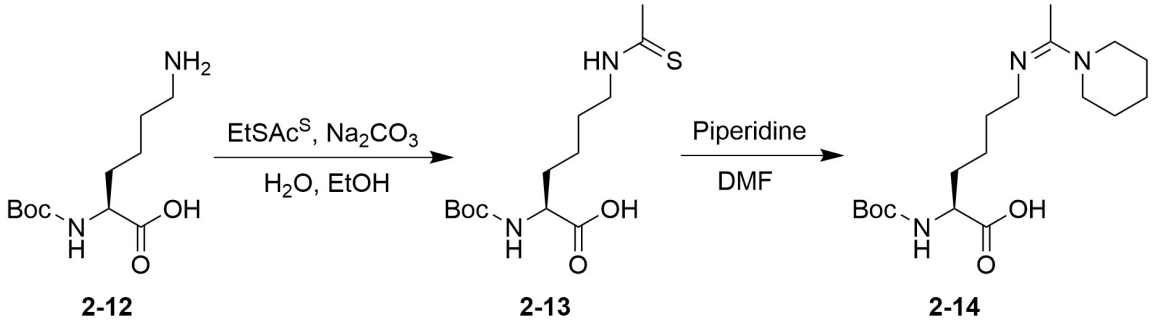




Species	m/z	
$[M+H]^+$ or $[M]^+$	Calculated	observed
2-7	571.3	571.4
2-8	555.3	555.4
2-8 + Pip. (**2-9**)	622.4	622.5
2-8 + Mor. (**2-10**)	624.4	624.7

Figure 2-2: Lys(Ac^S)-containing peptides used to study byproduct formation under piperidine deprotection conditions. MALDI-TOF MS masses of 2-7, 2-8 (using DBU deprotection), 2-9 (byproduct of 2-8 using piperidine deprotection), and 2-10 (byproduct of 2-8 using morpholine deprotection); Pip. = piperidine, Mor. = morpholine.

We hypothesized that the +51 mass corresponded to piperidine adduct **2-9**. As noted above, thioamides have been used previously to prepare amidines, as in Boger's syntheses of vancomycin derivatives[8]. This was confirmed with several experiments. We synthesized peptide **2-8** using either 50% (v/v) morpholine in DMF or 2% (v/v) DBU in DMF as deprotection solutions. We observed a +53 mass when Fmoc-deprotections were performed with morpholine, which we believe corresponds to peptide **2-10**. No side-product was observed in the peptide synthesized with DBU. We also synthesized peptide **2-11** (Figure 2-2), where N_{α}-Boc-N_{ε}-thioacetyl-lysine, or Boc-Lys(Ac^{S}), was installed at the N-terminus of the peptide and no +51 side product was observed after cleavage (TFA only to cleave Boc group, no base required).



2-12 **2-13** **2-14**

Figure 2-3: Synthesis of the piperidine adduct Boc-Lys(Pip)-OH (2-14).

To further verify adduct formation, we dissolved Boc-Lys(Ac^{S}) (**2-13**) in 50% (v/v) piperidine in DMF and isolated lysyl amidine (**2-14**) by HPLC (Figure 2-3). NMR, UV/VIS spectroscopy, and high-resolution mass spectrometry confirmed the identity of (**2-14**). Specifically, we observed cross peaks in a nuclear Overhauser effect spectroscopy (NOESY) experiment between the CH_3 group on the side-chain with the piperidinyl-ring as well as the protons attached to the δ-Carbon of the Lys sidechain. Additionally, we

observed correlations of that same methyl group with two nitrogen atoms using proton-nitrogen heteronuclear multi-bond correlation (HMBC) spectroscopy. An illustration of the observed correlations is shown in Figure 2-4 (only clearly distinguishable signals shown; only one set off 1H-13C HMBC correlations on Boc group shown to reduce complexity). These amino acid and peptide studies confirm the necessity of using DBU in Fmoc-deprotections when synthesizing Lys(Ac^{S}) containing peptides.

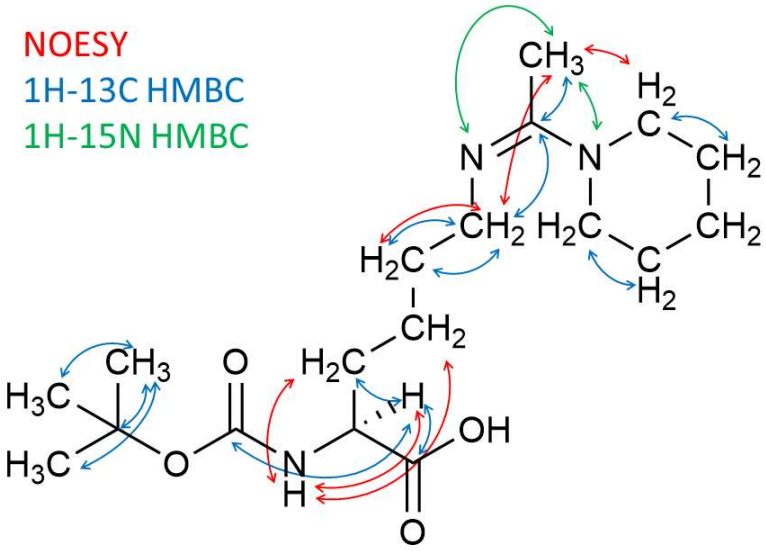


Figure 2-4: 2D correlations from NMR experiments on Boc-Lys(Pip)-OH (2-14).

2.3 Conclusion

With the preparation of each backbone thiopeptide system, we have demonstrated reduction of epimerization at the thioamino acid α-carbon by performing Fmoc deprotection reactions with DBU in the place of piperidine. Although only four different amino acids were tested (Ala^{S}, Phe^{S}, Glu^{S}, and Tyr^{S}), the positional and sequence variability with which this effect is demonstrated provides strong evidence that this result is general. Furthermore, we demonstrated that in at least one sidechain thioamide case,

the use of DBU as the deprotection reagent was superior to piperidine or morpholine. While we have not yet observed a piperidine adduct in a backbone thioamide, it may occur at sterically-accessible sites such as Gly^S residues. Such adducts would presumably not be stable, but could lead to Edman-type degradation, as well as hydrolysis of the peptide backbone or oxoamide formation by reaction with water under the acidic resin cleavage conditions. All of these side reactions would be avoided by using DBU for Fmoc-deprotections during peptide synthesis. We expect that these synthetic developments will help to improve the accessibility of thioamide containing peptides and proteins, further enabling the broad spectrum of biophysical and biochemistry experiments made possible by this versatile functional group.

2.4 Materials and Methods

General Information. N^{α}-Fmoc-4-cyanophenylalanine was purchased from PepTech (Burlington, MA, USA). N^{α}-Fmoc-7-methoxycoumarin-4-yl-alanine (Mcm) was purchased from Bachem (Basel, Switzerland). Piperidine was purchased from AmericanBio (Natick, MA). Triisopropylsilane (TIPS) was purchased from Santa Cruz Biotechnology (Dallas, TX, USA). All other Fmoc protected amino acids and peptide synthesis reagents were purchased from EMD Millipore (Billerica, MA, USA). All other reagents and solvents were purchased from Fisher Scientific (Waltham, MA, USA) or Sigma-Aldrich (St. Louis, MO, USA) and used without further purification unless otherwise specified.

High resolution electrospray ionization mass spectra (ESI-HRMS) were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA, USA). Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD)

mass spectrometer. UV-Vis absorption spectra were acquired on a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III instrument. Analytical HPLC was performed on an Agilent 1100 Series HPLC system. Preparative HPLC was performed on a Varian Prostar HPLC system (currently Agilent Technologies). HPLC columns were purchased from W. R. Grace & Company (Columbia, MD, USA) or Phenomenex (Torrence, CA, USA).

Synthesis of N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^6-ethanethioyl-L-lysine (Fmoc-Lys(Ac^S)-OH) Synthesis based on published procedure[132]. Briefly, Fmoc-Lys-OH (368 mg, 1.00 mmol, 1.0 equiv) was suspended in 2.2 mL EtOH and 2.0 mL 10% (w/v) Na_2CO_3 solution added. Ethyl-dithioacetate (126 µL, 1.10 mmol, 1.1 equiv) added and stirred overnight. Next day the solvent was removed in vacuo and redissolved in 10 mL H_2O. Reaction mixture was acidified with 3M HCl until solution became milky white (~pH 2). Aqueous phase extracted 3 times with 10 mL $CHCl_3$ each. The combined organic phase was dried over Na_2SO_4 and the solvent was removed in vacuo. The product yielded as off-white foam in high yield (423 mg, 0.99 mmol, 99.3%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.60 (s, 1H), 7.97 (s, 1H), 7.71 (d, *J* = 7.6 Hz, 2H), 7.53 (t, *J* = 8.1 Hz, 2H), 7.35 (td, *J* = 7.5, 2.9 Hz, 2H), 7.25 (td, *J* = 6.5, 5.7, 3.1 Hz, 2H), 5.77 (d, *J* = 8.2 Hz, 1H), 4.38 – 4.23 (m, 3H), 4.14 (t, *J* = 7.1 Hz, 1H), 3.60 – 3.46 (m, 2H), 2.45 (s, 3H), 1.84 (d, *J* = 10.5 Hz, 1H), 1.73 – 1.52 (m, 3H), 1.46 – 1.34 (m, 2H). ESI^+-HRMS: calculated for $C_{23}H_{27}N_2O_4S^+$: 427.1692; found $[M + H]^+$: 427.1705.

Synthesis of N^2-(*tert*-butoxycarbonyl)-N^6-ethanethioyl-L-lysine (Boc-Lys(Ac^S)-OH) (2-13) Synthesis was performed the same way as for the corresponding Fmoc protected compound. Briefly, Boc-Lys-OH **2-12** (369 mg, 1.50 mmol, 1.0 equiv) was suspended in 4.4 mL EtOH and 4.0 mL 10% (w/v) Na_2CO_3 solution added. Ethyl-dithioacetate (189 µL, 1.65 mmol, 1.1 equiv) added and stirred overnight. Next day the solvent was removed in vacuo and redissolved in 10 mL H_2O. Reaction mixture was acidified with 3M HCl until solution became milky white (~ pH 2). Aqueous phase extracted 3 times with 10 mL $CHCl_3$ each. The combined organic phase was dried over Na_2SO_4 and the solvent was removed in vacuo. The product yielded as yellow foam in high yield (423 mg, 1.39 mmol, 92.4%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.88 (s, 1H), 8.37 (s, 1H), 5.35 (d, *J* = 7.7 Hz, 1H), 4.07 (d, *J* = 75.4 Hz, 1H), 3.52 (s, 2H), 2.50 – 2.34 (m, 3H), 1.77 (s, 1H), 1.69 – 1.49 (m, 3H), 1.42 – 1.23 (m, 11H). ESI^+-HRMS: calculated for $C13_3H_{25}N_2O_4S^+$: 305.1535; found $[M + H]^+$: 305.1556.

Synthesis of Piperidine adduct (2-14) Boc-Lys(Ac^S)-OH **2-13** (133 mg, 0.438 mmols) was dissolved in 2 mL of 50% (v/v) piperidine in DMF. After stirring for 5 hours, reaction mixture was diluted with 0.1% TFA in H2O and purified by reverse phase HPLC using gradient **2A** on a Grace Vydac C18 Prep column with a flow rate of 15 mL/min (retention times for the product 2-14 and starting material **2-13** are 23.8 min and 27.8 min respectively). Fractions containing product **2-14** or unreacted starting material **2-13** were collected separately and lyophilized. After lyophilization starting material **2-13** was dissolved in 50% (v/v) piperidine in DMF and, after 5 hours the reaction was purified as before. This procedure was repeated one more time until enough product was collected for NMR analysis (1.90 mg, 4.05 µmols, 0.9 %). ^{1}H NMR (500 MHz, DMSO-d_6) δ 12.58 (s, 1H), 8.68 (s, 1H), 7.05 (d, *J* = 7.8 Hz, 1H), 3.84 (td, *J* = 8.8, 4.6 Hz, 1H), 3.57 (s, 4H), 3.32

(s, 2H), 2.28 (s, 3H), 1.67 – 1.54 (m, 8H), 1.49 (dt, *J* = 13.3, 7.0 Hz, 2H), 1.37 (s, 9H), 1.36 – 1.28 (m, 2H). ^{13}C NMR (126 MHz, DMSO) δ 174.28, 162.44, 155.65, 78.02, 53.36(+), 49.45(-), 46.48(-), 43.89(-), 30.38(-), 28.86(-), 28.25(+), 25.64(-), 24.74(-), 23.06(-), 22.65(-), 14.35(+). ESI$^+$-HRMS: calculated for $C_{23}H_{27}N_2O_4S^+$: 356.2549; found $[M+H]^+$: 356.2558. Additional 2D NMR correlations shown in Figure 2-4(only clearly distinguishable signals shown; only one set off 1H-13C HMBC correlations on Boc group shown to reduce complexity).

Synthesis of (thio)acetyl lysine containing peptides Peptides **2-7** – **2-11** were synthesized on 2-chlorotrityl resin (100-200 mesh, 10 μmol scale) in 3 mL fritted syringes. The resin was swelled in DMF for 20 minutes. All amino acids (5 equiv.) were dissolved in 2 mL DMF along with HBTU (5 equiv.) and DIPEA (10 equiv.), except for the first amino acid of each peptide, for which HBTU was omitted. Amino acid couplings were carried out as single couplings for 30 minutes. Fmoc deprotections were carried out with either 2% (v/v) DBU in DMF (3 x 2 min), 20% (v/v) piperidine in DMF (2 x 10 min) or 50% (v/v) morpholine in DMF (2 x 10 min). After complete synthesis, the resin was washed with DCM and dried under vacuum before being cleaved off. The standard cleavage solution contained DCM/TFA/TIPS/H_2O (10:8:1:1), although other cleavage conditions were tested as well (see Table 2.2). After cleavage for 45 minutes, the solvent was removed *in vacuo*. The crude peptide was dissolved in 1 mL of 50% (v/v) MeCN in H_2O and either analyzed directly by MALDI-MS or purified by HPLC.

Table 2.2: Tested cleavage conditions with peptide 2-8.

Entry	Cleavage Condition
1	AcOH / TFE / DCM (1:1:8)
2	AcOH / TFE / DCM (2:2:6)
3	0.5% (v/v) TFA in DCM
4	1% (v/v) TFA + 5% (v/v) TIPS in DCM
5	5% (v/v) TFA + 5% (v/v) TIPS in DCM

AcOH = acetic acid; TFE = Trifluoroethanol; DCM = Dichloromethane; TFA = Trifluoroacetic acid; TIPS = Triisopropylsilane

Table 2.3: HPLC purification methods and retention times.

Peptide/Compound	Gradient	Retention Time	Column
Boc-Lys(Ac^S)-OH (**2-13**)	2A	27.8 min	Vydac C18 Prep
Boc-Lys(Pip)-OH (**2-14**)	2A	23.8 min	Vydac C18 Prep
F*PPK(Ac)-OH (**2-7**)	2B	18.9 min	Vydac C18 Semiprep
F*PPK(AcS)-OH (**2-8**)	2C	18.1 min	Vydac C18 Semiprep

F* = 4-Cyanophenylalanine; Lys(Pip) = Piperidine adduct

Table 2.4: HPLC gradients used for small molecule/peptide purification.

No.	Time (min)	%B	No.	Time (min)	%B	No.	Time (min)	%B
2A	0:00	5	**2B**	0:00	2	**2C**	0:00	2
	5:00	5		3:00	2		3:00	2
	10:00	20		10:00	12		10:00	18
	31:00	26		25:00	22		25:00	28
	32:00	100		28:00	100		28:00	100
	35:00	100		31:00	100		31:00	100
	37:00	5		33:00	2		33:00	2

Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

Table 2.5: MALDI-TOF MS Characterization of Purified Peptides.

Peptide	$[M+H]^+$	
	Calculated	**Found**
F*PPK(Ac)-OH (**2-7**)	555.29	555.41
F*PPK(AcS)-OH (**2-8**)	571.27	571.39
F*PPK(Pip)-OH (**2-9**)	622.37 ($[M]^+$)	622.50 ($[M]^+$)
F*PPK(Mor)-OH (**2-10**)	624.35 ($[M]^+$)	624.69 ($[M]^+$)
K(AcS)PPF*-OH (**2-11**)	571.27	571.59

F* = 4-Cyanophenylalanine; K(Pip) = Piperidine adduct; K(Mor) = Morpholine adduct

Chapter 3 : Synthesis, Characterization, and Incorporation of Side-chain Thioamide derivatives Thioasparagine, Thioglutamine, and Thioacetyl Lysine

3.1 Introduction

While substitution of amide linkages in backbones has been pursued for decades[90], the incorporation of thioamides into side-chains of amino acids is a newer endeavor. Synthesis of side-chain bearing thioamides is synthetically more challenging and wasn't reported until 2000[133]. One could exploit the natural amide-containing amino acids asparagine or glutamine as starting point to create side-chain thioamide versions of asparagine ($Asn^{\gamma S}$) or glutamine ($Gln^{\delta S}$). Alternatively, the synthesis of *N*$^{\varepsilon}$-thioacetyl lysine ($Lys(Ac^{S})$) as a thioamide version of a posttranslationally modified amino acid would be possible[132]. These side-chain thioamides could be used in cases where backbone thioamides have been been shown to be disruptive[71].

The synthetic incorporation of $Lys(Ac^{S})$ into peptides using solid phase peptide synthesis (SPPS) and its effect on deacetylation have been described previously in the literature[132]. It has been shown that $Lys(Ac^{S})$ can act as non-hydrolyzable analogue of N$^{\varepsilon}$-acetyl lysine (Lys(Ac)) for Sirtuins, which represent the class III or NAD-dependent histone deacetylase family of proteins. The synthesis and incorporation of the other two side-chain thioamides is more challenging. Fmoc-protected precursors of $Asn^{\gamma S}$ for SPPS have been synthesized previously, but not successfully incorporated into peptides[133]. Synthesis of Fmoc-protected $Gln^{\delta S}$ has been attempted but was not successful.[133]

Here we investigate whether the fluorescence quenching behavior observed in backbone thioamides would be similar with the side-chain thioamide Lys(Ac^S). Furthermore, we report the first successful incorporation of $Asn^{\gamma S}$ into peptides and the first successful synthesis of Fmoc-protected precursors of $Gln^{\delta S}$.

3.2 Results and Discussion

We wanted to confirm that the fluorescence quenching behavior observed in backbone thioamides would hold up with the side-chain thioamide Lys(Ac^S). We synthesized Fmoc-Lys(Ac^S)-OH as previously described[132]. We synthesized rigid polyproline ruler peptides, where Lys(Ac^S) was incorporated on the C-terminus and a fluorophore was incorporated on the N-terminus. Based on the different quenching mechanisms of thioamides, we chose three different fluorophores: First, we chose *p*-Cyanophenylalanine (pCNF) because it has a significant spectral overlap with thioamides and is mostly quenched through Förster Resonance Energy Transfer (FRET), a non-radiative form of energy transfer. Our second fluorophore of choice was Tryptophan, since it has some spectral overlap, but can also be quenched through photoinduced electron transfer (PET). Lastly, we also incorporated (7-Methoxymcoumarinyl-)2-alanine (Mcm), since it has no spectral overlap with thioamides and all quenching is based upon PET.

We synthesized our peptides mostly using standard solid-phase peptide synthesis (SPPS) conditions. One of the crucial differences was the use of 1,8-Diazabicyclo[5.4.0]-undec-7-ene as base during Fmoc deprotections to suppress unwanted byproducts[77]. For all peptides we synthesized a thio version, containing Lys(Ac^S), as well as an oxo version containing Lys(Ac). We measured fluorescence spectra of the various peptides and were able to observe distance dependent quenching (see Figure 3-1). The strongest quenching can be observed for Trp, which can be quenched through FRET and PET based

mechanisms. Closer analysis of the data along with computational modelling are currently underway.

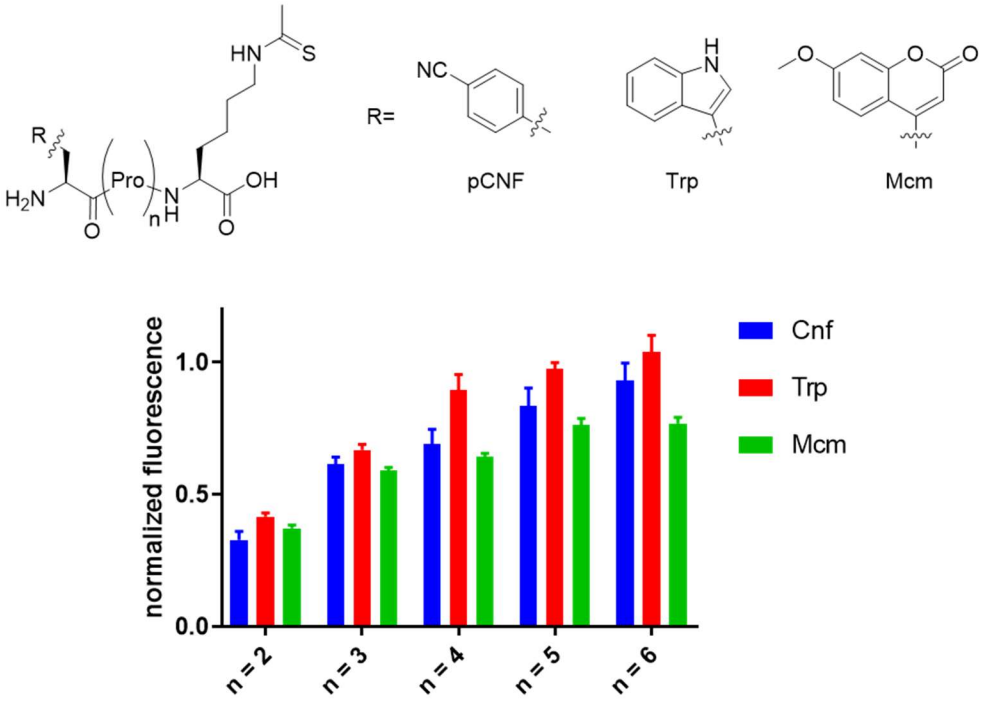


Figure 3-1: Fluorescence quenching by side-chain N$^{\varepsilon}$-thioacetyl lysine. Top: Structure of polyproline rulers with various fluorophores. Bottom: Quenching efficiency of peptides with various fluorophores (normalized to corresponding oxo peptide). The number of proline residues between fluorophore and quencher are indicated by n.

Furthermore, we wanted to synthesize Fmoc-protected thioamide precursors of Asn$^{\gamma S}$ and Gln$^{\delta S}$. We followed the synthesis described by Findlay *et al.*[133] to synthesize xanthyl protected Fmoc-Asn$^{\gamma S}$(Xan)-OH. Similar to their observation, we were not able to synthesize the corresponding Fmoc-Gln$^{\delta S}$(Xan)-OH, which we attribute to decomposition during xanthyl protection. Unlinke Findlay *et al.*, we were actually successful in incorporating Asn$^{\gamma S}$ into peptides. While they were able to demonstrate that the amino acid was incorporated into the peptide on resin, they were unable to isolate the resin using standard cleavage conditions with 95% trifluoroacetic acid (TFA). We have previously

observed with backbone thioamides that reducing the TFA concentration to 50% or lower and reducing cleavage time to <1 hour can significantly increase the isolated yield of thioamide containing proteins. We cleaved our peptides using lower TFA concentrations and were able to cleave thioamide containing peptides off of resin.

We wanted to demonstrate that these peptides are still capable of quenching fluorescence. We synthesized a small series of peptides similar to the ones above, where Cnf was the N-terminal residue and Asn or $Asn^{\gamma S}$ was the C-terminal residue. The two residues were separated by 2, 4, or 6 Pro residues. Fluorescence measurements of concentration matched peptides showed that $Asn^{\gamma S}$ is indeed capable of quenching fluorescence in a distance dependent manner that tracks well with $Lys(Ac^{S})$ (see Figure 3-2). Further analysis of the data along with computational modeling are currently underway.

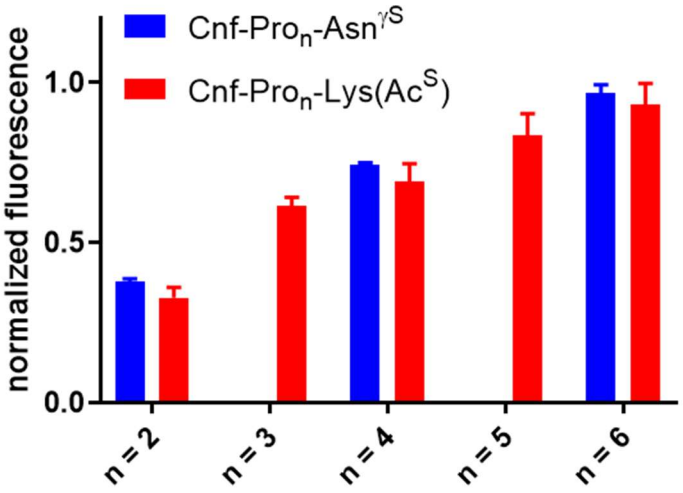


Figure 3-2: Distance dependent fluorescence quenching of $Asn^{\gamma S}$ is in good agreement with data obtained from $Lys(Ac^{S})$ measurements. Fluorescence intensity is normalized to corresponding oxo peptide. Peptides with n=3 and 5 were not synthesized for $Asn^{\gamma S}$. The number of proline residues between fluorophore and quencher are indicated by n.

Lastly, we were wondering if Fmoc-protected $Gln^{\delta S}$ might be obtained through direct thionation with phosphorous pentasulfide (P_4S_{10}) of commercially available derivatives as this would drastically reduce the number of steps needed to synthesize precursors. We hypothesized that the size of the side-chain protecting group might have an influence of thionation as large protecting groups might sterically obstruct the thionation site. We set out to attempt direct thionation of five differently protected Gln derivatives. Small scale reactions were set up and analyzed after 20 hours. As shown in Figure 3-3, direct thionation was successful with all tested derivatives. While 2,4,6-trimethoxybenzyl- (Tmob) and 4,4'-dimethyoxybenzhydryl-protected (Mbh) Fmoc-Gln showed the highest rate of conversion, they also showed additional peaks that had masses corresponding to species with multiple thionations. Since the presence of methoxy groups is the main chemical difference, we hypothesize that these groups can undergo thioetherification upon prolonged exposure of P_4S_{10}. While the presence of the thioether should not affect coupling or cleavage if one were to use these compounds, careful characterization by NMR spectroscopy is necessary to ensure that the desired thioamide moiety is present, as it cannot be determined by HPLC in combination with MS.

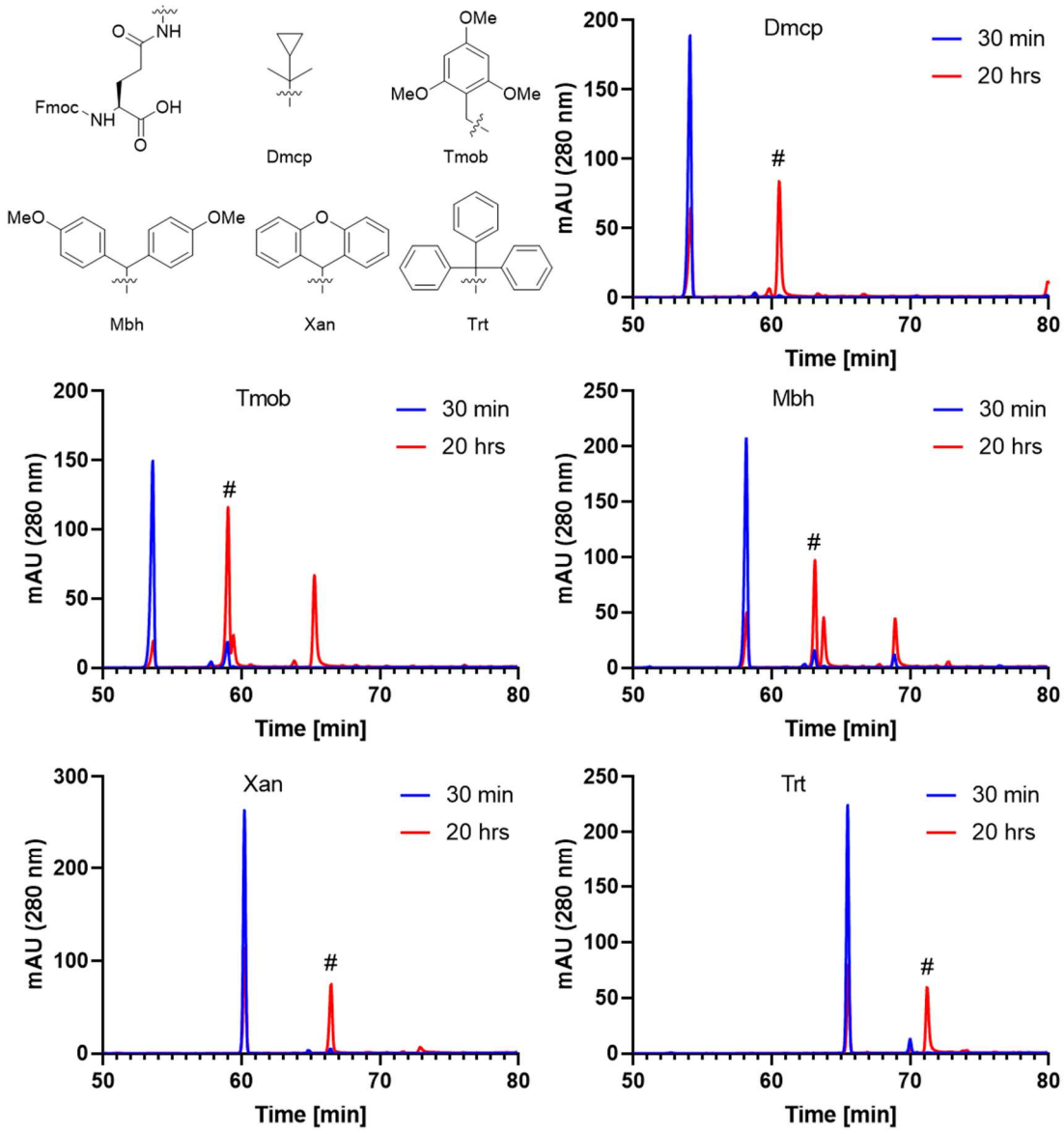


Figure 3-3: Direct thionation of Fmoc-Gln with various side-chain protecting groups. Analytical HPLC traces shown. # indicates peak with monothionated mass. Abbreviations: Dmcp = Dimethylcyclopropyl; Tmob = 2,4,6-Trimethoxybenzyl; Mbh = 4,4'-dimethyoxybenzhydryl; Xan = Xanthyl; Trt = Trityl.

Based on these results, we decided to proceed and set up a larger scale synthesis with a trityl (Trt) side-chain protecting group. We chose Trt because it had similar clean conversion as xanthyl- (Xan) protected Gln, but had better solubility. While the isolated yield was rather low, it is the first successful synthesis of a Fmoc protected precursor for

$Gln^{\delta S}$. Attempts to incorporate $Gln^{\delta S}$ into peptides will be performed shortly. Furthermore, we are currently trying to apply the same direct thionation to Asn derivatives with various side-chain protecting groups.

3.3 Conclusion

We wanted to see if the fluorescence quenching behavior observed in backbone thioamides would hold up with the side-chain thioamide $Lys(Ac^S)$. We were able to demonstrate that $Lys(Ac^S)$ is indeed able to quench fluorescence of various fluorophores in a distance dependent manner.

Previous attempts at precursor synthesis and peptide incorporation of $Asn^{\gamma S}$ and $Gln^{\delta S}$.have failed at various stages: A precursors for SPPS of $Asn^{\gamma S}$ was previously synthesized in 4 steps, but incorporation into peptides ultimately failed, while synthesis of a $Gln^{\delta S}$ SPPS precursor failed. We successfully repeated the reported synthesis of the $Asn^{\gamma S}$ precursor. We were successful in incorporating $Asn^{\gamma S}$ into peptides and demonstrate that similar to $Lys(Ac^S)$ it is able to quench Cnf fluorescence. Furthermore, we tested direct thionation of Gln with various side-chain protecting groups. We were able to show that all derivatives can get thionated. The scale-up of one of the Gln derivatives yielded the first isolated $Gln^{\delta S}$ precursor for SPPS. Attempts for direct thionation of Asn and incorporation of $Gln^{\delta S}$ into peptides are currently underway.

3.4 Materials and Methods

General Information. N^{α}-Fmoc-4-cyanophenylalanine was purchased from PepTech (Burlington, MA, USA). N^{α}-Fmoc-7-methoxycoumarin-4-yl-alanine (Mcm) was purchased from Bachem (Basel, Switzerland). Piperidine was purchased from AmericanBio (Natick, MA). Triisopropylsilane (TIPS) was purchased from Santa Cruz

Biotechnology (Dallas, TX, USA). Fmoc-Lys(Ac^S)-OH[132] and Fmoc-$Asn^{\gamma S}$(Xan)-OH[133] were synthesized as previously described. was All other Fmoc protected amino acids and peptide synthesis reagents were purchased from EMD Millipore (Billerica, MA, USA). All other reagents and solvents were purchased from Fisher Scientific (Pittsburgh, PA, USA) or Sigma-Aldrich (St. Louis, MO, USA) and used without further purification unless otherwise specified.

High resolution electrospray ionization mass spectra (ESI-HRMS) were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA, USA). Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. UV-Vis absorption spectra were acquired on a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Fluorescence data was acquired on a Photon Technologies International (PTI) QuantaMaster40 fluorometer (currently Horiba Scientific, Edison, NJ, USA). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III or Microflex LRF instrument. Reverse-phase purification of small molecules was performed on a Biotage Isolera System on Biotage SNAP Ultra C18 columns (Charlotte, NC, USA). Analytical HPLC was performed on an Agilent 1260 Infinity II series UHPLC system (Agilent, Santa Clara, CA, USA). Preparative HPLC was performed on an Agilent 1260 Infinity II Prep HPLC system. HPLC columns were purchased from Phenomenex (Torrance, CA, USA). HPLC columns used were Analytical (150 x 4.6 mm), Semiprep (250 x 10 mm), and Prep size (250 x 21.2 mm) Luna® Omega 5 μm PS C18 100 Å columns from Phenomenex.

Synthesis of (thio)acetyl lysine containing peptides Peptides were synthesized on 2-chlorotrityl resin (100-200 mesh, 10 µmol scale) in 3 mL fritted syringes. The resin was swelled in DMF for 20 minutes. All amino acids (5 equiv.) were dissolved in 2 mL DMF along with HBTU (5 equiv.) and DIPEA (10 equiv.), except for the first amino acid of each peptide, for which HBTU was omitted, and Mcm, for which only 2 equiv. amino acid/HBTU and 4 equiv. DIPEA were used. Amino acid couplings were carried out as double couplings for 30 minutes. Fmoc deprotections were carried out with 2% (v/v) DBU in DMF (3 x 2 min), After complete synthesis, the resin was washed with DCM and dried under vacuum before being cleaved off. The standard cleavage solution contained DCM/TFA/TIPS/H_2O (10:8:1:1). After cleavage for 45 minutes, the solvent was removed *in vacuo*. The crude peptide was dissolved in 1 mL of 50% (v/v) MeCN in H_2O purified by HPLC. Fractions containing pure peptide were combined and lyophilized.

Synthesis of (thio)asparagine containing peptides Peptides were synthesized on 2-chlorotrityl resin (100-200 mesh, 10 µmol scale) The resin was swelled in DMF for 20 minutes. All amino acids (5 equiv.) were dissolved in 2 mL DMF along with HBTU (5 equiv.) and DIPEA (10 equiv.), except for the first amino acid of each peptide, for which HBTU was omitted, and Fmoc- Asn^{vS}(Xan)-OH, for which only 2 equiv. amino acid and 4 equiv. DIPEA were used. Amino acid couplings were carried out as double couplings for 30 minutes. Fmoc deprotections were carried out with 20% (v/v) Piperidine in DMF (2 x 10 min), After complete synthesis, the resin was washed with DCM and dried under vacuum before being cleaved off. The cleavage solution contained 75% DCM, 10% TFA, 5% TIPS, 5% EDT and 5% thioanisole (all percentages are v/v). After cleavage for 30 minutes, the solvent was removed *in vacuo*. The crude peptide was dissolved in 1 mL of 50% (v/v) MeCN in H_2O purified by HPLC.

Fluorescence measurements. Peptide stocks were dissolved in 500 µL and dissolved in phosphate buffered saline (10 mM Na_2HPO_4, 150 mM NaCl, pH 7.0, sterile filtered) and quantified. Measurements were performed as triplicates at 10 µM concentration with the following settings: Cnf: excitation: 240 nm; emission: 260-400 nm; 5 nm slit widths; Trp: excitation: 278 nm; emission: 300-450 nm; 5 nm slit widths; Mcm: excitation: 325 nm; emission: 340-500 nm; 5 nm slit widths.

Small scale synthesis of Fmoc-Gln derivatives. Synthesis for all derivatives was performed on a 50 µmol scale. A thionation solution was prepared by suspending P_4S_{10} (134 mg, 300 µmoles) and Na_2CO_3 (31.8 mg, 300 µmoles) in 9 mL anhydrous tetrahydrofuran in an oven-dried 20 mL scintillation vial. The thionation mixture was stirred for 1 hour at room temperature under argon atmosphere. In the meantime, reagents were weighed out in 1 dram vials. After the thionation mixture has stirred for 1 hour, 1.5 mL thionation solution (equals 1 equiv. of P_4S_{10}) was added to each vial. The solution was stirred for 1 minute to dissolve amino acids and a sample was taken for HPLC analysis (10 µL sample added to 1 mL MeOH). Reactions were blanketed with Argon before being closed, sealed with parafilm and stirred for 20 hours. After 20 hours, another sample was taken and was analyzed by HPLC and LCMS. HPLC analysis was done using 20 µL injections on a Luna Omega analytical column using gradient **3O**.

Synthesis of Fmoc-Gln$^{\delta S}$(Trt)-OH. P_4S_{10} (233 mg, 0.5 mmoles) and Na_2CO_3 (53.0 mg, 0.5 mmoles) were suspended in 10 mL anhydrous tetrahydrofuran. The solution was stirred under argon atmosphere for 1 hour. Fmoc-Gln(Trt)-OH (122 mg, 0.2 mmoles) were added and the reaction was stirred under argon atmosphere overnight. The next day the solvent was removed in vacuo. Crude product was redissolved in 1 mL MeOH and purified on Biotage system using a SNAP Ultra C18 column (12g) on a gradient from 50%-

100% Solvent B (0.1% TFA in acetonitrile) over 20 column volumes. Fractions containing product were lyophilized and product yielded as white powder in low yield (22.0 mg, 35.1 μmoles, 17.6%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.77 (d, *J* = 7.6 Hz, 2H), 7.61 (d, *J* = 7.5 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.34 – 7.29 (m, 8H), 7.29 – 7.27 (m, 2H), 7.26 (d, *J* = 2.8 Hz, 1H), 7.26 – 7.21 (m, 6H), 6.84 (s, 1H), 6.02 (d, *J* = 7.6 Hz, 1H), 4.57 (dd, *J* = 10.6, 6.8 Hz, 1H), 4.36 (dd, *J* = 10.7, 7.2 Hz, 1H), 4.30 – 4.21 (m, 2H), 2.52 – 2.39 (m, 2H), 2.22 (d, *J* = 15.3 Hz, 1H), 2.04 – 1.83 (m, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 200.28, 171.33, 156.40, 144.50, 143.90, 141.47, 128.78, 128.18, 127.89, 127.87, 127.31, 127.27, 125.27, 120.12, 71.03, 67.31, 61.84, 47.36, 33.18, 27.30. ESI^{+}-HRMS calculated for $C_{39}H_{35}N_2O_4S^{+}$: 627.2318; found: $[M+H]^{+}$: 627.2313.

Table 3.1: HPLC purification methods and retention times.

Peptide/Compound	Gradient	Retention Time	Column
F*PPK(Ac)-OH	**3A**	19.0	Vydac C18 Semiprep
F*PPPK(Ac)-OH	**3A**	19.2	Vydac C18 Semiprep
F*PPPPK(Ac)-OH	**3A**	21.4	Vydac C18 Semiprep
F*PPPPPK(Ac)-OH	**3A**	22.3	Vydac C18 Semiprep
F*PPPPPPK(Ac)-OH	**3B**	16.0	Vydac C18 Semiprep
F*PPK(Ac^S)-OH	**3C**	19.2	Vydac C18 Prep
F*PPPK(Ac^S)-OH	**3C**	19.6	Vydac C18 Prep
F*PPPPK(Ac^S)-OH	**3D**	18.2	Vydac C18 Prep
F*PPPPPK(Ac^S)-OH	**3D**	18.4	Vydac C18 Prep
F*PPPPPPK(Ac^S)-OH	**3D**	18.4	Vydac C18 Prep
WPPK(Ac)-OH	**3E**	21.5	Vydac C18 Prep
WPPPK(Ac)-OH	**3F**	19.6	Vydac C18 Prep
WPPPPK(Ac)-OH	**3D**	17.4	Vydac C18 Prep
WPPPPPK(Ac)-OH	**3D**	18.0	Vydac C18 Prep
WPPPPPPK(Ac)-OH	**3D**	18.0	Vydac C18 Prep
WPPK(Ac^S)-OH	**3G**	20.5	Vydac C18 Prep
WPPPK(Ac^S)-OH	**3G**	20.7	Vydac C18 Prep
WPPPPK(Ac^S)-OH	**3G**	20.4	Vydac C18 Prep
WPPPPPK(Ac^S)-OH	**3G**	20.3	Vydac C18 Prep
WPPPPPPK(Ac^S)-OH	**3G**	20.7	Vydac C18 Prep
µPPK(Ac)-OH	**3H**	22.2	Vydac C18 Prep
µPPPK(Ac)-OH	**3H**	22.3	Vydac C18 Prep
µPPPPK(Ac)-OH	**3H**	25.2	Vydac C18 Prep

Table 3.1 continued.

Peptide/Compound	Gradient	Retention Time	Column
µPPPPPK(Ac)-OH	**3H**	25.2	Vydac C18 Prep
µPPPPPPK(Ac)-OH	**3H**	26.4	Vydac C18 Prep
µPPK(Ac^S)-OH	**3I**	21.8	Vydac C18 Prep
µPPPK(Ac^S)-OH	**3I**	23.0	Vydac C18 Prep
µPPPPK(Ac^S)-OH	**3I**	22.9	Vydac C18 Prep
µPPPPPK(Ac^S)-OH	**3I**	22.9	Vydac C18 Prep
µPPPPPPK(Ac^S)-OH	**3I**	22.7	Vydac C18 Prep
F*PPN-OH	**3J**	16.7	Vydac C18 Semiprep
F*PPPPN-OH	**3K**	19.5	Vydac C18 Semiprep
F*PPPPPPN-OH	**3L**	14.6	Vydac C18 Semiprep
F*PP$N^{\gamma S}$-OH	**3M**	25.3	Vydac C18 Semiprep
F*PPPP$N^{\gamma S}$-OH	**3M**	25.0	Vydac C18 Semiprep
F*PPPPPP$N^{\gamma S}$-OH	**3N**	27.2	Vydac C18 Semiprep

F* = 4-Cyanophenylalanine; µ = 7-Methoxycoumarin-4-yl-alanine

Table 3.2: Peptide Purification/Characterization Methods and Retention Times.

No.	Time (min)	%B	No.	Time (min)	%B	No.	Time (min)	%B
3A	0:00	2	**3B**	0:00	2	**3C**	0:00	2
	3:00	2		3:00	2		3:00	2
	10:00	12		12:00	23		10:00	18
	25:00	22		25:00	32		25:00	28
	28:00	100		28:00	100		30:00	100
	31:00	100		31:00	100		31:00	100
	33:00	2		33:00	2		33:00	2
3D	0:00	2	**3E**	0:00	2	**3F**	0:00	2
	3:00	2		3:00	2		3:00	2
	10:00	20		10:00	15		10:00	18
	25:00	30		25:00	30		25:00	25
	30:00	100		30:00	100		30:00	100
	31:00	100		31:00	100		31:00	100
	33:00	2		33:00	2		33:00	2
3G	0:00	2	**3H**	0:00	2	**3I**	0:00	2
	3:00	2		5:00	2		5:00	2
	10:00	20		10:00	15		10:00	22
	25:00	35		30:00	30		30:00	40
	30:00	100		32:00	100		32:00	100
	31:00	100		37:00	100		37:00	100
	33:00	2		40:00	2		40:00	2

Table 3.2 continued

No.	Time (min)	%B	No.	Time (min)	%B	No.	Time (min)	%B
3J	0:00	2	**3K**	0:00	2	**3L**	0:00	2
	3:00	2		3:00	2		3:00	2
	10:00	8		10:00	10		10:00	15
	24:00	16		24:00	16		24:00	21
	28:00	100		28:00	100		28:00	100
	31:00	100		31:00	100		31:00	100
	33:00	2		33:00	2		33:00	2
3M	0:00	2	**3N**	0:00	2	**3O**	0:00	5
	3:00	2		3:00	2		10:00	5
	35:00	20		35:00	25		105:00	100
	36:00	100		36:00	100		110:00	100
	41:00	100		41:00	100		111:00	5
	44:00	2		44:00	2			

Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

Table 3.3: MALDI-TOF MS Characterization of Purified Peptides.

Peptide	$[M+H]^+$	
	Calculated	Found
F*PPK(Ac)-OH	555.29	555.27
F*PPPK(Ac)-OH	652.35	652.56
F*PPPPK(Ac)-OH	749.41	749.56
F*PPPPPK(Ac)-OH	846.47	846.61
F*PPPPPPK(Ac)-OH	943.53	943.73
F*PPK(Ac^S)-OH	571.35	571.38
F*PPPK(Ac^S)-OH	668.41	668.45
F*PPPPK(Ac^S)-OH	765.47	765.52
F*PPPPPK(Ac^S)-OH	862.53	862.63
F*PPPPPPK(Ac^S)-OH	959.59	959.71
WPPK(Ac)-OH	569.31	569.60
WPPPK(Ac)-OH	666.37	666.70
WPPPPK(Ac)-OH	763.43	763.79
WPPPPPK(Ac)-OH	860.49	860.88
WPPPPPPK(Ac)-OH	957.55	957.94
WPPK(Ac^S)-OH	585.37	585.51
WPPPK(Ac^S)-OH	682.43	682.59
WPPPPK(Ac^S)-OH	779.49	779.67
WPPPPPK(Ac^S)-OH	876.55	876.73
WPPPPPPK(Ac^S)-OH	973.61	973.83
μPPK(Ac)-OH	628.30	628.93
μPPPK(Ac)-OH	725.36	725.83
μPPPPK(Ac)-OH	822.42	822.92
μPPPPPK(Ac)-OH	919.48	919.81

Table 3.3 continued

Peptide	$[M+H]^+$	
	Calculated	Found
μPPPPPPK(Ac)-OH	1,016.54	1,017.06
μPPK(Ac^S)-OH	644.36	644.65
μPPPK(Ac^S)-OH	741.42	741.74
μPPPPK(Ac^S)-OH	838.48	838.39
μPPPPPK(Ac^S)-OH	935.54	935.41
μPPPPPPK(Ac^S)-OH	1,032.60	1,033.06
F*PPN-OH	499.23	499.21
F*PPPPN-OH	693.34	693.34
F*PPPPPPN-OH	887.44	887.42
F*PP$N^{\gamma S}$-OH	515.21	515.21
F*PPPP$N^{\gamma S}$-OH	709.31	709.30
F*PPPPPP$N^{\gamma S}$-OH	903.42	903.41

F* = 4-Cyanophenylalanine; μ = 7-Methoxycoumarin-4-yl-alanine

Chapter 4 : Chemoselective Modifications for the Traceless Ligation of Thioamide-Containing Peptides and Proteins

4.1 Introduction

In our previous work, we showed that thioamides can also quench the fluorescence of a variety of fluorophores in a distance-dependent manner. The quenching takes place either through a Förster resonance energy transfer (FRET) mechanism for UV range fluorophores[83,134], or through a photoinduced electron transfer (PET) mechanism for red-shifted fluorophores[86,135]. We further demonstrated the utility of various fluorophore/thioamide pairs in monitoring protein thermal unfolding, protein substrate binding and protease activities[87]. To apply the fluorophore/thioamide dual labeling method to full-length proteins, we developed a semi-synthesis strategy to incorporate thioamides into proteins. As a backbone modification, the thioamide functional group can only be installed onto small molecules through solution phase thionation, or short peptides by fluorenylmethyloxycarbonyl (Fmoc) based solid phase peptide synthesis (SPPS)[65,68]. We extended this synthetic approach to full-length proteins by adopting native chemical ligation (NCL), where a small thioamide-containing fragment was first prepared by SPPS and then conjugated to an expressed protein fragment. NCL, as pioneered by Kent et al.[136,137], joins a C-terminal thioester and an N-terminal Cys through transthioesterification, after which an *S*-to-*N* acyl shift takes place to form a native amide bond at the junction. We showed that the thioamide can be placed in either the thioester or N-terminal Cys fragment, and successfully prepared several constructs of thioamide-labeled α-synuclein (αS), a 140 aa protein implicated in Parkinson's disease[69,73,114]. The detailed synthetic methodology has also been reviewed[1]. We note that while an in vitro translation strategy incorporating a thioamide-containing dipeptide was recently demonstrated[138], the NCL method is more general in terms of sequence and produces higher protein yields. One inherent limitation of the conventional NCL reactions used in our initial studies is the

necessity of a Cys at the ligation site. Cys is among the least abundant amino acids (2.4% in the human proteome[139]), and readily forms disulfide bonds under ambient conditions. For protein targets without a native Cys in their sequences (or at the desired ligation sites), NCL would leave an artifact of residual Cys in the ligated product, undermining the utility of the thioamide as a minimalist probe and introducing complications during biophysical studies. To circumvent the Cys limitation, various groups have developed traceless NCL methods by either masking or "erasing" the side chain thiol[121,140]. Masking techniques utilize alkyl halides to convert Cys into Lys/Glu/Gln analogs, or homocysteine (Hcs) into Met after ligation[141–144]. Alternatively, the side chain thiol can be "erased" by RANEY® nickel or radical initiated desulfurization, where either a Cys or a thiol analog of another amino acid is converted into the native residue (Ala in the case of Cys)[145–147]. Thus far, synthetic β- or γ-thiol analogs of 12 native amino acids have been reported[148–150]. The radical initiated method may also be extended to selenolcontaining residues such as selenocysteine (Sec) or β-selenophenylalanine, where the mild tris(2-carboxyethyl)phosphine (TCEP) initiated deselenization could preserve Cys residues in the sequence[148,151,152]. Having briefly explored the compatibility of thioamides with Hcs masking[69], we sought to further expand the scope of thioamide incorporation with the desulfurization/deselenization approach.

4.2 Results and Discussion

We started with a systematic evaluation of the compatibility of thioamides with various desulfurization/deselenization methods, namely TCEP assisted radical deselenization, 2,2′- azobis[2-(2-imidazolin-2-yl)propane] (VA-044) initiated radical desulfurization and RANEY® nickel desulfurization. Yanxin Wang performed all the work on radical deselenization and a detailed description can be found elsewhere[120]. We were

delighted to find that both the ligation and deselenization proceeded smoothly and chemoselectively. The positive results prompted us to further investigate selective desulfurization with VA-044 as radical initiator. When the Cys/thioamidecontaining model peptide **4-1** was treated with VA-044, the desired Ala peptide **4-2** was generated as the major product, which also preserved the characteristic absorption of the thioamide. In contrast, when the same peptide was treated with RANEY® nickel, a cleavage product **4-1*** was observed instead (see Figure 4-1). These results suggest that the radical initiated mechanisms are key to their thioamide compatibility. Consistent with our prior knowledge that a thioamide can function as an acceptor for PET (where a single electron is transferred from thioamide to a paired fluorophore, generating a transient thioamide radical cation in the process)[86,135], thioamides appeared to be stable in the radical-rich environment of the desulfurization/deselenization mixture. On the contrary, RANEY® nickel is well-established for its broad substrate scope where thiols, thioethers and thionoesters (thiocarbonyl esters) are all effectively desulfurized[153]; although the cleavage side reaction we observed in aqueous solution was somewhat different from the reductive desulfurization to an amine reported in organic solvent[154,155], it was not surprising that this metal–sulfur affinity based reaction resulted in desulfurization of the thioamide. Given that Cys and β-thiol analogs are much easier to synthesize and handle than Sec or β-selenol compounds, we decided to pursue VA-044 initiated desulfurization as the primary method for our traceless NCL incorporation of thioamides, and reserve TCEP assisted deselenization for applications where additional selectivity against Cys is needed.

