STUDIES TOWARD UNDERSTANDING THE BIOSYNTHESIS OF SACTIPEPTIDES AND THE CREATION OF PEPTIDE NATURAL PRODUCT LIBRARIES THROUGH MRNA DISPLAY

Paul Michael Himes

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Pharmaceutical Sciences in the Doctoral Program of the UNC Eshelman School of Pharmacy (Division of Chemical Biology and Medicinal Chemistry).

Chapel Hill
2017

Approved by:

Albert Bowers

Stephen Frye

Rihe Liu

Brian Kuhlman

Kevin Weeks

ABSTRACT

Paul Michael Himes: Studies toward Understanding the Biosynthesis of Sactipeptides and the Creation of Peptide Natural Product Libraries through mRNA Display
(Under the direction of Albert A. Bowers)

Ribosomally-synthesized and post-translationally modified peptides (RiPPs) are a class of natural products that are an attractive starting point for new antibiotics due to their wide range of structural diversity and biological activities. The post-translational modifications imparted upon the peptide substrate are carried out by promiscuous RiPP enzymes. Sactipeptides are members of the RiPPs family that are made through radical-mediated cysteine sulfur to α-carbon coupling reactions. The resulting thioether linkages give rise to sactipeptides defined structures and concomitant biological activities. The research presented here focuses on the biochemical and structural characterization of CteB, a radical SAM enzyme that imparts a single sactionine bridge, the development of an *E. coli* heterologous expression system for sactipeptides and the combination of RiPPs and mRNA display for the production of modified peptide libraries.

We have biochemically and structurally characterized CteB, a radical SAM enzyme that imparts a sactionine bridge on its corresponding peptide substrate. A crystal structure was obtained at 2.04 Å and showed a RiPPs recognition element connected to a $(\beta/\alpha)_6$-TIM barrel fold, followed by an SPASM domain that houses two auxiliary [4Fe-4S] clusters, one of which contains a free coordination site for potential peptide ligation.

We have developed an *E. coli* heterologous expression system for the production of sactipeptides based on subtilosin A from *Bacillus subtilis 168*. In the system, both the peptide

substrate and radical SAM enzyme (AlbA) are expressed together and the modified sactipeptide is produced and isolated. This system was used to probe the substrate promiscuity of AlbA, and determine what changes it can tolerate. Additionally, an unnatural amino acid, *O*-Me-tyrosine, was able to be incorporated into the peptide substrate while also forming a thioether bridge at that position.

We have also worked on combining the natural promiscuity of RiPPs enzymes with mRNA display to generate modified peptide libraries on a large scale (~5 x10^6). Using two previously described and characterized RiPPs systems, pantocin A and thiomuracin, we have used their respective RiPPs enzymes, PaaA and TbtF, to create RiPP peptide libraries to find elements important for binding and to further characterize the promiscuity of these modifying enzymes.

To my parents, my brother, my wife and my son.

ACKNOWLEDGMENTS

Graduate school has been a long and arduous task, and I would not have been able to complete it without the many people who helped me along the way. First, I would like to thank my wife, Avery, who has stood by my side through the good and bad times and tolerated the late days and nights in the lab, all I do is for ours and our son's future. Second, I would like to thank my parents and my brother who were always supportive, encouraging, and believed in me even when my own belief waned. I love you all.

I would like to thank my mentor and advisor, Dr. Albert Bowers, for his tutelage and guidance throughout my thesis work and graduate school studies. You have taught me the many things that go into being a successful and productive scientist, including diligence and the passion for the work we do. I am very honored to have learned as much as I could have from you and I hope to take the lessons you have conveyed to me onward into the next chapter of my life.

I would also like to acknowledge the people in my thesis committee, whose ideas and feedback were invaluable in my completion of my dissertation and thesis work. Thank you, Dr. Stephen Frye, for agreeing to be my committee chair and the insightful discussions on my work and my future scientific career. Thank you, Dr. Rihe Liu, for allowing me to learn from you and your lab the inner workings of mRNA display. Thank you, Dr. Brian Kuhlman, for your insights into protein and peptide modeling as well as your discussions on the CteB enzyme and model.

Lastly, thank you, Dr. Kevin Weeks, for bringing insightful discussion to each and every committee meeting and helping me grow as a critical thinker in those moments.

Thanks to all who have been members of the Bowers Laboratory while I was there: Nicoleta Economou, Scott Allen, Walter Wever, Rachel Bleich, Jonathan Bogart, Swapnil Ghodge, Sungwon Hwang, Kelly Bird, and Steven Fleming. Each one of you helped me become the scientist that I am today and I hope that I will leave having a positive impact on your lives and your work as scientists. Thank you to Dr. Bo Li and Dr. Eric Brustad, and their respective laboratories whose help with the LC-MS was always appreciated. Thank you to Dr. Rita Tamayo and her laboratory for their help with anaerobic procedures and continued use of the anaerobic chamber. A special thanks to Dr. Steve Almo and Dr. Tyler Grove at Albert Einstein School of Medicine, whose crystallography expertise was invaluable on the CteB paper and whose helpful discussions made working with them a true pleasure.

Thank you all for being a part of my life and everything you have done to help me accomplish this goal. I am eternally grateful.

PREFACE

Parts of this dissertation work were done in collaboration with other, talented scientists. Chapter 2 represents a submitted journal article for which I was one or two co-first authors. My contributions to the work focused on the cloning, expression, and reconstitution of activity of the sactionine synthase, CteB. I also characterized all activity of the enzyme by mass spectrometry while performing bioinformatic analysis of the enzyme and related proteins. These results are shown in Tables 2.1-2.2, Figures 2.2, and 2.5-2.6 and Appendix Figures A.1-A.3 and A.11-A.16. Dr. Tyler Grove performed all the crystallography on both the apo and peptide bound forms of CteB, structural comparisons to other known enzymes in the same class, as well as size exclusion chromatography. These experiments are shown in Figures 2.3, 2.4, 2.8 and Appendix Table A.1, Figures A.4-A.10. With the help of Dr. Bowers and Dr. Almo, the co-first authors designed experiments, communicated and divided the work, and then wrote the paper. This work has been submitted as a full article to JACS and has been through two rounds of review and resubmission:

> Grove, T.L., Himes, P.M., Hwang, S., Yumerefendi, H., Bonanno, J.B., Kuhlman, B., Almo, S.C., Bowers, A.A. Structural Insights into Thioether Bond Formation in the Biosynthesis of Sactipeptides. *JACS,* **2017**, *resubmission*

Dr. Tyler Grove and Dr. Steve Almo, co-first author on the paper and his PI respectively, have given permission for me to include this work in my dissertation.

Chapter 3 represents work done solely in the Bowers lab. I am the lead author on the paper but other members of the Bowers lab helped me to complete it. Dr. Scott Allen provided

helpful discussion, experimental design and mass spectrometry analysis. Sungwon Hwang helped me clone the library of peptides into the system I had developed for the production of sactipeptides in *E. coli*. The paper was published previous to the writing of this thesis with the following citation:

> Himes, P.M., Allen, S.E., Hwang, S., Bowers, A.A. Production of Sactipeptides in *Escherichia coli*: Probing the Substrate Promiscuity of Subtilosin A Biosynthesis. *ACS Chem Biol.* **2016**, *11*, 1737-1744

Permission to include the article in its entirety in this dissertation was retained from ACS Publications. Copyright (2016) American Chemical Society

Chapter 4 represents unpublished research that was designed and preformed primarily by myself with help from Steven Fleming also of the Bowers lab.

All copyrighted material included in this dissertation is used with permission from the relevant copyright holders.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5'-dA•	5'-deoxyadensoyl radical
Å	angstrom (1Å = $1x10^{-10}$ meters)
amu	atomic mass unit
anSME	anaerobic sulfatase maturating enzyme
ATP	adenosine 5'-triphosphate
Aux	auxiliary cluster
bp(s)	base pair(s)
BME	2-mercaptoethanol
Cα	alpha-carbon atoms
cte	thermocellin
DIPEA	N,N-diisopropylethyl amine
DCM	dichloromethane
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DT	dithionite
DTT	dithiothreitol
EIC	extracted ion chromatogram
ESI	electrospray ionization
FA	formic acid
FP	fluorescence polarization assay
GSH	reduced glutathione
GSSG	oxidized glutathione

HATU	2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HCl	hydrochloric acid
HEPES	N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_d	dissociation constant
kDa	kilodalton
LC-MS	liquid chromatography–mass spectrometry
LIC	ligation independent cloning
MCS1	multiple cloning site 1
MCS2	multiple cloning site 2
MBP	maltose binding protein
min	minute(s)
MS/MS	yandem mass spectrometry
NEM	N-ethylmalemide
NGS	next-generation sequencing
O-Me-Tyr	*O*-methyl tyrosine
ppm	parts per million
P-linker	DNA-Puromycin linker
RBS	ribosomal binding site
RiPPs	ribosomally synthesized and post-translationally modified peptides
RMSD	root mean square deviation
RRE	RiPPs recognition element
RS	radical SAM

rSAMs	radical SAM enzymes
sactionine	cysteine-sulfur to alpha-carbon thioether
sactipeptides	sulfur-to-alpha carbon thioether cross-linked peptides
SAM	*S*-adenosyl-L-methionine
SCIFF	six cysteines in forty-five residues
SEC	size exclusion chromatography
Skf	sporulation killing factor
SPASM	subtilosin A/PQQ/anaerobic sulfates maturing
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
SSN	sequence similarity network
TAMRA	5-(and 6)-carboxytetramethylrhodamine
TCEP	tris(2-carboxyethyl)phosphine
TIM	triose phosphate isomerase
TEV	tobacco etch virus
TFA	trifluoroacetic acid
Tris	tris(hydroxymethyl)aminomethane
UAA(s)	unnatural amino acid(s)
UV	ultra-violet
wHTH	winged helix-turn-helix
WT	wild-type
Z	charge state (mass spectrometry)

CHAPTER 1

INTRODUCTION

1.1 The Promiscuity of RiPPs Biosynthesis and the Potential for Natural Product Libraries

Natural products, for more than a century, have advanced the understanding of biology and have been at the forefront of the development of novel medicines for the world's most pressing diseases.[1] From the discovery of penicillin and use of Chinese herbal medicine, to the use of more complicated molecules for the treatment of cancer, diabetes, parasitic and bacterial infections, natural products have been used as medication or the first step in the development of life-saving medicines.[2-6] These natural products house privileged scaffolds that convey their unique activities but are very difficult to produce synthetically due to their large size and very specific combinations of stereochemistry found within the molecule.[7,8] The natural products, however, come with their own drawbacks. To harness the power of these natural products, the producer (bacteria, fungi, etc.) needs to be cultivated in specific conditions to facilitate the production of the natural product and in sufficient enough quantity. This can be extremely difficult due to laboratory conditions not resembling the native environments of microbial communities and their constant fight for resources where the production of the natural product would give an advantage.[9] Thus a method to produce enough fully formed, biologically active natural products is required to meet this pressing need.

Ribosomally-synthesized and post-translationally modified peptides (RiPPs) are a class of natural products defined by their unique biosynthetic pathways as well as their modifying enzymes **(Figure 1.1)**.[1] Unlike polyketide synthases (PKS) or non-ribosomal peptide synthases (NRPS), both of which use separate active module-like assembly to create their natural products, RiPPs use the ribosome to create the precursor peptide.[10,11] This precursor peptide can house up to three domains, termed the leader, core, and follower peptides respectively. The leader peptide is used by the RiPPs modifying enzymes to recognize the peptide substrate, while the core peptide is where the modifications are imparted by the enzyme. The follower, if the peptide has one, can play the same role as the leader peptide in terms of recognition. After the modifications are imparted, the leader and/or follower sequence is removed by a peptidase and the modified core is released to give the biologically active product.[1,12] RiPPs modifying enzymes impart extensive post-translational/co-translational modifications that give these peptides structures that are not directly accessible by natural ribosomal synthesis or by the modular synthesis related to PKS or NRPS. These modifications, which are typically conformationally constraining, allow a) better target recognition and higher binding affinity, b) metabolic and chemical stability, and c) a change in chemical functionality by altering the side chains of the canonical amino acids.[1] Due to their structural diversity, wide range of biological activities, and conformational constraining structures, RiPPs are an attractive starting point for novel therapeutics for anti-cancer and antibiotic therapies.

A member of the RiPPs family of natural products is a class of diverse modified peptides known as sactipeptides. Sactipeptides are characterized by their unique thioether bridges that form intramolecular bridges between the sulfurs of cysteine residues and the unreactive α-carbon of a bridging residue amino acid. This forms a new quaternary carbon because of the bridging

partner's amino acid side chain. Therefore, sactipeptides tend to have highly defined regions of secondary structure due to the distribution and number of thioether or sactionine bridges as well as the stereochemistry at the α,α-disubstituted bridging partner residues.[13-20] Subtilosin A, a founding member of sactipeptides, adopts a 3_{10} helix within its structure while other sactipeptide can adopt α-helical structures. The amphipathic helicity of these regions within sactipeptides is thought to grant subtilosin A and others narrow spectrum activity through the ability to interact and disrupt bacterial cell walls resulting in cell death through membrane disruption [16,21,22] This activity, as well as their added stability to heat and proteases due to their thioether bridges, make sactipeptide an attractive biological scaffold for the development of novel therapeutics and chemical probes. Previously in our lab, a system was developed to predict and estimate these highly defined secondary structures using analyses generated from replica exchange molecular dynamics (REMD) trajectories using AMBER 14 and various constraints and implicit solvent conditions.[23-25] In REMD, multiple molecular dynamic simulations are run simultaneously at varying temperatures, and these temperatures are exchanged between replicas at set intervals over the course of the entire simulation. This exchange of temperatures can allow the simulation to overcome energy wells and barriers that cannot be overcome at lower temperatures. After the simulation, the likelihood of each residue adopting a particular secondary structure over the course of the simulation will be identified by hydrogen bonding patterns and angles of that residue. We did this simulation for subtilosin A and found it agreed well with the NMR structure reported by Vederas and co-workers in 2004 **(Figure 1.2)**.[19] Using these simulations, we found that continuous stereochemistry (all D or all L) is required to propagate helicity (either α or 3_{10}) through subtilosin A, an important finding due to subtilosin A switching stereochemistry in the thioether bridges (L, D, D). We therefore hypothesize that this simulation tool for sactipeptides

can be used to help design and predict highly defined secondary structures that could be used as biologically active scaffolds for the grafting of known epitopes for desired function and biologically activity (i.e. inhibition, binding, cell-death, etc.) if it could be paired with a robust expression system.

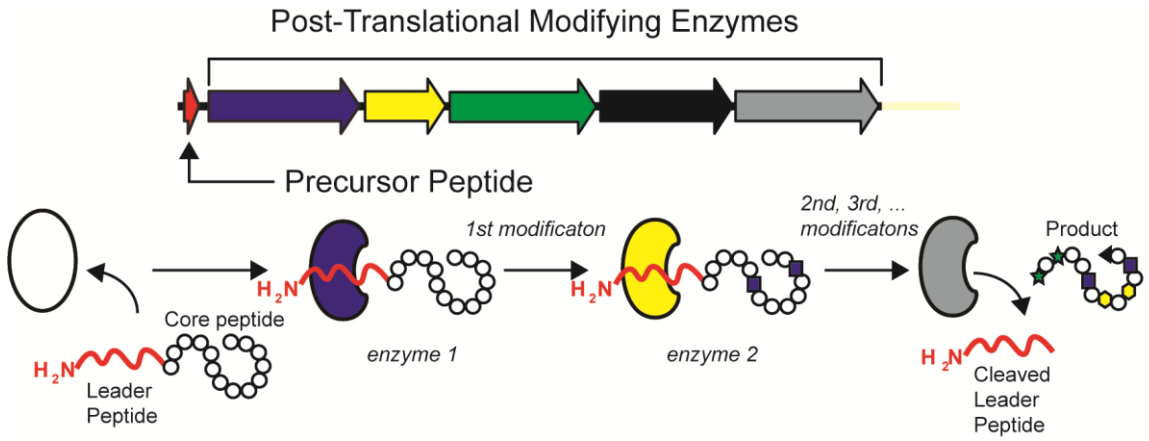


Figure 1.1. Overview of ribosomally synthesized and post-translationally modified peptides (RiPPs).

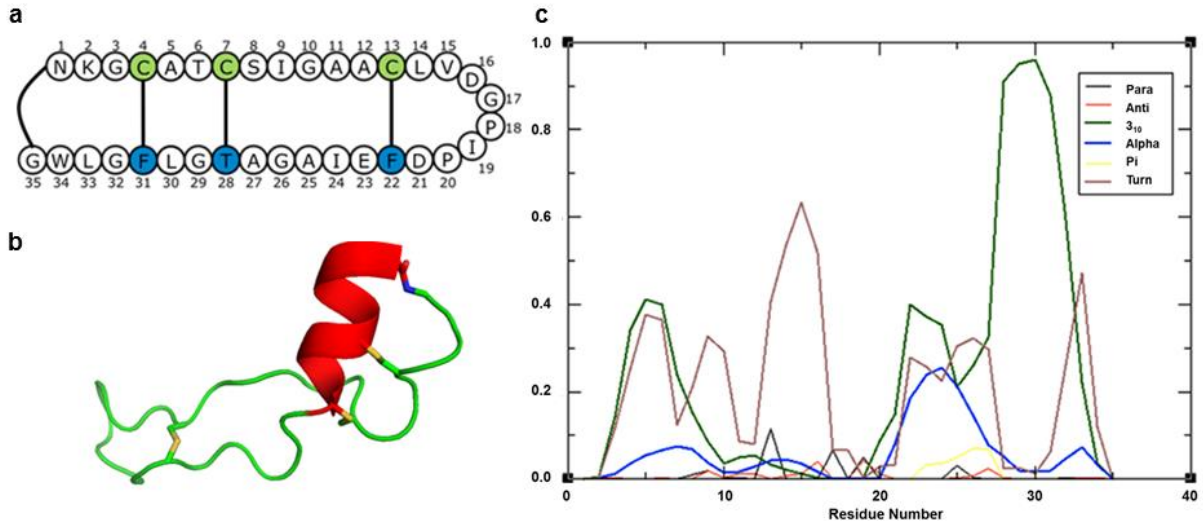


Figure 1.2. Sactipeptide secondary structure. a) Pictorial representation of fully modified and cyclized Subtilosin A. b)NMR structure showing the helix at the C-terminus.[19] c) Secondary structure of each residue of Subtilosin A over the entire simulation. Para = parallel β-sheet; Anti = anti-parallel β-sheet; 3-10= 310 helix; alpha = α-helix; pi = π-helix; turn = some other hydrogen bonding pattern.

RiPPs machinery, including the radical SAM enzymes in sactipeptide biosynthesis, that impart the modifications necessary for biological activity have only recently been isolated and studied in a way that sheds light on how the RiPPs modifying enzyme recognizes and imparts the aforementioned modifications. Through crystallography of these enzymes as well as mutational analysis (*in vitro* and *in vivo*), it has been determined there is a specialized recognition domain termed the RiPPs recognition element (RRE) that allows the RiPPs enzymes to recognize and coordinate to their intended peptide substrate.[26] This RRE recognizes sequences at either the leader or follower sequence of the precursor peptide. This gives RiPPs one of their most impressive abilities, their promiscuity within their own biosynthesis. It has been shown that as long as the recognition element within the leader peptide is intact, the core peptide can be mutated and the RiPPs modifying enzymes can still impart their modification on this "new" core peptide.[27-32] While not every change is allowed, this system houses much more flexibility in the identity of its substrates than most enzymes could tolerate. This gives RiPPs the advantage of creating a wide range of distinct, yet similar peptides that can be tested and altered for specific activities or properties.

With the promiscuity of RiPPs biosynthesis, comes a method to test these libraries of compounds for the desired efficacy and activity toward a therapy. mRNA display happens to be uniquely compatible with RiPPs due to a) RiPPs being encoded genetically in the genome, b) RiPPs being translated by the ribosome, c) mRNA display can tolerate chemical post-translational modifications, and d) mRNA display houses the capability to be performed in vitro and test libraries on the order of $10^{12\text{-}13}$ unique members[33-41], giving a comprehensive study of the promiscuity of a certain RiPPs enzyme related to its substrate. Utilizing the power of mRNA display, RiPPs biosynthetic promiscuity can be tested and unique RiPPs can be made with

differing activities **(Figure 1.3)** than its natural counterpart giving promising new leads for potential therapeutics against cancer and bacterial infections.

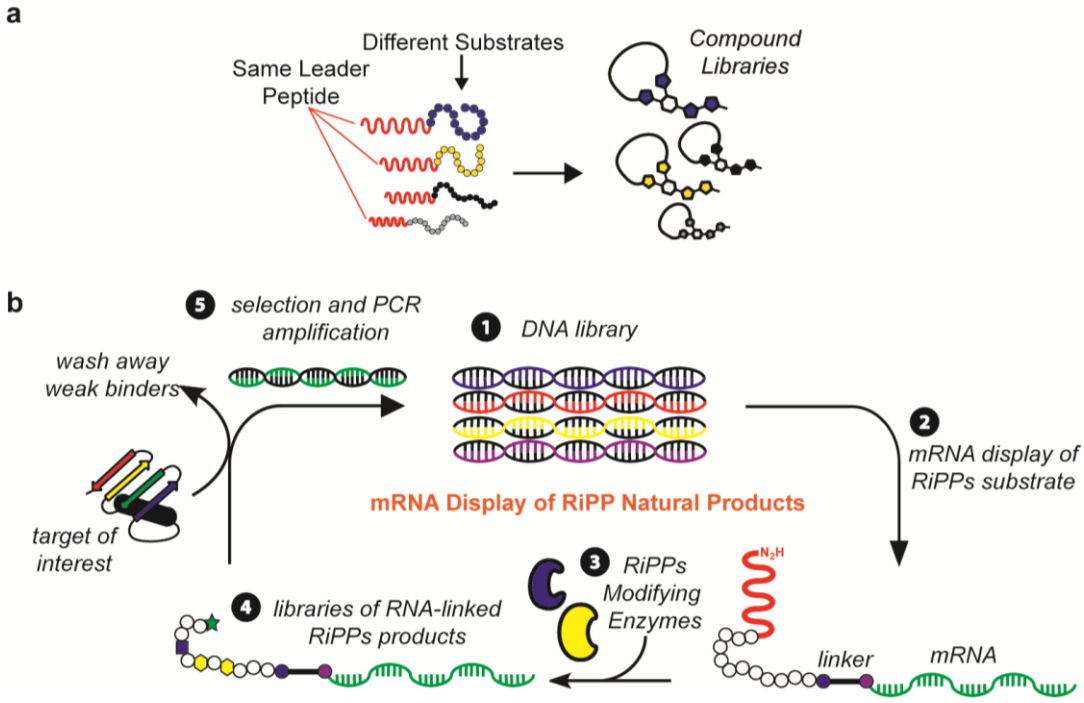


Figure 1.3. Overview of natural product peptide libraries. a) By using the same leader peptide and the innate promiscuity of RiPPs biosynthesis, a natural compound library can be created. b) Proposed workflow for the modification and selection of RiPPs natural products by mRNA display.

New antibiotics are desperately needed due to the rise of antibiotic resistance and the severe lack of new antibiotics. A recent study by the PEW Charitable Trust reported that there has not been a new class of antibiotics registered since 1984.[42] A potential work-around would be to use RiPPs as a starting point and using the power of mRNA display, test on the order of trillions molecules for activity against bacterial species. This can give rise to novel therapeutics in a high-throughput manner.

In the presented work, the promiscuity of RiPPs biosynthesis has been probed and its potential for natural product peptide libraries through mRNA display has been tested. We have studied the biosynthesis of sactipeptides and probed the promiscuity in specialized heterologous

expression system. More specifically, we have biochemical and structural characterized CteB, a radical SAM enzyme that imparts a single sactionine bridge in its corresponding sactipeptide thermocellin. We also describe the development of an *E. coli* heterologous expression system for sactipeptides based on subtilosin A. Using this system, we probe the promiscuity of sactipeptide biosynthesis related to subtilosin A and its sactionine synthase AlbA. Lastly, we will report the progress that has been made with combining RiPPs and mRNA display for the production of modified peptide libraries. We used the pantocin A and thiomuracin biosynthetic pathways that have been previously characterized in our mRNA display studies as cases to study both a) binding affinity and b) RiPP biosynthetic enzyme modification.[43-46]

1.2 REFERENCES

1. 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.; Muller, 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.* **2012**, *30*, 108

2. Sengupta, S.; Chattopadhyay, M.; Grossart, H. *Front. Microbiol.* **2013**, *4*, 1

3. Fleming, A. Br. *J. Exp. Pathol.* **1929**, *10*, 226

4. Brown D.; Lister, T.; May-Dracka, T. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 413

5. Silver, L. *Future Microbiol.* **2015**, *10*, 1711

6. Newman, D.J.; Cragg, G.M. *J. Nat. Prod.* **2016**, *79*, 629

7. Maier, M.E. Org. Biomol. Chem. 2015, 13, 5302

8. Nicolaou, K.C.; Sorensen, E.J.; Winssinger, N. J. Chem. Educ. 1998, 75, 1225

9. Forsberg, K.; Patel, S.; Gibson, M.; Lauber, C.; Knight, R.; Fierer, N.; Dantas, G. Nature. 2014, 509, 612

10. Dutta, S.; Whicher, J.; Hansen, D.; Hale, W.; Chemler, J.; Congdon, G.; Narayan, A.; Håkansson, K.; Sherman, D.; Smith, J.; Skiniotis, G. *Nature*. **2014**, *510*, 512

11. Reimer, J.; Aloise, M.; Harrison, P.; Schmeing, T. *Nature.* **2016**, *529*, 239

12. Dunbar, K. L.; Mitchell, D. A. *ACS Chem. Biol.* **2013**, *8*, 473

13. Jarrett, J. T. *J. Biol. Chem.* **2015**, *290*, 3972

14. Sit, C. S.; van Belkum, M. J.; McKay, R. T.; Worobo, R. W.; Vederas, J. C. *Angew. Chem. Int. Ed.* **2011**, *50*, 8718

15. Flühe, L.; Marahiel, M. A. *Curr. Opin. Chem. Biol.* **2013**, *17*, 605

16. Rea, M. C.; Sit, C. S.; Clayton, E.; O'Connor, P. M.; Whittal, R. M.; Zheng, J.; Vederas, J. C.; Ross, R. P.; Hill, C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9352

17. Sit, C. S.; McKay, R. T.; Hill, C.; Ross, R. P.;Vederas, J. C. *J. Am. Chem. Soc.* **2011**, *133*, 7680

18. Babasaki, K.; Takao, T.; Shimonishi, Y.; Kurahashi, K. J. *Biochem.* **1985**, *98*, 585

19. Kawulka, K. E.; Sprules, T.; Diaper, C. M.; Whittal, R. M.; McKay, R. T.; Mercier, P.; Zuber, P.;Vederas, J. C. *Biochem*. **2004**, *43*, 3385

20. Liu, W.T.; Yang, Y.L.; Xu, Y.; Lamsa, A.; Haste, N. M.; Yang, J.Y.; Ng, J.; Gonzalez, D.; Ellermeier, C. D.; Straight, P. D.; Pevzner, P. A.; Pogliano, J.; Nizet, V.; Pogliano, K.;Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16286

21. Thennarasu, S.; Lee, D.K.; Poon, A.; Kawulka, K. E.; Vederas, J. C.;Ramamoorthy, A. *Chem. Phys. Lipids.* **2005**, *137*, 38

22. Wang, G.; Feng, G.; Snyder, A. B.; Manns, D. C.; Churey, J. J.;Worobo, R. W. *FEMS Microbiol. Lett.* **2014**, *357*, 69

23. Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. AMBER 2015, **2015.**

24. (a) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins*. **2006**, *65*, 712 (b) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J*. **2009**, *97*, 853

25. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327

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

27. Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729

28. Himes, P. M.; Allen, S. E.; Hwang, S.; Bowers, A. A. *ACS Chem. Biol.* **2016**, *11*, 1737

29. Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. *J. Am. Chem. Soc.* **2012**, *134*, 5309

30. Melby, J. O.; Nard, N. J.; Mitchell, D. A. *Curr. Opin. Chem. Biol.* **2011**, *15*, 369

31. Tianero, M. D.; Donia, M. S.; Young, T. S.; Schultz, P. G.; Schmidt, E. W. *J. Am. Chem. Soc.* **2012**, *134*, 418

32. Velasquez, J. E.; van der Donk, W. A. *Curr. Opin. Chem. Biol.* **2011**, *15*,11

33. Bashiruddin, N.K.; Suga, H.; *Curr. Opin. Chem. Biol.* **2015**, *24*, 131

34. Takahashi, T.T.; Austin, R.J.; Roberts, R.W. *Trends Biochem. Sci.* **2003**, *28*, 159

35. Barendt, P.A.; Ng, D.T.W.; McQuade, C.N.; Sarkar, C.A.; *ACS Comb. Sci.* **2013**, *15*, 77

36. Guillen Schlippe, Y.V.; Hartman, M.C.T.; Josephson, K.; Szostak, J.W.; *J. Am. Chem. Soc.* **2012**, *134*, 10469

37. Horiya, S.; Bailey, J.K.; Temme, J.S.; Guillen Schlippe, Y.V.; Krauss, I.J. *J. Am. Chem. Soc.* **2014**, *136*, 5407

38. Millward, S.W.; Takahasi, T.T.; Roberts, R.W. *J. Am. Chem. Soc.* **2005**, *127*, 14142

39. Hayashi, Y.; Morimoto, J.; Suga, H. *ACS Chem. Biol.* **2012**, *7*, 607

40. Jalai-Yazdi, F.; Lai, L.H.; Takahasi, T.T.; Roberts, R.W. *Angew. Chem. Int. Ed.* **2016**, *55*, 4007

41. Maini, R.; Umemoto, S.; Suga, H. *Curr. Opin. Chem. Biol.* **2016**, *34*, 44

42. The PEW charitable trust. http://www.pewtrusts.org/~/media/assets/2016/05/ascientificroadmapforantibioticdiscovery.pdf (Accessed June 6, 2017)

43. Ghodge, S.V.; Biernat, K.A.; Bassett, S.J.; Redinbo, M.R.; Bowers, A.A. *J. Am. Chem. Soc.* **2016**, *138*, 5487

44. Hudson, G.A; Zhang, Z.; Tietz, J.I.; Mitchell, D.A.; van der Donk, W.A. *J. Am. Chem. Soc.* **2015**, *137*, 16012

45. Wever, W.J.; Bogart, J.W.; Bowers, A.A. *J. Am. Chem. Soc.* **2016**, *138*, 13461

46. Zhang. Z.; Hudson, G.A.; Mahanta, N.; Tietz, J.I.; van der Donk, W.A.; Mitchell, D.A. *J. Am. Chem. Soc.* **2016**, *138*, 15511

CHAPTER 2

STRUCTURAL INSIGHTS INTO THIOETHER BOND FORMATION IN THE BIOSYNTHESIS OF SACTIPEPTIDES

2.1 Introduction

Enzymes that belong to the *S*-adenosylmethionine (SAM) radical superfamily are capable of catalyzing a wide array of radical mediated reactions utilizing the 5'-deoxyadenosyl radical (5'-dAdo•) as a radical intermediate. These reactions include modification to not only DNA and RNA, but also complex peptide modifications such as the formation of carbon-carbon bonds and quaternary carbon-sulfur bonds.[1,2] These radical SAM (RS) enzymes (rSAMs) contain conserved domains and motifs that unite the family. rSAMs bind several [4Fe-4S] clusters that carry out the chemistry of the enzyme. One [4Fe-4S] cluster is bound by a $CX_3CX\varphi C$ motif where φ is an aromatic residue. This motif is present in a conserved partial $(\beta/\alpha)_6$ triose-phosphate isomerase (TIM) barrel and provides three cysteines to coordinate with the iron atoms in present in the cluster, while the fourth iron atom is ligated by the amine-nitrogen and carboxyl oxygen from the methionine present in SAM.[1,3,4] This direct ligation of SAM to the [4Fe-4S] cluster allows reductive cleavage of the C-S bond upon electron transfer to the σ*-antibonding orbital of the SAM sulfonium group leading to the formation of methionine and the5'-dAdo• intermediate. This radical abstracts a hydrogen atom from the enzyme substrate and facilitates the particular chemical transformation carried out by the enzyme.[5] The [4Fe-4S] cluster is then typically

regenerated by a chemical reductant such as dithionite, or, in some cases, by the enzymatic NADPH/flavodoxin-flavodoxin reductase system.[1] Sequence homology suggests that many rSAMs contain a unique C-terminal extension, termed a SPASM domain in addition to a conserved RS domain.[2,3,6-9] The SPASM domain (named for the biochemically characterized members, AlbA, PqqE, anSME, and MftC which are involved in subtilosin A, pyrroquinoline quinone, anaerobic sulfatase, and mycofactocin maturation respectively) is involved in the coordination of auxiliary [4Fe-4S] clusters by cysteine residues, which are thought to expand and enhance the range of chemistries accessible by the RS domain.[2,7,8] All known SPASM domain-containing enzymes catalyze overall oxidation of their respective substrates by two electrons, yet there appears to be significant sequence and structural variation among SPASM domains, namely in the state and arrangement of cysteine residues that coordinate to the auxiliary iron-sulfur clusters.[7,8] The crystal structure of anSME was solved in 2013 and showed that the SPASM domain housed two additional, fully ligated auxiliary [4Fe-4S] clusters that were important for the enzymatic reaction of anSME. anSME co-translationally catalyzes the formal 2-electron oxidation of a cysteine residue found in the active site of its sulfatase substrate to yield a formyl glycine residue **(Figure 2.1b)**.[9-11] To date this is the only modifying enzyme belonging to the rSAM superfamily, with a full SPASM domain, whose structure has been solved.

Some of the founding members of the SPASM domain, AlbA and PqqE, and thus their peptide substrate products, subtilosin A and pyrroquinoline quinone respectively, also belong to the natural product class known as ribosomally synthesized and post-translationally peptides (RiPPs). RiPPs, which are peptides that are first synthesized by the ribosome and then modified later by tailoring enzymes, have gained attention due to their structural diversity and biological activities.[12-15] The modifications involve leader peptide-directed, enzymatic transformations by

promiscuous enzymes that are readily exploited for combinatorial biosynthesis as well as other applications.[16-20] The leader peptide is responsible for binding to the post-translational modifying enzymes while the chemical modification is done on the core peptide. Recently it has been shown that leader peptide interacts with a conserved motif present in RiPPs enzymes known as a RiPP precursor peptide recognition element (RRE). These RREs have been found in a wide variety of RiPPs enzymes such as LynD, a cyclodehyratase involved in cyanobactin biosynthesis, and NisB, a dehydratase involved in the biosynthesis of the lantibiotic nisin.[21] These domains are based on the structure of PqqD which associates with the rSAM PqqE to allow the formation of the carbon-carbon bond between a glutamic acid and tyrosine residues.

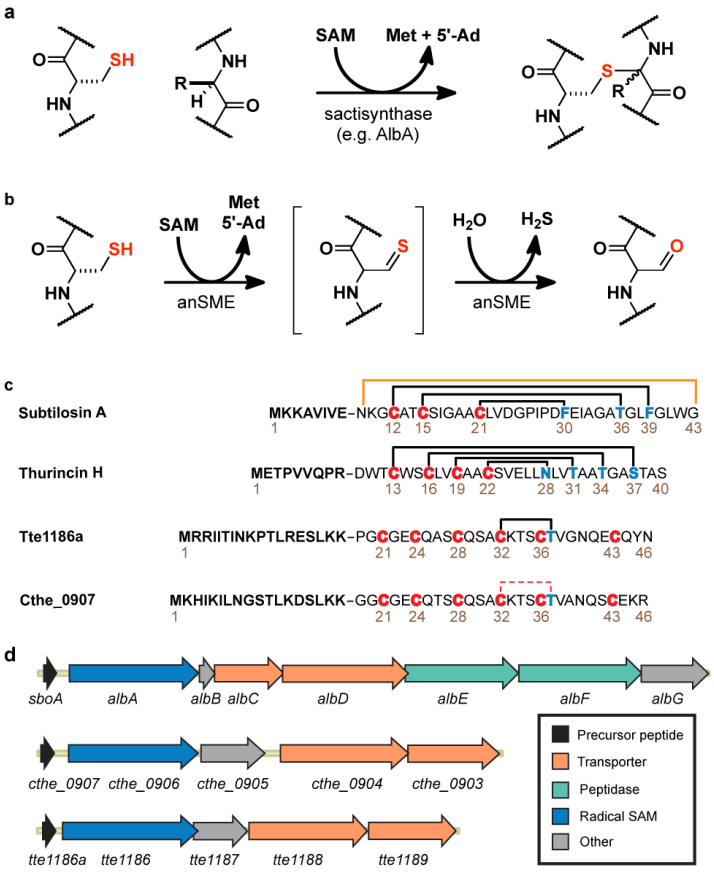


Figure 2.1. Introduction to sactipeptides. a) Formation of sactionine thioether linkages found in sactipeptides. b) Formation of formyl glycine from cysteine by anSME c) Comparison of known sactipeptides to the bridge formed in CteA. d) Gene clusters of some known sactipeptide producers.

Not only can rSAMs catalyze the formation of carbon-carbon bonds such as those found in PQQ biosynthesis by PqqE and the recently characterized StrB from streptide biosynthesis[22-25], they can also form sulfur to α-carbon bonds to form thioether (sactionine) bridges, a characteristic of the subclass of RiPPs known as sactipeptides. Sulfur-to-α-carbon-antibiotics (sactibiotics), also known as sactipeptides, are in the family of sulfur bridged bacteriocins, but unlike their lantipeptide cousins, sactipeptides are formed by making sulfur to α-carbon bonds (sactionine bridges) through the use of rSAMs, termed sactionine synthases.[26-28] These sactionine bridges impart conformational strain on the peptide, giving rise to unique secondary structures forming natural "stapled" helices. Sactipeptide may contain one or more of these sactionine linkages that form their constrained macrocyclic peptide backbones, making them resistant to heat, proteolysis and degradation.[29,30] These sactipeptides have been shown to have broad spectrum activity against gram positive bacteria. Thuricin CD, a particular two component sactipeptide, has even shown nanomolar activity against *Clostridium difficile,* a common hospital secondary infection.[30-36] These properties make sactipeptides attractive scaffolds for antibiotic development. Recent work has shown that sactionine synthases cleave SAM to generate 5'-dA•, which is used to catalyze thioether bond formation by an incompletely understood and not well-defined mechanism **(Figure 2.1a)**.[4,37-42] Elegant work by Marahiel and co-workers characterized the first member of sactionine synthases from the biosynthesis of the sactipeptide subtilosin A. Marahiel demonstrated that the sactionine synthase AlbA radically cleaves SAM and subsequently catalyzes the formation of three sactionine linkages on its precursor peptide.[37] By means of an *in vivo* engineered expression system, we recently showed that AlbA exhibits broad substrate promiscuity and that SboA cysteines and their attendant cross ring bridging partners can be repositioned within the precursor peptide sequence while still undergoing enzymatic

modification to sactionine linkages.[43] These efforts are limited to substrates that fortuitously undergo modification by the native enzyme, but engineering and rational design of sactipeptide libraries will require better understanding of both sactisynthase structure and mechanism of thioether bond formation. Not only are these rSAMs predicted to hold SPASM domains (only a few of which has been structurally characterized)[2,3,6-9], but they are also RiPPs enzymes and predicated to hold RRE domains attached to, not separate, of the active site of the enzyme. That makes any structural information on these thioether bond forming sactionine synthases extremely important. A crystal structure of a sactionine synthase can shed light on a) how the SPASM domain of a peptide modifying enzyme compares to known SPASM domains, b) how the RRE domain interacts with the peptide substrate, c) insights into the mechanism of sactionine bond formation, and d) what this can tell us about other known sactionine synthases.

Recent bioinformatic efforts predicted a number of sactipeptide clusters in a wide array of bacterial genomes, including several from thermophiles.[7,8,44] We anticipated that sactionine synthases from thermophilic bacteria might have the desired stability for efficient heterologous expression and crystallization. In particular, the sactionine synthase from *Clostridium thermocellum* ATCC 27405, Cthe_0906, here referred to as CteB, looked to be a member of the newly defined family of sactipeptides being called SCIFF (or six cysteines in forty-five residues) peptides.[7] CteB is co-located with the short peptide Cthe_0907, here referred to as CteA, in the *C. thermocellum* ATCC 27405 genome and is therefore predicted to chemically modify CteA to form its SCIFF peptide known as thermocellin (cte). Although no native natural products belonging to this family have been isolated to date, Bandarian and co-workers reconstituted enzymatic activity of Tte1186 from a putative SCIFF pathway in *Caldanaerobacter subterraneus subsp. tengcongensis* MB4.[41] We therefore considered CteB a strong candidate for

enzymatic reconstitution and potential crystallographic investigation aimed at understanding the mechanism of sactionine synthases. In this study, we reconstitute the activity of the sactisynthase, CteB and through a combination of chemical modification and tandem mass spectrometry, we demonstrate that CteB installs a single sactionine thioether linkage between Cys32 and Thr37 of its precursor peptide, CteA, and that the remaining five cysteines in CteA go unmodified. We also report two structures of CteB: a 2.70-Å-resolution structure of CteB with SAM bound and a 2.04-Å-resolution structure of CteB with both SAM and the leader peptide of CteA bound. These structures define all three [4Fe-4S] clusters predicted by bioinformatics, one of which, auxiliary cluster I (Aux I), displays a novel open coordination site on one of its constituent iron ions. These structures, together with substrate binding assays and a computational model based on the crystal structure, provide insights into the mechanism of thioether bond formation for CteA and other members of the sactisynthase family.

2.2 *In vitro* reconstitution of CteB: a sactionine synthase

The genes that encode CteB and its peptide substrate CteA were codon optimized for expression in *E. coli* and separately cloned into different expression vectors. Previously we had reported improved yields of peptide when co-expressed with its sactisynthase, presumably due to added protection from proteolysis.[43] Therefore, the precursor peptide, CteA, co-expressed with CteB in a pETDuet vector in multiple cloning sites 1 and 2 respectively. Only CteA was 6xHis-tagged in the construct and could be readily purified from inclusion bodies formed during expression at 18°C with generous aeration. No modification of CteA was observed under these aerobic conditions, making recombinant CteA obtained in this manner suitable for an *in vitro* enzymatic assay. CteB could be expressed and purified in a manner similar to other radical SAM enzymes and sactisynthases **(see Experimental section 2.7.2)**.[45]

Using Fluhe *et al*. as a basis[37], we ran a series of anaerobic assays to determine a) whether CteB could reductively cleave SAM to generate 5'-deoxyadenosine or 5'-dA and b) if/how it modified CteA. When there is no peptide substrate for a sactisynthase, the 5'-dA• abstracts a hydrogen from nearby solvent, forming 5'-dA. The production of 5'-dA is a clear indication of proper folding, redox state, and activity of the enzyme. When CteB was incubated in the presence of SAM and the strong non-physiological reductant sodium dithionite, we observed a distinctive mass (252.1108) corresponding to 5'-dA (within <10 ppm error) in LC-MS traces of the assay supernatants **(see Appendix Figure A.1)**. This product mass was not observed in control reactions without CteB or SAM, suggesting that reconstituted CteB carries out this characteristic reductive cleavage of radical SAM enzymes.

The next step was to confirm whether or not CteB was in fact a sactionine synthase, capable of forming sactionine bridges within its peptide substrate CteA. With this positive result, we went forward with the peptide modification assay. We ran a series of assays incubating CteB together with CteA, SAM and dithionite followed by analysis of the products by HPLC coupled to an Agilent 6520 Accurate-Mass Q-TOF spectrometer. In the presence of stoichiometric amounts of enzyme, we observed complete conversion of CteA to a mass that was 2.0 atomic mass units (amu) lower than the starting mass, consistent with the loss of two hydrogen atoms corresponding to the formation of a single sactionine thioether bond. **Figures 2.2a and 2.2b** show examples of the mass shift in the envelope corresponding to the +6 charge state of CteA. To confirm that the 2.0 amu loss resulted from a thioether linkage and not a disulfide bond (also a 2.0 amu loss) , we quenched the reactions under reducing conditions and reacted with N-ethylmaleimide (NEM), in order to alkylate all free cysteines **(Figures 2.2a and 2.2c)**. In a control reaction, where CteA was not treated with CteB but treated with NEM, the

m/z values for the various charge states corresponded to the mass of the peptide plus six molecules of NEM **(Figure 2.2c, black trace and Table 2.1)**. In contrast, CteA that was modified by CteB before being quenched with NEM exhibited masses corresponding to m/z for peptide with five alkylated cysteine residues and a loss of two hydrogens **(Figure 2.2c, red trace and Table 2.1)**, confirming that a single thioether had been installed by CteB under these conditions.

In order to identify the location of the single sactionine thioether linkage, we used tandem mass spectrometry. CteA that was modified with CteB and treated with NEM (CteA-mod-5NEM) was fragmented by collision induced dissociation (CID). Based on the pattern of b- and y- ions, the newly formed thioether bridge was found to reside between residues Cys32 and Thr37 of CteA **(Figure 2.2d)**. We found that the b- and y- ions for fragments containing Cys32 lack one NEM group and two hydrogens corresponding to the formation of the sactionine thioether linkage at this position. A full table of observed masses and the residues to which they correspond is provided in **Table 2.2**.

As mentioned before, as we were working on our system, Bandarian and co-workers reconstituted the activity of a sactisynthase, Tte1186, from a SCIFF pathway in *Caldanaerobacter subterraneus subsp. tengcongensis* MB4.[41] Upon comparison to our system, we found that the peptide substrates (CteA and Tte1186a) had 65.2% shared identity and 76.1% consensus sequences while the enzymes (CteB and Tte1186) had 55.0% shared identity and 70.0% consensus sequences. Thus when comparing our tandem MS-MS spectra for CteB-modified CteA to their corresponding tandem MS-MS spectra for Tte1186-modified Tte1186a, it was interesting that we found the same pattern of b- and y- ions that they had seen, further giving credence to the sactionine linkage placement between Cys32 and Thr37. To further corroborate

the assignment of the sactionine linkage, the Cys32Ala mutant of CteA was prepared via Gibson Assembly mutagenesis, and purified similar to wild type. Assays with CteA-C32A in presence of CteB and SAM yielded only the unmodified precursor peptide, consistent with thioether formation at this position **(see Appendix Figure A.3)**. Peptide products with only one sactionine bridge were observed regardless of whether CteB was limited or used in large excess. It cannot be completely ruled out that multiple thioether bridges may be formed in the cellular environment of the native producer with the native reductant. Whether this is the active form of CteA *in vivo* remains to be determined.

Table 2.1. Mass Spec. table for peptide modification assays treated with NEM

SNA-CteA-WT				
	Exact Mass	Thioether Bridge	1-Thioether Bridge 5-NEM	No Thioether Bridge 6-NEM
	5175.4484	5173.4328	5798.6711	5925.7345
z	m/Z	m/Z	m/Z	m/Z
1	5176.4562	5174.4406	5799.6789	5926.7423
2	2588.7320	2587.7242	2900.3434	2963.8751
3	1726.1573	1725.4854	1933.8982	1976.2527
4	1294.8699	1294.3660	1450.6756	1482.4415
5	1036.0975	1035.6944	1160.7420	1186.1547
6	863.5826	863.2466	967.4530	988.6302
7	740.3576	740.0696	829.3894	847.5413
8	647.9389	647.6869	725.8417	741.7246
9	576.0576	575.8337	645.3046	659.4228
10	518.5527	518.3511	580.8749	593.5813

Table 2.2. Tandem MS/MS table for peptide modification assays treated with NEM

MS/MS of Modified CteA + NEM											
		b-ions						y-ions			
Residue	b-ion #	Z=1	Z=2	Z=3	Z=4	Z=5	y-ion #	Z=1	Z=2	Z=3	Z=4
Gly22	24	2650.4167	1325.7123	884.1441	663.3600	530.8896	25	3150.2690	1575.6384	1050.7615	788.3231
Glu23	25	2707.4382	1354.2230	903.1513	677.6154	542.2939	24	3093.2476	1547.1277	1031.7544	774.0678
Cys24 + NEM	26	2836.4808	1418.7443	946.1655	709.8761	568.1024	23	2964.2050	1482.6064	988.7402	741.8071
Gln25	27	3064.5377	1532.7728	1022.1845	766.8903	613.7138	22	2736.1481	1368.5779	912.7212	684.7929
Thr26	28	3192.5962	1596.8020	1064.8706	798.9049	639.3255	21	2608.0895	1304.5486	870.0350	652.7782
Ser27	29	3293.6439	1647.3259	1098.5532	824.1668	659.5350	20	2507.0418	1254.0248	836.3525	627.5163
Cys28 + NEM	30	3380.6750	1690.8414	1127.5636	845.9246	676.9413	19	2420.0098	1210.5088	807.3418	605.7583
Gln29	31	3608.7328	1804.8703	1203.5828	902.9391	722.5528	18	2191.9529	1096.4803	731.3228	548.7441
Ser30	32	3736.7914	1868.8996	1246.2690	934.9537	748.1645	17	2063.8944	1032.4511	688.6367	516.7295
Ala31	33	3823.8234	1912.4156	1275.2797	956.7117	765.5709	16	1978.8779	989.9429	660.2979	495.4753
Cys32*	34	3894.8605	1947.9342	1298.9587	974.4710	779.7784	15	1907.8408	954.4243	636.6188	477.7161
Lys33	35	4122.9174	2061.9626	1374.9777	1031.4852	825.3897	14	1679.7840	840.3959	560.5999	420.7019
Thr34	36	4251.0124	2126.0101	1417.6760	1063.5090	851.0087	13	1551.6890	776.3484	517.9016	388.6781
Ser35	37	4352.0600	2176.5339	1451.3586	1088.7709	871.2183	12	1450.6413	725.8246	484.2190	363.4162
Cys36 + NEM	38	4439.0921	2220.0500	1480.3693	1110.5289	888.6247	11	1363.6093	682.3086	455.2083	341.6582
Thr37*	39	4542.1012	2271.5545	1514.7056	1136.2812	909.2265	10	1260.6001	630.8040	420.8719	315.9059
Val38	40	4641.1333	2321.0705	1547.7163	1161.0392	929.0329	9	1159.5524	580.2801	387.1894	290.6440
Ala39	41	4740.2017	2370.6047	1580.7391	1185.8063	948.8466	8	1060.4840	530.7459	354.1666	265.8769
Asn40	42	4811.2388	2406.1233	1604.4181	1203.5656	963.0540	7	989.4469	495.2274	330.4875	248.1176
Gln41	43	4925.2818	2463.1448	1642.4325	1232.0763	985.8626	6	875.4040	438.2059	292.4732	219.6069
Ser42	44	5053.3404	2527.1741	1685.1187	1264.0910	1011.4743	5	747.3454	374.1766	249.7870	187.5922
Cys43 + NEM	45	5140.3724	2570.6901	1714.1293	1285.8490	1028.8807	4	660.3134	330.6606	220.7764	165.8342
Glu44	46	5368.4292	2684.7185	1790.1483	1342.8632	1074.4921	3	432.2565	216.6322	144.7574	108.8200
Lys45	47	5497.4718	2749.2398	1833.1625	1375.1238	1100.3006	2	303.2139	152.1109	101.7432	76.5593
Arg46	48	5625.5668	2813.2873	1875.8608	1407.1476	1125.9196	1	175.1190	88.0634	59.0449	44.5356

* Cys32 and Thr37 are bridged residues

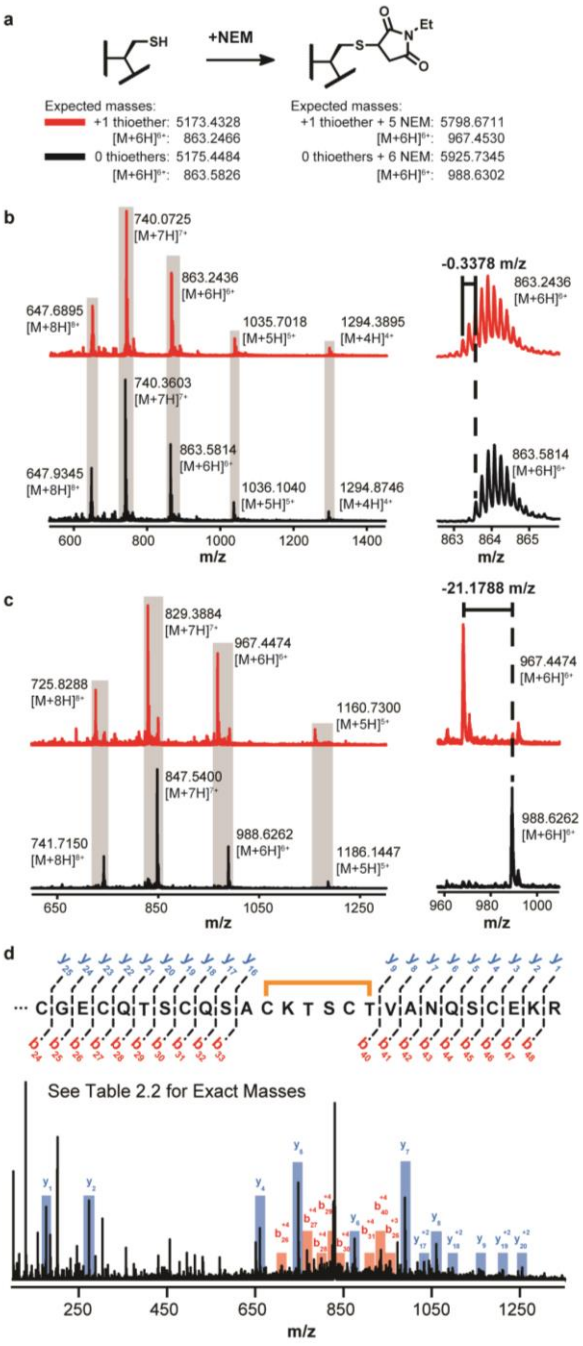


Figure 2.2. MS analysis of CteA modified by CteB. a) Expected masses of CteA modified by CteB or alkylated with NEM. b) MS of CteA product modified by CteB. In red is 1 eq. of CteA treated with 1 eq. of CteB while in black is CteA unmodified. The difference between the corresponding charge states is that of one sactionine bridge or two hydrogen atoms. c) MS of CteA product alkylated with NEM after first being modified by CteB. In red is 1 eq. of CteA treated with 1 eq. of CteB then NEM, while in black is CteA treated with just NEM. The difference between the corresponding charge states is that of one sactionine bridge and one NEM modification See Table 2.1 for expected exact masses. d) MS/MS analysis (+7 charge state) of where the sactionine bridge is forming in modified CteA.

2.3 Crystal structure of CteB

We solved the crystal structures of CteB, from anomalous iron edge datasets, in two different states: CteB bound to SAM and no substrate was solved at 2.7 Å resolution while CteB bound to both SAM and a 21-residue N-terminal fragment of CteA (M_1-C_{21}) at 2.04 Å resolution. We attempted co-crystallization with the full-length CteA precursor peptide, but were unable to obtain diffraction quality crystals. As of the time of writing, there have been no structures reported for any RiPP enzyme and its full-length precursor peptide substrate bound, presumably due to the dynamic nature of the interactions between the core peptide and the RiPP enzyme.

The two CteB structures, CteA-bound and unbound, superimpose with a root mean square deviation (RMSD) of 1.3 $Å^2$ based on 415 Cα atoms **(see Appendix Figure A.11)**. Because of this close structural similarity, our discussion focuses mainly on the higher resolution structure, CteB+SAM+CteA-M_1-C_{21} (Pep-Bound-CteB), with reference to the unbound structure where relevant. Crystals of the enzyme-peptide complex exhibit diffraction consistent with the orthorhombic space group $P2_12_12$, with a monomer in the asymmetric unit **(Figure 2.3a)**. The final model consists of CteB residues 1 to 449 (out of 450), residues 1 – 9, and 20- 21 of the CteA-M_1-C_{21} peptide, 12 iron ions and 12 sulfide ions which correspond to 3-[4Fe-4S] clusters, 2 calcium ions, 1 SAM molecule and 146 water molecules. CteB residues 115-121 are not defined by electron density, and reside in a disordered loop immediately following the RS [4Fe-4S] cluster binding motif. A similar disordered loop was also seen in the structure of anSME.[9] In addition, CteB residues 334-336 are not defined in a disordered loop that joins α6 in the RS domain to the SPASM domain. This region lies on a symmetry axis and is difficult to model **(see Appendix Figure A.8)**. This region contains a conserved cysteine (Cys336) found in SCIFF

maturases **(Appendix A.15)**, which may form a disulfide bond with the adjacent Cys336 from a symmetry mate **(see Appendix Figure A.8)**. In addition, this entire region (CteB residues 330-341) is disordered in the CteB+SAM structure. Interestingly, the crystallization solution contained about 500 μM dithiothretol (DTT) that carried over with the CteB added to the solution and all predicted [4Fe-4S] clusters are intact in the crystals indicating a lack of oxidative damage. To determine if a disulfide exists in the crystalline state, crystals of CteB+SAM+CteA-M_1-C_{21} were dissolved in buffer and the solution chromatographed on a size exclusion column equilibrated in buffer that did not contain reductants **(see Appendix Figure A.4)**. The majority (~ 65 %) of protein in this sample migrated with an apparent molecular weight of ~ 95 kDa, which is consistent with a dimer of CteB, while the remaining protein migrated as a monomer. The monomer fraction of this solution is likely CteB protein that was present in the crystallization drop but did not form crystals. We next tested the oligomeric state of CteB in solution, both in presence and absence of full-length CteA substrate, under reducing conditions by size exclusion chromatography (SEC). Under these conditions, CteB migrates with an apparent mass of 42 kDa, consistent with a monomer **(see Appendix Figure A.5)**. In the presence of CteA, the apparent mass of CteB increases by ~ 4 kDa, consistent with a CteA-CteB complex. Thus, in solution, under reducing conditions CteB does not seem to form a dimer. Next, we removed the reductant, DTT, from CteB and mixed protein with varying ratios of reduced glutathione (GSH)/oxidized glutathione (GSSG) to survey the redox potential of the mixture from – 377 mV to – 223 mV. These mixtures were subsequently electrophoresed on a non-reducing SDS-PAGE gel **(see Appendix Figure A.6)**. Importantly, increasing the redox potential by increasing the ratio of GSSG to GSH does not lead to significant intermolecular disulfide bond formation in solution as can be seen by the lack of a dimer band at ~100 kDa in

the non-reducing SDS-PAGE gel. With these results, we concluded that this disulfide is most likely an artifact of crystallization. Due to the weak electron density and the evidence that in solution, the enzyme behaves as a monomer, we have decided not to model this disulfide bond forming a dimer **(see Appendix Figure A.8)**.

However this leaves the question of what is the role of this highly conserved cysteine in SCIFF maturases? We set out to answer this by mutating Cys336 to an alanine and testing its ability to cleave SAM radically and modify CteA. Shown in **Appendix A.1 and A.2**, CteB-C336A cleaves SAM radically on par with that of CteB. However, unlike CteB, CteB-C336A does not modify CteA to form a single sactionine bridge, implicating this conserved residue plays a role in the activity of CteB on the substrate. Looking at the structure, we hypothesized that the cysteine rich CteA substrate could form a disulfide with CteB at Cys336. Using the same conditions to prove that CteB did not form a dimer in solution, we were able to show that CteB-CteA complex forms under oxidizing conditions **(see Appendix A.7)** indicating the presence of a disulfide bond formed between the enzyme and substrate. We hypothesize that Cys336 in CteB may form a disulfide with one of the other N-terminal cysteines (Cys21, Cys24, Cys28) in CteA and orient the peptide substrate within the active site for modification at Cys32. Therefore, we made the individual cysteine to alanine mutants of CteA (C21A, C24A, and C28A) and tested whether CteB could modify each substrate. Shown in **Appendix A.3**, all cysteine to alanine mutants (except for C32A) were modified to a single sactionine bridge by CteB indicating that a single residue change did not affect the ability of CteB to form a disulfide complex with CteA and form the sactionine bridge. A possible explanation for this observation is that the possible cysteines, within CteA, to form a disulfide bond with Cys336 in CteB are all equal distant from Cys336A and therefore more than one possible bridge can be formed between

the substrate and enzyme, replacing a disulfide loss upon mutation. Multiple, subsequent mutations within the CteA substrate would be required to answer this question. Further experiments are underway to elucidate the role Cys336 plays in complex formation and activity of CteB.

The structure of CteB exhibits three discernable domains **(Figure 2.3a and b)**: (1) a partial $(\beta/\alpha)_6$ triose phosphate isomerase (TIM) barrel (residues 95-319) containing one [4Fe-4S] cluster (canonical radical SAM domain) in green, which is flanked by (2) an N-terminal winged helix-turn-helix (wHTH) motif (residues 1-71) in purple and (3) a C-terminal extension (residues 338-450), which chelates two additional [4Fe-4S] clusters in orange. These are discussed individually below.

The central portion of the CteB structure exhibits the characteristic $(\beta/\alpha)_6$-TIM barrel (residues 95-319), common to nearly all other members of the radical SAM superfamily. This barrel is also known as the AdoMet or radical SAM (RS) domain for the fact is holds the [4Fe-4S] cluster that binds and reductively cleaves SAM. The [4Fe-4S] cluster motif (CX3CXφC) is found within the RS domain, in the loop between α1 helix and β1 loop (residues 100-125). This cluster is ligated by three cysteines (residues 104, 108, and 111), leaving one site open to chelate the α-aminonitrogen and α-carboxyl oxygen of the SAM co-factor.[3,46] The rest of the SAM binding pocket is similar to that of the SAM binding pocket of anSME and exhibits the four common SAM binding motifs: the GGE motif (residues 153-156), the ribose motif (residues Ser210 and Asp212), the GXIXGXXE motif (residues 254-262), and the β6 or adenine-binding motif (residues 281-284). In addition, Tyr110 forms a hydrogen bond to the N6 of the adenine present in SAM and Arg222 stabilizes the ribosyl and carboxyl moieties of the AdoMet; these interactions are also present in the anSME SAM binding pocket **(see Appendix Figure A.9)**.

Interestingly there are two new SAM binding pocket interaction found in the CteB structure: Arg253 and Thr255 in the β5 strand form hydrogen bonds to N3 of the adenine base within SAM. These residues reside in a highly conserved RGT motif found in thermophilic sactisynthases. In summary, a total of eight residues make side chain or backbone polar contacts with SAM **(see Appendix Figure A.9)**. Presumably these numerous interactions and motifs correctly position and orient SAM for radical-based hydrogen abstraction from its substrate in a very specific manner.

Through a partially ordered loop, the RS domain is connected to the C-terminal SPASM domain which spans the residues 338-450. It holds the conserved seven-cysteine motif found in SPASM domains, $CX_{9-15}GX_4CX_nCX_2CX_5CX_3CX_nC$ and coordinates two additional [4Fe-4S] clusters known as auxiliary clusters. The CteB SPASM domain exhibits structural homology (**see Appendix Figure A.9**, R.M.S.D. of 2.3 Å over 113 Cα) to the SPASM domain from anSME with several notable differences between the two. The SPASM domain in CteB extends from the C-terminus of the TIM barrel RS domain via a partially ordered loop to coordinate the first auxiliary [4Fe-4S] cluster (Aux I) at Cys344 and Cys362. A short hairpin loop (β1'/β2') separate these two coordinating residues. In anSME, a short amino acid insertion harbors Cys261, which is the fourth ligand to Aux I. This amino acid insertion and thus the fourth cysteine, is absent in CteB. As in anSME, the $CX_2CX_5CX_3C$ motif in the central region of SPASM domain provides three ligands for the second auxiliary [4Fe-4S] cluster (Aux II) and one additional cysteine ligand for Aux I in CteB. Cys400, Cys403, and Cys409 from CteB all coordinate Aux II, while the fourth cysteine of the motif, Cys413, crosses back to provide a third ligand for Aux I. Cys432 provides the fourth and final ligand for Aux II **(Figure 2.3b)**. Aux II is also exposed to the surface by a small channel on the back side of the protein. In contrast to

ansME, where both Aux I and Aux II are fully ligated, Aux I of CteB is left with an open coordination site. The fourth coordinating ligand, present in anSME but absent in CteB, to Aux I, besides leaving an open coordination site, results in the positioning of the [4Fe-4S] cluster AuxI closer to the RS cluster in CteB. The RS cluster resides 14.4 Å from the open coordination site of Aux I **(Figure 2.3c)**, which is ~ 2.5 Å closer than seen in the structure of anSME (16.9 Å). The distance between Aux I and Aux II in CteB is 11.6 Å, which is slightly compressed compared to that seen in the SPASM domain of anSME (12.9 Å). These differences indicate that the overall arrangement and separation of all [4Fe-4S] clusters within these SPASM family proteins is likely to support the different chemistries that are catalyzed by these different proteins.

In the CteB+SAM+CteA-M_1-C_{21} structure, electron density is observed around the open coordination site of Aux I. We first attempted to model this electron density as a weakly bound DTT (present in the crystalizing conditions) molecule as observed in the crystal structure of lipoyl synthase[47] but this model did not result in a satisfactory fit of the electron density. We next hypothesized that the free thiol of Cys21, from the CteA-M_1-C_{21} peptide, could reach into the active site and coordinate to the open ligation site in Aux I. The modeling of residues of Gly20 and Cys21 of the CteA-M_1-C_{21} peptide into this density provided a plausible fit **(Figure 2.3d)**. The lack of electron density for the peptide sequence between Gly9 and Gly20 is most likely due to innate flexibility of the peptide and the lack of hydrogen bonding or hydrophobic contacts between CteA and CteB in this region. Because only a fragment of the precursor peptide (residues 1-21) was found to co-crystallize with CteB, it is possible that the Cys21-Fe ligation seen in the reported structure is not mechanistically relevant, but rather represents a thermodynamically stable state of the peptide in the absence of the full, native sequence. Therefore, we propose that this Cys21 coordination may be analogous and similar to the enzyme-

substrate interactions that would occur during the catalytic cycle, involving one of the six cysteines from CteA and Aux I.

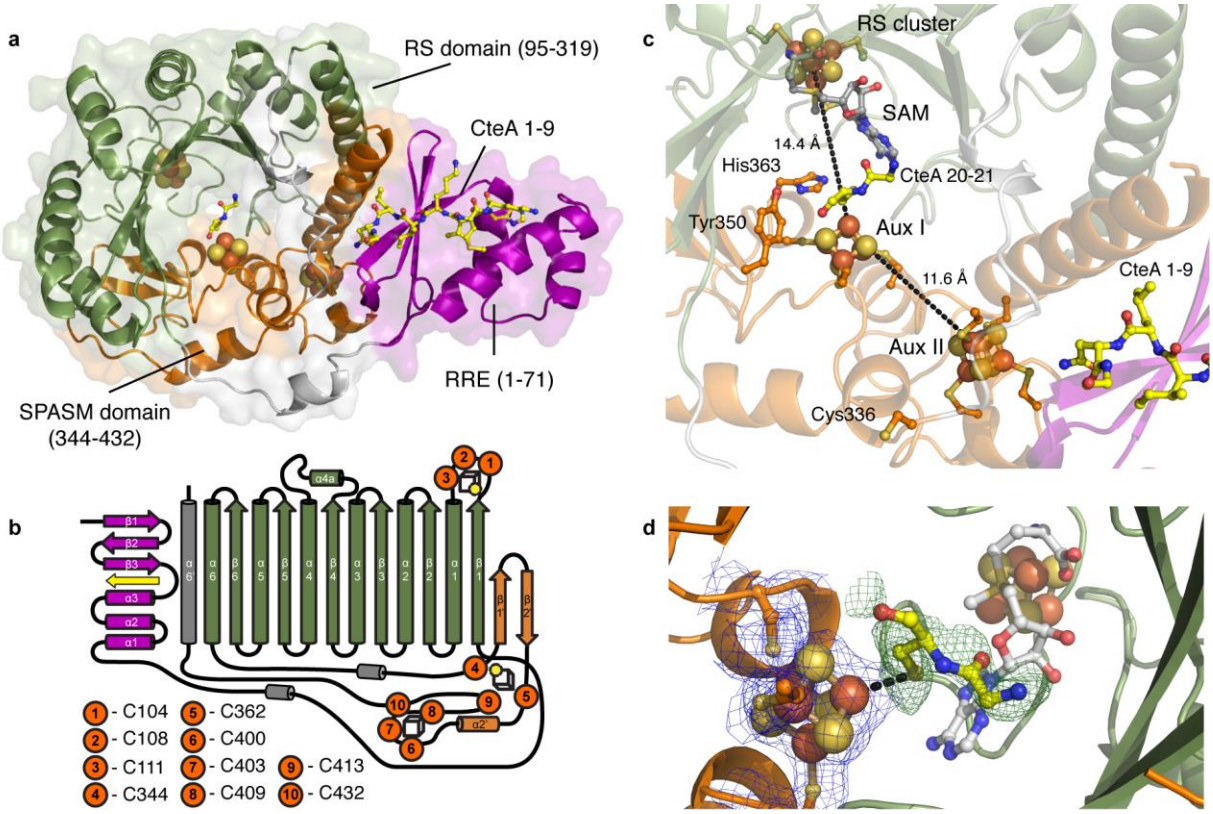


Figure 2.3. Structure of CteB. a) Overall structure of CteB. The β_6/α_6 core of the RS domain (green) contains one [4Fe-4S] cluster that coordinates one molecule of SAM. The C-terminal SPASM domain (orange) contains the [4Fe-4S] clusters Aux I and Aux II and comprises of residues 344 - 432. The N-terminal RRE domain (purple) of CteB provides the binding specificity for the peptide substrate leader sequence of CteA (yellow, stick representation). b) Topology figure of CteB with matching color scheme as a). c) Zoom of [4Fe-4S] clusters present in CteB along with their distances from one another. The distance from RS to Aux I is 14.4 Å while the distance from Aux I to Aux II is 11.6Å. RS, radical SAM cluster, Aux I, and, Aux II d) Omit map ($2F_o$-F_c) contoured to 1.5 σ of Gly20 and Cys21 from CteA-M_1-C_{21} substrate bound to Aux I. The distance between the Fe and S_Y of Cys21 is 2.7 Å.

One structural feature that is absent in anSME, but has been predicted to occur in AlbA and other SPASM domain proteins, is the RiPPs precursor peptide recognition element (RRE). The winged helix-turn-helix (wHTH) motif present in CteB is structurally homologous to the RREs, which have recently been identified in the crystal structures of other RiPP modifying enzymes, such as LynD, PaaA, and NisB.[21,48-54] The wHTH in CteB spans residues 4-71 and

represents just the fourth reported structure of a leader-bound RRE structure **(Figure 2.4a and d)**.[48,49] These reported structures all exhibit a common pattern in the conserved wHTH domain architecture and β-strand conformation when bound to peptide substrate **(Figure 2.4a-c)**. While the RREs vary greatly in sequence, they are predicted based upon secondary structure. Despite sharing only 13% sequence identity to the RRE domains of LynD and NisB, the RRE domain of CteB exhibits a relatively small overall R.M.S.D at 2.16 Å and 3.05 Å, respectively over 71 Cα. The RRE domain provides one of the primary structural motifs for leader peptide recognition. The three-stranded β-sheet, or wing, of the RRE interacts with the backbone of the N-terminus of the CteA fragment in the co-crystal structure in a manner similar to LynD and NisB **(Figure 2.4a-c)**. A very extensive hydrogen bond network is formed by backbone carbonyl and amide interaction of the RRE with CteA **(Figure 2.4d and Table 2.3)**. Hydrogen bonds can be seen between side-chain and main-chain atoms of CteA. His3 from CteA forms a series of salt bridges with the CteB residues Asp27, Glu60 and Glu64. CteA also makes favorable van der Waals interactions with the RRE domain *via* Ile4 and Ile6, both of which fit into hydrophobic pockets found in the cleft between α3 and β3 strands. The RRE is connected to the N-terminus of β1 of the partial $(\beta/\alpha)_6$ TIM barrel by a long, flexible linker, which passes across the face of the SPASM domain to position the RRE next to the α6' helix **(Figure 2.3a and b)**. β1 and β2 of the RRE also make hydrophobic contacts with the α6' helix coming from the C-terminus of the SPASM domain, which weakly stabilizes its position relative to the active site. In addition, the RRE domain makes limited crystallographic contacts with symmetry molecules and, as a result, shows higher than average β-factors than the core of CteB. This explains why the density for the leader peptide is weaker than the resolution would predict.

Table 2.3. List of CteB(RRE) and CteA(M1-G9) hydrogen bond interactions

CteA H-bond atom	CteB H-bond atom	Distance
1 – N-term NH	64 – Glu OE2	3.1
2 – backbone NH	64 – Glu OE1	3.3
3 – backbone NH	64 – Glu OE1	2.8
3 – His NE2	60 – Glu OE1	3.3
3 – His ND1	27 – Asp OD2	2.7
3 – backbone CO	27 – backbone NH	3.2
5 – backbone NH	25 – backbone CO	2.9
5 – backbone CO	25– backbone NH	2.7
7 – backbone NH	23 – backbone CO	2.6
8 – Asn OD1	23– backbone NH	3.1

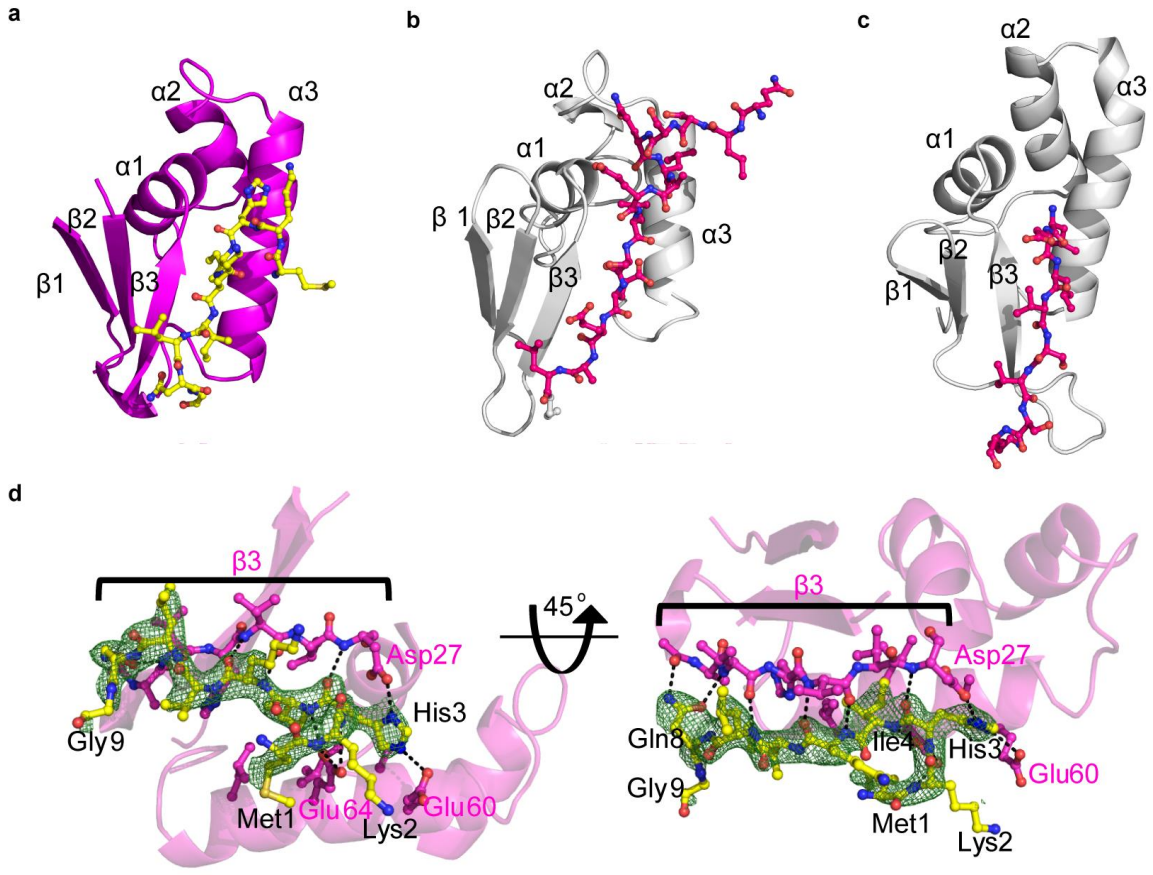


Figure 2.4. Leader peptide and binding to RRE of CteB. Comparison of RRE domains from CteB (a), LynD (b), NisB (c). d) Simulated annealing omit composite map ($2F_o$-F_c) contoured to 1.0 σ of residues 1 – 9 of the leader peptide (yellow sticks) of CteA. Residues from CteA involved in binding of the leader peptide are shown in yellow. Hydrogen bond interactions are shown as dashed lines. For full list of interactions and distances see Table 2.3.

2.4 Homology and Comparison to other SPASM and Twitch Domains

There are not many proteins to which CteB can be compared to as it is the first of its kind to be structurally characterized. CteB is only the second example of a SPASM domain to be structurally characterized, the first being anSME. In addition, BtrN[55] and MoaA[56] exhibit smaller, single [4Fe-4S] cluster domains dubbed "Twitch" domains.[2] Taken together, the four structures of the SPASM and Twitch domains provide four different coordination architectures for Aux I **(Figure 2.5)**. All four of these enzymes use the two conserved cysteines present on either side of the β1'/β2' hairpin loop (Cys344 and Cys362 in the case of CteB) but differ in the positioning of the remaining coordinating cysteines. MoaA has an open coordination site on Aux I, similar to that of CteB. However, the open iron sites in these two structures differ as they are on alternate sides of Aux I. While Cys413 from the $CX_2CX_5CX_3C$ SPASM motif loops back to provide the third coordination in CteB, this cysteine motif is not present in MoaA's twitch domain, and MoaA's corresponding Aux I is instead ligated by an additional cysteine, Cys264, upstream of the β1'/β2' hairpin loop. Cys264 in MoaA is analogous to the cysteine, Cys261, present in anSME, but absent in CteB. The difference in coordination pattern results in the open coordination site of CteB's Aux I being oriented towards the active site entrance, favorably positioned for coordination by an incoming peptide substrate. In contrast, the open coordination site of Aux I in MoaA is oriented towards the interior of the active site, potentially to aid in capturing its smaller substrate. This open coordination site is analogous to how the RS cluster ligates to the amine-nitrogen and carboxyl oxygen from the methionine present in SAM. The specific orientation of these [4Fe-4S] clusters also impacts their distance from the SAM activating cluster: this distance is 14.4 Å in CteB **(Figure 2.3c)**, whereas it is ~17.0 Å in anSME, BtrN, and MoaA (16.8, 16.9, and 17.3 Å, respectively). The more compact architecture in CteB

could facilitate the pseudo-intermolecular bond formation reaction between Cys32 and Thr37 of the CteA peptide substrate.