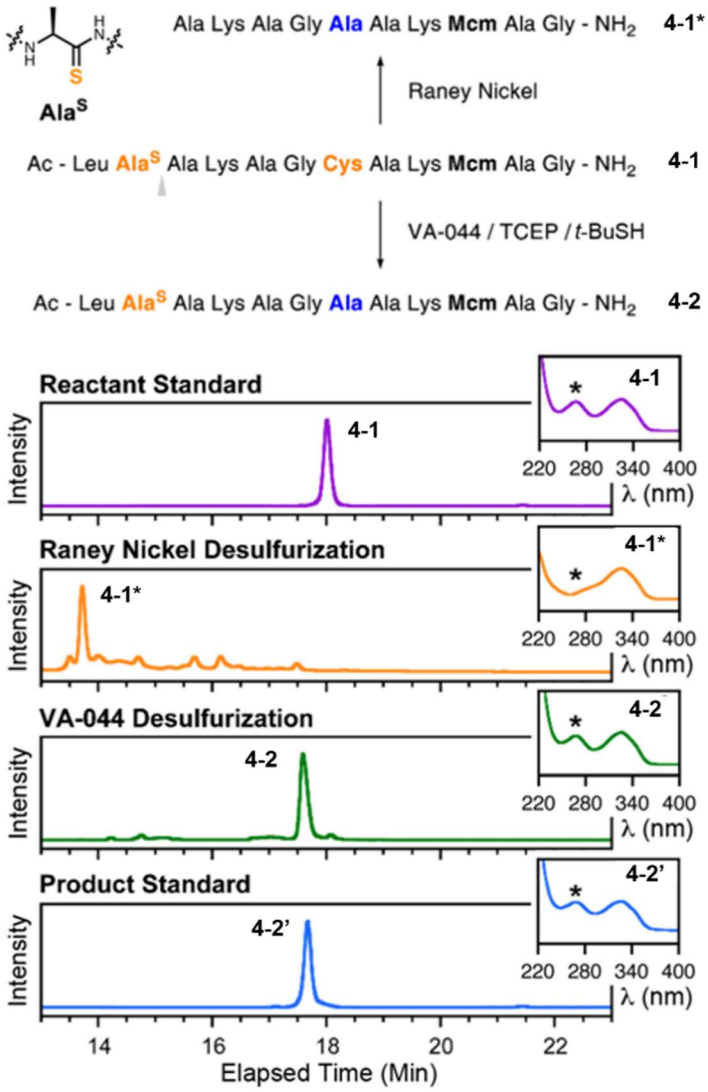


Figure 4-1: Thioamide Compatibility with different Desulfurization Methods. Top: reaction scheme; Bottom: HPLC chromatogram monitored at 325 nm and UV-Vis absorption profiles for major peak in each chromatogram.

Selective desulfurization using thioacetamide scavenger. To further characterize the VA-044 initiated desulfurization reaction, we systematically varied key parameters to identify the optimal conditions. We found that a high TCEP concentration (empirically above 40 mM) was necessary to avoid disulfide bond formation, which likely resulted from the collisional quenching of thiol radicals and was reversed by TCEP

reduction. We also observed that the reaction was tolerant to a wide range of VA-044 concentrations; above a critical amount (empirically 10 equivalents), the desulfurization reaction proceeded smoothly with no complications for as high as 1000 equivalents of VA-044, which is consistent with our current understanding that VA-044 merely served as a radical initiator and that propagation primarily occurred via the thiol additive *t*-BuSH. Under optimal conditions, the reaction completed in less than 10 min, with no major complications for up to 4 h. In addition, we showed that dissolved oxygen – which may function as collisional radical quencher – was well-tolerated, with only minor alterations in reaction kinetics. Denaturants such as guanidinium (Gdn) and urea (which may become necessary for proteins or long peptides) were also shown to be compatible with the radical initiated desulfurization. When investigating the robustness of the chemoselectivity, we observed a small amount of desulfurization of the thioamide, resulting in the accumulation of the thioamide-to-oxoamide conversion side product **4-5** after prolonged treatment with VA-044. Although **4-5** only represented a 6% side reaction after 2 h, it became significant after 18 h (15–20% side product formation). Although the thioamide C=S bond is much stronger than the Cys C–S bond, increased stability of the thioamide allylic-type radical may contribute to its susceptibility to desulfurization. We therefore turned to thioacetamide as a "suicide scavenger" to prevent desulfurization of the thioamide in the peptide. When the model peptide **4-4′** (prime symbols denote genuine peptide standards to distinguish them from species identified in reactions) was treated with VA-044 in the absence of thioacetamide, side product **4-5** was generated at 16% by 18 h; in the presence of thioacetamide, **4-4′** was fully stable over the same time period. When we applied this additive to the Cys/thioamide peptide **4-3**, the same suppression was observed, improving the yield of desired product **4-4** from 58% to 88% after an 18 h reaction (Figure 4-2). As a further validation, we demonstrated that thioacetamide was effective over a wide range of

concentrations as a scavenger, and that the oxygen and denaturant tolerance was unaltered by its addition. It is worth noting that the residual peak at 21.8 min represents a 2–3% Cys-to-Ser conversion byproduct **4-6** (based on MALDI-TOF MS). The formation of the isomeric dethioamidated peptide **4-7** was not observed. These studies were carried out together with Yanxin Wang.

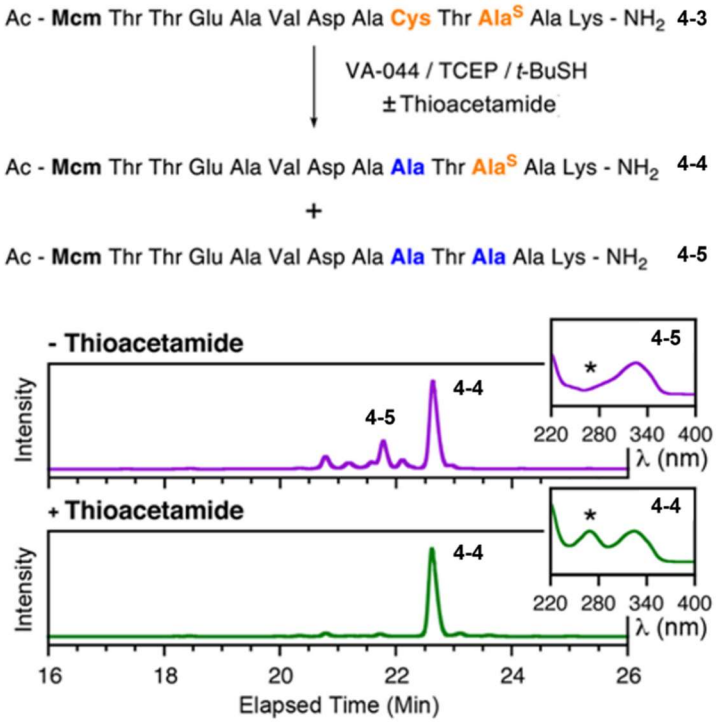


Figure 4-2: Thioacetamide as Small Molecule Scavenger. Top: Reaction scheme; Bottom: HPLC chromatogram monitored at 325 nm, and UV-Vis absorption profiles for selected peaks in each chromatogram.

One pot ligation/desulfurization. Having established a robust method for selective Cys desulfurization in the presence of thioamides, we next sought to apply this method in traceless NCL, where two peptide fragments are first joined through a Cys, which is subsequently “erased” to form an Ala peptide. In particular, we were interested in achieving one-pot ligation-desulfurization without an intermediate purification step to

isolate the ligated Cys product. The major hurdle was aromatic thiol additives such as thiophenol, which was used to accelerate the NCL reaction[156], but may also function as a radical quencher to sequester subsequent desulfurization[157]. For large proteins, one may simply remove the residual PhSH by dialysis after ligation; for small proteins and peptides, however, a different strategy must be devised. Knowing that the pK_a of PhSH is 6.5, we hypothesized that we could remove PhSH by lyophilization after NCL if the pH was kept sufficiently low to maintain its protonated form (vapor pressure 1.8 mBar[158]). In contrast, if the ligation mixture (at neutral or basic pH) was lyophilized directly, the majority of the PhSH would exist as non-volatile thiophenolate, and interfere with the desulfurization. Indeed, when we acidified a neutral PhSH solution to pH 1.6 and subjected it to lyophilization, we observed a significant reduction in residual PhSH concentration. Using thioester **4-8a** and Cys-containing peptide **4-8b**, we demonstrated that one-pot ligation-desulfurization can be successfully performed in the presence of PhSH as an aromatic thiol additive, by adopting the simple procedure of acidification and lyophilization (see Figure 4-3). As a further validation, we compared the desulfurization product 5 to the authentic peptide standard 5′, and found them to be perfect matches in both mass and UV-Vis absorption profiles. With thioacetamide present as a scavenger, no thioamide-to-oxoamide conversion was observed. One-pot ligation-desulfurization experiments were carried out together with Yanxin Wang.

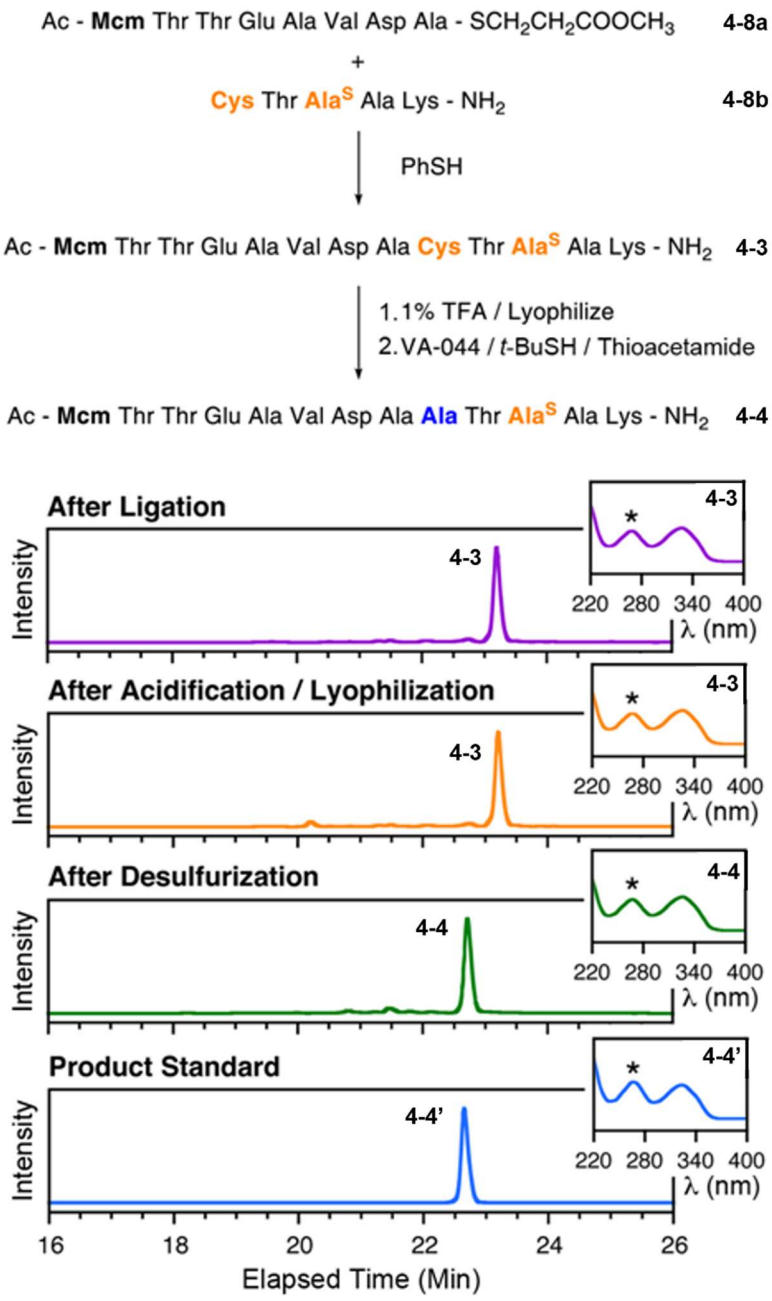


Figure 4-3: One-Pot Ligation-Desulfurization. Top: Reaction Scheme; Bottom: HPLC chromatogram monitored at 325 nm, and UV-Vis absorption profiles for major peaks in each chromatogram.

Desulfurization of Cys analogs. To further expand the scope of traceless NCL for thioamide incorporation, we also investigated the desulfurization of β-thiol amino acid

analogs, selective deselenization in the presence of both thioamide and Cys residues, as well as the incorporation of Sec into expressed proteins. Similar to the conversion of Cys into Ala, synthetic β- or γ-thiol analogs of other amino acids may also be used as ligation handles and then “erased” to form the native residues. As a proof-of-concept example, we performed the VA-044 initiated desulfurization reaction on Pen-containing peptide **4-9** (Pen = penicillamine). Pen is the β-thiol analog of Val[159]; we chose it because its desulfurization involves the formation of a tertiary carbon radical close to the peptide backbone, and is thus the most sterically demanding of all β-thiol analogs. Upon treatment with VA-044 in the presence of thioacetamide, **4-9** was converted to desired **4-4** in 83% yield, with its thioamide intact (Figure 4-4). The reaction was somewhat slower than Cys desulfurization, presumably due to the steric constraints near the reaction center. The successful expansion of the VA-044 initiated radical desulfurization to include Pen clearly demonstrated the versatility of this method for thioamide incorporation, providing a gateway to other synthetic β- or γ-thiol analogs.

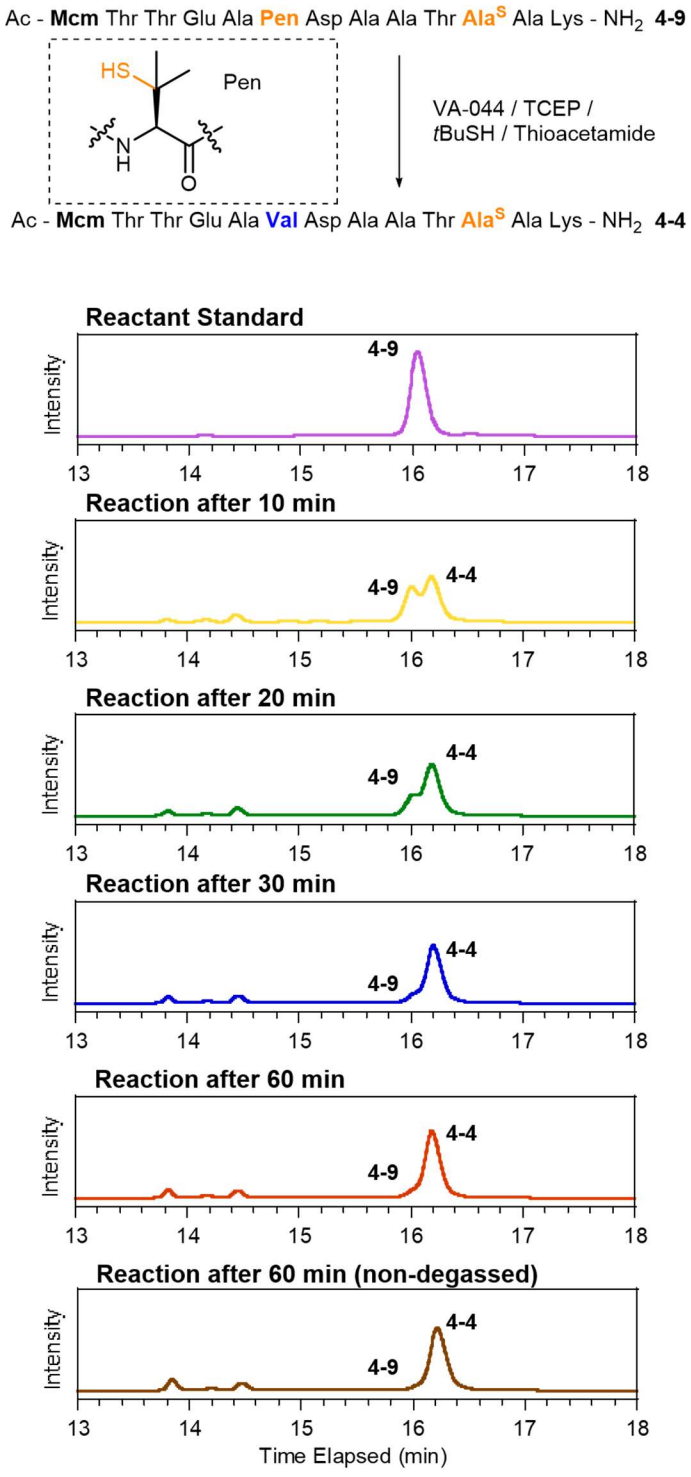


Figure 4-4: Desulfurization time-course of Penicillamine containing peptides. Top: Reaction scheme; Bottom: HPLC chromatograms monitored at 325 nm at various time points.

Synthesis of GB1 protein by ligation/desulfurization. We then applied the ligation–desulfurization method to the synthesis of a thioamide-containing version of the B1 domain of Streptococcal protein (GB1)[160]. GB1 is small protein, but has been studied extensively as a model biophysical system and has been used in protein engineering efforts involving both natural and unnatural amino acids[161–166]. Therefore, we are interested in synthesizing thioamide analogs of GB1 as part of a longer term effort to study the impact of thioamides on the folding and function of proteins. To synthesize GB1 using NCL and desulfurization, we chose a central Ala residue as a point of disconnection. The GB1 N-terminus was synthesized as an acyl hydrazide ($GB1_{1–23}$ L^{S}_{5}-N_2H_3, **4-10a**) and converted to the phenyl thioester in situ using Liu's method ($GB1_{1–23}L^{S}_{5}$-SPh, **4-10b**)[119,167]. The GB1 C-terminus was synthesized by conventional SPPS with an Ala-to-Cys mutation at its N-terminus ($GB1_{24–56}$ A_{24}C-OH, **4-11**). Ligation under standard conditions lead to quantitative conversion to form full length $GB1_{1–56}$ $L^{S}_{5}A_{24}$C- OH (**4-12a**), which was isolated by HPLC purification and desulfurized using VA-044 to give $GB1_{1–56}L^{S}_{5}$-OH (**4-12b**) in 71% yield (see Figure 4-5).

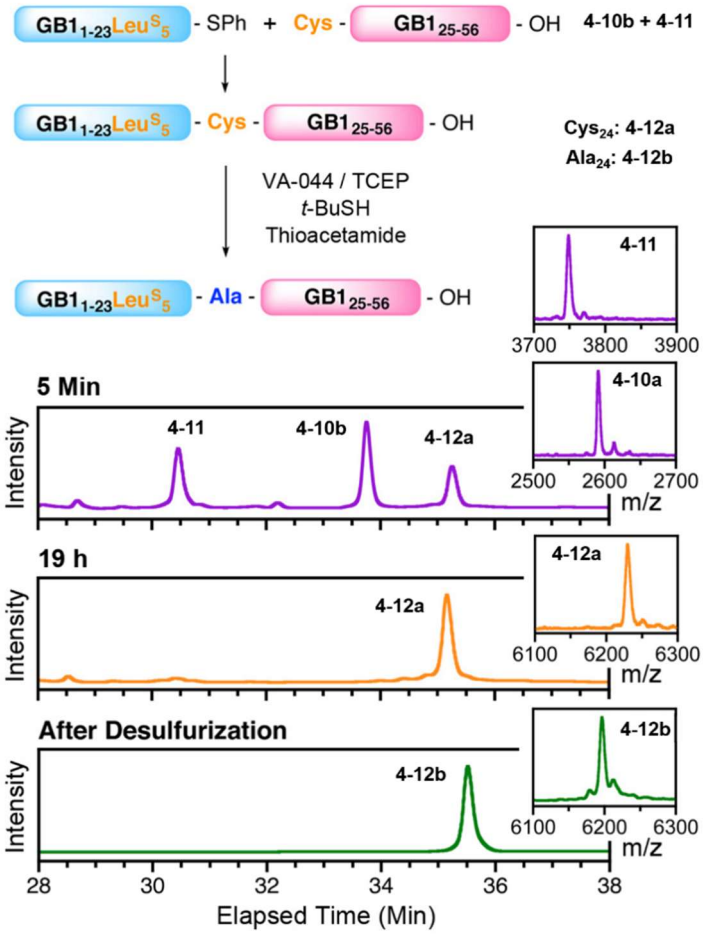


Figure 4-5: Synthesis of Thioamide GB1 by Ligation and Desulfurization. Top: reaction scheme; Bottom: HPLC chromatograms monitored at 272 nm, and MALDI-TOF MS profiles for selected peaks in each chromatogram.

We note that attempts to acidify the reaction and desulfurize after lyophilizing gave a number of byproducts which we attribute to the reactivity of residual $NaNO_2$ (from acyl hydrazide activation) in highly acidic media (see Figure 4-6). Therefore, the one pot lyophilization/desulfurization procedure may not be applicable to the products of ligations using the Liu method, and other additives such as trifluoroethanethiol should be considered[157].

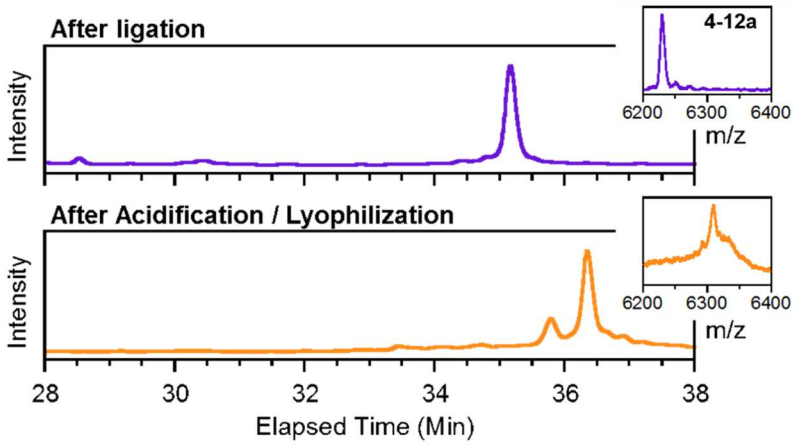


Figure 4-6: Attempted One-Pot Ligation-Desulfurization is Incompatible with Hydrazide Ligation Chemistry. Bottom shows HPLC chromatogram monitored at 272 nm and MALDI-TOF MS profile of unidentified by-products after Acidification/Lyophilization.

4.3 Conclusion

We have devised a method for the traceless incorporation of thioamides into peptides and proteins using chemoselective desulfurization/deselenization in combination with NCL. We demonstrated that Cys can be selectively converted into Ala using VA-044 as a radical initiator in the presence of a backbone thioamide, and that thioacetamide can be used as an additive to improve the robustness of this chemoselectivity. We also showed that by acidifying the ligation mixture, PhSH can be effectively removed by lyophilization, enabling one-pot ligation–desulfurization that had not previously been achieved. As a proof-of-concept example, we synthesized thioamidecontaining GB1 through traceless NCL at a Leu site. We will use the methods developed here to synthesize other thioamide variants of GB1 in order to better understand the impact of thioamide incorporation on this small, well-folded protein. We note that Liu's acyl hydrazide method gave high yields of the thioester protein fragment with no degradation of the thioamide and we will explore

this method in the context of other thioamide ligations. To expand the scope of ligation sites, we also showed the selective desulfurization of β-thiol analogs using conversion of Pen into Val as an example, demonstrated selective deselenization in the presence of both Cys and a thioamide, and devised a chemoenzymatic approach to incorporate ligation handles other than Cys onto the N-termini of expressed proteins. With these methods, we are now able to perform traceless ligation at sites other than Cys for thioamide-containing peptides and proteins, making thioamide incorporation by semi-synthesis fully general.

4.4 Materials and Methods

General Information VA-044 was purchased from Wako Pure Chemical Industries (Osaka, Japan). Fmoc protected amino acids and peptide synthesis reagents were purchased from EMD Millipore (Billerica, MA, USA). *N*α-Fmoc-L-(*S*-Trityl)-Penicillamine (Fmoc-Pen(Trt)-OH) was purchased from Sigma-Aldrich (St. Louis, MO, USA). *N*α-Fmoc-L-Thioalanine-benztriazolide (Fmoc-Ala^{S}-Bt) was synthesized by Yanxin Wang. Hydrazine Hydrate solution was purchase from Oakwood Chemical (Estill, SC, USA). GB1$_{24\text{-}56}$ A$_{24}$C-OH was purchased from Genscript (Piscataway, NJ, USA) and was supplied as lyophilized powder. Amicon Ultra centrifugal filter units (3k Da MWCO) were purchased from EMD Millipore. All other reagents and solvents were purchased from Fisher Scientific or Sigma-Aldrich unless otherwise specified.

High resolution electrospray ionization mass spectra (ESI-HRMS) were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA, USA). Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. UV-Vis absorption spectra were acquired on a Hewlett-Packard

8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III instrument. Analytical HPLC was performed on an Agilent 1100 Series HPLC system. Preparative HPLC was performed on a Varian Prostar HPLC system (currently Agilent Technologies). HPLC columns were purchased from W. R. Grace & Company (Columbia, MD, USA).

***N*α-Fmoc-L-Thioleucine-nitrobenztriazolide (Fmoc-Leu^{S}-Nbt) synthesis:**

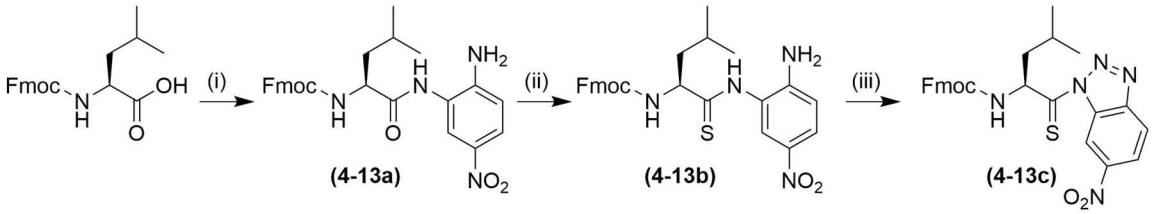


Figure 4-7: Synthesis of *N*α-Fmoc-L-Thioleucine-nitrobenztriazolide: (i) NMM, IBCF, 4-nitrobenzene-1,2-diamine, THF; (ii) P_4S_{10}, Na_2CO_3, THF; (iii) $NaNO_2$, 95% $AcOH_{(aq)}$.

***N*α-Fmoc-L-Leucine-2-amino-5nitroanilide (4-13a):** Fmoc-Leu-OH (2.82g, 8.0 mmoles) were dissolved in 70 mL anhydrous THF. N-methylmorpholine (1.76 mL, 16.0 mmoles) was added and the reaction was cooled to -10°C while the flask was purged with Argon. Isobutylchloroformate (1.04 mL, 8.00 mmoles) were added dropwise under stirring. Upon complete addition the reaction mixture was stirred for 15 min at -10°C and 4-nitrobenzene-1,2-diamine (1.22 g, 8.00 mmoles) was added. The reaction was stirred under Argon for 2 hours at -10°C, then at room temperature overnight. After removing the solvent *in vacuo*, the residue was dissolved in 50 mL DMF. Upon addition of 250 mL sat. aqueous KCl solution and 250 mL Milli-Q water, a yellow solid precipitated. The solid was filtered, washed with Milli-Q water and dried under vacuum. The title compound was isolated in quantitative yield and 94.9**%** purity, as determined by HPLC.

^{1}H NMR (500 MHz, DMSO-d_6) δ 9.57 (s, 1H), 8.24 (d, *J* = 2.2 Hz, 1H), 7.91 – 7.84 (m, 3H), 7.79 (d, *J* = 7.6 Hz, 1H), 7.75 (d, *J* = 7.5 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.32 (t, *J* = 7.5 Hz, 2H), 6.79 (d, *J* = 9.1 Hz, 1H), 6.56 (s, 2H), 4.37 – 4.21 (m, 3H), 1.76 – 1.65 (m, 1H), 1.67 – 1.56 (m, 1H), 0.95 (d, *J* = 6.4 Hz, 3H), 0.92 (d, *J* = 6.5 Hz, 3H).

^{13}C NMR (126 MHz, DMSO) δ 172.13, 156.22, 149.19, 143.85, 143.73, 140.70, 135.45, 127.61, 127.03, 125.31, 122.97, 121.54, 121.29, 120.07, 113.67, 65.68, 53.73, 46.69, 40.25, 24.33, 23.03, 21.56.

HRMS (ESI): $[M+H]^+$ calc.: 489.2138, obs.: 489.2148

***N*α-Fmoc-L-Thioleucine-2-amino-5nitroanilide (4-13b):** Anhydrous Na_2CO_3 (334 mg, 3.15 mmoles) and P_4S_{10} (1.40 g, 3.15 mmoles) were suspended in anhydrous THF in an oven-dried round bottom flask and stirred under Argon atmosphere for 30 minutes until solution became clear. Compound **4-13a** (2.05 g, 4.20 mmoles) added and reaction stirred over night at room temperature under Argon atmosphere. The next day the solution was removed *in vacuo*. The resulting residue was resuspended in EtOAc and filtered over a pad of Celite to remove insoluble P_4S_{10} adducts. The filtrate was washed twice with 5% $NaHCO_3$ and once with brine. The aqueous layer was backextracted with once with EtOAc. The combined organic layers were dried over $MgSO_4$, filtered and the solvent was removed *in vacuo*. The crude product was redissolved in CH_2Cl_2 and purified over silica to afford pure product as a yellow solid in 68.4% yield. R_f = 0.5 in EtOAc/Hexanes (1:1)

^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.65 (s, 1H), 7.99 (s, 1H), 7.92 (dd, *J* = 9.0, 2.6 Hz, 1H), 7.72 (t, *J* = 6.9 Hz, 2H), 7.48 (d, *J* = 7.5 Hz, 1H), 7.43 (d, *J* = 7.6 Hz, 1H), 7.40 – 7.30 (m, 2H), 7.29 – 7.18 (m, 2H), 6.56 (d, *J* = 9.0 Hz, 1H), 5.69 (d, *J* = 7.5 Hz, 1H), 4.66 (d, *J* = 8.5 Hz, 1H), 4.61 (s, 2H), 4.26 (d, *J* = 7.3 Hz, 2H), 4.11 (t, *J* = 7.2 Hz, 1H),

1.87 – 1.80 (m, 1H), 1.79 – 1.72 (m, 2H), 0.99 (t, *J* = 6.0 Hz, 6H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 206.93, 157.20, 148.61, 143.37, 143.28, 141.29, 141.23, 138.18, 128.00, 127.93, 127.36, 127.22, 125.57, 124.92, 124.85, 122.27, 120.20, 120.14, 115.17, 67.57, 60.45, 46.92, 44.28, 25.05, 22.80, 22.27. HRMS (ESI): $[M+H]^+$ calc.: 505.1910, obs.: 505.1916

N_α-Fmoc-L-Thioleucine-nitrobenztriazolide (4-13c): Compound **4-13b** (1.46 g, 2.89 mmoles) were dissolved in 25 mL of 95% glacial acetic $acid_{(aq)}$. The solution got cooled to 0°C and $NaNO_2$ (299 mg, 4.34 mmoles) were added slowly. Upon complete addition, the reaction mixture was stirred for 30 min at 0°C, after which 125 mL of cold Milli-Q water was added. The precipitating orange solid was filtered, washed with cold Milli-Q water and dried *in vacuo*. The resulting crude product was afforded in 85.9% yield. The product can be used crude or purified on a silica column to afford pure product. R_f = 0.17 in CH_2Cl_2.

^{1}H NMR (500 MHz, Chloroform-*d*, major rotomer only) δ 9.66 (s, 1H), 8.44 (dd, *J* = 8.9, 2.1 Hz, 1H), 8.31 (d, *J* = 8.9 Hz, 1H), 7.82 – 7.71 (m, 2H), 7.61 (d, *J* = 7.5 Hz, 2H), 7.40 (t, *J* = 5.6 Hz, 2H), 7.32 (q, *J* = 7.2 Hz, 2H), 6.37 – 6.24 (m, 1H), 5.63 (d, *J* = 9.6 Hz, 1H), 4.54 (dd, *J* = 10.8, 6.7 Hz, 1H), 4.40 (dd, *J* = 10.9, 6.9 Hz, 1H), 4.22 (t, *J* = 6.8 Hz, 1H), 1.96 – 1.86 (m, 1H), 1.86 – 1.65 (m, 2H), 1.26 (s, 2H), 1.13 (d, *J* = 6.4 Hz, 3H), 0.97 (d, *J* = 6.6 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 210.70, 156.05, 149.70, 149.14, 143.74, 141.48, 132.04, 127.87, 127.20, 125.19, 125.09, 122.31, 121.61, 120.13, 112.94, 67.06, 47.40, 46.05, 25.87, 23.45, 21.45. HRMS (ESI): $[M+Na]^+$ calc.: 538.1525, obs.: 538.1535

Synthesis and Purification of model peptides Peptides were synthesized using standard Fmoc solid phase peptide synthesis (SPPS) procedure on either Rink Amide or 2-chlorotrityl chloride resin (100 - 200 mesh; 0.6 mmol substitution/g). For coupling, 5

equiv of amino acid and 5 equiv of HBTU were dissolved in dimethylformamide (DMF), pre-activated for 1 min in the presence of 10 equiv of *N*,*N*-diisopropylethylamine (DIPEA), and then stirred with the resin for 30 min at room temperature. For deprotection, 20% piperidine in DMF was stirred with the resin for 20 min. Thioalanine (denoted Ala^S or A^S) was introduced through activated benzotriazole precursors (S4); 2 equiv of the precursor was dissolved in anhydrous CH_2Cl_2, and stirred with the resin for 45 min in the presence of 2 equiv DIPEA. For N-terminally acetylated peptides, an acetylation cocktail of 1:1:8 acetic anhydride/N-methylmorpholine/DMF was stirred with the resin twice, each for 10 min, after the last deprotection. Upon completion of SPPS, resins were rinsed thoroughly with CH_2Cl_2 and dried under vacuum.

For cleavage, resins were treated with a cleavage cocktail (12:1:1:26 trifluoroacetic acid/triisopropylsilane/H_2O/CH_2Cl_2) for 30 min. The solution was then collected by filtration, and dried by rotary evaporation. For purification, the crude residues were brought up in 1:10 MeCN/H_2O, and then purified by reverse phase HPLC with acidified (with 0.1% trifluoroacetic acid) MeCN/H_2O gradients. Individual fractions were characterized by MALDI-TOF MS, and dried by lyophilization.

Synthesis and Purification of GB1$_{1\text{-}23}$ Leu$^S_5$-N_2H_3 (4-10a) Hydrazide resin for peptide synthesis was prepared according to Liu and coworkers[119]. For coupling, 5 equiv of amino acid and 5 equiv of HBTU were dissolved in DMF, pre-activated for 1 min in the presence of 10 equiv of DIPEA, and then stirred with the resin for 30 min at room temperature. For deprotection, 2% (v/v) 1,8- diazabicyclo[5.4.0]undec-7-ene (DBU) in DMF was stirred with the resin three times for 2 min each. Thioleucine (denoted Leu^S or L^S) was introduced through activated nitrobenztriazole precursor; 2 equiv of the precursor was dissolved in anhydrous CH_2Cl_2, and stirred with the resin for 45 min in the presence

of 2 eq DIPEA. Upon completion of SPPS, resin was rinsed thoroughly with CH_2Cl_2 and dried under vacuum. For cleavage, resin was treated with a cleavage cocktail (6:1:1:12 trifluoroacetic acid/triisopropylsilane/H_2O/CH_2Cl_2) for 30 min. The solution was then collected by filtration, and dried by rotary evaporation. For purification, the crude residues were brought up in 1:1 MeCN/H_2O, diluted with H_2O, and then purified by reverse phase HPLC using gradient **4G**. Individual fractions were characterized by MALDI-TOF MS, and dried by lyophilization.

Purification of $GB1_{24\text{-}56}$ A_{24}C-OH (4-11) Crude peptide was purchased from Genscript and received as a lyophilized white powder. Crude product was dissolved in 1:1 MeCN/H_2O and purified by reverse phase HPLC using gradient **4H**. Individual fractions were characterized by MALDI-TOF MS, and dried by lyophilization.

Table 4.1: Peptide Purification Methods and Retention Times.

Peptide/Compound	Gradient	Retention Time	Column
Ac-L**A**^{S}AKAG**C**AKμAG-NH_2 (**4-1**)	4A	23.1 min	Vydac C18 Semiprep
Ac-L**A**^{S}AKAG**A**AKμAG-NH_2 (**4-2**)	4A	22.3 min	Vydac C18 Semiprep
Ac-μTTEAVDA**C**T**A**^{S}AK-NH_2 (**4-3**)	4B	16.0 min	Vydac C18 Semiprep
Ac-μTTEAVDA**A**T**A**^{S}AK-NH_2 (**4-4**)	4C	21.6 min	Vydac C18 Semiprep
Ac-μTTEAVDA**A**T**A**AK-NH_2 (**4-5**)	4C	19.9 min	Vydac C18 Semiprep
Ac-μTTEAVDA**S**T**A**^{S}AK-NH_2 (**4-6**)	4C	19.7 min	Vydac C18 Semiprep
Ac-μTTEAVDA**C**T**A**AK-NH_2 (**4-7**)	4C	21.2 min	Vydac C18 Semiprep
Ac-μTTEAVDA-$SC_2H_4COOCH_3$ (**4-8a**)	4D	17.6 min	YMC-Pack Pro C8 Semiprep
CT**A**^{S}AK-NH_2 (**4-8b**)	4E	12.8 min	Vydac C18 Semiprep
Ac-μTTEAVDA**ψ**T**A**^{S}AK-NH_2 (**4-9**)	4F	34.6 min	Vydac C18 Prep
$GB1_{1\text{-}23}$ L^{S}_{5}-N_2H_3 (**4-10a**)	4G	21.9 min	Vydac C18 Prep
$GB1_{24\text{-}56}$ A_{24}C-OH (**4-11**)	4H	22.3 min	Vydac C18 Prep
GB1 L^{S}_{5} (**4-12b**)	4I	21.3 min	Vydac C18 Semiprep

μ = 7-methoxycoumarinylalanine; ψ = penicillamine

Table 4.2: HPLC Gradients for Peptide Purification and Characterization.

No.	Time (min)	%B	No.	Time (min)	%B	No.	Time (min)	%B
4A	0:00	2	**4B**	0:00	2	**4C**	0:00	2
	5:00	2		5:00	2		5:00	2
	10:00	15		10:00	25		10:00	20
	25:00	30		25:00	40		25:00	35
	30:00	100		30:00	100		30:00	100
	35:00	100		35:00	100		35:00	100
	40:00	2		40:00	2		40:00	2
4D	0:00	2	**4E**	0:00	2	**4F**	0:00	5
	5:00	2		5:00	2		5:00	5
	10:00	25		10:00	10		15:00	22
	30:00	45		20:00	20		45:00	28
	35:00	100		25:00	100		50:00	100
	40:00	100		30:00	100		55:00	100
	45:00	2		35:00	2		60:00	5
4G	0:00	5	**4H**	0:00	2	**4I**	0:00	2
	5:00	5		5:00	2		10:00	2
	10:00	20		10:00	23		15:00	30
	25:00	30		45:00	35		35:00	50
	27:00	100		48:00	35		37:00	50
	30:00	100		50:00	100		39:00	100
	32:00	5		53:00	100		43:00	100
				55:00	2		45:00	2

Table 4.2 continued.

No.	Time (min)	%B	No.	Time (min)	%B	No.	Time (min)	%B
4J	0:00	2	**4K**	0:00	2	**4L**	0:00	2
	5:00	2		8:00	2		15:00	2
	10:00	20		10:00	10		20:00	20
	25:00	50		30:00	50		50:00	50
	30:00	100		35:00	100		55:00	100
	35:00	100		40:00	100		60:00	100
	40:00	2		45:00	2		65:00	2
4M	0:00	2						
	5:00	2						
	10:00	25						
	40:00	40						
	45:00	100						
	50:00	100						
	55:00	2						

Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

Table 4.3: MALDI-TOF MS Characterization of Peptides.

Peptide	[M+H]$^+$	
	Calculated	Found
Ac-L**A**^{S}AKAG**C**AKμAG-NH_2 (**4-1**)	1262.59	1262.63
AKAG**A**AKμAG-NH_2 (**4-1***)	988.53	988.46
Ac-L**A**^{S}AKAG**A**AKμAG-NH_2 (**4-2**)	1230.62	1230.78
Ac-μTTEAVDA**C**T**A**^{S}AK-NH_2 (**4-3**)	1482.62	1482.42
Ac-μTTEAVDA**C**^{StBu}T**A**^{S}AK-NH_2 (**4-3b**)	1570.65	1570.74
Ac-μTTEAVDA**A**T**A**^{S}AK-NH_2 (**4-4**)	1450.64	1450.80
Ac-μTTEAVDA**A**T**A**AK-NH_2 (**4-5**)	1434.67	1434.81
Ac-μTTEAVDA**S**T**A**^{S}AK-NH_2 (**4-6**)	1466.64	1466.74
Ac-μTTEAVDA**C**T**A**AK-NH_2 (**4-7**)	1466.64	1466.80
Ac-μTTEAVDA-$SC_2H_4COOCH_3$ (**4-8a**)	1117.40 ([M+Na]$^+$)	1117.45 ([M+Na]$^+$)
CT**A**^{S}AK-NH_2 (**4-8b**)	508.23	508.37
Ac-μTTEAVDA**ψ**T**A**^{S}AK-NH_2 (**4-9**)	1482.62	1482.56
$GB1_{1-23}$ L^{S}_{5}-N_2H_3 (**4-10a**)	2512.89	2512.25
$GB1_{1-23}$ L^{S}_{5}-SPh (**4-10b**)	2591.01	2591.31
$GB1_{24-56}$ A_{24}C-OH (**4-11**)	3749.01	3748.85
GB1 $L^{S}_{5}A_{24}$C-OH (**4-12a**)	6228.84	6229.73
GB1 L^{S}_{5} (**4-12b**)	6196.78	6196.39

μ = 7-methoxycoumarinylalanine; ψ = penicillamine

Desulfurization for Thioamide Compatibility Studies Initial compatibility studies for desulfurization of thioamide containing peptides were carried out by Yanxin Wang with Peptide **4-1** (10 nmol, ε_{325} = 12,000 M^{-1} cm^{-1}). Peptides were either desulfurized

by using VA-044 or RANEY-Nickel and results were analyzed by analytical RP-HPLC on a Luna C8 analytical column using gradient **4J**. A more detailed description can be found elsewhere[120].

VA-044 Desulfurization for Condition Optimization Yanxin Wang performed condition optimization for VA-044 based desulfurization with regards to TCEP concentration, VA-044 concentration, oxygen tolerance, denaturant tolerance and side-reaction suppression. A detailed description of the experiments can be found elsewhere[120]. Peptide **4-3** (10 nmol, ε_{325} = 12,000 M^{-1} cm^{-1}) was used for all optimizations and reactions were analyzed by reverse phase HPLC on a Luna C8 analytical column using gradient **4K**. Fractions were collected, and then analyzed by MALDI-TOF MS and UV-Vis absorption spectroscopy. Yield was quantified by integrating peak areas in the HPLC chromatograms. The optimized standard condition is as follows: Peptide **4-3** was dissolved in 80 µL of argon-purged 1.25x buffer stock (50 mM TCEP, 125 mM thioacetamide, 125 mM Na_2HPO_4, pH 7.0). 10 µL of *t*-BuSH was added, followed by 10 µL of freshly prepared 10x VA-044 stock (100 mM in argon-purged water). The reaction was allowed to proceed at 37 °C under argon atmosphere. Upon completion the reaction was quickly quenched by chilling to 0 °C on ice. Excess *t*-BuSH was removed by a stream of argon, and then the crude mixture was diluted into H_2O for HPLC analysis or purification.

Desulfurization of Pen/Thioamide-Containing Peptide. Reaction was conducted using standard desulfurization procedure with thioacetamide. Briefly, Peptide **4-9** (10 nmol, ε_{325} = 12,000 M^{-1} cm^{-1}) was dissolved in 80 µL of argon-purged 1.25x buffer stock (50 mM TCEP, 125 mM thioacetamide, 125 mM Na_2HPO_4, pH 7.0). 10 µL of *t*-BuSH was added, followed by 10 µL of freshly prepared 10x VA-044 stock (100 mM in argon-purged water). The reaction was allowed to proceed at 37 °C for 10-60 min under argon

atmosphere. Upon completion the reaction was quickly quenched by chilling to 0 °C on ice. Excess *t*-BuSH was removed by a stream of argon, and then the crude mixture was diluted into H_2O and analyzed by reverse phase HPLC on a Luna C8 analytical column using gradient **4M**, and characterized by MALDI-TOF MS. After 60 minutes the reaction was complete as shown in Figure 4-4.

Native Chemical Ligation to obtain GB1 $L^S_5A_{24}$C-OH (4-12a). $GB1_{24-56}$ A_{24}C-OH (**4-11**) and $GB1_{1-23}$ L^S_5-N_2H_3 (**4-10a**) were aliquoted in 50 nmol aliquots using ε_{280} = 8,480 M^{-1} cm^{-1} and ε_{274} = 11,569 M^{-1} cm^{-1}, the sum of contributions from the thioamide (ε_{274} = 10,169 M^{-1} cm^{-1}) and tyrosine (ε_{274} = 1,400 M^{-1} cm^{-1}), respectively. Activation buffer (200 mM Na_2HPO_4, 6 M GnHCl, pH 3.0) and ligation buffer (200 mM Na_2HPO_4, 6 M GnHCl, pH 7.0) were degassed by purging with argon for 15 min. C-terminal fragment **4-11** was dissolved in 48 µL ligation buffer and 2 µL thiophenol added. N-terminal hydrazide peptide **4-10a** was dissolved in 45 µL activation buffer and stirred at -15°C for 5 min. 5 µL of a 0.5 M $NaNO_2$ solution in activation buffer was added and stirred at -5°C. After 15 min reaction allowed to warm to room temperature for 1 min, C-terminal peptide added and pH was adjusted to 7.0. Reaction was stirred overnight at room temperature for 19 h. Analytical HPLC runs were performed on a Luna C8 column using gradient **4L**.

Desulfurization to obtain GB1 L^S_5–OH (4-12b). To the crude ligation reaction, 400 µL degased desulfurization buffer (200 mM Na_2HPO_4, 6 M GnHCl, 500 mM TCEP, 250 mM Thioacetamide, pH 7.0) were added and spin-concentrated to about 80 µL using a Millipore Amincon Ultracel 3K Spin Filter Unit at 13.2k rpm for 60 min at 4°C. This procedure was repeated one more time. To 80 µL of buffer exchanged ligation product **4-12a**, 10 µL t-BuSH and 10 µL of a 0.5 M VA-044 solution in desulfurization buffer were added. The reaction was allowed to proceed at 37°C and was monitored by MALDI-TOF

MS (Table 4-3), since the desulfurized product **4-12b** has the same retention time as **4-12a**. After 23 h the reaction was complete, and the crude reaction mixture was diluted to 3 mL with 0.1% TFA in H2O and purified by reverse phase HPLC using gradient **4l**. Analytical HPLC runs were performed on a Luna C8 column using gradient 14. The product yield (71%) was determined from UV/Vis spectra of isolated products using ε_{274} = 19,736 M^{-1} cm^{-1}, the sum of contributions from the thioamide (ε_{274} = 10,169 M^{-1} cm^{-1}), tryptophan (ε_{274} = 5,357 M^{-1} cm^{-1}), tyrosine (ε_{274} = 1,400 M^{-1} cm^{-1}) and phenylalanine (ε_{274} = 5 M^{-1} cm^{-1}).

GB1 Sequences

Full Length Thioamide GB1 ($GB1_{1\text{-}56}$ L^{S}_{5}-OH, **4-12b**)

DTYKL^{S}ILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Thioamide GB1 N-Terminus Hydrazide ($GB1_{1\text{-}23}$ L^{S}_{5}-N_2H_3, **4-10a**)

DTYKL^{S}ILNGK TLKGETTTEA VDA-N_2H_3

Thioamide GB1 N-Terminus Thioester ($GB1_{1\text{-}23}$ L^{S}_{5}-SPh, **4-10b**)

DTYKL^{S}ILNGK TLKGETTTEA VDA-SPh

GB1 C-Terminus ($GB1_{24\text{-}56}$ A_{24}C-OH, **4-11**)

CTAEKVFKQY ANDNGVDGEW TYDDATKTFT VTE-OH

Full Length Thioamide GB1 Cys24 (GB1 $L^{S}_{5}A_{24}$C-OH, **4-12a**)

DTYKL^{S}ILNGK TLKGETTTEA VDACTAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Chapter 5 : The Effects of Thioamide Backbone Substitution on Protein Stability

Reprinted and adapted with permission from: Walters, C. R.; Szantai-Kis, D. M.; Zhang, Y.; Reinert, Z. E.; Horne, W. S.; Chenoweth, D. M.; Petersson, E. J., The effects of thioamide backbone substitution on protein stability: a study in *α*-helical, *β*-sheet, and polyproline II helical contexts. *Chemical Science* **2017**, *8*, 2868-2877[71]. Copyright 2017 The Royal Society of Chemistry

5.1 Introduction

The incorporation of functional molecules for monitoring and manipulating protein folding has greatly facilitated studies of biochemical phenomena[168]. The optimal choice of modification to a protein of interest depends on the type of process to be tracked, disrupted, strengthened, or isolated. Recently, there has been an emerging interest in making subtle alterations to the protein backbone to provide additional means to probe and mimic dynamic protein movements and interactions[165,169–172]. In particular, substituting a thioamide peptide linkage at single or multiple positions provides a small but powerful means to influence and/or monitor protein dynamics or function.

Thioamides (a.k.a. thionoamides, thioxoamides) have distinct properties from oxoamides, including a larger van der Waals radius, red shifted $\pi \rightarrow \pi^*$ and n $\rightarrow \pi^*$ transitions, lower oxidation potential, lower infrared (IR) stretching frequency and lower N–H pK_a[10,11,18,173,174]. These properties have allowed thioamides to be leveraged as photoswitches, Förster resonance energy transfer (FRET) acceptors, and photoinduced electron transfer quenchers, as well as IR and circular dichroism (CD) spectroscopic probes in biochemical systems[66,67,80,83,86,96,134,175,176]. Although thioamides have already proven to be valuable multifaceted probes, few studies have sought to elucidate the energetic impacts they have on peptide or protein folding. This consideration is of the utmost importance in assessing the thioamide's potential as a “minimalist” spectroscopic label.

Previous structural studies of thioamide substitutions largely relied upon short, isolated protein motifs in the absence of a tertiary structure context. In fact, when we began this work, there were no published studies of thioamides in proteins of defined tertiary structure. Fischer, Kiefhaber, and coworkers provided the most complete study of the

effects of thioamides in *α*-helices, using model helical polyalanine peptides bearing thioalanine (Ala^S) substitutions[96]. Their analysis revealed that the thioamide disrupts the helix if substituted at a central position, but is less destabilizing at the N-terminus. Their results suggest that the thioamide is similar to a glycine substitution at the same position in terms of helix disruption. Thioamides have also been incorporated into the 15-*mer* S peptide, which forms an *α*-helix upon binding to ribonuclease S[122]. Substitutions were made throughout the helix and were destabilizing at every position, with varying $\Delta\Delta G$ values in comparison to the WT complex (0.6–4.7 kcal mol^{-1}). However, in a different study, Miwa and coworkers observed that a Leu^S substitution in the central region of an *α*-helical coiled-coil dimer gave a similarly helical structure to the oxoamide counterpart (as observed by CD) and increased the melting temperature by 10°C[80]. This discrepancy, which likely arises from subtle geometric differences between an isolated *α*-helix and that in a coiled coil[177], implies that tertiary structure may serve to lessen or even reverse destabilization by a thioamide.

In an alternate secondary structure context, the thioamide was found to be tolerated between residues *i* + 2 and *i* + 3 in a type II' *β*-turn[66]. Here, the thiocarbonyl is solvent-exposed, and the interior of the turn does not require any reorganization to accommodate the larger sulfur atom. Thioamide substitutions have been made in a tryptophan-rich *β*-hairpin to interrogate the role of hydrogen bond formation in the folding transition state[67]. However, the results of this study may not be general to all *β*-sheets due to the highly engineered sequence of the tryptophan zipper peptide. Most recently, Raines and coworkers made thioamide substitutions in collagen model peptides (CMPs) to assess their impact in the Pro-Pro-Gly (PPG) repeats of an all PPG polyproline type II (PPII)

helix[123]. It was found that incorporation of Gly^S (*i.e.*, PPG^S) was significantly destabilizing, whereas substitution with Pro^S (*i.e.*, PP^{S}G) was slightly stabilizing to the triple helix.