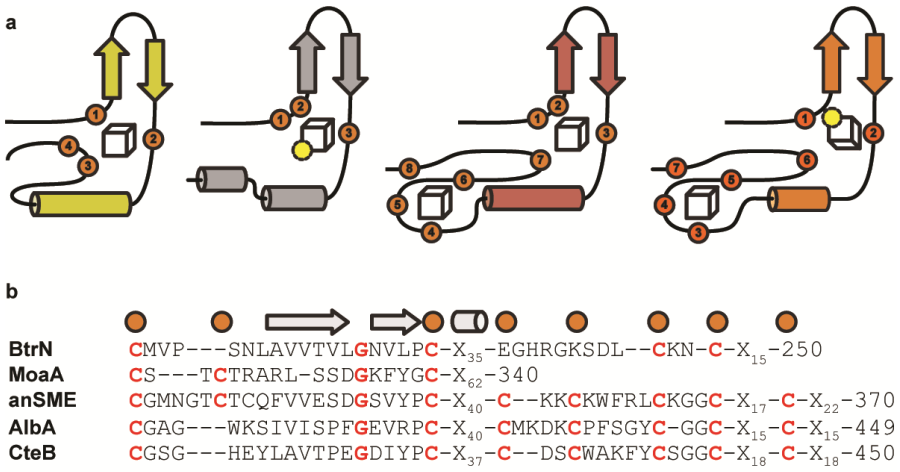
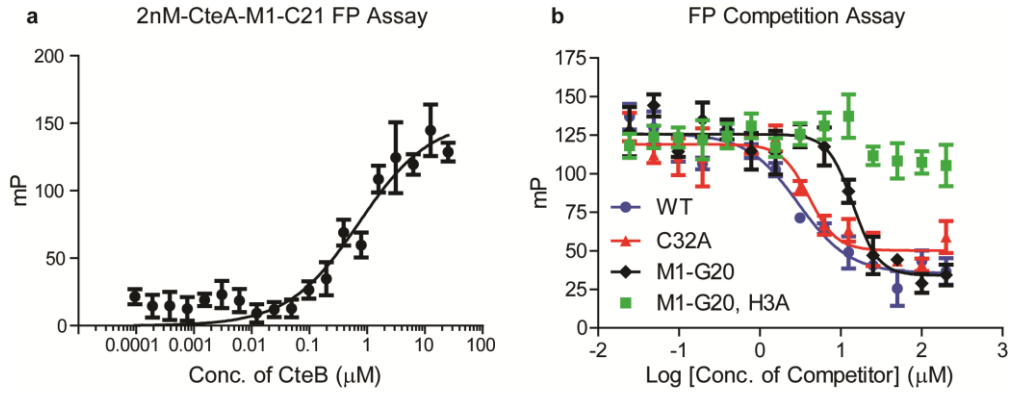


Figure 2.5. Comparison of Aux I and Aux II clusters. a) Topology diagrams of known crystallized enzymes that hold either one or both Aux I and Aux II clusters. Yellow-BtrN, gray-MoaA, red anSME, and orange-CteB. b) Sequence alignments of those domains.

Removing the RRE domain of CteB, the other domains of CteB and anSME (RS and SPASM) adopt similar structural organizations (R.M.S.D. of 2.52 Å over 300 Cα atoms) **(see Appendix Figure A.9)**, despite sharing only 20% sequence identity between the two proteins. Interestingly, the conserved Asp277 and Tyr24 active site residues of anSME are absent in CteB. In particular, Asp277 was shown to be absolutely required for anSME activity and was proposed to act as a base in the anSME reaction mechanism.[9] His363 and Tyr350 in CteB are within 8 Å of Aux I and we hypothesized that these residues could play analogous roles in the CteB reaction mechanism. However, when we prepared and tested the H363A and Y350A mutants, we observed formation of the thioether bridge on CteA **(see Appendix Figure A.2)**. These observations suggest that these residues do not act as essential bases during the reaction, unlike Asp277 in anSME. Studies are currently under way to determine which active site residues are critical for the activity of CteB within both the RS and SPASM domains.

2.5 Contributions to Binding Affinity of CteA

A series of CteA derivatives were prepared in order to assess the separate contributions of leader peptide, core, and cysteine residues to affinity of CteA for CteB. A fluorophore-labeled probe was prepared by SPPS; specifically CteA-M_1-C_{21} was synthesized with a TAMRA label on the N-terminus for use in fluorescence polarization (FP) binding assays. The leader peptide alone exhibits a 0.7 ± 0.2 μM binding affinity, in good agreement with affinities for similar leader peptide-RRE interactions.[21,57] The unlabeled peptides CteA-M_1-G_{20}, CteA-M_1-G_{20}-H3A, wildtype CteA, and CteA-C32A were used with the fluorophore-labeled peptide in competition assays. Both wild-type CteA and CteA-C32A exhibit a slightly weaker affinity, than the leader peptide alone (K_d 3.0 ±1.0 and 4.0 ±1.0, respectively). Notably, the C32A mutant did not substantially impact binding, but the H3A variant leader peptide was unable to compete off the labeled peptide (K_d >100). The fact that the H3A variant exhibits greatly reduced binding to CteB provides strong evidence that the histidine side chain interactions with the RRE domain of CteB are critical for CteA recognition by CteB. These observations also provide evidence that the peptide substrate sequence is modeled correctly within the reported structure. Taken together, these observations suggest that the leader peptide contributes significantly to binding in this system and that key interactions between the leader peptide and RRE (e.g. His3) can significantly impact binding. The results are summarized in **Figure 2.6** below.



Parameter	Value ± Std. Error
Hill Coeff.	0.7 ± 0.2
Kd	0.7 ± 0.3 μM
R-square	0.735

Construct	IC_{50} ± Std. Error (μM)
SNA-CteA-WT	3.0 ± 1.0
SNA-CteA-C32A	4.0 ± 1.0
GGSSG-CteA-M1-G20	14.0 ± 1.2
CteA-M1-G20, H3A	> 100
Construct	**K_d ± Std. Error (μM)**
SNA-CteA-WT	3.0 ± 1.0
SNA-CteA-C32A	4.0 ± 1.0
GGSSG-CteA-M1-G20	14.0 ± 1.2
CteA-M1-G20, H3A	> 100

Figure 2.6. Fluorescence Polarization binding of CteA to CteB. a) Binding curve of 2 nM of TAMRA-CteA-M1-C21 to CteB. To produce the curve, two replicates done in triplicate and analyzed by GraphPad Prism 5 (One site- Specific Binding with Hill Slope). b) Competition Assay with full length competitors (CteA (WT or C32A) and leader peptide truncates (GGSSG-CteA (M_1-G_{20}) or CteA (M_1-G_{20})-H3A). The fluorophore concentration was set at 5 nM while the protein concentration was set at 5 μM. To produce individual curves, one set of data was done in triplicate and analyzed by GraphPad Prism 5 (log (inhibitor) vs. response-Variable slope (four parameters)). K_d values were calculated using the following equation: $K_d=IC_{50}/(1+[L]/K_d,\text{labled})$ where K_d is the dissociation constant of for the unlabeled peptide, [L] is the concentration of labeled peptide (5 nM), and K_d,labeled is the dissociation constant for the labeled peptide (0.7 μM from Figure 2.6a).

2.6 Summary and Discussion

Perhaps the most important feature that comes from the CteB structure is the clear presence of an open coordination site on Aux I in the substrate free-state, which appears to be filled by a cysteine from the peptide substrate in the peptide bound structure. The coordination state of this particular predicted [4Fe-4S] cluster within sactisynthases has been the subject of debate for years. Berteau[6] and Drennan[2,9] both note number of cysteines for complete ligation of the two predicted [4Fe-4S] clusters is insufficient in the sactisynthase AlbA. Berteau postulated that the ligation state may be fulfilled by a serine or arginine as in LipA[58] and BioB[59],

respectively, whereas Drennan and co-workers hypothesized that an open site on the [4Fe-4S] cluster might be involved in substrate binding as in MoaA.[2] The structures of CteB are consistent with a mechanism in which the open coordination of Aux I in CteB is involved in substrate binding, namely through one of the six cysteines present in CteA. Substrate coordination at this open site also appears to be consistent with spectrophotometric data reported by Marahiel *et al*. for AlbA, where substrate titration into a solution of enzyme was accompanied by a shift in the UV-spectrum where the [4Fe-4S] clusters absorb (300-500 nm), which is absent in enzyme mutants that disrupt the predicted Aux I present in AlbA.[37] Although the current structure shows a terminal cysteine, Cys21 from the peptide fragment, coordinating to Aux I, we hypothesize that in the full-length, native substrate, coordination of the reacting cysteine (Cys32) would serve to orient and activate the cysteine for thioether bridge formation.

Two mechanisms have been proposed for enzymatic formation of sactionine bridges **(Figure 2.7a)**. The first mechanism, involves separate activation of the bridging partner α-carbon and the cysteine sulfur by distinct [4Fe-4S] clusters, followed by attack of the carbon centered radical on the coordinated/activated sulfur atom **(Figure 2.7a, Mechanism A)**. An alternative mechanism, in which the intermediate α-carbon radical undergoes a one-electron oxidation to the ketoimine **(Figure 2.7a, Mechanism B)**, has been proposed by Bandarian and co-workers in their work on Tte1186.[41] In this mechanism, the thioether is formed by nucleophilic attack of the cysteine sulfur on the ketoimine. Both mechanisms would be catalytic if the SAM cleavage cluster acts as an oxidant or electron acceptor via chain transfer to return to its active, reduced state. Justification for mechanism B is based on the observance of sactionine linkages with both D- and L- stereochemistry in sactipeptides like Subtilosin A. The polar mechanism would clearly allow for attack on either *re*- or *si*-face of the ketoimine; however, the

low barrier to inversion of a carbon centered radical, especially when it may proceed via the enol radical tautomer cannot realistically rule out Mechanism A. The open coordination site on Aux I neither refutes nor supports either of the proposed mechanisms. Both mechanisms can reasonably be drawn, as in **Figure 2.7a**, with a substrate Cys-ligated Aux I.

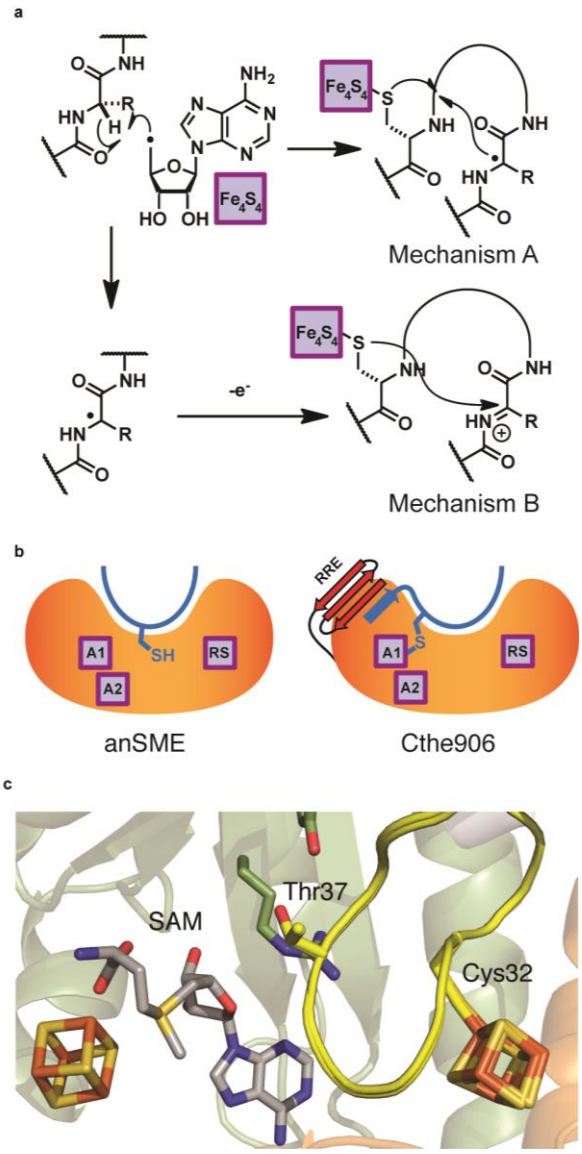


Figure 2.7. Proposed mechanisms of sactionine bridge formation. a) Mechanisms describing either separate activation of the bridging partner α-carbon and the cysteine sulfur by distinct [4Fe-4S] clusters followed by attack of the carbon centered radical on the coordinated sulfur atom (Mechanism A) or the intermediate α-carbon radical undergoes a one-electron oxidation to the ketoimine which is then subject to nucleophilic attack of the cysteine sulfur (Mechanism B). b) Proposed binding of substrates in their enzymes. c) Rosetta model of CteA-CteB complex

with Cys32 ligated to Aux I. This computational model shows possible interactions between CteA (yellow) and CteB. In the model Cys32 from CteA ligates the free coordination site on Aux I and Thr37 is placed in close proximity to where the 5'-dA radical is formed from SAM (gray)

This newly discovered free ligation site on Aux I has other potential ramifications for the mechanism of thioether bond formation. Based on coordination of Cys21 in the peptide bound structure of CteB, a site that does not make a thioether bridge in CteA *in vitro*, it seems possible that the observed coordinating cysteine could be artificial. An alternative is that the bound Cys21 in the current structure mimics the physiologically relevant cysteine in CteA, Cys32, which would be activated for crosslinking by coordinating to this [4Fe-4S] cluster. By analogy to anSME, CteA would bind the RRE with its N-terminus (residues 1-9) and likely project down into the bowl-like active site, where the reactive Cys32 sulfur may coordinate to Aux I **(Figure 2.7b)**. Substrates then make an abrupt turn and climb out of the active site, aided by a number of conserved H-bonding residues on the β5 and β6 strands of the TIM barrel. The peptide trajectory presumably places the bridging partner residue in front of the SAM binding pocket for activation, but there are no obvious pockets capable of dictating stereochemistry. The open coordination site on Aux I present in CteB could also provide either an electron sink for the radical mechanism (Mechanism A) or an oxidant and intermediate Lewis acid for the polar/ketoimine intermediate mechanism (Mechanism B). The CteB structure demonstrates that Aux I and Aux II of CteB are in sufficient proximity to act as electron transfer partners.[60] Patches of highly conserved surface residues border the RS and Aux II clusters of CteB **(Figure 2.8)**, showing possible recognition surfaces for single electron donors and acceptors, such as ferrodoxins or the flavodoxin-flavodoxin reductase system.[1]

With the help of the Kuhlman Lab at UNC, Rosetta3 macromolecular modeling suite was used to build the full length CteA substrate peptide into the active site of CteB based on the CteB+SAM+CteA-M_1-C_{21} (Pep-Bound-CteB) structure to generate a computational model **(see**

Experimental 2.7.8, Figure 2.7c and Appendix A.17 and 18). Distance constraints were used to ensure Cys32 from CteA remains in ligation proximity to the [4Fe-4S] cluster AuxI and Thr37 from CteA was immediately next to C5 of the SAM cofactor. Another single, loose distance constraint between the C-terminal Arg46 of CteA and residue Thr342 of CteB at the periphery of the binding pocket was used to keep the substrate in the active site. Two representative lowest energy structures from the model are shown in **Figure 2.7c and Appendix A.18**. These models show that Cys32 of CteA can ligate to the open coordination site of the [4Fe-4S] cluster of AuxI and that Thr37 of CteA does not clash with the SAM cofactor bound in CteB. Thus this model gives a representation of a possible interaction between CteA and CteB. Thr37 of CteA is within ~5 Å of the C5 of the SAM cofactor making it plausible for the radical formed at this position to abstract the α-hydrogen from Thr37. The model also puts Thr37 in close proximity to Cys32 (~5 Å), making it reasonable for a sactionine bridge to form, just as described and confirmed in the *in vitro* system. Another interesting development from the model is the proximity of Cys336 from CteB to Cys21, 24, and 28 from CteA. In the model, Cys21, 24, and 28 reside on the same face of the α-helix formed by CteA, orienting them in such a way as to possible interact with Cys336 from CteB. While the distances in the model are much too great to form a disulfide (average ~13 Å), it is interesting that the lowest possible energy states indicate an interaction in the region between cysteines may be favorable shedding light on how the actual enzyme interacts with its substrate. There are several other polar interactions between the substrate and enzyme in the model that could be tested to ascertain the validity of the model presented. Besides the numerous hydrogen bonding contacts between the amide backbone of CteA and residues in CteB as well as intramolecular hydrogen bonding between residues and the backbone within CteA, there were two key hydrogen bonding interactions between amino acid

side chains of CteA and CteB. First there is a polar interaction predicted between the carboxylic acid in Asp150 from CteB and the terminal amine in Lys33 from CteA. Second there is another polar interaction predicted between the guanidinium group from Arg182 of CteB and the carbonyl in the terminal amide group of Gln41. The model can be tested by canceling out and/or flipping these interactions by exchanging the interacting residues between CteA and CteB. If activity or binding is altered, it is a strong possibility that the interactions predicted by the model are real and the model is a good representation of how the substrate sits in the active site of the enzyme. Experiments are ongoing to determine if the aforementioned interactions are a true representation of substrate-enzyme interaction.

Based on sequence analysis, the open site on Aux I is predicted to be conserved in several sactisynthases, suggesting that coordination may be a common feature of substrate recognition within the sactisynthases. For example, multiple sequence alignments **(see Appendix Figures A.13 and 14)** indicate that sactisynthases from thurincin H (ThnB, which forms 4 bridges) and subtilosin A (AlbA, which forms 3 bridges) biosynthesis should have very similar architectures to CteB, but unlike CteB capable of making multiple sactionine linkages. Two other SPASM-containing enzymes, PqqE and StrB, also align well with the CteB SPASM architectures; however, the chemistry carried out by PqqE and StrB differ significantly from CteB, forming C-C bonds instead of C-S, and it is unclear how a coordination state might contribute in these enzymes. However, not all known sactisynthases align well with CteB. TrnC and TrnD, sactisynthases responsible for the biosynthesis of two component system thuricin CD, and SkfB responsible for the biosynthesis of sporulation killing factor, vary significantly in the SPASM architecture. The connection of these structural changes to the chemistry these enzymes carry out remains to be defined. It remains unclear how CteB-related sactisynthases catalyze the

formation of multiple nested thioether linkages. For example, AlbA catalyzes formation of three sactionine linkages in subtilosin A and ThnB makes four in thurincin H biosynthesis. Dynamics in the active site and a degree of substrate control could both play roles in the formation of additional thioethers linkages. Initial substrate coordination may act to "set the register" for thioether positioning in these multiply bridged systems. Along with this, the long RRE linker would presumably allow greater flexibility of the N-terminus, and enable a more diverse ensemble of approaches to the catalytic site.

A sequence similarity network (SSN) of the radical SAM protein family (PF04055) was made using the EFI-Enzyme Similarity Tool online **(Appendix A.12)**. This SSN shows the similarity and relationship between different rSAM enzymes belonging to the same family. Even though there is sequence similarity in the RS and SPASM domains **(Appendix A.13 and 14)**, the SSN shows that the known, biochemically characterized sactionine synthases are widely different from one another, with the exception of CteB and Tte1186. These differences could play a part in the number of sactionine bridges each enzyme imparts. AlbA, ThnB and TrnD, enzymes that impart three or more sactionine bridges, are away from big clusters of proteins unlike CteB and Tte1186. This could be an indication that AlbA, ThnB, and TrnD are outliers in sactionine synthases and thus so is there unique activity of more than one thioether bridge. CteB is firmly in the center of a large cluster of proteins that represents the SCIFF maturases. This could be an indication that CteB, along with its activity, is not so alone and that they belong to a large group of enzymes with the same activity, making the more promiscuous sactionine synthases the outliers. This hypothesis can only be tested once more sactionine synthases are isolated and characterized.

The RRE is a fairly new motif to be structurally characterized in RiPPs enzymes; however its prevalence can be seen in predictions and bioinformatic analyses. In Burkhart *et al.*[21], using HHpred, RRE's based on PqqD were predicted in myriad of RiPPs biosynthetic proteins including the cyclodehydratases (LynD), dehydratases (NisB), adenylases (PaaA), proteases, methyltransferases, epimerases, and of course thioether bond formation (AlbA), and carbon-carbon bond formation (PqqD/PqqE). Some of these RRE's have been structurally characterized, such as LynD, NisB, PaaA, and now CteB a sactionine synthase, but there are significant differences between the recognition of peptidyl substrates of these enzymes.[21,48-54] In LynD and NisB, the RRE binds to the N-terminal LP portion of their substrates through mostly hydrophobic interactions but in the middle of the leader peptide sequence, leaving the very N-terminus of the peptide free. CteB's RRE binds the very N-terminus of leader peptide of CteA through amide backbone hydrogen bonding, but there is a very significant polar interaction between His3 and the wing in the RRE that is not present in the other structurally characterized RRE motifs. These changes make the RRE of CteB unique among other RREs that have been structurally characterized. The RRE is important for peptide recognition, but it is only a recognition element and has no catalytic activity, this is done by the active site, the partial TIM barrel and SPASM domain in the case of CteB. It has been shown that the core peptide can be mutated very rigorously and the enzyme itself is very tolerant to those changes as long as the leader peptide is intact and can bind to the RRE.[16-20,43] This makes RiPPs enzymes very promiscuous and valuable in biosynthesis of new and natural products. CteB may also have this inherent promiscuity but further experiments and analysis are needed to fully probe this.

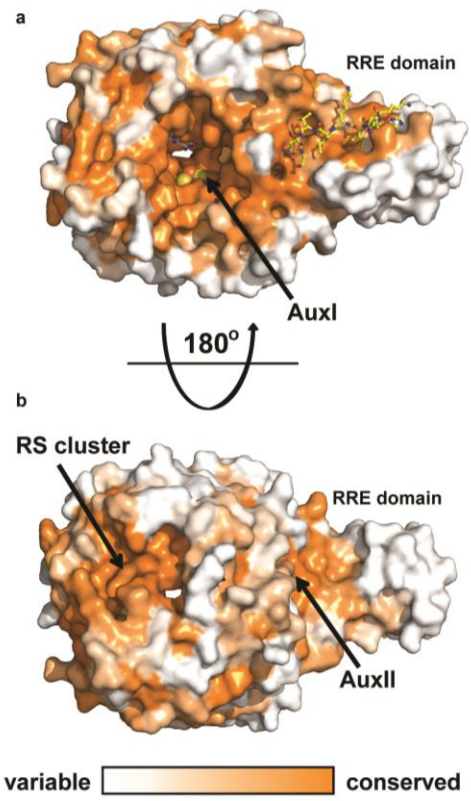


Figure 2.8. Conservation of CteB homologs. Surface map (ConSURF server) of sequence conservation based on 150 sequences with homology ranging from 35% to 90% identity. Conservation scores are based on Bayesian method. a) The highest sequence conservation can be found around the active site and peptide binding surface of the RRE domain. b) 180 ° rotation showing the bottom of CteB. A patch of highly conserved residues are found around the RS and Aux II clusters. These sites may have a role in the recognition of redox partners.

In conclusion, we have biochemically characterized a new sactionine synthase, CteB from *C. thermocellum* ATCC 27405, which installs a single sactionine bridge on the thermocellin precursor peptide. We also determined the X-ray crystal structures of CteB both in the presence and in absence of a fragment of the peptide substrate CteA, which represent the first structures of a sactionine synthase and PqqE-like enzyme. The structures reveal a conserved SAM activating domain, as well as a new SPASM domain motif displaying a single open coordination site on the internal auxiliary iron-sulfur cluster (Aux I). These structures provide valuable insight into the enzymatic mechanism of sacti-bridge formation and, by analogy, into the mechanisms of related, PqqE-like enzymes. In particular, these structures provide evidence

for the role of SPASM auxiliary clusters in direct substrate ligation and potential residue activation required to facilitate product formation. We anticipate that this structure will have utility for the continued mechanistic understanding and engineering of sactionine synthases and other PqqE-like RiPP enzymes.

2.7 Experimental

2.7.1 General Cloning and Molecular Biology Techniques

Cloning of cteA and cteB into pMCSG7

The genes *cteA* (encoding the 51 amino acid precursor peptide) and *cteB* (encoding the RS sactisynthase) were codon-optimized for expression in *E. coli* and synthesized by Integrated DNA Technologies as gene-blocks **(Table 2.4)**. The codon-optimized *cteA* and *cteB* were amplified from their respective gene-block templates by PCR using primers from **Table 2.5** and Q5® High-Fidelity DNA Polymerase with Q5 Reaction Buffer following the manufacturer's manual (NEB). The purified PCR product was phosphorylated with T4-PNK and then treated with T4-DNA Polymerase to create ligation independent cloning (LIC)-overhangs. In parallel, pMCSG7 (ampicillin resistant) was linearized with SspI and dephosphorylated with Antarctic Phosphatase and then treated with T4-DNA Polymerase to create complimentary LIC overhangs. The digested PCR products and vector were combined and allowed to anneal for 10 min at 22 ^{o}C, then were transformed into One-Shot ® Top 10 cells. A single colony we used to inoculate 5 mL of LB culture. The recombinant plasmids were purified with the GeneJET Plasmid Miniprep kit following the manual. The final constructs were verified by DNA sequencing. The cloning cteA and cteB into pMCSG7 added a hexa-histidine (His) tag to the N-terminus of both genes with a TEV protease site that allows removal of the His tag after purification of the proteins.

Generation of CteA and CteB variants

The *cteA* or *cteB* gene-blocks were used as templates to produce mutations with primers from **Table 2.5**. PCR was performed with Q5® High-Fidelity DNA Polymerase following the manufacturer's manual. Primer 1 and corresponding reverse mutant primer **(Table 2.5)** were used to create Piece I for CteA. Primer 2 and corresponding forward mutant primer **(Table 2.5)** were used to create Piece II for CteA. Primer 3 and corresponding reverse mutant primer **(Table 2.5)** were used to create Piece I for CteB. Primer 4 and corresponding forward mutant primer **(Table 2.5)** were used to create Piece II for CteB. The two PCR pieces were purified and kept in water. In parallel, pMCSG7 was prepared as above. Piece I and II were mixed with linearized pMCSG7, then ligated using Gibson Assembly® Master Mix according to the manufacturer's protocol (NEB). Then, 3 μL of reaction mixture was then transformed into 50 μL of One-Shot ® Top 10 cells. A single colony we used to inoculate 5 mL of LB culture. The plasmid was purified as above.

Cloning cteA and cteB into Duet plasmid

His-cteA was cloned into MCS1 (NcoI and HindIII) from the pMCSG7-*Cthe0907* using primers 5 and 6 **(Table 2.5)** while *cteB* (no His-tag) was cloned into MCS2 (NdeI and XhoI) from the codon-optimized gene-block using primers 7 and 8 **(Table 2.5)**. PCR was conducted with Q5® High-Fidelity DNA Polymerase following the manufacturer's manual. The purified PCR products and pETDuet plasmid were digested using the corresponding restriction enzymes for each site **(Table 2.5)**. The purified digested plasmid was treated with Antarctic Phosphatase and then combined with the purified digested PCR product in the presence of T4 ligase. The ligation was allowed to go overnight at 16 ^{o}C. After ligation, T4 ligase was heat inactivated at 65 ^{o}C for 10 min before the ligation reaction was transformed into One-Shot ® Top 10 cells.

The resulting plasmids were purified as above. *cteB* was cloned into MCS2 first and once the resulting plasmid was sequencing confirmed, *His-cteA* was cloned into MCS1.

Expression of His-CteA peptides

His-CteA and its variant plasmids were transformed into BL-21 (DE3) or BL-21 (DE) cells harboring the pPH151 corrector plasmid by electroporation. The cells stocks were made electrocompetent according to standard molecular biology protocols found in Green *et al*.[61] The electroporation was carried out in a 0.1cm cuvette, at 1.8 kV, 200 Ω, and 20 μFD. His-CteA precursor peptide (in pETDuet-*His-CteA*-1, *CteB*-2) was heterologously expressed in *E. coli* (pPH-151/BL21 DE3) while His-CteA (in pMCSG7) variants were heterologously expressed in E. coli (BL21 DE3) cells. LB media was supplemented with ampicillin (100 μg/mL) with or without chloramphenicol (34 μg/mL). A 5 mL LB overnight culture was used to inoculate a 1 L LB culture. Cultures are grown at 37 ^{o}C and 200 rpm to an OD_{600}~0.6-0.7, at which point IPTG was added to a final concentration of 0.5 mM and the culture was grown at 18 ^{o}C, 200 rpm for 22-24 hours.

Expression of His-CteB proteins

His-CteB and its mutant plasmids were transformed into BL-21 (DE) cells harboring the pPH151 corrector plasmid by electroporation. His-CteB enzyme (in pMCSG7-*CteB*) was heterologously expressed in *E. coli* (pPH-151/BL21 DE3) cells using 1 L of auto-induction media, adapted from Studier.[62] Auto-induction media was supplemented with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL). A 5 mL overnight culture of ZYP-0.8G was used to inoculate 1 L of ZYP-5052. Cultures were grown at 37 ^{o}C and 200 rpm to an OD_{600} ~ 0.6-0.8, at which point the culture was cooled to 30 ^{o}C for 30 minutes. After cooling, cysteine was added to a final concentration of 300 μM. The culture was allowed to grow at 30 ^{o}C, 200 rpm for 18-24

hours before harvest.

Table 2.4. Gene-blocks ordered from Integrated DNA Technologies (IDT), codon optimized used in CteB study

CteA	5'-ATG AAG CAC ATT AAA ATT TTG AAC GGG TCA ACA CTG AAA GAC AGC CTG AAA AAA GGT GGG TGT GGG GAA TGT CAA ACC TCT TGC CAG TCA GCT TGC AAG ACC TCA TGT ACC GTT GCT AAT CAG TCA TGC GAA AAG CGT TAA -3'
CteB	5'-ATG GCG ATG ATC CAC AAA TTC TCG ATG ATG GGC ACA AAC ATT GTG GTG GAC GTA AAT TCA GGT GCT GTA CAC GTG GTT GAT GAT ATC AGT TTT GAT ATC CTT GAT TAT TAC AAG AAT TTT ACC GCG GGT GAG ATC AAG AAC AAG TTG GCG CAT AAG TAC AAT GCC GAC GAA ATC GAC GAA GCG TTA CGC GAA ATC GAG TCA TTA GAA GCT GAG GGC CTG TTA TTT TCA GAG GAC CCG TAT AAA GAA TAC GTA TCA TCT ATG GAC CGC AAG TCC GTC GTA AAA GCG TTG TGT CTT CAT ATC TCA CAC GAC TGT AAT CTG CGC TGC AAA TAT TGT TTT GCT TCG ACA GGA AAT TTC GGG GGC CAG CGT AAT ATG ATG TCC CTG GAG GTT GGA AAG AAG GCT ATT GAC TTC CTT ATT TCG GAA TCA GGT AAC CGC AAG AAT CTT GAG ATC GAT TTC TTT GGG GGC GAG CCC ATG ATG AAC TTC GAC GTC GTA AAG GGT ATT ATT GAG TAT GCC CGT CAG AAA GAG AAG GAG CAT AAT AAA AAC TTT CGC TTT ACA TTG ACT ACT AAT GGT CTG CTT CTG AAT GAT GAA AAT ATT AAG TAC ATT AAC GAA AAC ATG CAG AAT ATC GTT TTA TCG ATC GAC GGT CGC AAG GAA GTC AAT GAC CGT ATG CGC ATT CGC ATT GAC GGA TCC GGT TGT TAT GAT GAC ATT CTG CCC AAA TTC AAA TAT GTA GCC GAA AGC CGC AAT CAA GAC AAT TAC TAT GTT CGT GGC ACG TTC ACA CGC GAG AAT ATG GAC TTT TCA AAT GAC GTG TTA CAC TTG GCC GAC GAA GGG TTC CGT CAA ATT AGC GTT GAA CCG GTG GTT GCT GCT AAA GAC TCT GGT TAC GAC CTT CGT GAA GAA GAT CTG CCT CGT CTT TTT GAG GAA TAT GAA AAG CTG GCG TAC GAG TAC GTG AAA CGT CGT AAG GAG GGA AAT TGG TTT AAT TTC TTC CAC TTC ATG ATT GAC TTA ACA CAA GGT CCA TGT ATT GTA AAG CGC CTT ACC GGA TGT GGT AGC GGA CAC GAA TAT TTG GCC GTC ACG CCT GAA GGG GAT ATT TAC CCA TGC CAC CAA TTC GTA GGG AAT GAG AAG TTC AAG ATG GGC AAT GTA AAG GAG GGC GTC CTT AAC CGC GAT ATC CAA AAC TAC TTC AAA AAC AGC AAT GTA TAC ACT AAG AAG GAA TGT GAT TCC TGT TGG GCT AAA TTC TAT TGC AGT GGA GGC TGT GCA GCG AAC TCC TAC AAT TTC CAC AAA GAC ATT AAT ACG GTG TAC AAA GTT GGT TGT GAA TTG GAA AAG AAA CGT GTG GAG TGC GCT TTA TGG ATC AAG GCG CAA GAG ATG TAA -3'

Table 2.5. Plasmids, sites, and primers (ordered from Eton Bioscience, Inc.) used in CteB study

LIC-plasmid	pMCSG7	SspI
MCS1-Duet	pETDuet	NcoI, HindIII
MCS2-Duet	pETDuet	NdeI, XhoI
Primer 1	pMCSG7-CteA-F	5'-TAC TTC CAA TCC AAT GCG ATG AAG CAC ATT AAA ATT TTG AAC GG- 3'
Primer 2	pMCSG7-CteA-R	5'-TTA TCC ACT TCC AAT GCG CTA TTA ACG CTT TTC GCA TGA CTG ATT AG -3'
Primer 3	pMCSG7-CteB-F	5'- TAC TTC CAA TCC AAT GCG ATG GCG ATG ATC CAC AAA -3'
Primer 4	pMCSG7-CteB-R	5'- TTA TCC ACT TCC AAT GCG CTA TTA CAT CTC TTG CGC CTT GAT CCA T -3'
Primer 5	His-CteA-into Duet-MCS1-F	5'-GAT CGA TCC CAT GGG CAT GCA CCA TCA TCA TCA TCA TTC TT -3'

Primer 6	His-CteA-into Duet-MCS1-R	5'-GAT CGA TCA AGC TTT TAA CGC TTT TCG CAT GAC -3'
Primer 7	CteB-into Duet-MCS2-F	5'-GAT CGA TCC ATA TGG CGA TGA TCC ACA AA -3'
Primer 8	CteB-into Duet-MCS2-R	5'-GAC TGA TCC TCG AGT TAC ATC TCT TGC GC -3'
Primer 9	CteB-Y350A-F	5'-GTG GTA GCG GAC ACG AAG CGT TGG CCG TCA CGC CTG-3'
Primer 10	CteA-Y350A-R	5'-CAG GCG TGA CGG CCA ACG CTT CGT GTC CGC TAC CAC -3'
Primer 11	CteB-H363A-F	5'-GGA TAT TTA CCC ATG CGC GCA ATT CGT AGG GAA TG-3'
Primer 12	CteB-H363A-R	5'-CAT TCC CTA CGA ATT GCG CGC ATG GGT AAA TAT CC -3'
Primer 13	CteA-C32A-F	5'-CCT CTT GCC AGT CAG CTG CTA AGA CCT CAT GTA CCG -3'
Primer 14	CteA-C32A-R	5'-CGG TAC ATG AGG TCT TAG CAG CTG ACT GGC AAG AGG -3'

2.7.2 Purification of Substrates and Enzymes

Purification of His-CteA (WT and variants) peptide and cleavage to SNA-CteA (WT and variants)

Purification procedures were modified from Li *et al*.[63] A 5 mL cell pellet was suspended in 30 mL of Start Buffer (20 mM $Na_2(PO_4)$ pH 7.5, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) supplemented with 0.5 mL of 25 mg/mL T4 lysozyme and 0.5 mL of 150 mM PMSF and then incubated on ice for 10 minutes. The cell pellet was then sonicated. The cell debris was pelleted by centrifuging the lysate at 15,000 rpm at 4 ^{o}C for 10 minutes. The supernatant was then discarded and the pellet was washed and resuspended in 30 mL of IB Buffer (20 mM $Na_2(PO_4)$ pH 7.5, 6 M Guanidinium-HCl, 500 mM NaCl, 50 mM imidazole) by using a spatula to break up the pellet coupled with extensive vortexing. The debris was again pelleted and the supernatant collected and filtered through a 0.45 μm syringe filter. The flow through from the filter was then passed over a Ni^{2+} IMAC column (HISTrap^{TM} HP 5mL GE Healthcare) coupled

to an FPLC (NGC-Quest-10 Bio-Rad). The Ni^{2+} IMAC column was washed with 6 column volumes (CV) of IB Buffer. The peptide was eluted with a gradient of 0-100% of elution buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 500 mM imidazole) over 10 CV. The peptide eluted between 15-25% elution buffer. Fractions containing the peptide were combined and dialyzed against 25 mM Tris-HCl, pH 7.0 and 150 mM NaCl in a 2000 MWCO cassette from Thermo Fischer. The buffer exchange was repeated three times, while being maintained at 4 ^{o}C. The peptide solution was then collected and tobacco etch virus (TEV) protease was added at approximately a 1:15 ratio of TEV to peptide. The protease reaction was incubated overnight at 4 ^{o}C. The cleaved CteA peptide was further purified by preparative HPLC. Preparative HPLC was performed on a Shimadzu UFLC CBM-20A with a dual channel wavelength detector at 220 nm and 280 nm with a Luna® 10 μm, 100 Å, 250 x 30 mm) AXIA™ (Phenomenex®) semipreparatory column. Purification was carried out with a two solvent system (solvent A = 0.1% trifluoroacetic acid (TFA) in water; solvent B = 0.1% TFA in acetonitrile) using gradient of 30-60% B over 20 min at a flow rate of 15 mL/min. The peptide eluted from the column between 40-45% solvent B. The peptide product after these steps is CteA with three additional amino acids (SNA) at the N-terminus, and is hereby denoted as CteA. The fractions containing the CteA (or its variants) were pooled and partially concentrated with a rotary evaporator, followed by flash freezing and lyophilization to obtain the purified solid product. The yield of peptide was ~ 1mg of CteA per 1 L of culture. The peptide was then dissolved in dimethyl sulfoxide (DMSO) to a final concentration of 1.25 mM (based on mass).