Taken together, the above studies present a limited understanding about where thioamides can be utilized most efficiently in biological systems. Thioamides are likely to have much more nuanced effects in proteins with complex tertiary folds and dynamic regions. Thus far, only three full-length thioamide proteins have been described in the literature: the semi-synthetic constructs a-synuclein and dihydrofolate reductase, and the natural protein methyl-coenzyme M reductase[52,114,138]. Consequently, there have been no systematic studies of thioamides in folded proteins, which will be essential to guiding their future use as spectroscopic probes or as modulators of protein structure and function.

Herein, we describe an in-depth study of the effects of thioamide backbone substitution in three benchmark protein systems: (1) the C-terminal loop and helix of calmodulin (CaM), an *α*-helical protein; (2) the *β*-sheet of the B1 domain of protein G (GB1), a compact *α*/*β* tertiary structure; and (3) the PPII helix of a Pro-Hyp-Gly (POG) based CMP. For each protein, we performed CD spectroscopy studies to elucidate the structural and thermodynamic stability changes resulting from thioamide insertion. The first two systems represent new hosts for examination of the effects of thioamides on protein secondary structure in a tertiary fold context. In CMPs, our results build on prior published work through complete positional scanning substitutions in both PPG and POG subunits. Overall, the findings reported here lay the groundwork for the rational implementation of thioamides as biophysical probes in diverse protein systems.

5.2 Results and Discussion

Design and semi-synthesis of CaM thioproteins. CaM is a 148 amino acid, a-helical calcium signaling protein that is ubiquitous in eukaryotes. It is comprised of two structurally similar domains (N- and C-terminal), each containing two calcium binding sites[178]. Upon binding calcium, CaM undergoes a conformational change to expose a *trans*-domain helix, which acts as a binding platform for many of its regulatory target proteins[179,180]. Previously, a semi-synthesis of CaM was performed to modify the C-terminal EF-hand[181]. Since these authors observed some destabilization of CaM, we decided to limit thioamide substitutions in our investigations to the C-terminal loop and helix, near the highest affinity calcium binding site. Thus, we performed native chemical ligation (NCL) reactions between the expressed fragment $CaM_{1–134}$ and the synthetic fragment $CaM_{135–148}$. The native residue at position 135 in CaM is a Gln. Therefore, Cys at this position was masked after NCL to mimic Gln (denoted as Cys^{Q}) by iodoacetamide treatment. All CaM (thio)proteins were expressed, synthesized, purified, characterized, and analyzed by Christopher R. Walters. Details on purification are described elsewhere[71].

Positions for thioamide substitutions were selected with three criteria in mind: importance to local secondary structure, functional importance for Ca^{2+} binding, and ease of synthesis of peptides and thioamide precursors. Of particular interest were residues in the C-terminal helix of Ca^{2+}-bound CaM (Tyr_{138}, Glu_{139}, Glu_{140}, Phe_{141}, and Val_{142}). Glu_{140} is functionally important as it undergoes significant conformational change from the apo protein in order to directly chelate a Ca^{2+} ion in the holo protein[179]. Tyr_{138}, Glu_{139}, Phe_{141}, and Val_{142} reside in the N-terminal and central portions of the helix. We hypothesized that thioamide substitution at these sites would be destabilizing due to the weaker hydrogen bond acceptor capacity of the thiocarbonyl. Two additional sites were chosen to assess

the impact of substitutions in loops and solvent exposed areas proximal to this helix (Val_{136} and Ala_{147}, the penultimate residue in CaM). Proteins bearing each of these thioamide substitutions were synthesized by semi-synthesis using NCL. Wild-type (WT) CaM and a Cys_{135} mutant were expressed in *E. coli* as controls for proper refolding of the thioproteins. The Cys_{135} mutant was subjected to denaturing NCL conditions similar to the thioproteins and subsequently underwent the same purification and capping reaction.

CaM thioprotein folding thermodynamics. Each CaM variant was characterized by CD spectroscopy under two different buffer conditions: one for the calcium-bound holo protein (10 mM Tris pH 7.5, 2 mM $CaCl_2$) and another for the calcium-free apo form (10 mM Tris pH 7.5, 0.5 mM ethylenediaminetetraacetic acid, EDTA). CaM CD spectra show a prototypical α-helical signature with minima at 208 and 222 nm. Spectra for the thioprotein variants include a small minimum between 260 and 280 nm resulting from the thioamide $\pi \rightarrow \pi^*$ transition. This peak is weaker than the comparable transition observed in GB1 and collagen, or in previous reported small thiopeptides[66,80,96]. The attenuation of the thiocarbonyl CD signal is likely an environmental and concentration-based effect, as the thioamide in CaM is a small fraction of the overall amide content and the region containing the thioamide is destabilized in several variants. The thioamide $n \rightarrow \pi^*$ transition at 340 nm is not visible in any of the CaM CD spectra (see Figure 5-1).

Among the holo proteins, the spectrum of Cys^{Q}_{135} is similar in shape and magnitude to that of WT CaM, indicating the side chain modification necessary to enable NCL has no significant effect on the fold. Most of the thioamide substitutions appear well tolerated in the holo protein; however, the Tyr^{S}_{138} and Phe^{S}_{141} modifications considerably alter the folded state, with a nearly three-fold decrease in molar residue ellipticity (MRE, $[\theta]$) at 222 and 208 nm. These mutants show a similar helical signature to the N-terminal

truncation mutant CaM_{1-71}, suggesting that the C-terminal domain may be completely disordered as a result of thioamide incorporation. To understand more about the nature of each thioamide substitution, we subjected the CaM variants to thermal denaturation. Like WT CaM, each of the thioamide variants and the Cys^{Q}_{135} oxo control have melting temperatures greater than 80°C in the presence of Ca^{2+} and did not reach a plateau that allowed detailed analysis.

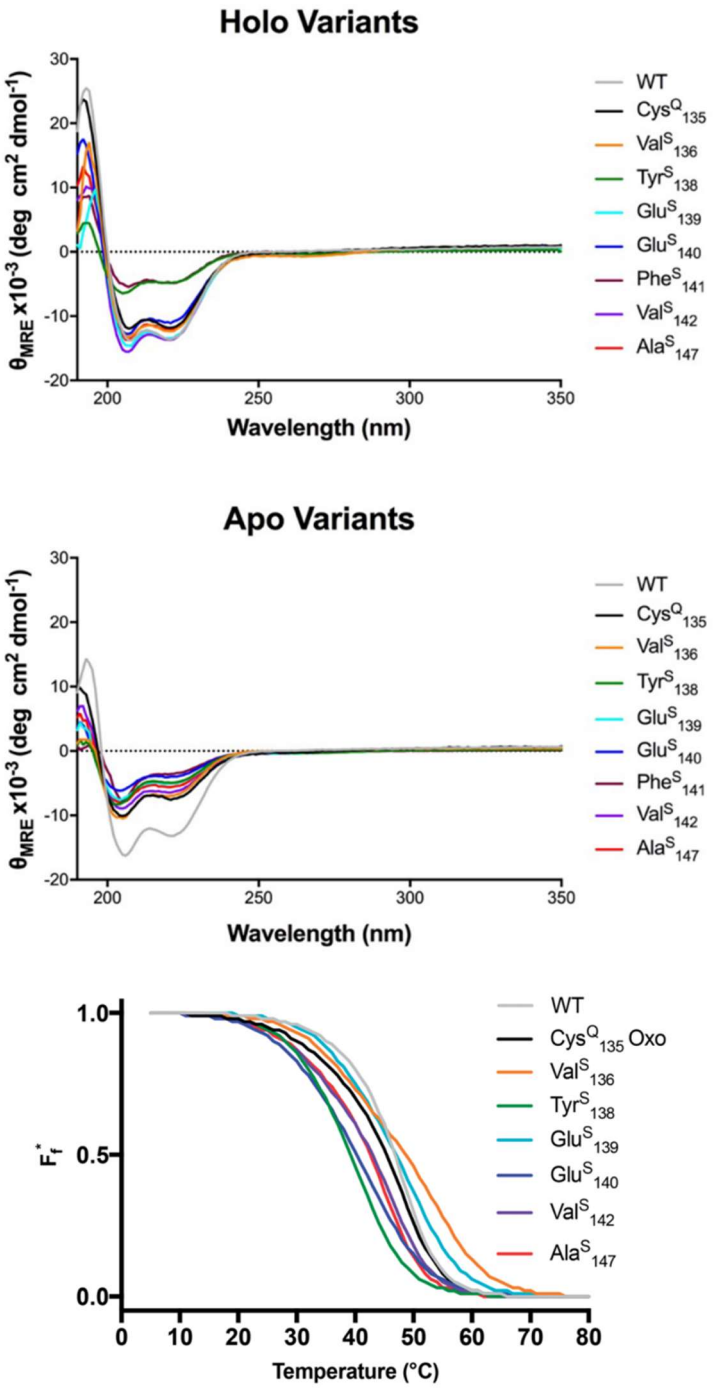


Figure 5-1: Characterization of WT, CysQ135 and thioCaM variants. CD wavelength scan of holo (top) and apo (middle) forms of CaM. Bottom: CD thermal melts of apo forms plotted as pseudo-Fraction folded (F^*_f) generated from fitted data.

Upon the removal of calcium by EDTA, CaM reverts to the apo structure, which lacks the stable transdomain helix and open EF hands observed in the holo form[182,183]. We hypothesized that the substantial structural differences between the apo and holo folds may lead a thioamide at a given site to have distinct effects in those two contexts. Thus, we carried out an analogous series of CD experiments to characterize folding in the apo CaM variants. Surprisingly, the Cys^{Q}_{135} modification appears to be significantly perturbing to the apo form, reducing helicity by half relative to apo WT CaM. While the origins of the effect of side chain modification at residue 135 are not clear, we used the data observed for the Cys^{Q}_{135} oxo control as a benchmark to probe the impact of thioamide incorporation on the apo fold.

Each of the apo proteins show complete unfolding transitions over the temperature range of 5 to 95°C. Initial apo thermal melts were performed on WT CaM and the Cys^{Q}_{135} control. The melting temperature for our WT CaM sample is lower than values previously reported in literature[184–187], likely due to different buffer conditions. Although we attempted to fit our data to a two-state model, we ultimately chose to analyze our results using a three-state model as it is well established that CaM unfolding proceeds through a semi-stable intermediate[187]. We therefore report thermodynamic data as ΔG_U, which is the sum of the individual unfolding free energies for each transition at 25°C (see Table 5.1). For comparison of three-state CaM unfolding to the two-state unfolding of GB1 and CMPs, we generated pseudo fraction folded (F^*_f) plots based on this three-state model, where the intermediate state makes a weighted contribution to the total fraction of folded protein. We denote the half-point of these weighted unfolding curves as T^*_M. Using this fitting model, we determined that Cys^{Q}_{135} CaM is slightly less stable than WT CaM in T^*_M and ΔG_U; however, this minor change in stability should not prevent it from serving as an appropriate

control to study the effects of the thioamide substitutions. A more detailed list including thermodynamic data for each transition (T_{M1}, ΔH_1, T_{M2}, and ΔH_2) can be found elsewhere[71].

Table 5.1: Apo CaM thermodynamic values.

CaM variant	T_M^*[a] (°C)	ΔT_M^* (°C)	ΔG_U[b] (kcal mol^{-1})	$\Delta\Delta G_U$ (kcal mol^{-1})
WT	46.8 ± 0.3	—	6.4 ± 0.2	—
Cys^{Q}_{135} Oxo	45.5 ± 0.4	—	5.9 ± 0.1	—
Val^{S}_{136}	48.6 ± 0.7	+ 3.2	6.5 ± 0.1	+ 0.6
Tyr^{S}_{138}	39.3 ± 0.8	- 6.2	4.1 ± 0.2	- 1.8
Glu^{S}_{139}	47.0 ± 0.1	+ 1.6	6.4 ± 0.1	+ 0.5
Glu^{S}_{140}	40.4 ± 0.8	- 5.1	3.9 ± 0.2	- 2.0
Val^{S}_{142}	44.0 ± 0.7	- 1.5	5.8 ± 0.2	- 0.1
Ala^{S}_{147}	43.7 ± 0.7	- 1.8	5.5 ± 0.2	- 0.3

[a]Melting temperature weighted for contributions from the unfolding of each domain.
[b]ΔG_U calculated from T_{M1}, ΔH_1, T_{M2}, and ΔH_2 as described elsewhere[71].

Using the Cys^{Q}_{135} protein as a basis for comparison, we see that Ala^{S}_{147} slightly destabilizes the apo fold, lowering ΔG_U by 0.3 kcal mol^{-1}. In agreement with the CD spectra, Tyr^{S}_{138} and Glu^{S}_{140} appear to be the most destabilizing mutations, decreasing ΔG_U by 1.8 and 2.0 kcal mol^{-1}, respectively. Val^{S}_{142} is well-tolerated, giving rise to a ΔG_U within error of the oxoamide. Unexpectedly, both Val^{S}_{136} and Glu^{S}_{139} show elevated melting temperatures and an increase of ΔG_U by ~0.5 kcal mol^{-1}. The unstable Phe^{S}_{141} construct was prone to aggregation during thermal denaturation, precluding rigorous thermodynamic analysis. Nevertheless, the thioamide substitutions that we could study span a range of thermodynamic effects, which we can interpret in terms of known

structures. A detailed discussion and analysis of the observed effects on thioamide substitutions can be found elsewhere[71].

Design and synthesis of GB1 thioproteins. The immunoglobulin-binding B1 domain of protein G from *Streptococcus* bacteria is 56 amino acids in length and has a tertiary structure comprised of three of the most common secondary structural motifs in proteins[160]. The compact GB1 fold consists of a protein-spanning *α*-helix packed against a four-stranded *β*-sheet with both parallel and anti-parallel strands, making it an ideal model to study local folding dynamics in a tertiary structure context. Indeed, NMR, X-ray crystallography, and an array of computational studies have been used to great effect in understanding the order in which individual secondary structure elements fold to form the final tertiary structure of GB1[188,189]. More recently, various unnatural backbone substitutions have been examined in GB1[166,190–192]. While these substitutions were thermodynamically destabilizing to the folded state, many of the proteins showed a native-like fold by X-ray crystallography and CD. Given these precedents, GB1 is an excellent platform to examine effects of thioamides on folding in a *β*-sheet, a structural motif not yet studied in detail with thioamide substitutions.

GB1 and three thioamide variants were synthesized using a combination of automated and manual SPPS, followed by NCL. We originally intended to synthesize thioamide GB1 through SPPS only, but encountered low yields due to suboptimal coupling reactions after thioamide insertion and Edman degradation-type cleavage of the amide bond C-terminal to the thioamide under acidic deprotection conditions. Thioprotein production by NCL allowed us to work with shorter thiopeptide fragments where the number of couplings subsequent to thioamide insertion and the length of the deprotection reaction can be limited. In the synthesis of GB1 and variants, we performed a ligation

between a thioamide containing $GB1_{1–23}$ thioester and $GB1_{24–56}Cys_{24}$. After desulfurization of Cys_{24} using VA-044 and sacrificial thioacetamide, we obtained the thioamide GB1 constructs with no trace of the ligation point and yields ranging between 10 and 20%[120].

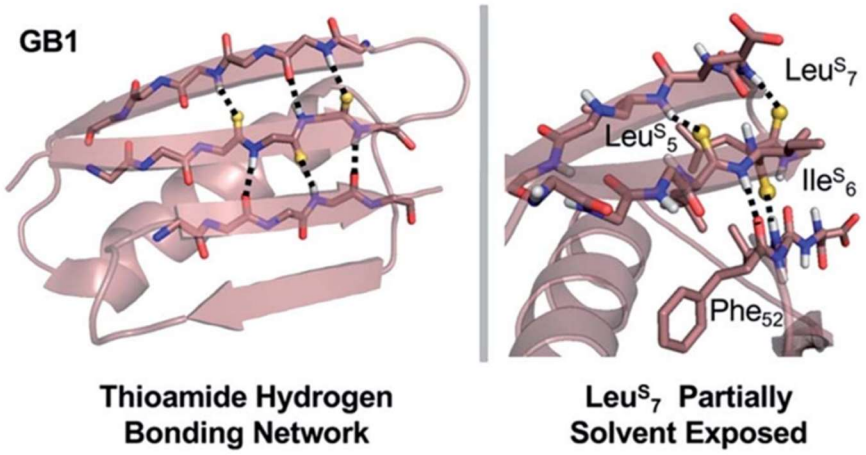


Figure 5-2: Structural analysis of GB1 thioamide variants. Left: Overall structure of GB1 with modeled thioamide substitutions. Right: Zoomed in region showing that the Leu$^S_7$ thiocarbonyl is skewed out of the plane of the anti-parallel *β*-sheet, providing a potential explanation for why it is less destabilizing than Leu$^S_5$ or Ile$^S_6$. Structures rendered from PDB entry 2QMT[189]

Using this NCL and desulfurization strategy, we synthesized three GB1 variants with single thioamide substitutions at Leu_5, Ile_6, or Leu_7. The Leu_5 residue acts as a hydrogen bond acceptor in an anti-parallel *β*-sheet interaction and as a hydrogen bond donor in a parallel *β*-sheet interaction. In contrast, Ile_6 acts as a hydrogen bond acceptor in a parallel *β*-sheet interaction and as a hydrogen bond donor in an anti-parallel *β*-sheet interaction. The hydrogen bonding pattern of Leu_7 is similar to that of Leu_5 (anti-parallel acceptor, parallel donor), but the carbonyl points outward slightly, increasing the hydrogen bond distance by 0.2 Å and placing it out of plane with respect to its partner in the sheet. Thus, we anticipated that the increased size and C=X bond length of the thiocarbonyl might be better accommodated at Leu_7 than at Leu_5 or Ile_6. These three substitutions,

while not comprehensive, provide examples of how thioamides may be accommodated into both types of *β*-sheet folds within the context of a complex tertiary domain.

GB1 thioprotein folding thermodynamics. GB1 CD spectra typically contain a broad minimum between 208 and 222 nm with few defined features in this range due to the combination of the *α*-helical and *β*-sheet contributions to the signal. The Leu^{S}_{5} and Ile^{S}_{6} substitutions have nearly identical spectra and maintain similar curve features as the GB1 Oxo control, but with significantly reduced MRE values. It is likely that the internal hydrogen bonding networks are perturbed by these substitutions, leading to a disruption in the packing of the tertiary structure. Leu^{S}_{7} seems to be the most tolerated thioamide substitution with a CD signature closest in magnitude to the control at 208 nm. However, the MRE values are still significantly reduced across the entire set of spectra.

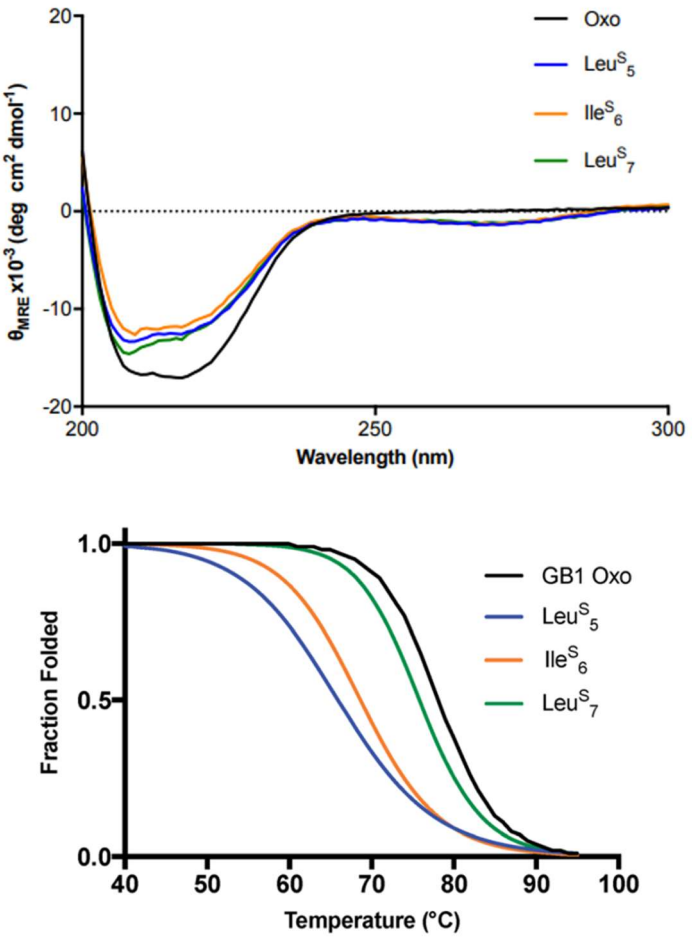


Figure 5-3: Characterization of WT and thio-GB1 variants. Top: CD wavelength scan of GB1 variants. Bottom: CD thermal melts plotted as Fraction folded (F_f) generated from fitted data.

For GB1, the thermal denaturation curves fit well to a two-state unfolding model. The T_M of GB1 Oxo is 78.2 C and includes a sharp unfolding transition that is preceded by only a modest initial unfolding event occurring gradually between 5 and 65°C. The thermodynamic differences for each thioamide substitution track with the changes in signal observed in the wavelength scans. Ile$^S_6$ leads to a 9.5°C destabilization in melting temperature while Leu$^S_5$ leads to an even greater destabilization of 12.6°C. Leu$^S_7$ is the least destabilizing thioamide substitution with a melting temperature only 2.3°C lower than WT GB1. The corresponding unfolding free energies ($\Delta\Delta G_U$ calculated according to

Equation 5[193]) show that Leu^{S}_{7} is destabilized by only 0.5 kcal mol^{-1}. Thus, we see that the precise context of thioamide substitution is also important for *β*-sheet systems.

$$\Delta\Delta G_U = \frac{\Delta T_M * \Delta H_{oxo}}{T_{M,oxo}}$$

Equation 5: Equation used to calculate $\Delta\Delta G_U$[193].

In GB1, the thermodynamic and structural changes from thioamide incorporation can be readily rationalized based on existing structural data (PDB entry 2QMT)[189]. Leu^{S}_{5} thiocarbonyl substitution likely disrupts the hydrogen bond with the N–H of Thr_{16}, altering packing of only the outer strand of the *β*-sheet (see Figure 5-3). The Leu^{S}_{5} thioamide N–H should maintain the hydrogen bond to the Phe_{52} carbonyl in the core parallel *β*-sheet interface and might even strengthen this interaction. Computational modeling indicates that Leu_5 forms part of the folding nucleus of GB1[188], so perhaps it is not surprising that the Leu_5 substitution has the most significant effect on the T_M and reduced the cooperativity of the folding transition. Ile^{S}_{6} reverses this hydrogen bonding pattern, maintaining the N–H hydrogen bond to the carbonyl of Gly_{14} in the outer strand and disrupting hydrogen bonding with the Val_{54} amide N–H in the core strand. Disturbance of the core of the *β*-sheet region would explain why the Ile^{S}_{6} substitution induces such a significant change in the T_M. Leu^{S}_{7} causes almost no disruption in folding as determined by CD. It seems that the longer hydrogen bond distance (3.1 Å to ˚ the Lys_{13}/Gly_{14} amide N) and out-of-plane orientation of the Leu_7 carbonyl are permissive of the sulfur substitution, as we anticipated. If there is any minor disturbance, it could be compensated for by the increased strength of the Leu^{S}_{7} thioamide N–H hydrogen bond with the carbonyl of Val_{54}. Taken together, these analyses reinforce our observation that the effects of

thioamide incorporation in *β*-sheets can depend dramatically on the location of the substitution (see Table 5.2 for detailed thermodynamic values).

Table 5.2: GB1 thermodynamic values.

GB1 variant	T_M (°C)	ΔT_M (°C)	$\Delta\Delta G_U$[a] (kcal mol^{-1})
WT	78.2 ± 0.0	—	—
Leu$^S_5$	65.6 ± 1.6	- 12.6	- 2.5 ± 0.3
Ile$^S_6$	68.7 ± 0.5	- 9.5	- 1.9 ± 0.1
Leu$^S_7$	75.9 ± 1.2	- 2.3	- 0.5 ± 0.2

[a]Change in unfolding free energy $\Delta\Delta G_U$, calculated from ΔT_M using Equation 5[193].

Design and synthesis of collagen model thiopeptides. Collagen provides an opportunity to explore the thioamide's potential as a modulator of key stabilizing hydrogen bonds and n→*π** interactions. The protein is comprised of three monomeric left-handed PPII helices that are made up of XaaYaa-Gly repeats. These monomers anneal to one another to form the native structure, a right-handed triple helix[194]. Although, ProProGly (PPG) is one of the most prevalent repeat elements in collagen, nearly every other amino acid has been observed in the Xaa or Yaa position[195]. Within this PPG unit, a variety of substitutions have been made on the 4 position of the pyrrolidine ring[196–201]. The most common alteration found physiologically is (2*S*,4*R*)-4-hydroxyproline (Hyp) at the Yaa position (Pro-Hyp-Gly, POG)[195]. Backbone ester, alkene, and aza-glycine substitutions have also been incorporated to modify and study the PPII triple helix[202–204]. Recently Raines *et al.* reported the effects of thioamide incorporation in two positions (ProPro^{S}Gly and ProProGly^S) by thermal denaturation in a PPG based CMP[123]. Here, we build on this precedent by providing detailed kinetics and thermodynamics for single substitutions in all Xaa, Yaa, and Gly positions in a 21-*mer* POG host peptide system.

To probe the effect of the thioamide on CMP self-assembly, we installed the moiety near the central positions of the 21-*mer* host system Ac-$(POG)_3(XYG)(POG)_3$-NH_2. Peptides were synthesized through SPPS, coupling Fmoc-POG-OH trimers at all but the central POG subunit. Here, individual residues were installed, including the suitable building block for incorporation of the thioamide. Typical yields of completed thioamide CMPs range between 4–6%. To analyze the thioamide impact on structure, thermodynamics, and kinetics of folding, we subjected each CMP variant to CD scans, thermal melts, and kinetic refolding experiments. All experiments with CMPs were performed by Yitao Zhang in the Chenoweth Lab at the University of Pennsylvania.

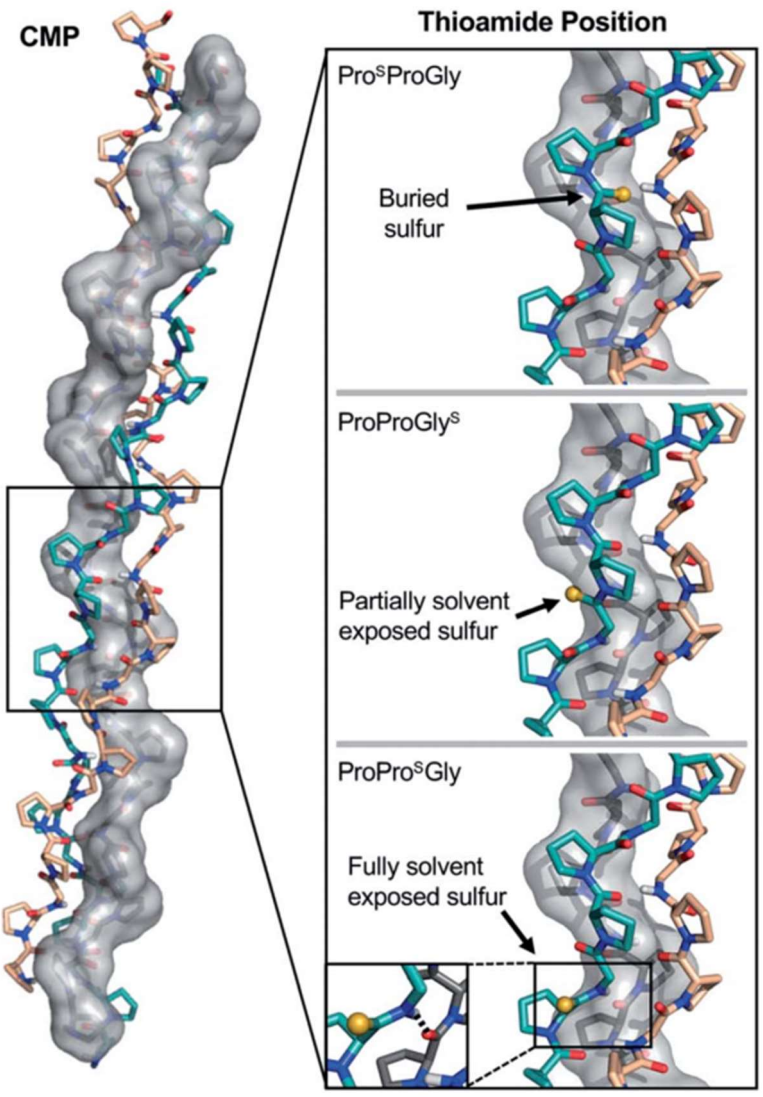


Figure 5-4: Structural analysis of CMP thioamide variants. Pro^{S}ProGly and ProProGly^{S} have buried and partially buried sulfur atoms, respectively (Top and Middle), accounting for the disruptions observed in the CD experiments. The ProPro^{S}Gly (Bottom) sulfur atom is fully solvent exposed and the N–H of the thioamide forms a hydrogen bond with the strand shown in grey (inset), accounting for the stabilization observed in the CD experiments. Structures rendered from PDB entry 2CUO[205].

Collagen model thiopeptide folding thermodynamics. CD scans of each thioamide variant display a characteristic collagen spectrum with a minimum at 198 nm and maximum around 225 nm, along with a broad minimum centered around 265 nm,

representing the $\pi \rightarrow \pi^*$ contribution from the thioamide. The magnitude of the minima at 198 nm and 265 nm depend upon the position of the thioamide, with the P^{S}PG substitution leading to the greatest loss in MRE signal. Unsurprisingly, the P^{S}PG replacement massively destabilizes the protein (Figure 5-5 and Table 5.3), likely because the sulfur points inward towards the core of the trimer, imposing a steric obstacle to the packing of the strands (see Figure 5-4). Additionally, the lower electronegativity of the elongated carbon–sulfur bond weakens key interchain H-bonding between the glycine N–H and the C=O of the Xaa proline. The PP^{S}G modification shows a stabilization of the triple helix with a significant increase in T_M and 0.4 kcal mol^{-1} increase in ΔG_U, while PPG^S shows a moderate destabilization, in agreement with previous studies[123]. Hysteresis studies showed that the free energy differences are in good agreement with the melting temperature profiles. Additionally, each thioamide CMP folded on a similar timescale to the PPG control, demonstrating that the kinetics of folding are not altered by thioamide substitution. Folding kinetics were carried out by Yitao Zhang in the Chenoweth Lab using CD spectroscopy.

Table 5.3: CMP thermodynamic and kinetic values.

Collagen variant	T_M (°C)	$t_{1/2}$ (min)	ΔG_U[a] (kcal mol^{-1})	$\Delta\Delta G_U$[a] (kcal mol^{-1})
Pro-Pro-Gly Oxo	36.7 ± 0.4	32.5 ± 2.0	10.8	—
Pro^S-Pro-Gly	11.0 ± 0.9	25.0 ± 4.5	5.1	- 5.7
Pro-Pro^S-Gly	38.5 ± 0.5	35.3 ± 5.2	11.2	+ 0.4
Pro-Pro-Gly^S	30.3 ± 0.5	40.3 ± 7.6	9.2	- 1.6
Pro-Hyp-Gly Oxo	39.9 ± 0.5	24.0 ± 4.5	12.0	—
Pro-Hyp^S-Gly	41.0 ± 0.2	17.7 ± 4.0	12.6	+ 0.6
Pro-Hyp-Gly^S	34.5 ± 0.0	23.8 ± 2.8	11.1	- 0.9

To assess thioamide backbone compatibility with substitutions at the 4 position of the pyrrolidine ring in proline, we incorporated Hyp into the Yaa position. This POG mutant is well known to stabilize collagen triple helices through stereoelectronic effects and hydrogen bonding[206–210]. Results show that peptides PO^{S}G and POG^{S} exhibit T_M increases of 2.5 and 4.2°C, respectively, relative to their corresponding Pro congeners. Interestingly, the stabilizing effects from the Hyp substitution and thioamide appear additive when compared to appropriate PPG and POG controls. Thioamide substitution at the central Yaa position contributes about 0.5 kcal mol^{-1} of stability and Hyp substitution contributes about 1.3 kcal mol^{-1}. Furthermore, the PO^{S}G substitution leads to faster folding compared to that of any thioamide variant or control peptide examined here. Like for CaM and GB1, thioamide substitutions in collagen have effects that strongly depend on the position of incorporation. Further combinations of thioamide and pyrrolidine modifications may also garner additive gains in collagen stability, or even act to compensate for other modications previously observed to be destructive.

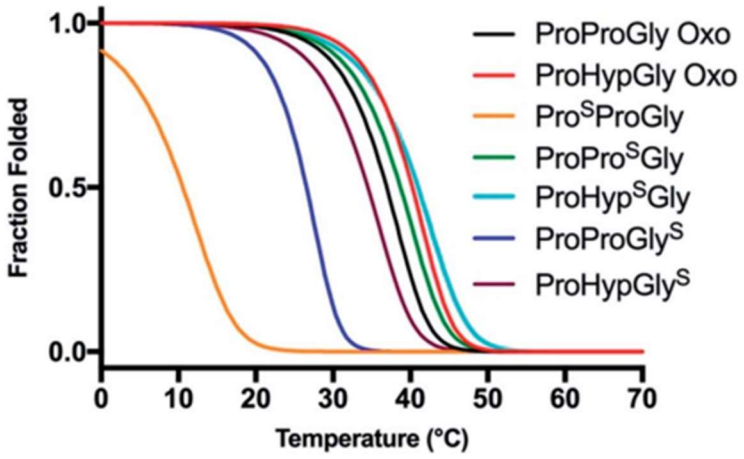


Figure 5-5: Thermal stability of CMP thioamide variants. CD thermal melts of collagen plotted as fraction folded (F_f) values.

Analysis of thioamide substitutions in PPG and POG CMPs illustrates the three fundamental effects that thioamides are expected to have on proteins. In PP^{S}G, the combination of weaker hydrogen bonding and steric clashes due to the larger van der Waals radius of sulfur and longer C=S bond length significantly destabilizes (ΔT_M = -25.7°C) the protein (see Figure 5-4). In contrast, both PP^{S}G and PO^{S}G collagen peptides are moderately stabilized compared to their oxoamide counterparts. At the Yaa position, the thioamide N–H can form a stabilizing interstrand hydrogen bond, while the thiocarbonyl projects into solvent. Thus, thioamide substitution at the Yaa position confers the energetic benefits of the stronger hydrogen bond donor without the penalties of the weaker acceptor or steric clashes. The disruption introduced by the PPG^{S} thioamide is not as drastic as the P^{S}PG replacement because it does not pack against the core of the triple helix, but it may provide general steric interference with neighboring strand contacts. The PO^{S}G and POG^{S} peptides recapitulate the positional effects of the thioamide while retaining the stabilizing effects of Hyp. The additivity of the backbone and sidechain substitutions supports the idea that the conformations of the thioamide collagen variants are similar to the parent collagen model peptides (with the exception of P^{S}PG, for which few conclusions can be drawn). In this regard, our data show that a thioamide substitution can act synergistically with a proline ring substitution for potential applications in thiopeptide based collagen materials.

5.3 Conclusion

Thioamides have the potential to be one of the most multifunctional probes in the large repertoire of unnatural amino acid substitutions available for protein labeling. They can be used as functional handles in IR, CD, or fluorescence spectroscopy, as photoswitches, or as perturbants in structure/function studies. With thioamide

substitutions across three protein systems, CaM, GB1, and CMPs, the data set amassed here represents the most comprehensive study to date bearing on the question of how thioamides affect protein thermostability. While virtually all previous reports of thioamides in secondary structure contexts have found the modification to be disruptive, we show that, within CaM, the replacement is tolerated in some positions within the C-terminal a-helix as well as the preceding loop. While all thioamide substitutions in the *β*-sheet of GB1 were destabilizing, at one position the impact was minimal. Our CMP results are consistent with findings by Raines, and although we do not yet have high resolution structures, the additivity of the effects of thioamidation and proline hydroxylation suggests that our CMPs adopt a fold consistent with structures like PDB entry 2CUO, shown in Figure 5-4. Overall, our analysis based on existing crystal and NMR structures allows us to rationalize many of our findings in terms of the physical properties of the thioamide bond and to consider how both perturbing and non-perturbing thioamide locations might be useful. Thioamide incorporation in the first turn of an *α*-helix is a strong helix breaker, and it is disruptive in sheets where it replaces a short hydrogen bond that is in the plane of the sheet. These findings can be explained by the longer C=S bond length and larger van der Waals radius of sulfur, and highlight the potential of thioamides as tools to modulate protein folding. Thioamides will be tolerated in helices where the substituted amide oxygen participates in longer H-bonds and in sheets where the carbonyl bond is at a more acute angle. We have also seen that thioamides are tolerated, or even stabilizing, when the disruptive effect of the thiocarbonyl is compensated by the stronger hydrogen bond donation of the thioamide N–H. Minimally perturbing substitutions are ideal for fluorescence studies using the thioamide as a quenching moiety. Increased thermal stability of collagen $PP^{S}G$ and $PO^{S}G$ substitutions, as well as Val^{S}_{136} and Glu^{S}_{139} in CaM, highlight the fact that thioamides can serve as stabilizing backbone replacements for interrogating hydrogen bonding networks.

The results reported here have set a foundation for rational thioprotein design to realize the above benefits to biophysics and protein engineering. A growing database of thioamide “mutants” could, in conjunction with appropriate computational models, allow one to predict the effects of thioamide substitution to achieve the desired destabilizing, neutral, or stabilizing effects on a protein of interest. Efforts are underway in our laboratories to obtain high resolution structural and dynamic information on thioamide substitution. In addition to the generation of thioamide proteins by NCL, Hecht's recent cotranslational incorporation of thioamide dipeptides using mutant ribosomes offers a potentially more facile semisynthetic route to thioproteins[138]. Deciphering the mechanism of incorporation of the natural thioglycine residue in the archael methyl-coenzyme M reductase could also permit the *in vivo* biosynthetic incorporation of backbone thioamides[52]. As thioamide proteins become more synthetically accessible, the results reported here should provide valuable insights into the design of appropriate thioproteins for diverse applications

5.4 Materials and Methods

General Information. Fmoc protected amino acids and peptide coupling reagents as well as Rink amide and 2-chlorotrityl resins were purchased from EMD Millipore (Billerica, MA, USA) with the exception of Fmoc Hyp(tBu)-OH, which was purchased from Advanced Chemtech (Louisville, KY, USA). Hydrazide hydrate solution was purchased from Oakwood Chemical (Estill, SC, USA). Piperidine was purchased from American Bioanalytical (Natick, MA, USA). VA-044 was purchased from Wako Pure Chemical Industries (Osaka, Japan). Triisopropylsilane (TIPS) was purchased from Santa Cruz Biotechnology, Inc (Dallas, TX, USA). $GB1_{24\text{-}56}$ $A_{24}C$ was purchased from Genscript (Piscataway, NJ, USA). Restriction enzymes and chitin purification resin were purchased

from New England Biolabs (Ipswich, MA, USA) and Pfu Turbo DNA Polymerase was purchased from Agilent Technologies (Santa Clara, CA, USA). DNA oligomers were purchased from Integrated DNA Technologies, Inc (Coralville, IA, USA). The pTXB1 plasmid containing $CaMS_{17}C_{,1\text{-}116}$ was a gift from the Linse Laboratory of the Department of Biochemistry and Structural Biology at Lund University. Isopropyl *β*-D-1-thiogalactopyranoside (IPTG) was purchased from LabScientific, Inc (Highlands, NJ, USA). Protease inhibitor tablets were purchased from Roche Boehringer Mannheim (Indianapolis, IN, USA). Amicon Ultra centrifugal filter units were purchased from EMD Millipore (Billerica, MA, USA). Sodium 2-mercaptoethanelsulfonate (MES•Na) was purchased from TCI America (Portland, OR, USA). β-mercaptoethanol (βME) was purchased from Bio-Rad Laboratories (Hercules, CA, USA). FPLC purification of proteins was performed on an AKTA system with HiTrap Q columns (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA), while peptide HPLC was performed on a Varian Prostar system (currently Agilent Technologies, Santa Clara, CA, USA) for CaM and GB1 peptides and a Jasco HPLC instrument (Easton, MD, USA) for collagen. Thiopropyl sepharose 6b resin was also purchased from GE Healthcare (Princeton, NJ, USA). Cuvettes for CD spectroscopy were purchased from Hellma Analytics USA (Plainview, NY, USA). All other reagents, solvents, and materials were purchased from either Fisher Scientific (Pittsburgh, PA, USA) or Sigma-Aldrich (St. Louis, MO, USA) unless otherwise specified. Matrix assisted laser desorption/ionization mass spectrometry (MALDI) was performed on a Bruker Ultraflex III mass spectrometer (Billerica, MA). High Resolution Mass Spectrometry (HRMS) for small molecules were obtained on a Waters LCT Premier XE LC/MS system (Milford, MA, USA). Nuclear magnetic resonance spectra were obtained on a Bruker DRX 500 MHz instrument. Common abbreviations for all other chemicals are as follows: 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluoro-

phosphate (HATU), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU), N-methylmorpholine (NMM), isobutyl chloroformate (ICBF), N,N-diisopropylethylamine (DIPEA), trifluoroacetic acid (TFA), 1,2-ethanedithiol (EDT), hydroxybenzotriazole (HOBt), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), tris(2-carboxyethyl)phosphine (TCEP), and guanidinium hydrochloride (Gdn•HCl). All work on Calmodulin was performed by Christopher Walters. All work on Collagen model peptides was performed by Yitao Zhang in the laboratory of Prof. David Chenoweth at the University of Pennsylvania. Solid phase peptide synthesis of full-length thio-GB1 was done by Zachary Reinert in the laboratory of Prof. W. Seth Horne at the University of Pittsburgh. Experimental details for those experiments is described elsewhere[71].

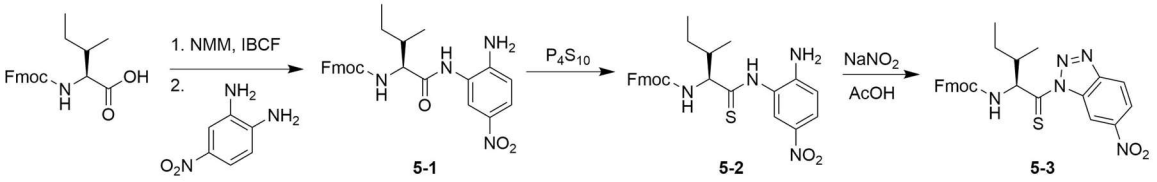


Figure 5-6: Synthesis of Thio-Isoleucine nitrobenztriazolide monomer.

Fmoc-Leu^{S}-Nbt was synthesized as previously described[72,120]. Fmoc-Ile^{S}-Nbt synthetic scheme is shown in Figure 5-6 and is described in detail here.

(9*H*-fluoren-9-yl)methyl ((2*S*,3*S*)-1-((2-amino-5-nitrophenyl)amino)-3-methyl-1-oxopentan-2-yl)carbamate (5-1). Fmoc-Ile-OH (3.53 g, 10.0 mmol) was dissolved in anhydrous THF (90 mL). N-methylmorpholine (2.20 mL, 20.0 mmol) was added and the reaction was cooled to -10 °C while the flask was purged with argon. Isobutyl chloroformate (1.30 mL, 10.0 mmol) was added dropwise while stirring. The reaction mixture was stirred for 15 min at -10 °C, then 4-nitro-*o*-phenylenediamine (1.53 g, 10.0

mmol) was added. The reaction was stirred under argon for 2 hours at -10 °C, then at room temperature overnight. After removing the solvent *in vacuo*, the residue was dissolved in DMF (50 mL). Upon addition of saturated aqueous KCl (250 mL) solution and Milli-Q water (250 mL), a yellow solid precipitated. The solid was filtered, washed with Milli-Q water and dried under vacuum. Compound **4** was isolated in 78.5% yield (3.84 g, 7.85 mmol) and 79.4% purity, as determined by HPLC. HRMS (ESI) m/z calculated for $C_{27}H_{28}N_4O_5$ $[M+H]^+$ is 489.2138, found 489.2130. ^{1}H NMR (500 MHz, C_2D_6OS): δ 9.54 (s, 1H), 8.29 (d, *J* = 2.7 Hz, 1H), 7.91 – 7.83 (m, 3H), 7.80 – 7.72 (m, 3H), 7.41 (t, *J* = 7.2 Hz, 4H), 7.32 (t, *J* = 7.5 Hz, 2H), 6.79 (d, *J* = 9.1 Hz, 1H), 6.52 (s, 2H), 4.35 – 4.20 (m, 4H), 4.09 (t, *J* = 8.2 Hz, 1H), 1.90 (ddt, *J* = 12.4, 8.5, 4.2 Hz, 1H), 1.56 (ddd, *J* = 13.7, 7.4, 3.3 Hz, 1H), 1.30 – 1.17 (m, 1H), 0.94 (d, *J* = 6.8 Hz, 3H), 0.88 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (126 MHz, C_2D_6OS): δ 171.2, 156.4, 148.8, 143.7, 140.7, 135.5, 127.6, 127.0, 125.3, 122.9, 121.2, 121.1, 120.1, 113.7, 65.8, 59.8, 46.7, 35.8, 24.7, 15.4, 10.7.

(9*H*-fluoren-9-yl)methyl ((2*S*,3*S*)-1-((2-amino-5-nitrophenyl)amino)-3-methyl-1-thioxopentan-2-yl)carbamate (5-2). Anhydrous Na_2CO_3 (0.239 g, 2.25 mmol) and P_4S_{10} (1.00 g, 2.25 mmol) were suspended in anhydrous THF (30 mL) in an oven-dried round bottom flask and stirred under argon for 30 minutes until solution became clear. Compound **5-1** (1.10 g, 3.00 mmol) was added and the reaction stirred over night at room temperature under argon. The next day the solvent was removed *in vacuo*. The resulting residue was resuspended in ethyl acetate and filtered over a pad of Celite® (Sigma-Aldrich, St. Louis, MO) to remove insoluble P_4S_{10} adducts. The filtrate was washed twice with 5% $NaHCO_3$ and once with brine. The aqueous layer was acidified with HCl and back extracted twice with ethyl acetate. The combined organic layers were dried over $MgSO_4$, filtered and the solvent was removed *in vacuo*. The crude product was redissolved in

dichloromethane and purified over silica to afford pure product as a yellow foam in 50.8% yield (0.770 g, 1.52 mmol). R_f = 0.5 in ethyl acetate/hexanes (1:1). HRMS (ESI) m/z calculated for $C_{27}H_{28}N_4O_4S$ $[M+H]^+$ is 505.1910, found 505.1909. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.57 (s, 1H), 7.98 (s, 1H), 7.95 (d, *J* = 8.9 Hz, 1H), 7.77 – 7.70 (m, 2H), 7.49 (d, *J* = 7.5 Hz, 1H), 7.44 (d, *J* = 7.4 Hz, 1H), 7.42 – 7.33 (m, 2H), 7.27 (s, 2H), 7.30 – 7.20 (m, 2H), 6.58 (d, *J* = 9.0 Hz, 1H), 5.70 (s, 1H), 4.27 (d, *J* = 7.7 Hz, 3H), 4.14 (t, *J* = 7.1 Hz, 1H), 2.10 – 2.02 (m, 1H), 1.29 – 1.20 (m, 2H), 1.02 (d, *J* = 6.7 Hz, 3H), 0.97 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 206.2, 157.4, 148.5, 143.3, 141.3, 138.4, 128.1, 128.0, 127.4, 127.3, 125.6, 125.0, 125.0, 124.9, 122.3, 120.3, 120.2, 115.3, 77.4, 67.6, 66.89, 47.0, 39.1, 29.83, 15.9, 11.1.

(9*H*-fluoren-9-yl)methyl ((2*S*,3*S*)-3-methyl-1-(6-nitro-1*H*-benzo[*d*][1,2,3]-triazol- 1-yl)-1-thioxopentan-2-yl)carbamate (5-3, Fmoc-Ile^S-NBt). Compound **5-2** (0.770 g, 1.52 mmol) was dissolved in 95% glacial acetic acid diluted with 5 % water (20 mL). The solution was cooled to 0 °C and $NaNO_2$ (156 mg, 2.26 mmol) was added slowly. Upon complete addition, the reaction mixture was stirred for 30 min at 0 °C, after which cold Milli-Q water (125 mL) was added. The precipitated orange solid was filtered, washed with cold Milli-Q water and dried *in vacuo*. The crude yield of the reaction was 57.7% (0.453 g, 0.88 mmol) and was used directly in solid phase peptide synthesis without further purification. R_f = 0.45 in dichloromethane. HRMS (ESI) m/z calculated for $C_{27}H_{25}N_5O_4S$ $[M+Na]^+$ is 538.1525, found 538.1508. ^{1}H NMR (500 MHz, $CDCl_3$) δ 9.70 (s, 1H), 8.46 (d, *J* = 8.7 Hz, 1H), 8.32 (d, *J* = 8.9 Hz, 1H), 7.76 (d, *J* = 7.6 Hz, 2H), 7.60 (d, *J* = 7.5 Hz, 2H), 7.40 (t, *J* = 7.5 Hz, 2H), 7.31 (q, *J* = 8.1 Hz, 2H), 6.22 (dd, *J* = 9.9, 6.2 Hz, 1H), 5.70 (d, *J* = 9.9 Hz, 1H), 4.49 (dd, *J* = 10.9, 6.9 Hz, 1H), 4.41 (dd, *J* = 10.9, 6.8 Hz, 1H), 4.22 (t, *J* = 6.9 Hz, 1H), 2.13 – 2.04 (m, 1H), 1.72 – 1.61 (m, 1H), 1.30 – 1.19 (m, 1H), 1.03 (d, *J* = 6.7

Hz, 3H), 0.91 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 209.5, 156.0, 149.5, 149.0, 143.6, 143.6, 141.3, 131.7, 127.7, 127.0, 127.0, 125.0, 124.9, 122.2, 121.5, 119.9, 112.8, 66.9, 65.7, 47.2, 41.1, 24.1, 16.2, 11.2.

Purification of $GB1_{24\text{-}56}$ Cys_{24} (5-4). Crude peptide **5-4** was purchased from Genscript (Piscataway, NJ). Crude peptide (20 mg) was dissolved in CH_3CN/H_2O (1:1) containing 20 µL of 0.5 M TCEP bond breaker™ and purified by reverse phase HPLC (gradient **5A**, Table 5.4) at a flow rate of 15 mL/min on a Grace Vydac C18 preparatory column. Individual fractions were characterized by MALDI- MS (see Table 5.5), and dried by lyophilization. Dried peptide was dissolved in CH_3CN/H_2O (1:1), quantified by UV/Vis (ε_{280} = 8,480 M^{-1} cm^{-1}), aliquoted into 100 nmol portions and lyophilized for later use.

Synthesis and Purification of $GB1_{1\text{-}23}$ -N_2H_3 (5-5). The N-terminal fragment of GB1 (**5-5**) was synthesized on a CEM Liberty 1 Automated Microwave Peptide Synthesizer (Matthews, NC) on a 100 µmol scale. The hydrazide resin was prepared according to Liu and coworkers[119] and the first residue was coupled manually. 20% piperidine in DMF was used as deprotection reagent, 0.5 M HBTU in DMF was used as the activating reagent, and 2 M DIPEA in *N*-methyl-2-pyrrolidone (NMP) was used as the activator base. Five molar equivalents of each amino acid were used for each coupling step. Ala_{23} was coupled twice using Method 1 (listed below). Leu_{12}, Glu_{19}, and Asp_{22} were coupled using a single coupling step as in Method 2. All other residues were coupled twice using Method 3. After synthesis was completed, a final deprotection was performed, consisting of an initial deprotection for 30 sec (35 W, 50 °C), followed by a 3 min deprotection (35 W, 50°C).

Method 1: Two consecutive couplings for 10 min each (25 W, 50°C) were used.

Method 2: Initial deprotection for 30 sec (35 W, 50°C), followed by a 3 min deprotection (35 W, 50°C). A single coupling for 10 min (25 W, 50°C) was used.

Method 3: Initial deprotection for 30 sec (35 W, 50°C), followed by a 3 min deprotection (35 W, 50°C). Two consecutive couplings for 10 min each (25 W, 50°C) were used.

Peptide cleavage from resin was performed by treating the resin with a cleavage cocktail (18:1:1 Trifluoroacetic acid(TFA):Triisoproylsilane(TIPS):H_2O) for 45 min. The solution was then collected by filtration, and dried by rotary evaporation. For purification, the crude residues were dissolved in CH_3CN/H_2O (1:1), diluted with H_2O, and then purified by reverse phase HPLC (gradient **5B**, Table 5.4) at a flowrate of 15 mL/min on a Grace Vydac C18 preparatory column. Individual fractions were characterized by MALDI-MS (Table 5.5), and dried by lyophilization. Dried peptide was dissolved in CH_3CN/H_2O (1:1), quantified by UV/Vis (ε_{274} = 1,400 M^{-1} cm^{-1}), aliquoted into 100 nmol aliquots, and lyophilized for later use.