Purification of His-CteB protein

A 10 mL cell pellet was suspended in 30 mL of Buffer A (50 mM HEPES, pH 7.5, 150 mM NaCl, 50 mM imidazole) supplemented with 0.5 mL of 25 mg/mL T4 lysozyme, 0.5 mL of

150 mM PMSF, 80 µl of DNaseI (1u/µl), one tablet of Pierce™ Protease Inhibitor Tablets (EDTA Free from Thermo Scientific), and then incubated on ice for 10 minutes. The cell pellet was then sonicated. The lysate was clarified by centrifugation at 15,000 rpm at 4 ^{o}C for 15 minutes. The supernatant was collected and filtered through a 0.45 µm syringe filter. The flow through from the filter was then passed over a Ni^{2+} IMAC column (HISTrapTM HP 5mL GE Healthcare) coupled to an FPLC (NGC-Quest-10 Bio-Rad). The Ni^{2+} IMAC column was washed with 5 column volumes (CV) of Buffer A. CteB was eluted from the Ni^{2+} IMAC column with a gradient of 0-100% of elution buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 500 mM imidazole) over 10 CV. The protein eluted between 20-35% elution buffer. Fractions containing CteB were pooled and concentrated in a 30,000 MWCO filter in an Amicon stirred cell. The concentrated protein was then exchanged into reconstitution (RC) buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 10 mM dithiothreitol (DTT)) by passage over a Sephadex™ PD-10 column (GE Healthcare) equilibrated in RC bufer. The protein concentration was estimated by A_{280} using an extinction coefficient of 0.98 mg•mL^{-1}•AU^{-1}.

2.7.3 *In vitro* Reconsitution. Assays and Characterization of Products

All the following were done in strictly anaerobic conditions in a Coy anaerobic chamber from a procedure Modified from Flühe *et al.*[37]

In vitro reconstitution of CteB or variants

All solutions (1 M DTT, 100 mM ammonium ferric citrate, and 100 mM lithium sulfide) were made fresh in the anaerobic chamber with degassed RC buffer (described above). The protein solution was placed in the anaerobic chamber and passively degassed at RT for approximately one hour. Then, 100 equiv. of DTT were added to the protein solution and incubated on ice for an additional hour. After, 10 equiv. of ammonium ferric citrate were added

dropwise to the protein solution with gentile mixing. The solution was then allowed to incubate on ice for five minutes. During this time the solution turned reddish to dark red in color. Then 10 equiv. of lithium sulfide were added dropwise to the protein solution and the reaction was further incubated overnight (16-18 hours) on ice. The final color of the reaction mixture was a dark brown solution. The excess iron and sulfide was removed by buffer exchange with a PD-10 gel filtration column equilibrated in RC buffer. This protein is denoted as RC CteB.

SAM cleavage assays with CteB or its variants

All assays were carried with freshly RC CteB and its variants under anaerobic conditions. Solutions of S-adenosylmethionine (SAM) and dithionite (DT) were made fresh in the anaerobic chamber with RC Buffer. RC CteB was added to a final concentration of 20 μM in H_2O and was incubated with 300 μM DT for five minutes at RT, followed by the addition of 300 μM SAM in a total reaction volume of 47.5 μL. The reaction was allowed to proceed at RT for five hours. The reaction was then removed from the anaerobic chamber and quenched by adding 2.5 μL of neat formic acid (final concentration 5% v/v). The samples were then centrifuged at 15,000 rpm at 4 ^{o}C for 15 minutes to remove precipitated protein. The supernatant was collected and analyzed by LC-MS Method 1.

Peptide modification assays

All assays were carried with freshly prepared reagents and RC CteB as above. CteA (in DMSO) was allowed to passively degas in anaerobic chamber overnight at RT. CteB (20 μM) was added to a reaction mixture consisting of 20 μM CteA (or variants), and 1mM DT in a total volume of 49 μL. The reaction was incubated for five minutes at RT, followed by the addition of 1 mM SAM. The reaction (16 % DMSO, final volume 50 μL) was allowed to proceed for five hours, after which the reaction was removed from the anaerobic chamber. The reaction products

then worked up with method 1 or method 2.

Workup method 1: If the free cysteines of CteA were not to be modified, tris(2-carboxyethyl)phosphine (TCEP) was added to the reaction at a concentration of 10 mM and subsequently incubated at 37 ^{0}C for 10 minutes. Methanol was added in a 1:1 ratio (volume) to precipitate the protein. The mixture was centrifuged at 15,000 rpm at 4 ^{o}C for 15 minutes to pellet the enzyme. The supernatant was collected and analyzed by the LC-MS Method 2.

Workup method 2: If the free cysteines of CteA were to be modified with n-ethylmaleimide (NEM), as above, TCEP was added to a final concentration of 10 mM and the reaction was incubated at 37 °C for 10 minutes. Then 20 mM NEM (dissolved in ethanol) was added and the reaction was further incubated at 37 ^{o}C for 30 minutes. Methanol was added in a 1:1 ratio (volume) to precipitate the protein. The mixture was centrifuged at 15,000 rpm at 4 ^{0}C for 15 minutes to pellet the enzyme. The supernatant was collected and analyzed by the LC-MS Method 2. This procedure was modified from Thibodeaux *et al*.[64]

LC-MS Method 1

LC-MS analysis was performed on an Agilent 1200 HPLC system coupled to an Agilent 6520 Accurate-Mass Q-TOF spectrometer with an electrospray ionization (ESI) source operated in positive ion mode. The products were separated with a Phenomenex® Kinetex C_{18} column (2.1 mm x 50 mm, 100 Å, 2.6 μm) with the gradient program described below. Solvent A consisted of H_2O with 0.1 % formic acid (FA) and solvent B consisted of acetonitrile with 0.1% FA. Analytes from the column were sent to the MS and spectra were acquired in centroid mode using a gas temperature of 350 ^{o}C and a fragmentor voltage of 70 V.

Flow Rate	0.5 ml/min
Post time	2 min
Time (min)	**%B**
0.00	1
10.00	1
20.00	100
22.00	100
22.01	1

LC-MS Method 2

LC-MS analysis was performed as above except spectra were acquired in profile mode using a gas temperature of 350 ^{o}C and a fragmentor voltage of 250 V. The gradient program is described below.

Flow Rate	0.5 ml/min
Post time	4 min
Time (min)	**%B**
0.00	2
2.00	2
15.00	100
16.00	100
18.01	2

When running tandem MS/MS, the +7 charge state (Z) of modified CteA product was used as the target ion. A specific retention time of 7 minutes +/- 0.5 minutes and used for the MS/MS and an isolation width of 1.3 m/Z. Collision energies of 25-35 eV were used to obtain MS/MS spectra.

2.7.4 Fluoresence Polarization Assays

Synthesis of TAMRA-CteA-M_1-C_{21}, GGSSG-CteA-M_1-G_{20}, CteA-M_1-G_{20} H3A

All syntheses were carried out by microwave assisted solid-phase peptide synthesis (SPPS). ChemMatrix solid support (0.47 mmol/g) on a 0.047 mmol scale was used. The solid support was initially swollen in DMF (1.5 mL) for 20 min at 70 °C. Fmoc-Amino Acids-OH (0.5 M in DMF) or 5,6 TAMRA-OH (0.5 M in DMF, VWR) at 7.0 - 10.0 equiv were coupled with HATU (0.2 M in DMF) at 6.86 equiv., and DIEA (0.2 M in DMF) at 14.0 equiv. The reagents were added to the swollen resin in the above. The resulting suspension was heated under microwave irradiation for 5 min at 75 °C. The reaction vessel is then drained and resin is thoroughly washed with DMF four times. Removal of the Fmoc protecting group was accomplished after amino acid coupling using excess 20% piperidine. 20% piperidine was added to the reaction vessel and allowed to incubate at RT for 3 min with constant stirring. The reaction vessel was then drained, washed with DMF and excess 20% piperidine was again added and the reaction was incubated for another 10 min at RT. The reaction vessel was then drained and the resin thoroughly washed with DMF four times. After washing the resin with DCM, the resin was dried and cleaved using the standard cleavage cocktail (TFA/TIPS/H2O, 95:2.5:2.5) to yield the fully deprotected peptides. The peptides were concentrated by precipitation with cold diethyl ether. Preparative HPLC was performed as described above for purification of the CteA peptide. The TAMRA-CteA-M_1-C_{21} eluted from the column between 40-45% B, while GGSSG-CteA-

M_1-G_{20} and CteA-M_1-G_{20} H3A eluted between 30-35%B. Relevant fractions were collected and partially concentrated in a rotary evaporator, followed by flash freezing and lyophilization to obtain the purified solid product. The identities were confirmed by LC-MS by the method described above.

FP assay set-up and measurement

Polarization measurements were performed on a LJL Biosystems Acquest Plate Reader using Greiner low volume black-384 well plates. Excitation was set with the Rho/TAMRA at 530/25 nm and emission was set with the Rho/TAMRA at 580/10 nm. For every measurement, there is one read per well, with an integration time of 100 ms with the lamp source continuous and the Z-height set for the bottom of the well. All mixtures were prepared in triplicate in the Coy anaerobic chamber to ensure that His-CteB was kept in its active state. All data analyses were done in GraphPad Prism 5. The affinity of TAMRA-CteA-M_1-C_{21} peptide binding to CteB was conducted by adding 2 nM of the peptide, while varying the concentration of the CteB from 50 μM to 0.0954 nM by halving the concentration twenty times for twenty separate readings. The peptide and protein were mixed together in RC Buffer containing 0.005% Tween 20 and 1% DMSO. All reagents were passively degassed for at least one hour in the Coy anaerobic chamber. To carry out the assay, first the TAMRA-CteA-M_1-C_{21} peptide was added to the wells, then the varying concentration His-CteB protein solutions. The plate was then sealed with an adhesive cover and allowed to incubate at room temperature for 30 minutes. After incubation, the sealed plate was removed from the anaerobic chamber. The plate was briefly spun in the centrifuge to remove any bubbles. The cover was then removed and readings were taken within five minutes to limit the amount of oxygen that entered the sample. To measure the ability of CteA (or variants), GGSSG-CteA-M_1-$G_{20,}$ or CteA-M_1-G_{20} H3A to compete with the binding of

the TAMRA-CteA-M_1-C_{21} peptide to CteB, 5 nM of the TAMRA-CteA-M_1-C_{21} peptide was first mixed with 5 μM of the His-CteB in RC buffer and incubated for 30 min at RT. The different concentrations of each peptide was added to the plate (peptide solutions varied from 200 μM to 24.4 nM), followed addition of His-CteB/ TAMRA-CteA-M_1-C_{21} protein solution. The plate was then sealed with an adhesive cover and allowed to incubate at room temperature for 45 minutes. After incubation, the plate was removed from the anaerobic chamber. The plate was briefly spun in the centrifuge to remove any bubbles. The cover was then removed and readings were taken within five minutes to limit the amount of oxygen that entered the sample.

2.7.5 Oxidation of CteB with Glutathione

CteB (200 μM) in 200 μL was passed over a Superdex 200 Increase 10/300 GL column with an ÄKTA express FPLC system (GE Life Sciences) housed in a MBraun anaerobic chamber. The column was equilibrated in running buffer consisting of 25 mM HEPES, pH 7.5, 300 mM KCl, 10% glycerol. CteB was then collected and concentrated with vivaspin 20 concentrator (Sartorius Stedium Biotech). CteB (35 μM) was mixed 50 mM Tris, pH 80, 200 KCl, and a 10 mM series of GSH:GSSG ratios ranging from 95:5 to 2:98. This yields a solution redox potential of – 377 mV to – 223 mV. The mixture was incubated at RT for 30 min, then 100 mM iodoactemide was added and the reaction was incubated for another 10 min. The reactions were quenched by adding 2 X non-reducing SDS-PAGE loading buffer (SDS-PAGE loading buffer without BME or DTT). The samples were separated on a Bio-Rad AnykD™ Criterion™ TGX™ Precast Midi Protein Gel.

2.7.6 Size Exclusion Chromatography

Determination of oligomeric state of CteB

All size exclusion experiments were conducted with an ÄKTA express FPLC system (GE Life Sciences) coupled to a Superdex 200 Increase 10/300 GL column housed in a MBraun anaerobic chamber. The column was equilibrated in running buffer consisting of 25 mM HEPES, pH 7.5, 300 mM KCl, 10% glycerol, and 5 mM DTT. A 100 µl injection of Bio Rad Molecular Weight Standards mix (cat no. 151-1901) were used to create a standard curve. TEV cleaved CteB (100 µM) was incubated with 500 µM SAM and in the presence or absence of *CteA-M_1-C_{21}* (150 µM) at a final volume of 110 µL. The complete mixture was injected and separated with a flow rate of 0.5 mL • min^{-1}.

Determination of intermolecular disulfide state of CteB+SAM+CteA-M_1-C_{21} crystals

To determine the disulfide state of crystals of CteB+SAM+CteA-M_1-C_{21}, the same protocol as above was followed, except running buffer did not include 5 mM DTT. Approximately 50 crystals of CteB+SAM+-M_1-C_{21} were looped from 1 µL crystallization drops and dissolved in 100 µL of 10 mM HEPES, pH 7.5. This solution was analyzed as above.

2.7.7 Crystallography Methods

Preparation of CteB for crystallography

The pMCSG-7 plasmid containing His-CteB Wt was transformed into BL-21(DE3) cells that already harbored the pPH151 plasmid. The transformants were selected on an LB/agar plate containing 100 µg•mL^{-1} carbenicillin (a substitute for ampicillin) and 34 µg•mL^{-1} chloramphenicol. A single colony was used to inoculate 20 mL of LB overnight culture containing the above antibiotics. The overnight culture was used to inoculate 2 L of Studier's auto induction media (ZYP-5052 supplemented with 200 µM $FeCl_3$) housed in a 2 L PYREX®

media bottle. The culture was grown at 37 °C in a water bath with constant aeration using a sparging stone attached to a pressurized, 0.22 μm filtered air source. After 5 hr, aeration was stopped and the culture was placed in an ice bath for 1 hr. The culture was returned to a 22 °C water bath and light aeration was resumed. After 5 min, cysteine was added to a final concentration of 600 μM. The culture was grown at 22 °C for ~ 20 hr before being harvest by centrifugation at 10,000 x g. Cell pellets were flash frozen and stored in LN_2 until purification. All subsequent steps were carried out in an MBraun anaerobic chamber maintained at < 0.1 ppm oxygen (MBraun, Stratham, NH). Plastics were brought into the chamber and allowed to sit for two weeks before use. All solvents and buffer stocks were degassed by sparging with argon gas for 4 hr before being taken into the chamber. To purify His-CteB, ~ 30 grams of cell paste were resuspended in 30 mL of lysis buffer containing 50 mM HEPES, pH 7.5, 300 mM KCl, 4 mM imidazole, 10 mM 2-mercaptoethanol (BME), 10% glycerol, and 1 % Triton-X305. The resuspension was subjected to 50 rounds of sonic disruption (80% output, 3 s pulse on, 12 s pulse of) at 4 °C. The lysate was cleared by centrifugation at 4 °C for 1 hour at 15,000 × g. The supernatant was loaded onto a 5 mL fast-flow HisTrap™ column (GE Healthcare Life Sciences) equilibrated in lysis buffer lacking Triton-X305 with an ÄKTA express FPLC system. The column was washed with 10 CV of lysis buffer before elution with 5 mL of buffer containing 50 mM HEPES, pH 7.5, 300 mM KCl, 300 mM imidazole, 10 mM BME, and 10% glycerol. The protein fractions were immediately passed over a HiPrep 16/60 Sephacryl S-200 HR column equilibrated in 20 mM HEPES, pH 7.5, 300 mM KCl, 5 mM DTT, and 10% glycerol. The brown fractions were pooled and RC as previously described. The RC His-CteB was concentrated to 1 mL with a vivaspin 20 concentrator (Sartorius Stedium Biotech). The protein was again passed over a HiPrep 16/60 Sephacryl S-200 HR column as above and the fraction

corresponding to RC CteB were pooled and concentrated to 100 mg•L^{-1}. The protein concentration was estimated by A_{280} using an extinction coefficient of 0.98 mg•mL^{-1}•AU^{-1}.

Structure determination of CteB+SAM+CteA-M_1-C_{21} peptide

RC His-CteB (100 mg) was first treated with 0.1 mg (1:1000 ratio) of MHT238Δ TEV[65] at 12 °C for ~15 hr. The protein was then passed over a 5 mL HisTrap™ HP IMAC column equilibrated in 20 mM HEPES, pH 7.5, 300 mM KCl, 5 mM DTT, and 10% glycerol to remove the cleaved His-tag and the TEV protease. The flow-through was collected and concentrated to ~ 100 mg•mL^{-1} with vivaspin 20 concentrator. Diffraction quality crystals of CteB+SAM+CteA-M_1-C_{21} peptide were obtained by sitting-drop vapor diffusion at 20 °C in an anaerobic chamber maintained at < 0.1 ppm oxygen (MBraun, Stratham, NH) by mixing 0.5 µL protein solution (10 mg•mL^{-1} of CteB in 10mM HEPES, pH 7.5, 1 mM SAM, 0.5 mM CteA-M_1-C_{21} peptide, (MKHIKILNGSTCKDSLKKGGC) with 0.5 µL precipitant (0.1 M Tris-HCl, pH 7.0, 0.2 M calcium acetate, 20% polyethylene glycol 3,000) equilibrated against a well solution of 0.5 M LiCl. Bar-shaped crystals would appear after two to five days and grow to dimensions of ~ 50 x 50 x 200-400 µm. Crystals were removed from the original drop and soaked for 1 min in mother liquor that contained 1 mM SAM and 1 mM CteA-M_1-C_{21} peptide, mounted on nylon loops and flash-cooled in liquid nitrogen, inside the anaerobic chamber, and stored in liquid nitrogen prior to data collection. All diffraction data were integrated and scaled using the HKL3000 suite.[66] A dataset was collected at 1.3776 Å to exploit the intrinsic iron-sulfur clusters of CteB. Phases were determined by SAD with autoSHARP[67], and an initial poly-alanine model was built with ARP/wARP.[68] This model was subjected to subsequent rounds of automated model building performed by AutoBuild[69], interspersed with manual model building and refinement against a native dataset collected at 1.0333 Å X-ray wavelength using Coot[70], phenix.refine[69], and

Refmac5.[71] All figures were produced using PyMOL (Shrodinger, LLC). The final model consists of residues 1 to 449 (450) of CteB, 12 iron ions, 12 sulfide ions, 2 calcium ions, 1 SAM, and 52 water molecules. Residues 115-121 are missing in a disordered loop immediately following the radical SAM cluster binding motif, while residues 334-336 are missing from a flexible loop region. Simulated annealing composite omit maps were used to verify the final model. Data collection and refinement statistics are shown in **Appendix A Table A.1**.

Structure determination of CteB+SAM

To obtain crystals of the CteB in the absence of CteA-M_1-C_{21} peptide, a solution containing 10 mg/mL of the above TEV-cleaved, reconstituted CteB was mixed with 10 mM HEPES pH 7.5, 1 mM AdoMet and was subsequently incubated overnight at room temperature in the anaerobic chamber. Diffraction quality crystals of CteB+SAM were obtained by sitting-drop vapor diffusion, as above, by mixing 0.5 µL protein solution (10 mg•mL^{-1} of CteB in 10mM HEPES, pH 7.5, 1 mM SAM) with 0.5 µL precipitant (0.1 M Tris-HCl, pH 8.5, 0.2 M calcium chloride, 25% polyethylene glycol 4,000) equilibrated against a solution of 0.5 M LiCl. Bar-shaped crystals would appear after thirty days and grow to dimensions of ~ 10 x 10 x 75 µm. Crystals were removed from the original drop and soaked for 1 min in mother liquor that contained 1 mM SAM, mounted on nylon loops and flash-cooled in liquid nitrogen, inside the anaerobic chamber, and stored in liquid nitrogen prior to data collection. The CteB+SAM+CteA-M_1-C_{21} peptide without ligands was used as a model to phase a dataset collected at 1.0333 Å using isomorphous replacement with subsequent rounds of automated model building performed by AutoBuild[69], interspersed with manual model building and refinement using Coot[70], phenix.refine[69], and Refmac5.[71] The final model consists of two monomers in the asymmetric unit. Chain A contains residues 1-77, 91-115, 121-330, and 341-448 (of 450), while Chain B

contains residues 1-79, 90-114, 122-329, and 342-448 (of 450). The model also contains 24 iron ions, 24 sulfide ions, 2 calcium ions, 2 SAM molecules, and 10 water molecules. Data collection and refinement statistics are shown in **Appendix A Table A.1**.

2.7.8 Structual Modeling of the CteB Catalytic Site

Several modules were used in the Rosetta3 macromolecular modeling suite to build the CteA substrate peptide into the active site of CteB.[72-74] The input structure for the modeling was CteB+SAM+CteA-M_1-C_{21}, which lacked electron density for most of the CteA except for residues 1-9 and two amino acids including a Gly-Cys dipeptide in proximity to the FeS cluster AuxI. Firstly, the input structure was refined and ligated the Cys amino acid present in the structure to AuxI Fe2 imposing AtomPair distance constraints to be under 2 Å between the Fe2 and SG atoms. From this point onwards the ligated Cys is modeled as Cys32. At the next stage RosettaRemodel was used to build the missing section of the peptide covering residues 10-31. Kinematic closure (KIC) using fragments was run for 100 build cycles followed by 5 outer containing 200 inner cycles each of cyclic coordinate descent (CCD) refinement. The lowest energy model was taken for the next step of modeling to build the remaining missing amino acids from 33-46. The missing amino acids were initially constructed using pymol and then modeled using the FloppyTail protocol allowing flexibility from residue 29 to the C-terminus. AtomPair distance constraints were used to ensure Cys32 remains in ligation proximity to FeS (AuxI), to position Thr37 immediately next to C5 of the SAM cofactor and finally a single very loose distance constraint between the C-terminal Arg46 and residue Thr342 at the periphery of the binding pocket. Finally, three representative lowest energy structures were refined assuring that Cys32 can ligate to Fe2 of AuxI and that Thr37 does not clash with the SAM cofactor.

Code and Steps in creating CteA-CteB Model

1. First step modeling (preparing the input structure and ligating Cys32 to Fe2 (AuxI):

The following command was used to generate 100 models:

```
~/rosetta/main/source/bin/rosetta_scripts.linuxgccrelease \
 -database ~/rosetta/main/database \
 -s ../CthE_loopClosure.pdb \
 -nstruct 1 \
 -parser:protocol ../relax.xml \
 -nblist_autoupdate \
 -ignore_unrecognized_res \
 -load_PDB_components 1 \
 -PDB_components_file chemical/components.cif \
```

relax.xml file:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
          <talaris_w_csts weights="talaris2014_cst.wts" />
          <talaris weights="talaris2014.wts" />
            </SCOREFXNS>
            <RESIDUE_SELECTORS>
          <Chain name="chA" chains="A" />
              <Index name="chB" resnums="1B-9B" />
            </RESIDUE_SELECTORS>
        <TASKOPERATIONS>
```

```
        <OperateOnResidueSubset name="fixA" selector="chA" >
       <PreventRepackingRLT/>
      </OperateOnResidueSubset>
        <OperateOnResidueSubset name="fixB" selector="chB" >
       <PreventRepackingRLT/>
      </OperateOnResidueSubset>
       </TASKOPERATIONS>
       <FILTERS>
    </FILTERS>
    <MOVERS>
        <ConstraintSetMover name="cst" add_constraints="1" cst_file="constraints"/>
        <AddConstraintsToCurrentConformationMover name="coord_constr1" CA_only="0"
bb_only="1" coord_dev="0.5" cst_weight="1" task_operations="fixA"/>
        <AddConstraintsToCurrentConformationMover name="coord_constr2" CA_only="0"
bb_only="1" coord_dev="0.5" cst_weight="1" task_operations="fixB"/>
        <AddChainBreak name="chainbreak" change_foldtree="1" find_automatically="1"
distance_cutoff="2.5" />
        <FastRelax name="relax" repeats="1" scorefxn="talaris_w_csts">
          <MoveMap name="removejumps">
            <Jump number="1" setting="0"/>
            <Jump number="2" setting="0"/>
            <Jump number="3" setting="0"/>
            <Jump number="4" setting="0"/>
```

```
            <Jump number="5" setting="0"/>
          </MoveMap>
        </FastRelax>
    </MOVERS>
      <APPLY_TO_POSE>
    </APPLY_TO_POSE>
      <PROTOCOLS>
<Add mover_name="chainbreak" />
        <Add mover_name="coord_constr1" />
        <Add mover_name="coord_constr2" />
        <Add mover_name="cst" />
        <Add mover_name="relax" />
    </PROTOCOLS>
      <OUTPUT scorefxn="talaris" />
</ROSETTASCRIPTS>
```

constraints file:

AtomPair FE2 445 SG 459 BOUNDED 0 2 0.50 0.50 NOE

2. Second step modeling (building the peptide covering residues 10-31):

1000 models were generated using the following command:

```
~/rosetta/main/source/bin/remodel.linuxgccrelease \
 -database ~/rosetta/main/database/ \
 -s ../CthE_loopClosure_0026.pdb \
 -remodel:blueprint ../blueprint \
```

-run:chain B \

-remodel:num_trajectory 100 \

-nstruct 1

blueprint file:

1 M .
2 A .
3 M .
4 I .
5 H .
6 K .
7 F .
8 S .
9 M .
10 M .
11 G .
12 T .
13 N .
14 I .
15 V .
16 V .
17 D .
18 V .
19 N .
20 S .
21 G .
22 A .
23 V .
24 H .
25 V .
26 V .
27 D .
28 D .
29 I .
30 S .
31 F .
32 D .
33 I .
34 L .
35 D .
36 Y .
37 Y .
38 K .
39 N .
40 F .
41 T .
42 A .
43 G .
44 E .
45 I .
46 K .
47 N .
48 K .
49 L .
50 A .
51 H .
52 K .
53 Y .
54 N .
55 A .
56 D .
57 E .
58 I .
59 D .
60 E .
61 A .
62 L .
63 R .
64 E .
65 I .
66 E .
67 S .
68 L .
69 E .
70 A .
71 E .
72 G .
73 L .
74 L .
75 F .
76 S .
77 E .
78 D .
79 P .
80 Y .
81 K .
82 E .
83 Y .
84 V .
85 S .
86 S .
87 M .
88 D .
89 R .
90 K .
91 S .
92 V .
93 V .
94 K .
95 A .
96 L .
97 C .
98 L .
99 H .
100 I .
101 S .
102 H .
103 D .
104 C .
105 N .
106 L .
107 R .
108 C .
109 K .
110 Y .
111 C .
112 F .
113 A .
114 S .
115 Q .
116 R .
117 N .
118 M .
119 M .
120 S .
121 L .
122 E .
123 V .
124 G .
125 K .
126 K .
127 A .
128 I .
129 D .
130 F .
131 L .
132 I .
133 S .
134 E .
135 S .
136 G .
137 N .
138 R .
139 K .
140 N .
141 L .
142 E .
143 I .
144 D .
145 F .
146 F .
147 G .
148 G .
149 E .
150 P .
151 M .
152 M .

153 N .
154 F .
155 D .
156 V .
157 V .
158 K .
159 G .
160 I .
161 I .
162 E .
163 Y .
164 A .
165 R .
166 Q .
167 K .
168 E .
169 K .
170 E .
171 H .
172 N .
173 K .
174 N .
175 F .
176 R .
177 F .
178 T .
179 L .
180 T .
181 T .
182 N .
183 G .
184 L .
185 L .
186 L .
187 N .
188 D .
189 E .
190 N .
191 I .
192 K .
193 Y .
194 I .
195 N .
196 E .
197 N .
198 M .
199 Q .
200 N .
201 I .
202 V .
203 L .
204 S .
205 I .
206 D .
207 G .
208 R .
209 K .
210 E .
211 V .
212 N .
213 D .
214 R .
215 M .
216 R .
217 I .
218 R .
219 I .
220 D .
221 G .
222 S .
223 G .
224 C .
225 Y .
226 D .
227 D .
228 I .
229 L .
230 P .
231 K .
232 F .
233 K .
234 Y .
235 V .
236 A .
237 E .
238 S .
239 R .
240 N .
241 Q .
242 D .
243 N .
244 Y .
245 Y .
246 V .
247 R .
248 G .
249 T .
250 F .
251 T .
252 R .
253 E .
254 N .
255 M .
256 D .
257 F .
258 S .
259 N .
260 D .
261 V .
262 L .
263 H .
264 L .
265 A .
266 D .
267 E .
268 G .
269 F .
270 R .
271 Q .
272 I .
273 S .
274 V .
275 E .
276 P .
277 V .
278 V .
279 A .
280 A .
281 K .
282 D .
283 S .
284 G .
285 Y .
286 D .
287 L .
288 R .
289 E .
290 E .
291 D .
292 L .
293 P .
294 R .
295 L .
296 F .
297 E .
298 E .
299 Y .
300 E .
301 K .
302 L .
303 A .
304 Y .
305 E .
306 Y .
307 V .
308 K .
309 R .
310 R .
311 K .
312 E .
313 G .
314 N .
315 W .
316 F .
317 N .
318 F .
319 F .
320 H .
321 F .
322 M .
323 I .
324 D .
325 L .
326 T .
327 Q .
328 G .
329 P .
330 C .
331 I .
332 V .
333 K .
334 R .
335 L .
336 T .

337 G .
338 C .
339 G .
340 S .
341 G .
342 H .
343 E .
344 Y .
345 L .
346 A .
347 V .
348 T .
349 P .
350 E .
351 G .
352 D .
353 I .
354 Y .
355 P .
356 C .
357 H .
358 Q .
359 F .
360 V .
361 G .
362 N .
363 E .
364 K .
365 F .
366 K .
367 M .
368 G .
369 N .
370 V .
371 K .
372 E .
373 G .
374 V .
375 L .
376 N .
377 R .
378 D .
379 I .
380 Q .
381 N .
382 Y .
383 F .
384 K .
385 N .
386 S .
387 N .
388 V .
389 Y .
390 T .
391 K .
392 K .
393 E .
394 C .
395 D .
396 S .
397 C .
398 W .
399 A .
400 K .
401 F .
402 Y .
403 C .
404 S .
405 G .
406 G .
407 C .
408 A .
409 A .
410 N .
411 S .
412 Y .
413 N .
414 F .
415 H .
416 K .
417 D .
418 I .
419 N .
420 T .
421 V .
422 Y .
423 K .
424 V .
425 G .
426 C .
427 E .
428 L .
429 E .
430 K .
431 K .
432 R .
433 V .
434 E .
435 C .
436 A .
437 L .
438 W .
439 I .
440 K .
441 A .
442 Q .
443 E .
444 M .
445 M .
446 K .
447 H .
448 I .
449 K .
450 I E PIKAA I
0 x L PIKAA L
0 x L PIKAA N
0 x L PIKAA G
0 x L PIKAA S
0 x L PIKAA T
0 x L PIKAA L
0 x L PIKAA K
0 x L PIKAA D
0 x L PIKAA S
0 x L PIKAA L
0 x L PIKAA K
0 x L PIKAA K
0 x L PIKAA G
0 x L PIKAA G
0 x L PIKAA C
0 x L PIKAA G
0 x L PIKAA E
0 x L PIKAA C
0 x L PIKAA Q
0 x L PIKAA T
0 x L PIKAA S
0 x L PIKAA C
0 x L PIKAA Q
0 x L PIKAA S
451 A E PIKAA A
452 C

3. Third step modeling (building the peptide covering residues 32-46):

The following command was run to generate 400 models:

~/rosetta/main/source/bin/FloppyTail.linuxgccrelease \

-database ~/rosetta/main/database/ @../options3

options3 file:

#input PDB

-s ../CthE_cterExt_0026_0799.pdb

#ex flags give extra rotamers for packing; use_input_sc allows the pre-existing rotamer when packing (useful when paired with sidechain minimization)

-ex1

-ex2

-packing:repack_only

#start of tail

-FloppyTail:flexible_start_resnum 473

-FloppyTail:flexible_chain B

#used for preventing loss of compactness at centroid/fa switch; see documentation

-FloppyTail:short_tail:short_tail_off 0

-FloppyTail:short_tail:short_tail_fraction 1.0

#shear does nothing for extended tails; see documentation

-FloppyTail:shear_on .33333333333333333

#constraints

-constraints::cst_file ../constraints

-constraints::cst_weight 10

-constraints::cst_fa_file ../constraints

-constraints::cst_fa_weight 10

-FloppyTail:refine_repack_cycles 10

-FloppyTail:perturb_cycles 500

-FloppyTail:refine_cycles 300

-nstruct 1

-ignore_unrecognized_res

constraints file:

AtomPair CA 278 SG 476 BOUNDED 8 10 0.50 0.50 NOE

AtomPair CA 338 SG 476 BOUNDED 6 8 0.50 0.50 NOE

AtomPair CA 357 SG 476 BOUNDED 7 9 0.50 0.50 NOE

AtomPair CA 111 CA 481 BOUNDED 8 10 0.50 0.50 NOE

AtomPair CA 145 CA 481 BOUNDED 8 10 0.50 0.50 NOE

AtomPair CA 201 CA 481 BOUNDED 8 10 0.50 0.50 NOE

AtomPair CA 336 CA 490 BOUNDED 8 10 0.50 0.50 NOE

4. Fourth step modeling (reintroducing the ligands with pymol and refining using Rosetta):

The following command was run over the three of the lowest energy structures from the previous step to generate 100 models:

~/rosetta/main/source/bin/rosetta_scripts.linuxgccrelease \

 -database ~/rosetta/main/database \

 -s ../CthE_cterExt_0026_0799_0177_withCofactors.pdb \

 -nstruct 1 \

 -parser:protocol ../relax.xml \

-nblist_autoupdate \

-ignore_unrecognized_res \

-load_PDB_components 1 \

-PDB_components_file chemical/components.cif \

-auto_setup_metals

relax.xml file:

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
       <ScoreFunction name="talaris_w_csts" weights="talaris2014_cst.wts" />
       <ScoreFunction name="talaris" weights="talaris2014.wts" />
        </SCOREFXNS>
        <RESIDUE_SELECTORS>
          <Index name="res"
resnums="112A,278A,338A,339A,340A,344A,355A,356A,357A,358A,359A,383A,406A,407A,
408A,409A,410A,451B,452B,452B,453B,454B,456B,457B,458B,459B,460B,461B,462B,463B,
464B,465B,466B,467B,468B,469B,470B,471B,472B,473B,474B,475B,476B,477B,478B,479B,
480B,481B,482B,483B,484B,485B,486B,487B,488B,489B,490B" />
          <Neighborhood name="nghb" selector="res" distance="4.5" />
          <Not name="fix" selector="nghb" />
        </RESIDUE_SELECTORS>
        <TASKOPERATIONS>
          <OperateOnResidueSubset name="fix_outside" selector="fix" >
        <PreventRepackingRLT/>
```

```
  </OperateOnResidueSubset>
    </TASKOPERATIONS>
    <FILTERS>
  </FILTERS>
  <MOVERS>
      <ConstraintSetMover name="cst" add_constraints="1" cst_file="../constraints"/>
      <AddConstraintsToCurrentConformationMover name="coord_constr" CA_only="0"
bb_only="1" coord_dev="0.5" cst_weight="1" task_operations="fix_outside"/>
      <AddChainBreak name="chainbreak" change_foldtree="1" find_automatically="1"
distance_cutoff="2.5" />
      <FastRelax name="relax" repeats="1" scorefxn="talaris_w_csts">
        <MoveMap name="removejumps">
          <Jump number="2" setting="0"/>
          <Jump number="3" setting="0"/>
          <Jump number="4" setting="0"/>
          <Jump number="5" setting="0"/>
        </MoveMap>
      </FastRelax>
  </MOVERS>
    <APPLY_TO_POSE>
  </APPLY_TO_POSE>
    <PROTOCOLS>
      <Add mover_name="chainbreak" />
```

```
        <Add mover_name="coord_constr" />
        <Add mover_name="cst" />
        <Add mover_name="relax" />
    </PROTOCOLS>
      <OUTPUT scorefxn="talaris" />
</ROSETTASCRIPTS>
```

constraints file:

```
AtomPair FE1 445 SG 338 BOUNDED  0 2  0.50 0.50 NOE
AtomPair FE4 445 SG 407 BOUNDED  0 2  0.50 0.50 NOE
AtomPair FE2 445 SG 476 BOUNDED  0 2  0.50 0.50 NOE
```

2.8 REFERENCES

1. Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. *Chem. Rev.* **2014**, *114*, 4229

2. Grell, T. A. J.; Goldman, P. J.; Drennan, C. L. *J. Biol. Chem.* **2015**, *290*, 3964

3. Vey, J. L.; Drennan, C. L. *Chem. Rev.* **2011**, *111*, 2487

4. Frey, P. A.; Booker, S. *J. Adv. Protein Chem.* **2001**, *58*, 1

5. Dey, A.; Peng, Y.; Broderick, W. E.; Hedman, B.; Hodgson, K. O.; Broderick, J. B.; Solomon, E. I. *J. Am. Chem. Soc.* **2011**, *133*, 18656

6. Benjdia, A.; Guillot, A.; Lefranc, B.; Vaudry, H.; Leprince, J.; Berteau, *O. Chem. Commun.* **2016**, *52*, 6249

7. Haft, D. H.; Basu, M. K. *J. Bacteriol.* **2011**, *193*, 2745

8. Haft, D. H. *BMC Genomics.* **2011**, *12*, 1

9. Goldman, P. J.; Grove, T. L.; Sites, L. A.; McLaughlin, M. I.; Booker, S. J.; Drennan, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8519

10. Benjdia, A.; Subramanian, S.; Leprince, J.; Vaudry, H.; Johnson, M. K.; Berteau, O. *FEBS J.* **2010**, *277*, 1906

11. Benjdia, A.; Leprince, J.; Guillot, A.; Vaudry, H.; Rabot, S.; Berteau, O. *J. Am. Chem. Soc.* **2007**, *129*, 3462

12. 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.; Muller, 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*. **2012**, *30*, 108

13. Ortega, M. A.; van der Donk, W. A. *Cell Chem. Biol.* **2016**, *23*, 31

14. Dunbar, K. L.; Mitchell, D. A. *ACS Chem. Biol.* **2013**, *8*, 47

15. McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26* , 537

16. Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. *J. Am. Chem. Soc.* **2010**, *132*, 7519

17. Menzella, H. G. H.; Reeves, C. D. C. *Curr. Opin. Microbiol.* **2007**, *10*, 8

18. Sardar, D.; Schmidt, E. W. *Curr. Opin. Chem. Biol.* **2015**, *31*, 15

19. Kim, E.; Moore, B. S.; Yoon, Y. J. *Nat. Chem. Biol.* **2015**, *11*, 649

20. Ruffner, D. E.; Schmidt, E. W.; Heemstra, J. R. *ACS Synth. Biol*. **2015**, *4*, 482

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

22. Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. *Nat. Chem.* **2015**, *7*, 431

23. Barr, I.; Latham, J. A.; Iavarone, A. T.; Chantarojsiri, T.; Hwang, J. D.; Klinman, J. P. *J. Biol. Chem.* **2016**, *291*, 8877

24. Puehringer, S.; Metlitzky, M.; Schwarzenbacher, R. *BMC Biochem.* **2008**, *9*, 8

25. Schramma, K. R.; Seyedsayamdost, M. R. *ACS Chem. Biol.* **2017**, *12*, 922

26. Flühe, L.; Marahiel, M. A. *Curr. Opin. Chem. Biol.* **2013**, *17*, 605

27. Lohans, C. T.; Vederas, J. C. *J. Antibiot.* **2014**, *67*, 23

28. Jarrett, J. T. *J. Biol. Chem.* **2015**, *290*, 3972

29. Babasaki, K.; Takao, T.; Shimonishi, Y.; Kurahashi, K. *J. Biochem.* **1985**, *98*, 585

30. Shelburne, C. E.; An, F. Y.; Dholpe, V.; Ramamoorthy, A.; Lopatin, D. E.; Lantz, M. S. *J. Antimicrob. Chemother.* **2007**, *59*, 297