Synthesis and Purification of $GB1_{1\text{-}23}$-N_2H_3 Thiopeptides (5-6 – 5-8). Residues 8-23 were synthesized on a peptide synthesizer on a 150 µmol scale as described for $GB1_{1\text{-}23}$-N_2H_3 (**5-5**). Resin was split into 3 equal parts and remaining amino acids were coupled manually. For each coupling, 5 equiv of amino acid and 5 equiv of HBTU were dissolved in DMF (2 mL), pre-activated for 1 min in the presence of 10 equiv of DIPEA, and then stirred with the resin for 30 min at room temperature. For deprotections, 2% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in DMF (2 mL) was stirred with the resin for 2 min. The vessel was then drained, the resin was washed, and the process was repeated twice more. For thioamide couplings, 2 equiv of either Fmoc-Ile^S-NBt or Fmoc-Leu^S-NBt were dissolved in dry DCM (1 mL), and stirred with the resin for 45 min in the presence of 2

equiv DIPEA. Upon completion of SPPS, resin was rinsed thoroughly with DCM and dried under vacuum. For cleavage, resin was treated with a cleavage cocktail (18:1:1 TFA:TIPS:H_2O) for 45 min. The solution was then collected by filtration, and dried by rotary evaporation. For purification, the crude residues were brought up in CH_3CN/H_2O (1:1), diluted with H_2O to 5% CH_3CN, and then purified by reverse phase HPLC using (gradient **5B**, Table 5.4) at a flowrate of 15 mL/min on a Grace Vydac C18 preparatory column. Individual fractions were characterized by MALDI-MS (Table 5.5), and dried by lyophilization. Dried peptide was dissolved in CH_3CN/H_2O (1:1), quantified by UV/Vis (ε_{274} = 11,569 M^{-1} cm^{-1}), aliquoted into 100 nmol portions and lyophilized for later use.

Native Chemical Ligation of $GB1_{24\text{-}56}$ Cys_{24} with $GB1_{1\text{-}23}$-N_2H_3. Activation buffer (200 mM Na_2HPO_4, 6 M Gdn•HCl, pH 3.0) and ligation buffer (200 mM Na_2HPO_4, 6 M Gdn•HCl, pH 7.0) were degassed by purging with argon for 15 min. 100 nmol of the C-terminal fragment (**5-4**) was dissolved in 95 µL ligation buffer and 5 µL thiophenol was added. Then, 100 nmol of the N-terminal fragment (**5-5**, **5-6**, **5-7**, or **5-8**) was dissolved in 90 µL activation buffer and stirred at -15 °C for 10 minutes. Next, 10 µL of a 1 M solution of $NaNO_2$ in activation buffer was added and the reaction was stirred for 15 minutes at -15 °C. After 15 minutes, the C-terminal fragment was added to the N-terminal fragment, the pH was adjusted to 7.0, and the solution stirred at room temperature overnight. Upon completion, the reaction mixture was diluted to a volume of 3 mL with H_2O, filtered through a 0.22 µm syringe filter and purified by HPLC (gradient **5C**, Table 5.4) at a flowrate of 3.5 mL/min on a Grace Vydac C18 semi-preparatory column. Individual fractions were characterized by MALDI-MS (Table 5.5), and pure fractions dried by lyophilization. Typical isolated yields for NCL reactions are 30 to 40 nmol (30-40% yield).

Desulfurization of Cys_{24}. Desulfurization buffer (6 M Gdn•HCl, 0.2 M Na_2HPO_4, 0.5 M TCEP, 0.2 M thioacetamide, pH 7.2) and water were degassed by purging with argon for 15 min. A 0.5 M solution of VA-044 was freshly prepared in degassed water. The GB1 NCL products were dissolved in 160 µL desulfurization buffer. Next, 20 µL of tBuSH and 20 µL of VA-044 solution were added and the microcentrifuge tube was put under Ar atmosphere, sealed with Parafilm and incubated in a 37 °C water bath for 10 hours. Once the reaction was complete, the tube was removed from the water bath and put on ice. The reaction was diluted with 450 µL H_2O, then 150 µL CH_3CN was added and the reaction mixture was purged with Ar for 10 minutes to remove excess tBuSH. The reaction was then diluted to 3 mL with H_2O and purified by reverse phase HPLC (gradient **5D**, Table 5.4) at a flowrate of 3.5 mL/min on a Grace Vydac C18 semi-preparatory column. Individual fractions were characterized by MALDI-MS (Table 5.5), and dried by lyophilization.

Table 5.4: HPLC Gradients for Peptide Purification and Characterization.

No.	Time (min)	%B	No.	Time (min)	%B	No.	Time (min)	%B
5A	0:00	2	**5B**	0:00	5	**5C**	0:00	2
	5:00	2		5:00	5		15:00	2
	10:00	20		10:00	20		18:00	25
	30:00	30		25:00	30		48:00	35
	32:00	30		27:00	100		52:00	35
	33:00	100		30:00	100		53:00	100
	38:00	100		42:00	5		58:00	100
	39:00	2					60:00	2
5D	0:00	2						
	15:00	2						
	18:00	15						
	48:00	45						
	52:00	45						
	53:00	100						
	58:00	100						
	60:00	2						

Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

Table 5.5: MALDI-TOF MS Characterization of Peptides.

Peptide	$[M+H]^+$	
	Calculated	Found
$GB1_{24\text{-}56}$ Cys_{24} (**5-4**)	3746.68	3746.74
$GB1_{1\text{-}23}$-N_2H_3 (**5-5**)	2495.33	2495.40
$GB1_{1\text{-}23}$ Leu^S_5-N_2H_3 (**5-6**)	2511.39	2511.37
$GB1_{1\text{-}23}$ Ile^S_6-N_2H_3 (**5-7**)	2511.39	2511.30
$GB1_{1\text{-}23}$ Leu^S_7-N_2H_3 (**5-8**)	2511.39	2511.36
GB1 Cys_{24} (**5-9**)	6212.78	6212.38
GB1 Leu^S_5 Cys_{24} (**5-10**)	6228.84	6228.71
GB1 Ile^S_6 Cys_{24} (**5-11**)	6228.84	6228.23
GB1 Leu^S_7 Cys_{24} (**5-12**)	6228.84	6228.90
WT Oxo GB1 (**5-13**)	6180.72	6181.22
GB1 Leu^S_5 (**5-14**)	6196.78	6197.19
GB1 Ile^S_6 (**5-15**)	6196.78	6197.27
GB1 Leu^S_7 (**5-16**)	6196.78	6198.87

Sample Preparation for CD. Dried peptides **5-13** – **5-16** were dissolved in 500 µL ligation Buffer (200 mM Na_2HPO_4, 6 M GnHCl, pH 7.0) and dialyzed against 2 L 20 mM Na_2HPO_4, pH 7.0. The dialysis buffer was exchanged for fresh buffer twice over a period of 24 hours. After dialysis, samples were concentrated to 300-400 uL, using Amicon Ultra centrifugal filter units (3 kDa MWCO).

Concentration Determination of GB1. The absorbance for each GB1 sample was measured at 274 nm and Beer's law was used to determine the concentration of each sample by using extinction coefficients of ε_{274} = 9,567 M^{-1} cm^{-1} and ε_{274} = 19,736 M^{-1} cm^{-1} for the GB1 Oxo and GB1 thioamide proteins, respectively.

CD Wavelength Measurements. Each sample (300 µL) was added to a 1 mm path length Helma Analytics 110-QS CD cuvette and loaded into a Jasco J-1500 CD Spectrometer. CD wavelength scans were taken at 25 °C by scanning at a 1 nm step between 350 nm and 190 nm with a bandwidth of 1 nm and an averaging time of 8 s per measurement. Blank spectra were obtained by measuring the CD signal of the relevant buffer in each cuvette with an averaging time of 2 s. The raw signal (θ_d, mdeg) from each sample was background subtracted against the blank buffer spectrum and converted to molar residue ellipticity (θ_{MRE}) using Equation 6, where c is the sample concentration, l is the path length and n_R is the number of residues.

$$\theta_{MRE} = \frac{\theta_d}{c * l * n_R}$$

Equation 6: Equation to calculate θ_{MRE}.

CD Thermal Melts. Thermal denaturation was performed by monitoring θ_d at 220 nm over a temperature range of 5 to 95 °C. The temperature increase was set to 0.2 °C per min with an equilibration time of 10 s. Measurements were made with an averaging time of 8 s per sample and a bandwidth of 1 nm.

Two-State Fitting of GB1 Thermal Melts. The two-state model assumes that there is a single transition between the folded and unfolded states, with no intermediate state populated. For this model, the folded and unfolded linear baselines were fit by using the low temperature $\theta_f = m_fT+b_f$ and high temperature $\theta_u = m_uT+b_u$ equations. The linear folded and unfolded baselines for GB1 Oxo were fit from 5 to 60 °C and 90 to 95 °C, respectively. The thioprotein folded and unfolded baselines were fit from 5 to 50 °C and 85 to 95 °C, respectively. The entire data set was then fit to Equation 7, where ΔH and ΔS are adjustable parameters and R = 1.9872 cal mol^{-1} K^{-1}. The fraction unfolded curve is

generated by minimizing the root mean square difference (RMSD) between F_{calc} and the experimental fraction unfolded against the parameters listed above over the entire temperature range.

$$\theta_{fit} = \theta_f(T)(1 - F_{calc}) + \theta_u(T)(F_{calc}) \quad F_{calc} = \frac{e^{-(\Delta H + T\Delta S)/RT}}{1+e^{-(\Delta H + T\Delta S)/RT}}$$

Equation 7: Equation used for two-state fit of CD data.

The F_{calc} fits of θ_{MRE} data and fraction folded (1-F_{calc}; F_f) plots for all constructs can be seen in Figure 5-7. The T_M of each construct was calculated by $T_M = \frac{\Delta H}{\Delta S}$, assuming the free energy $\Delta G_U = 0$ at the midpoint of the transition. The free energy at 25 °C ($\Delta G_U(25)$) was determined by $\Delta G = \Delta H - T\Delta S$ using the ΔH and ΔS calculated from the fits and T = 298.15 K. All data fitting was performed using the solver function in Microsoft Excel (Redmond, WA) and the plots were rendered in GraphPad Prism (LaJolla, CA).

Changes in free energy of unfolding between GB1 Oxo and the thioprotein variants were calculated using Equation 7 as described by Becktel *et al.* and previously used in GB1 studies[190,193]. ΔT_M is the difference in melting temperatures between GB1 Oxo and a particular thioprotein analogue. These calculations are made under the assumption that a single amino acid change only provides a small perturbation to the system and that the structural and functional characteristics of the system largely remain intact[193].

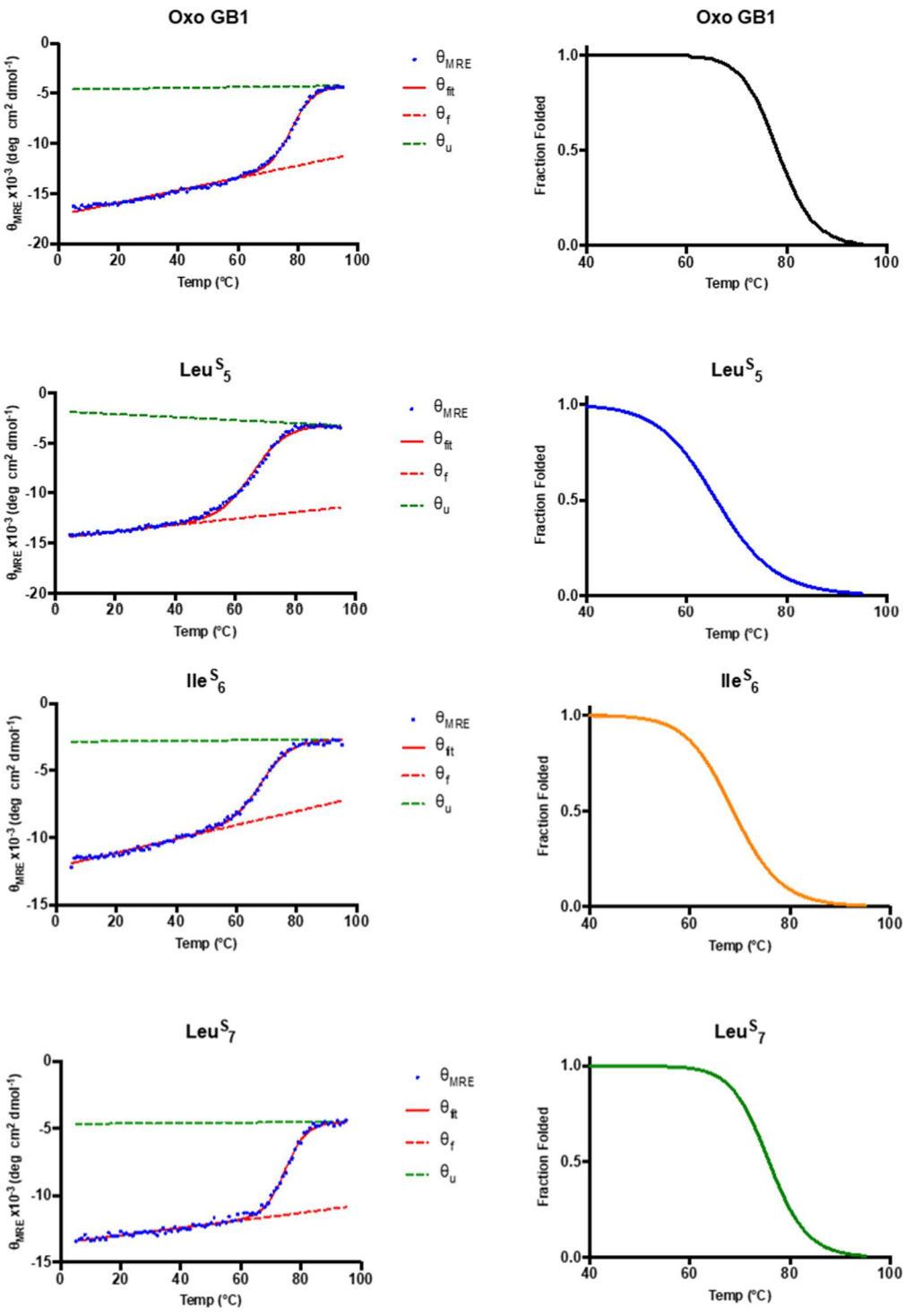


Figure 5-7: Plots generated using the two-state fitting method for GB1 variants. Left: For each GB1 variant, a single set of raw data (θ_{MRE}) and fits from equations for θ_{fit}, θ_f, and θ_u. Right: The averaged fraction folded plots (1-F_{calc}) from two replicated melts.

Sequences

GB1 C-Terminus (GB1$_{24-56}$ A$_{24}$C-OH, **5-4**)

CTAEKVFKQY ANDNGVDGEW TYDDATKTFT VTE-OH

GB1 N-Terminus Hydrazide (GB1$_{1-23}$ -N$_2$H$_3$, **5-5**)

DTYKLILNGK TLKGETTTEA VDA-N$_2$H$_3$

Thioamide (L$^S_5$) GB1 N-Terminus Hydrazide (GB1$_{1-23}$ L$^S_5$-N$_2$H$_3$, **5-6**)

DTYKL^{S}ILNGK TLKGETTTEA VDA-N$_2$H$_3$

Thioamide (I$^S_6$) GB1 N-Terminus Hydrazide (GB1$_{1-23}$ I$^S_6$-N$_2$H$_3$, **5-7**)

DTYKLI^{S}LNGK TLKGETTTEA VDA-N$_2$H$_3$

Thioamide (L$^S_7$) GB1 N-Terminus Hydrazide (GB1$_{1-23}$ L$^S_7$-N$_2$H$_3$, **5-8**)

DTYKLIL^{S}NGK TLKGETTTEA VDA-N$_2$H$_3$

Full Length GB1 Cys24 (GB1 A$_{24}$C-OH, **5-9**)

DTYKLILNGK TLKGETTTEA VDACTAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length Thioamide GB1 Cys24 (GB1 L$^S_5$ A$_{24}$C-OH, **5-10**)

DTYKL^{S}ILNGK TLKGETTTEA VDACTAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length Thioamide GB1 Cys24 (GB1 I$^S_6$ A$_{24}$C-OH, **5-11**)

DTYKLI^{S}LNGK TLKGETTTEA VDACTAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length Thioamide GB1 Cys24 (GB1 L$^S_7$ A$_{24}$C-OH, **5-12**)

DTYKLIL^{S}NGK TLKGETTTEA VDACTAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length GB1 (GB1$_{1-56}$ L$^S_5$-OH, **5-13**)

DTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length Thioamide GB1 (GB1$_{1-56}$ L$^S_5$-OH, **5-14**)

DTYKL^{S}ILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length Thioamide GB1 ($GB1_{1\text{-}56}$ I^{S}_{6}-OH, **4-12b**)

DTYKLI^{S}LNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Full Length Thioamide GB1 ($GB1_{1\text{-}56}$ L^{S}_{7}-OH, **4-12b**)

DTYKLIL^{S}NGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE-OH

Chapter 6 : Improved Semi-Synthetic Strategy for Thioamide-Incorporation into Proteins

6.1 Introduction

Previously we have observed that thioamides can perturb protein structure, specifically in *β*-sheet structures[71]. We were able to rationalize the observed changes in protein stability of thioamide-containing proteins based on available high-resolution structures of their cognate oxo-versions. However, the effects of dithioamide incorporation into proteins are less predictable and in some cases hard to rationalize based on available structural data[78]. High-resolution structures of thioamide-containing proteins would enable more detailed characterization and understanding of the effects of thioamide substitutions in proteins. Currently, most high-resolution structures of thioamides are small-molecule structures. While these structures provided valuable information on the geometry and some physicochemical properties of thioamides, they provide very limited information on their behavior in more complex settings of proteins. Thus far, the only protein structure that contains a backbone thioamide substitution is Methyl-Coenzyme M Reductase (MCR) from *Methanosarcina barkeri*[52].

We set out to scale up the production of thioamide-containing B1 domain of protein G (GB1). While we were able obtain sufficient quantities to perform protein stability measurements[71], the generation of well diffracting crystals for high-resolution structures usually requires screening of optimal conditions and therefore requires significantly larger quantities of protein. We chose GB1 as model protein for this due to its low polymorphism of observed structures and relatively high thermal melting temperature of thioamide-

containing derivatives[71], indicating that they might be stable for extended periods of time during crystallization. We previously observed low isolated yields of C-terminal acylhydrazide modified peptides and therefore decided to switch to a semi-synthetic approach, where a shorter N-terminal fragment was to be synthesized, while the C-terminal fragment would be expressed. After native chemical ligation (NCL) of the two fragments, we imagined the ligation site to be masked as an amino acid mimic to make the whole ligation 'traceless'.

6.2 Results and Discussion

We envisioned the generation of the C-terminal fragment of GB1 with an N-terminal cysteine through recombinant expression. The recombinant expression and purification of tag-less short protein fragments in high yield can be a challenging task. Various methods often employ fusion proteins and specific proteases to generate these fragments. Intein fusion proteins have been shown to be compatible with the generation of very short protein fragments of <40 amino acids[211]. We decided to use engineered inteins, which are inducible, self-excising protein domains, as purification handle. These have been widely recognized as useful tools in protein semi-synthesis and we have used them successfully in our lab[212–214]. Furthermore, we would rely on methionine aminopeptidase to expose N-terminal cysteine, as our lab has successfully utilized this technique in the past[69,215]. The ligation site we initially chose was the Gly_9-Lys_{10} amide bond for various reasons: First, the designated ligation bond is in the loop region between the first two *β*-sheets, making the ligation site fairly accessible. Furthermore, the size of the N-terminal amino acid at the ligation site has been shown to have an influence on reaction rate. Glycine as the smallest amino acid has been found to be the fastest reacting one[216]. Additionally, cysteine can be masked to mimic the cognate lysine by alkylating

cysteine with 2-bromoethylamine post ligation[144,217]. We hereafter denote the resulting thialysine as C^K.The full semisynthetic scheme is shown in Figure 6-1.

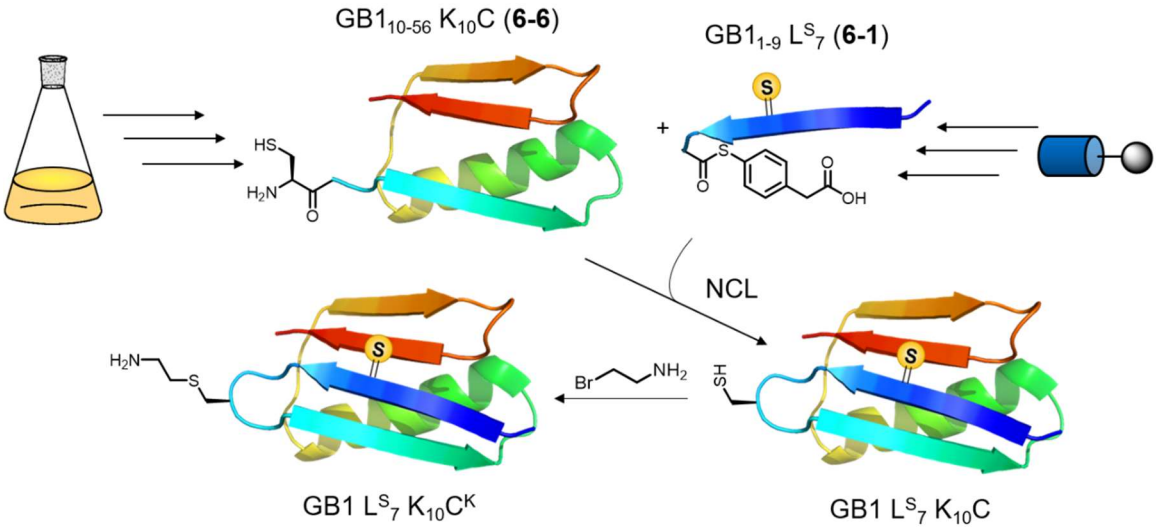


Figure 6-1: Semi-synthetic scheme to obtain thioamide labeled GB1. C-terminal fragment will be recombinantly expressed while the N-terminal fragment will be synthesized on solid phase. After NCL the cysteine at the ligation site will be alkylated with 2-bromoethylamine to yield thialysine.

Native chemical ligation between Gly_9 and Lys_{10}. We started with the synthesis of the N-terminal peptide GB1$_{1-9}$ L_7^S-N_2H_3. As before, we utilized Lei Liu's acyl hydrazide method, which serve as latent thioesters[71,119,120]. We noticed during synthesis that the main product of our peptide synthesis still contained a trityl group after cleavage. This is unusual, since trityl protecting groups are considered more acid labile than other protecting groups. Only two amino acids were incorporated as trityl protected precursors into the peptide: Gln_2 and Asn_8. MS-MS fragmentation analysis revealed that the trityl group was still attached to Asn_8 and not to Gln_2. Switching to 2,4,6-Trimethoxybenzoyl (Tmob)-protected Asn as a different protecting group resolved this problem and our desired peptide was the main isolated product.

The recombinant expression of the C-terminal fragment proved to be somewhat more challenging. We started by cloning full length GB1 into a pTXB1 vector, where it would be expressed with a C-terminal *Mycobacterium xenopi* GyrA Intein (MxeGyrA) and His_6-tag fusion. We used this construct as starting point to clone the NCL control construct (GB1 $K_{10}C$) and the N-terminal deletion construct ($GB1_{10\text{-}56}$ $K_{10}C$). Expression and purification of the full length GB1-MxeGyrA-His_6 fusion using standard protein expression conditions worked well. However, we were unable to induce the self-splicing reaction, even with variations of pH, temperature, denaturant and 2-mercaptoethanol (BME) concentration (see Figure 6-2).

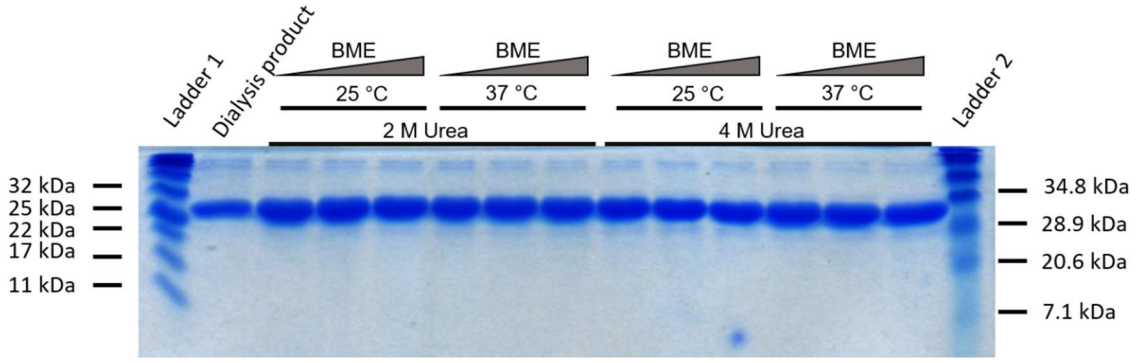


Figure 6-2: SDS-PAGE (16% Tris-Glycine gel) showing attempted intein cleavage of GB1-MxeGyrA-His_6 under varying conditions. BME concentrations used were 0.1 M, 0.25 M, and 0.5 M. Data shown here is for the reaction at pH 9.0. Gel at pH 8.0 looked identical. MW: GB1-MxeGyrA-His_6: 29.1 kDa; MxeGyrA-His_6: 22.9 kDa; GB1: 6.2 kDa.

Upon looking in the literature, we found that this observation is not unprecedented. The Iwaï lab has investigated the splicing efficiency of 20 different inteins with GB1 as one of their test substrates[218]. MxeGyrA was one of the least efficient inteins for GB1 with 0-10% splicing efficiency, depending on the condition. However, the second best performing intein with GB1 was the DnaE intein from *Nostoc punctiforme* (NpuDnaE) with about 90% spicing efficiency in their experiments. This was an intein we utilized previously in our lab[213] and were able to use to clone GB1-NpuDnaE-His_6 into a pTXB1 vector. Expression

and purification of the full length GB1-NpuDnaE-His_6 fusion using standard protein expression conditions worked well. After a preliminary test that cleavage was indeed occurring, we again varied BME concentration, temperature and pH to optimize cleavage efficiency. All cleavages were carried out for 24 hours and a representative gel is shown in Figure 6-3. Residual Intein-fusion protein was quantified based on band intensity. Temperature and BME concentration were the major contributors with pH having a minor effect. The best cleavage conditions concluded to be cleavage with 1 M BME at pH 8.5 and 37°C, leaving 61.6% of the intein-fusion protein intact after 24 hours.

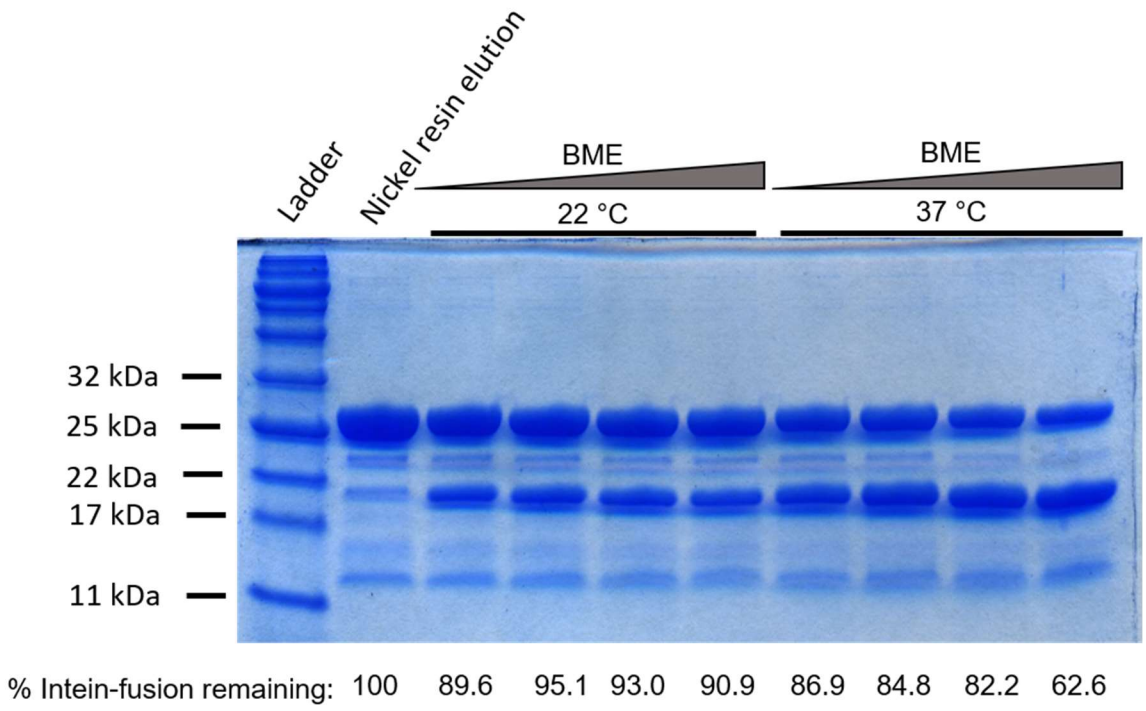


Figure 6-3: SDS-PAGE (16% Tris-Glycine gel) showing intein cleavage of GB1-NpuDnaE-His_6 under varying conditions. BME concentrations used were 0.1 M, 0.2 M, 0.5 M and 1.0 M. Data shown here is for the reaction at pH 7.5. Gels at pH 8.0, 8.5, and 9.0 looked similar. MW: GB1-NpuDnaE-His_6: 23.1 kDa; MxeGyrA-His_6: 16.9 kDa; GB1: 6.2 kDa. GB1 stains poorly and runs higher than expected at around 12 kDa.

Using the optimized intein cleavage conditions (1 M BME, 60 hrs, 37°C, pH 8.5), we were able to express and purify full-length WT GB1 from an NpuDnaE fusion in high

yield. The same procedure was applicable to the expression and purification of the NCL control construct GB1 $K_{10}C$. Unfortunately, the purification of the N-terminal deletion construct GB1 $\Delta_{1\text{-}9}$ $K_{10}C$ from NpuDnaE fusion was not successful as the intein did not cleave. Before testing intein fusions with NpuDnaE and MxeGyrA, we cloned two additional fusion proteins of GB1 $\Delta_{1\text{-}9}$ $K_{10}C$. It was previously reported that a 56 amino acid disordered region of MxeGyrA can be replaced with a $(GS)_4$-linker and would result in an intein we termed MxeGyrA(Δ) with altered cleavage capabilities[211]. Additionally, a T_3C mutation in MxeGyrA can make the intein less susceptible to spontaneous cleavage. We therefore also cloned GB1 $\Delta_{1\text{-}9}$ $K_{10}C$ as a fusion with MxeGyrA(Δ) and MxeGyrA(Δ, T_3C). Small scale expression and purification allowed testing to see which of these inteins are able to efficiently cleave (see Figure 6-4). Suprisingly, MxeGyrA and MxeGyrA(Δ) fusion constructs were deemed to be the most efficient. As to why a 9 amino acid deletion at the N-terminus causes a C-terminal intein-fusion protein to be cleaved while the full-length protein will not and *vice versa* is not intuitive and, without structural data, purely speculative. Further investigation of this phenomenon with attempts to crystalize various intein fusion constructs is currently underway.

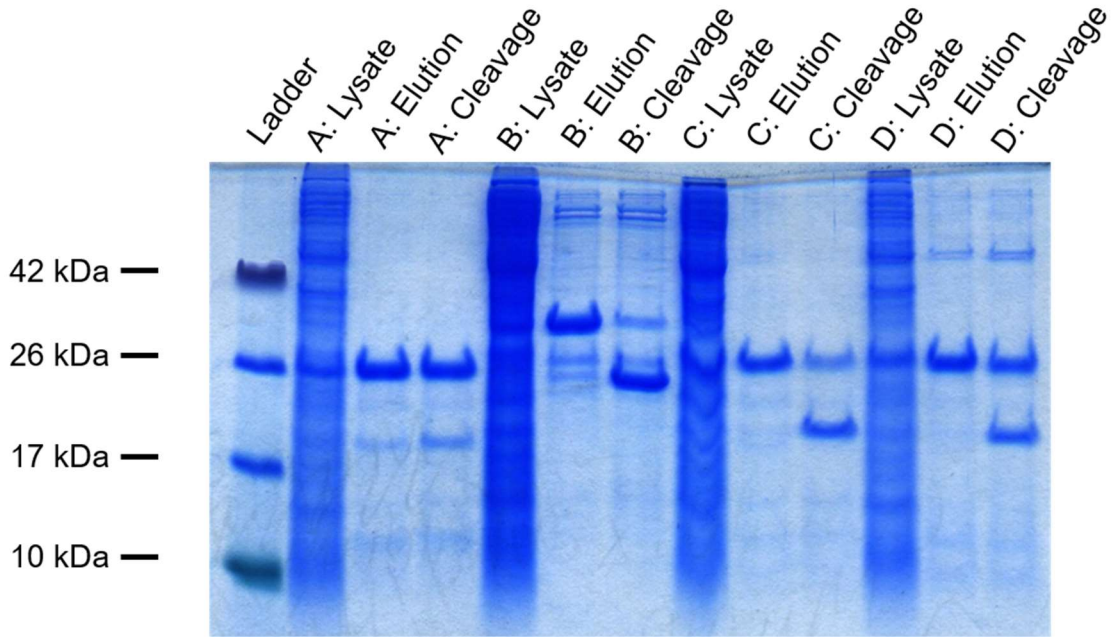


Figure 6-4: SDS-PAGE (14% Tris-Tricine gel) of intein cleavage test for $GB1_{10\text{-}56}$ $K_{10}C$-Intein-His_6 fusion constructs. All cleavages performed with 1 M BME at 37°C for 48 hrs. Inteins: A: NpuDnaE; B: MxeGyrA; C: MxeGyrA(Δ); D: MxeGyrA(Δ, T_3C)

With all controls and NCL fragments in our hands, we were finally able to attempt our ligation. To our surprise we found that the major products was a $GB1_{1\text{-}9}$ L_7^S dimer. Activation of acyl hydrazide as well as the reaction were carried out as in previously for other GB1 ligations[71,120] except for the use of 4-Mercaptophenylacetic acid (MPAA) instead of thiophenol. Figure 6-5 shows how the majority of the thioester peptide seemed to have reacted with itself. Although the N-terminus does not contain a cysteine, we hypothesize that the two fragments were able to align in a way where the N-terminal amine was in close enough proximity to undergo the nucleophilic attack. Interestingly the resulting dimer was only observed in its hydrolyzed form and not as thioester. Based on these results we decided to redesign our constructs and to move the ligation site.

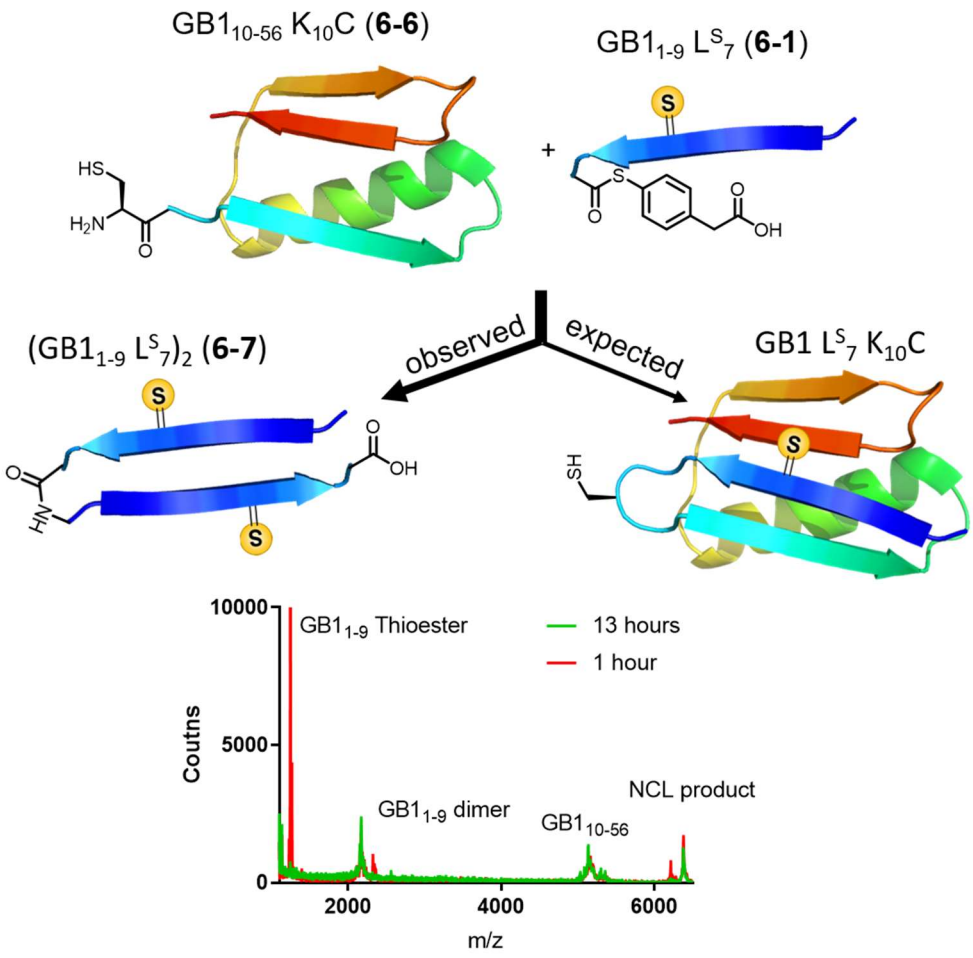


Figure 6-5: NCL with ligation site between Gly$_9$ and Lys$_{10}$. Top: Scheme showing the expected vs. observed reaction product. Botton: MALDI-TOF MS of NCL reaction showing formation of GB1 dimer with only minimal product formation.

Native chemical ligation between Leu$_{12}$ and Lys$_{13}$. We moved our ligation site over by three residues between residues Leu$_{12}$ and Lys$_{13}$. The ligation site is not ideal due to Leu as the C-terminal residue on the thioester fragment. Systematic studies on a model peptide have shown that ligation at Leu was amongst the longest with taking more than 48 hours to go to completion[216]. However, we hoped to avoid the association we observed previously and therefore prevent dimer formation, since the new disconnection site would be in the second *β*-sheet strand instead of in the loop (see Figure 6-6).

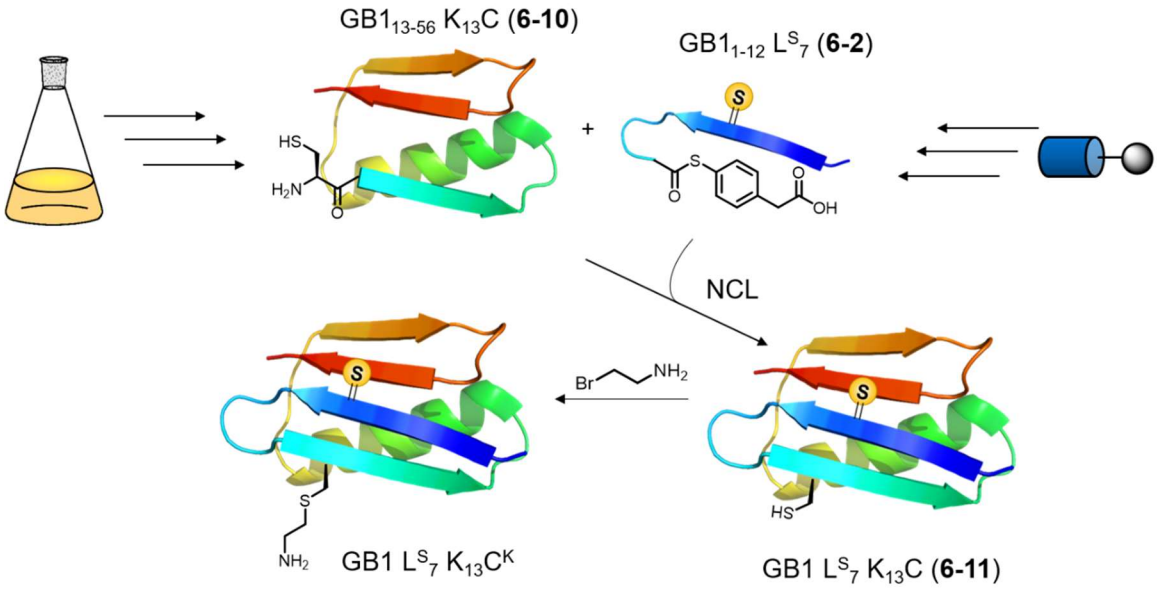


Figure 6-6: Semi-synthetic scheme to generate thio-GB1. Key difference to the previous scheme is the different ligation site to prevent the previously observed dimer formation.

We also switched our system to generate the C-terminal thioester in order to move away from Met aminopeptidase. The issue with Met aminopeptidase is that N-terminal cysteines can condensate with intracellular aldehydes and ketones, especially pyruvic acid, to form thiazolidines[215]. While this reaction is reversible through the treatment with methoxyamine at pH 4.0, this adds an extra step in the preparation of the protein[215]. Furthermore, it has been shown that Met aminopeptidase removes not only methionine, but also the penultimate amino acid, depending on the antepenultimate residue[215], which in our case would result in an unreactive protein fragment.

To achieve a higher sample homogeneity, we decided to switch to a protease-based method to generate our C-terminal fusion protein. Various proteases have been utilized for this purpose in the past, including Factor Xa and Tobacco Etch Virus (TEV) protease. Factor Xa cleaves the C-terminal of an IEGR recognition sequence while TEV protease recognizes the sequence ENLYFQG and cuts between Q and G, although it has

been reported that G can be replaced by C and used to generate N-terminal cysteine fragments[219,220]. However, both proteases have known off-target effects upon prolonged incubation and we have previously experienced protein loss through aggregation with Factor Xa[73]. We decided to switch to a rather new protease system that does not just rely on a recognition sequence, but also on tertiary structure[221]. Ulp-1 is a naturally occurring enzyme in yeast that removes Small Ubiquitin-like Modifier (SUMO), a post-translational modification, from amino acid side chains. SUMO-fusions have been utilized to stabilize proteins for structural investigation and as purification tag[222,223]. Ulp-1 can be used to selectively cleave SUMO fusion proteins as it will recognize the tertiary fold of SUMO and cleave after the last residues of SUMO as long as there is not proline residue in the next two positions.

We expressed and purified Ulp-1 as previously published without problems[222]. We cloned the gene for His6-SUMO-GB1 $\Delta_{1\text{-}12}$ $K_{13}C$ into a pET vector. Expression and purification went smoothly. Compared to the Met aminopeptidase-based method the SUMO/Ulp-1 pair gave protein in lower (but still good) yield (8.5 mg isolated protein / L expression), but in high homogeneity and in significantly shorter time, as it cut out several dialysis steps. The synthesis of $GB1_{1\text{-}12}$ L^{S}_{7}-N_2H_3 was carried out under identical conditions as the $GB1_{1\text{-}9}$ fragment.

Using fragments with the new ligation site, we carried out our NCL reaction. As expected, the reaction took significantly longer and was deemed to be complete after 72 hours, when no thioester peptide was detectable by MALDI-TOF MS (see Figure 6-7). During the reaction we observed significant drift in pH over time from 7.0 up to 7.6. While we readjusted the pH back to 6.9-7.0, the increased pH resulted in hydrolyzed peptide, lowering the overall yield. We initially attempted FPLC purification of the ligated protein,

but were not able to get sufficient separation between the ligated product and the unligated C-terminal fragment. HPLC purification on the other hand gave us good separation and allowed us to isolate our desired product in 21.2% yield (based on peptide as limited reagent). Further experiments for scale-up and aminoethylation are in progress.

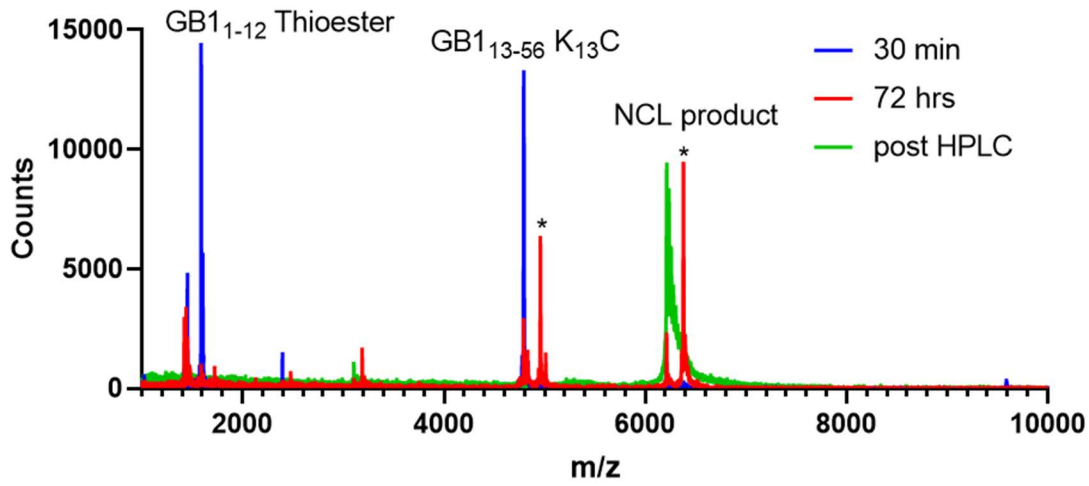


Figure 6-7: Progress of NCL reaction based on MALDI-TOF MS. Traces show reaction at the beginning (30 minutes; blue), at completion (72 hours, red), and post HPLC purification (green). * denotes MPAA disulfide adducts.

6.3 Conclusion

We have previously shown that incorporation of thioamides into *β*-sheet positions GB1 can have disruptive effect on the protein fold, but the extent of disruption is positionally dependent[71]. To provide further evidence for this hypothesis, high-resolution structures are necessary. We set out to synthesize thio-GB1 through semi-synthetic methods in order to be able to produce thioprotein in quantities sufficient for high-resolution structures by NMR and crystallography.

While our first choice of ligation site was not suitable, we were successful in redesigning and optimizing our semisynthetic scheme. Using a protease-based method to generate our C-terminal ligation fragment with an N-terminal cysteine, we were able to obtain it in good yield and high purity. While our overall isolated yield for the ligation was only modest, we hope to increase it in the near future by monitoring the pH more carefully, increasing the concentration of reactants in the ligation, and by performing the reaction at 37°C. Additionally, we still need to carry out aminoethylation reactions on both, the recombinantly expressed NCL control protein (GB1 $K_{13}C$) as well as the NCL product. We anticipate that these reactions will proceed smoothly as it has with the previous construct GB1 $K_{10}C^{K}$.

With this approach we anticipate that in the near future it will be possible to obtain mg quantities of thioamide labelled GB1 enabling high-resolution structures by NMR as well as X-ray crystallography.

6.4 Materials and Methods

General Information Fmoc protected amino acids, 2-chlorotrityl resin and peptide synthesis reagents were purchased from EMD Millipore (Billerica, MA, USA) or ChemImpex (Wood Dale, IL, USA). *N*α-Fmoc-L-Thioleucine-nitrobenztriazolide (Fmoc-Leu^{S}-Nbt) was synthesized as previously described[72]. Hydrazine Hydrate solution was purchase from Oakwood Chemical (Estill, SC, USA). Isopropyl *β*-D-1-thiogalacto-pyranoside (IPTG) was purchased from LabScientific, Inc. (Highlands, NJ, USA). cOmplete Mini, EDTA-free protease inhibitor tablets were purchased from Roche Diagnostics (Mannheim, Germany). High-Density Nickel Agarose Beads were purchased from GoldBio (St. Louis, MO, USA). Guanidine Hydrochloride was purchased from Invitrogen (Waltham, MA, USA). Amicon Ultra Spin Filters were purchased from EMD

Millipore. Slide-A-Lyzer G2 dialysis casettes were purchased from Thermo Fisher Scientific (Waltham, MA, USA). All other reagents, solvents and materials were purchased from Fisher Scientific (Pittsburgh, PA, USA) or Sigma-Aldrich (St. Louis, MO, USA) and used without further purification unless otherwise specified.

High resolution electrospray ionization mass spectra (ESI-HRMS) were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA, USA). Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. UV-Vis absorption measurements were performed on a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA) or Thermo Fisher Scientific Genesys 150 UV/Vis Spectrophotometer (Waltham, MA, USA). Since OD_{600} measurements are instrument specific (measured scattering intensity depends on the distance between sample and detector), values obtained on the Genesys 150 were multiplied by a factor of 1.5 (empirically determined) to match values obtained on HP 8452A. Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III or Microflex LRF instrument. Fast protein liquid chromatography (FPLC) was performed on an Äkta Explorer system (GE Healthcare, Chicago, IL, USA). HiTrap Q HP (5mL) anion exchange columns and HiLoad Superdex 75 pg 16/600 size exclusion chromatography column were purchased from GE Healthcare. Reverse-phase purification of crude peptides was performed on a Biotage Isolera System on Biotage SNAP Ultra C18 columns (Charlotte, NC, USA). Analytical HPLC was performed on an Agilent 1260 Infinity II series UHPLC system (Agilent, Santa

Clara, CA, USA). Preparative HPLC was performed on an Agilent 1260 Infinity II Prep HPLC system. HPLC columns were purchased from Phenomenex (Torrance, CA, USA). HPLC columns used were Analytical (150 x 4.6 mm), Semiprep (250 x 10 mm), and Prep size (250 x 21.2 mm) Luna® Omega 5 μm PS C18 100 Å columns from Phenomenex.

Synthesis of Peptides used in Native Chemical Ligation. Peptides were synthesized using standard Fmoc solid phase peptide synthesis (SPPS) procedure on 2-chlorotrityl chloride resin (100 - 200 mesh; 0.6 mmol substitution/g). Hydrazide resin for peptide synthesis was prepared according to Liu and coworkers[119]. For coupling, 5 equiv of amino acid and 5 equiv of HBTU or PyAOP were dissolved in DMF, pre-activated for 1 min in the presence of 10 equiv of DIPEA, and then stirred with the resin for 30 min at room temperature. For deprotection, 2% (v/v) 1,8- diazabicyclo[5.4.0]undec-7-ene (DBU) in DMF was stirred with the resin three times for 2 min each. Thioleucine (denoted Leu^S or L^S) was introduced through activated nitrobenztriazole precursor and was synthesized as previously described[71,72,120]. 2 equiv of the precursor was dissolved in anhydrous CH_2Cl_2, and stirred with the resin for 45 min in the presence of 2 eq DIPEA. Upon completion of SPPS, resin was rinsed thoroughly with CH_2Cl_2 and dried under vacuum. For cleavage, resin was treated with 5 mL of a cleavage cocktail solution (35% trifluoroacetic acid, 5% ethane-1,2-dithiol, 5% thioanisole, 2.5% triisopropylsilane, 2.5% water, and 50% dichloromethane; all percentages as v/v) for 60 min. The peptide was precipitated by the addition of cold diethyl ether. Various peptides were synthesized and/or purified by or with help of Kristen E. Fiore.

Synthesis and purification of $GB1_{1\text{-}9}$ L^S_7-N_2H_3 (6-1). Peptide was synthesized as described above on a 100 µmol scale. The crude peptide was dissolved in 1 mL dimethyl sulfoxide (DMSO) and purified by reverse phase chromatography on a Biotage SNAP

Ultra C18 (30g) cartridge using Gradient **6A** at a flowrate of 25 mL/min. Pure fractions were combined and dried by lyophilization. Mixed fractions containing the desired product as well as other peptide masses were combined, dried by lyophilization and re-purified by reverse phase HPLC on a Phenomenex Luna Omega Prep column using Gradient **6B** at a flowrate of 25 mL/min. Pure fractions were combined and dried by lyophilization.

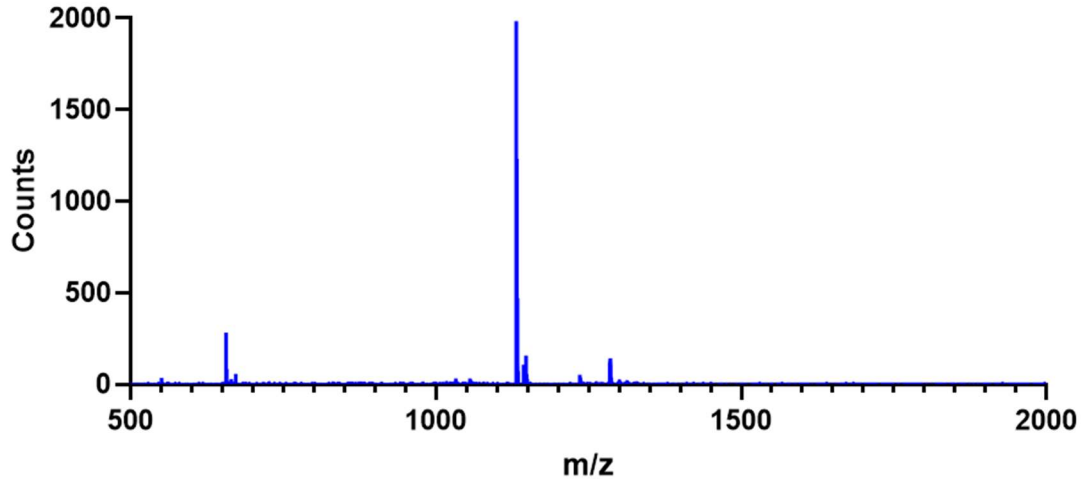


Figure 6-8: MALDI-TOF MS of purified GB1$_{1\text{-}9}$ L$^{S}_{7}$-N$_2$H$_3$ (6-1). [M+Na]$^+$: expected: 1,131.58; observed: 1,131.86.

Synthesis and purification of GB1$_{1\text{-}12}$ L$^{S}_{7}$-N$_2$H$_3$ (6-2). Peptide was synthesized as described above on a 100 µmol scale. The crude peptide was dissolved in 1 mL dimethyl sulfoxide (DMSO) and purified by reverse phase chromatography on a Biotage SNAP Ultra C18 (30g) cartridge using Gradient **6C** at a flowrate of 25 mL/min. Pure fractions were combined and dried by lyophilization. Mixed fractions containing the desired product as well as other peptide masses were combined, dried by lyophilization and re-purified by reverse phase HPLC on a Phenomenex Luna Omega Prep column using Gradient **6D** at a flowrate of 25 mL/min. Pure fractions were combined and dried by lyophilization.

We observed an inseparable mass adduct of +104 in our purified peptide. We analyzed this peak with m/z of 1,556 by MALDI-TOF/TOF MS. For our analysis we assumed that this is a modification of our desired peptide. We were able to assign 8 out of 12 b-ions, missing only b_1 (too small), b_7 (usually not observed for thioamides), b_9 and b_{12} (see Figure 6-9). However, we were not able to assign any of the expected y-ions to any of the observed peaks. This indicated that the modification appeared on the most C-terminal residue. Given that the C-terminal residues is Leu, we concluded that this mass adduct is modification of the C-terminal acyl-hydrazide. We deemed fractions containing this mass adduct as “pure”, since the modification will either be reversible, so that the peptide could be converted to a thioester in later steps, or irreversible, in which case it would be unreactive during later NCL reactions and can be purified away in later steps.

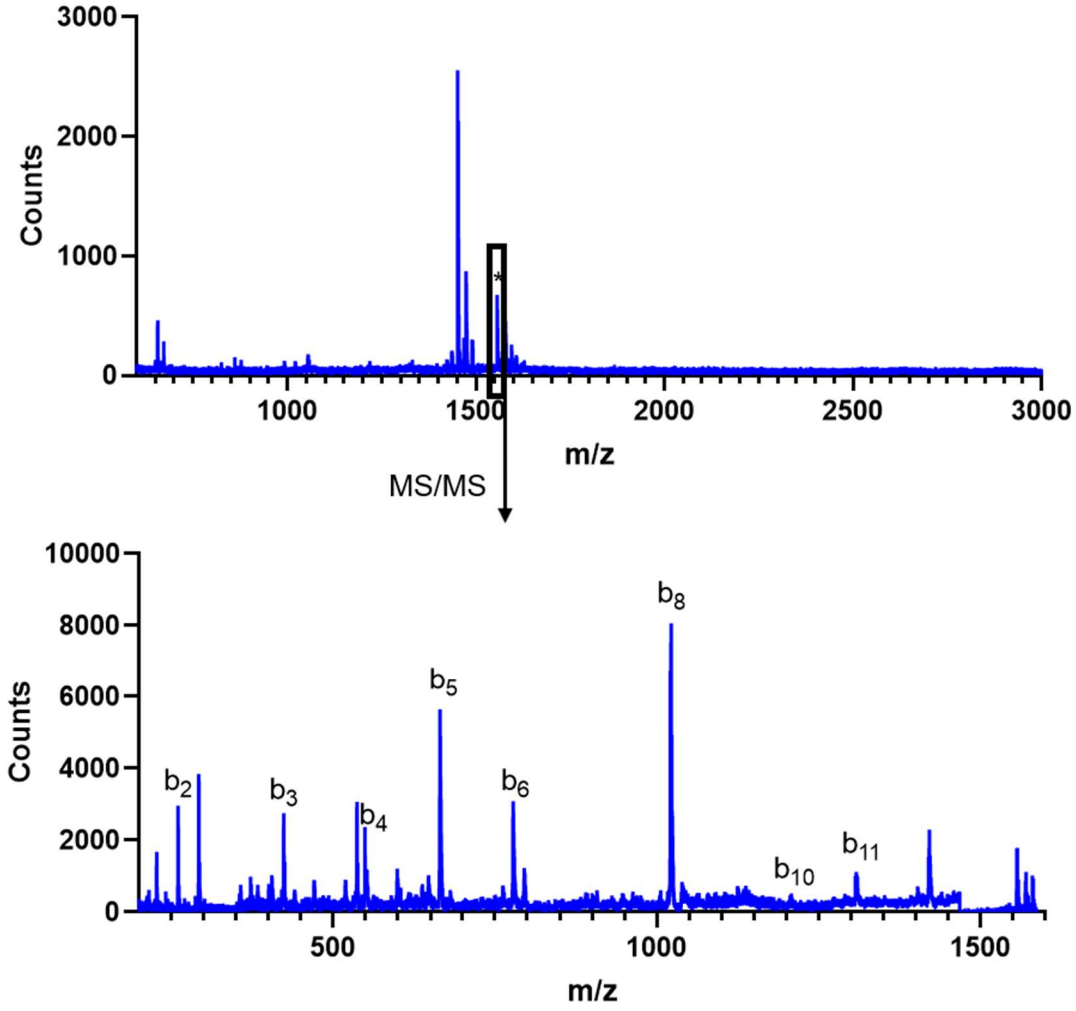


Figure 6-9: Top: MALDI-TOF MS of purified GB1$_{1-12}$ L$^{S}_{7}$-N$_2$H$_3$ (6-2). [M+H]$^+$: expected: 1,451.82; observed: 1,452.21. *denotes inseparable product with +104 mass difference. Bottom: MS/MS analysis of peak at m/z = 1,556 shows most expected b-ions of 6-2 up to b$_{11}$ but none of the expected y-ions, indicating modification of the C-terminal acyl-hydrazide.

General information about cloning. All PCR reactions were carried out on a T100 thermocycler from Bio-Rad (Hercules, CA, USA). Primers and double stranded DNA fragments (gBlocks) were purchased from Integrated DNA Technologies (Coralville, IA, USA). dNTPs, SYBR Safe DNA gel stain and Agarose were purchased from Invitrogen (Waltham, MA, USA). Phusion High-Fidelity DNA polymerase, Q5 High-Fidelity DNA

polymerase, Q5 High-Fidelity DNA polymerase Master Mix, HiFi DNA Assembly Master Mix, Restriction Enzymes, competent High Efficiency *E. coli* cells (NEB 5-alpha and NEB 10-beta), T4 DNA Polynucleotide Kinase, T4 DNA Ligase and Monarch Gel Extraction Kit were purchased from New England Biolabs (Ipswich, MA, USA). Gel Green DNA gel stain was purchased from Biotium (Freemont, CA, USA). DNA Clean & Concentrator Kit and Plasmid Miniprep Kit were purchased from Zymo Research (Irvine, CA, USA). DNA concentrations were determined with a TECAN Nanoquant plate on an Infinite M1000Pro plate reader (Tecan, Männedorf, Switzerland). DNA sequencing (Sanger sequencing) was performed at the University of Pennsylvania DNA Sequencing Facility (Philadelphia, PA, USA). Agarose gels were visualized on a Typhoon Imager (GE Healthcare, Marlborough, MA, USA) or a SmartBlue transilluminator (Southern Labware, Cumming, GA 30028). PCR of constructs containing $K_{13}C$ mutation was performed by Kristen E. Fiore. pET His6-SUMO-TEV (2S-T) plasmid was a gift from Scott Gradia (Addgene plasmid # 29711; http://n2t.net/addgene:29711; RRID: Addgene_29711). pFGET19_Ulp1 was a gift from Hideo Iwai (Addgene plasmid #64697; http://n2t.net/addgene:64697; RRID: Addgene_64697).

General procedure for PCR. Primers were dissolved in Milli-Q grade water to be at a concentration of 50 μM. The amount of plasmid used ranged from 0.1 – 1 ng total DNA for amplifications using Q5 polymerase and 0.5-10 ng for amplifications using Phusion polymerase. Primers, plasmid and buffer components were mixed according to manufacturer recommendations. PCR was carried out for 35 cycles at annealing temperatures calculated with the New England Biolabs Tm calculator tool (https://tmcalculator.neb.com). Alternatively, a ‘Touchdown protocol’ was used, where

annealing started with either 70° or 72°C in the first cycle and was reduced in each of the 19 subsequent cycles by 0.5°C each, followed by 15 amplifications at 55°C.

Following amplification, the PCR product was isolated using a DNA Concentrator & Cleanup Kit. PCR product was quantified to verify successful PCR amplification. PCR product was circularized by mixing 2 µL PCR product with 5 µL Milli-Q grade water, 1 µL 10x T4 DNA Ligase Buffer, 1 µL T4 Polynucleotide Kinase, 0.5 µL T4 DNA Ligase and 0.5 µL DpnI. This mixture was incubated at 37°C for 45 minutes, cooled on ice and 50 µL of competent NEB 5-alpha cells were added. Cells were transformed and plated on an LB-Agar plate containing appropriate antibiotic. Individual colonies were picked, grown up in 5 mL LB-medium and their DNA was extracted using a Plasmid Miniprep Kit according to manufacturer's protocol. Quantified DNA was submitted for sanger sequence analysis to verify correct sequence. Sequencing primer was provided by Sequencing Center and aligns with T7 promoter unless otherwise noted.

General procedure for Gibson Assembly. Primers were dissolved in Milli-Q grade water to be at a concentration of 50 µM. The amount of plasmid used ranged from 0.1 – 1 ng total DNA for amplifications using Q5 polymerase and 0.5-10 ng for amplifications using Phusion polymerase. Vector and insert were amplified simultaneously using a 'Touchdown protocol', where annealing started with 70° in the first cycle and was reduced in each of the 19 subsequent cycles by 0.5°C each, followed by 15 amplifications at 55°C. PCR product was purified using a 1% Agarose gel containing SYBR Safe or GelGreen Dye. Gel was visualized using a Blue light transilluminator and bands of the expected size were excised. DNA was extracted using a DNA gel extraction kit according to the manufacturer's protocol and isolated DNA was quantified. DNA assembly was performed with 100 ng of vector DNA and 2 equiv of insert DNA (5 equiv for insert <300

bp) using a HiFi DNA Assembly Mix according to manufacturer's protocol. 2 µL of the ligated product were transformed in 50 µL competent NEB10*β* cells according to manufacturer's instructions and plated on an LB-Agar plate containing appropriate antibiotic. Individual colonies were picked, grown up in 5 mL LB-medium and their DNA was extracted using a Plasmid Miniprep Kit according to manufacturer's protocol. Quantified DNA was submitted for sanger sequence analysis to verify correct sequence. Sequencing primer was provided by Sequencing Center and aligns with T7 promoter unless otherwise noted.

Cloning of pTXB1 GB1-MxeGyrA-His$_6$. Plasmid was constructed using Gibson Assembly. The Vector donor plasmid was pTXB1 αS-MxeGyrA-His$_6$ and construction of this plasmid is described elsewhere[224]. The insert was purchased as gBlock from IDT.

gBlock (5' to 3' strand only):
TTAACTTTAA GAAGGAGATA TACATATGCA GTATAAGTTA ATCCTTAATG
GGAAGACTCT GAAGGGGGAA ACAACAACTG AAGCGGTAGA CGCCGCCACG
GCAGAAAAGG TCTTTAAGCA GTATGCCAAT GATAATGGGG TTGATGGTGA
GTGGACTTAC GACGATGCGA CAAAGACGTT CACCGTTACC GAGTGCATCA
CGGGAGATGC ACTAGTTG

Forward primer for vector:
5′ – TGCATCACGG GAGATGCA – 3′

Reverse complement primer for vector:
5′ – ATGTATATCT CCTTCTTAAA GTTAAAC – 3′

Cloning of pTXB1 GB1(K_{10}C)-MxeGyrA-His$_6$. Plasmid was generated through site-directed mutagenesis PCR from plasmid pTXB1 GB1-MxeGyrA-His$_6$.

Forward primer:
5′ – TGTACTCTGA AGGGGGAAA – 3′

Reverse complement primer:
5′ – CCCATTAAGG ATTAACTTAT ACTGC – 3′

Cloning of pTXB1 GB1($\Delta_{2\text{-}9}$ K_{10}C)-MxeGyrA-His$_6$. Plasmid was generated through deletion PCR from plasmid pTXB1 GB1(K_{10}C)-MxeGyrA-His$_6$.

Forward primer:
5′ – TGTACTCTGA AGGGGGAAA – 3′

Reverse complement primer:
5′ – CATATGTATA TCTCCTTCTT AAAGTTAAAC – 3′

Cloning of pTXB1 GB1-PacI-MxeGyrA-His$_6$. Plasmid was generated through insertion PCR from plasmid pTXB1 GB1-MxeGyrA-His$_6$. The PacI restriction enzyme site was introduced between the DNA sequence of GB1 and the MxeGyrA Intein.

Forward primer:
5′ – TTAATGCATC ACGGGAGATG CA – 3′

Reverse complement primer:
5′ – TTAACTCGGT AACGGTGAAC GT – 3′

Cloning of pTXB1 GB1-PacI-NpuDnaE-His$_6$. Plasmid was generated through restriction enzyme cloning. The vector donor plasmid was pTXB1-GB1-PacI-MxeGyrA-His$_6$ and the insert donor plasmid was pTXB1-Ub-PacI-NpuDnaE-His$_6$ (construction of this plasmid is described elsewhere)[213]. Both plasmids were digested using PacI and NdeI in CutSmart buffer according to manufacturer recommendations. Vector and insert were purified using a 1% Agarose gel. After gel extraction DNA fragments were combined in a 1:1 ratio and ligated using T4 DNA Ligase. The sequence was confirmed by DNA sequencing analysis.

Cloning of pTXB1 GB1-NpuDnaE-His$_6$. Plasmid was generated by deleting the PacI restriction enzyme site from plasmid pTXB1 GB1-PacI-NpuDnaE-His$_6$ by the means of deletion PCR.

Forward primer:
5′ – TGCCTGAGCT ATGAAACGG – 3′

Reverse complement primer:
5′ – CTCGGTAACG GTGAACGT – 3′

Cloning of pTXB1 GB1(K_{10}C)-NpuDnaE-His$_6$. Plasmid was generated through site-directed mutagenesis PCR from plasmid pTXB1 GB1-NpuDnaE-His$_6$.

Forward primer:
5′ – TGTACTCTGA AGGGGGAAA – 3′

Reverse complement primer:
5′ – CCCATTAAGG ATTAACTTAT ACTGC – 3′

Cloning of pTXB1 GB1($\Delta_{2\text{-}9}$ K_{10}C)-NpuDnaE-His_6. Plasmid was generated through site-directed mutagenesis and simultaneous deletion PCR from plasmid pTXB1 GB1-NpuDnaE-His_6.

Forward primer:
5′ – TGTACTCTGA AGGGGGAAA – 3′

Reverse complement primer:
5′ – CATATGTATA TCTCCTTCTT AAAGTTAAAC – 3′

Cloning of pTXB1 GB1-MxeGyrA(Δ)-His_6. It was previously reported that MxeGyrA can be truncated to replace residues 107-160 with a $(GS)_4$ linker[211]. Plasmid was generated through simultaneous deletion and insertion PCR from plasmid pTXB1 GB1-MxeGyrA-His_6.

Forward primer:
5′ – GGTTCAGGCT CGTTCTACTA CGCGAAAGTC – 3′

Reverse complement primer:
5′ – ACTGCCGCTA CCTTGAATCA CCGCGTAATC – 3′

Cloning of pTXB1 GB1-MxeGyrA(Δ, T_3C)-His_6. It was previously reported that a T3C mutation in MxeGyrA can suppress premature hydrolysis of the thioether intermediate by tying up the cysteine responsible for thioester formation in a disulfide bond[211]. Plasmid was generated through site-directed mutagenesis PCR from plasmid pTXB1 GB1-MxeGyrA(Δ)-His_6.

Forward primer:
5′ – TGTGGAGATG CACTAGTTGC – 3′

Reverse complement primer:
5′ – GATGCACTCG GTAACGGT – 3′

Cloning of pTXB1 GB1($\Delta_{2\text{-}9}$ $K_{10}C$)-MxeGyrA(Δ)-His_6. Plasmid was generated through site-directed mutagenesis and simultaneous deletion PCR from plasmid pTXB1 GB1-MxeGyrA(Δ)-His_6.

Forward primer:
5′ – TGTACTCTGA AGGGGGAAA – 3′

Reverse complement primer:
5′ – CATATGTATA TCTCCTTCTT AAAGTTAAAC – 3′

Cloning of pTXB1 GB1($\Delta_{2\text{-}9}$ $K_{10}C$)-MxeGyrA(Δ, T_3C)-His_6. Plasmid was generated through site-directed mutagenesis and simultaneous deletion PCR from plasmid pTXB1 GB1-MxeGyrA(Δ, T_3C)-His_6.

Forward primer:
5′ – TGTACTCTGA AGGGGGAAA – 3′

Reverse complement primer:
5′ – CATATGTATA TCTCCTTCTT AAAGTTAAAC – 3′

Cloning of pTXB1 GB1($K_{13}C$)-MxeGyrA(Δ)-His_6. Plasmid was generated through site-directed mutagenesis PCR from plasmid pTXB1 GB1-MxeGyrA(Δ)-His_6.

Forward primer:
5′ – TGCGGGGAAC ACACAACTG – 3′

Reverse complement primer:
5′ – CAGAGTCTTC CCATTAAGGA TTAACTTATA – 3′

Cloning of pTXB1 GB1($K_{13}C$)-NpuDnaE-His_6. Plasmid was generated through site-directed mutagenesis PCR from plasmid pTXB1 GB1-NpuDnaE-His_6.

Forward primer:
5′ – TGCGGGGAAC ACACAACTG – 3′

Reverse complement primer:
5′ – CAGAGTCTTC CCATTAAGGA TTAACTTATA – 3′

Cloning of pTXB1 GB1($\Delta_{2\text{-}12}$ $K_{13}C$)-MxeGyrA(Δ)-His_6. Plasmid was generated through site-directed mutagenesis and simultaneous deletion PCR from plasmid pTXB1 GB1-MxeGyrA(Δ)-His_6.

Forward primer:
5′ – TGCGGGGAAC ACACAACTG – 3′

Reverse complement primer:
5′ – CATATGTATA TCTCCTTCTT AAAGTTAAAC AAAATTATTT CT – 3′

Cloning of pTXB1 GB1($\Delta_{2\text{-}12}$ $K_{13}C$)-NpuDnaE-His_6. Plasmid was generated through site-directed mutagenesis and simultaneous deletion PCR from plasmid pTXB1 GB1-NpuDnaE-His_6.

Forward primer:
5′ – TGCGGGGAAC ACACAACTG – 3′

Reverse complement primer:
5′ – CATATGTATA TCTCCTTCTT AAAGTTAAAC AAAATTATTT CT – 3′

Cloning of pET His_6-SUMO-TEV -EcoRI. pET His6-SUMO-TEV (2S-T) plasmid was a gift from Scott Gradia (Addgene plasmid # 29711 ; http://n2t.net/addgene:29711 ; RRID:Addgene_29711). One of the EcoRI restriction enzyme sites in an empty region of the plasmid about 400 bp 5' from the Amp-R promoter was deleted by deletion PCR to enable EcoRI specific restriction enzyme cloning in the future.

Forward primer:
5′ – TTCTTGAAGA CGAAAGGGC – 3′

Reverse complement primer:
5′ – TCATGTTTGA CAGCTTATCA T – 3′

Cloning of pET His_6-SUMO-$GB1_{13\text{-}56}$ $K_{13}C$. Plasmid was constructed using Gibson Assembly. The Vector donor plasmid was pET His6-SUMO-TEV -EcoR1. The insert donor plasmid was GB1($\Delta_{2\text{-}12}$ $K_{13}C$)-MxeGyrA(Δ)-His_6. Primers were designed to introduce 10 nt overlaps, i.e. the 5' and 3' and of insert contained 10 nt of the vector and *vice versa*. This created the necessary 20 nt overlap for Gibson assembly.

Forward primer for vector:
5′ – CGTTACCGAG TAACGGATCC GCGATCG – 3′

Reverse complement primer for vector:
5′ – TTTCCCCGCA CCCAGCAATC TGTTCTCT – 3′

Forward primer for insert:
5′ – GATTGGTGGG TGCGGGGAAA CAACA – 3′

Reverse complement primer for insert:
5′ – GGATCCGTTA CTCGGTAACG GTGAACG – 3′

Amino acid sequences of various fragments used:

GB1

MQYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK
TFTVTE

GB1 $K_{10}C$

MQYKLILNGC TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK
TFTVTE

$GB1_{10\text{-}56}$ $K_{10}C$

CTLKGETTTE AVDAATAEKV FKQYANDNGV DGEWTYDDAT KTFTVTE

GB1 $K_{13}C$

MQYKLILNGK TLCGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK
TFTVTE

$GB1_{13\text{-}56}$ $K_{13}C$)

CGETTTEAVD AATAEKVFKQ YANDNGVDGE WTYDDATKTF TVTE

MxeGyrA-His_6 (Note: Segment in red is replaced by $(GS)_4$ linker in MxeGyrA(Δ)-His_6)

CITGDALVAL PEGESVRIAD IVPGARPNSD NAIDLKVLDR HGNPVLADRL
FHSGEHPVYT VRTVEGLRVT GTANHPLLCL VDVAGVPTLL WKLIDEIKPG
DYAVIQRSAF SVDCAGFARG KPEFAPTTYT VGVPGLVRFL EAHHRDPDAQ
AIADELTDGR FYYAKVASVT DAGVQPVYSL RVDTADHAFI TNGFVSHATG
LTGLKLHHHH HH

NpuDnaE

CLSYETEILT VEYGLLPIGK IVEKRIECTV YSVDNNGNIY TQPVAQWHDR
GEQEVFEYCL EDGSLIRATK DHKFMTVDGQ MLPIDEIFER ELDLMRVDNL
PNIKIATRKY LGKQNVYDIG VERDHNFALK NGFIASAAFN HHHHHH

MxeGyrA(Δ)-His_6

CITGDALVAL PEGESVRIAD IVPGARPNSD NAIDLKVLDR HGNPVLADRL
FHSGEHPVYT VRTVEGLRVT GTANHPLLCL VDVAGVPTLL WKLIDEIKPG
DYAVIQGSGS GSGSFYYAKV ASVTDAGVQP VYSLRVDTAD HAFITNGFVS
HATGLTGLKL HHHHHH

MxeGyrA(Δ, T_3C)-His_6

```
CICGDALVAL PEGESVRIAD IVPGARPNSD NAIDLKVLDR HGNPVLADRL
FHSGEHPVYT VRTVEGLRVT GTANHPLLCL VDVAGVPTLL WKLIDEIKPG
DYAVIQGSGS GSGSFYYAKV ASVTDAGVQP VYSLRVDTAD HAFITNGFVS
HATGLTGLKL HHHHHH
```

His6-SUMO

```
MKSSHHHHHH GSSMASMSDS EVNQEAKPEV KPEVKPETHI NLKVSDGSSE
IFFKIKKTTP LRRLMEAFAK RQGKEMDSLR FLYDGIRIQA DQTPEDLDME
DNDIIEAHRE QIGG
```

His-Ulp1

```
MRGSHHHHHH GLVPRGSLVP ELNEKDDDQV QKALASRENT QLMNRDNIEI
TVRDFKTLAP RRWLNDTIIE FFMKYIEKST PNTVAFNSFF YTNLSERGYQ
GVRRWMKRKK TQIDKLDKIF TPINLNQSHW ALGIIDLKKK TIGYVDSLSN
GPNAMSFAIL TDLQKYVMEE SKHTIGEDFD LIHLDCPQQP NGYDCGIYVC
MNTLYGSADA PLDFDYKDAI RMRRFIAHLI LTDALK
```

Generation of *E. coli* Glycerol stocks for recombinant protein expression. Desired plasmid was transformed into BL21(DE3) *E.* coli cells by heat-shock and grown on an LB-Agar plate containing kanamycin (50μg/mL) for pFGET19 plasmid or ampicillin (100μg/mL) for all other plasmids. Plates were incubated at 37°C for at least 16 hours until colonies formed. An individual colony was picked and grown in 3 mL of sterile LB medium containing the appropriate antibiotic overnight at 37°C with 250 RPM shaking until saturation. 500 μL of this saturated culture were mixed with 500 μL 50% (v/v) sterile glycerol solution and frozen at -80°C.

General Procedure for recombinant expression and purification of proteins from Intein fusion proteins. The protocol is based on a previously published procedure[213]. For each liter of expression, four 5 mL primary cultures of sterile LB-media containing the ampicillin (100 μg/mL) were inoculated by scraping the surface of the frozen glycerol stock with a sterile pipette tip. Primary cultures were grown overnight at 37°C with 250 RPM shaking until saturation. Each liter of sterile LB-media containing ampicillin (100 μg/mL) was inoculated by addition of three primary cultures and grown at 37°C with 250

RPM shaking. Upon reaching an OD_{600} of 0.6-0.8, culture was switched to an incubator operating at 18°C with 250 RPM shaking. After allowing the media temperature to adjust for 15 minutes, IPTG was added to a final concentration of 1 mM to induce protein expression. Protein expression was continued at 18°C with 250 RPM shaking for 20 hours. Cells were harvested by centrifugation at 4,000 RPM in a GS3 rotor and Sorvall RC-5 centrifuge for 20 min at 4°C. The supernatant was discarded and the cell pellet was suspended in 30 mL lysis buffer (40 mM Tris, 1 mM PMSF, pH 8.3) containing a broad-spectrum protease inhibitor tablet. Resuspended cells were then lysed on ice by sonication (35 amps power, 1 second pulse, 2 second rest, 5 minutes total sonication time) and then pelleted at 14,000 RPM in an SS-34 rotor (Sorvall RC-5 centrifuge) for 30 min at 4°C. In the meantime, 4 mL (settled volume) of Nickel Agarose beads were washed with 20 mL equilibration buffer (50 mM HEPES, pH 7.5). The crude cell lysate was incubated with the washed nickel resin for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through. The resin was then washed with 20 mL of equilibration buffer and 25 mL of wash buffer (50 mM HEPES, 5 mM imidazole, pH 7.5). Intein fusion protein was then eluted from the resin in 10 mL of elution buffer (50 mM Tris, 300 mM imidazole, pH 8.5). Intein cleavage was induced by addition 758 µL of *β*-Mercaptoethanol (1 M final concentration). Cleavage was carried out at room temperature, 4°C, or 37°C for 60-75 hours before it was dialyzed into 20 mM Tris, pH 8.0 using Slide-A-Lyzer 2,000 MWCO cassettes. If protein has precipitated during cleavage, 8 M Urea was added immediately before dialysis until protein was dissolved again. Additionally, 20 µL of TCEP Bond-Breaker solution was added to cysteine containing proteins. After 2 rounds of dialysis at 4°C against at least 100 times the sample volume, sample was removed from dialysis cassette. 4 mL (settled volume) of Nickel Agarose beads were washed with 20 mL equilibration buffer (50 mM HEPES, pH 7.5) and dialyzed

sample was incubated with the washed nickel resin for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through and collected. The desired cleaved protein was contained in the flow-through fraction, while uncleaved fusion protein and cleaved intein product were bound to the resin. The crude product was dialyzed twice against 10 mM NH_4HCO_3 before quantification and lyophilization. Various protein expressions and/or purifications were carried out by or with help of Kristen E. Fiore.

2-Bromoethylamine labeling of Cysteine containing proteins. Proteins were dissolved in alkylation buffer (8 M Urea, 0.55 M Tris, pH 8.5). 5 equiv. of Tris(2-carboxyethyl)phosphine (TCEP) were added and solution was incubated at room temperature for 15 minutes. 20 equiv of 2-Bromoethylamine were dissolved in a minimal amount of 0.55 M Tris, pH 8.5 and added. Reactions were incubated in a waterbath at 37°C for 3-4 hours until reaction was deemed complete by MALDI-TOF MS. Crude reaction was diluted by 5 fold and dialyzed twice against 20 mM Tris, pH 8.0 at 4°C before FPLC purification.

Recombinant expression of GB1 (6-3). Protein was expressed as described above as GB1-NpuDnaE-His_6 fusion protein. Intein was cleaved at 37°C for 60 hours. After 2nd nickel column, protein was dialyzed against 20 mM Tris, pH 8.0 and purified by FPLC using gradient **6E**. Quantification post FPLC by UV-Vis (ε_{280} = 9,970 M^{-1} cm^{-1}) determined the yield as 14.5 mg/L expression. Expression and purification are shown in Figure 6-10.

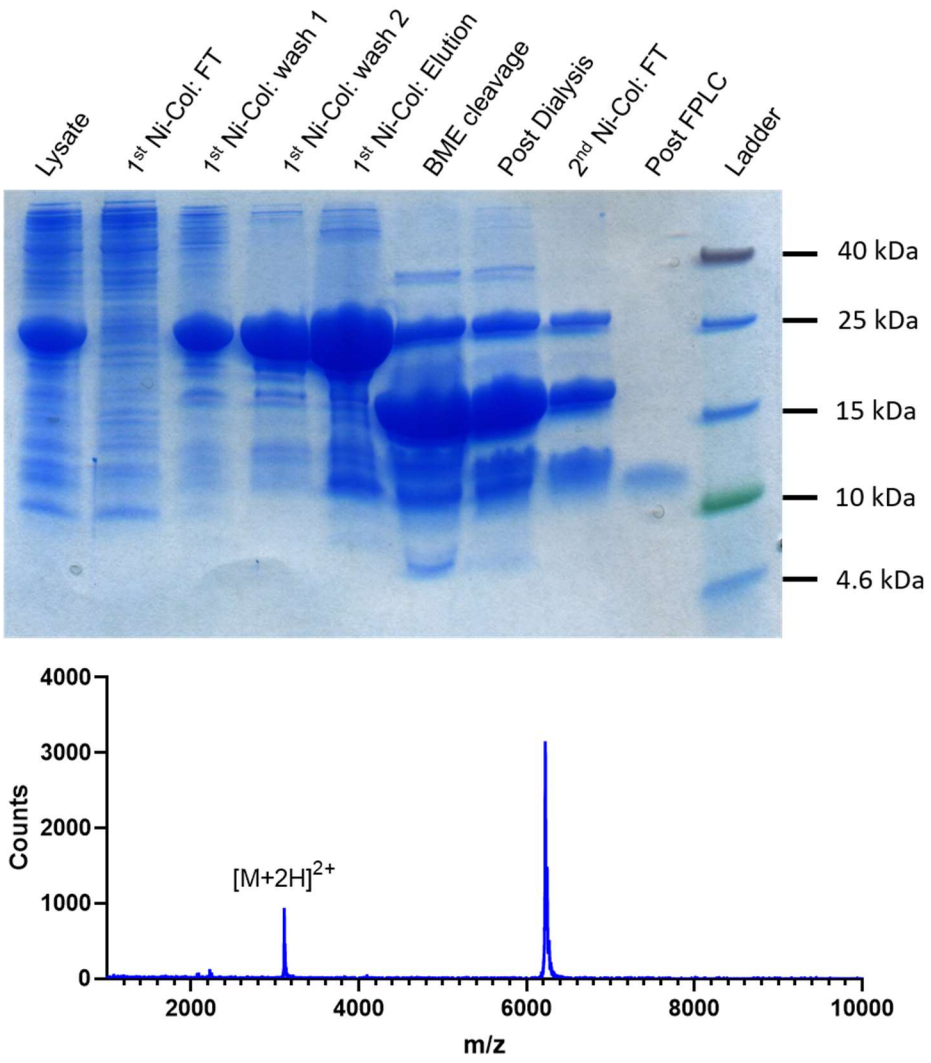


Figure 6-10: Characterization of GB1 expression. Top: SDS-PAGE (14% Tris-Tricine gel) of GB1 (WT) expression/purification. MW: GB1-NpuDnaE-His$_6$: 23.1 kDa; NpuDnaE-His$_6$: 19.6 kDa; GB1: 6.2 kDa. GB1 runs at a higher than expected molecular weight at about 11 kDa. Bottom: MALDI-TOF analysis of GB1. $[M+H]^+$: expected: 6,223.85 observed: 6,222.37.

Recombinant expression of GB1 $K_{10}C^K$ (6-5). Protein was expressed as described above as GB1 K_{10}C-NpuDnaE-His$_6$ fusion protein. Intein was cleaved at 37°C for 60 hours. After dialysis, protein **6-4** was quantified by UV-Vis (ε_{280} = 9,970 M^{-1} cm^{-1}; yield = 14.2 mg/L expression) before lyophilization. Lyophilized protein from 2 L

expression was alkylated as described above and purified on FPLC using gradient **6E**. Quantification post FPLC by UV-Vis (ε_{280} = 9,970 M^{-1} cm^{-1}) determined yield to be 6.7 mg / 48% (some protein was spilled post dialysis). Expression and purification are shown in Figure 6-11.

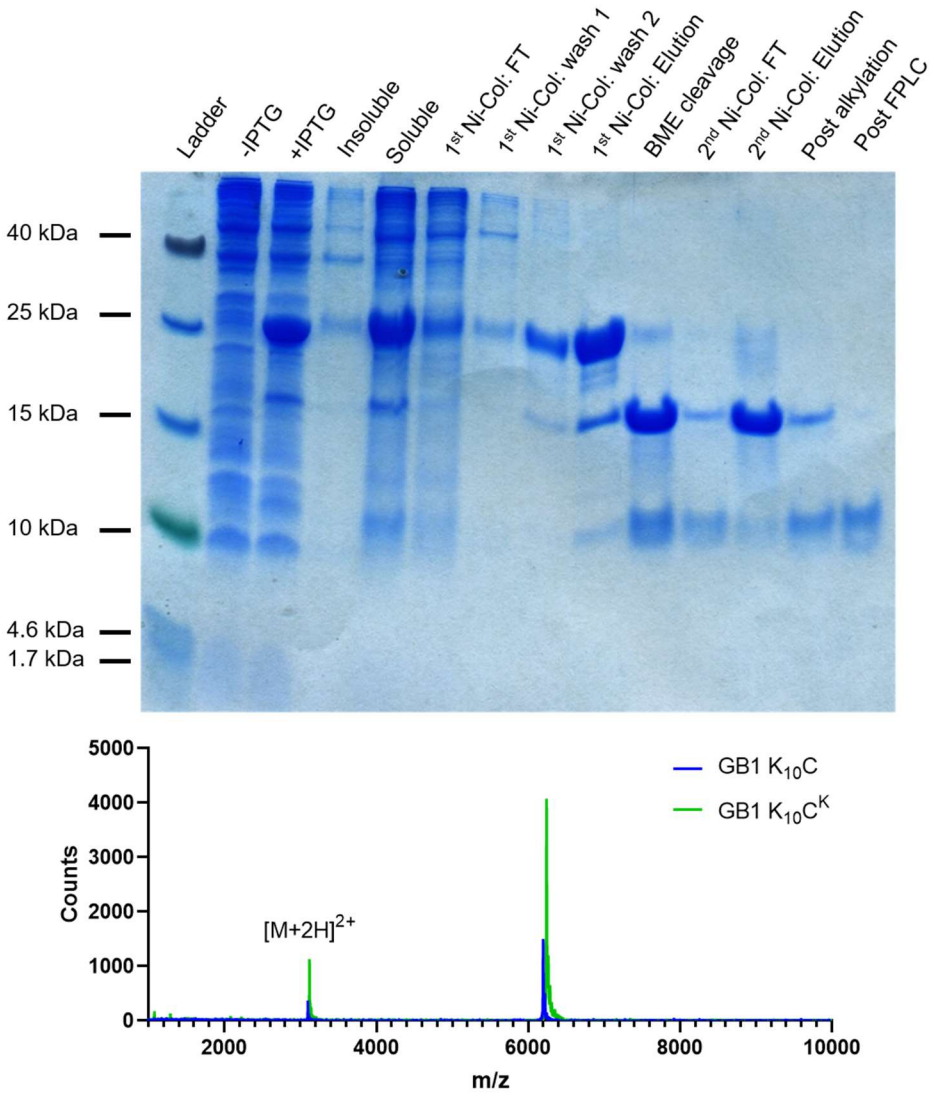


Figure 6-11: Characterization of GB1 $K_{10}C^K$ (6-5) expression. Top: SDS-PAGE (14% Tris-Tricine gel) of expression/purification. MW: GB1 $K_{10}C$-NpuDnaE-His_6: 23.1 kDa; NpuDnaE-His_6: 19.6 kDa; GB1: 6.2 kDa. GB1 $K_{10}C$ runs at a higher than expected molecular weight at about 11 kDa. Bottom: MALDI-TOF analysis of protein before (blue) and after (green) labelling. GB1 $K_{10}C$ (6-4): $[M+H]^+$: expected: 6,198.81 observed: 6,197.30. GB1 $K_{10}C^K$ (6-5): $[M+H]^+$: expected: 6,241.88 observed: 6,242.53.

Recombinant expression of $GB1_{10\text{-}56}$ $K_{10}C$ (6-6). Protein was expressed as described above as $GB1_{10\text{-}56}$ $K_{10}C$-MxeGyrA-His_6 fusion protein. Intein was cleaved at 4°C for 72 hours. After dialysis, protein was quantified by UV-Vis (ε_{280} = 9,970 M^{-1} cm^{-1}; yield = 20.9 mg/L expression) before lyophilization. N-terminal cysteine has been observed to form +70 mass adducts from pyruvate addition. This was also the main observed species in MALDI-TOF MS. Reversal of this thiazolidine (Thz) modification is achieved by treatment thiazolidine deprotection buffer (6 M Gdn*HCl, 0.2 M sodium phosphate, 0.2 M Methoxyamine, pH 4.0) overnight[215]. After Thz deprotection, protein was dialyzed twice against 10 mM NH_4HCO_3 and lyophilized. Expression and purification are shown in Figure 6-12

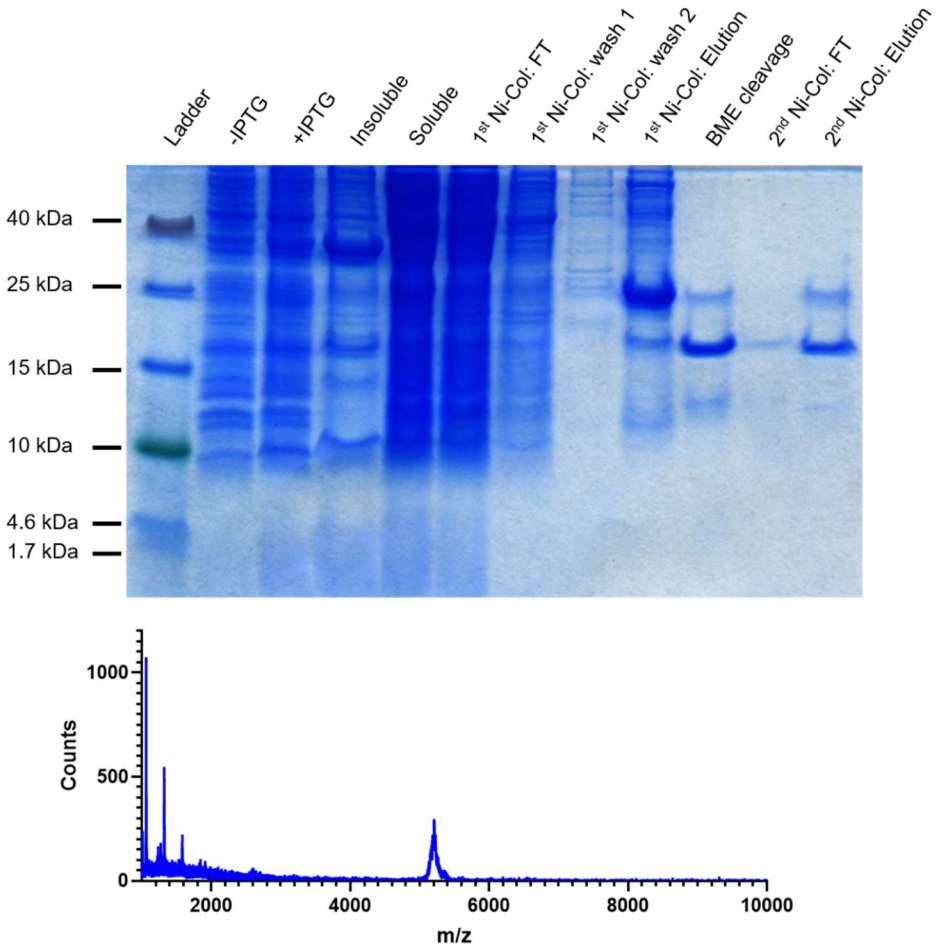


Figure 6-12: Characterization of $GB1_{10\text{-}56}$ $K_{10}C$ (6-6) expression. Top: SDS-PAGE (14% Tris-Tricine gel) of expression/purification. MW: $GB1_{10\text{-}56}$ $K_{10}C$-MxeGyrA(Δ)-His_6: 22.8 kDa; MxeGyrA(Δ)-His_6: 17.7 kDa; $GB1_{10\text{-}56}$ $K_{10}C$: 5.1 kDa. $GB1_{10\text{-}56}$ $K_{10}C$ runs at a higher than expected molecular weight at about 11 kDa and stains poorly. Bottom: MALDI-TOF analysis. $GB1_{10\text{-}56}$ $K_{10}C$ (Pyruvate adduct): $[M+H]^+$: expected: 5,207.50 observed: 5,208.79

Recombinant expression of GB1 $K_{13}C$ (6-8). Protein was expressed as described above as GB1 $K_{13}C$-NpuDnaE-His_6 fusion protein. Intein was cleaved at room temperature for 74 hours. After 2nd nickel column, protein was dialyzed against 20 mM Tris, pH 8.0 and purified by FPLC using gradient **6F**. Quantification post FPLC by UV-Vis (ε_{280} = 9,970 M^{-1} cm^{-1}) determined the yield as 12.7 mg/L expression. Expression and purification are shown in Figure 6-13.

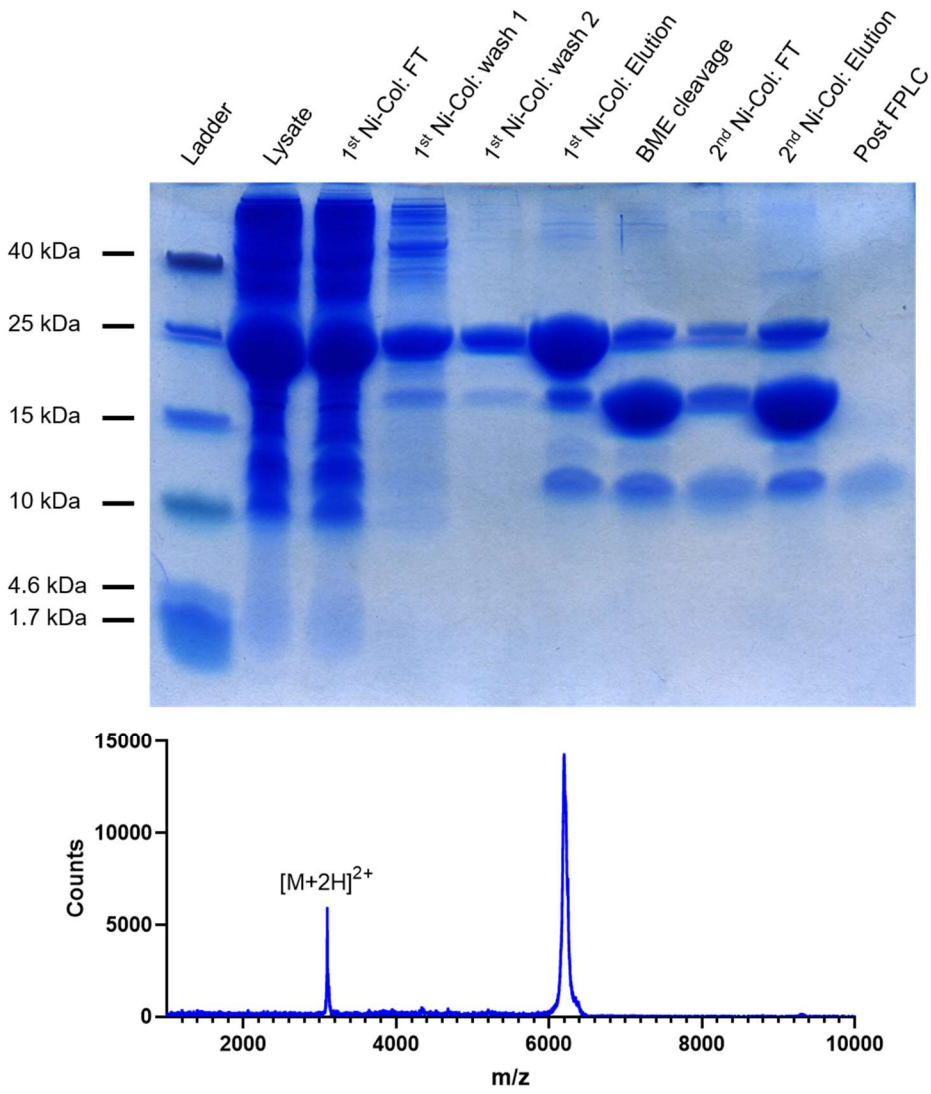


Figure 6-13: Characterization of GB1 $K_{13}C$ (6-8) expression. Top: SDS-PAGE (14% Tris-Tricine gel) of expression/purification. MW: GB1 $K_{13}C$-NpuDnaE-His_6: 23.1 kDa; NpuDnaE-His_6: 19.6 kDa; GB1: 6.2 kDa. GB1 $K_{13}C$ runs at a higher than expected molecular weight at about 11 kDa. Bottom: MALDI-TOF analysis. GB1 $K_{13}C$ $[M+H]^+$: expected: 6,198.81 observed: 6,198.26.

Recombinant expression of Ulp-1. The protocol is based on a previously published procedure[222]. For each liter of expression, four 5 mL primary cultures of sterile LB-media containing the kanamycin (50 µg/mL) were inoculated by scraping the surface of the frozen glycerol stock with a sterile pipette tip. Primary cultures were grown overnight

at 37°C with 250 RPM shaking until saturation. Each liter of sterile LB-media containing ampicillin (50 µg/mL) was inoculated by addition of three primary cultures and grown at 37°C with 250 RPM shaking. Upon reaching an OD_{600} of 0.6-0.8, IPTG was added to a final concentration of 1 mM to induce protein expression. Protein expression was continued at 37°C with 250 RPM shaking for 5 hours. Cells were harvested by centrifugation at 4,000 RPM in a GS3 rotor and Sorvall RC-5 centrifuge for 20 min at 4°C. The supernatant was discarded and the cell pellet was suspended in 40 mL lysis buffer (40 mM Tris, 500 mM NaCl, 1 mM PMSF, pH 8.3) containing a broad-spectrum protease inhibitor tablet. Resuspended cells were then lysed on ice by sonication (35 amps power, 1 second pulse, 2 second rest, 5 minute total sonication time) and then pelleted at 14,000 RPM in an SS-34 rotor (Sorvall RC-5 centrifuge) for 30 min at 4°C. In the meantime, 5 mL (settled volume) of Nickel Agarose beads were washed with 20 mL equilibration buffer (50 mM HEPES, 20 mM imidazole, pH 7.5). The crude cell lysate was incubated with the washed nickel resin for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through. The resin was then washed with 30 mL of equilibration buffer and was then eluted from the resin in 12 mL of elution buffer (50 mM HEPES, 300 mM imidazole, pH 7.5). The crude protein was FPLC purified by size-exclusion chromatography using gradient **6G** on a HiLoad Superdex 75pg 16/600 column. Fractions were analyzed by SDS-PAGE and pure fractions from peak were combined and spin concentrated using an Amicon Ultra-4 10k MWCO Spin Filter until the volume was about 1 mL. 1 mL of protein solution was mixed with 20 uL of 1 M Dithiothreitol (DTT; 10 mM final concentration) and 980 µL 50% (v/v) glycerol, aliquoted and frozen at -80°C until further use.

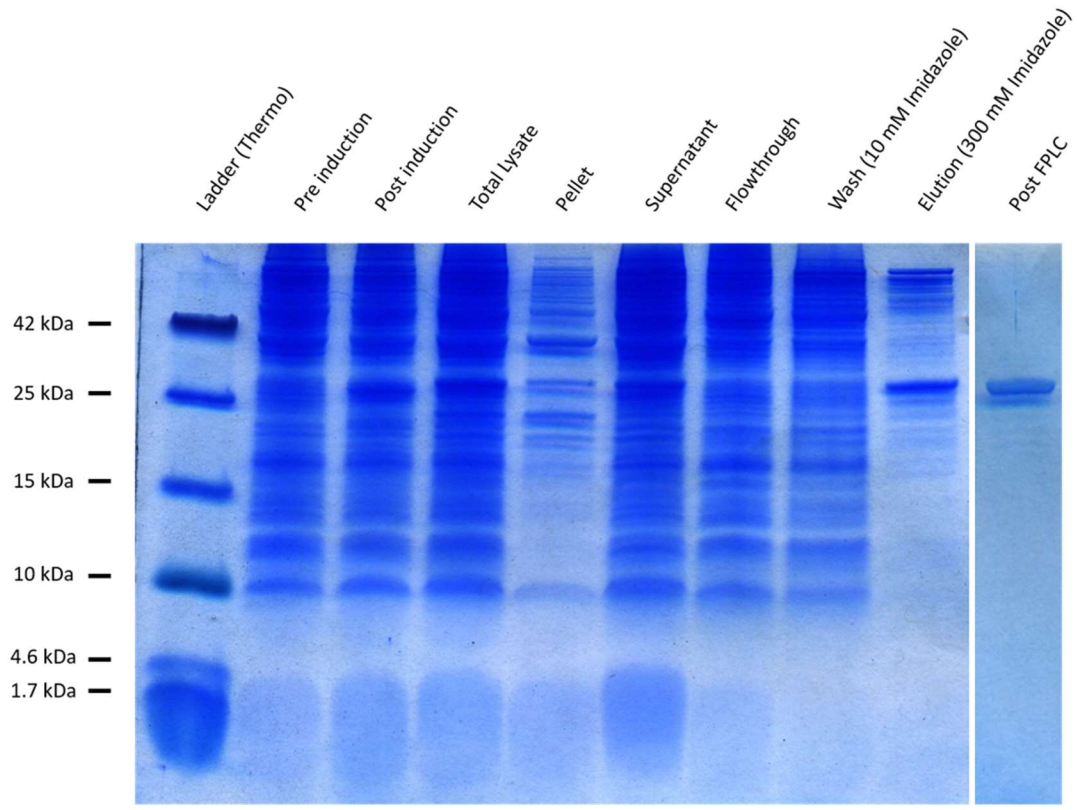


Figure 6-14: SDS-PAGE (14% Tris-Tricine gel) following expression and purification of Ulp-1 (6-9). The post FPLC band was run on a separate gel with identical composition and run conditions. The molecular weight of Ulp-1 is 27.4 kDa.