31. Rea, M. C.; Sit, C. S.; Clayton, E.; O'Connor, P. M.; Whittal, R. M.; Zheng, J.; Vederas, J. C.; Ross, R. P.; Hill, C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9352

32. Sit, C. S.; McKay, R. T.; Hill, C.; Ross, R. P.; Vederas, J. C. *J. Am. Chem. Soc.* **2011**, *133*, 7680

33. Wang, G.; Feng, G.; Snyder, A. B.; Manns, D. C.; Churey, J. J.; Worobo, R. W. *FEMS Microbiol. Lett.* **2014**, *357*, 69

34. Sit, C. S.; van Belkum, M. J.; McKay, R. T.; Worobo, R. W.; Vederas, J. C. *Angew. Chem. Int. Ed.* **2011**, *50*, 8718

35. Sutyak, K. E.; Wirawan, R. E.; Aroutcheva, A. A.; Chikindas, M. L. *J. Appl. Microbiol.* **2008**, *104*, 1067

36. Thennarasu, S.; Lee, D.-K.; Poon, A.; Kawulka, K. E.; Vederas, J. C.; Ramamoorthy, A. *Chem. Phys. Lipids* **2005**, *137*, 38

37. Flühe, L.; Knappe, T. A.; Gattner, M. J.; Schäfer, A.; Burghaus, O.; Linne, U.; Marahiel, M. A. *Nat. Chem. Biol.* **2012**, *8*, 350

38. Flühe, L.; Burghaus, O.; Wieckowski, B. M.; Giessen, T. W.; Linne, U.; Marahiel, M. A. *J. Am. Chem. Soc.* **2013**, *135*, 959

39. Wieckowski, B. M.; Hegemann, J. D.; Mielcarek, A.; Boss, L.; Burghaus, O.; Marahiel, M. A. *FEBS Lett.* **2015**, *589*, 1802

40. Bruender, N. A.; Bandarian, V. *Biochem.* **2016**, *55*, 4131

41. Bruender, N. A.; Wilcoxen, J.; Britt, R. D.; Bandarian, V. *Biochem.* **2016**, *55*, 2122

42. Benjdia, A.; Guillot, A.; Lefranc, B.; Vaudry, H.; Leprince, J.; Berteau, *O. Chem. Commun.* **2016**, *52*, 6249

43. Himes, P. M.; Allen, S. E.; Hwang, S.; Bowers, A. A. *ACS Chem. Biol.* **2016**, *11*, 1737

44. Murphy, K.; O'Sullivan, O.; Rea, M. C.; Cotter, P. D.; Ross, R. P.; Hill, C. *PLoS ONE* **2011**, *6*, e20852

45. Lanz, N. D.; Grove, T. L.; Gogonea, C. B.; Lee, K.-H.; Krebs, C.; Booker, S. *J. Meth. Enzymol.* **2012**, *516*, 125

46. Dowling, D. P.; Vey, J. L.; Croft, A. K.; Drennan, C. L. *Biochim. Biophys. Acta* **2012**, *1824*, 1178

47. Harmer, J. E.; Hiscox, M. J.; Dinis, P. C.; Fox, S. J.; Iliopoulos, A.; Hussey, J. E.; Sandy, J.; Van Beek, F. T.; Essex, J. W.; Roach, P. L. *Biochem. J.* **2014**, *464*, 123

48. Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. *Nature* **2015**, *517*, 509

49. Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W. E.; Jaspars, M.; Naismith, J. H. *Nat. Chem. Biol.* **2015**, *11*, 558

50. Ghodge, S. V.; Biernat, K. A.; Bassett, S. J.; Redinbo, M. R.; Bowers, A. A. *J. Am. Chem. Soc.* **2016**, *138*, 5487

51. Ortega, M. A.; Hao, Y.; Walker, M. C.; Donadio, S.; Sosio, M.; Nair, S. K.; van der Donk, W. A. *Cell Chem. Biol.* **2016**, *23*, 370

52. Tsai, T.-Y.; Yang, C.-Y.; Shih, H.-L.; Wang, A. H. J.; Chou, S.-H. *Proteins*. **2009**, *76*, 1042

53. Regni, C. A.; Roush, R. F.; Miller, D. J.; Nourse, A.; Walsh, C. T.; Schulman, B. A. *EMBO J.* **2009**, *28*, 1953

54. Cheung, W. L.; Chen, M. Y.; Maksimov, M. O.; Link, A. J. *ACS Cent. Sci.* **2016**, *2*, 702

55. Hänzelmann, P.; Schindelin, H. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 6829

56. Hänzelmann, P.; Schindelin, H. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12870

57. Koehnke, J.; Bent, A. F.; Zollman, D.; Smith, K.; Houssen, W. E.; Zhu, X.; Mann, G.; Lebl, T.; Scharff, R.; Shirran, S.; Botting, C. H.; Jaspars, M.; Schwarz-Linek, U.; Naismith, J. H. *Angew. Chem. Int. Ed.* **2013**, *52*, 13991

58. McLaughlin, M. I.; Lanz, N. D.; Goldman, P. J.; Lee, K.- H.; Booker, S. J.; Drennan, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9446

59. Berkovitch, F.; Nicolet, Y.; Wan, J. T.; Jarrett, J. T.; Drennan, C. L. *Science* **2004**, *303*, 76

60. Moser, C. C.; Anderson, J. L. R.; Dutton, P. L. *Biochim. Biophys. Acta* **2010**, *1797*, 157

61. Green, M. R.; Sambrook, J. *Molecular cloning: a laboratory manual 4th ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. **2012**

62. Studier, F.W. *Protein Expr Purif.* **2005**, *41*, 207

63. Li, B.; Cooper, L. E.; van der Donk, W. A. *Method Enzymol.* **2009**, *48*, Chapter 21, 533

64. Thibodeaux, C.J.; Ha, T.; van der Donk, W.A. *J. Am. Chem. Soc.* **2014**, *136*, 17513

65. Blommel, P. G.; Fox, B. G. *Protein Expr. Purif.* **2007**, *55*, 53

66. Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307

67. Vonrhein, C.; Blanc, E.; Roversi, P.; Bricogne, G. *Methods Mol. Biol.* **2007**, *364*, 215

68. Langer, G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A. *Nat Protoc.* **2008**, *3*, 1171

69. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213

70. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486

71. Vagin, A. A.; Steiner, R. A.; Lebedev, A. A.; Potterton, L.; McNicholas, S.; Long, F.; Murshudov, G. N. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2184

72. Leaver-Fay, A.; Tyka, M.; Lewis, S.M.; Lange, O.F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P.D.; Smith, C.A.; Sheffler, W.; Davis, I.W.; Cooper, S.; Treuille, A.; Mandell, D.J.; Richter, F.; Ban, Y.E.; Fleishman, S.J.; Corn, J.E.; Kim, D.E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J.J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J.J.; Kuhlman, B.; Baker, D.; Bradley, P. *Methods Enzymol.* **2011**, *487*, 545

73. Kleiger, G.; Saha, A.; Lewis, S.; Kuhlman, B.; Deshaies, R.J.; *Cell*, **2009**, *139*, 957

74. Huang, P.S.; Ban, Y.E.; Richter, F.; Andre, I.; Vernon, R.; Schief, W.R.; Baker, D.; *PLoS One*, **2011**, *6*, e24109

CHAPTER 3

PRODUCTION OF SACTIPEPTIDES IN *ESCHERICHIA COLI*: PROBING THE SUBSTRATE PROMISCUITY OF SUBTILOSIN A BIOSYNTHESIS

3.1 Introduction

Sactipeptides are a growing class of modified peptide natural products. The characteristics of these compounds are the signature intramolecular thioether bridges between cysteine residue sulfurs and the unreactive α-carbons of the bridging partner amino acids, known as sactionine bridges or linkages **(Figure 3.1a)**. As of the time of this writing, only five sactipeptides have been isolated and structurally characterized from their native producers. This list includes subtilosin A and sporulation killing factor (Skf), both from *Bacillus subtilis* 168, thurincin H from *Bacillus thuringiensis* SF361, and the two component sactipeptide thuricin CD from *Bacillus thuringiensis* DPC 6431 **(Figure 3.1b)**.[1–8] These sactionine bridges, unlike the β-thioether bridges found in lantipeptides, are quaternary due to the bridging partner side chain. Also, sactipeptide rings tend to be co-axial in regards to the peptide backbone constrained in a U-like formation. Sactipeptide tend to have highly defined regions of secondary structure due to the distribution and number of sactipeptide bridges as well as the stereochemistry of the α,α-disubstituted bridging partner residues. Specifically, the N-terminus of subtilosin A adopts a 3_{10} helix while thuricin CD has two α-helical faces **(Figure 3.1c)**. The amphipathic helicity of regions within these peptides is thought to grant subtilosin A and other sactipeptides narrow

spectrum activity through the ability to interact and disrupt bacterial cell walls resulting in often cell death through membrane disrubtion.[4,9,10] Still, the precise mechanisms of action of many sactipeptides are not yet fully understood.

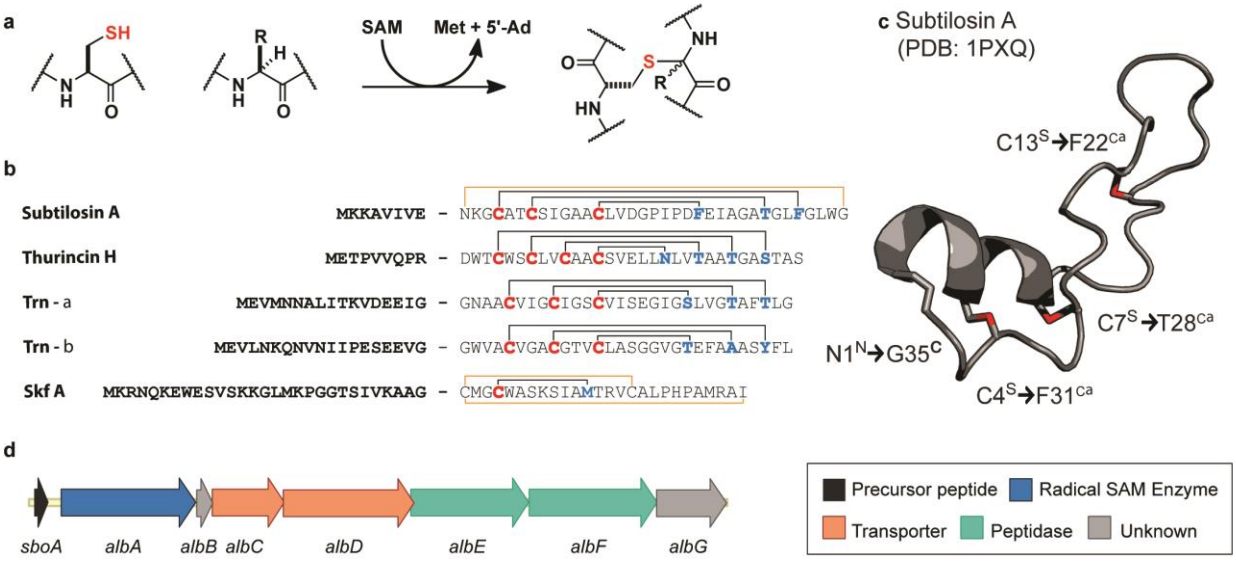


Figure 3.1. Sactipeptide biosynthesis. a) Sactionine linkage formation from cysteines and cross-ring amino acid coupling partners. b) Comparison of amino acid sequences of known sactipeptides with sactionine cysteines in red and bridging partners in blue. c) NMR structure of subtilosin A, indicating sactionine linkages and head to tail cyclization. d) The subtilosin A biosynthetic cluster from *Bacillus subtilis* sp. 168 including the structural gene *sboA* and radical SAM sactionine synthase-encoding gene *albA*.

In terms of biosynthesis, sactipeptides are ribosomally synthesized and post-translationally modified peptide (RIPPs) natural products.[11] RiPPs are produced from gene-encoded precursor peptides, which are then subsequently modified, post-translationally, by enzymes to install the characteristic sactionine linkages. Recently, Marahiel and co-workers characterized the enzymes responsible for installing the sactionine linkages in subtilosin A, Skf, and thurincin H, specifically, the radical SAM enzymes, AlbA, SkfB, and ThnB, respectively.[12–14] These enzymes called sactionine synthases are members of the newly defined SPASM domain, named for named for the biochemically characterized enzymes involved in subtilosin A, pyrroloquinoline quinone, anaerobic sulfatase and mycofactocin maturation. The

SPASM domain in this family of enzymes are predicted to have two [4Fe-4S] clusters, one each for activating the sulfhydryl and for coupling partner carbon center which results in a radical−radical hetero-coupling during cyclization **(Figure 3.1a)**.[15] However, there are numerous large gaps that still exist in our understanding, specifically regarding the mechanism and promiscuity of these sactionine synthases and what the effect changes in precursor peptide sequence and structure may have on their innate bactericidal activity. Some particular questions that remain unanswered are how the sactionine synthases, such as AlbA or ThnB, select multiple different bridging partner residues specifically at each bridge site and to what extent this can be manipulated or modified.

A homologous expression system has already been reported for thurincin H, using a plasmid encoded copy of the precursor peptide in the background of a precursor-knockout of the native producer, *B. thuringiensis* SF361.[16] Mutants of subtilosin A have also been genetically encoded on the pDG-148 plasmid and co-expressed with the native substrate. The mutants were isolated in the native producer *B. subtilis* 168.[13] Homologous expression does allow limited mutational sampling, but the study of the promiscuity of radical SAM (rSAM) could be obscured by multiple factors found in homologous expression systems. These factors include (1) poor downstream processing by other pathway enzymes potentially also contributing to (2) inefficient export by the dedicated natural product transporters in the clusters, (3) decreased transcriptional amplification by the known feedback regulation mechanisms or (4) rapid proteolytic degradation of partially modified peptides by the potent intracellular protease activity inherent to many strains of *Bacilli*.

Because of these problems, we thought that the development of a heterologous system for *in vivo* overexpression of partially to fully-modified sactipeptides would propel future

investigations of their biosynthesis forward, as well as, allow for quick and large-scale production of modified variants.[17–23] We therefore sought a system that would allow (1) robust overexpression of precursor peptides, (2) allow modification by modifying enzymes, and (3) allow isolation of sufficient materials for downstream characterization and/or potential application. We decided to set out and develop a system to allow heterologous expression of processed (modified) sactipeptides in *E. coli* under the control of the strong T7 promoter and also allow purification from cell pellets. By implementing *E. coli* expression, we could adapt the mature amber stop-codon suppression technology which has been employed to varying successes with other RiPPs natural products.[20,24–28] We chose to focus these first efforts on the sactipeptide subtilosin A made from its precursor peptide SboA, and sactionine synthase AlbA **(Figure 3.1d)**. Fully matured subtilosin A contains three sactionine thioether bridges and, as such, provides a number of opportunities to examine the interdependency of the different bridge formation sites.

3.2 Heterologous Production of Pre-subtilosin A

To ensure Ni^{2+} IMAC purification would be easily achieved with our heterologously expressed sactipeptides, we first tested AlbA's compatibility with both N-terminal and C-terminal hexahistidine (6xHis) tags. The subtilosin A precursor gene *sboA* was separately cloned into the pMCSG7 plasmid, which houses an N-terminal-6xHis tag and pET28c (C-terminal-6xHis-tag) followed by expression and purification from *E. coli* BL21 (DE3) cells. Full length peptides were purified from inclusion bodies using HISTrap, Ni-NTA, columns according to the procedure of Li *et al*.[29] In parallel, N-terminal-6xHis-tagged sactionine synthase AlbA was expressed, purified, and reconstituted as described by Flühe et al.[13] Upon incubation of peptides, enzyme, and essential co-factors (dithionite and *S*-adenosylmethionine), we could confirm complete formation of all three sactionine bridges by LC/MS with either 6xHis-tag. This result

suggested that tandem, heterologous expression of the AlbA sactionine synthase and the 6xHis-tagged SboA substrate could in theory yield mature sactipeptides. We chose to move forward with the N-terminal 6xHis-tag as we postulated that the placement of several lysine residues in SboA would be advantageous and allow easy removal of the 6xHis-tag together with the leader peptide of SboA *via* trypsin digestion.

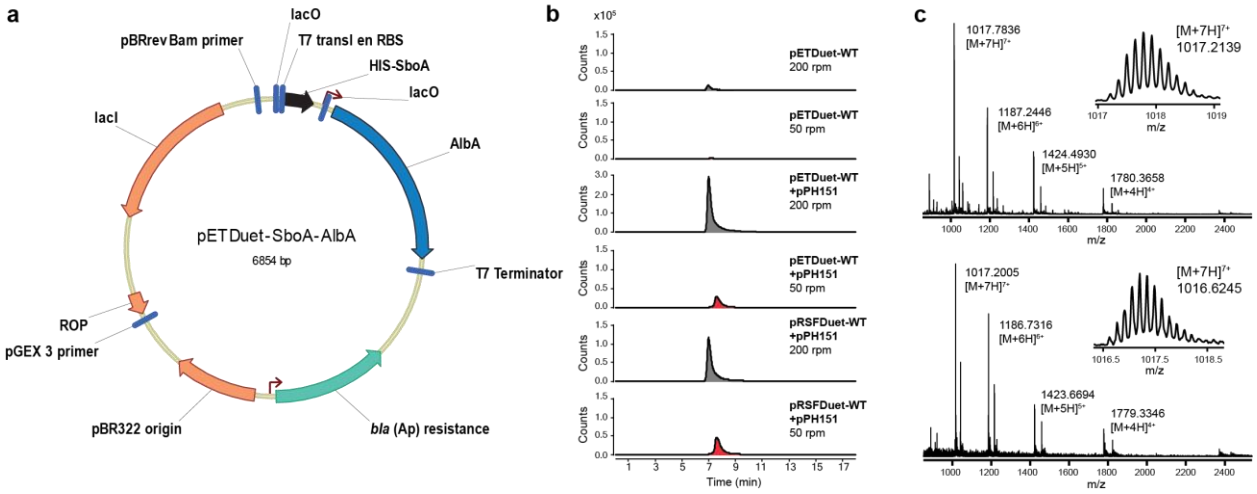


Figure 3.2. Expression of sactipeptides in *Escherichia coli*. a) Vector map of pETDuet-SboA-AlbA used in this study showing organization of genes in the two multiple cloning sites. b) LC-MS comparison of *in vivo* production of unmodified precursor peptide (gray) and modified (red) from constructs under varied, listed, conditions. pPH151 containing the *E. coli suf ABCDSE* genes aids in expression and repair of [4Fe-4S] clusters of sactionine synthase (AlbA) required for activity. (c) QTOF-MS data of unmodified (top) and modified (bottom) SboA peptide from the *E. coli* heterologous expression system.

N-terminal, 6xHis-tagged precursor peptide, SboA and the enzyme, AlbA were cloned into multiple cloning sites 1 (MCS1) and 2 (MCS2), respectively, of the pETDuet-1 plasmid to generate the bicistronic plasmid **(Figure 3.2a)**. Unlike the native producer *Bacillus subtilis* 168, this particular expression system would not be capable of producing the fully modified head-to-tail cyclized product as designed. We did not include any candidate protease capable of forming this last linkage and therefore no head-to-tail cyclized product could be formed. However, it has already been shown that the lack of head-to-tail cyclization does not affect the ability of the

sactionine synthase, AlbA, to modify the precursor peptide. Head-to-tail cyclization is often not present in a number of other sactipeptides, such as thurincin H and thuricin CD.[13] A wide range and variety of conditions were tested to facilitate and improve sactionine formation *in vivo* **(Figure 3.2b)**. We ultimately found that co-expression in the presence of pPH151, termed corrector plasmid, which contains the *E. coli suf ABCDSE* genes along with dropping the shake rate significantly improved expression of the modified sactipeptide.[30,31] These two measures together ensure proper AlbA expression and activity by (1) proper assembly and repair of the [4Fe-4S] clusters, assisted by suf ABCDSE, and (2) limiting aeration and possible deactivation of the enzyme by oxidation through oxygen. SDS-PAGE gels of crude lysates showed remarkable improved expression of AlbA under these optimized conditions **(see Appendix Figure B.28).** The high copy duet vector pRSFDuet-1 was also tested but showed negligible improvement over our pETDuet-1 construct. We therefore employed this, pETDuet-SboA-AlbA, construct in our further examination of AlbA sactionine substrate promiscuity.

3.3 Design and Evaluation of SboA Mutants

Upon confirming production of modified sactipeptides by our system, the next step was to examine the ability of AlbA to accept alternate SboA substrates *in vivo* using our heterologous expression system. Sactipeptide mutants have largely been limited to simple alanine scans or conservative mutations at bridging partner residues in the sactionine linkage. Preliminary efforts by Marahiel and co-workers suggested that the sactionine synthase AlbA is relatively non-tolerant to substrate changes which stand in direct contrast to the vast majority of reported RiPP enzymes characterized to date. Because of these previously reported results, we designed our initial mutants with three primary goals in mind: (1) to test the promiscuity at bridging partner residues, (2) to examine requirements for substrate flexibility or “preorganization”, and (3)

cysteine spacing relative to the brief leader peptide (8 residues) sequence. Saturation mutagenesis would require a library size too large to allow specific probing of each of these areas; we therefore pursued a more conservative rational design approach compared to saturation mutagenesis. In the case of the three bridging positions present in subtilosin A, Cys4^{S}-Phe31$^{C\alpha}$, Cys7^{S}-Thr28$^{C\alpha}$, and Cys13^{S}-Phe22$^{C\alpha}$ **(Figure 3.1c)**, we incorporated a small subset of sterically and electronically distinct amino acids that are observed at bridging partner positions in other known sactipeptides, specifically the amino acids Ala, Gln, Met, Phe, Ser, and Thr **(Table 3.1a, entries 1−14)**. The second set of mutants targeted the flexible glycine residues that are adjacent to sactionine bridging residues by alanine mutation (Gly's 26, 29, and 32), as well as two potential turn-inducing prolines at residues 18 and 20 respectively. The two prolines were also targeted for direct deletion in order to fully probe the effects of the residues 14−21 loop on priming sactionine bridge formation. **(Table 3.1a, entries 15-27)** The last set of mutations focused on the movement of the cysteine residues themselves, involved in sactionine bridge formation, around the "N-terminal side" or first half of SboA. In order to minimize the number of potential changes within this set, we chose to move cysteines by "swapping" or exchanging them with residues at their respective positions. **(Table 3.1a, entries 28-40)** The analogous Ala-swap could also be considered for further investigation.

For this initial study, we designed and fully characterized a total 40 sactipeptide mutants. Several of these mutants were accessed by QuikChange mutagenesis; however, the majority could be readily obtained by gene synthesis at equivalent or lesser cost and time (see Experimental Section 3.9.1). In each case, the mutant gene was incorporated with 6xHis-tag on the N-terminus into MCS1 of the pETDuet vector containing AlbA in MCS2 and expressed in the pPH151 background under the same optimized conditions as the wild-type SboA using small

(100 mL culture size) grow-ups. The peptides were induced with IPTG, as with the wild type, and isolated in a medium-throughput manner by using small spin Ni-NTA columns. To facilitate characterization by LC-MS, the peptides were digested with the use of trypsin to remove the 6x-His-tag plus a portion of the leader peptide. Trypsin seemed to leave the major product as the peptide cleaved at Lys,-6 of the leader peptide **(Figure 3.1b)**.

Table 3.1a. SboA mutants analyzed and sites of linkages identified by MS-MS

Sequence and connectivity of wild-type SboA:

Entry	N-term changes	C-term changes	No. Bridges	Bridging Partners
1		F22A	3	A22, T28, F31
2		F22M	3	M22, T28, F31
3		F22S	no product	--
4		F22T	no product	--
5		T28A	3	F22, A28, F31
6		T28F	3	F22, F28, F31
7		T28N	3	F22, N28, F31
8		T28S	3	F22, S28, F31
9		T27/A28	3	F22, A28, F31
10		F31A	3	F22, T28, A31
11		F31M	no product	--
12		F31N	no product	--
13		F31S	3	F22, T28, S31
14		F31T	no product	--
15		G26A	3	F22, T28, F31
16		G29A	3	--
17		G32A	3	F22, T28, F31
18		G26A,G29A	3	--
19		G26A,G32A	3	F22, T28, F31
20		G29A,G32A	3	--
21		G26A,G29A,G32A	3	F22, T28, F31
22		P18A	3	F22, T28, F31
23		P20A	3	F22, T28, F31
24		P18A, P20A	3	F22, T28, F31
25		ΔP18	3	P19, T27, F30
26		ΔP20	3	F21, T27, F30
27		ΔP18, ΔP20	no product	
28	C1/N4		no product	--
29	C2/K4		no product	--
30	C3/G4		1	F31, 1 disulfide
31	A4/C5		no product	--
32	C6/T7		no product	--
33	S7/C8		1	F31, 1 disulfide
34	I7/C9		no product	--
35	G7/C10		1	F31, 1 disulfide
36	C10/G13		3	F22, T28, F31
37	C11/A13		no product	--
38	C12/A13		3	F22, T28, F31
39	L13/C14		3	F22, T28, F31
40	V13/C15		3	F22, T28, F31

A very distinctive tandem mass spec. fragmentation pattern of sactionine linkages **(Figure 3.3a)** was used to identify bridge partnering residues of cysteines in the sactionine linkages.[4,32] The thioamidals in sactionine bridges have been shown to undergo facile retro-elimination and tautomerization to the corresponding dehydro-amino acid at low collision energies as presented in **Figure 3.3a−d**; the new amide bond formed from this dehydro-amino acid is much more labile than a typical peptide bond due to the resonance-stabilized enamine leaving group. Thus, at low collision voltages (~20-35 eV), only cleavage of these bonds is observed by the instrument while the other amide bonds stay intact, allowing the identification of the bridge partnering residues. This MS-generated formal dehydrogenation at bridging partners has previously been reported by Vederas *et al*., who exploited it to correctly assign the connectivity in the structure of the two-component sactipeptide system thuricin CD. The alternative method, employing nickel-borodeuteride desulfurization, did work in our hands but often gave variable and incomplete results, even with WT-subtilosin A itself.[7,8] We hypothesize that this was potentially due to metal contaminants carried over from Ni- NTA purification of the sactipeptides or other catalyst poisons from the *E. coli* heterologous system. In the tandem mass spec. method, sactionine linkages were easily the most labile at lower collision energies, and thus provided strong qualitative confirmation of bridging partner residues with low parts-per-million (ppm) error. In all cases with mutant sactipeptide, the presence of less than three bridges could be readily confirmed by reductive treatment with N-ethylmaleimide (NEM) and LC-MS to show masses of the relevant NEM adducts. Although sactionine linkages can be formed with differing stereochemistry at the bridge partner α-carbon (D or L as is the case with subtilosin A), we did not examine the effects of mutants on stereochemistry within this system. Outside of nuclear

magnetic resonance (NMR), there are no good methods to measure sactionine stereochemistry, in regard to our system, reported in the literature to date.

We did compared production of our modified peptides to native production of subtilosin A in *B subtilis* 168. *B. subtilis* is reported to produce subtilosin A at roughly 5.5 mg/L.[6] We obtained between 1 and 2 mg/L dry weight or ~20−40% of native production from the pETDuet system (see Appendix B). Production levels of mutants were highly variable, and many of them could only be detected very faintly by UV-vis absorption. Based on extracted ion chromatograms (EICs) of the products, yields of mutants varied from 10% to greater than 300% when compared to EIC of the modified wild type peptide on the same scale **(see Appendix Figure B.30-B.59)**. In general, many of the cysteine swap mutants were produced at comparable or higher levels than the wild-type SboA peptide in the pETDuet system **(see Appendix Table B.1)**. In several instances (12 out of 40), we observed no product from induction of a given mutant. The variable or complete lack of production may happen for any number reasons. Intrinsic destabilization of the given mutant peptide, its transcript causing degradation, unanticipated metabolic processing such as glutathionylation, other detoxifying post-translational modifications could lead to low or no production and might cause the sactipeptide product to escape our isolation procedure for this system. We expected that post-translation installation of sactionine linkages would improve overall peptide stability. Therefore, in the analysis, instances of "no product" **(Table 3.1)** were actually interpreted as a weak proclivity of the enzyme to accept those substrates. However, it should be noted that these are the results of peptide expression under the current and reported *in vivo* conditions. It is reasonable to speculate that while expression of some mutants may be lower than wild-type, even production at these low levels has been improved by the presence of strong overexpression of the sactionine synthase relative to levels in the native producer.

3.4 Substrate Tolerance at Bridging Partners

Several different amino acids, with different steric and electrochemical properties, appear at bridging partner residues with in sactipeptides. Even within subtilosin A, Phe and Thr appear at different positions, suggesting that the sactionine synthase, AlbA, might have some level of tolerance for different bridging partners. A set of alanine mutants at the three bridging partners of subtilosin A was tested in our *E. coli* expression system **(Table 3.1a, entries 1, 5, and 10)** and in all three Ala mutants, all three bridges were formed and confirmed by MS-MS showing that the site of modification remained the same, i.e. cross-linking occurred at the new Ala22, Ala28, or Ala31 residue. Marahiel and co-workers had reported a lack of production of F22A, T28A, and F31A mutants expressed in parallel with the native cluster in *B. subtilis*.[13] However, in an *E. coli* heterologous expression system, these three mutants were well tolerated by the AlbA machinery, demonstrating a potential benefit of this approach. Beyond alanine residues, only the +28 position appeared particularly permissive to any other amino acid residue substitution. Thr28 could also be substituted with Ser, Phe, and Asn and still be activated and incorporated into the resulting sactionine bridge. Meanwhile, Phe22 could be substituted with Met and still be processed to a sactionine bridge, but neither Ser nor Thr substitutions at this position yielded identifiable peptide product. Similarly, the F31S was processed to a three bridge product, but no product could be observed with Thr, Met, or Asn substitutions at this position. Given the obvious promiscuity at the +28 residue, we wondered if there could be any bias for the native amino acid Thr. We therefore exchanged Thr28 with the upstream Ala27 to see if this would drive bridge formation toward the newly positioned Thr. This, however, was not the case and the sactionine bridge was instead formed at Ala28, suggesting minimal residue level control over steering modifications, at least at this position.

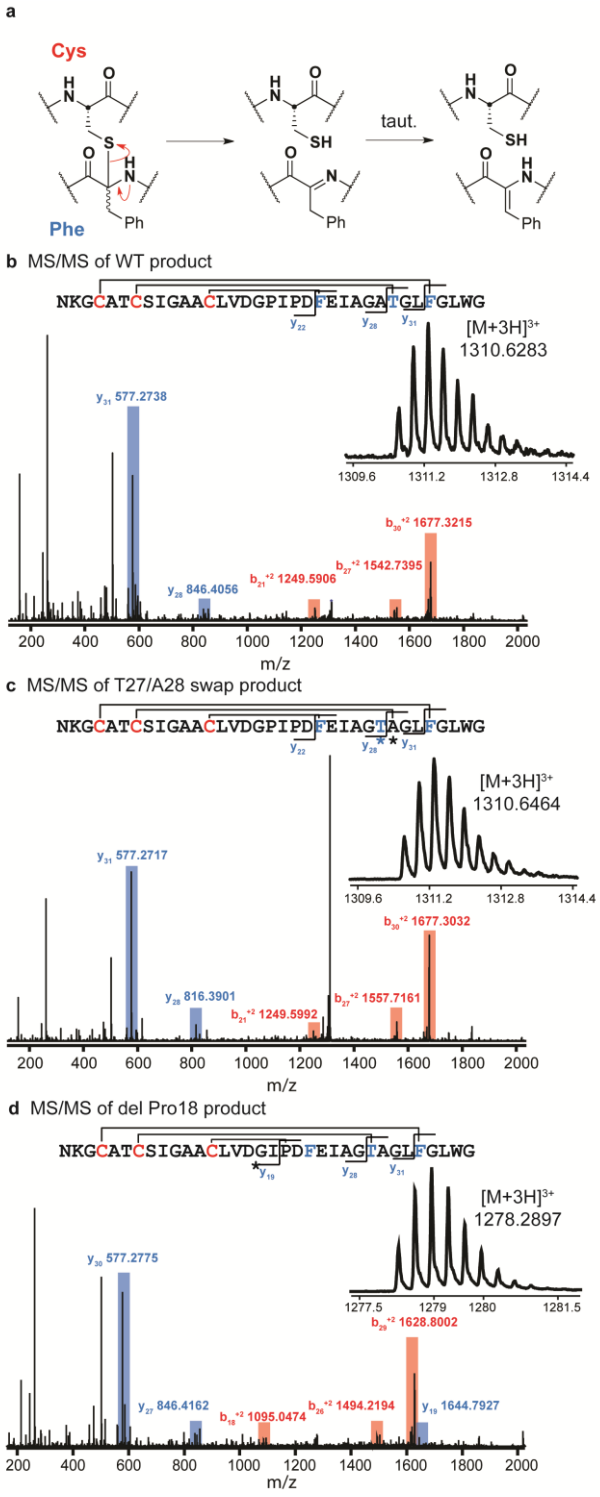


Figure 3.3. Confirming bridge formation in mutant sactipeptides. a) Mechanism of sactionine fragmentation. b) MS/MS on modified "wild-type" SboA peptide from 50 rpm overnight culture. c) MS-MS on modified T27/A28 swap SboA mutant peptide from 50 rpm overnight culture. d) MS/MS on modified Δ Pro 18 SboA mutant peptide from 50 rpm overnight culture. Insets show $[M+3H]^{3+}$ ion targeted in MS-MS.

3.5 Substrate Tolerance at Unmodified Positions

The precursor peptides for subtilosin and the two sactipeptides that make up thuricin CD, Trn-α and Trn-β, have relatively high glycine content (22%, 26%, and 22% of non-cysteine residues respectively). We sought to find out whether substrate flexibility arising from these constitutive glycine residues was necessary to allow enzymatic modification by the sactionine synthase or if Cα-branching might drive other alternative bridging partners. An alanine scan of the three C-terminal glycine residues, Gly26, Gly29, and Gly32, demonstrated that these positions can tolerate at least small, sterically unhindered alanine side chains at these positions **(Table 3.1a, entries 15−21)**. Besides modification occurring with these single mutations the same as WT modification, all three positions could be substituted with alanines at once, with no detriment to modified product formation. All products isolated from these mutation experiments formed three, complete sactionine bridges with the native residues of WT (Phe22, Thr28, and Phe31), suggesting that AlbA is at least somewhat indifferent to secondary structure or flexibility at these specific glycine residues.

Subtilosin A has the largest loop of the known sactipeptides (eight amino acids compared to the seven in thuricin CD and the four in thurincin H), and none of these other known sactipeptides have prolines, much less two prolines, in their designated loop regions. We therefore sought to determine whether Pro18 and/or Pro20 were necessary for inducing a turn and allowing AlbA to work on the sets of bridging residues **(Table 3.1a, entries 22−27)**. Again, alanine scans at these two Pro positions demonstrated that AlbA was indifferent to the presence of prolines at these positions similar to the results seen in the glycine mutations. However, complete deletion of Pro18 did appear to shift the bridging partner of Cys13 in SboA for the first time in our experiments **(Table 3.1a, entry 25)**. Careful study and examination of the tandem

mass spec. data indicated that the Cys13 sactionine thioether is formed at the α carbon of the new Pro19 in isolates from this mutant while the locations of the other two sactionine bridges are conserved and not changed. In contrast, deletion of Pro20 did not change the bridging partners of the cysteines involved in thioether bridge formation again **(Table 3.1a, entry 26)**. Overall, these mutants demonstrated substrate promiscuity in regards to AlbA activity and suggest that the substrate-enzyme interactions may play a role in dictating the position of bridge formation. Alterations to the loop proline residues could also conceivably be expected to have an impact on stereochemistry, especially at the nearby Cys13−Phe21 sactionine linkage, but this a question for future investigations.

3.6 Substrate Tolerance for Cysteine Placement

Interestingly, subtilosin A has seemingly relaxed specificity for the bridging partner residues and is able to accept Ala, Gln, Met, Phe, Ser, and Thr residues at the bridging partners to varying degrees of success. Therefore, two major issues can be probed with respect to cysteine placement: (1) the possibility of cysteines being moved and still form bridges and (2) what will the bridging partner residue be, if the cysteine is moved. We set out to probe these issues with a series of 14 cysteine swap mutants: four swap mutants at both Cys4 and Cys7, five swap mutants at Cys13, and also a tandem mutant. Under our heterologous expression conditions, described above, sactionine bridges at Cys4 and Cys7 seemed very resistant to changes in placement **(Table 3.1a, entries 28−35)**. Cys4 did tolerate a single move one residue toward the leader peptide **(Table 3.1a, entry 30)**, but this move came at the expense of the formation of the remaining two sactionine bridges, which did not form in this product. Similarly, Cys7 could be swapped with Ser8 or Gly10, while still allowing formation of the sactionine

bridge at Cys4 **(Table 3.1a, entries 33 and 35)**, but neither of the other two bridges were formed.

In stark contrast to the Cys4 and Cys7 bridge placements, Cys13 proved remarkably tractable. Of the five positional swaps tested **(Table 3.1, entries 36−40)**, only the Cys11/Ala13 swap (entry 37) did not yield a fully modified three-bridge sactipeptide. Each placement of Cys13, either at the +10, + 12, + 14, or +15 positions within SboA, resulted in a sactionine linkage to the same bridging partner as the WT sactipeptide linkage (Phe22). Interestingly, the Gly10 position, which did not allow a swap with Cys7, is capable of tolerating a sactionine bridge when a +7 bridge is also formed in the final product **(Table 3.1, entry 36)**.

The relative success of the mutants that were tested may betray the order of AlbA-catalyzed sactionine bridge formation in this *E. coli*, heterologous host. While the system does not appear to tolerate cysteine movement at the +4 and +7 positions, the third bridge at Cys13 seems to tolerate a wide variety of movements. This could be due to the first two sactionine bridges being formed quickly, allowing the rest of the peptide to reach a type of conformational equilibrium before the third bridge can be formed; further kinetic analysis and experiments of the bridge-forming reaction will be necessary to confirm this hypothesis. The formation of a bridge to Phe22 regardless of the position of the third cysteine indicates a degree of enzymatic control over regioselectivity. This control, however, is overridden in the case of Pro18 deletion, which suggests that this residue is critical in both the spacing and in the conformation of the loop.

3.7 Unpublished Mutants Tested

The 40 mutants tested and published were not the only mutants cloned and tested. In **Table 3.1b**, another 31 mutants, designed to have 2, 3 or even 4 sactionine bridges, are listed with their changes and their preliminary results by LC-MS. These mutants were designed to

have different secondary structure characteristics than the native sactipeptide. The mutants were not included in the other table of mutants **(Table 3.1a)** due to the low quality of data obtained during the initial testing of production. While some mutants (F22N and T28M, not listed) were not initially successfully cloned into out heterologous expression system, other mutants did not produce or simply gave inconclusive MS data after purification of the product from inclusion bodies. Others mutant SboA peptides did produce and an initial run on the LC-MS gave an indication of the number of sactionine bridges or disulfides formed. However, upon trypsin digestion, NEM modification, and injection onto the LC-MS to run tandem mass spectrometry, no product ions could be analyzed successfully. This led us to the conclusion that while the mutants were produced, they were produced as such low levels that accurate determination of the bridging partners was impossible. Low or no production could be due to a number of factors including intrinsic destabilization and degradation, unaccounted for metabolic processing, other detoxifying post-translational modifications. These modifications might cause the sactipeptide product to escape our designed isolation procedure for this heterologous system.