Recombinant expression and purification of $GB1_{13\text{-}56}$ $K_{13}C$ (6-10) from SUMO fusion protein. The protocol is based on the expression and purification protocol of a previously published procedure[222]. For each liter of expression, four 5 mL primary cultures of sterile LB-media containing the ampicillin (100 µg/mL) were inoculated by scraping the surface of the frozen glycerol stock with a sterile pipette tip. Primary cultures were grown overnight at 37°C with 250 RPM shaking until saturation. Each liter of sterile LB-media containing ampicillin (100 µg/mL) was inoculated by addition of three primary cultures and grown at 37°C with 250 RPM shaking. Upon reaching an OD_{600} of 0.6-0.8, culture was switched to an incubator operating at 18°C with 250 RPM shaking. After allowing the

media temperature to adjust for 15 minutes, IPTG was added to a final concentration of 1 mM to induce protein expression. Protein expression was continued 18°C with 250 RPM shaking for 20 hours. Cells were harvested by centrifugation at 4,000 RPM in a GS3 rotor and Sorvall RC-5 centrifuge for 20 min at 4°C. The supernatant was discarded and the cell pellet was suspended in 30 mL lysis buffer (40 mM Tris, 1 mM PMSF, pH 8.3) containing a broad-spectrum protease inhibitor tablet. Resuspended cells were then lysed on ice by sonication (35 amps power, 1 second pulse, 2 second rest, 5 minute total sonication time) and then pelleted at 14,000 RPM in an SS-34 rotor (Sorvall RC-5 centrifuge) for 30 min at 4°C. In the meantime, 4 mL (settled volume) of Nickel Agarose beads were washed with 20 mL equilibration buffer (50 mM HEPES, pH 7.5). The crude cell lysate was incubated with the washed nickel resin for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through. The resin was then washed with 20 mL of equilibration buffer and 25 mL of wash buffer (50 mM HEPES, 5 mM imidazole, pH 7.5). Intein fusion protein was then eluted from the resin in 10 mL of elution buffer (50 mM HEPES, 300 mM imidazole, pH 7.5). 50 µL of thawed Ulp-1 and DTT (1 mM final concentration) were added and incubated on a rotisserie at room temperature overnight. The next day, protein was dialyzed twice against 10 mM NH_4HCO_3 using a Slide-A-Lyzer 2,000 MWCO cassette. 4 mL (settled volume) of Nickel Agarose beads were washed with 20 mL equilibration buffer (50 mM HEPES, pH 7.5) and dialyzed sample was incubated with the washed nickel resin for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through and collected. The desired cleaved protein was contained in the flow-through fraction, while uncleaved fusion protein and cleaved SUMO product were bound to the resin. The resulting protein was >95% pure based on SDS-PAGE and MALDI and therefore lyophilized post quantification via UV-Vis (ε_{280} = 8,480 M^{-1} cm^{-1}). The isolated yield was

determined to be 8.5 mg/L expression. Figure 6-15 shows the cleavage of the SUMO fusion protein by Ulp-1. $GB1_{13\text{-}56}$ $K_{13}C$ runs at a higher molecular weight than expected and stains poorly with Coomassie stain. It is visible as faint band at around 11 kDa.

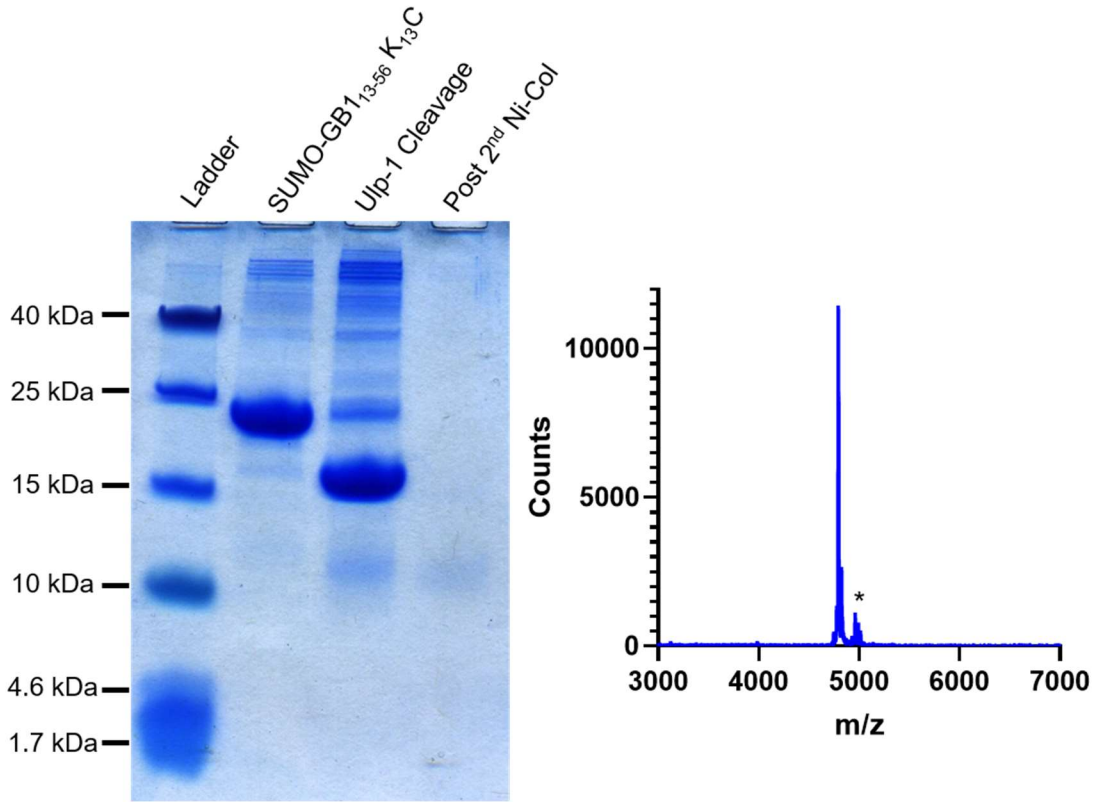


Figure 6-15: Characterization of $GB1_{13\text{-}56}$ $K_{13}C$ (6-10) expression. Left: SDS-PAGE (14% Tris-Tricine gel) following purification of 6-10 from a SUMO fusion protein. MW: His_6-SUMO-$GB1_{13\text{-}56}$ $K_{13}C$: 17.8 kDa; His_6-SUMO: 13.0 kDa; $GB1_{13\text{-}56}$ $K_{13}C$: 4.8 kDa. $GB1_{13\text{-}56}$ $K_{13}C$ runs at a higher than expected molecular weight at about 11 kDa. Right: MALDI-TOF analysis. $GB1_{13\text{-}56}$ $K_{13}C$ $[M+H]^+$: expected: 4,795.07 observed: 4,793.57. * denotes MPAA adduct.

Native Chemical Ligation of $GB1_{10\text{-}56}$ $K_{10}C$ with $GB1_{1\text{-}10}$-N_2H_3. C-terminal fragment (**6-6**, 27.5 mg, 5.36 mmoles) was dissolved in 1 mL ligation buffer (200 mM Na_2HPO_4, 6 M Gdn•HCl, pH 7.5) and 2.5 equiv TCEP bond-breaker solution was added. N-terminal fragment (**6-1**, 1.94 mg, 1.75 umol) was dissolved in 500 µL activation buffer

(200 mM Na_2HPO_4, 6 M Gdn•HCl, pH 3.0) and stirred at -15 °C for 10 minutes. Next, 80.4 µL of a 1 M solution of $NaNO_2$ in water (15 equiv.) was added and the reaction was stirred for 15 minutes at -15 °C. After 15 minutes, the reaction was allowed to warm to room temperature and 150 equiv. MPAA in 1.2 mL ligation buffer were added. The N-terminal fragment was stirred for 1 hour to allow formation of the thioester before the C-terminal fragment was added to the N-terminal fragment. The pH was adjusted to 7.5, and the solution stirred at room temperature for 19 hours.

Native Chemical Ligation of $GB1_{13\text{-}56}$ $K_{13}C$ with $GB1_{1\text{-}12}$-N_2H_3. C-terminal fragment (**6-10**, 8.5 mg, 1.77 mmoles) was dissolved in 2 mL ligation buffer (200 mM Na_2HPO_4, 0.1 M MPAA, 6 M Gdn•HCl, pH 7.0). N-terminal fragment (**6-2**, 2.1 mg, 1.44 umol) was dissolved in 480 µL activation buffer (200 mM Na_2HPO_4, 6 M Gdn•HCl, pH 3.0) and stirred at -15 °C for 10 minutes. Next, 20 µL of a 1 M solution of $NaNO_2$ in water was added and the reaction was stirred for 15 minutes at -15 °C. After 15 minutes, the C-terminal fragment was added to the N-terminal fragment, the pH was adjusted to 7.0, and the solution stirred at room temperature for 72 hours. Occasionally, the reaction progress was checked by MALDI-TOF MS and the pH was adjusted to pH 7.0. Upon completion, the reaction mixture was diluted to a volume of 12 mL with H_2O and dialyzed twice against 20 mM Tris, pH 8.0. The product was then attempted to be purified by FPLC on a HiTrap Q HP column using gradient **6F**. However, product **6-11** and unreacted C-terminal fragment **6-10** coeluted. All fractions containing product were combined and lyophilized. Crude product was dissolved in 0.1% TFA in water and purified on HPLC using gradient **6H**. We initially used a Luna Omega C18 semiprep column at a flowrate of 4 mL/min. After observing good separation, we switched to a Luna Omega C18 prep column operating at

a flowrate of 25 mL/min, resulting in a slight shift of retention times as indicated in Table 6.1.

Table 6.1: Peptide/Protein Purification Methods and Retention Times.

Peptide/Compound	Gradient	Retention Time/ Elution Volume	Column
$GB1_{1\text{-}9}$ L^{S}_{7}-N_2H_3 (**6-1**)	**6A**	11.5 min	SNAP Ultra C18 30g
	6B	21.1 min	Luna Omega Prep
$GB1_{1\text{-}12}$ L^{S}_{7}-N_2H_3 (**6-2**)	**6C**	19.0 min	SNAP Ultra C18 30g
	6D	17.6 min	Luna Omega Prep
GB1 (WT) (**6-3**)	**6E**	50 mL	HiTrap Q HP 5 mL
GB1 $K_{10}C$ (**6-4**)	**6E**	71 mL	HiTrap Q HP 5 mL
GB1 $K_{10}C^{K}$ (**6-5**)	**6E**	53 mL	HiTrap Q HP 5 mL
GB1 $K_{13}C$ (**6-8**)	**6F**	56 mL	HiTrap Q HP 5 mL
Ulp-1 **(6-9)**	**6G**	70 mL	HiLoad Superdex 75 pg 16/600
$GB1_{13\text{-}56}$ $K_{13}C$ **(6-10)**	**6H**	38.4 min	Luna Omega Semiprep
		37.5 min	Luna Omega Prep
GB1 L^{S}_{7} $K_{13}C$ (**6-11**)	**6H**	41.9 min	Luna Omega Semiprep
		41.5 min	Luna Omega Prep

Table 6.2: FPLC and HPLC gradients used for Peptide/Protein Purification and Characterization.

No.	Time (min) or CV	%B	No.	Time (min) or CV	%B	No.	Time (min) or CV	%B
6A[^]	0 CV	15	**6B**[^]	0:00	5	**6C**[^]	0 CV	15
	2 CV	15		2:00	5		3 CV	15
	12 CV	35		4:00	15		23 CV	45
	12 CV	100		34:00	30		23 CV	100
	14 CV	100		36:00	30		26 CV	100
	14 CV	50		37:00	100		26 CV	50
	17 CV	50		40:00	100		28 CV	50
6D[^]	0:00	5	**6E**[#]	0.0 CV	0	**6F**[#]	0.0 CV	0
	5:00	5		2.0 CV	0		2.0 CV	0
	8:00	20		14.0 CV	20		22.0 CV	30
	18:00	25		15.6 CV	20		23.6 CV	30
	20:00	25		25.6 CV	100		23.6 CV	100
	21:00	100		20.6 CV	100		28.6 CV	100
	26:00	100		20.6 CV	0		28.6 CV	0
6G[#]	0.0 CV	15	**6H**[^]	0:00	2			
	1.2 CV	15		10:00	2			
				15:00	20			
				45:00	35			
				47:00	35			
				48:00	100			
				53:00	100			
				55:00	2			

[^]HPLC: Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

[#]FPLC: Solvent A: 20 mM Tris, pH 8.0; Solvent B: 20 mM Tris, 1 M NaCl, pH 8.0

Table 6.3: MALDI-TOF MS Characterization of Peptides and Proteins.

Peptide	$[M+H]^+$	
	Calculated	**Found**
$GB1_{1\text{-}9}\ L^S{}_7\text{-}N_2H_3$ (**6-1**)	1,131.58 ($[M+Na]^+$)	1,131.86 ($[M+Na]^+$)
$GB1_{1\text{-}12}\ L^S{}_7\text{-}N_2H_3$ (**6-2**)	1,451.82	1,452.21
GB1 (WT) (**6-3**)	6,223.85	6,222.37
GB1 $K_{10}C$ (**6-4**)	6,198.81	6,197.30
GB1 $K_{10}C^K$ (**6-5**)	6,241.88	6,242.53
$GB1_{10\text{-}56}\ K_{10}C$ (**6-6**)	5,207.50 (pyruvate adduct)	5,208.79 (pyruvate adduct)
$(GB1_{1\text{-}9}\ L^S{}_7)_2$-OH (**6-7**)	2,712,12	2,172.30
GB1 $K_{13}C$ (**6-8**)	6,198.81	6,198.26
$GB1_{13\text{-}56}\ K_{13}C$ (**6-10**)	4,795.07	4,793.57
GB1 $L^S{}_7\ K_{13}C$ (**6-11**)	6,214.81	6,214.23

Chapter 7 : Synthesis, Characterization and Attempted Genetic Incorporation of N^{ε}-alkyl lysine and N^{ε}-Thioaalkyl lysine derivatives

7.1 Introduction

While our lab has built significant expertise in incorporating thioamides into full-length proteins using semi-synthetic methods, it is still cumbersome and not high-yielding. Genetic incorporation of thioamides would be the 'holy grail', as it would increase yields, facilitate purification and enable less specialized labs to express thioamide labelled proteins. Unfortunately, it is still a long way to go to reach this goal. Thus far, there are only two examples of ribosomal incorporation of thioamides into full-length proteins, both utilizing *in vitro* protein translation with *in vitro* aminoacylated tRNAs. The first example is from the Hecht lab[138]. In this study, they introduced random mutations in two different regions of the 23S rRNA to generate modified ribosomes that would allow incorporation of various dipeptides and dipeptide analogues, including Phe^{S}-Gly dipeptide. After aminoacylating tRNA *in vitro* with the Phe^{S}-Gly, they used this tRNA for *in vitro* protein expression with a cell extract from *E. coli* containing the modified ribosomes. While they were able to provide some evidence for successful thioamide incorporation, based on fluorescence quenching, it is not known if the desired dipeptide was incorporated quantitatively. Furthermore, the isolated yield was less than 100 ng, making it difficult to obtain protein quantities needed for biochemical assays.

The other example of ribosomal thioamide incorporation is similar in nature and comes from the Söll lab[225]. Similar to the study by the Hecht lab, this study uses *in vitro*

aminoacylated tRNA in cell-free *in vitro* translation experiments. Unlike Hecht's aminoacylation approach, which relied on synthesis of the aminoacylated dinucleotide and ligating it to truncated tRNA with T4 RNA Ligase, the Söll lab utilizes Flexizyme. Flexizyme is a catalytic RNA developed by the Suga lab that is capable of aminoacylating tRNA[226]. It recognizes an activated leaving group, in this case 3,5-dinitrobenzyl esters, and aminoacylates tRNA, virtually independent from the rest of the dinitrobenzyl ester (DBE) molecule. The Söll lab synthesized 3,5-dinitrobenzyl esters of N^{ε}-acetyl-lysine (Lys(Ac)-DBE), N^{ε}-thioacetyl-lysine (Lys(Ac^{S})-DBE), and N^{ε}-selenoacetyl-lysine (Lys(Ac^{Se})-DBE) and were able to aminoacylate tRNA in a flexizyme reaction using their substrates. While they were unable to incorporate Lys(Ac^{Se}) into any of their tested constructs, they successfully incorporated up to two unnatural amino acids into full-length Histone 3, as supported by MS/MS characterization of tryptic digests. However, similar to the other study their yield was very low only yielding low µg amounts of expressed protein. Together these two studies show that once thioamides are attached to tRNA, they can be successfully incorporated into proteins and don't seem to interfere with any of the associated processes. However, while being very valuable proof-of-concept studies, none of these approaches yield protein amounts that would allow further biochemical assays. The low yields are inherent to techniques that rely on *in vitro* aminoacylated tRNA, as these tRNAs have only a single turnover while also slowly undergoing hydrolysis of the aminoacylated tRNA under protein expression conditions.

We were interested in testing a different approach to unnatural amino acid (UAA) incorporation that relies on an tRNA synthetase (RS). This method, pioneered by the Schultz Lab, requires the synthetase/tRNA pair to be orthogonal in the expression system used, meaning that the orthogonal synthetase doesn't aminoacylate any cognate tRNA

and that the orthogonal tRNA doesn't get aminoacylated by any of the cognate synthetases[227,228]. Once this orthogonality is ensured, the synthetase can be mutated to accommodate unnatural amino acids, effectively expanding the genetic code by adding an additional amino acid/tRNA pair (see Figure 7-1). Hundreds of unnatural amino acids have been incorporated into various proteins using this approach[229]. Usually the amber stop codon UAG is used as recognition site for the tRNA, since the amber codon is the rarest of the three stop codons. This implies a competition between release factor, specifically release factor I for amber codons, and the orthogonal, aminoacylated tRNA. The two main products are usually the truncated protein and full-length desired protein containing the unnatural amino acid. Our lab has utilized this approach previously to incorporate various unnatural amino acids[224,230–232].

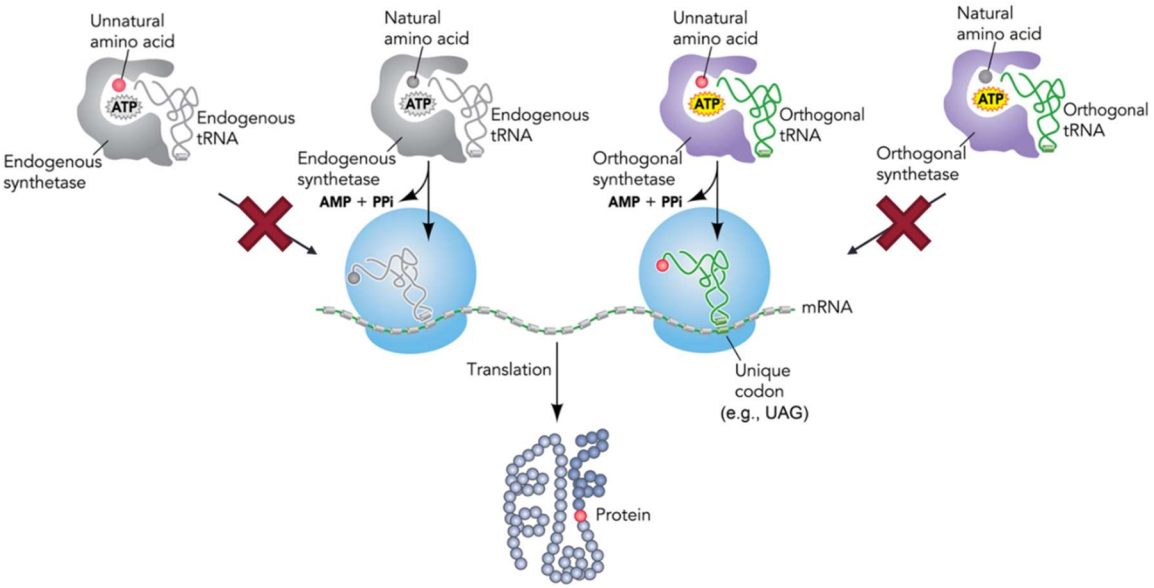


Figure 7-1: Codon reassignment strategy for the site-specific incorporation of unnatural amino acids: An orthogonal synthetase/tRNA pair are responsible for site-specific of the unnatural amino acid. Figure adapted from Lei Wang[233].

We were interested in using this approach to incorporate N^{ε}-thioacetyl-lysine (Lys(Ac^{S})) and N^{ε}-thioheptyl-lysine (Lys($Hept^{S}$)). Lys(Ac^{S}) was of interest since

various versions of synthetases able to incorporate Lys(Ac) have been reported. Lys(Hept^{S}) incorporation was suggested by our collaborator, Prof. Wenshe Liu, as they have evolved a synthetase that aminoacylates tRNA with N$^{\varepsilon}$-octyl-lysine (Lys(Oct)). However, they noted that there is a variety of N$^{\varepsilon}$-alkyl-lysine derivatives it can utilize and as part of their substrate scope, they were curious if their synthetase would be able to incorporate Lys(Hept^{S}). The synthetases for both are based on pyrrollysine synthetase (PylRS) from *Methanosarcina barkeri* or *Methanosarcina mazei*. The PylRS/tRNA^{Pyl} pair is often used since it is orthogonal in *E. coli* as well as in mammalian cell lines[234]. The synthetase is unique in that recognition of its cognate tRNA does not occur through recognition of the anticodon[234,235]. Therefore the Pyl system can be modified to recognize various codons, although the UAG amber stop codon is most commonly used since it is its cognate recognition codon[235].

7.2 Results and Discussion

We chose super-folder green fluorescent protein (sfGFP) as our test system. It is a well-established test system for unnatural amino acid incorporation. The advantage of sfGFP as test system that the protein is expressed at high levels, very stable, and yield can be determined based on absorption or fluorescence. In the sfGFP system the amber stop codon, here denoted as Z in 1-letter code, is usually placed in one of the sheets of the *β*-barrel structure. Truncation at the amber codon site results in a shortened sfGFP fragment that has the amino acids needed for chromophore maturation, but is unable to fold properly and is therefore non-fluorescent.

Initial attempts were carried out at the Unnatural Amino Acid workshop at the Oregon State University in Corvallis, OR. We attempted to incorporate Lys(Ac) and Lys(Ac^{S}) into GFP using the AcK3 synthetase developed by Chin and coworkers[236]. While

Lys(Ac) is commercially available, we had to synthesize Lys(Ac^S) ourselves. We based our synthesis on a published procedure for a protected Lys(Ac^S) derivative and were able to obtain our desired compound over 2 steps in high purity and high yield[132]. We expressed in our protein in autoinducing media. WT full-length sfGFP was used as positive control and an expression not containing any unnatural amino acid was used as negative control. As evident from Figure 7-2, the synthetase is not very efficient at incorporating Lys(Ac). Unfortunately the tiny amount of full-length protein with Lys(Ac^S) is comparable to the amount of full-length protein in the absence of UAA, indicating that it might just be non-specific incorporation. However, the amount of protein obtained from the oxo-control Lys(Ac) were discouraging.

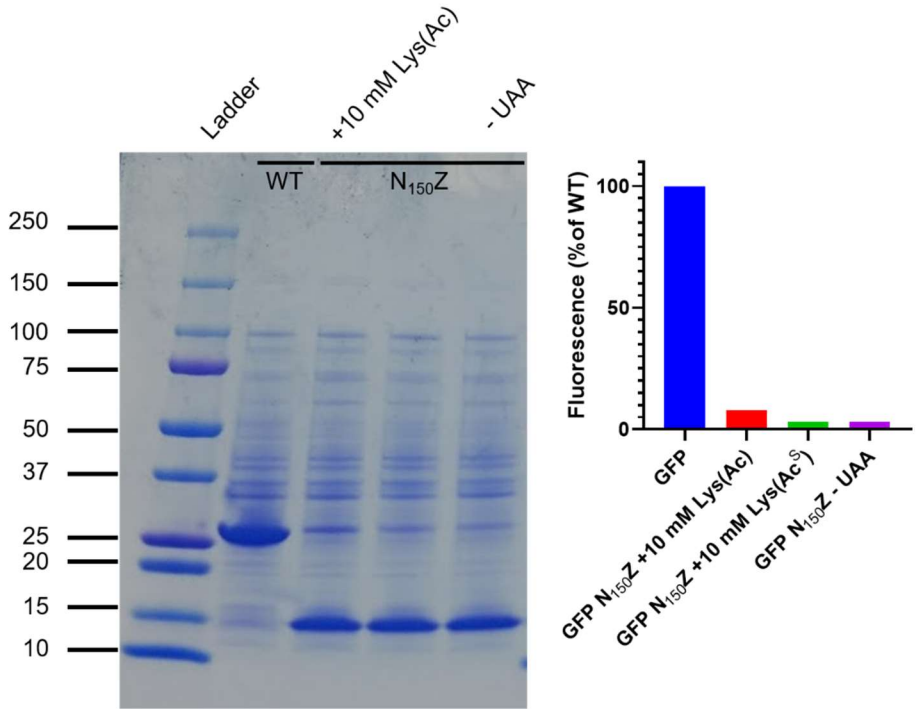


Figure 7-2: Attempted incorporation of Lys(Ac) and Lys(Ac^S) into sfGFP. Left: SDS-PAGE analysis of cells shows the main product to be truncated protein. Right: Fluorescence measurement supports that only a small fraction of Lys(Ac) got incorporated to generate fluorescent full-length GFP.

Recently we started a collaboration with the lab of Prof. Wenshe Liu (Texas A&M, College Station, TX, USA). His lab has evolved a new synthetase that is capable of incorporating a variety of N^{ε}-alkyl lysine derivatives. Upon receiving their plasmid, we decided to utilize the same plasmid backbone to test a newly described N^{ε}-acetyl lysine synthetase[237] and introduced the new synthetase into the plasmid using Gibson Assembly. The new synthetase was evolved using the phage-assisted continuous evolution (PACE) system for rapid directed evolution developed by Liu and coworkers[238]. With their optimized synthetase they were able to obtain 20-25% of WT GFP expression (based on fluorescence), compared to about 2% with the previously used synthetase.

To generate a GFP test system, we used plasmids with sfGFP and sfGFP($Y_{151}Z$) from our collaborator Prof. Abhishek Chatterjee (Boston College, Chestnut Hill, MA, USA). However, these plasmids contained a T5 promoter site that is recognized by endogenous *E. coli* RNA polymerases. We were interested in a T7 system, where our gene of interest would only be recognized by T7 RNA polymerase present in *E. coli BL21* (DE3). We used Gibson assembly to create a pET based T7-compatible vector where sfGFP is under control of the lac operon.

Since none of the amino acids we planned to incorporate are commercially available, we synthesized Lys($Hept^{S}$)-OH (**7-6**) and its oxo analogue Lys(Hept)-OH (**7-4**) ourselves. While the synthesis was rather low yielding, we obtained sufficient material for initial test expressions and did not bother optimizing synthetic conditions.

We set up 25 mL expressions to test incorporation of unnatural amino acids into sfGFP. Proteins were expressed for 24 hours in 25 mL autoinducing medium containing 20 mM nicotinamide to suppress, endogenously expressed CobB, a NAD-dependent lysine deacylase[236]. After 24 hours, OD_{600} and fluorescence of the crude suspension was

measured before harvesting and purifying cells. As shown in Figure 7-3, cells produced significantly lower amount of protein (estimated by fluorescence), especially if one were to take the number of cells (estimated by optical density) into account: Only 9.0 % of WT fluorescence were observed with Lys(Ac) incorporation and 18.6% with Lys(Hept). Measured fluorescence for thio-amino acid incorporation is roughly 1/3 of the corresponding oxo version (4.0% and 6.2%) and only slightly above the negative control without UAA (3.6% and 4.0%), which accounts for non-specific readthrough. We were wondering if the decrease in observed fluorescence is due to lower amount of full-length protein or if incorporated thioamides could potentially partially quench fluorescence.

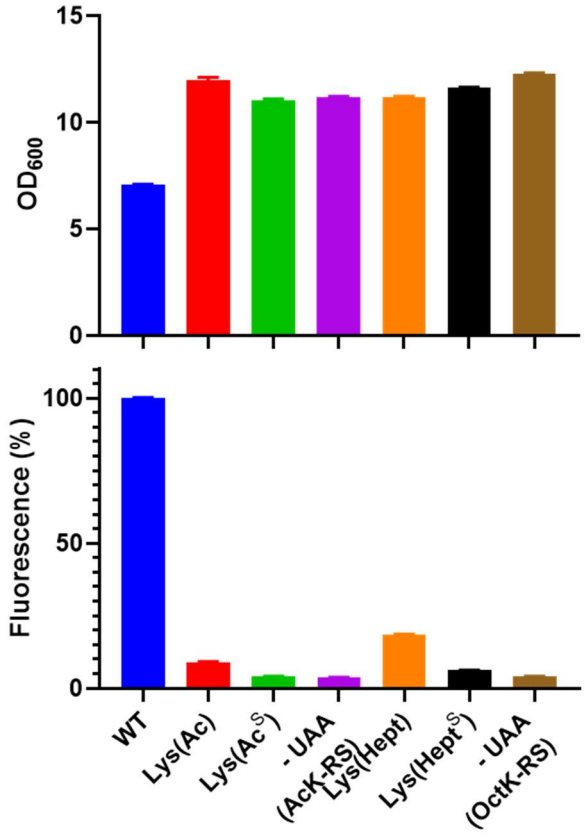


Figure 7-3: OD_{600} (top) and fluorescence (bottom) measurements (in triplicate) of cells grown for 24 hours in autoinducing medium. Non-WT expressions contained 10 mM Lys(Ac)/Lys(Ac^S) or 1 mM Lys(Hept)/Lys($Hept^S$).

To further analyze the amount of amino acid incorporated into proteins, we harvested the cells and isolated the protein using nickel affinity columns. Since the His tag is on the C-terminal end of the protein, only full-length protein will be purified, while truncated protein will be washed away. Purified proteins were desalted and further analyzed by Trypsin digests. SDS-PAGE and results of tryptic digest results are shown in Figure 7-4 for incorporation experiments using acetyllysine synthetase mmchAcK3RS, and Figure 7-5 for incorporation experiments using octyllysine synthetase mmOctKRS. In neither case, we did not observe incorporation of the desired thio-amino acid; in all cases the purified protein seems to solely contain the oxo-analogue. The mechanism of this is not clear. A likely alternative, is that some of the thio-amino acid undergoes S-to-O exchange in the growth medium or in the cell and that the synthetase is only able to incorporate oxo-amino acid. This hypothesis is supported by the observed reduced protein yield, as the concentration of oxo-amino acid in this case is significantly lower. Furthermore, Söll *et al.*[225] have shown that when tRNA is aminoacylated with Lys(Ac^S), the amino acid will get incorporated into the protein and it is possible to isolate the thioamide containing protein. This, in our opinion, excludes the possibility of enzymatic processing after thioamide incorporation. However, further systematic investigation would be necessary to elucidate the mechanism.

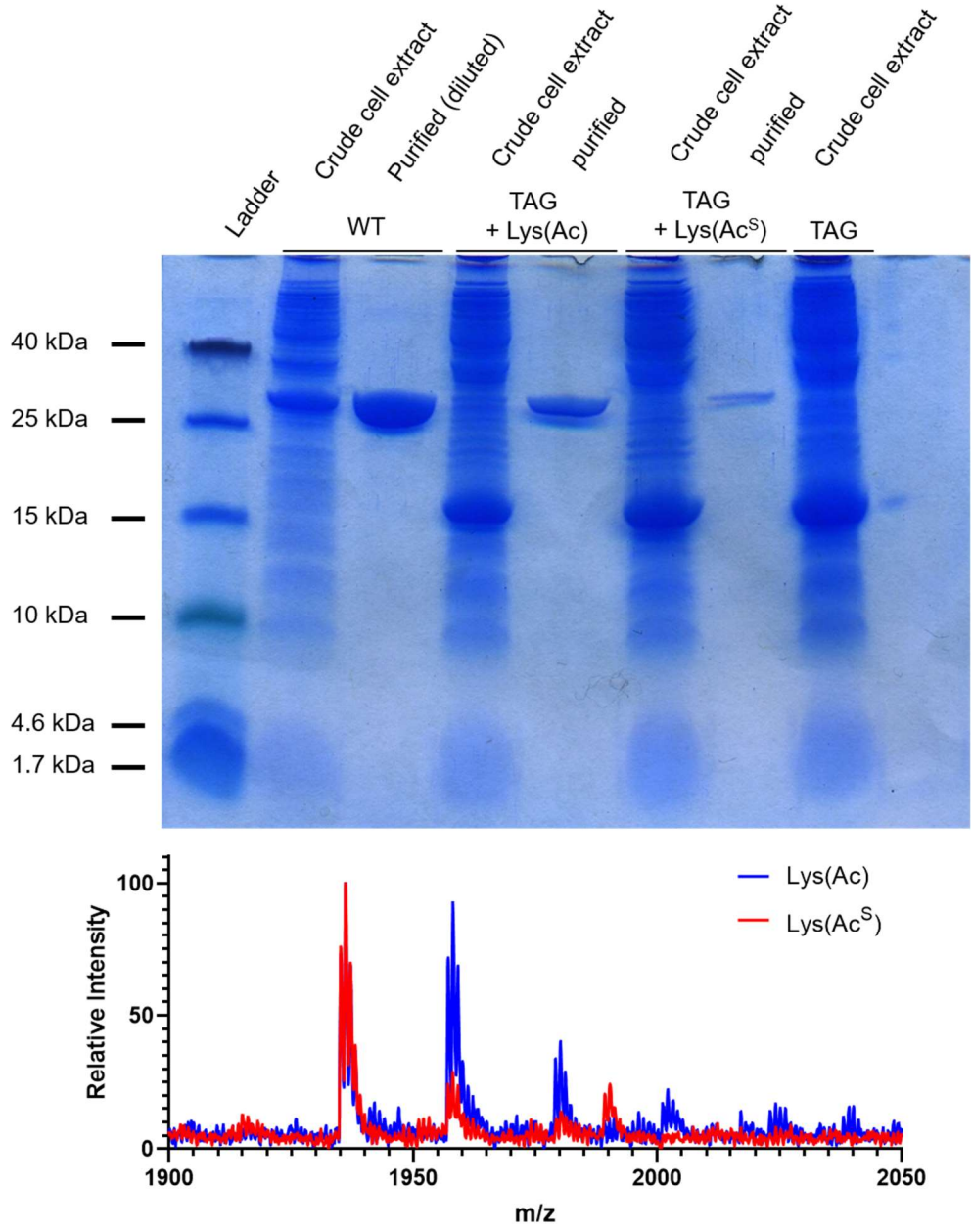


Figure 7-4: Result of attempted UAA incorporation using mmchAcK3 synthetase. Top: SDS-PAGE of crude cells extract and purified protein for WT (sfGFP) and TAG (sfGFP $Y_{151}Z$) constructs. Expression were performed in the presence of 10 mM UAA. Purified WT protein was diluted (1:5) before loading to not overload gel. MW: sfGFP: 27.7 kDa, $sfGFP_{1-150}$: 16.9 kDa. Bottom: MALDI-TOF MS of tryptic digest shows that in both cases fragment 141-156 didn't contain thioamino acid. Expected masses: $sfGFP_{141-156}$ $Y_{151}K(Ac)$: 1,934.96; : $sfGFP_{141-156}$ $Y_{151}K(Ac^S)$: 1,950.94. observed: 1,935.13 for Lys(Ac) incorporation; 1,935.14 for Lys(Ac^S) incorporation.

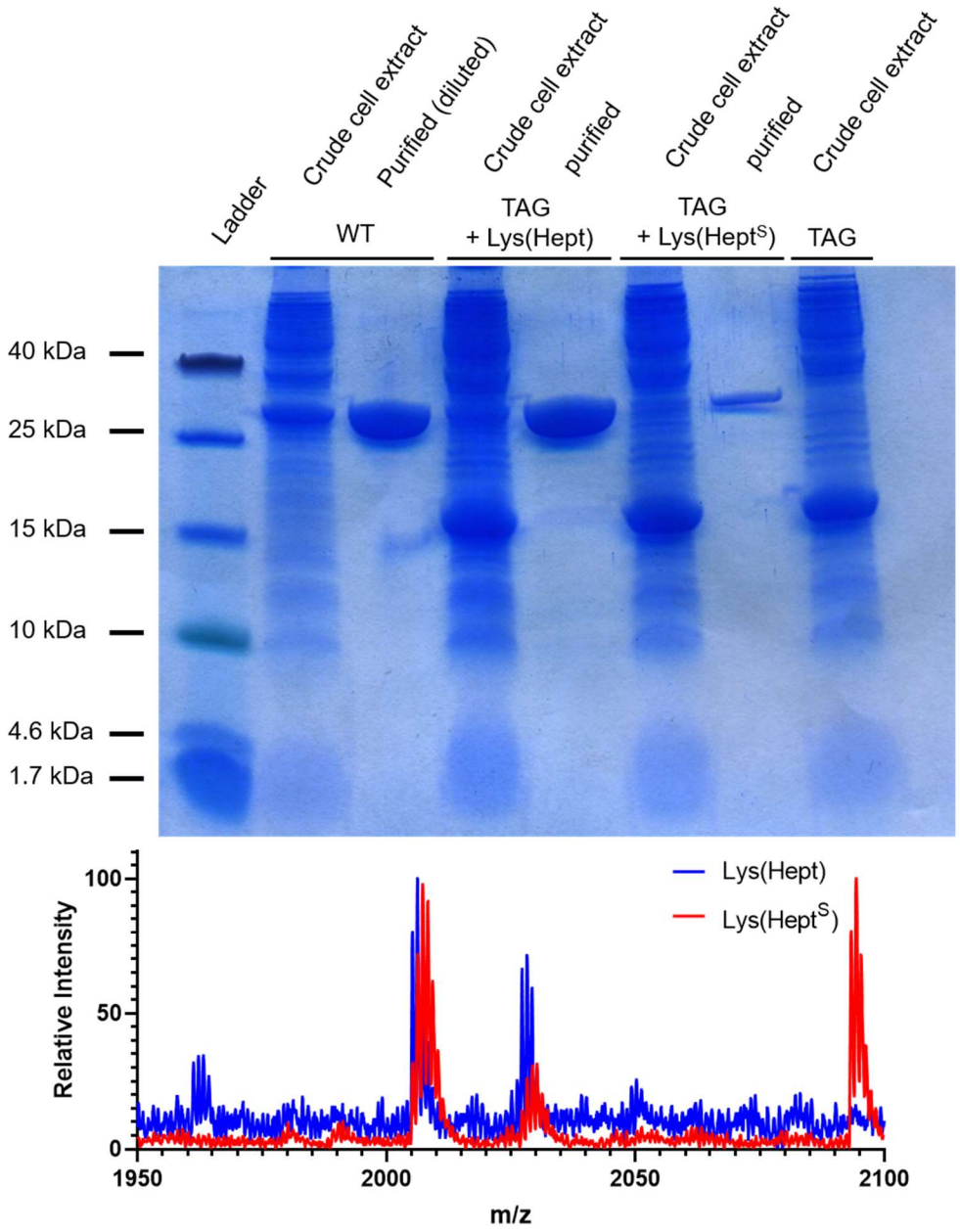


Figure 7-5: Result of attempted UAA incorporation using mmOctK synthetase. Top: SDS-PAGE of crude cells extract and purified protein for WT (sfGFP) and TAG (sfGFP $Y_{151}Z$) constructs. Expression were performed in the presence of 1 mM UAA. Purified WT protein was diluted (1:5) before loading to not overload gel. MW: sfGFP: 27.7 kDa, $sfGFP_{1-150}$: 16.9 kDa. Bottom: MALDI-TOF MS of tryptic digest shows that in both cases fragment 141-156 didn't contain thioamino acid. Expected masses: $sfGFP_{141-156}$ $Y_{151}K$(Hept): 2,005.04; : $sfGFP_{141-156}$ $Y_{151}K(Hept^S)$: 2,021.02. observed: 2,005.19 for Lys(Hept) incorporation; 2,005.25 for Lys($Hept^S$) incorporation.

7.3 Conclusion

We attempted genetic incorporation of side-chain thioamide Lys(Ac^S) and Lys($Hept^S$) using newly evolved aminoacyl-tRNA synthetases and amber suppression. The thio-amino acids and their cognate oxo-control were synthesized, since only Lys(Ac) was commercially available. While the synthesis of Lys(Ac^S) was straightforward, the synthesis of Lys($Hept^S$) was low-yielding and might need optimization in the future. However, we were able to obtain the amounts needed for small scale test expressions.

Test expression showed good growth and high cell densities after 24 hours of expression in autoinducing media, indicating that amino acids did not have an obvious negative impact on cell viability. Fluorescence measurements as well as SDS-PAGE analysis of crude cell extracts and purified proteins indicate low amounts of protein. Incorporation of thio-amino acids further reduced protein yields by about 2/3.

Tryptic digest of purified peptides followed by MALDI-TOF MS analysis indicates that none of the thio-amino acids were incorporated and only their cognate oxo version was incorporated. While we are uncertain of the source of this amino acid, we hypothesize that the thio-amino acid undergoes S-to-O exchange and that only the oxo version can get recognized by the synthetase.

The field of genetic code expansion is rapidly evolving with new and improved synthetases being described frequently. We hope that in the future a synthetase capable of incorporating N^{ε}-thioalkyl lysine derivatives will become available.

7.4 Materials and Methods

General Information Lysine derivatives and PyAOP were purchased from ChemImpex (Wood Dale, IL, USA). cOmplete Mini, EDTA-free protease inhibitor tablets

were purchased from Roche Diagnostics (Mannheim, Germany). High-Density Nickel Agarose Beads were purchased from GoldBio (St. Louis, MO, USA). Amicon Ultra Spin Filters were purchased from EMD Millipore. Sephadex G-25 PD-10 desalting columns were purchased from GE Healthcare (Chicago, IL, USA). All other reagents, solvents and materials were purchased from Fisher Scientific (Pittsburgh, PA, USA) or Sigma-Aldrich (St. Louis, MO, USA) and used without further purification unless otherwise specified.

High resolution electrospray ionization mass spectra (ESI-HRMS) were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA, USA). Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. UV-Vis absorption measurements were performed on a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA) or Thermo Fisher Scientific Genesys 150 UV/Vis Spectrophotometer (Waltham, MA, USA). Since OD_{600} measurements are instrument specific (measured scattering intensity depends on the distance between sample and detector), values obtained on the Genesys 150 were multiplied by a factor of 1.5 (empirically determined) to match values obtained on HP 8452A. Fluorescence data was acquired on a Photon Technologies International (PTI) QuantaMaster40 fluorometer (currently Horiba Scientific, Edison, NJ, USA). Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Matrix assisted laser desorption/ionization with time-of-flight detector (MALDI-TOF) mass spectra were acquired on a Bruker Ultraflex III or Microflex LRF instrument. Reverse-phase purification of small molecules was performed on a Biotage Isolera System on Biotage SNAP Ultra C18 columns (Charlotte, NC, USA).

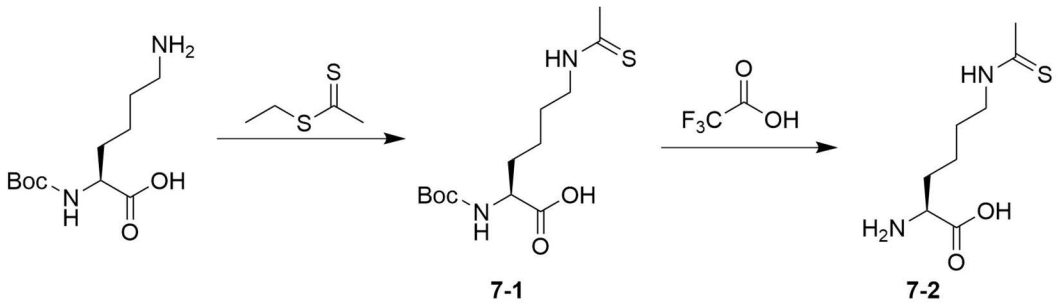


Figure 7-6: Synthetic scheme for the synthesis of Lys(Ac^S)-OH (7-2).

Synthesis of Boc-Lys(Ac^S)-OH (7-1). Boc-Lys-OH (492 mg, 2.0 mmoles) was suspended in 4.4 mL EtOH and 4.0 mL of a 10% (w/v) Na_2CO_3 solution were added. Ethyl dithioacetate (252 µL, 2.2 mmoles) were added and stirred overnight. Next day the solvent was removed *in vacuo*. The crude reaction mixture was dissolved in 10 mL H_2O and 3 M HCl was added dropwise while stirring until solution was milky white and pH was ~2. The product was extracted 3 times with CH_2Cl_2 before being dried over $MgSO_4$. After filtration the solvent was removed in vacuo to yield the product as orange foam (567 mg, 1.86 mmoles, 93.3%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 10.88 (s, 1H), 8.37 (s, 1H), 5.35 (d, *J* = 7.7 Hz, 1H), 4.07 (d, *J* = 75.4 Hz, 1H), 3.52 (s, 2H), 2.50 – 2.34 (m, 3H), 1.77 (s, 1H), 1.69 – 1.49 (m, 3H), 1.42 – 1.23 (m, 11H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 200.42, 176.09, 155.94, 81.99, 80.35, 77.42, 77.16, 76.91, 54.37, 52.99, 45.88, 33.50, 31.90, 28.16, 26.94, 22.68.

Synthesis of Lys(Ac^S)-OH (7-2). Boc-Lys(AcS)-OH (**7-1**, 567 mg, 1.86 mmoles) were dissolved in 3 mL CH_2Cl_2. 3 mL Trifluoroacetic acid were added and the reaction was stirred at room temperature for 45 minutes, after which the solvent was removed *in vacuo*. The crude product was dissolved in 2 mL MeCN and precipitated with 90 mL of cold ether. The precipitate was pelleted by centrifugation at 4,000 RPM for 15 minutes and the supernatant was discarded. The pellet was allowed to air dry for 15 minutes before

being redissolved in 3 mL MeCN/H_2O (1:1) and dried by lyophilization. The product was obtained in high purity as white powder (542 mg, 1.70 mmoles, 91.6%). ^{1}H NMR (500 MHz, Deuterium Oxide) δ 3.77 (t, *J* = 6.1 Hz, 1H), 3.64 (t, *J* = 7.1 Hz, 2H), 2.52 (s, 3H), 2.02 – 1.84 (m, 2H), 1.80 – 1.67 (m, 2H), 1.55 – 1.37 (m, 2H). ESI$^+$-HRMS calculated for $C_8H_{16}N_2O_2SNa^+$: 259.2022; found: $[M+Na]^+$: 259.2004.

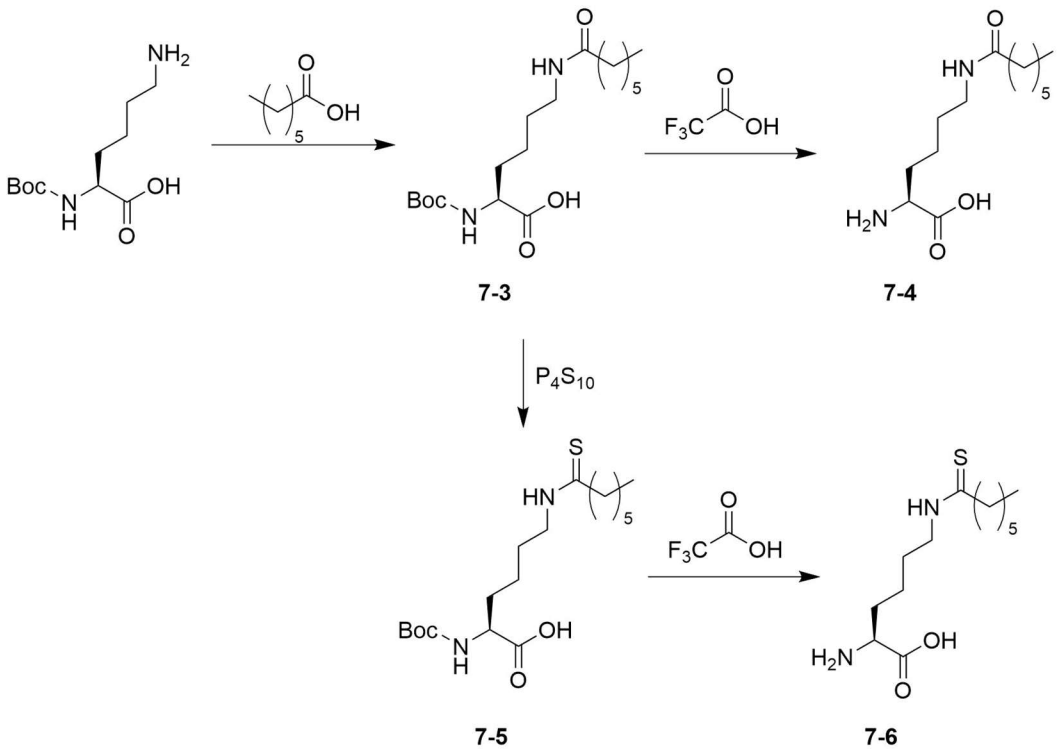


Figure 7-7: Synthetic scheme for the synthesis of Lys(Hept)-OH (7-4) and Lys(Hept^S)-OH (7-6).

Synthesis of Boc-Lys(Hept)-OH (7-3). Heptanoic acid (312 µL, 2.2 mmoles) and DIPEA (348 µL, 2.0 mmoles) dissolved in 10 mL CH_2Cl_2. PyAOP (1.04 g, 2.0 mmoles) added and solution turned yellow. Reaction stirred for 1 hour before Boc-Lys-OH (492 mg, 2.0 mmoles) was added. 2 mL of MeOH added to aid solubility and solution was sonicated until clear. Reaction was stirred at room temperature overnight. Next day the solvent was removed *in vacuo* before being redissolved in MeOH and purified by reverse phase

chromatography on Biotage system using a SNAP Ultra C18 (60g) column and gradient **7A** at a flowrate of 60 mL/min. Fractions containing pure product were combined and lyophilized. Product yielded as clear oil (653 mg /1.82 mmoles / 91.2%). ESI^+-HRMS calculated for $C_{18}H_{35}N_2O_5^+$: 359.2546; found: $[M+H]^+$: 359.2552.

Synthesis of Lys(Hept)-OH (7-4). Boc-Lys(Hept)-OH (**7-3**; 326 mg, 0.91 mmoles) was dissolved in 2 mL CH_2Cl_2. 2 mL of TFA added and stirred for 1 hour. Solvent removed *in vacuo* and redissolved in minimal amount of MeOH for reverse phase purification on Biotage system using a SNAP Ultra C18 (12g) column and gradient **7B** at a flowrate of 12 mL/min. Pure fractions were combined and lyophilized to yield a clear oil. Product redissolved in MeCN/H_2O (1:1) and lyophilized overnight again to give a white powder (212 mg / 0.57 mmoles / 62.6%). For NMR, 50 mg of the sample were suspended in 600 µL D_2O and 15 µL NaOD (40% (w/v) in D_2O) were added for solubility. ^1H NMR (500 MHz, Deuterium Oxide) δ 3.22 (t, *J* = 6.4 Hz, 1H), 3.18 (t, *J* = 6.8 Hz, 2H), 2.23 (t, *J* = 7.2 Hz, 2H), 1.69 – 1.47 (m, 6H), 1.39 – 1.31 (m, 2H), 1.29 (s, 6H), 0.86 (t, *J* = 6.6 Hz, 3H). ^{13}C NMR (126 MHz, D_2O) δ 183.55, 177.03, 55.91, 39.13, 35.84, 34.33, 30.69, 28.33, 27.84, 25.39, 22.48, 21.87, 13.35. ESI^+-HRMS calculated for $C_{13}H_{27}N_2O_3^+$: 259.2022; found: $[M+H]^+$: 259.2004.

Synthesis of Boc-Lys(Hept^S)-OH (7-5). P_4S_{10} (609 mg, 1.37 mmoles) and Na_2CO_3 (106 mg, 1.37 mmoles) were suspended in THF and stirred under Ar for 1 hour until solution became clear and yellow. Since starting material was a thick oil, thionation solution was added to vial containing Boc-Lys(Hept)-OH (**7-3**; 326 mg, 0.91 mmoles). After solubilizing everything by sonication, the reaction was stirred overnight under Ar. Next day, the solvent was removed *in vacuo*, redissolved in minimal DMSO and purified by reverse phase on Biotage system using a SNAP Ultra C18 (30g) column and gradient **7C**

at a flowrate of 12 mL/min.. Fractions containing pure product were combined and lyophilized. Product yielded as reddish, gum-like solid (51.7 mg, 138 μmoles, 15.2%). ESI$^+$-HRMS calculated for $C_{18}H_{35}N_2O_4S^+$: 375.2318; found: $[M+H]^+$: 375.2552.

Synthesis of Lys(Hept^S)-OH (7-6). Boc-Lys(Hept^S)-OH (**7-5**; 51.7 mg, 138 μmoles) was dissolved in 2 mL CH_2Cl_2. 2 mL TFA added and stirred for 1 hour and solvent removed in vacuo. Crude product redissolved in a minimal amount of DMSO and purified on Biotage system using a SNAP Ultra C18 (12g) column and gradient **7B** at a flowrate of 12 mL/min. Fractions containing product combined and lyophilized. Product yielded as fluffy white powder (18.0 mg / 46.4 μmoles / 33.6%). For NMR, 10 mg of the sample were suspended in 600 μL D_2O and 15 μL NaOD (40% (w/v) in D_2O) were added for solubility. ^{1}H NMR (500 MHz, Deuterium Oxide) δ 3.55 (t, *J* = 7.1 Hz, 2H), 3.20 (t, *J* = 6.3 Hz, 1H), 2.61 (t, *J* = 7.4 Hz, 2H), 1.73 – 1.48 (m, 6H), 1.35 (p, *J* = 7.7 Hz, 2H), 1.26 (s, 6H), 0.83 (t, *J* = 6.5 Hz, 3H). ^{13}C NMR (126 MHz, D_2O) δ 202.59, 183.62, 55.88, 46.81, 45.77, 34.37, 30.74, 28.70, 27.47, 27.11, 22.65, 21.85, 13.32. ESI$^+$-HRMS calculated for $C_{13}H_{27}N_2O_2S^+$: 275.1793; found: $[M+H]^+$: 275.1782.

General information about cloning. All PCR reactions were carried out on a T100 thermocycler from Bio-Rad (Hercules, CA, USA). Primers and double stranded DNA fragments (gBlocks) were purchased from Integrated DNA Technologies (Coralville, IA, USA). dNTPs, SYBR Safe DNA gel stain and Agarose were purchased from Invitrogen (Waltham, MA, USA). Q5 High-Fidelity DNA polymerase, Q5 High-Fidelity DNA polymerase Master Mix, HiFi DNA Assembly Master Mix, Restriction Enzymes, competent High Efficiency *E. coli* cells (NEB 5-alpha and NEB 10-beta), T4 DNA Polynucleotide Kinase, T4 DNA Ligase and Monarch Gel Extraction Kit were purchased from New England Biolabs (Ipswich, MA, USA). Gel Green DNA gel stain was purchased from

Biotium (Freemont, CA, USA). DNA Clean & Concentrator Kit and Plasmid Miniprep Kit were purchased from Zymo Research (Irvine, CA, USA). DNA concentrations were determined with a TECAN Nanoquant plate on an Infinite M1000Pro plate reader (Tecan, Männedorf, Switzerland). DNA sequencing (Sanger sequencing) was performed at the University of Pennsylvania DNA Sequencing Facility (Philadelphia, PA, USA). Agarose gels were visualized on a Typhoon Imager (GE Healthcare, Marlborough, MA, USA) or a SmartBlue transilluminator (Southern Labware, Cumming, GA 30028). pET His6-SUMO-TEV (2S-T) plasmid was a gift from Scott Gradia (Addgene plasmid # 29711; http://n2t.net/addgene:29711; RRID: Addgene_29711). pEVOL-mmOctKRS plasmid was a gift from Prof. Wenshe Liu (Department of Chemistry, Texas A&M University, College Station, TX, USA). pET22b-T5 sfGFP and pET22b-T5 sfGFP(Y151Z) plasmids were a gift from Prof. Abhishek Chatterjee (Department of Chemistry, Boston College, Chestnut Hill, MA, USA).

General procedure for PCR. Primers were dissolved in Milli-Q grade water to be at a concentration of 50 µM. The amount of plasmid used ranged from 0.1 – 1 ng total DNA. Primers, plasmid and buffer components were mixed according to manufacturer recommendations. PCR was carried out for 35 cycles at annealing temperatures calculated with the New England Biolabs Tm calculator tool (https://tmcalculator.neb.com). Alternatively, a ‘Touchdown protocol’ was used, where annealing started with 70° in the first cycle and was reduced in each of the 19 subsequent cycles by 0.5°C each, followed by 15 rounds of amplification at 55°C.

Following amplification PCR product was isolated using a DNA Concentrator & Cleanup Kit. PCR product was quantified to verify successful PCR amplification. PCR product was circularized by mixing 2 µL PCR product with 5 µL Milli-Q grade water, 1 µL

10x T4 DNA Ligase Buffer, 1 µL T4 Polynucleotide Kinase, 0.5 µL T4 DNA Ligase and 0.5 µL DpnI. This mixture was incubated at 37°C for 45 minutes, cooled on ice and 50 µL of competent NEB 5-alpha cells were added. Cells were transformed and plated on an LB-Agar plate containing appropriate antibiotic. Individual colonies were picked, grown up in 5 mL LB-medium and their DNA was extracted using a Plasmid Miniprep Kit according to manufacturer's protocol. Quantified DNA was submitted for sanger sequence analysis to verify correct sequence. Sequencing primer was provided by Sequencing Center and aligns with T7 promoter unless otherwise noted.

General procedure for Gibson Assembly. Primers were dissolved in Milli-Q grade water to be at a concentration of 50 µM. The amount of plasmid used ranged from 0.1 – 1 ng total DNA. Vector and insert were amplified simultaneously using a 'Touchdown protocol', where annealing started with 70° in the first cycle and was reduced in each of the 19 subsequent cycles by 0.5°C each, followed by 15 amplifications at 55°C. PCR product was purified using a 1% Agarose gel containing SYBR Safe or GelGreen Dye. Gel was visualized using a Blue light transilluminator and bands of the expected size were excised. DNA was extracted was extracted using a DNA gel extraction kit according to manufacturer's protocol and isolated DNA was quantified. DNA assembly was performed with 100 ng of vector DNA and 2 equiv of insert DNA (5 equiv for insert <300 bp) using a HiFi DNA Assembly Mix according to manufacturer's protocol. 2 µL of the ligated product were transformed in 50 µL competent NEB10*β* cells according to manufacturer's instructions and plated on an LB-Agar plate containing appropriate antibiotic. Individual colonies were picked, grown up in 5 mL LB-medium and their DNA was extracted using a Plasmid Miniprep Kit according to manufacturer's protocol. Quantified DNA was submitted

for sanger sequence analysis to verify correct sequence. Sequencing primer was provided by Sequencing Center and aligns with T7 promoter unless otherwise noted.

Cloning of pEVOL-mmchAcK3RS. Plasmid was constructed using Gibson Assembly. The Vector donor plasmid was pEVOL-mmOctKRS and was a present from Prof. Wenshe Liu. The insert sequence was obtained from the published amino acid sequence for chAcKRS3[237]. The sequence was optimized for *E. coli* using IDT's codon optimization tool and was purchased as gBlock from IDT. Sequencing was submitted with custom sequencing primers from 3' and 5' end to cover the whole range of interest.

gBlock (5' to 3' strand only):

```
GCTAACAGGA GGAATTACTA GTATGGACAA AAAACCGCTG GATGTTCTGA
TTAGCGCAAC CGGTCTGTGG ATGAGCCGCA CCGGTACACT GCATAAAATC
AAACATTATG AAATCAGCCG CAGCAAAATC TATATTGAAA TGGCATGTGG
TGATCATCTG GTGGTGAATA ATAGCCGTAG CTGTCGTCCG GCACGTGCAT
TTCGTTATCA CAAATATCGT AAAACCTGCA AACGTTGTCG TGTTAGCGGT
GAAGATATTA ACAATTTTCT GACCCGTAGC ACCGAAGGTA AAACCAGCGT
TAAAGTTAAA GTTGTGAGCG AGCCGAAAGT GAAAAAAGCA ATGCCGAAAA
GCGTTAGCCG TGCACCGAAA CCTCTGGAAA ATCCGGTTAG CGCAAAAGCA
AGCACCGATA CCAGCCGTAG CGTTCCGAGT CCGGCAAAAA GCACCCCGAA
TAGTCCGGTT CCGACCAGCG CAAGCGCACC GGCACTGACC AAAAGCCAGA
CCGATCGCCT GGAAGTTCTG CTGAATCCGA AAGATGAAAT TAGCCTGAAT
AGCGGTAAAC CGTTTCGTGA ACTGGAAAGC GAACTGCTGA GCCGTCGTAA
AAAAGATCTG CAGCAGATTT ATGCCGAAGA ACGCGAAAAC TATCTGGGTA
AACTGGAACG TGAAATCACC CGTTTTTTTG TGGATCGTGG TTTCCTGGAA
ATCAAAAGCC CGATTCTGAT TCCGCTGGAA TATATTGAAC GTATGGGCAT
TGATAACGAT ACCGAACTGA GCAAACAAAT CTTTCGCGTG GATAAAAACT
TTTGTCTGCG TCCGATGATG GCACCGAACA TTTTTAACTA TGCACGCAAA
CTGGATCGTG CACTGCCGGA TCCGATTAAA ATCTTTGAAA TTGGTCCGTG
CTACCGCAAA GAAAGTGATG GTAAAGAACA CCTGGAAGAA TTTACGATGC
TGAACTTTTT TCAGATGGGT AGCGGTTGTA CCCGTGAAAA TCTGGAAAGC
ATTATTACCG ATTTTCTGAA CCATCTGGGC ATCGATTTCA AAATTGTTGG
TGATAGCTGC ATGGTGTATG GTGATACCCT GGATGTTATG CATGGTGATC
TGGAACTGAG TAGCGCAGTT GTTGGTCCGA TTCCTCTGGA TCGCGAATGG
GGTATTGATA AACCGTGGAT TGGTGCCGGT TTTGGTCTGG AACGCCTGCT
GAAAGTTAAA CACGACTTCA AAAACATTAA ACGTGCAGCA CGTAGCGAGA
GCTATTACAA TGGTATTAGC ACCAACCTGT AAGTCGACCA TCATCATCAT
```

Forward primer for vector:

```
5′ – ACCAACCTGT AAGTCGAC – 3′
```

Reverse complement primer for vector:

```
5′ – CATACTAGTA ATTCCTCCTG TTAGC – 3′
```

Forward primer for sequencing:
5′ – CTACCTGACG CTTTTTATCG CA – 3′

Reverse complement primer for sequencing:
5′ – TTTATCAGAC CGCTTCTGCG TT – 3′

Cloning of pET sfGFP. Plasmid was constructed using Gibson Assembly. The Vector donor plasmid was pET His6-SUMO-TEV -EcoR1. The insert donor plasmid was pET22b-T5-sfGFP and was a gift from Prof. Abhishek Chatterjee. Primers were designed to introduce 10 nt overlaps, i.e. the 5' and 3' and of insert contained 10 nt of the vector and *vice versa*. This created the necessary 20 nt overlap for Gibson assembly.

Forward primer for vector:
5′ – CCACCACTAA CGGATCCGCG ATCG – 3′

Reverse complement primer for vector:
5′ – CTTTGCTCAT ATGTATATCT CCTTCTTAAA GTTAAACAAA ATTATTTCTA G– 3′

Forward primer for insert:
5′ – AGATATACAT ATGAGCAAAG GAGAAGAAC – 3′

Reverse complement primer for insert:
5′ – CGCGGATCCG TTAGTGGTGG TGGTGGT – 3′

Cloning of pET sfGFP($Y_{151}Z$). Plasmid was constructed using Gibson Assembly. The Vector donor plasmid was pET His6-SUMO-TEV -EcoR1. The insert donor plasmid was pET22b-T5-sfGFP($Y_{151}Z$) and was a gift from Prof. Abhishek Chatterjee. Primers were designed to introduce 10 nt overlaps, i.e. the 5' and 3' and of insert contained 10 nt of the vector and *vice versa*. This created the necessary 20 nt overlap for Gibson assembly.

Forward primer for vector:
5′ – CCACCACTAA CGGATCCGCG ATCG – 3′

Reverse complement primer for vector:
5′ – CTTTGCTCAT ATGTATATCT CCTTCTTAAA GTTAAACAAA ATTATTTCTA G– 3′

Forward primer for insert:
5′ – AGATATACAT ATGAGCAAAG GAGAAGAAC – 3′

Reverse complement primer for insert:
5′ – CGCGGATCCG TTAGTGGTGG TGGTGGT – 3′

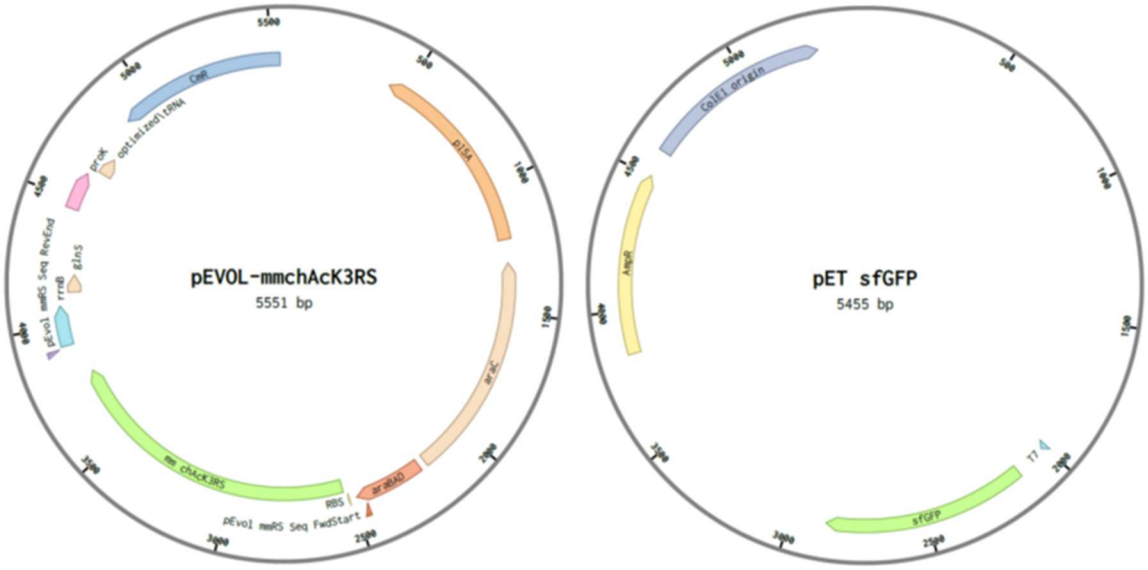


Figure 7-8: Plasmid maps showing vector backbones used. Left: pEVOL plasmid containing mmchAcK3RS gene (green) under control of araBAD. Right: pET plasmid contining sfGFP gene (green) under control of the lac operon.

Expression of sfGFP (WT or Y151Z). pET plasmid containing sfGFP (WT) was transformed into BL21 (DE3) cells and plated on LB-Agar plates containing Ampicillin (100 µg/mL). pET plasmid containing sfGFP ($Y_{151}Z$) and pEVOL plasmid containing desired synthetase (mmchAcK3RS or mmOctKRS) were co-transformed into BL21 (DE3) cells and plated on LB-Agar plates containing Ampicillin (100 µg/mL) and Chloramphenicol (25 µg/mL). Non-inducing medium (NIM) and autoinducing medium (AIM) containing Lactose and Arabinose were prepared as described by MehI[239], except for the use 25x M-salts instead of 50x M-salts (to extend shelf life). A single colony was picked and grown in 3 mL LB media containing Ampicillin (100 µg/mL) for WT sfGFP expression or Ampicillin (100 µg/mL) and Chloramphenicol (25 µg/mL) for sfGFP ($Y_{151}Z$) expression at 37°C with 250 RPM shaking overnight. Next day, 25 mL cultures with AIM in 125 mL Erlenmeyer flasks were prepared by adding 20 mM nicotinamide, UAA (1 mM for Lys($Hept^{O/S}$) and 10 mM

for Lys($Ac^{O/S}$)), and antibiotics (Ampicillin (100 μg/mL) for WT sfGFP expression or Ampicillin (100 μg/mL) and Chloramphenicol (25 μg/mL) for sfGFP ($Y_{151}Z$) expression). Secondary cultures were inoculated with 0.5% (v/v) overnight culture and grown at 37°C with 250 RPM shaking for 24 hours. At this point for each culture three 1:10 dilutions (1 mL each) were prepared and OD_{600} and fluorescence (exitation: 395 nm; emission: 510 nm; 2 nm slit width, 0.1 sec integration time) measured in triplicates. For SDS-PAGE analysis, cells from 500 μL were pelleted by spinning at 13.2 k RPM in a tabletop microcentrifuge for 3 minutes. The supernatant was discarded and cell pellet frozen until further use. When needed, cell pellets were thawed on ice and resuspended in 50 μL 1x LDS buffer containing 200 mM DTT and sample was boiled at >90°C for 5 minutes before loading 15 μL sample.

Cells were harvested by centrifugation at 4,000 RPM in a GS3 rotor and Sorvall RC-5 centrifuge for 20 min at 4°C. The supernatant was discarded and the cell pellet was suspended in 5 mL lysis buffer (40 mM Tris, 500 mM NaCl, 1 mM PMSF, pH 8.3) containing a broad-spectrum protease inhibitor. Resuspended cells were then lysed on ice by sonication (30 amps power, 1 second pulse, 2 second rest, 3 minutes total sonication time) and then pelleted at 14,000 RPM in an SS-34 rotor (Sorvall RC-5 centrifuge) for 30 min at 4°C. In the meantime, 1 mL (settled volume) of Nickel Agarose beads were washed with 5 mL equilibration buffer (50 mM HEPES, pH 7.5). The crude cell lysate was incubated with the washed nickel resin for 1 h on ice with shaking. The slurry was then added to a fritted syringe and the liquid was allowed to flow through. The resin was then washed with 5 mL of equilibration buffer, 5 mL of wash buffer 1 (50 mM HEPES, 5 mM imidazole, pH 7.5), 5 mL wash buffer 2 (50 mM HEPES, 50 mM imidazole, pH 7.5) before being eluted from the resin in 2.5 mL of elution buffer (50 mM HEPES, 300

mM imidazole, pH 7.5). The crude protein was buffer exchanged using PD-10 desalting columns equilibrated with 10 mM NH_4HCO_3 according to manufacturer's instruction (gravity protocol) and stored at -20°C until further use.

Table 7.1: HPLC Gradients used for Small Molecule Purification.

No.	CV	%B	No.	CV	%B	No.	CV	%B
7A	0 CV	25	**7B**	0 CV	10	**7C**	0 CV	10
	2 CV	25		2 CV	10		2 CV	10
	12 CV	60		12 CV	50		3 CV	30
	13 CV	60		13 CV	50		13 CV	65
	13 CV	100		13 CV	100		14 CV	65
	15 CV	100		15 CV	100		14 CV	100
							16 CV	100

Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

Chapter 8 APPENDIX

Appendix A: Synthesis of flexizyme substrates to probe the ribosomal exit tunnel

The Deutsch Lab was interested in studying the effects of steric bulk on the elongation kinetics of nascent peptides in the ribosomal exit tunnel. As part of our collaboration, we synthesized dinitrobenzylesters of various modified cysteines. These dinitrobenzylesters were used to aminoacylate tRNA in an *in vitro* aminoacylation assay using Flexizyme, a catalytic RNA developed by the Suga Lab[240]. The synthesis of these compounds and their use have been published[241].

The overall synthetic scheme is shown in Figure 8-1: The synthesis starts with the *tert*-Butyloxycarbonyl (Boc) protection of the two N_{α}-groups of Cystine to yield **8-4**. In the second step, the dintrobenzyl ester was formed resulting in compound **8-5**, before Boc-deprotection to yield cystine dinitrobenzyl ester **8-6**. After reduction of the disulfide bond, the resulting thiol was modified with *N*-etyhlmaleimide or with maleimides **8-1**, **8-2** or **8-3** resulting in compounds **8-7a-d**. Tryptophan cyanomethyl ester was synthesized as well and used as a control in experiments.

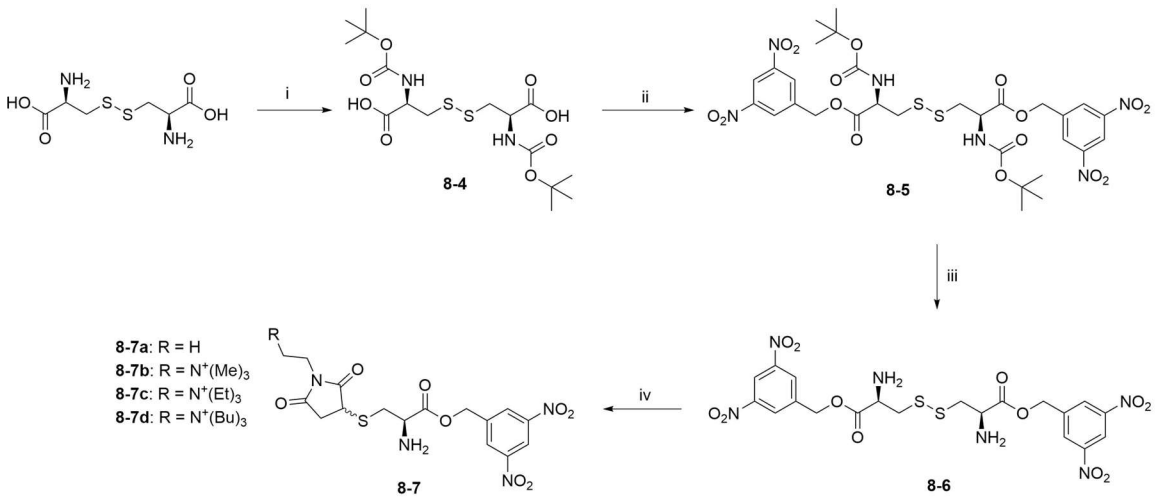


Figure 8-1: General synthesis scheme for dinitrobenzyl esters of (trialkyl-ammonium)maleimide modified cysteine. i) Boc_2O, TEA, H_2O; ii) 3,5-Dinitrobenzylchloride, NaI, DIPEA, DMF; iii) TFA; iv) TCEP, *N*-ethylmaleimide/2-8a-c, H_2O, pH 6.0

General information: All chemicals were purchased from Sigma-Aldrich or Fisher Scientific and used without further purification. Electrospray ionization mass spectra (ESI-MS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Preparative HPLC was performed on a Varian Prostar HPLC system (currently Agilent Technologies, Santa Clara, CA, USA. HPLC column (Sunfire C18 OBD, 19x150mm, 5µm) was purchased from Waters Corporation (Milford, MA, USA).

Synthesis of maleimido-quatenary ammoniums: A synthesis of maleimido-quaternary ammoniums was developed by modification of the published work of from Lu *et al*[242]. These modifications include the replacement of saturated sodium bicarbonate solution for reaction workup by a mixture of saturated sodium bicarbonate solution/water (1:3), extension of the reaction time to 2 h and performing the reaction at 0 °C. These modifications significantly reduced the amount of hydrolyzed, ring-opened byproduct **8-3**

(see Figure 8-2). Adjustment of the buffer pH to 7.8 or lower was attempted but did further improve the reaction outcome. After 2 h, the reaction was quenched by addition of 25 mL of 0.1 % trifluoroacetic acid (TFA) in water. The crude reaction mixture was purified by HPLC with the following conditions: Solvent A: 0.1% TFA in water; Solvent B: 0.1% TFA in acetonitrile (HPLC grade); Flowrate: 8 mL/min (isocratic) or 14 mL/min (gradient). Trimethylamino(TMA)-maleimide **8-2a** and Triethylamino(TEA)-maleimide **8-2b** eluted isocratically with 0% and 2% Solvent B respectively. Tributylamino(TBA)-maleimide **8-2c** eluted on gradient **8A** at approximately 28% Solvent B. Note: Purified TEA maleimide contained unreacted N-Methoxycarbonylmaleimide, which was not separable on HPLC. However, this contaminant could be separated after reaction of **8-2b** with cysteine-dinitrobenzyl ester **8-6**.

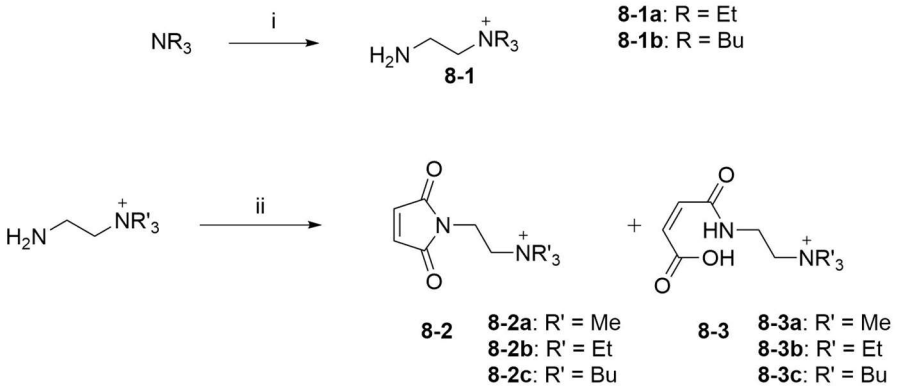


Figure 8-2: Synthesis of maleimido-quaternary ammoniums. i) (a) *N*-(Bromoethyl)-phtalimide, MeCN; (b) HBr (48%); ii) *N*-methoxycarbonylmaleimide, ¼ sat. $NaHCO_3$.

Nuclear magnetic resonance (NMR) spectroscopy and electrospray ionization mass spectrometry (ESI-MS):

8-2a: ^{1}H NMR (500 MHz, D_2O): δ 3.25 (*m*, 9H), 3.63 (*m*, 2H), 4.05 (m, 2H), 6.95 (*m*, 2H); ESI-MS: *m/z* ($[M]^+$; calculated: 183.11; found: 183.19)

8-2b: ^{1}H NMR (500 MHz, D_2O): δ 1.37 (*t*, 9H, J=7.5 Hz), 3.41 (*m*, 8H), 3.96 (m, 2H), 6.96 (*s*, 2H); ESI-MS: *m/z* ($[M]^+$; calculated: 225.16; found: 225.27)

8-2c: ^{1}H NMR (500 MHz, CD_3CN): δ 0.98 (*t*, 9H, J=7.5 Hz), 1.37 (*sextet*, 6H, J= 7.5 Hz), 1.66 (*m*, 6H), 3.17 (*m*, 6H), 3.29 (*m*, 2H), 3.80 (*m*, 2H), 6.86 (*s*, 2H); ESI-MS: *m/z* ($[M]^+$; calculated: 309.25; found: 309.44)

Synthesis of *N*-(trialkylammonium)ethylmaleimido)cysteine dinitrobenzyl esters: L-Cystine (2.08 mmol) and Triethylamine (TEA) (6.24 mmol) were dissolved in 10 mL H_2O. After a few min of stirring, the solution became clear and di-tert-butyl dicarbonate (6.24 mmol) was added. The reaction was stirred for 3 h before being partioned between 100 mL EtOAc and H_2O (50 mL each). The aqueous layer was acidified to pH 2 and extracted 3 times with EtOAc. All organic layers were combined, dried over $MgSO_4$ and filtered. Solvent was removed *in vacuo* to yield **8-4** as a white powder (1.55 mmol, 55.5%).

Disulfide **8-4** (0.70 mmol), dinitrobenzylchloride (2.8 mmol), sodium iodide (3.4 mmol) and *N*,*N*-diisopropylethylamine were dissolved in 5 mL anhydrous dimethyl-formamide and stirred under inert atmosphere overnight. The crude reaction mixture was partioned between 100 mL EtOAc and H_2O containing 10% brine (50 mL each). The organic phase was washed 3 times with 15 mL saturated ammonium chloride solution, then concentrated and the resulting residue was purified over silica (50% EtOAc in Hexanes). Fractions containing the product **8-5** were combined and solvent was removed *in vacuo*. The crude product was triturated with 20% EtOAc in Hexanes, then filtered and the filtrate was washed first with 50% EtOAc in hexanes, then with 100% EtOAc. The filtrate was dried to yield **8-5** as white powder in high purity and modest yield (0.125 mmol, 18.0%).

Disulfide **8-5** (10 μmol) was dissolved in 1 mL TFA and stirred at room temperature for 30 min, after which solvent was removed *in vacuo.* **8-6** was dissolved in 1 mL water. Tris(2-carboxyethyl)phosphine hydrochloride (15.0 μmol) added and the pH was adjusted to 6.0 with 0.1 M sodium hydroxide. A solution of maleimide (30.0 μmol) in 600 μL acetonitrile was added to the reaction and the pH was readjusted to 6.0. After 45 min, the reaction was quenched by reducing the pH through addition of 12 mL 0.1% TFA in water. The crude reaction mixture was purified by HPLC on a C18 reversed-phase column to yield **8-7a-d** in the following amounts: **8-7a**: 7.05 mg, 52.2%, **8-7b**: 7.60 mg, 46.4%, **8-7c**: 7.68 mg, 49.8%, **8-7d**: 9.55 mg, 55.3%. Solvent A: 0.1% TFA in water; Solvent B: 0.1% TFA in acetonitrile; Flowrate: 14 mL/min; Gradient **8A**. Note: HPLC retention times listed below include two peaks for **8-7b**, **8-7c**, and **8-7d** since the diasteromers were resolvable. NMR spectra were obtained for mixtures of the two diasteromers, which appear to slowly equilibrate.

NMR spectroscopy and ESI-MS:

8-4: ^{1}H NMR (500 MHz, DMF-d_7): δ 1.42 (*s*, 18H), 3.10 (*dd*, 2H, J=13.6/9.8 Hz), 3.31 (*m*, 2H), 4.45 (*dt*, 2H, J=9.0/4.2 Hz), 7.09 (d, 2H, J=8.0 Hz); ESI-MS: *m/z* ($[M+H]^+$; calculated: 441.14; found: 441.39; $[M+Na]^+$; calculated: 463.12; found: 463.38)

8-5: ^{1}H NMR (500 MHz, DMF-d_7): δ 1.39 (*s*, 18H), 3.19 (*dd*, 2H, J=14.0/9.5 Hz), 3.37 (*dd*, 2H, J=14.0/5.0 Hz), 4.60 (*dt*, 2H, J=8.5/4.5 Hz), 5.55 (*dt*, 4H, J=13.5/10.0 Hz), 7.47 (d, 2H, J=8.0 Hz), 8.80 (s, 4H), 8.89 (s, 2H); ESI-MS: *m/z* ($[M+HCO_2]^-$; calculated: 845.16; found: 845.41)

8-7a: ^{1}H NMR (500 MHz, MeCN-d_3): δ 1.08 (*dt*, 3H, J=7.2/2.5 Hz), 2.45 (*m*, 1H), 3.14 (*m*, 1H), 3.31 (*dd*, 0.5H, J=14.8/7.6 Hz), 3.46 (*m*, 2H), 3.52 (*d*, 1H, J=5.4 Hz) , 3.61 (*dd*, 0.5H, J=14.8/5.1 Hz), 3.98 (*m*, 0.5H) , 4.03 (*m*, 0.5H) , 4.44 (*dd*, 0.5H, J=5.1/2.5 Hz)

, 4.53 (*t*, 0.5H, J=5.3 Hz) , 5.44 (*m*, 2H), 8.65 (*t*, 2H, J=2.7 Hz), 8.88 (*t*, 1H, J=2.0 Hz); ESI-MS: *m/z* ($[M+H]^+$; calculated: 429.09; found: 429.39)

8-7b: ^{1}H NMR (500 MHz, MeCN-d_3): δ 2.53, (*dt*, 0.5H, J=19.0/4.5 Hz) 3.09 (*s*, 9H), 3.21 (*m*, 1.5H), 3.29 (*m*, 0.5H), 3.49 (*m*, 3.5H), 3.85(*m*, 2H), 4.14 (*dt*, 1H, J=10.0/5.0 Hz), 4.38 (*dt*, 1H, J=10.3/4.9 Hz), 5.41 (m, 2H), 8.80 (*s*, 2H), 8.88 (*s*, 1H); ESI-MS: *m/z* ($[M]^+$; calculated: 484.15; found: 484.50)

8-7c: ^{1}H NMR (500 MHz, MeCN-d_3): δ 1.267 (*t*, 9H, J=7.2 Hz), 2.54 (*dt*, 1H, J=18.7/5.5 Hz), 3.27 (*m*, 9.5H), 3.48 (*m*, 1H), 3.58 (*dt*, 0.5H), 3.76 (*m*, 2H), 4.13 (*m*, 1H) 4.44 (*m*, 1H), 5.43 (*m*, 2H), 8.66 (*s*, 2H), 8.88 (*t*, 1H, J=1.8 Hz); ESI-MS: *m/z* ($[M]^+$; calculated: 526.20; found: 526.56)

8-7d: ^{1}H NMR (500 MHz, DMSO-d_6): δ 0.94 (*dt*, 9H, J=7.5/1.5 Hz), 1.33 (*q*, 6H, J=7.5 Hz), 1.62 (*m*, 6H), 2.56 (*m*, 1H), 3.21 (*m*, 1.5H), 3.28 (*m*, 8.5H), 3.42 (*dd*, 0.5H, J=14.5/5.5 Hz), 3.50 (*dd*, 0.5H, J=14.5/6.0 Hz), 3.74 (*m*, 2H), 4.13 (*m*, 1H) 4.55 (*m*, 1H), 5.54 (*s*, 2H), 8.77 (*m*, 2H), 8.83 (*t*, 1H, J=1.8 Hz); ESI-MS: *m/z* ($[M]^+$; calculated: 610.29 found: 610.69)

Synthesis of L-Tryptophan cyanomethyl ester: N_α-Boc-L-Tryptophan (0.5 mmol), sodium iodide (1.50 mmol) and *N*,*N*-diisopropylethylamine (2.0 mmol) were dissolved in 6 mL Chloroacetonitrile and stirred overnight at room temperature. The solvent was removed in vacuo and residue dissolved in CH_2Cl_2. Crude product was partioned between 100 mL CH_2Cl_2 and H_2O containing 10% brine (50 mL each). Aqueous phase was extracted one more time with 30 mL CH_2Cl_2. The combined organic layers were washed twice with saturated ammonium chloride solution and once with H_2O, before being dried over $MgSO_4$ and filtered. The crude product was purified over silica using 2% MeOH in CH_2Cl_2 to yield **8-8** in decent yield (0.25 mmol, 50%).

Boc-L-Tryptophan cyanomethyl ester **8-8** (166 µmol) was dissolved in 1 mL $CHCl_3$ and 1 mL TFA. Reaction mixture was stirred at room temperature for 45 minutes after which the solvent was removed *in vacuo* and the crude product was rinsed twice with MeCN. The product was purified by HPLC on a C18 reversed-phase column using gradient **8B** to yield compound **8-9** in good yield (86.9 µmol, 52%).

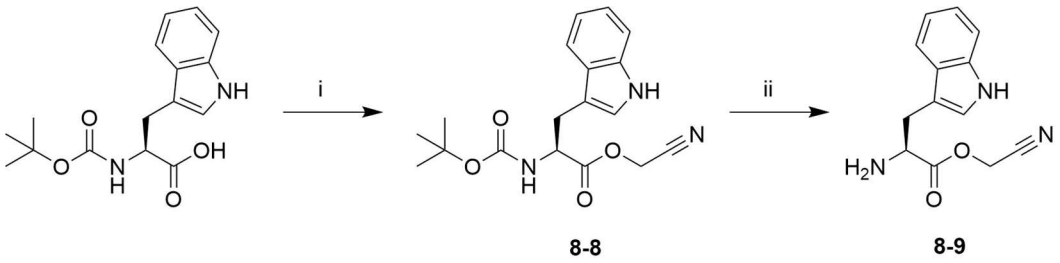


Figure 8-3: Synthesis of L-Tryptophan cyanomethyl ester. i) NaI, DIPEA, $ClCH_2CN$; ii) TFA, $CHCl_3$.

NMR spectroscopy and ESI-MS:

8-8: ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.39 (s, 1H), 7.56 (d, *J* = 7.9 Hz, 1H), 7.36 (d, *J* = 8.1 Hz, 1H), 7.22 (t, *J* = 7.4 Hz, 1H), 7.15 (t, *J* = 7.4 Hz, 1H), 7.02 (s, 1H), 5.12 (d, *J* = 8.1 Hz, 1H), 4.72 (q, *J* = 6.9, 6.5 Hz, 1H), 4.66 (d, *J* = 15.7 Hz, 1H), 4.52 (d, *J* = 15.7 Hz, 1H), 3.31 (d, *J* = 5.8 Hz, 2H), 1.45 (s, 9H). ESI-MS: *m/z* ($[M+H]^+$; calculated: 344.16; found: 344.26; $[M+Na]^+$; calculated: 366.14; found: 366.29)

8-9: ^{1}H NMR (500 MHz, DMSO-d_6) δ 11.14 (d, *J* = 1.4 Hz, 2H), 8.65 (s, 3H), 7.54 (dd, *J* = 7.9, 1.1 Hz, 1H), 7.38 (dt, *J* = 8.2, 0.9 Hz, 1H), 7.23 (d, *J* = 2.5 Hz, 1H), 7.11 (ddd, *J* = 8.1, 7.0, 1.2 Hz, 1H), 7.02 (ddd, *J* = 8.0, 7.0, 1.1 Hz, 1H), 5.07 (d, *J* = 0.7 Hz, 2H), 4.41 (t, *J* = 6.5 Hz, 1H), 3.36 – 3.24 (m, 2H). ESI-MS: *m/z* ($[M+H]^+$; calculated: 244.11; found: 244.18; $[M+Na]^+$; calculated: 266.09; found: 266.17)

Table 8.1: HPLC gradients used for compound purification.

No.	Time (min)	%B	No.	Time (min)	%B
8A	0:00	15	**8B**	0:00	2
	3:00	15		10:00	2
	7:00	25		15:00	20
	17:00	35		30:00	25
	20:00	100		33:00	100
	25:00	100		37:00	100
	26:00	15		38:00	2

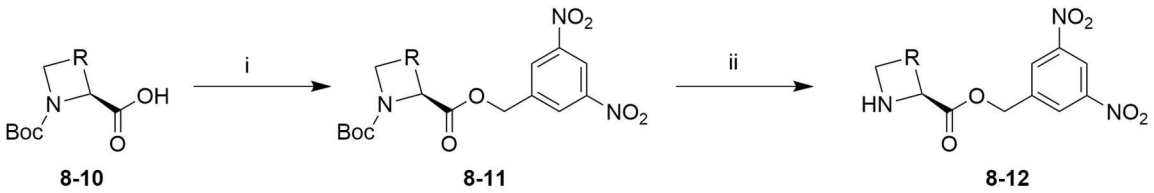
Solvent A: 0.1% (v/v) TFA in water; Solvent B: 0.1% (v/v) TFA in acetonitrile

Table 8.2: Compound purification retention times.

Compound	Gradient	Retention Time
(*N*-ethylmaleimido)-cysteine	8A	2.9 min
(*N*-ethylmaleimido)cysteine dinitrobenzyl ester (**8-7a**)	8A	13.8 min
(*N*-(trimethylammonium)ethylmaleimido)cysteine dinitrobenzyl ester (**8-7b**)	8A	7.3 / 7.8 min
(*N*-(triethylammonium)ethylmaleimido)cysteine dinitrobenzyl ester (**8-7c**)	8A	8.7 / 9.0 min
(*N*-(tributylammonium)ethylmaleimido)cysteine dinitrobenzyl ester (**8-7d**)	8A	17.7 / 18.0 min
(*N*-Methoxycarbonylmaleimido)cysteine dinitrobenzyl ester	8A	10.1 min
Cysteine dinitrobenzyl ester	8A	10.5 min
Cystine dinitrobenzyl ester (**8-6**)	8A	14.6 min
Dinitrobenzyl alcohol	8A	15.5 min
L-Tryptophan cyanomethyl ester (**8-9**)	8B	19.2 min

Appendix B: Synthesis of flexizyme substrates to study the mechanism of a quadruplet-codon reader

The Hou Lab is interested in studying the mechanism of a quadruplet codon reader tRNA. For this purpose they are generated various aminoacylated tRNA constructs using the Flexizyme method developed by the Suga Lab[240]. We provided them with purified dinitrobenzyl esters for *in vitro* aminoacylation reactions. This study has not yet been published.

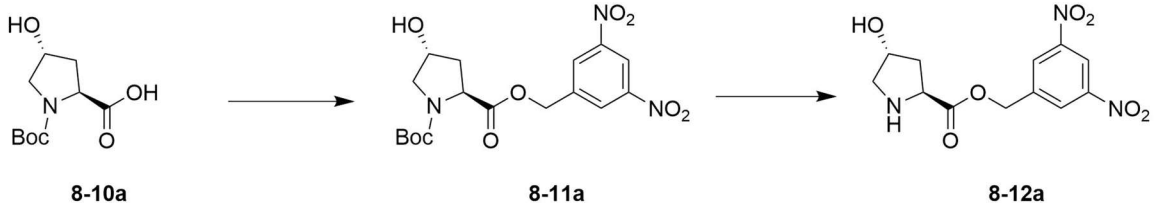


8-10 8-11 8-12

Figure 8-4: Generalized synthesis scheme for Dinitrobenzyl esters. Reagents and conditions: i) (3,5)Dinitrobenzyl chloride, NaI, DIPEA, THF; ii) TFA, $CHCl_3$

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA) or Fisher Scientific (Waltham, MA, USA). Boc protected proline derivatives were purchased from ChemImpex (Wood Dale, IL). All reagents were used without further purification. High resolution electrospray ionization mass spectra (ESI-HRMS) were collected with a Waters LCT Premier XE liquid chromatograph/mass spectrometer (Milford, MA, USA). Low resolution electrospray ionization mass spectra (ESI-LRMS) were obtained on a Waters Acquity Ultra Performance LC connected to a single quadrupole detector (SQD) mass spectrometer. Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker DRX 500 MHz instrument (Billerica, MA, USA). Preparative HPLC was performed on a Varian Prostar HPLC system (currently: Agilent Technologies, Santa Clara, CA, USA) using with water and acetonitrile each containing 0.1% trifluoroacetic acid (TFA) as mobile phase. A Waters Sunfire C18 OBD Prep column was used to purify all final products.

Synthesis of *trans*-L-4-Hydroxyproline dinitrobenzyl ester (8-12a):



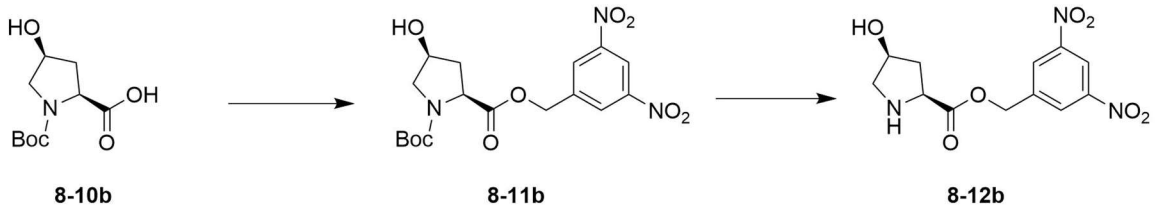
8-10a **8-11a** **8-12a**

Figure 8-5: Synthesis of *trans*-L-4-Hydroxyproline dinitrobenzyl ester (8-12a).

Boc-*trans*-L-4-hydroxyproline **8-10a** (231 mg, 1.00 mmols), 3,5-Dinitrobenzyl-chloride (543 mg, 2.5 mmols) and sodium iodide (450 mg, 3.00 mmols) dissolved in Tetrahydrofuran (10 mL). Diisopropylethylamine (695 µL, 4.00 mmols) added and stirred at room temperature overnight. The next day the crude reaction mixture was partioned between 100 mL H_2O/EtOAc (1:1). The organic phase was washed twice with saturated NH_4Cl solution and once with brine before dried over $MgSO_4$. The solution was filtered and the solvent was removed *in vacuo*. Crude product was dissolved in CH_2Cl_2 and purified over silica using 5% MeOH in CH_2Cl_2 (R_f = 0.3). Pure fractions were combined and solvent removed in vacuo. Purified compound **8-11a** was yielded in 46.0% yield as orange foam.

Boc-*trans*-L-4-hydroxyproline dinitrobenzyl ester **8-11a** (95 mg, 0.23 mmols) was dissolved in $CHCl_3$ (1mL) and Trifluoroacetic acid (TFA; 1mL) was added under stirring. After 1 hour stirring at room temperature, the solvent was removed in vacuo. The crude product was dissolved in 50% MeCN in H_2O and purified by reverse phase HPLC. The desired product **8-11a** eluted with approximately 24% MeCN was isolated in high purity. After lyophilization the product was dissolved in DMSO to yield a 25 mM stock solution used in later experiments.

Synthesis of *cis*-L-4-Hydroxyproline dinitrobenzyl ester (8-12b):



8-10b **8-11b** **8-12b**

Figure 8-6:Synthesis of *cis*-L-4-Hydroxyproline dinitrobenzyl ester (8-12b).

Boc-*cis*-L-4-hydroxyproline **8-10b** (231 mg, 1.00 mmols), 3,5-Dinitrobenzylchloride (543 mg, 2.5 mmols) and sodium iodide (450 mg, 3.00 mmols) dissolved in Tetrahydrofuran (10 mL). Diisopropylethylamine (695 µL, 4.00 mmols) added and stirred at room temperature overnight. The next day the crude reaction mixture was partioned between 100 mL H_2O/EtOAc (1:1). The organic phase was washed twice with saturated NH_4Cl solution and once with brine before dried over $MgSO_4$. The solution was filtered and the solvent was removed *in vacuo*. Crude product was dissolved in CH_2Cl_2 and purified over silica using 5% MeOH in CH_2Cl_2 (R_f = 0.3). Pure fractions were combined and solvent removed in vacuo. Purified compound **8-11b** was yielded in 30.4% yield as orange foam.

Boc-*cis*-L-4-hydroxyproline dinitrobenzyl ester **8-11b** (62 mg, 0.15 mmols) was dissolved in $CHCl_3$ (1mL) and TFA (1mL) was added under stirring. After 1 hour stirring at room temperature, the solvent was removed in vacuo. The crude product was dissolved in 50% MeCN in H_2O and purified by reverse phase HPLC. The desired product **8-12b** eluted with approximately 26% MeCN was isolated in high purity. After lyophilization the product was dissolved in DMSO to yield a 25 mM stock solution used in later experiments.

Synthesis of 2-Carboxyazetidine dinitrobenzyl ester (8-12c):

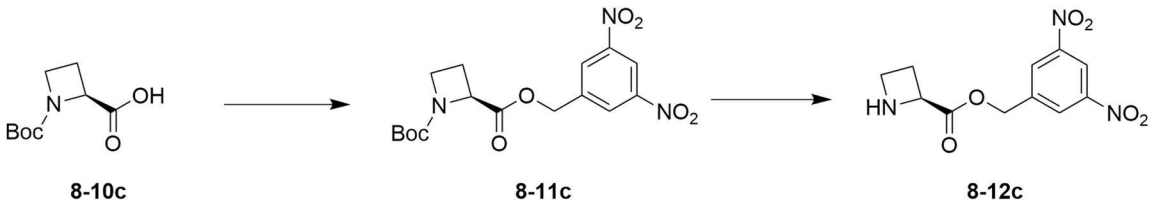
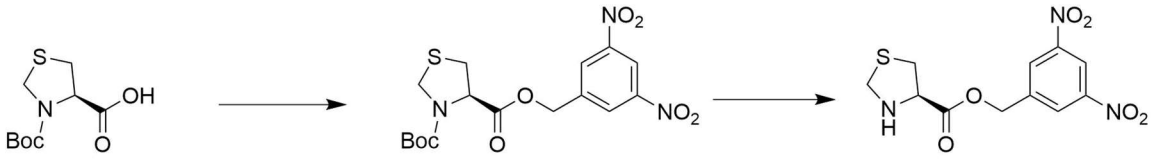


Figure 8-7: Synthesis of 2-Carboxyazetidine dinitrobenzyl ester (8-12c).

Boc-L-azetidine-2-carboxylic acid **8-10c** (201 mg, 1.00 mmols), 3,5-Dinitrobenzylchloride (543 mg, 2.5 mmols) and sodium iodide (450 mg, 3.00 mmols) dissolved in Tetrahydrofuran (10 mL). Diisopropylethylamine (695 µL, 4.00 mmols) added and stirred at room temperature overnight. The next day the crude reaction mixture was partioned between 100 mL H_2O/EtOAc (1:1). The organic phase was washed twice with saturated NH_4Cl solution and once with brine before dried over $MgSO_4$. The solution was filtered and the solvent was removed *in vacuo*. Crude product was dissolved in CH_2Cl_2 and purified over silica using 2% MeOH in CH_2Cl_2 (R_f = 0.3). Pure fractions were combined and solvent removed in vacuo. Purified compound **8-11c** was yielded in 70.3% yield as red oil.

Boc-L-azetidine-2-carboxylic acid dinitrobenzyl ester **8-11c** (134 mg, 0.35 mmols) was dissolved in $CHCl_3$ (1mL) and TFA (1mL) was added under stirring. After 1 hour stirring at room temperature, the solvent was removed in vacuo. The crude product was dissolved in 50% MeCN in H_2O and purified by reverse phase HPLC. The desired product **8-12c** eluted with approximately 27% MeCN was isolated in high purity. After lyophilization the product was dissolved in DMSO to yield a 25 mM stock solution used in later experiments.

Synthesis of Thiaproline dinitrobenzyl ester (8-12d):



8-10d **8-11d** **8-12d**

Figure 8-8: Synthesis of Thioproline dinitrobenzyl ester (8-12d).

Boc-L-Thiaproline (**8-10d**) (116 mg, 0.50 mmols) and 3,5-Dinitrobenzylchloride (217 mg, 1.0 mmols) were dissolved in Tetrahydrofuran (5 mL). Diisopropylethylamine (348 µL, 2.00 mmols) added and stirred at room temperature overnight. The next day the crude reaction mixture was partioned between 100 mL H_2O/EtOAc (1:1). The organic phase was washed twice with saturated NH_4Cl solution and once with brine before dried over $MgSO_4$. The solution was filtered and the solvent was removed *in vacuo*. Crude product was dissolved in CH_2Cl_2 and purified over silica using 1% MeOH in DCM (R_f = 0.4). Pure fractions were combined and solvent removed in vacuo. Purified compound **8-11d** was yielded in 77.5% yield as yellow oil.

Boc-L-Thiaproline dinitrobenzyl ester **8-11d** (160 mg, 0.39 mmols) was dissolved in $CHCl_3$ (1mL) and Trifluoroacetic acid (TFA; 1mL) was added under stirring. After 1 hour stirring at room temperature, the solvent was removed in vacuo. The crude product was dissolved in 50% MeCN in H_2O and purified by reverse phase HPLC. The desired product **8-11d** eluted with approximately 30% MeCN was isolated in high purity. After lyophilization the product was dissolved in DMSO to yield a 25 mM stock solution used in later experiments.

Nuclear magnetic resonance (NMR) spectroscopy and electrospray ionization mass spectrometry (ESI-MS) characterization:

1-(tert-butyl) 2-(3,5-dinitrobenzyl) (2S,4R)-4-hydroxypyrrolidine-1,2-dicarboxylate (**8-11a**): ^{1}H NMR (500 MHz, Chloroform-d) δ 8.93 (dt, J = 18.1, 2.1 Hz, 1H), 8.54 (dd, J = 14.5, 2.1 Hz, 2H), 5.39 – 5.24 (m, 2H), 4.51 – 4.42 (m, 2H), 3.56 (ddd, J = 15.8, 11.6, 4.2 Hz, 1H), 3.47 (ddt, J = 34.5, 11.6, 1.8 Hz, 1H), 2.38 – 2.24 (m, 1H), 2.08 – 1.98 (m, 1H), 1.33 (d, J = 38.8 Hz, 9H). ESI$^+$-HRMS: calculated for $C_{17}H_{21}N_3O_9Na^+$: 434.1175; found [M + Na]$^+$: 434.1171.

1-(tert-butyl) 2-(3,5-dinitrobenzyl) (2S,4S)-4-hydroxypyrrolidine-1,2-dicarboxylate (**8-11b**): ^{1}H NMR (500 MHz, Chloroform-d) δ 8.96 (dt, J = 10.0, 2.1 Hz, 1H), 8.59 (dd, J = 9.2, 2.1 Hz, 2H), 5.43 (t, J = 13.4 Hz, 1H), 5.34 (dd, J = 13.7, 7.3 Hz, 1H), 4.52 – 4.44 (m, 1H), 4.44 – 4.39 (m, 1H), 3.61 – 3.51 (m, 2H), 3.03 (s, 1H), 2.42 – 2.30 (m, 1H), 2.23 – 2.13 (m, 1H), 1.39 (d, J = 35.2 Hz, 9H). ESI$^+$-HRMS: calculated for $C_{17}H_{21}N_3O_9Na^+$: 434.1175; found [M + Na]$^+$: 434.1205.