3.8 Incorporation of Unnatural Amino Acids into Sactipeptides

Unnatural amino acids (UAAs) have been incorporated in a number of heterologously expressed RiPP natural products by amber codon suppression technology, including both the lantipeptides and the cyanobactins.[20,24–28] In the case of the lantipeptides, UAAs allowed access to mutant lanitpeptides with improved potency and solubility. We anticipated that similar effects could be obtained with UAA incorporation into bioactive sactipeptides. UAA incorporation may also allow access to new chemistries from sactionine forming sactionine synthases: since the enzyme putatively generates a reactive radical intermediate, at the α carbon of the bridging residue , a specific placement of a radical trap or other reactive group on a bridging partner

amino acid side chain could allow side-tracking into alternative reactions. This would only be possible if the sactionine synthases prove permissive enough for UAAs at the bridging partner residue. We thus tested UAA incorporation via the orthogonal *O*-methyl tyrosine tRNA synthetase/tRNA (*O*-Me-Tyr) pair adapted from *M. janaschii*.

Table 3.1b. Unpublished SboA mutant analysis

Sequence and connectivity of wild-type SboA:

Entry	N-terminal changes	C-terminal changes	No. Bridges
41		I24F, A27W	1 bridge
42	C4A,C6/T7,L13/C14	F21/D22,G28/T29	1 bridge or disulfide
43	A4/C5,C7A	F30/L31	no product
44	C7A,C12/A13	E22/F23	not determined (N.D.)
45	A4/C5,C7A	E22/F23	2 bridges or 1 bridge and 1 disulfide
46	A4/C5,C7A	F20/P22	1 bridge or disulfide
47	C7A,C12/A13	F30/L31	1 bridge or disulfide
48	C2/K4,C7A,C10/G13	F30/L31	1 bridge or disulfide
49	C3/G4,C13A	G31/F32	1 bridge or disulfide
50	S7/C8,C13A	T27/A28	2 bridges or 1 bridge and 1 disulfide
51	S7/C8,C13A	G31/F32	1 bridge or disulfide
52	S7/C8,C13A	G28/T29,L31/F33	N.D.
53	C3/G4,C13A	T27/A28	N.D.
54	C2/K4,C6/T7,C13A	T27/A28	no product
55	C13A	T27/A28	N.D.
56	T4/C6,G7/C10,C13A	G31/F32	N.D.
57	S7/C8,C12/A13	E22/F23,T27/A28	2 bridges or 1 bridge and 1 disulfide
58	C3/G4,C11/A13	I22/F24,G31/F32	2 bridges or 1 bridge and 1 disulfide
59	C3/G4,V13/C15	F21/D22,G28/T29,L31/F33	2 bridges or 1 bridge and 1 disulfide
60	S7/C8,C12/A13	E22/F23,T27/A28	2 bridges or 1 bridge and 1 disulfide
61	C2/K4,C6/T7,L13/C14	F20/P22,G31/F32	no product
62	C3/G4,C11/A13	E22/F23,T27/A28	no product
63		E22/F23,T27/A28	N.D.
64		T26/G28,F30/L31	N.D.
65		F21/D22,G28/T29,L31/F33	1 bridge or disulfide
66	C3/G4,C11/A13	T26/G28,F30/L31	no product
67	C1/N4,C5/A7	F19/I22,T27/A28	2 bridges or 1 bridge and 1 disulfide
68		F20/P22,G31/F32	1 bridge or disulfide
69	G10C	A25T	N.D.
70	G10C	A25F	no product
71	A5_T6insA	G29_L30insA	N.D.
72	deltaG10,deltaA11	deltaA25,deltaG26,	no product

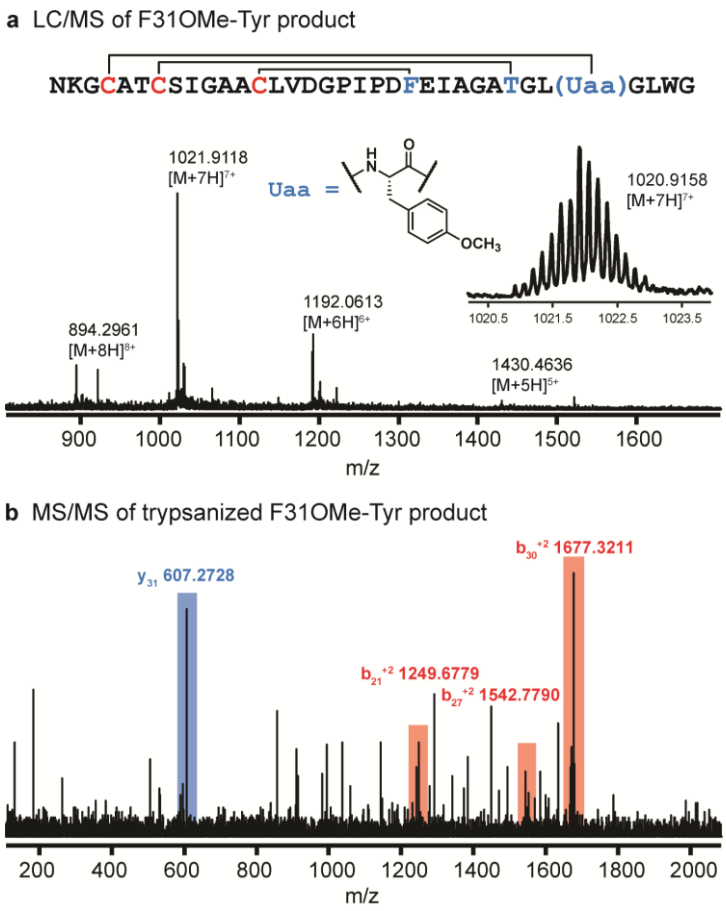


Figure 3.4. Incorporation of unnatural amino acids (UAAs) into sactipeptides in *E. coli*. a) LC-MS of isolate from F31*O*-Me-Tyr mutant, structure of *O*-Me-Tyr, and blow-up of $[M+7H]^{7+}$ envelope.(b) MS-MS of trypsinized F31*O*-Me-Tyr sactipeptide mutant.

pUltra-*O*-Me-Tyr plasmid, replicated by an orthogonal pCDF origin of replication and bearing streptomycin resistance, was incorporated into BL21 (DE3) cells on top of pPH151 and a pETDuet-SboA-AlbA vector mutated to the amber stop codon (TAG) at Phe31. While we could observe unmodified peptide production/UAA incorporation at 200 rpm, this was accompanied by substantial amounts of truncation arising from disrupted translation at the newly incorporated TAG codon. No modified sactipeptide could be found at the lower, 50 rpm shake rate either. We therefore turned to the recently reported"TAGless" *E. coli* strain, C321.ΔA.exp, developed by Isaacs *et al*.[33] In order to drive expression of our same pETDuet-SboA-AlbA, the λDE3 prophage

was integrated into C321.ΔA.exp. After these changes and subsequent growth under our optimized conditions, cyclic precursor peptide exhibiting fully modified cysteine thioether bridges at all three native positions **(Figure 3.4**) was produced. Tandem mass spec. confirmed the $C4^{S}$-(*O*-Me−Tyr)$^{C\alpha}$ linkage at the newly installed UAA position.

3.9 Design, production, and analysis of potential sactipeptide MDM2-p53 inhibitor

With the proclivity of our heterologous expression system to accept changes made to the precursor peptide, we decided to test if grafting a known biologically relevant epitope onto a modified sactipeptide would result in a novel inhibitor or binder to a biologically important protein. We decided to test whether a modified sactipeptide could inhibit the well-known protein-protein interaction (PPI) MDM2-p53. We thought sactipeptide would be an attractive alternative to traditional approaches for targeting and disrupting PPIs due to their propensity for helical secondary structures and their overall larger, stable structure.[34,35] These points can prove useful due to two facts. The first is based on the multiple observations of helices participating in PPIs.[36] While these helices can be present before or after the conformational change upon binding, helices are crucial in the interactions present MDM2-p53 interaction. The second is the fact that PPIs usually encompass a large surface area, $\geq$2000 Å^2.[37] This area cannot be mediated efficiently by traditional small molecules unless it can mimic key residues (hot spots)[36, 38-41], but a larger molecule such as a peptide should prove to be more beneficial in this regard by not only mimicking hot spots but also making other favored interactions throughout the hydrophobic pocket of the interaction. Thus we sought to design sactipeptides specifically to mimic secondary motifs and hot spot residues crucial for binding in MDM2-p53 interaction and therefore creating a natural peptidomimetic.

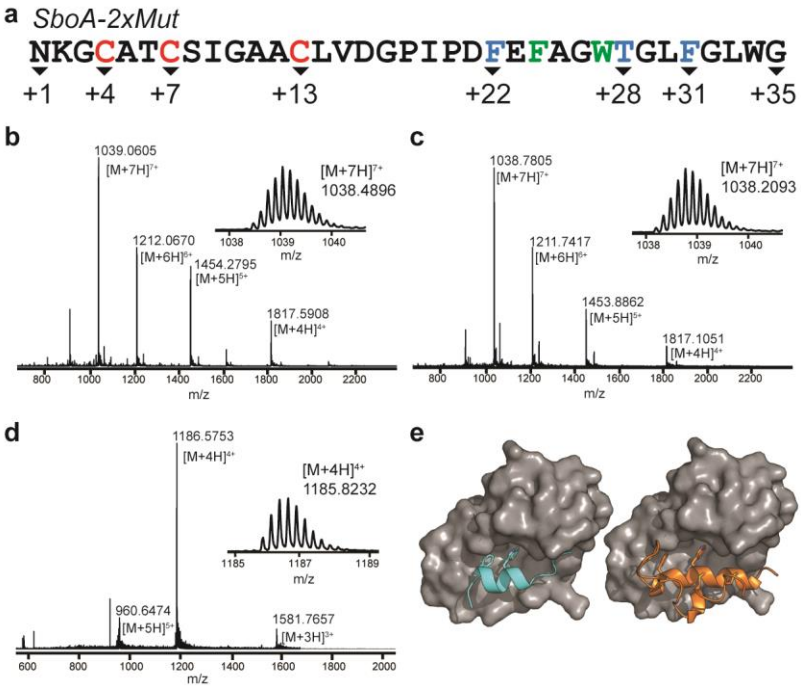
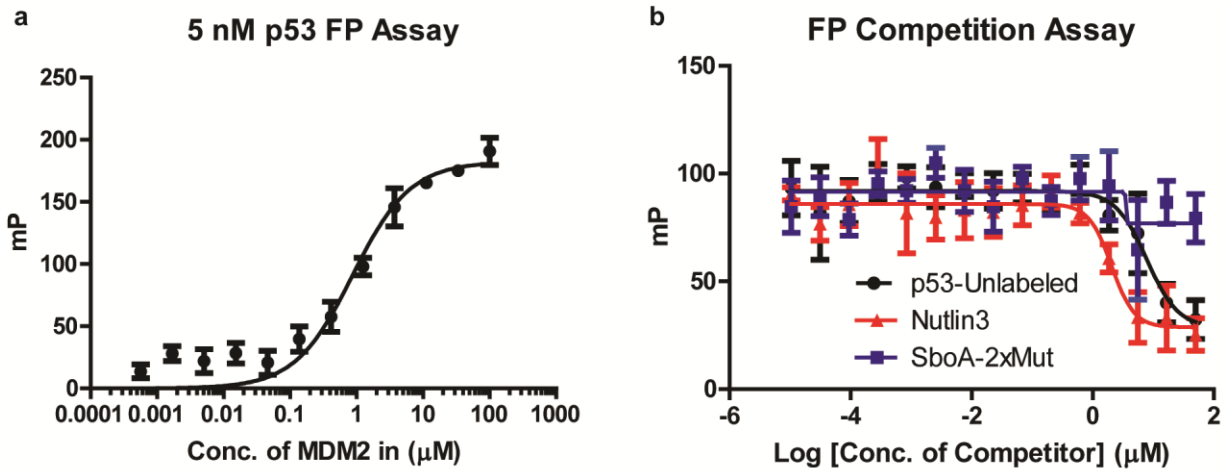


Figure 3.5. Overview of proposed inhibitor SboA-2xMut. a) Sequence of SboA-2xMut with mutated residues in green, corresponding to the Phe, Trp epitope inserted. b-c) Mass spectrum of SboA-2xMut produced in pETDuet system. b was at 200 rpm while c was at 50 rpm. The difference between the two mass spectrums is a loss of 2.0 atomic mass units (amu). d) Mass spectrum of TEV-cleaved SboA-2xMut used in inhibition assays. e) left-crystal structure of p53 interacting with MDM2, right-proposed binding of SboA-2xMut to MDM2.

Phe19 and Trp23 have been characterized as hot spot residues in the MDM2-p53 interaction and thus make a suitable epitope to graft onto a sactipeptide. We did this by mutating Ile24 and Ala27 to Phe and Trp respectively **(Table 3.1b entry 41, and Figure 3.5)**. The peptide was expressed and modified within our heterologous expression system as before and yielded a sactipeptide with only one sactionine linkage **(Figure 3.5b and c)**. The location of this bridge has yet to be determined but we decided to go on with our modified SboA-124F-A27W (referred to from now on as SboA-2xMut). The 6xHis tag was removed by TEV protease and the peptide further purified by HPLC yielding the peptide that would be used in our inhibition assays **(Figure 3.5d)**.



Parameter	Value ± Std. Error
K_d	0.8 ± 0.1 μM
R-square	0.936

Construct	IC_{50} ± Std. Error (μM)
p53-Unlabled	8.0 ± 1.0
Nutlin3	2.0 ± 1.0
SboA-2xMut	Ambiguous

Figure 3.6. Fluorescence Polarization assay of MDM2-p53 interaction. a) FP curve showing p53 truncate (15-29) binding to MDM2 yielding a K_d of ~0.8 μM. b) FP competition assay with unlabled-p53 truncate, nutlin3, and proposed inhibitor SboA-2xMut. IC50 values are shown in the table below the graph.

There are many ways to measure inhibition and binding *in vitro* and while several have been used to measure the inhibition of the p53-MDM2 interaction, fluorescence polarization (FP) and surface plasmon resonance (SPR) were used to measure the inhibitory affect that SboA-2xMur has on the MDM2-p53 interaction.[42,43] A MDM2 truncate (1-188) was expressed and purified from an adapted procedure[41] as a 6xHIS-MBP fusion. Both 5-(and-6)-carboxytetramethylrhodamine-p53 (TAMRA-p53) and Biotin-p53 truncates (residues 15-29) were synthesized using a microwave peptide synthesizer using conventional Fmoc solid phase peptide synthesis (SPPS) **(Experimental 3.11.7)**. The TAMRA labeled peptide was used in the FP assays from an adapted procedure[41], while the Biotin labeled peptide was used and bound to the streptavidin SPR chip for the SPR assays.[42] **In Figure 3.6 and 3.7**, whether by FP or SPR, the results were the same. Both assays show that MDM2 can bind to p53, a control to ensure proper folding of both the MDM2 truncate and the p53. Both the FP and SPR assays gave a

binding constant for p53 to MDM2 that corresponded well to literature values (~1 μM). We were also able to show that nutlin 3 inhibited binding of p53 to MDM2 in both assays, but our proposed inhibitor (SboA-2xMut) showed no binding or affect in either assay. This could be due to only one sactionine bridge being formed in our heterologous expression system. This could inhibit any meaningful secondary structure from forming and prevent the hot spot residues of Phe and Trp from adopting the proper orientation to interact with MDM2 in the p53 binding pocket. Another possibility is the size of SboA-2xMut compare to the p53 truncate. SboA-2xMut is over three times the size of the p53 truncate used in our study. SboA-2xMut, once modified, may adopt a conformation that cannot access the p53 binding pocket due to its large size. More testing and characterization of the system is required to fully realize the goal of sactipeptide inhibitor.

3.10 Other Sactipeptide Systems: Thurincin H and 4BD1

After seeing successful production and modification of variant subtilosin A sactipeptides, we believed that our system could be adapted for other known and predicted sactipeptides. We attempted to produce one known sactipeptide system, Thurincin H **(Figure 3.1b)** and one predicted sactipeptide system from *Bacillus thuringiensis serovar huazhongensis* BGSC 4BD1 in our heterologous expression system and determine whether peptide could be produced and modified by their respective sactionine synthases. A sactipeptide gene cluster has been bioinformatically predicted in *Bacillus thuringiensis serovar huazhongensis* BGSC 4BD1 **(Figure 3.8a)** that houses two copies of its precursor peptide termed 4BD1sg and two predicted sactionine synthases belonging to the radical SAM superfamily of enzymes, named 4BD1-rSAM1 and 4BD1-rSAM2. This system is similar to Thurincin H in that the precursor peptide has four cysteines for a potential of four sactionine bridges **(Figure 3.8b)**. We were able to

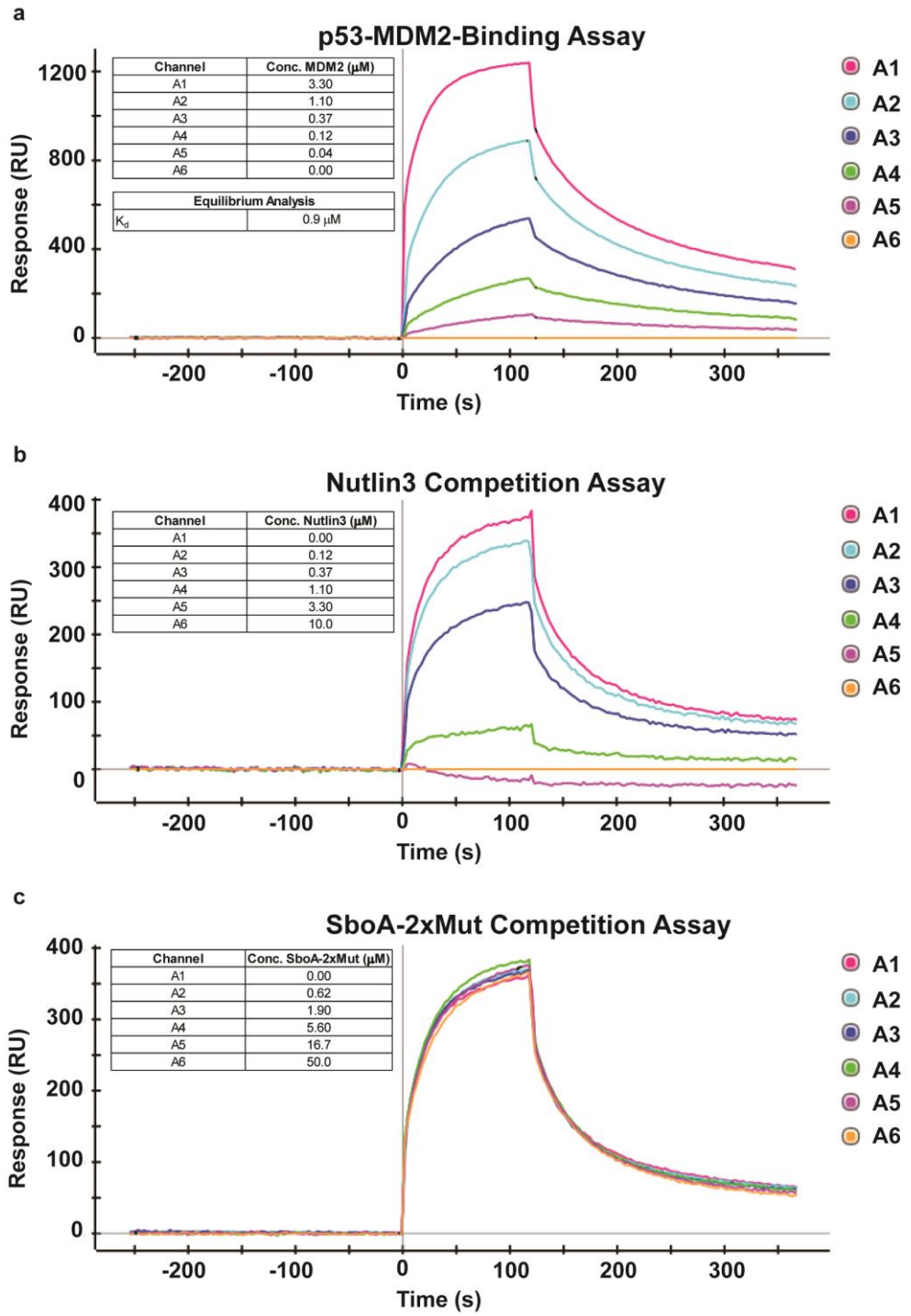
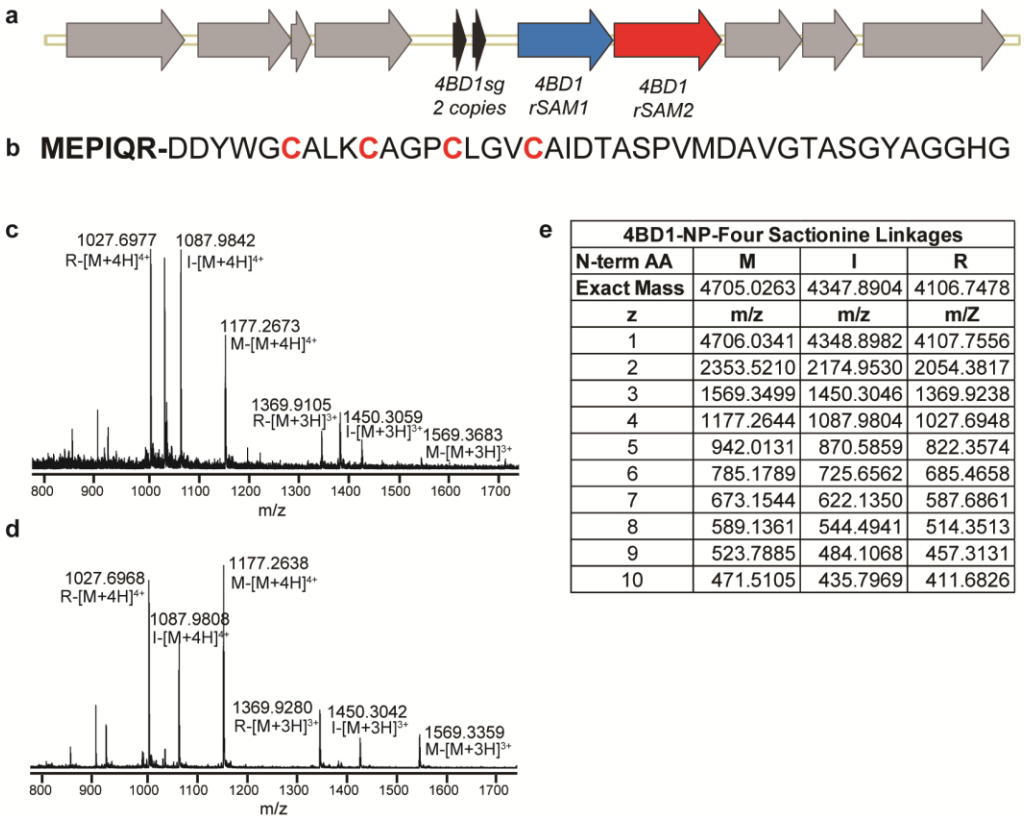


Figure 3.7. Surface Plasmon Resonance binding assays of MDM2-p53 interaction. a) Curves showing p53 truncate (15-29) binding to MDM2 yielding a K_d of ~0.9 µM at equilibrium. b) Curves showing nutlin3 is able to bind MDM2 and prevent p53 truncate binding reducing the response. c) Curves showing the proposed inhibitor SboA-2xMut has no effect on MDM2-p53 truncate binding.

isolate a product from the native producer whose mass corresponded to the precursor peptide of the cluster but with a loss of eight hydrogens. These eight hydrogens correspond to a possible four sactionine bridges formed between the cysteine residues and their yet unidentified bridging partners. This mass was able to be isolated in both the pellet and supernatant of the culture with a variety of truncates at the N-terminus **(see Experimental 3.12.3 and Figure 3.8c-e)**. We termed this isolated product 4BD1-NP for natural product. Experiments are ongoing to determne the placement of the sactionine bridged positions by tandem mass spectrometry.

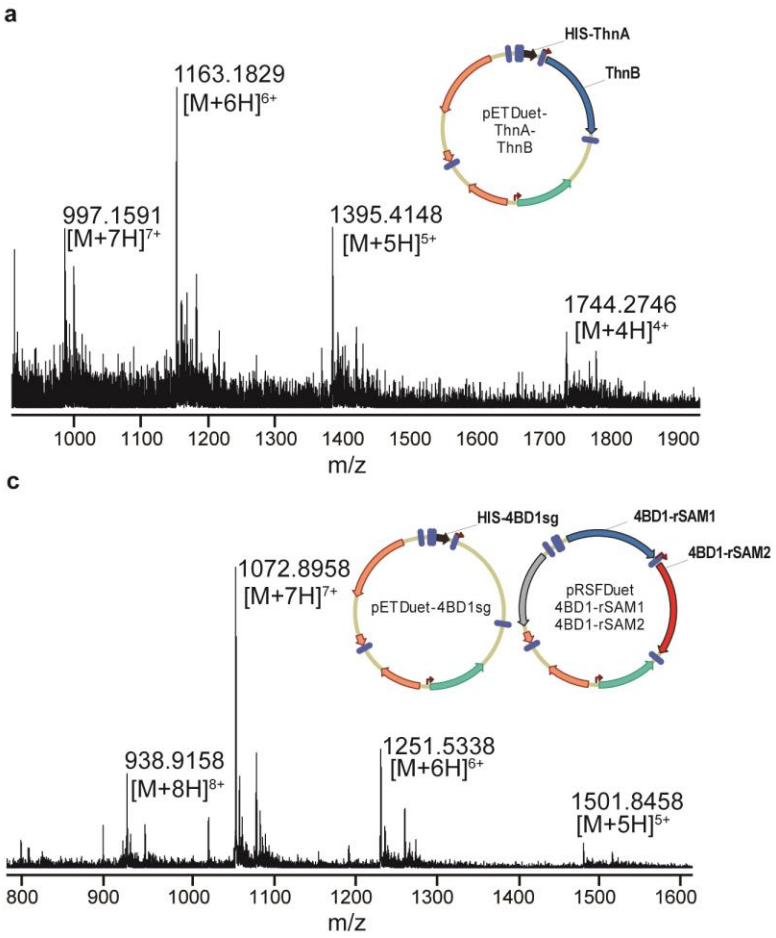


e

4BD1-NP-Four Sactionine Linkages			
N-term AA	**M**	**I**	**R**
Exact Mass	4705.0263	4347.8904	4106.7478
z	**m/z**	**m/z**	**m/Z**
1	4706.0341	4348.8982	4107.7556
2	2353.5210	2174.9530	2054.3817
3	1569.3499	1450.3046	1369.9238
4	1177.2644	1087.9804	1027.6948
5	942.0131	870.5859	822.3574
6	785.1789	725.6562	685.4658
7	673.1544	622.1350	587.6861
8	589.1361	544.4941	514.3513
9	523.7885	484.1068	457.3131
10	471.5105	435.7969	411.6826

Figure 3.8. Isolation of 4BD1 Natural Product (4BD1-NP). a) Predicted sactipeptide gene cluster from *Bacillus thuringiensis serovar huazhongensis* BGSC 4BD1. b) Amino acid sequence of 4BD1sg with cysteines highlighted for formation of sactionine linkages. c) Mass spectrum of 4BD1-NP isolated from the cell pellet showing a range of N-terminal truncates and four sactionine linkages. d) Mass spectrum of 4BD1-NP isolated from the supernatant of the culture showing a range of N-terminal truncates and four sactionine linkages. e) Table of expected masses for potential truncates of 4BD1-NP forming four sactionine linkages.

Next, for both Thurincin H and 4BD1 satipeptide systems, we placed both their precursor peptides and their respective sactionine synthases into our Duet system as described in Experimental 3.12.1. For Thurincin H, we followed a strategy similar to subtilosin A where the precursor peptide, ThnA, was 6xHis-tagged and placed in MCS1 of pETDuet-1 while the sactionine synthase, ThnB, was cloned into MCS2 without an affinity tag **(Figure 3.9a)**. For the 4BD1 sactipeptide system, a second Duet plasmid was employed. We found that by cloning the precursor peptide, 4BD1sg with a 6xHis-tag, into MCS1 of pETDuet-1 and cloning the predicted sactionine synthases 4BD1-rSAM1 and 4BD1-rSAM2 into MCS1 and 2 respectively of pRSFDuet-1 we were able to see production of the peptide **(Figure 3.9b)**. Both systems were transformed and expressed in systems containing the pPH151 corrector plasmid, just as the subtilosin A system described above. However, unlike our previous system, both Thurincin H and the predicted 4BD1 sactipeptide system never yielded fully modified products. During aerobic conditions, both peptides can be isolated with masses indicating two sactionine bridges or two disulfides **(Figure 3.9a-d)**. There are ongoing efforts to definitively answer if these are disulfides are true sactionine bridges, but being as these peptides were produced aerobically, disulfide bonds seem likely to be responsible for the corresponding four hydrogen loss. Unlike our subtilosin A-Duet system, we were unable to see production of peptide at 50 rpm anaerobic conditions. These systems may be less robust than the subtilosin A sactipeptide system and further optimization of expression conditions could remedy this production problem. These systems show promise within our heterologous system and indicate our system can be adopted for producing other sactipeptide modified products.



b

Thurincin H-Duet System			
	No Modifcation	1 Thioether Bridge or 1 Disulfide	2 Thioether Bridges or 2 Disulfides
Exact Mass	6976.1771	6974.1615	6972.1458
z	m/z	m/z	m/z
1	6977.1854	6975.1698	6973.1541
2	3489.0968	3488.0890	3487.0812
3	2326.4006	2325.7288	2325.0569
4	1745.0525	1744.5486	1744.0447
5	1396.2437	1395.8406	1395.4374
6	1163.7044	1163.3685	1163.0326
7	997.6050	997.3170	997.0291
8	873.0304	872.7784	872.5265
9	776.1390	775.9151	775.6911
10	698.6260	698.4244	698.2228

d

4BD1-Duet System			
	No Modifcation	1 Thioether Bridge or 1 Disulfide	2 Thioether Bridges or 2 Disulfides
Exact Mass	7506.3104	7504.2947	7502.2791
z	m/z	m/z	m/z
1	7507.3184	7505.3027	7503.2871
2	3754.1632	3753.1553	3752.1475
3	2503.1114	2502.4395	2501.7677
4	1877.5856	1877.0816	1876.5777
5	1502.2701	1501.8669	1501.4638
6	1252.0597	1251.7238	1251.3878
7	1073.3380	1073.0501	1072.7621
8	939.2968	939.0448	938.7929
9	835.0425	834.8185	834.5945
10	751.6390	751.4374	751.2359

Figure 3.9. Duet systems cloned and expressed for Thurincin H and 4BD1 sactipeptides. a) Mass spectrum of 6xHis-tagged ThnA expressed from pETDuet at 200 rpm. The insert shows the Duet construct used to produce the potential sactipeptde. b) Table of expected masses for 6xHis-ThnA modification. c) Mass spectrum of 6xHis-tagged 4BD1sg expressed from pETDuet at 200 rpm. The insert shows the Duet constructs used to produce the potential sactipeptde. d) Table of expected masses for 6xHis-4BD1sg modification.

3.11 Summary and Discussion

In summary, we have created a system for the heterologous expression of modified sactipeptide derivatives from subtilosin A in *E. coli* and demonstrated unexpected biosynthetic promiscuity of the sactionine synthetase, AlbA. We anticipate that a similar strategy could be utilized to access other predicted sactipeptides or else chemistry from the ever-widening library of rSAM enzymes in RiPP pathways.[44-46] The pronounced promiscuity, especially in the loop region and at unmodified positions on the solvent-exposed exterior of the sactipeptide

macrocycle, could be exploited for the grafting of peptide epitopes as has been seen in lasso peptides and conotoxins, such as integrin binding motifs. We did try and use our promiscuous heterologous expression system, to insert a biologically relevant epitope into our sactipeptide natural products. While the production of a possible MDM2-p53 sactipeptide inhibitor did not yield a functioning peptide, it did indicate and show the potential for the production of natural peptidomimetics by our system. Moreover, we observed for the first time promiscuity with respect to sactionine bridge placement and, in specific instances, bridging partner selection. An expanded understanding of control over sactionine bridge placement could allow for the design of constrained sactipeptide conformations for display of peptide epitopes; as such, our *E. coli* expression system should facilitate such efforts. Results with sactionine bridge movement suggest interdependence between the positioning of these three bridges; this may further be compounded by the overall flexibility of the substrate. Therefore, many more modified sactipeptides may be possible from compound mutants that combine multiple mutations. Again, this is another place where the heterologous expression system may prove beneficial: based on the changes reported here, including the demonstrated mobility of the bridge partners, we estimate libraries far in excess of 10^{12} non-native subtilosin A analogs to be possible with use of the right system. This work suggests that sactipeptides systems are capable of producing vast libraries with significantly high structural complexity. Finally we have also introduced unnatural amino acids (UAAs) into sactipeptides using stop-codon suppression technology, specifically at an otherwise less permissive bridging partner position. This demonstrates the robustness of UAA incorporation in our system and may open up new chemistry and new applications for unnatural sactipeptides.

3.12 Experimental

3.12.1 Cloning and Generation Duet System for the Production of Sactipeptides

LIC cloning precursor peptides genes into pMCSG7

The precursor gene *sboA* was amplified from *B. subtilis* 168 genomic DNA by PCR using Phusion® High-Fidelity Master Mix with HF-buffer following the manufacturer's manual. Primers 1 and 2 were used to accomplish this **(see Table 3.3)**. The purified PCR product was phosphorylated with T4-PNK and then treated with T4-DNA Polymerase to create LICoverhangs. In parallel, pMCSG7 was linearized with SspI and dephosphorylated with Antarctic Phosphatase (AP) and then treated with T4-DNA Polymerase to create LIC overhangs. Treated PCR product and vector were combined and transformed into One-Shot ® Top 10 cells. The plasmid was harvested from cells utilizing the GeneJET Plasmid Miniprep kit following the manual. This process was repeated from cloning *4BD1sg* and *thnA* in to *pMCSG7* as well. *4BD1sg* was amplified from *Bacillus thuringiensis serovar huazhongensis* BGSC 4BD1 genomic DNA by PCR using primers 21 and 22 **(See Table 3.3)** while *thnA* was amplified from a codon optimized template obtained from Integrated DNA Technologies (Gene-Block 1, **Table 3.3**) using primers 23 and 24 **(See Table 3.3)**.

QuikChange Site-Directed Mutagenesis of His-sboA (pMCSG7-SboA)

Site-directed mutagenesis of the *His-sboA* construct was performed following instructions from the QuikChange Site-Directed Mutagenesis kit (Agilent). Briefly, the pMCSG7-*sboA* plasmid was amplified with primers designed to incorporate desired mutations **(see Table 3.3)**. Template plasmid was digested away using DpnI and mutant plasmid was transformed into One-Shot ® Top 10 cells. The mutant plasmids were harvested from cells by GeneJET Plasmid Miniprep kit.

Cloning of sactionine synthases (albA, thnB, 4BD1-rSAM1, and 4BD1-rSAM2) into multiple cloning sites (MCS1 or 2) of Duet plasmids

The gene *albA* was amplified from *B. subtilis* 168 genomic DNA by PCR using Phusion® High-Fidelity Master Mix with HF-buffer following the manufacturer's manual. Primers 19 and 20 were used to accomplish this **(see Table 3.3)**. The purified PCR product and Duet plasmid (pETDuet-1 or pRSFDuet-1) were digested using NdeI and XhoI (MCS2). The purified digested plasmid was treated with AP and then combined with the purified digested PCR product in the presence of T4 ligase. The ligation was allowed to go overnight at 16 ^{o}C. After ligation, the T4 ligase was heat inactivated at 65 ^{o}C for 10 min before being transformed into One-Shot ® Top 10 cells. The plasmid was harvested from cells by GeneJET Plasmid Miniprep kit. This process was repeated for the cloning of *thnB, 4BD1-rSAM1, and 4BD1-rSAM2* into the Duet plasmids as well. The gene *thnB* was amplified by PCR from a codon optimized template obtained from Integrated DNA Technologies (Gene-Block 2, **Table 3.3**) using primers 33 and 34 **(see Table 3.3)**, while both *4BD1-rSAM1, and 4BD1-rSAM2* were amplified from *Bacillus thuringiensis serovar huazhongensis* BGSC 4BD1 genmoic DNA by PCR using primers 29/30 and 31/32 respectively **(see Table 3.3)**. *thnB* was placed in MCS2 of pETDuet-1. *4BD1-rSAM1* was placed in MCS1 of pRSFDuet-1 using NcoI and NotI while *4BD1-rSAM2* was placed in MCS2 of pRSFDuet-1 with NdeI and XhoI after successful cloning of *4BD1-rSAM1*.

Cloning of His-precursor genes (pMCSG7-precursor gene) and mutant sboA into multiple cloning site 1 (MCS1) of Duet plasmids

The construct *His-sboA* was amplified from the pMCSG7-sboA plasmid by PCR using Phusion® High-Fidelity Master Mix with HF-buffer following the manufacturer's manual. Relevant primers are listed in **Table 3.3** below. Primers 16, 17, and 18 were used to amplify

mutants from synthetic genes. The purified PCR product and Duet plasmid with *albA* in MCS2 (pETDuet-1 or RSFDuet-1) were digested using NcoI and HindIII. The purified digested plasmid was treated with AP and then combined with the purified digested PCR product in the presence of T4 ligase. The ligation was allowed to incubate overnight at 16 ^{o}C. After ligation, the T4 ligase was heat inactivated at 65 ^{o}C for 10 min before being transformed into One-Shot ® Top 10 cells. The plasmid was harvested from cells by GeneJET Plasmid Miniprep kit. The above process was used for *His-thnA* and *His-4BD1sg* into MCS1 of pETDuet-1. Primers 27 and 28 were used for *His-thnA* and Primers 25 and 26 were used for *His-4BD1sg* **(see Table 3.3)**.

Generation of mutant sactipeptide library

All other mutants were purchased from General Biosystems. Mutants were based on the template: 5'- CAA TCC AAT GCG ATG AAA AAA GCT GTC ATT GTA GAA AAC AAA GGT TGT GCA ACA TGC TCG ATC GGA GCC GCT TGT CTA GTG GAC GGT CCT ATC CCT GAT TTT GAA ATT GCC GGT GCA ACA GGT CTA TTC GGT CTA TGG GGG TAA AAG CTT-3'. Swaps were generated by directly swapping the codons at the relevant positions. Alternatively, the following codons, adopted from other sites in *sboA*, were used for direct replacement of a native codon: Phe-TTT, Met-ATG, Ala-GCC, Ser-TCG, Thr-ACA, Asn-AAC. The mutant gene templates received from General Biosystems were then amplified using relevant primers **(see Table 3.3)**. These purified PCR constructs were used as templates for the cloning into Duet-MCS1 described above.

Table 3.3. Plasmids, sites, and primers used in pETDuet-SboA-AlbA heterologous system

LIC-plasmid	pMCSG7	SspI
MCS1-Duet	pETDuet, RSFDuet	NcoI, HindIII for all expcept 4BD1-rSAM1-MCS1, then use NotI
MCS2-Duet	pETDuet, RSFDuet	NdeI, XhoI
Primer 1	pMCSG7-SboA-F	5'-TAC TTC CAA TCC AAT GCG ATG AAA AAA GCT GTC ATT GTA GAA AAC AAA GG-3'
Primer 2	pMCSG7-SboA-R	5'-TTA TCC ACT TCC AAT GCG CTA TTA CCC CCA TAG ACC GAA TAG ACC T-3'
Primer 3	Quick-change-SboA-C3/G4 swap-F	5'-ATT GTA GAA AAC AAA TGT GGT GCA ACA TGC TCG ATC-3'
Primer 4	Quick-change-SboA--C3/G4 swap-R	5'-GAT CGA GCA TGT TGC ACC ACA TTT GTT TTC TAC AAT-3'
Primer 5	Quick-change-SboA--S7/C8 swap-F	5'-AAA GGT TGT GCA ACA TCG TGC ATC GGA GCC GCT TGT-3'
Primer 6	Quick-change-SboA--S7/C8 swap-R	5'-ACA AGC GGC TCC GAT GCA CGA TGT TGC ACA ACC TTT-3'
Primer 7	Quick-change-SboA-C12/A13 swap-F	5'-TGC TCG ATC GGA GCC TGT GCT CTA GTG GAC GGT CCT-3'
Primer 8	Quick-change-SboA-C12/A13 swap-F	5'-AGG ACC GTC CAC TAG AGC ACA GGC TCC GAT CGA GCA-3'
Primer 9	His-SboA-into Duet-MCS1-F	5'-GAT CGA TCC CAT GGG CAT GCA CCA TCA TCA TCA TCA TTC TT-3'
Primer 10	His-SboA-into Duet-MCS1-R	5'-GAT CGA TCA AGC TTT TAC CCC CAT AGA CCG AAT AGA CCT-3'
Primer 11	His-SboA-P18A into Duet-MCS1-R	5'-GAT CGA TCA AGC TTT TAC CCC CAT AGA CCG AAT AGA CCT GTT GCA CCG GCA ATT TCA AAA TCA GCG ATA GGA CCG TC-3'
Primer 12	His-SboA-P20A into Duet-MCS1-R	5'-GAT CGA TCA AGC TTT TAC CCC CAT AGA CCG AAT AGA CCT GTT GCA CCG GCA ATT TCA AAA TCA GGG ATA GCA CCG TC-3'
Primer 13	His-SboA-P18A and P20A into Duet-MCS1-R	5'-GAT CGA TCA AGC TTT TAC CCC CAT AGA CCG AAT AGA CCT GTT GCA CCG GCA ATT TCA AAA TCA GCG ATA GCA CCG TC-3'
Primer 14	His-SboA-F31TAG into Duet-MCS1-R	5'-GAT CGT ACA AGC TTT TAC CCC CAT AGA CCC TAT AGA CCT GT-3'
Primer 15	SboA-add His into Duet-MCS1-F-Library	5'-GAT CGA TCC CAT GGG CAT GCA CCA TCA TCA TCA TCA TTC TTC TGG TGT AGA TCT GGG TAC CGA GAA CCT GTA CTT CCA ATC CAA TGC GAT GAA A-3'

Primer 16	His-SboA-into Duet-MCS1-F-Library	5’-GAT CGA TCC CAT GGG CAT GCA CCA T-3’
Primer 17	His-SboA-into Duet-MCS1-R-Library-1	5’-GAT CGT ACA AGC TTT TAC CCC CAT AG-3’
Primer 18	His-SboA-into Duet-MCS1-R-Library-2	5’-GAT CGT ACA AGC TTT TAC CCC CA-3’
Primer 19	AlbA-into Duet-MCS2-F	5’-GAT CGA TCC ATA TGT TGT TTA TAG AGC AGA TGT TTC CAT-3’
Primer 20	AlbA-into Duet-MCS2-R	5’-GAT CGA TCC TCG AGC TAA ATA AGC TGG ACC ACG TCT TC-3’
Primer 21	pMCSG7-4BD1sg-F	5’-TAC TTC CAA TCC AAT GCG ATG GAA CCA ATT CAA CGT GA-3’
Primer 22	pMCSG7-4BD1sg-R	5’-TTA TCC ACT TCC AAT GCG CTA TTA ACC ATG TCC TCC AGC AT-3’
Primer 23	pMCSG7-ThnA-F	5’-TAC TTC CAA TCC AAT GCG ATG GAA ACA CCA GTA GTA CAA-3’
Primer 24	pMCSG7-ThnA-R	5’-TTA TCC ACT TCC AAT GCG CTA TTA GCT TGC AGT ACT AGC CCC TGT-3’
Primer 25	His-4BD1sg-into Duet-MCS1-F	5’-GAT CGA TCC CAT GGG CAT GCA CCA TCA TCA TCA TCA TTC TT-3’
Primer 26	His-4BD1sg-into Duet-MCS1-R	5’-GAT CGA TCA AGC TTT TAA CCA TGT CCT CCA GCA T-3’
Primer 27	His-ThnA-into Duet-MCS1-F	5’-GAT CGA TCC CAT GGG AAT GCA CCA TCA TCA TCA TCA TTC TTC TGG T-3’
Primer 28	His-ThnA-into Duet-MCS1-R	5’-GAT CGA TCA AGC TTT TAG CTT GCA GTA CTA GCC CCT GT-3’
Primer 29	4BD1-rSAM1 into Duet-MCS1-F	5’-GAT CGA TCC CAT GGG CAT GCA AAC TGG TAC TGC TAA AG-3’
Primer 30	4BD1-rSAM1 into Duet-MCS1-R	5’-GAT CGA TCG CGG CCG CTT ACA TAT TTA ATA TAT CAT-3’
Primer 31	4BD1-rSAM2 into Duet-MCS2-F	5’-GAT CGA TCC ATA TGA TGT TCA TGA ATA AAT ACT T-3’
Primer 32	4BD1-rSAM2 into Duet-MCS2-R	5’-GAT CGA TCC TCG AGT TAA GCA TAA GAG CAT AAT G-3’
Primer 33	ThnB into Duet-MCS1-F	5’-GAT CGA TCC ATA TGA TGA ATG GTT ACC TGT TTT GGA AGG AAA AAC TGG AAA TCC G-3’
Primer 34	ThnB into Duet-MCS1-R	5’-GAT CGA TCC TCG AGC TAA TTC TGA CCA ACC ATG AGT GGC TCT TT-3’

Gene-Block 1	ThnA	5’-ATG GAA ACA CCA GTA GTA CAA CCA AGG GAT TGG ACT TGT TGG AGT TGC TTA GTA TGT GCA GCA TGT TCT GTG GAA TTA TTA AAT TTA GTT ACT GCG GCA ACA GGG GCT AGT ACT GCA AGC TAA -3’
Gene-Block 2	ThnB	5’-ATG AAT GGT TAC CTG TTT TGG AAG GAA AAA CTG GAA ATC CGC AAA TTT ACC TCT AAT TAC GAG AGC ATG CTG GTA GTT CAT AAA AAC CCT AAC GAA AGC GCA CCG ACC CTG AAG AAT GAG AAT ACG TTT ACG ATC AAC AAG ACG GCG ACC GAA ATC ATT GAA TTA ATT GAT GGC ACA AAG ACG TAC GGC CAA GTG GTG AGT TTT CTT TCA CTG AAA TAC TCC GAA GAT CCT ATC AGT ATT GAA AAG AAA CTG AAC GCC TTT CTT AAC AAC GTG TCC AAA GTG TAC AAC ATG AAT ATC GGC ACG CAA GAA GAA CCT ATT AAT GTT CCG GTG AAC CTG ATC GAG GAA CAG ACC ATT TAC CCA AAA GTG GCT AGC ATT GAG ATT ACA AAT CGC TGC AAC GTT CGT TGT CGC CAC TGC TAT GGG GAT TTC GGC GCG GTA AAA CCG AAG GTT ATG TCG CTG GAT CAG ATC AAA TCG CTT TTA GAT GAT CTG AAC AAC ATT GGA GTT AAA CTG ATC GAA CTG ACG GGC GGT GAT ATC ACT GTT CAC CCG AAT TTA AAA GAA ATC CTG CTG TAT GCC CTG AAT CTG GAT TTC AGC CAG ATT ACA TTA CTG ACA AAC GGT ATT GCC CTG AGT GAT AAA GTC ATG GAC ATT ATC ATT AAA AAC AAA AGT AAA ACT TTC GTC CAA ATT GAT ATG CAC TCT TTA GAT GAT AAC TAC CTG ACG TGG TTC TTC AAA GTA CCG AAC ACA CTG CAT AAA ATC AAA AAC AAT ATT ATG AAG TTG GCA GAA AAT GAC GTG CGT CTT CGT ATT GCT ACG ATC GTT ACC CAT CTG AAT GTT CAT GAA GTG GAA GAC ATC GCC GAA TGG GTT CAT AAC CTG GGC ATT GAT TCT ATC GGT GTG AGC CCG GTG ATC CCG ATG GGG CGC GCG CTG GGC TGC AGC GAC CTC TAT CTT AAC GAG GAA GAT GTC AAA ACC TAC GGC GAA GCT CTG TTA AAG ATT AAT AAG AAA TAT CCG AAA TTCG TCT CTT TAT ATG AGG GCG CTC GTG CAG AGA TTC GCA ATT GTG GTG CCA TCA CTA GTC ATA TTG TAA TCG CAC CGG ATG GCG AAA TTA AGA TGT GCA CTA TGC ATA GTC TGG ATG ATC TCA AAA ACT CCA TTG GTA ATG TGT TTG AAC AGA ACA TCA AAG ACA TTT ATG ATG AAA AAT TCA AGT ACA TCA ATG CCT TCT TTA ACC TGC AAG CTC CGC AGA TGG ATA GTG AAG AAT GCA AAG AGT GCG AAA ATA AGC GTT TCT GCA GTA CCT GCT TTT TGC GCA GTT TCA TTA AAG CGC AGG AAA TTG GCG ATA AAT GCA AAT GGT TCA AGA ACC ATG TGC CTG AAA TTA TCA AAG AGC CAC TCA TGG TTG GTC AGAAT TAG-3’

3.12.2 Expression and purification of sactipeptides from Duet System

Production of cell lines and expression of sactipeptides from Duet system

pPH151/BL21 DE3, Duet-*sboA*-*alba*/pPH-151/Bl21 DE3, and Duet-*sboA*-F31TAG-*alba*/pPH151/pUltra-*O-Me-Tyr-aaRS*/TAGless DE3, pETDuet-*thnA*-*thnB*/pPH-151/Bl21 DE3, pETDuet-*4BD1sg*/pRSFDuet-*4BD1-rSAM1,4BD1-rSAM2*/pPH-151/Bl21 DE3 cell lines were

created by transforming plasmids into electrocompetent cells generated according to standard molecular biology protocols found in Green *et al.*[47] The electroporation was carried out in a 0.1cm cuvette, at 1.8 kV, 200 Ω, and 20 μFD. His-tagged sactipeptides were heterologously expressed in *E. coli* (pPH151/BL21 DE3) cells in either 100 mL or 1 L of LB media. Media was supplemented with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) and either 0.5 mL (100 mL grow-up) or 5 mL (1 L grow-up) of an overnight culture. Cultures are then grown at 37 ^{o}C and 200 rpm to an OD_{600}~0.6-0.7, at which point IPTG was added to a final concentration of 0.5 mM and the culture was grown at 18 ^{o}C for 22-24 hours. For unmodified peptides the shake was left at 200 rpm but for modified peptides the shake was dropped down to 50 rpm.

Purification of His-tagged sactipeptides from inclusion bodies (IB) (100mL scale)

Modified from Li *et al*.[29] Cell pellet was suspended in 1 mL of Start Buffer (20 mM $Na_x(PO_4)_y$ pH=7.5, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and then sonicated. Cell debris was pelleted at 15,000 rpm, 4 ^{o}C, for 10 minutes. The supernatant was removed and the cell pellet was washed with 0.5 mL of IB Buffer (20 mM $Na_x(PO_4)_y$ pH=7.5, 6 M Guanidinium HCl, 500 mM NaCl, 50 mM imidazole) and vortexed to ensure the pellet was washed thoroughly. Cell debris was pelleted again and the supernatant was collected. Using HisPur^{TM} Ni-NTA Spin Columns-0.2 mL (Thermo Scientific), the supernatant was incubated in contact with the resin for 30 minutes at a gentle shake at 4 ^{o}C. After incubation, the spin column was spun at 2000 x g for 2 minutes at 4 ^{o}C. The flow through was discarded and the resin washed twice with 0.4 mL of IB buffer twice while spinning and removing flow through. The resin was washed with 0.4 mL of water and then eluted with 0.3-0.4 mL of Elution Buffer (25 mM Tris/HCl pH=8.0, 150 mM NaCl, 500 mM imidazole). The elution was used as is in other experiments.

Purification of His-tagged sactipeptides from Duet system (1L scale)

Modified from Li *et al.*[29] Cell pellet was suspended in 30 mL of Start Buffer supplemented with 0.5 mL of 25 mg/mL T4 lysozyme and 0.5 mL of 150 mM PMSF and then incubated on ice for 10 minutes. The cell pellet was then sonicated. The cell pellet was pelleted at 15,000 rpm at 4 ^{o}C for 10 minutes. The supernatant was then removed and the pellet was washed and suspended in 30 mL of IB Buffer. The cell pellet was scrapped and broken up by spatula and using the vortexer. The cell debris was pelleted again and the supernatant collected and filtered through a 0.45 μm syringe filter. The flow through from the filter was then passed over a Ni^{2+} IMAC column (HISTrap^{TM} HP 5mL GE Healthcare) utilizing a FPLC (NGC-Quest-10 Bio-Rad). The Ni^{2+} IMAC column was washed with 6 column volumes (CV) of IB Buffer and 2.5 CV of water. To elute the peptide off the Ni^{2+} IMAC column, a binary elution gradient of 0-100% of Elution buffer (water the other solvent) over 10 CV was used. The peptide came off between 30-40% Elution Buffer. The elution was used as is in other experiments.

3.12.3 Production and Extraction of 4BD1-Natural Product (NP)

The 4BD1-natural product (4BD1-NP) sactipeptide was produced by inoculating 1 L of LB media with 1 mL of a previous overnight of *Bacillus thuringiensis serovar huazhongensis* BGSC 4BD1. This culture was allowed to grow at 37 ^{o}C and 200 rpm overnight. After ~24 hours of growth, the cells were separated from the supernatant by centrifugation (5000 rpm, 4 ^{o}C, 20 minutes) and the supernatant collected. The natural product was extracted from the pellet with ~10 mL of methanol and breaking the pellet up with vigorous shaking and vortexing. Then the methanol was allowed to incubate in contact with the pellet with gentle agitation at room temperature for two hours. The cell debris was removed by centrifugation and the methanol supernatant collected. The methanol supernatant was concentrated in a rotary evaporator and a

small sample was diluted and analyzed by mass spectrometry with the method listed below. To the first supernatant collected (spent LB media), this was passed over a small C18-SepPak® Plus column (Waters, WAT023635) with the use of a syringe. Once all the spent LB had been passed over the column and the flow-through discarded, the column was washed once with 10 mL nuclease free water followed by elution of the natural product in 5 mL of methanol. A small sample was diluted and analyzed by mass spectrometry with the method listed below

3.12.4 NEM Modification of free cysteines

Used a procedure adapted from Thibodeaux *et al.*[48] First, the sactipeptide solutions were changed to a pH~6.5-7 by the addition of HCl. Then the sactipeptide solutions were incubated with 16.67 mM TCEP at 25 ^{o}C and 500 rpm for 10 minutes. Then NEM (in ethanol) was added to a final concentration of 10 mM. The reaction was carried out at 37 ^{o}C and 500 rpm for 10 minutes.

3.12.5 Incorporation of unnatural amino acids into Duet System

His-SboA-F31OMe-Tyr was heterologously expressed in *E. coli* (pPH151/pUltra-*O-Me-Tyr aaRS*/TAGless DE3) cells. The C321.ΔA.exp cell line (https://www.addgene.org/49018/) was modified to allow expression of constructs *via* the T7 promoter. Site specific integration of λDE3 prophage was performed with λDE3 Lysogenization Kit (Novagen). Sactipeptide expression was carried out as above with the following exceptions: 1) media was supplemented with ampicillin (100 μg/mL) chloramphenicol (34 μg/mL) and spectinomycin (100 μg/mL); 2) in addition to IPTG, when OD_{600} reached 0.6, *O*-Me-Tyr in 0.1 M NaOH was added to a final concentration of 2 mM. To limit the amount of base added, minimal amounts were used to dissolve the unnatural amino acid (2 mL for a 100 mL culture, 15 mL for a 1 L culture). All other purification and analysis as above.

3.12.6 Characterization of sactipeptides by mass spectrometry

Trypsin digestion of sactipeptide

An aliquot of 300-600 µl of sactipeptide solution was digested by adding 2 µl of 10 mg/ml trypsin in 0.1 mM HCl. The digestion was carried out at 37 ^{o}C and 500 rpm for 0.75-2 hours. 800 µl of methanol was added and the solution was pelleted at 15,000 rpm and 4 ^{o}C for 5 minutes. The supernatant was collected and the solvent removed by speed-vac till the volume was 100-200 µl. This solution was injected onto the LC-MS.

LC-MS and MS-MS methods

LC-MS analysis was performed on an Agilent 1200 HPLC system coupled to a 6520 Accurate-Mass Q-TOF spectrometer with an electrospray ionization (ESI) source in positive ion mode. LC used a 2.1 mm x 50 mm Phenomenex® Kinetex C18 column (100 Å, 2.6 µm) with the gradient program described below. Water with 0.1 % formic acid (FA) was used as solvent A and acetonitrile with 0.1% FA was used as solvent B. Analytes were eluted directly into the MS and spectra were acquired in profile mode using a gas temperature of 350 ^{o}C and a fragmentor voltage of 250 V.

Flow Rate	0.5 ml/min
Post time	4 min
Time (min)	**%B**
0.00	2
2.00	2
15.00	100
16.00	100
18.01	2

When running tandem MS-MS, the +3 (Z) state of the trypsin digest product (K-6) was used as the target ion. A specific retention time for each ion was collected and used for the MS-MS with a delta time of 0.5 min and an isolation width of 1.3 m/Z. When only one bridge was

produced, the +3 (Z) state of the trypsin digest product (K-6) of the doubly n-ethylmaleimide (NEM) modified product was targeted. A collision energy of 25-35 eV was used to obtain MS/MS spectra.

3.12.7 Synthesis of $p53^{15-29}$ and derivatives by Solid Phase Peptide Synthesis

All syntheses were carried out by microwave assisted solid-phase peptide synthesis (SPPS). Unlabled-p53 truncate (aa. 15-29), N-terminal TAMRA- p53 truncate (aa. 15-29), and N-terminal Biotin- p53 truncate (aa. 15-29) were all synthesized for FP and SPR experiments. ChemMatrix solid support (0.47 mmol/g) on a 0.047 mmol scale was used. The solid support was initially swollen in DMF (1.5 mL) for 20 min at 70 °C. Fmoc-Amino Acids-OH (0.5 M in DMF), 5,6 TAMRA-OH (0.5 M in DMF, VWR) or Biotin, free acid (0.5 M in DMF) at 5.0 - 10.0 equiv were coupled with HATU (0.2 M in DMF) at 6.86 equiv., and DIEA (0.2 M in DMF) at 14.0 equiv. The reagents were added to the swollen resin in the above. The resulting suspension was heated under microwave irradiation for 5 min at 75 °C. The reaction vessel is then drained and resin is thoroughly washed with DMF four times. Removal of the Fmoc protecting group was accomplished after amino acid coupling using excess 20% piperidine. 20% piperidine was added to the reaction vessel and allowed to incubate at RT for 3 min with constant stirring. The reaction vessel was then drained, washed with DMF and excess 20% piperidine was again added and the reaction was incubated for another 10 min at RT. The reaction vessel was then drained and the resin thoroughly washed with DMF four times. After washing the resin with DCM, the resin was dried and cleaved using the standard cleavage cocktail (TFA/TIPS/H2O, 95:2.5:2.5) to yield the fully deprotected peptides. The peptides were concentrated by precipitation with cold diethyl ether. Preparative HPLC was performed for purification of the $p53^{15-29}$ peptide and its derivatives. Relevant fractions were collected and

partially concentrated in a rotary evaporator, followed by flash freezing and lyophilization to obtain the purified solid product. The identities were confirmed by LC-MS by the method described above.

3.13 REFERENCES

1. Jarrett, J. T. *J. Biol. Chem.* **2015**, *290*, 3972

2. Sit, C. S.; van Belkum, M. J.; McKay, R. T.; Worobo, R. W.; Vederas, J. C. *Angew. Chem. Int. Ed.* **2011**, *50*, 8718

3. Flühe, L.; Marahiel, M. A. *Curr. Opin. Chem. Biol.* **2013**, *17*, 605

4. Rea, M. C.; Sit, C. S.; Clayton, E.; O'Connor, P. M.; Whittal, R. M.; Zheng, J.; Vederas, J. C.; Ross, R. P.; Hill, C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9352

5. Sit, C. S.; McKay, R. T.; Hill, C.; Ross, R. P.;Vederas, J. C. *J. Am. Chem. Soc.* **2011**, *133*, 7680

6. Babasaki, K.; Takao, T.; Shimonishi, Y.; Kurahashi, K. J. *Biochem.* **1985**, *98*, 585

7. Kawulka, K. E.; Sprules, T.; Diaper, C. M.; Whittal, R. M.; McKay, R. T.; Mercier, P.; Zuber, P.;Vederas, J. C. *Biochem.* **2004**, *43*, 3385

8. Liu, W.T.; Yang, Y.L.; Xu, Y.; Lamsa, A.; Haste, N. M.; Yang, J.Y.; Ng, J.; Gonzalez, D.; Ellermeier, C. D.; Straight, P. D.; Pevzner, P. A.; Pogliano, J.; Nizet, V.; Pogliano, K.;Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16286

9. Thennarasu, S.; Lee, D.K.; Poon, A.; Kawulka, K. E.; Vederas, J. C.;Ramamoorthy, A. *Chem. Phys. Lipids.* **2005**, *137*, 38

10. Wang, G.; Feng, G.; Snyder, A. B.; Manns, D. C.; Churey, J. J.;Worobo, R. W. *FEMS Microbiol. Lett.* **2014**, *357*, 69

11. 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.; Muller, 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.* **2012**, *30*, 108

12. Wieckowski, B. M.; Hegemann, J. D.; Mielcarek, A.; Boss, L.; Burghaus, O.; Marahiel, M. A. *FEBS Lett.* **2015**, *589*, 1

13. Flühe, L.; Knappe, T. A.; Gattner, M. J.; Schäfer, A.; Burghaus, O.; Linne, U.; Marahiel, M. A. *Nat. Chem. Biol.* **2012**, *8*, 350

14. Flühe, L.; Burghaus, O.; Wieckowski, B. M.; Giessen, T. W.; Linne, U.; Marahiel, M. A. *J. Am. Chem. Soc.* **2013**, *135*, 959

15. Grell, T. A. J.; Goldman, P. J.; Drennan, C. L. *J. Biol. Chem.* **2015**, *290*, 3964

16. Wang, G.; Manns, D. C.; Churey, J. J.; Worobo, R. W. *Appl. Environ. Microbiol.* **2014**, *80*, 3576

17. Deane, C. D.; Melby, J. O.; Molohon, K. J.; Susarrey, A. R.; Mitchell, D. A. *ACS Chem. Biol.* **2013**, *8*, 1998

18. Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. *J. Am. Chem. Soc.* **2010**, *132*, 7519

19. Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2011**, *133*, 5016

20. Tianero, M. D. B.; Donia, M. S.; Young, T. S.; Schultz, P. G.; Schmidt, E. W. *J. Am. Chem. Soc.* **2012**, *134*, 418

21. Bowers, A. A.; Acker, M. G.; Young, T. S.; Walsh, C. T. *J. Am. Chem. Soc.* **2012**, *134*, 10313

22. Young, T. S.; Dorrestein, P. C.; Walsh, C. T. *Chem. Biol.* **2012**, *19*, 1600

23. Li, C.; Zhang, F.; Kelly, W. L. *Mol. BioSyst.* **2011**, *7*, 82

24. Shi, Y.; Yang, X.; Garg, N.; van der Donk, W. A. *J. Am. Chem. Soc.* **2011**, *133*, 2338

25. Piscotta, F. J.; Tharp, J. M.; Liu, W. R.; Link, A. J. *Chem. Commun.* **2015**, *51*, 409

26. Budisa, N. *Curr. Opin. Biotechnol.* **2013**, *24*, 591

27. Al Toma, R. S., Kuthning, A., Exner, M. P., Denisiuk, A.; Ziegler, J.; Budisa, N.; Süssmuth, R. D. *Chem Bio Chem*. **2015**, *16*, 503

28. Bindman, N. A.; Bobeica, S. C.; Liu, W. R.; van der Donk, W. A. *J. Am. Chem. Soc.* **2015**, *137*, 6975

29. Li, B.; Cooper, L. E.; van der Donk, W. A. *Methods Enzymol.* **2009**, *458*, 533

30. Wecksler, S. R.; Stoll, S.; Tran, H.; Magnusson, O. T.; Wu, S.P.; King, D.; Britt, R. D.; Klinman, J. P. *Biochem.* **2009**, *48*, 10151

31. Hänzelmann, P.; Hernández, H. L.; Menzel, C.; García-Serres, R.; Huynh, B. H.; Johnson, M. K.; Mendel, R. R.; Schindelin, H. *J. Biol. Chem*. **2014**, *279*, 34721

32. Lohans, C. T.; Vederas, J. C. *J. Antibiot*. **2014,** *67*, 23

33. Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. *Science*. **2013**, *342*, 357

34. Ding, W.; Li, Y.; Zhang, Q. *ACS Chem. Biol.* **2015**, *10*, 1590

35. Tanaka, M. *Chem. Pharm. Bull.* **2007**, *55*, 349

36. Jochim, A. L.; Arora, P. S. *ACS Chem. Biol*. **2010**, *5*, 919

37. Smith, M. C.; Gestwicki J. E. *Expert Rev. Mol. Med*. **2012**, *14*, 1

38. Wang S.; Zhao, Y; Bernard, D.; Aguilar, A; Kumar, S. *Top Med. Chem*. **2012**, *Chapter 8*, pgs. 57-80

39. Chene ,P. *Nat. Rev. Cancer*. **2003**, *3*, 102

40. Watkins, A. M.; Arora, P. S. *Eur. J. Med. Chem*. **2015**, *94*, 480

41. Lao, B. B.; Drew, K. Guarracino, D. A.; Brewer, T. F.; Heindel, D. W.; Bonneau, R.; Arora, P.S. *J. Am. Chem. Soc*. **2014**, *136*, 7877

42. Reed, D.; Shen, Y.; Shelat, A. A.; Arnold, L. A.; Ferreira, A. M.; Zhu, F.; Mills, N.; Smithson, D. C.; Regni, C. A.; Bashford, D.; Cicero, S. A.; Schulman, B. A.; Jochemsen, A. G.; Guy, R.K.; Dyer, M. A. *J. Biol. Chem*. **2010**, *285*, 10786

43. Smith, J. M.; Frost, J. R.; Fasan, R. *Chem. Comm*. **2014**, *50*, 5027

44. Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. *Nat. Chem.* **2015**, *7*, 431

45. Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H.G.; Matsunaga, S.; Piel, J. *Science.* **2012**, *338*, 387

46. Murphy, K.; O'Sullivan, O.; Rea, M. C.; Cotter, P. D.; Ross, R. P.; Hill, C. *PLoS One*, **2011**, *6*, e20852

47. Green, M. R., and Sambrook, J. *Molecular cloning: a laboratory manual 4th ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. **2012**

48. Thibodeaux, C.J.; Ha, T.; van der Donk, W.A. *J. Am. Chem. Soc.* **2014**, *136*, 17513

CHAPTER 4

PROBING THE PROMISCUITY OF RIPPS ENZYMES USING MRNA DISPLAY TECHNOLOGIES AND NEXT-GENERATION SEQUENCING

4.1 Introduction

From its initial conception and subsequent publication in 1997, mRNA display has become a robust *in vitro* selection method that has been used to select peptides and protein with designed and desired properties.[1-4] The power of mRNA display has been exploited many times over its ~20 years as an *in vitro* selection and display technology. It has been used to address and answer a wide range of complicated and diverse biochemical questions and problems. Being as it is a natural extension of directed evolution and its high enrichment fidelity, mRNA display has been used to a) identify important residues required for protein binding and affinity and improve upon them, b) elucidate important enzyme-substrate interactions, c) identify ligands for receptors, d) identify drug-binding targets, and e) evolve novel protein sequences with unique enzymatic activities.[5-24] mRNA display works by covalently linking the translated protein or peptide to the mRNA that encodes the amino acid sequence, essentially connecting genotype to phenotype. The connection between the C-terminal portion of the polypeptide chain and the 3'-end of the mRNA occurs because of a DNA linker that has been grafted onto the mRNA. This DNA linker terminates in a puromycin moiety that mimics an aminoacyl tRNA. During translation, the ribosome pauses at the RNA-DNA junction thus allowing the puromycin moiety

to enter the "A" site of the ribosome and form an amide bond with the growing polypeptide chain translated from the mRNA. When using a library template, the result is the display of the entire compendium of all possible polypeptide sequences attached to its corresponding mRNA "barcode" used for the identification of library members. After a selection process has been carried out (i.e. binding to a target), the "positive" selection members can separated from the "negative" selection members, reverse transcribed back into DNA, sequenced, and identified on a genomic level resulting in genomic and phenotypic information on what residues are important for what property is being selected.

Compared to other display technologies (phage, yeast, ribosome, etc.), mRNA display and selection have several unique advantages that make it a compelling system for interrogating biological systems and questions. *In vivo* steps in a selection procedure typically limit the size of the library that can be studied, resulting in low complexity compared to the completely *in vitro* mRNA display. This can be due to low efficiency of transformation or transfecting the library cDNA into the organism of choice. Phage display allows the complexity of the library to range from 10^9-10^{10}, while other cell-based selections (yeast-two hybrid, bacteria and yeast display) are limited to ~10^6. [25,26] Unlike the previously mentioned systems, mRNA display is completely *in vitro*, allowing large libraries to be generated, on the order of 10^{13} unique members, reflecting the large diversity present within the proteome. Another advantage of mRNA display is that each library sequence will be represented multiple times, allowing the greater likelihood of capturing rare sequences. Rapid identification of these sequences is achieved due to the genotype and phenotype being covalently linked through the puromycin moiety. mRNA display also allows greater flexibility in selection methods than other display methods due to it being an *in vitro* system. By being an *in vitro* system, adjustments in cofactors (i.e. ATP), temperature,

metals (i.e. Mg^{+2}), pH and detergents, which are useful in selection procedures, can be made to optimize selection and post-translational modifications if so desired.[5-7,27,28]

Ribosomally-synthesized and post-translationally modified peptides (RiPPs) are a growing class of natural products that have garnered substantial attention because of their structural diversity and biological activities due to the enzymatic modifications changing the canonical amino acids into moieties not accessible by ribosomal synthesis.[29-32] The biosynthesis of RiPPs involves leader peptide-directed enzymatic transformations on the core peptide which can be readily exploited for combinatorial biosynthesis and other applications.[33-37] The biosynthetic enzymes and machinery that impart the desired modifications for activity are unique due to several properties they share. First, most of these enzymes house a specialized recognition domain termed the RiPPs recognition element (RRE) that allows the RiPPs enzymes to recognize and coordinate to their intended peptide substrate. This RRE recognizes specific amino acids and motifs present in the leader peptide of the precursor peptide.[38-45] This gives RiPPs one of their most impressive abilities, their promiscuity within their own biosynthesis. Since the recognition of the peptide substrate happens due to the leader peptide-RRE interaction, the core peptide can be altered from its wild-type amino acid sequence and the enzymatic modifications still imparted on the altered core peptide. While not every change is allowed, this system houses much more flexibility in the identity of its substrates than most enzymes could tolerate.[46-51] This gives RiPPs the advantage of creating a wide range of distinct, yet similar peptides that can be tested and altered for specific activities or properties. mRNA display happens to be uniquely compatible with RiPPs due to several factors. RiPPs are genetically encoded and translated by the ribosome both of which are required by mRNA display. Also, mRNA display can tolerate chemical post-translational modifications and has capability to test

libraries on the order of $10^{12\text{-}13}$ unique members giving a comprehensive study of the promiscuity of a certain RiPPs enzyme related to its substrate.

We set out to accomplish two things using the powerful combination of mRNA display and RiPPs natural products **(see Figure 4.1a)**. The first was to tease out leader peptide requirements for binding to the RRE of two RiPPs enzymes, PaaA and TbtF. PaaA is a RiPPs enzyme from the pantocin A biosynthesis pathway that catalyzes the double dehydration and decarboxylation of two glutamic acid residues in its precursor peptide PaaP **(Figure 4.1b)**. The structure of PaaA was solved in 2016 by Bowers and co-workers and a RRE motif can clearly be seen at the N-terminus of each one of the monomers present in the dimer.[52] TbtF is an accessory protein in the biosynthesis of thiomuracin that acts as the recognition element for peptide and enzymatic modification **(Figure 4.1b)**. Although no structural information is known for TbtF, many biochemical analyses have been done to determine a) TbtF is required for full maturation of the precursor peptide, TbtA, into thiomuracin[53] and b) TbtF binds to TbtA with a K_d of ~43 nM and several residues within the leader peptide have been identified to be critical to binding by fluorescence polarization.[54] If key residues of RRE binding could be identified, then chimeric leader peptides, leader peptides that allow the recognition of more than one RiPPs enzyme from different pathways, can be designed to allow the generation of hybrid-RiPPs natural products to be modified by more than one RiPP enzyme from more than one pathway, combining two distinct RiPPs pathways into one.[55] The second was to fully probe the promiscuity of PaaA modification in regards to changes in PaaP. While small libraries can give some sort of indication of RiPP enzyme biosynthetic promiscuity, the library size accomplished through mRNA display can give a completely picture of what residues are necessary for modification and what residues can be modified. With mRNA display, a complete promiscuity profile can be

done on a single RiPPs enzyme in mere days. This study will build upon the results gathered in Ghodge *et al*.[52]

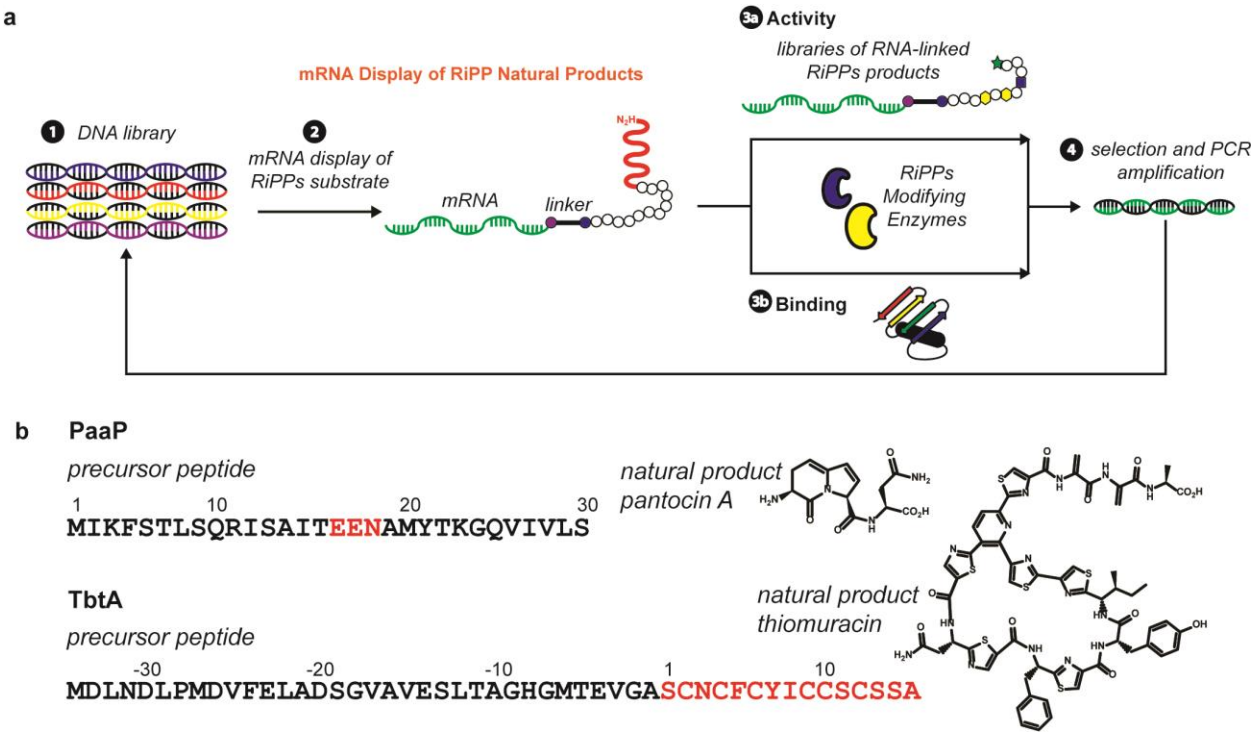


Figure 4.1. Overview of mRNA display combined with RiPPs. a) Workflow of mRNA display to test enzymatic modification and binding to RiPPs enzymes. b) Two precursor peptides and their corresponding natural products used in this study. The core peptides are highlighted in red that form the natural products to the left.