1-(tert-butyl) 2-(3,5-dinitrobenzyl) (S)-azetidine-1,2-dicarboxylate (**8-11c**): ^{1}H NMR (500 MHz, Chloroform-d) δ 8.88 (s, 1H), 8.53 (d, J = 2.1 Hz, 2H), 5.36 (d, J = 2.8 Hz, 2H), 4.67 (dd, J = 9.2, 5.4 Hz, 1H), 3.91 (dtd, J = 45.7, 8.4, 5.9 Hz, 2H), 2.59 – 2.41 (m, 1H), 2.25 – 2.08 (m, 1H), 1.30 (s, 9H). ESI$^+$-HRMS: calculated for $C_{16}H_{21}N_3O_8Na^+$: 404.1070; found [M + Na]$^+$: 404.1086

3-(tert-butyl) 4-(3,5-dinitrobenzyl) (R)-thiazolidine-3,4-dicarboxylate (**8-11d**): 1H NMR (500 MHz, Chloroform-d) δ 8.92 (d, J = 6.6 Hz, 1H), 8.53 (d, J = 2.1 Hz, 2H), 5.47 – 5.28 (m, 2H), 4.93 – 4.76 (m, 1H), 4.56 (dd, J = 22.7, 8.9 Hz, 1H), 4.42 (t, J = 8.0 Hz, 1H), 3.40 – 3.26 (m, 1H), 3.22 – 3.13 (m, 1H), 1.37 (d, J = 36.1 Hz, 9H). ESI$^+$-HRMS: calculated for $C_{16}H_{19}N_3O_8SNa^+$: 436.0791; found [M + Na]$^+$: 436.0796 .

3,5-dinitrobenzyl (2S,4R)-4-hydroxypyrrolidine-2-carboxylate (**8-12a**): ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.75 (s, 2H), 8.82 (t, *J* = 2.1 Hz, 1H), 8.75 (d, *J* = 2.1 Hz, 2H), 5.60 (s, 1H), 5.49 (s, 2H), 4.66 (dd, *J* = 10.7, 7.6 Hz, 1H), 4.48 – 4.44 (m, 1H), 3.38 (dd, *J* = 12.1, 4.2 Hz, 1H), 3.13 (dt, *J* = 12.1, 1.6 Hz, 1H), 2.30 – 2.14 (m, 2H). ESI$^+$-HRMS: calculated for $C_{12}H_{14}N_3O_7^+$: 312.0832; found $[M + H]^+$: 312.0833.

3,5-dinitrobenzyl (2S,4S)-4-hydroxypyrrolidine-2-carboxylate (**8-12b**): ^{1}H NMR (500 MHz, DMSO-d_6) δ 9.74 (s, 2H), 8.81 (t, *J* = 2.1 Hz, 1H), 8.74 (d, *J* = 2.1 Hz, 2H), 5.51 (d, *J* = 2.1 Hz, 2H), 4.71 (dd, *J* = 9.7, 3.4 Hz, 1H), 4.40 (dt, *J* = 4.2, 2.1 Hz, 1H), 3.73 (s, 1H), 3.28 (dd, *J* = 11.9, 4.0 Hz, 1H), 3.21 (dt, *J* = 11.5, 1.5 Hz, 1H), 2.36 (ddd, *J* = 13.8, 9.7, 4.3 Hz, 1H), 2.24 (ddt, *J* = 13.4, 3.4, 1.8 Hz, 1H). ESI$^+$-HRMS: calculated for $C_{12}H_{13}N_3O_7Na^+$: 334.0651; found $[M + Na]^+$: 334.0673.

3,5-dinitrobenzyl (S)-azetidine-2-carboxylate (**8-12c**): ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.82 (t, J = 2.2 Hz, 1H), 8.75 (d, J = 2.1 Hz, 2H), 5.50 (s, 2H), 5.28 (t, J = 9.0 Hz, 1H), 4.01 (dt, J = 10.0, 8.8 Hz, 1H), 3.81 (ddd, J = 10.0, 8.6, 7.1 Hz, 2H), 2.73 – 2.64 (m, 2H). ESI$^+$-HRMS: calculated for $C_{11}H_{12}N_3O_6^+$: 282.0726; found $[M + H]^+$: 282.0728.

3,5-dinitrobenzyl (R)-thiazolidine-4-carboxylate (**8-12d**): ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.80 (t, *J* = 2.1 Hz, 1H), 8.70 (d, *J* = 2.1 Hz, 2H), 5.43 (d, *J* = 4.5 Hz, 2H), 4.31 (dd, *J* = 7.1, 5.7 Hz, 1H), 4.17 (q, *J* = 9.1 Hz, 2H), 3.13 (dd, *J* = 10.5, 7.1 Hz, 1H), 3.03 (dd, *J* = 10.4, 5.7 Hz, 1H). ESI$^+$-HRMS: calculated for $C_{11}H_{12}N_3O_6S^+$: 314.0447; found $[M + H]^+$: 314.0443.

Synthesis of fMet-Pro-Pro-DBE (8-16):

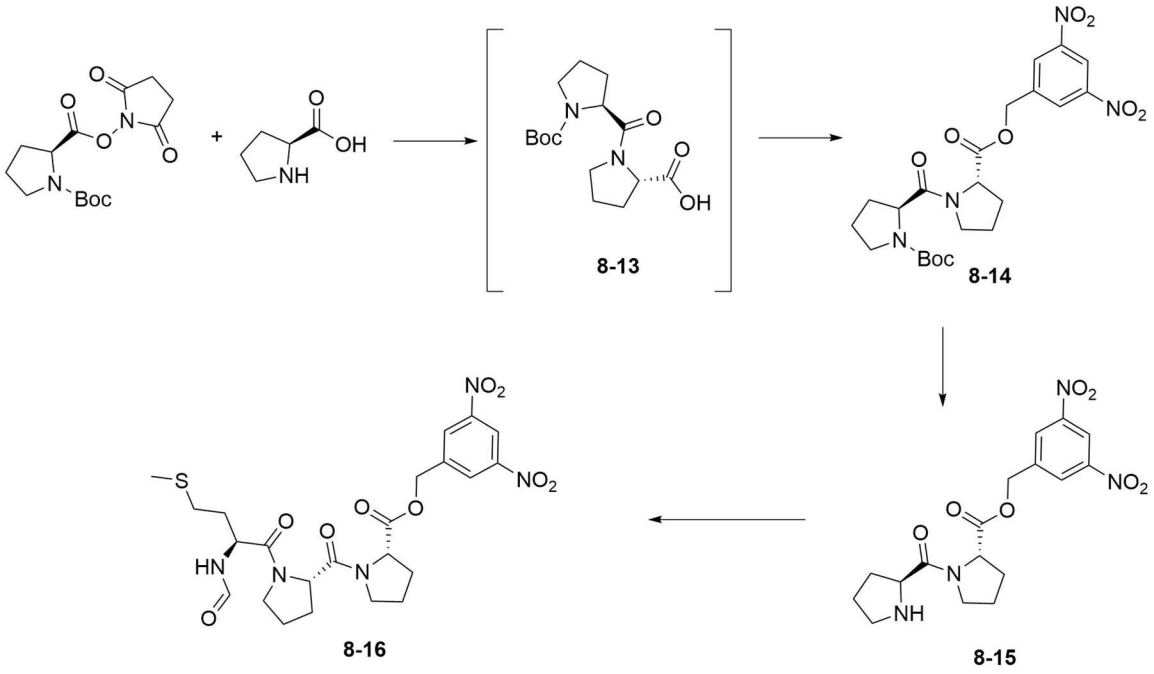


Figure 8-9: Synthetic scheme for the synthesis of fMet-Pro-Pro-DBE.

Tert-butyl (S)-2-((S)-2-(((3,5-dinitrobenzyl)oxy)carbonyl)pyrrolidine-1-carbonyl)-pyrrolidine-1-carboxylate (Boc-Pro-Pro-DBE; **8-14**): Boc-Pro-OSu (343 mg, 1.10 mmols) and L-Proline (115 mg, 1.00 mmol) dissolved in 10 mL $CHCl_3$ and 1 mL MeOH. DIPEA (348 µL, 2.00 mmols) added and stirred at room temperature for 3 hours until formation of Boc-Pro-Pro-OH (**8-13**) was complete. 3,5-Dinitrobenzylchloride (432 mg, 2.00 mmols) and more DIPEA (174 µL, 1.00 mmol) added and stirred overnight at room temperature. The next day solvent was removed in vacuo and the reaction mixture was purified over silica using 1% MeOH in CH_2Cl_2 (R_f = 0.2). The crude product was around 40% pure and was further purified by reverse phase HPLC. The desired product **8-14** eluted with approximately 65% MeCN was isolated in high purity. After lyophilization the product yielded as white powder (41.5 mg, 0.092 mmols, 9.2%). ^{1}H NMR (500 MHz, Chloroform-d) δ 8.98 – 8.95 (m, 1H), 8.53 (s, 2H), 5.32 (d, J = 10.6 Hz, 2H), 4.85 (s, 1H), 4.62 (ddd, J

= 11.0, 8.4, 4.6 Hz, 1H), 4.44 (ddd, J = 52.7, 8.4, 3.9 Hz, 1H), 3.86 – 3.78 (m, 1H), 3.71 – 3.41 (m, 2H), 3.42 – 3.33 (m, 1H), 2.31 – 2.21 (m, 1H), 2.20 – 1.87 (m, 6H), 1.87 – 1.77 (m, 1H), 1.40 (d, J = 25.9 Hz, 9H). ESI^+-HRMS: calculated for $C_{22}H_{28}N_4O_9Na^+$: 515.1754; found $[M + Na]^+$: 515.1730.

3,5-dinitrobenzyl L-prolyl-L-prolinate (Pro-Pro-DBE; **8-15**): Compound **8-14** was dissolved in 2 mL CHCl3. 2 mL TFA added and stirred at room temperature for 1 hour. The solvent was removed in vacuo and the product was rinsed twice to remove excess TFA. The product was dissolved in 50% (v/v) MeCN in H_2O and lyophlized to yield product **8-15** in quantitative yield as white powder. ^1H NMR (500 MHz, Chloroform-d) δ 8.97 (t, J = 2.1 Hz, 1H), 8.53 (d, J = 2.0 Hz, 2H), 7.55 (s, 1H), 5.43 (d, J = 13.6 Hz, 1H), 5.30 (d, J = 13.6 Hz, 1H), 4.73 (s, 1H), 4.65 (dd, J = 8.6, 5.4 Hz, 1H), 3.75 – 3.67 (m, 1H), 3.58 (dt, J = 9.9, 6.9 Hz, 1H), 3.55 – 3.46 (m, 3H), 2.61 – 2.49 (m, 1H), 2.44 – 2.33 (m, 1H), 2.22 – 1.99 (m, 6H). ESI^+-HRMS: calculated for $C_{17}H_{21}N_4O_7^+$: 393.1410; found $[M + H]^+$: 393.1414.

3,5-dinitrobenzyl formyl-L-methionyl-L-prolyl-L-prolinate (fMet-Pro-Pro-DBE; **8-16**): Formyl-methionine (26.6 mg, 150 µmols) and PyBOP (78.0 mg, 150 µmols) dissolved in 3 mL CH_2Cl_2 and DIPEA (52.2 µL, 300 µmols) added. After 10 minutes, Compound **8-15** (15.4 mg, 30.0 µmols) added and stirred at room temperature for 30 minutes, after which the reaction was complete. The solvent was removed in vacuo and the crude product was purified by reverse phase HPLC. The desired product **8-16** eluted with approximately 43% MeCN was isolated in high purity. After lyophilization the product yielded as white powder (11.0 mg, 20.0 µmols, 66.7%) and was dissolved in DMSO to yield a 25 mM stock solution used in later experiments. ESI^+-HRMS: calculated for

$C_{23}H_{29}N_5O_9SNa^+$: 574.1584; found $[M + Na]^+$: 574.1569. The ^1H NMR of compound **8-16** is shown in Figure 8-10.

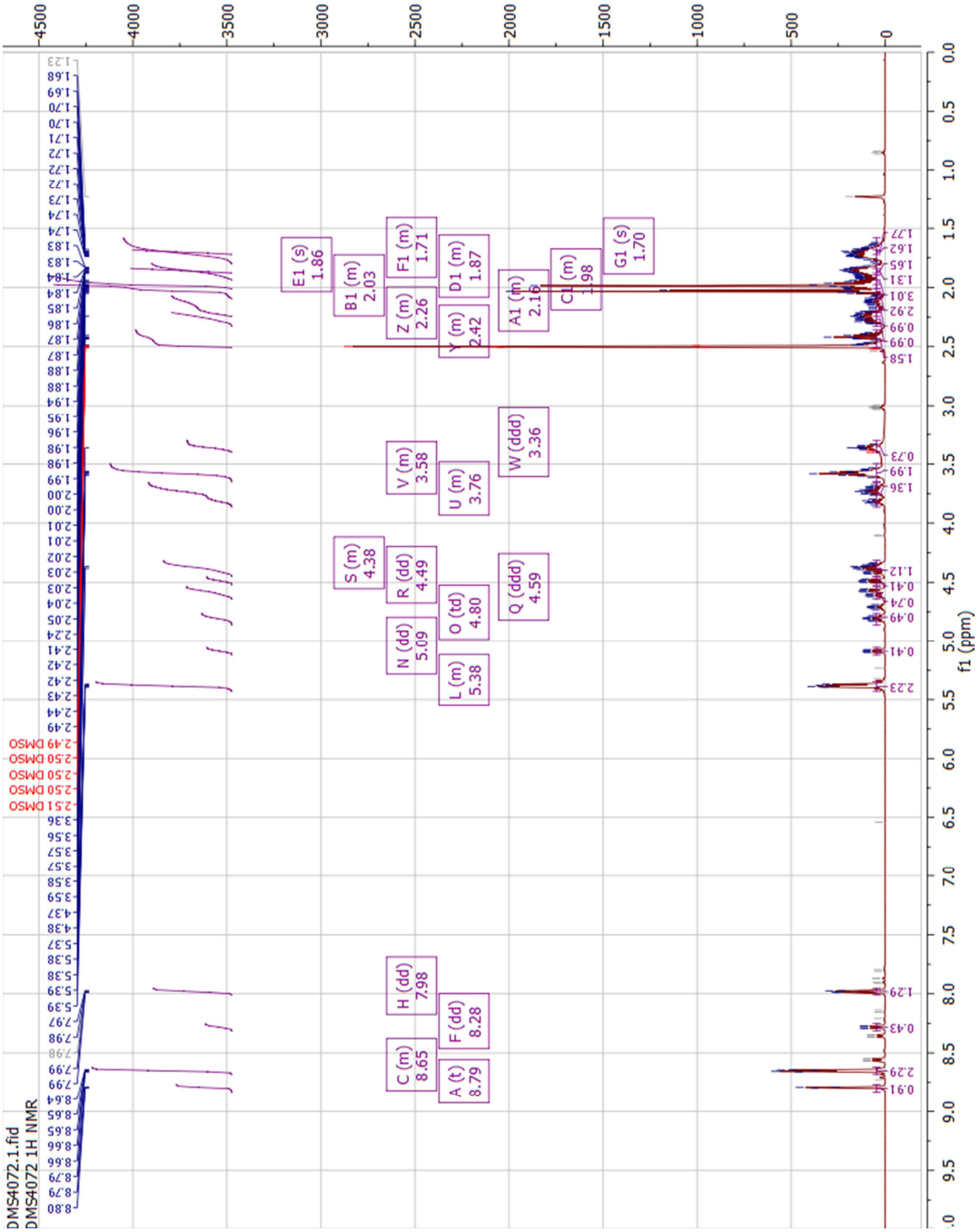


Figure 8-10: NMR of fMet-Pro-Pro-DBE.

BIBLIOGRAPHY

(1) Wang, Y. J.; Szantai-Kis, D. M.; Petersson, E. J. *Org. Biomol. Chem.* **2015**, *13*, 5074–5081.

(2) Mahanta, N.; Szantai-Kis, D. M.; Petersson, E. J.; Mitchell, D. A. *ACS Chem. Biol.* **2019**, *14*, 142–163.

(3) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. *Nat. Prod. Rep.* **2013**, *30*, 108–160.

(4) Jagodziński, T. S. *Chem. Rev.* **2002**, *103*, 197–228.

(5) Cour, T. F. M.; Hansen, H. A. S.; Clausen, K.; Lawesson, S.-O. *Int. J. Pept. Protein Res.* **1983**, *22*, 509–512.

(6) Sifferlen, T.; Rueping, M.; Gademann, K.; Jaun, B.; Seebach, D. *Helv. Chim. Acta* **1999**, *82*, 2067–2093.

(7) Kim, H. J.; Graham, D. W.; DiSpirito, A. A.; Alterman, M. A.; Galeva, N.; Larive, C. K.; Asunskis, D.; Sherwood, P. M. A. *Science* **2004**, *305*, 1612–1615.

(8) Okano, A.; James, R. C.; Pierce, J. G.; Xie, J.; Boger, D. L. *J. Am. Chem. Soc.* **2012**, *134*, 8790–8793.

(9) Prabhu, G.; Nagendra, G.; Sagar, N. R.; Pal, R.; Guru Row, T. N.; Sureshbabu, V. V. *Asian J. Org. Chem.* **2016**, *5*, 127–137.

(10) Truter, M. R. *J. Chem. Soc.* **1960**, 997–1007.

(11) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.

(12) Wiberg, K. B.; Rush, D. J. *J. Am. Chem. Soc.* **2001**, *123*, 2038–2046.

(13) Bordwell, F. G. *Acc. Chem. Res.* **1988**, *21*, 456–463.

(14) Dudek, E. P.; Dudek, G. O. *J. Org. Chem.* **1967**, *32*, 823–824.

(15) Hollósi, M.; Zewdu, M.; Kollót, E.; Majer, Z.; Kajtár, M.; Batta, G.; Kövér, K.; Sándor, P. *Int. J. Pept. Protein Res.* **1990**, *36*, 173–181.

(16) Helbing, J.; Bregy, H.; Bredenbeck, J.; Pfister, R.; Hamm, P.; Huber, R.; Wachtveitl, J.; De Vico, L.; Olivucci, M. *J. Am. Chem. Soc.* **2004**, *126*, 8823–8834.

(17) Banala, S.; Süssmuth, R. D. *ChemBioChem* **2010**, *11*, 1335–1337.

(18) Bordwell, F. G.; Algrim, D. J.; Harrelson, J. A. *J. Am. Chem. Soc.* **1988**, *110*, 5903–5904.

(19) Wiberg, K. B.; Rablen, P. R. *J. Am. Chem. Soc.* **1995**, *117*, 2201–2209.

(20) Reiner, A.; Wildemann, D.; Fischer, G.; Kiefhaber, T. *J. Am. Chem. Soc.* **2008**, *130*, 8079–8084.

(21) Lincke, T.; Behnken, S.; Ishida, K.; Roth, M.; Hertweck, C. *Angew. Chemie* **2010**, *122*, 2055–2057.

(22) Dunbar, K. L.; Scharf, D. H.; Litomska, A.; Hertweck, C. *Chem. Rev.* **2017**, *117*, 5521–5577.

(23) Schwalen, C. J.; Hudson, G. A.; Kille, B.; Mitchell, D. A. *J. Am. Chem. Soc.* **2018**, *140*, 9494–9501.

(24) Nayak, D. D.; Mahanta, N.; Mitchell, D. A.; Metcalf, W. W. *Elife* **2017**, *6*, e29218.

(25) Mahanta, N.; Liu, A.; Dong, S.; Nair, S. K.; Mitchell, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 3030–3035.

(26) Dunbar, K. L.; Büttner, H.; Molloy, E. M.; Dell, M.; Kumpfmüller, J.; Hertweck, C. *Angew. Chemie Int. Ed.* **2018**, *57*, 14080–14084.

(27) Mueller, E. G. *Nat. Chem. Biol.* **2006**, *2*, 185–194.

(28) Kenney, G. E.; Dassama, L. M. K.; Pandelia, M.-E.; Gizzi, A. S.; Martinie, R. J.; Gao, P.; DeHart, C. J.; Schachner, L. F.; Skinner, O. S.; Ro, S. Y.; Zhu, X.; Sadek, M.; Thomas, P. M.; Almo, S. C.; Bollinger, J. M.; Krebs, C.; Kelleher, N. L.; Rosenzweig, A. C. *Science* **2018**, *359*, 1411–1416.

(29) Miari, V. F.; Solanki, P.; Hleba, Y.; Stabler, R. A.; Heap, J. T. *Antimicrob. Agents Chemother.* **2017**, *61*, e00929-17.

(30) Kloss, F.; Chiriac, A. I.; Hertweck, C. *Chem. - A Eur. J.* **2014**, *20*, 15451–15458.

(31) Chiriac, A. I.; Kloss, F.; Krämer, J.; Vuong, C.; Hertweck, C.; Sahl, H.-G. *J. Antimicrob. Chemother.* **2015**, *70*, 2576–2588.

(32) Behnken, S.; Lincke, T.; Kloss, F.; Ishida, K.; Hertweck, C. *Angew. Chemie Int. Ed.* **2012**, *51*, 2425–2428.

(33) Kloss, F.; Pidot, S.; Goerls, H.; Friedrich, T.; Hertweck, C. *Angew. Chemie Int. Ed.* **2013**, *52*, 10745–10748.

(34) Süssmuth, R. D.; Mainz, A. *Angew. Chemie Int. Ed.* **2017**, *56*, 3770–3821.

(35) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. *Nat. Prod. Rep.* **2014**, *31*, 61–108.

(36) Finking, R.; Marahiel, M. A. *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.

(37) Llewellyn, N. M.; Li, Y.; Spencer, J. B. *Chem. Biol.* **2007**, *14*, 379–386.

(38) Hayakawa, Y.; Sasaki, K.; Adachi, H.; Furihata, K.; Nagai, K.; Shin-ya, K. *J. Antibiot. (Tokyo).* **2006**, *59*, 1.

(39) Izumikawa, M.; Kozone, I.; Hashimoto, J.; Kagaya, N.; Takagi, M.; Koiwai, H.; Komatsu, M.; Fujie, M.; Satoh, N.; Ikeda, H.; Shin-ya, K. *J. Antibiot. (Tokyo).* **2015**, *68*, 533.

(40) Frattaruolo, L.; Lacret, R.; Cappello, A. R.; Truman, A. W. *ACS Chem. Biol.* **2017**, *12*, 2815–2822.

(41) Sit, C. S.; Yoganathan, S.; Vederas, J. C. *Acc. Chem. Res.* **2011**, *44*, 261–268.

(42) Izawa, M.; Kawasaki, T.; Hayakawa, Y. *Appl. Environ. Microbiol.* **2013**, *79*, 7110–7113.

(43) Burkhart, B. J.; Schwalen, C. J.; Mann, G.; Naismith, J. H.; Mitchell, D. A. *Chem. Rev.* **2017**, *117*, 5389–5456.

(44) Dunbar, K. L.; Melby, J. O.; Mitchell, D. A. *Nat. Chem. Biol.* **2012**, *8*, 569.

(45) Dunbar, K. L.; Chekan, J. R.; Cox, C. L.; Burkhart, B. J.; Nair, S. K.; Mitchell, D. A. *Nat. Chem. Biol.* **2014**, *10*, 823.

(46) Dunbar, K. L.; Tietz, J. I.; Cox, C. L.; Burkhart, B. J.; Mitchell, D. A. *J. Am. Chem. Soc.* **2015**, *137*, 7672–7677.

(47) Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. *Nat. Chem. Biol.* **2015**, *11*, 564.

(48) Schwalen, C. J.; Hudson, G. A.; Kosol, S.; Mahanta, N.; Challis, G. L.; Mitchell, D. A. *J. Am. Chem. Soc.* **2017**, *139*, 18154–18157.

(49) Franz, L.; Adam, S.; Santos-Aberturas, J.; Truman, A. W.; Koehnke, J. *J. Am. Chem. Soc.* **2017**, *139*, 18158–18161.

(50) Kjaerulff, L.; Sikandar, A.; Zaburannyi, N.; Adam, S.; Herrmann, J.; Koehnke, J.; Müller, R. *ACS Chem. Biol.* **2017**, *12*, 2837–2841.

(51) Grabarse, W.; Mahlert, F.; Shima, S.; Thauer, R. K.; Ermler, U. *J. Mol. Biol.* **2000**, *303*, 329–344.

(52) Ermler, U.; Grabarse, W.; Shima, S.; Goubeaud, M.; Thauer, R. K. *Science* **1997**, *278*, 1457 LP-1462.

(53) Scheller, S.; Goenrich, M.; Boecher, R.; Thauer, R. K.; Jaun, B. *Nature* **2010**, *465*,

606.

(54) Thauer, R. K.; Kaster, A.-K.; Seedorf, H.; Buckel, W.; Hedderich, R. *Nat. Rev. Microbiol.* **2008**, *6*, 579.

(55) Moore, S. J.; Sowa, S. T.; Schuchardt, C.; Deery, E.; Lawrence, A. D.; Ramos, J. V.; Billig, S.; Birkemeyer, C.; Chivers, P. T.; Howard, M. J.; Rigby, S. E. J.; Layer, G.; Warren, M. J. *Nature* **2017**, *543*, 78.

(56) Wongnate, T.; Sliwa, D.; Ginovska, B.; Smith, D.; Wolf, M. W.; Lehnert, N.; Raugei, S.; Ragsdale, S. W. *Science* **2016**, *352*, 953 LP-958.

(57) Selmer, T.; Kahnt, J.; Goubeaud, M.; Shima, S.; Grabarse, W.; Ermler, U.; Thauer, R. K. *J. Biol. Chem.* **2000**, *275*, 3755–3760.

(58) Kahnt, J.; Buchenau, B.; Mahlert, F.; Krüger, M.; Shima, S.; Thauer, R. K. *FEBS J.* **2007**, *274*, 4913–4921.

(59) Wagner, T.; Kahnt, J.; Ermler, U.; Shima, S. *Angew. Chemie Int. Ed.* **2016**, *55*, 10630–10633.

(60) Horng, Y.-C.; Becker, D. F.; Ragsdale, S. W. *Biochemistry* **2001**, *40*, 12875–12885.

(61) Grabarse, W.; Mahlert, F.; Duin, E. C.; Goubeaud, M.; Shima, S.; Thauer, R. K.; Lamzin, V.; Ermler, U. *J. Mol. Biol.* **2001**, *309*, 315–330.

(62) Goenrich, M.; Duin, E. C.; Mahlert, F.; Thauer, R. K. *JBIC J. Biol. Inorg. Chem.* **2005**, *10*, 333–342.

(63) Kloss, F.; Lincke, T.; Hertweck, C. *European J. Org. Chem.* **2011**, *2011*, 1429–1431.

(64) Zacharie, B.; Sauvé, G.; Penney, C. *Tetrahedron* **1993**, *49*, 10489–10500.

(65) M. Ashraf Shalaby; Christopher W. Grote, and; Rapoport*, H. **1996**.

(66) Miwa, J. H.; Patel, A. K.; Vivatrat, N.; Popek, S. M.; Meyer, A. M. *Org. Lett.* **2001**, *3*, 3373–3375.

(67) Culik, R. M.; Jo, H.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2012**, *134*, 8026–8029.

(68) Mukherjee, S.; Verma, H.; Chatterjee, J. *Org. Lett.* **2015**, *17*, 3150–3153.

(69) Wissner, R. F.; Wagner, A. M.; Warner, J. B.; Petersson, E. J. *Synlett* **2013**, *24*, 2454–2458.

(70) Mukherjee, S.; Chatterjee, J. *J. Pept. Sci.* **2016**, *22*, 664–672.

(71) Walters, C. R.; Szantai-Kis, D. M.; Zhang, Y.; Reinert, Z. E.; Horne, W. S.; Chenoweth, D. M.; Petersson, E. J. *Chem. Sci.* **2017**, *8*, 2868–2877.

(72) Huang, Y.; Ferrie, J. J.; Chen, X.; Zhang, Y.; Szantai-Kis, D. M.; Chenoweth, D. M.; Petersson, E. J. *Chem. Commun.* **2016**, *52*, 7798–7801.

(73) Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J. *J. Am. Chem. Soc.* **2013**, *135*, 6529–6540.

(74) Batjargal, S.; Huang, Y.; Wang, Y. J.; Petersson, E. J. *J. Pept. Sci.* **2014**, *20*, 87–91.

(75) Chen, X.; Mietlicki-Baase, E. G.; Barrett, T. M.; McGrath, L. E.; Koch-Laskowski, K.; Ferrie, J. J.; Hayes, M. R.; Petersson, E. J. *J. Am. Chem. Soc.* **2017**, *139*, 16688–16695.

(76) Stolowitz, M. L.; Paape, B. A.; Dixit, V. M. *Anal. Biochem.* **1989**, *181*, 113–119.

(77) Szantai-Kis, D.; Walters, C.; Barrett, T.; Hoang, E.; Petersson, E. *Synlett* **2017**, *28*, 1789–1794.

(78) Walters, C. R.; Ferrie, J. J.; Petersson, E. J. *Chem. Commun.* **2018**, *54*, 1766–1769.

(79) Yang, J.; Wang, C.; Xu, S.; Zhao, J. *Angew. Chemie Int. Ed.* **2019**, *58*, 1382–1386.

(80) Miwa, J. H.; Pallivathucal, L.; Gowda, S.; Lee, K. E. *Org. Lett.* **2002**, *4*, 4655–4657.

(81) Sandström, J. *Chirality* **2000**, *12*, 162–171.

(82) Fŏrster, T. *Discuss. Faraday Soc.* **1959**, *27*, 7–17.

(83) Goldberg, J. M.; Batjargal, S.; Petersson, E. J. *J. Am. Chem. Soc.* **2010**, *132*, 14718–14720.

(84) Petersson, E. J.; Goldberg, J. M.; Wissner, R. F. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6827–6837.

(85) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; 3rd Editio.; Springer Science+Business Media: New York, 2006.

(86) Goldberg, J. M.; Batjargal, S.; Chen, B. S.; Petersson, E. J. *J. Am. Chem. Soc.* **2013**, *135*, 18651–18658.

(87) Goldberg, J. M.; Chen, X.; Meinhardt, N.; Greenbaum, D. C.; Petersson, E. J. *J. Am. Chem. Soc.* **2014**, *136*, 2086–2093.

(88) Cekic, N.; Vocadlo, D. Development of chemical tools for studying human O-GlcNAcase activity, Simon Fraser University, 2015.

(89) Perley-Robertson, G. E.; Vocadlo, D. Development of tools and methods for studying glycan processing proteins in living systems, Simon Fraser University, 2016.

(90) Jones, W. C.; Nestor, J. J.; Du Vigneaud, V. *J. Am. Chem. Soc.* **1973**, *95*, 5677–5679.

(91) Okano, A.; Nakayama, A.; Wu, K.; Lindsey, E. A.; Schammel, A. W.; Feng, Y.; Collins, K. C.; Boger, D. L. *J. Am. Chem. Soc.* **2015**, *137*, 3693–3704.

(92) Okano, A.; Isley, N. A.; Boger, D. L. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E5052 LP-E5061.

(93) Thombare, V. J.; Hutton, C. A. *Angew. Chemie Int. Ed.* **2019**, *0*.

(94) Frank, R.; Jakob, M.; Thunecke, F.; Fischer, G.; Schutkowski, M. *Angew. Chemie Int. Ed.* **2000**, *39*, 1120–1122.

(95) Huang, Y.; Cong, Z.; Yang, L.; Dong, S. *J. Pept. Sci. J. Pept. Sci* **2008**, *14*, 1062–1068.

(96) Wildemann, D.; Schiene-Fischer, C.; Aumüller, T.; Bachmann, A.; Kiefhaber, T.; Lücke, C.; Fischer, G. *J. Am. Chem. Soc.* **2007**, *129*, 4910–4918.

(97) Cong, Z.; Huang, Y.; Yang, L.; Dong, D. Y. and S. Solution Structure of a Photo-Switchable Insect Kinin Thioxo-Analog. *Protein & Peptide Letters*, 2010, *17*, 343–346.

(98) Broichhagen, J.; Podewin, T.; Meyer-Berg, H.; von Ohlen, Y.; Johnston, N. R.; Jones, B. J.; Bloom, S. R.; Rutter, G. A.; Hoffmann-Röder, A.; Hodson, D. J.; Trauner, D. *Angew. Chemie Int. Ed.* **2015**, *54*, 15565–15569.

(99) Bond, M. D.; Holmquist, B.; Vallee, B. L. *J. Inorg. Biochem.* **1986**, *28*, 97–105.

(100) Schutkowski, M.; Neubert, K.; Fischer, G. *Eur. J. Biochem.* **1994**, *221*, 455–461.

(101) Komatsu, R.; Matsuyama, T.; Namba, M.; Watanabe, N.; Itoh, H.; Kono, N.; Tarui, S. *Diabetes* **1989**, *38*, 902 LP-905.

(102) Kim, W.; Egan, J. M. *Pharmacol. Rev.* **2008**, *60*, 470 LP-512.

(103) Aertgeerts, K.; Ye, S.; Tennant, M. G.; Kraus, M. L.; Rogers, J.; Sang, B.-C.; Skene, R. J.; Webb, D. R.; Prasad, G. S. *Protein Sci.* **2004**, *13*, 412–421.

(104) Whalen, E. J.; Rajagopal, S.; Lefkowitz, R. J. *Trends Mol. Med.* **2011**, *17*, 126–139.

(105) Zhang, H.; Sturchler, E.; Zhu, J.; Nieto, A.; Cistrone, P. A.; Xie, J.; He, L.; Yea, K.; Jones, T.; Turn, R.; Di Stefano, P. S.; Griffin, P. R.; Dawson, P. E.; McDonald, P. H.; Lerner, R. A. *Nat. Commun.* **2015**, *6*, 8918.

(106) Liang, Y.-L.; Khoshouei, M.; Glukhova, A.; Furness, S. G. B.; Zhao, P.; Clydesdale, L.; Koole, C.; Truong, T. T.; Thal, D. M.; Lei, S.; Radjainia, M.; Danev, R.; Baumeister, W.; Wang, M.-W.; Miller, L. J.; Christopoulos, A.; Sexton, P. M.; Wootten, D. *Nature* **2018**, *555*, 121.

(107) Zacharie, B.; Lagraoui, M.; Dimarco, M.; Penney, C. L.; Gagnon, L. *J. Med. Chem.* **1999**, *42*, 2046–2052.

(108) Zhang, W.; Li, J.; Liu, L.-W.; Wang, K.-R.; Song, J.-J.; Yan, J.-X.; Li, Z.-Y.; Zhang, B.-Z.; Wang, R. *Peptides* **2010**, *31*, 1832–1838.

(109) Verma, H.; Khatri, B.; Chakraborti, S.; Chatterjee, J. *Chem. Sci.* **2018**, *9*, 2443–2451.

(110) Bock, J. E.; Gavenonis, J.; Kritzer, J. A. *ACS Chem. Biol.* **2013**, *8*, 488–499.

(111) Mas-Moruno, C.; Kessler, F. R. and H. Cilengitide: The First Anti-Angiogenic Small Molecule Drug Candidate. Design, Synthesis and Clinical Evaluation. *Anti-Cancer Agents in Medicinal Chemistry*, 2010, *10*, 753–768.

(112) Xiong, J.-P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A. *Science* **2002**, *296*, 151 LP-155.

(113) Dawson, P. E. *Methods Enzymol.* **1997**, *287*, 34–45.

(114) Batjargal, S.; Wang, Y. J.; Goldberg, J. M.; Wissner, R. F.; Petersson, E. J. *J. Am. Chem. Soc.* **2012**, *134*, 9172–9182.

(115) Kawakami, T.; Aimoto, S. *Tetrahedron* **2009**, *65*, 3871–3877.

(116) Botti, P.; Villain, M.; Manganiello, S.; Gaertner, H. *Org. Lett.* **2004**, *6*, 4861–4864.

(117) George, E. A.; Novick, R. P.; Muir, T. W. *J. Am. Chem. Soc.* **2008**, *130*, 4914–4924.

(118) Blanco-Canosa, J. B.; Dawson, P. E. *Angew. Chemie Int. Ed.* **2008**, *47*, 6851–6855.

(119) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. *Nat. Protoc.* **2013**, *8*, 2483–2495.

(120) Wang, Y. J.; Szantai-Kis, D. M.; Petersson, E. J. *Org. Biomol. Chem.* **2016**, *14*, 6262–6269.

(121) Malins, L. R.; Payne, R. J. *Curr. Opin. Chem. Biol.* **2014**, *22*, 70–78.

(122) Bachmann, A.; Wildemann, D.; Praetorius, F.; Fischer, G.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3952 LP-3957.

(123) Newberry, R. W.; VanVeller, B.; Raines, R. T. *Chem. Commun.* **2015**, *51*, 9624–9627.

(124) Newberry, R. W.; VanVeller, B.; Guzei, I. A.; Raines, R. T. *J. Am. Chem. Soc.* **2013**, *135*, 7843–7846.

(125) Engel-Andreasen, J.; Wich, K.; Laursen, J. S.; Harris, P.; Olsen, C. A. *J. Org. Chem.* **2015**, *80*, 5415–5427.

(126) Bordwell, F. G.; Bartmess, J. E.; Hautala, J. A. *J. Org. Chem.* **1978**, *43*, 3095–3101.

(127) Bordwell, F. G.; Fried, H. E. *J. Org. Chem.* **1991**, *56*, 4218–4223.

(128) Stroud, E. D.; Fife, D. J.; Smith, G. G. *J. Org. Chem.* **1983**, *48*, 5368–5369.

(129) Huang, Y.; Jahreis, G.; Lücke, C.; Wildemann, D.; Fischer, G. *J. Am. Chem. Soc.* **2010**, *132*, 7578–7579.

(130) Zinieris, N.; Leondiadis, L.; Ferderigos, N. *J. Comb. Chem.* **2005**, *7*, 4–6.

(131) Wade, J. D.; Bedford, J.; Sheppard, R. C.; Tregear, G. W. *Pept. Res.* **1991**, *4*, 194–199.

(132) Fatkins, D. G.; Zheng, W. *Int. J. Mol. Sci.* **2008**, *9*, 1–11.

(133) Sanderson, J. M.; Singh, P.; Fishwick, C. W. G.; Findlay, J. B. C. *J. Chem. Soc., Perkin Trans. 1* **2000**, 3227–3231.

(134) Goldberg, J. M.; Wissner, R. F.; Klein, A. M.; Petersson, E. J. *Chem. Commun.* **2012**, *48*, 1550–1552.

(135) Goldberg, J. M.; Speight, L. C.; Fegley, M. W.; Petersson, E. J. *J. Am. Chem. Soc.* **2012**, *134*, 6088–6091.

(136) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science* **1994**, *266*, 776–779.

(137) Hackenberger, C. P. R.; Schwarzer, D. *Angew. Chemie Int. Ed.* **2008**, *47*, 10030–10074.

(138) Maini, R.; Dedkova, L. M.; Paul, R.; Madathil, M. M.; Chowdhury, S. R.; Chen, S.; Hecht, S. M. *J. Am. Chem. Soc.* **2015**, *137*, 11206–11209.

(139) Echols, N.; Harrison, P.; Balasubramanian, S.; Luscombe, N. M.; Bertone, P.; Zhang, Z.; Gerstein, M. *Nucleic Acids Res.* **2002**, *30*, 2515–2523.

(140) Dawson, P. E. *Isr. J. Chem.* **2011**, *51*, 862–867.

(141) Bochar, D. A.; Tabernero, L.; Stauffacher, C. V; Rodwell, V. W. *Biochemistry* **1999**, *38*, 8879–8883.

(142) Heinrikson, R. L. *J. Biol. Chem.* **1971**, *246*, 4090–4096.

(143) Kochendoerfer, G. G.; Chen, S.-Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B. H.; Bradburne, J. A. *Science* **2003**, *299*, 884 LP-887.

(144) Smith, H. B.; Hartman, F. C. *J. Biol. Chem.* **1988**, *263*, 4921–4925.

(145) Wan, Q.; Danishefsky, S. J. *Angew. Chemie - Int. Ed.* **2007**, *46*, 9248–9252.

(146) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

(147) Crich, D.; Banerjee, A. *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.

(148) Malins, L. R.; Payne, R. J. *Org. Lett.* **2012**, *14*, 3142–3145.

(149) Malins, L. R.; Payne, R. J. *Aust. J. Chem.* **2015**, *68*, 521–537.

(150) Wong, C. T. T.; Tung, C. L.; Li, X. *Mol. Biosyst.* **2013**, *9*, 826–833.

(151) Metanis, N.; Keinan, E.; Dawson, P. E. *Angew. Chemie Int. Ed.* **2010**, *49*, 7049–7053.

(152) Townsend, S. D.; Tan, Z.; Dong, S.; Shang, S.; Brailsford, J. A.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2012**, *134*, 3912–3916.

(153) Trost, B. M.; Flaming, I. *Comprehensive Organic Synthesis - Selectivity, Strategy and Efficiency in Modern Organic Chemistry*; Elsevier: Oxford, UK, 1991.

(154) Kornfeld, E. C. *J. Org. Chem.* **1951**, *16*, 131–138.

(155) Hendrickson, T. L.; Imperiali, B. *Biochemistry* **1995**, *34*, 9444–9450.

(156) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.

(157) Thompson, R. E.; Liu, X.; Alonso-García, N.; Pereira, P. J. B.; Jolliffe, K. A.; Payne, R. J. *J. Am. Chem. Soc.* **2014**, *136*, 8161–8164.

(158) Cox, B. G. *Acids and Bases: Solvent Effects on Acid-Base Strength*; Oxford University Press: Oxford, 2013.

(159) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. *Angew. Chemie Int. Ed.* **2008**, *47*, 8521–8524.

(160) Gronenborn, A.; Filpula, D.; Essig, N.; Achari, A.; Whitlow, M.; Wingfield, P.; Clore, G. *Science* **1991**, *253*, 657–661.

(161) Cao, Y.; Li, H. *Nat. Nanotechnol.* **2008**, *3*, 512.

(162) Malakauskas, S. M.; Mayo, S. L. *Nat. Struct. Biol.* **1998**, *5*, 470–475.

(163) Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D. *Protein Sci.* **2002**, *11*, 2924–2931.

(164) Odaert, B.; Jean, F.; Boutillon, C.; Buisine, E.; Melnyk, O.; Tartar, A.; Lippens, G. *Protein Sci.* **1999**, *8*, 2773–2783.

(165) Reinert, Z. E.; Horne, W. S. *Chem. Sci.* **2014**, *5*, 3325–3330.

(166) Reinert, Z. E.; Musselman, E. D.; Elcock, A. H.; Horne, W. S. *ChemBioChem* **2012**, *13*, 1107–1111.

(167) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angew. Chemie Int. Ed.* **2011**, *50*, 7645–7649.

(168) Spicer, C. D.; Davis, B. G. *Nat. Commun.* **2014**, *5*, 4740.

(169) Kreitler, D. F.; Mortenson, D. E.; Forest, K. T.; Gellman, S. H. *J. Am. Chem. Soc.* **2016**, *138*, 6498–6505.

(170) Matulef, K.; Annen, A. W.; Nix, J. C.; Valiyaveetil, F. I. *Structure* **2016**, *24*, 750–761.

(171) Gopalan, R. D.; Del Borgo, M. P.; Mechler, A. I.; Perlmutter, P.; Aguilar, M.-I. *Chem. Biol.* **2015**, *22*, 1417–1423.

(172) Gopalakrishnan, R.; Frolov, A. I.; Knerr, L.; Drury, W. J.; Valeur, E. *J. Med. Chem.* **2016**, *59*, 9599–9621.

(173) Judge, R. H.; Moule, D. C.; Goddard, J. D. *Can. J. Chem.* **1987**, *65*, 2100–2105.

(174) Lee, H.-J.; Choi, Y.-S.; Lee, K.-B.; Park, J.; Yoon, C.-J. *J. Phys. Chem. A* **2002**, *106*, 7010–7017.

(175) Bregy, H.; Heimgartner, H.; Helbing, J. *J. Phys. Chem. B* **2009**, *113*, 1756–1762.

(176) Hamm, P.; Helbing, J.; Bredenbeck, J. *Annu. Rev. Phys. Chem.* **2008**, *59*, 291–317.

(177) Goodman, E. M.; Kim, P. S. *Biochemistry* **1991**, *30*, 11615–11620.

(178) Chin, D.; Means, A. R. *Trends Cell Biol.* **2000**, *10*, 322–328.

(179) Kursula, P. *Acta Crystallogr. Sect. D* **2014**, *70*, 24–30.

(180) Vetter, S. W.; Leclerc, E. *Eur. J. Biochem.* **2003**, *270*, 404–414.

(181) Hellstrand, E.; Kukora, S.; Shuman, C. F.; Steenbergen, S.; Thulin, E.; Kohli, A.; Krouse, B.; Linse, S.; Åkerfeldt, K. S. *FEBS J.* **2013**, *280*, 2675–2687.

(182) Schumacher, M. A.; Crum, M.; Miller, M. C. *Structure* **2004**, *12*, 849–860.

(183) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **1995**, *2*, 768–776.

(184) Gibrat, G.; Assairi, L.; Craescu, C. T.; Hui Bon Hoa, G.; Loew, D.; Lombard, B.; Blouquit, L.; Bellissent-Funel, M.-C. *Biochim. Biophys. Acta - Proteins Proteomics* **2012**, *1824*, 1097–1106.

(185) Masino, L.; Martin, S. R.; Bayley, P. M. *Protein Sci.* **2000**, *9*, 1519–1529.

(186) Protasevich, I.; Ranjbar, B.; Lobachov, V.; Makarov, A.; Gilli, R.; Briand, C.; Lafitte, D.; Haiech, J. *Biochemistry* **1997**, *36*, 2017–2024.

(187) Sorensen, B. R.; Shea, M. A. *Biochemistry* **1998**, *37*, 4244–4253.

(188) Kmiecik, S.; Kolinski, A. *Biophys. J.* **2008**, *94*, 726–736.

(189) Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M. *J. Phys. Chem. B* **2007**, *111*, 14362–14369.

(190) Reinert, Z. E.; Lengyel, G. A.; Horne, W. S. *J. Am. Chem. Soc.* **2013**, *135*, 12528–12531.

(191) Lengyel, G. A.; Reinert, Z. E.; Griffith, B. D.; Horne, W. S. *Org. Biomol. Chem.* **2014**, *12*, 5375–5381.

(192) Tavenor, N. A.; Reinert, Z. E.; Lengyel, G. A.; Griffith, B. D.; Horne, W. S. *Chem. Commun.* **2016**, *52*, 3789–3792.

(193) Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859–1877.

(194) Chattopadhyay, S.; Raines, R. T. *Biopolymers* **2014**, *101*, 821–833.

(195) Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B. *J. Struct. Biol.* **1998**, *122*, 86–91.

(196) Choudhary, A.; Kamer, K. J.; Shoulders, M. D.; Raines, R. T. *Pept. Sci.* **2015**, *104*, 110–115.

(197) Erdmann, R. S.; Wennemers, H. *J. Am. Chem. Soc.* **2010**, *132*, 13957–13959.

(198) Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 9262–9263.

(199) Kotch, F. W.; Guzei, I. A.; Raines, R. T. *J. Am. Chem. Soc.* **2008**, *130*, 2952–2953.

(200) Shoulders, M. D.; Guzei, I. A.; Raines, R. T. *Biopolymers* **2008**, *89*, 443–454.

(201) Shoulders, M. D.; Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2006**, *128*, 8112–8113.

(202) Dai, N.; Wang, X. J.; Etzkorn, F. A. *J. Am. Chem. Soc.* **2008**, *130*, 5396–5397.

(203) Jenkins, C. L.; Vasbinder, M. M.; Miller, S. J.; Raines, R. T. *Org. Lett.* **2005**, *7*, 2619–2622.

(204) Zhang, Y.; Malamakal, R. M.; Chenoweth, D. M. *J. Am. Chem. Soc.* **2015**, *137*, 12422–12425.

(205) Hongo, C.; Noguchi, K.; Okuyama, K.; Tanaka, Y.; Nishino, N. *J. Biochem.* **2005**, *138*, 135–144.

(206) Bella, J.; Brodsky, B.; Berman, H. M. *Structure* **1995**, *3*, 893–906.

(207) Bella, J. *Biochem. J.* **2016**, *473*, 1001 LP-1025.

(208) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 777–778.

(209) Erdmann, R. S.; Wennemers, H. *Angew. Chemie Int. Ed.* **2011**, *50*, 6835–6838.

(210) Shoulders, M. D.; Satyshur, K. A.; Forest, K. T.; Raines, R. T. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 559 LP-564.

(211) Albertsen, L.; Shaw, A. C.; Norrild, J. C.; Strømgaard, K. *Bioconjug. Chem.* **2013**, *24*, 1883–1894.

(212) Shah, N. H.; Muir, T. W. *Chem. Sci.* **2014**, *5*, 446–461.

(213) Batjargal, S.; Walters, C. R.; Petersson, E. J. *J. Am. Chem. Soc.* **2015**, *137*, 1734–1737.

(214) Haney, C. M.; Cleveland, C. L.; Wissner, R. F.; Owei, L.; Robustelli, J.; Daniels, M. J.; Canyurt, M.; Rodriguez, P.; Ischiropoulos, H.; Baumgart, T.; Petersson, E. J. *Biochemistry* **2017**, *56*, 683–691.

(215) Gentle, I. E.; De Souza, D. P.; Baca, M. *Bioconjug. Chem.* **2004**, *15*, 658–663.

(216) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10068 LP-10073.

(217) Hopkins, C. E.; O'connor, P. B.; Allen, K. N.; Costello, C. E.; Tolan, D. R. *Protein Sci.* **2002**, *11*, 1591–1599.

(218) Ellilä, S.; Jurvansuu, J. M.; Iwaï, H. *FEBS Lett.* **2011**, *585*, 3471–3477.

(219) Tropea, J. E.; Cherry, S.; Waugh, D. S. Doyle, S. A., Ed.; Humana Press: Totowa, NJ, 2009; pp. 297–307.

(220) Tolbert, T. J.; Wong, C.-H. *Angew. Chemie* **2002**, *114*, 2275–2278.

(221) Mossessova, E.; Lima, C. D. *Mol. Cell* **2000**, *5*, 865–876.

(222) Guerrero, F.; Ciragan, A.; Iwaï, H. *Protein Expr. Purif.* **2015**, *116*, 42–49.

(223) Lee, C.-D.; Sun, H.-C.; Hu, S.-M.; Chiu, C.-F.; Homhuan, A.; Liang, S.-M.; Leng, C.-H.; Wang, T.-F. *Protein Sci.* **2008**, *17*, 1241–1248.

(224) Haney, C. M.; Wissner, R. F.; Warner, J. B.; Wang, Y. J.; Ferrie, J. J.; J. Covell, D.; Karpowicz, R. J.; Lee, V. M.-Y.; James Petersson, E. *Org. Biomol. Chem.* **2016**, *14*, 1584–1592.

(225) Xiong, H.; Reynolds, N. M.; Fan, C.; Englert, M.; Hoyer, D.; Miller, S. J.; Söll, D. *Angew. Chemie Int. Ed.* **2016**, *55*, 4083–4086.

(226) Goto, Y.; Katoh, T.; Suga, H. *Nat. Protoc.* **2011**, *6*, 779–790.

(227) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.

(228) Davis, L.; Chin, J. W. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 168.

(229) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **2015**, *6*, 50–69.

(230) Ferrie, J. J.; Haney, C. M.; Yoon, J.; Pan, B.; Lin, Y.-C.; Fakhraai, Z.; Rhoades, E.; Nath, A.; Petersson, E. J. *Biophys. J.* **2018**, *114*, 53–64.

(231) Speight, L. C.; Muthusamy, a K.; Goldberg, J. M.; Warner, J. B.; Wissner, R. F.; Willi, T. S.; Woodman, B. F.; Mehl, R. a; Petersson, E. J. *J. Am. Chem. Soc.* **2013**, *135*, 18806–18814.

(232) Jun, J. V; Haney, C. M.; Karpowicz, R. J.; Giannakoulias, S.; Lee, V. M.-Y.; Petersson, E. J.; Chenoweth, D. M. *J. Am. Chem. Soc.* **2019**, *141*, 1893–1897.

(233) Wang, L. *Acc. Chem. Res.* **2017**, *50*, 2767–2775.

(234) Wan, W.; Tharp, J. M.; Liu, W. R. *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844*, 1059–1070.

(235) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Söll, D.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11268 LP-11273.

(236) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232–234.

(237) Bryson, D. I.; Fan, C.; Guo, L.-T.; Miller, C.; Söll, D.; Liu, D. R. *Nat. Chem. Biol.* **2017**, *13*, 1253.

(238) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472*, 499.

(239) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. *Nat. Protoc.* **2007**, *2*, 2601.

(240) Ohuchi, M.; Murakami, H.; Suga, H. *Curr. Opin. Chem. Biol.* **2007**, *11*, 537–542.

(241) Po, P.; Delaney, E.; Gamper, H.; Szantai-Kis, D. M.; Speight, L.; Tu, L. W.; Kosolapov, A.; Petersson, E. J.; Hou, Y. M.; Deutsch, C. *J. Mol. Biol.* **2017**, *429*, 1873–1888.

(242) Lu, J.; Hua, Z.; Kobertz, W. R.; Deutsch, C. *J. Mol. Biol.* **2011**, *411*, 499–510.