4.2 Design of Targeted RiPP libraries for mRNA Display

We started by designing DNA templates that house the T7 promoter (necessary for T7 RNA polymerase), RBS (necessary for translation), precursor peptide of interest, and then a specialized spacer and annealing sequences **(see Experimental 4.6.1 and Table 4.1)**. The highly optimized and specialized annealing sequence is required to photo-crosslink to the DNA-puromycin linker, which is used to attach the translated peptide off the mRNA.[56] The precursor peptides of interest His-PaaP (used for PaaA activity), PaaP (used for PaaA binding) and TbtA (used for TbtF binding), were synthesized by Keck Oligonucleotide Synthesis facility at Yale School of Medicine, while the additional components were added by PCR and primers **(see**

Experimental 4.6.1 Table 4.1 and 4.2). Since this is a pilot study, we decided to limit the complexity of the libraries we would use in our selections and not try the top level of library members, ~$10^{12\text{-}14}$. For both the PaaP based libraries, we used bioinformatics to predict important residues for both binding and activity of PaaP to its modifying enzyme PaaA **(see Appendix C.1 and C.2)**. Based on these analyses, we decided to randomize codons Ile2, Leu7, Arg10, Ile11, Ser12, Asn18 **(Figure 4.1b)** to NNC, NNC, RVK, NNC, RVK, NNC respectively, where NNC represent a list of 15 amino acids (A, C, D, F, G, H, I, L, N, P, R, S, T, V, Y) and RVK represents (A, D, E, G, H, K, N, R, S, T) or charged hydrophilic amino acids. In the codon annotation, N=any nucleotide (A, C, G or T), R= any purine (A or G), V= any nucleotide not T (A, C or G), and K= keto nucleotide (G or T). The theoretical library complexity for PaaP is designed to contain 5,062,500 unique members. For TbtA, we looked to a recent article by Zhang *et al*. where they measured the binding affinity of full length TbtA to two thiomuracin biosynthetic enzymes (TbtF and TbtD).[54] They found three residues that when mutated to an alanine, severely impair binding affinity to both TbtF and TbtD. We randomize the codons of these residues, Leu-29, Asp-26, and Phe-24 **(Figure 4.1b)** to NNC. The theoretical library complexity for TbtA is designed to contain 3,375 unique members. For both PaaP and TbtA, it was imperative that for the codons randomized, there was the possibility for the wild-type (WT) residues to be selected for again, since these residues were already selected for in nature and may have the best affinity or be required activity. Also there was no possibility of stop codons so the full length precursor peptide could be translated. Using the NNC and RVK codons, we were able to lower the library complexity, include the WT residues in each randomized position, and prevent stop codons for terminating translation prematurely. In parallel, the true WT peptide

constructs for His-PaaP, PaaP, and TbtA were made as radio-labeled controls to test conditions and selections.

4.3 Creation of Displayed-Peptide Libraries

All experimental detail is explained in **Experimental sections 4.6.1, 4.6.3 and 4.6.4** but the entire process will be explained in brief here. After the DNA template of the randomized precursor peptides were synthesized and delivered by Keck, the prerequisite additions (T7 promoter, RBS, spacer, and annealing sequences) were added on using a series of primers and PCR reactions. After the PCR reaction and gel purification, the DNA template was used in a transcription reaction using T7 RNA polymerase followed by a purification step using phenol: chloroform: isoamyl alcohol extraction. After this purification, the mRNA was treated with DNaseI to remove the template DNA followed by another phenol: chloroform: isoamyl alcohol extraction and purification. Now that the mRNA is clean, we connect the mRNA to a DNA-Puromycin linker (P-linker). This specialized linker was synthesized by standard oligonucleotide solid-phase synthesis **(Experimental 4.6.2)** to allow four things to happen 1) anneal, complementary, to the mRNA anneal sequence, 2) photo-crosslink through a [2+2] photo-cycloaddition to one uracil nucleotide on the mRNA sequence and one *O*-Me-uracil nucleotide on the DNA-Puromycin linker, 3) has a region of poly-A sequence (A_{15}) that facilitates the purification by Oligo-dT_{25} magnetic resin and 4) house the Puromycin moiety that will attach to the C-terminus of the peptide translated off the mRNA sequence **(Figure 4.2a and b)**. To connect the mRNA to the P-linker, an annealing program was run to allow the complementary sequences on the mRNA and linker to come together. The annealing program was followed by the exposure to UV light (~365 nm) for 20 minutes. UV exposure facilitates the [2+2] photo-cycloaddition covalent attachment of the mRNA to the DNA puromycin linker. This mRNA-

DNA-Puromycin molecule was purified by LiCl precipitation. The cross-linking efficiency was tested by 8% denaturing urea gel compared to the mRNA template **(see Figure 4.2c)**. The efficiency ranged anywhere from 20-50% mRNA cross-linked. In **Figure 4.2d**, the crosslinking efficiency for the library constructs can be seen in order: PaaP, His-PaaP, TbtA.

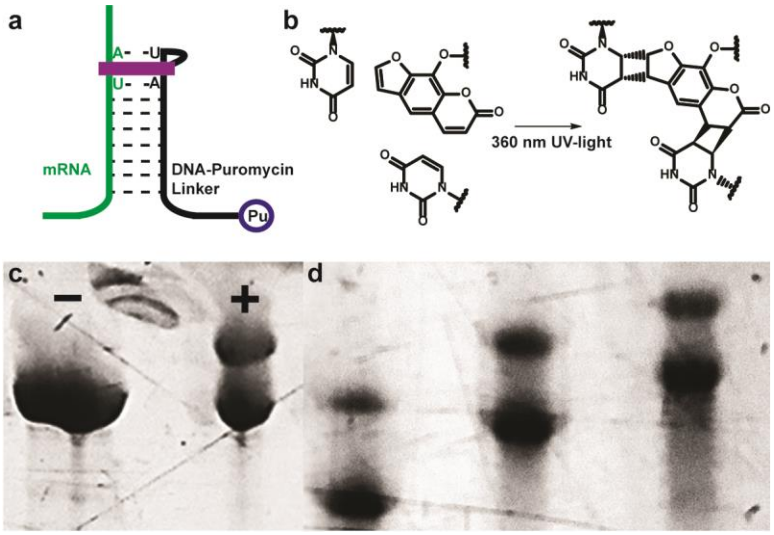


Figure 4.2. Covalent attachment of DNA-Puromycin linker to mRNA. a) Pictorial representation of mRNA attachment to DNA-Puromycin linker (P-linker). b) [2+2] photo-cycloaddition covalent attachment between two uracil bases. c) Example 8% denaturing urea gel showing the shift from mRNA (-) to mRNA-DNA-Puromycin molecule (+). d) 8% denaturing urea gel showing crosslinking efficiency of library constructs, PaaP (left), His-PaaP (middle), TbtA (right).

Using the mRNA-DNA-Puromycin as a template, the peptide was translated using the PURE *in vitro* translation system supplied by New England Biosciences (NEB). For every reaction (25 µl), ~1.0-1.2 µg of template was used in the reaction. If the desired molecule was to be radio-labeled, ^{35}S-Met was used to incorporate into the peptide of interest. After the translation, $MgCl_2$ and KCl was added to allow the puromycin to enter the "A" site of the ribosome and attach to the C-terminus of the recently translated peptide.[2] After the overnight salt incubation, the mRNA-DNA-Puromycin-Peptide fusion (display peptide) was purified by Oligo-

dT_{25} magnetic resin and eluted in simple Tris buffer. The purification results in the purified displayed peptide used in the mRNA display selection procedures **(Figure 4.3)**.

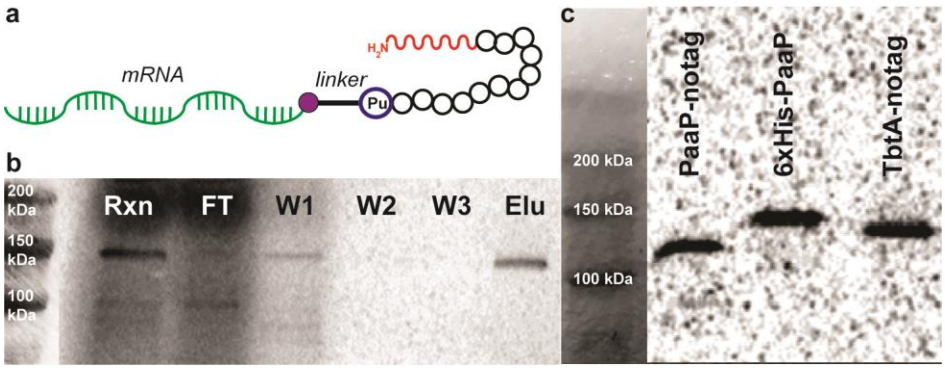


Figure 4.3. Translation and purification of "display" peptide. a) Pictorial representation of display peptide. b) Oligo-dT_{25} purification of display peptide. c) Radio-labeled (^{35}S), 8% SDS-PAGE gel showing purified WT-display peptides that correspond to library constructs.

4.4 Selection Procedures

We decided to not protect the mRNA portion of the displayed peptide molecule by reverse transcription before selection because upon doing so, the gel bands in the radio-labeled gels became diffuse and low quality. We tested thoroughly for RNase activity, included excess Murine RNase Inhibitor (0.5 U/μl), and we were satisfied by the quality of the data without the additional protection of the reverse transcription. PaaA and TbtF were expressed and purified based on previously reported procedures but adapted to ensure the removal of RNases **(see Experimental section 4.6.6)**.[52,53] The selection procedures differed significantly between binding and activity and are discussed separately below **(see Experimental section 4.6.5)**.

4.4.1 PaaA Activity

We used the display peptide His-PaaP and PaaA to test whether PaaA could modify a displayed peptide and what residue mutations from the above library criteria allow this modification. Using conditions set forth by Ghodge *et al.*[52], His-PaaP displayed peptide was incubated with or without PaaA overnight at room temperature for 16-18 hours. PaaA catalyzes

the double dehydration and decarboxylation of two glutamic acid residues in PaaP **(Figure 4.4a)**. After overnight modification assay, GluC was added to cleave after the glutamic acids present in the peptide. The beauty of PaaA is that upon modification of the glutamic acids, the side chains are no longer there or hidden and prevent GluC from cleaving the peptide and removing the 6xHis peptide **(Figure 4.4b)** Then the modified, digested His-PaaP displayed peptide is incubated with Ni-NTA to allow any 6xHis tag still present to attach and bind to the resin. After washing the resin, the bound displayed peptide is eluted off with high imidazole and heat. This is the positive, selected peptides, meaning the mutations allowed modification by PaaA. After elution, we used Superscript III to reverse transcribe the mRNA followed by PCR to create more of the selected DNA template **(Figure 4.6c)**.

By testing with radio-labeled, WT, His-PaaP, we can see a distinct difference between GluC digested PaaA treated vs non-treated display peptide. **In Figure 4.4c**, a clear shift corresponding to the loss of the 6xHis tag and leader peptide of His-PaaP can be seen when comparing treated vs. non-treated with PaaA. This is clear evidence that GluC cannot digest modified PaaP. It is unlikely that all the displayed peptide is modified by PaaA, but these results show that is it possible for a RiPPs modifying enzyme to impart its modifications on a displayed peptide.

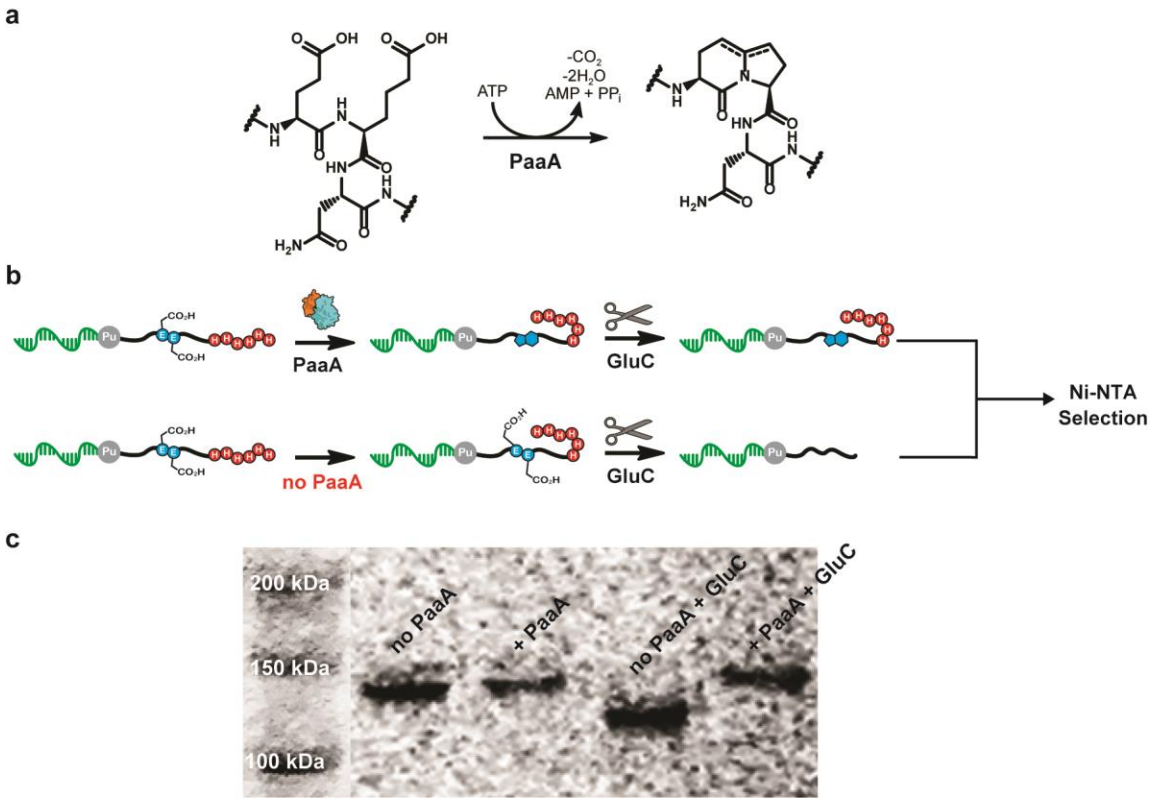


Figure 4.4. Selection for PaaA activity. a) PaaP modification by PaaA. b) Pictorial representation of selection procedures for PaaA activity. c) Radio-labeled (^{35}S), 8% SDS-PAGE gel showing shift from GluC digestion when not treated with PaaA.

4.4.2 PaaA and TbtF Binding

For binding assays, we had the 6xHis tags attached to PaaA (C-term) and TbtF (N-term with MBP) and not attached to the peptide substrate. Initially, we incubated the displayed peptide substrate with its corresponding RiPPs enzyme. After this "pre-binding" step, we incubated this peptide-bound-enzyme with Ni-NTA resin. We removed the flow-through and washed the resin once with buffer. To elute, we heat denatured at 95 ^{o}C in 1x reverse transcription (RT) buffer. This is the positive, selected peptides, meaning the mutations allowed binding to the enzymes. After elution, we used Superscript III to reverse transcribe the mRNA followed by PCR to create more of the selected DNA template **(Figure 4.6.b)**.

By testing with radio-labeled, WT, PaaP and TbtA, binding to their respective RiPPs enzymes (PaaA and TbtF) can clearly be seen **(Figure 4.5b and c)**. In these radio-labeled gels, little to no displayed peptide can be seen in the flow-through and wash lanes, but the displayed peptide can clearly be seen coming off in the elution. These gels show that displayed peptide can bind to its corresponding RiPPs enzymes. These results open the door to testing binding affinity and leader peptide requirements for interacting and binding to their corresponding RRE present in these enzymes. It has been shown by Roberts and co-workers that mRNA display can be used to determine binding kinetics which can be used and adapted for determining binding constants for leader peptides to RRE's.[23]

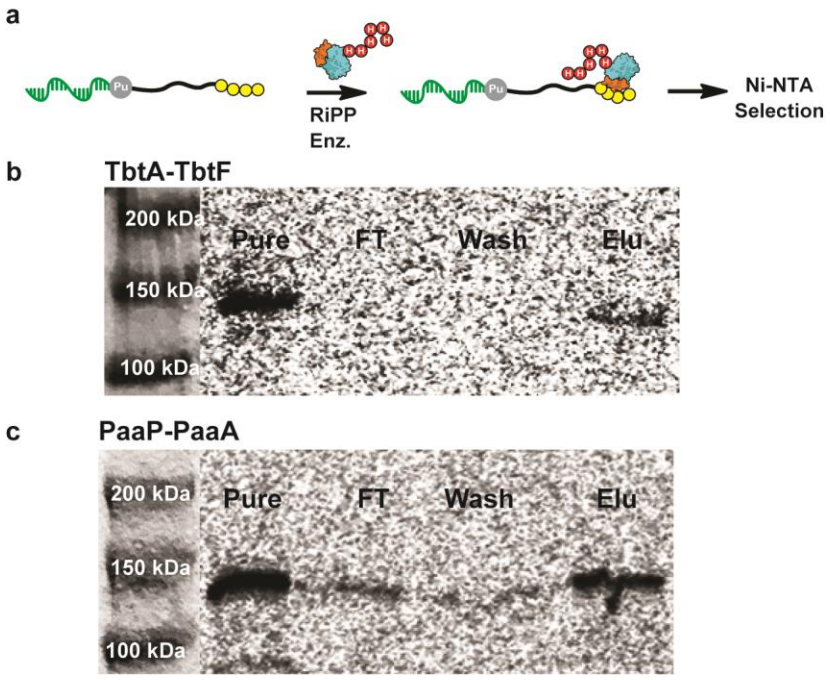


Figure 4.5. Selection for binding to RiPPs enzymes. a) Pictorial representation of selection for binding to RiPPs enzymes. b) Radio-labeled (^{35}S), 8% SDS-PAGE gel showing display-peptide (TbtA) binding and eluting with TbtF. c) b) Radio-labeled (^{35}S), 8% SDS-PAGE gel showing display-peptide (PaaP) binding and eluting with PaaA.

4.5 Summary and Discussion

The next step after the reverse transcription and PCR would be to sequence the positive selected sequences that correspond to either modification or binding depending on the selection procedure **(Figure 4.6)**. Traditionally, the DNA sequences would be cloned into a blunt or TA-TOPO vector and the insert sequenced normally by Sanger methods.[15,19-22] This approach does not give a full depiction of the library until many round of selection and a great number of colonies sequenced. Also, when we tried this approach, we received inconsistent data, namely partial reads and low quality data. There is another approach using next-generation sequencing (high-throughput sequencing) to sequence the DNA that was amplified after selection, called amplicons. Next-generation sequencing (NGS) has many advantages over traditional cloning and sequencing.[57] First, there is no cloning, so poor cloning, ligation, and transformation efficiency are not a problem. Second, the PCR product off the reverse transcription reaction is what is sent for sequencing, so there are no more necessary steps required by the scientist carrying out the selection. Lastly, instead of getting one read of a sequence like in traditional Sanger sequencing, NGS gives millions of reads over the amplicons, allowing full coverage and quantitation, by percentage, of what is present in the amplicon. Exact sequences, single-nucleotide polymorphisms (SNPs), and mutations can all be detected at the same time through NGS. More recently, NGS was used with mRNA display in a high-throughput manner to measure binding kinetics between a library of peptide and Bcl-x_L protein.[23] This was accomplished by adding on unique identifying barcodes and utilizing a Illumina[58] HiSeq 2500 platform to sequence the library. This approach gave both the frequency and factional composition of their library, statistics valuable for knowing the complexity and make-up of your selected library. We have planned on doing a similar NGS approach using the Illumina

technology and platform; however at the time of this writing, this has yet to be done. We are actively pursuing NGS through GENEWIZ to perform NGS on the amplicons after selection, both binding and activity **(Figure 4.6b and c)**.

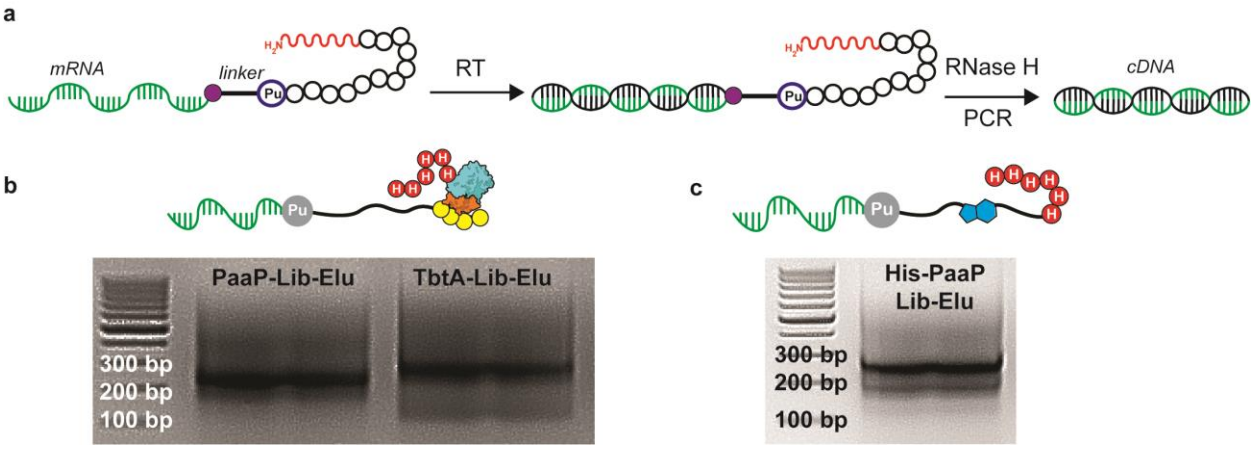


Figure 4.6. Reverse transcription and amplification of selected library members. a) Pictorial representation of workflow. b) Amplicons of selected members from PaaA and TbtF binding assays. c) Amplicon of selection from PaaA activity assay. The sizes of the corresponding amplicons are: PaaP: 225 bp, His-PaaP: 258 bp, and TbtA: 282 bp.

In the presented work, we have described a mRNA system for probing the promiscuity of the biosynthesis of a natural product class, RiPPs. We show that it is possible for displayed RiPPs to bind to their respective enzymes while attached to their genomic component, mRNA. We also show that it is indeed possible to impart RiPPs modifications onto a displayed RiPP precursor peptide and confirm that modification. We also outline the procedures necessary to probe both the biosynthetic activity as well as binding affinity of RiPPs to their enzymes by mRNA display. Sequencing, specifically NGS, is what remains to analyze the data collected after selection. With this system, we believe it is possible to fully probe the biosynthetic promiscuity of a RiPP enzyme as well as to determine the leader peptide requirements necessary for binding to the RRE of the enzyme. This can give valuable information on what residues are necessary and can be used to create chimeric leader peptides to combine separate and distinct RiPPs pathways into a single precursor peptide.

4.6 Experimental

4.6.1 Design of DNA templates and mRNA transcription and photo-crosslinking

The following tables describe the DNA prepared by Keck Oligonucleotide Synthesis facility at Yale School of Medicine and the primers prepared by Eton Bioscience. Also listed are the sequences to be added on by PCR to create the DNA templates for mRNA display. Library codons are in bold and underlined. The Universal RT primer was synthesized and prepared by Integrated DNA Technologies and where iSp9 is an internal triethylene glycol spacer prepared by them. The template DNA was amplified from their respective templates **(Table 4.1)** by PCR using primers **(Table 4.2)** and Q5® High-Fidelity DNA Polymerase with Q5 Reaction Buffer following the manufacturer's manual (New England Bioscience (NEB)). The template DNA was purified by 2% agarose gel and then gel extracted following the manufacturer's manual (Thermo Scientific GeneJET Gel Extraction Kit). The template DNA was eluted in nuclease free water and used as is in the RNA transcription step.

Table 4.1. DNA sequences used to create template DNA for mRNA display constructs.

PaaP-WT	5'-CTT TAA GAA GGA GAT ATA CAT ATG ATC AAG TTC TCT ACT CTG TCT CAG CGC ATT TCT GCT ATC ACT GAA GAA AAT GCT ATG TAT ACC AAA GGT CAG GTT ATC GTC CTT AGC GGT TCT GGT GGT AGC GGT ATG-3'
His-PaaP-WT	5'-CAT CAT CAT GGT TCT TCT GGT ATG ATC AAG TTC TCT ACT CTG TCT CAG CGC ATT TCT GCT ATC ACT GAA GAA AAT GCT ATG TAT ACC AAA GGT CAG GTT ATC GTC CTT AGC GGT TCT GGT GGT AGC GGT ATG-3'
TbtA-WT	5'-CTT TAA GAA GGA GAT ATA CAT ATG GAC CTG AAT GAT CTG CCG ATG GAT GTT TTT GAA CTG GCA GAT AGC GGT GTT GCA GTT GAA AGC CTG ACC GCA GGT CAT GGT ATG ACC GAA GTT GGT GCA AGC TGT AAT TGC TTT TGT TAT ATT TGT TGT AGC TGC AGC AGC GCC-3'
PaaP-Lib	5'-CTT TAA GAA GGA GAT ATA CAT ATG **NNC** AAG TTC TCT ACT **NNC** TCT CAG **RVK NNC RVK** GCT ATC ACT GAA GAA **NNC** GCT ATG TAT ACC AAA GGT CAG GTT ATC GTC CTT AGC GGT TCT GGT GGT AGC GGT ATG-3'
His-PaaP-Lib	5'-CAT CAT CAT GGT TCT TCT GGT ATG **NNC** AAG TTC TCT ACT **NNC** TCT CAG **RVK NNC RVK** GCT ATC ACT GAA GAA **NNC** GCT ATG TAT ACC AAA GGT CAG GTT ATC GTC CTT AGC GGT TCT GGT GGT AGC GGT ATG-3'

TbtA-Lib	5’-CTT TAA GAA GGA GAT ATA CAT ATG GAC CTG AAT GAT **NNC** CCG ATG **NNC** GTT **NNC** GAA CTG GCA GAT AGC GGT GTT GCA GTT GAA AGC CTG ACC GCA GGT CAT GGT ATG ACC GAA GTT GGT GCA AGC TGT AAT TGC TTT TGT TAT ATT TGT TGT AGC TGC AGC AGC GCC-3’
T7-Promoter (based on pMCSG7)	5’-TAA TAC GAC TCA CTA TAG G-3’
RBS (based on pMCSG7)	5’-TAA CTT TAA GAA GGA G-3’
Spacer Sequence	5’-GGT TCT GGT GGT AGC GGT ATG GGA ATG TC-3’
Anneal Sequence	5’-CAC CGG CTA TTA A-3’
Total 5’ Addition (based on pMCSG7)	5’-GAA ATT AAT ACG ACT CAC TAT AGG GGA ATT GTG AGC GGA TAA CAA TTC CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CAT-3’
Total 3’ Addition (Spacer+Anneal Seq.)	5’-GGT TCT GGT GGT AGC GGT ATG GGA ATG TCC ACC GGC TAT TAA-3’

Table 4.2. Primers used in mRNA display study

Primer #	Template Used	Sequence
Primer 1	PaaP-Fwd 1 and TbtA-Fwd 1	5’-TTC CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA-3’
Primer 2	His-PaaP-Fwd-1	5’-TTC CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CAT ATG CAC CAT CAT CAT CAT CAT GGT TCT TCT-3’
Primer 3	PaaP-Fwd 2 and His-PaaP-Fwd 2 and TbtA-Fwd 2	5’-GAA ATT AAT ACG ACT CAC TAT AGG GGA ATT GTG AGC GGA TAA CAA TTC CCC TCT AGA AAT-3’
Primer 4	TbtA-Rev-1	5’-TTA ATA GCC GGT GGA CAT TCC CAT ACC GCT ACC ACC AGA ACC GGC GCT GCT GCA GCT AC-3’
Primer 5	Universal Fwd	5’-GAA ATT AAT ACG ACT CAC TAT-3’
Primer 6	Universal Rev	5’-TTA ATA GCC GGT GGA CAT TCC-3’
Primer 7	Univeral RT	5’-TTT TTT TTT TTT TTT TTT TT/iSp9/G ACA TTC CCA TAC CGC TAC CAC CAG AAC C-3’

Generation of RNA

Following the manufacturer's manual (NEB), RNA was created with T7 RNA Polymerase and precipitated out of the reaction solution (white solid) using $MgCl_2$. All reagents were nuclease free (both RNA and DNA). The normal reaction consisted of:

Water (nuclease free)	X to a volume of 250 µL
10x RNA polymerase Buffer	25 µL
25 mM rNTP's	30 µL
1 M $MgCl_2$	3 µL
1 M DTT	2.5 µL
T7 RNA Polymerase (50 U/µL)	25 µL
cDNA	Y to an amount of ~1.0 µg
Total	250 µL

The reaction was incubated at 37 ^{o}C for 2 hrs. Then 18 µL of 0.5 M EDTA was added to suspend the RNA back in solution

Purification of RNA

RNA was purified using a standard Phenol: Chloroform: Isoamyl Alcohol extraction, keeping the aqueous layer. Then to the ~500 µl of aqueous solution, 250 µl of 8 M LiCl was added to precipitate the RNA. The solution was incubated at -20 ^{o}C for 0.5 hrs-overnight. The RNA was pelleted out of solution by centrifugation at 4 ^{o}C, 15,000 rpm, for 30 minutes. The supernatant was removed leaving a white pellet. The white pellet was suspended in 85 ul of water and DNaseI treatment was carried out as follows:

RNA solution	85 µL
10x DNaseI Buffer	10 µL
DNaseI (2 U/µL)	5 µL
Total:	100 µL

The solution was incubated at 37 ^{o}C for 2.5 hrs. Purification was done as before and the RNA was suspended in 100 µL of nuclease free water.

4.6.2 Synthesis of DNA-Puromycin Linker (P-linker)

Synthesis of the DNA-Puromycin (P-linker) was accomplished by standard solid-phase phosphoramidite chemistry using an Applied Biosystems 392/394 biosystems DNA/RNA Synthesizer with Software API. All reagents and building blocks were purchased from Glen Research. The overall sequence of the P-linker is as follows: 5'-Psoralen(C6)-$(UAGCCGGUC)_{2'OMe\text{-}RNA}$-$dA_{15}$-C9C9-$(ACC)_{RNA}$-Puromycin-3', where dA15 is a DNA spacer of 15 adenosines and C9 is Spacer Phosphoramidite 9. Detritylation was achieved with use of 3% trifluoroacetic acid (TFA) in dichloromethane (DCM). Activation of the phorphoramidite building block for coupling used 0.45M ^{1}H-tetrazole in acetonitrile. To prevent branching and missing nucleotides, capping of any un-coupled phosphoramidite building blocks was achieved using both acetic anhydride/pyridine/tetrahydrofuran(THF) in a 1:1:8 ratio and 16% N-methylimidazole in THF. Prior to the next coupling, the phosphite-triester (P(III)) formed in the previous coupling step must be converted to the more stable P(V) species. This is achieved by iodine oxidation in the presence of water and pyridine using 0.02M iodine in THF/pyridine/water in a ratio of 70:20:10. This converts the backbone into a protected-phosphate DNA backbone. Then the deprotection and coupling cycle continues. For our coupling procedures the following steps were used to couple each building block. The steps were modified from 0.1 μM RNA standard coupling procedure from ABI 392/394.

Step	Time (seconds)
Acetonitrile Wash	30
Trityl-deprotection	50
Coupling (Tetrazole+ Base)	720
Capping (Capping solution A and B)	30
Oxidizing	45
Reverse Flush (to remove reagents)	10

The building blocks were dissolved in acetonitrile (anhydrous) to a final concentration of 0.1M and the starting solid support was Puromycin-CPG (1 μmole scale). All couplings were done in an inert (argon gas) atmosphere under anhydrous conditions. The trityl-deprotection was monitored by detecting the release of the chromophore trityl cation at absorbance of 498 nm. If average deprotection count was $\geq$ 90% the synthesis was allowed to proceed forward without change or restarting. Once Psoralen(C6) was added, light was avoided with the use of foil. Cleavage and deprotection was accomplished with 1 mL of ethanolic ammonium hydroxide (3 parts ammonium hydroxide to 1 part 100% ethanol. The solution was passed over the column 4-5 times and then allowed to incubate in contact with the resin at room temperature for 24 hours. After deprotection and cleavage, the solution was collected and resin washed with 0.25 mL of RNase free water and combined with the precious supernatant. The solution was further dried under vacuum with the use of a speed-vac with no heat. This was to remove as much ammonium hydroxide, ethanol, and water as possible and to yield a white solid. Further deprotection was carried out to remove the 2'-t-butyl-dimethylsilyl groups (Ultramild TBDMS) from the RNA bases added in P-linker sequence. The white solid was dissolved/suspended in 0.5 mL of

anhydrous dimethyl sulfoxide (DMSO) by heating the solution at 65 ^{o}C for 5 minutes. Add 125 μL of triethylamine trihydrofluoride (TEA.3HF) and heat at 65 ^{o}C, 750 rpm mix, for 2 ½ hours. After, the solution was cooled at -20 ^{o}C for 3 minutes followed by the addition of 25 μL of 3 M sodium acetate and 1 mL of 1-butanol. The solution was then incubated at -80 ^{o}C for 1 hour. The P-linker solid was then pelleted at 4 ^{o}C at 15,000 rpm for 10 minutes. A white pellet formed at the bottom of the tube and the supernatant was removed. The pellet was washed twice with 0.75 mL of 100% ethanol repeating the pelleting step above and removing the supernatant. The pellet was further dried under vacuum using a speed-vac with no heat for 30 minutes. The P-linker solid was dissolved/suspended in nuclease free water and loading dye and loaded onto a preparatory-20% denaturing urea gel. Before loading, the gel was pre-warmed at 200 V (6 watts) for 45 minutes. The gel after loading was ran at 250 V (6 watts) for 2 hours. The loading dye used held xylene cyanol FF (XCFF) dye which runs at ~28 bp in a 20% denaturing urea gel. Therefore XCFF was used as a ladder being the P-linker is 31 bp. Also, the P-linker was detected, without over exposure, with UV-light. With the addition of the Psoralen(C6), the band is bright blue. The band was cut and removed from the gel, mashed to pieces/powder and 1x TE (10 mM Tris-HCl pH=7.5, 1 mM EDTA) added in a ratio of 3 mL for every 0.5 grams of gel. The gel solution was vortexed vigorously and then frozen at -80 ^{o}C for 15 minutes or until frozen solid. The frozen solution was thawed in a water bath (42 ^{o}C) for 10 minutes followed by being wrapped in foil and place on rotary shaker overnight at room temperature. After overnight extraction, the gel pieces were pelleted and the supernatant removed and collected. The supernatant was filtered through a 0.22 μm filter to remove any other solid. The volume of the supernatant was reduced by the removal of water with 1:1 butanol extraction. Briefly, the butanol was added, shaken for 30 seconds and then pelleted at 4000 rpm, room temperature, for

two minutes and the top layer removed. This was repeated until the bottom layer, the layer with the P-linker, was ~ 1 mL. Next, 3 M sodium acetate (pH=5.2) was added to a final concentrate of 0.3M. Then 2 mL of ice cold 100% ethanol was added followed by incubation at -20 ^{o}C for 30minutes. When the sodium acetate was added, the solution turned white and cloudy, but when the ethanol was added, the solution turned clear and colorless again. After incubation at -20 ^{o}C, the P-linker was pelleted at 4 ^{o}C, 15,000 rpm, for 20 minutes. A white pellet formed at the bottom of the tube and the supernatant was removed. The pellet was washed with 70%, ice cold, ethanol and the pelleted the same way above. To further dry the pellet, the pellet was dried under vacuum using a speed-vac without heat. The pellet was suspended in nuclease free water. Purity was tested by diagnostic 20% denaturing urea gel and nanodrop UV-vis measurements. The final concentration of the P-linker solution was designed to be ~150 μM. The overall yield of the P-linker synthesis was calculated to be 6.4% from a 1.0 μmole scale.

4.6.3 Preparation of puromycin-fused mRNA libraries

A reaction solution was made to perform a UV-crosslink between the RNA of interest and the DNA-Puromycin linker (P-linker). All reagents were nuclease free (both RNA and DNA). The solution contained:

1 M HEPES	5 μL
2 M KCl	12 μL
P-linker	Z μL (~5.6 μM)
RNA	Y μL (~1.0 μM)
Water (nuclease free)	X μL (final volume 250 μl)
Total:	250 μl

The solution was split into 5 x 50 μl reactions. An anneal program, in a thermocycler, was ran starting at 85 ^{o}C and dropping 0.5 ^{o}C every 30 seconds, finally holding at 25 ^{o}C for 3 minutes. After anneal program is finished, the reaction solution was exposed to UV light (~365 nm) for 20 minutes in cold room (4 ^{o}C). After UV crosslink was formed, the reactions were

combined and 118 μL of 8 M LiCl was added. The reaction was incubated overnight at -20 ^{o}C. After the overnight incubation at -20 ^{o}C, the RNA-DNA-Puromycin construct was pelleted by centrifugation at 4 ^{o}C, 15,000 rpm, for 30 minutes. A pellet formed, and the supernatant was removed. The pellet was washed twice with 0.5 mL of 75% ethanol. Once all supernatant was removed from the wash, the pellet was suspended in 50 μL of nuclease water.

4.6.4 PURE System Translation of Displayed Peptide Libraries

Expression of Display Molecules using PURE System

Expression was done according to the manufacturer's manual (NEB) that comes with the PURExpress® *In Vitro* Protein Synthesis Kit. All reagents were nuclease free (both RNA and DNA). The reaction is as follows and scales linearly:

Solution A	10 μL
Solution B	7.5 μL
Water	X (final volume 25 μL)
Murine RNase Inh. (40 U/μL)	1 μL
^{35}S-Met (if desired)	1 μL
Template	Y (~1-2 μg of RNA)
Total:	25 μL

The solution was incubated at 37 ^{o}C for 3 hours followed by incubation at 4 ^{o}C for 5 minutes to halt translation. After expression a salt incubation was performed by adding 2 μL of 1M $MgCl_2$ and 11 μL of 2M KCl (final concentrations: ~50 mM and ~580 mM respectively. The reaction was incubated at room temp for 30 minutes then -20 ^{o}C overnight.

Purification of Display Molecules by Oligo-dT_{25} Magnetic Beads

Purification was done according to the manufacturer's manual (NEB) that comes with Oligo dT_{25} Magnetic Beads and magnetic separation rack. The buffer used were as follows: Lysis/Binding Buffer (20 mM Tris-HCl pH=7.5, 500 mM LiCl, 0.5% LiDS, 100 mM EDTA, 5 mM DTT), Wash Buffer I (20 mM Tris-HCl pH=7.5, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA, 5

mM DTT), Wash Buffer II (20 mM Tris-HCl pH=7.5, 500 mM LiCl, 1 mM EDTA), Low Salt Buffer (20 mM Tris-HCl pH=7.5, 200 mM LiCl, 1 mM EDTA), and Elution Buffer (20 mM Tris-HCl pH=7.5). Briefly, the Oligo-dT_{25} resin was equilibrated and washed with lysis buffer. For purification, the amount of resin used is roughly the same volume as the translation reaction. Lysis buffer was added in a 1:1 ratio to that of the translation reaction and vortexed briefly. Add the translation reaction with lysis buffer to the resin and incubate at room temperature for 30 minutes with light mixing. The supernatant was removed with magnetic separation rack, followed by washing (with one translation reaction volume) the resin twice with Wash Buffer I, twice with Wash Buffer II, and once with Low Salt Buffer with removal of the supernatant after each wash. The display molecule was eluted in Elution Buffer twice with 0.5 x translation reaction volumes after heating the elution buffer and resin at 70 ^{o}C, 500 rpm, for 2 minutes. The elutions are combined, cooled and used as is in the selection procedures.

4.6.5 *In vitro* selection methods and Reverse Transcription/PCR

PaaA or TbtF Binding Assays

The binding assays were the same for either peptide-protein interactions tested except for the buffers used. PaaA used the buffer 50 mM HEPES at pH= 7~8 tested by pH paper, 0.01% Triton X-100, 0.5 mM TCEP, and 0.5 U/μL of murine RNase inhibitor while TbtF used the buffer 50 mM HEPES at pH= 7~8 tested by pH paper, 300 mM NaCl, 2.5% glycerol, 0.01% Triton X-100, 0.5 mM TCEP, and 0.5 U/μL of murine RNase inhibitor. Protein (20 μM) was incubated in buffer with 25 μL of purified display molecule in a total of 40 μL. This solution was allowed to incubate at 4 ^{o}C in end-over-end mixer for one hour (pre-bind). As the incubation was happing, 40 μL of a 50% slurry of Nuvia^{TM} IMAC resin from Bio-rad was prepared and washed with the corresponding buffer above. The resin was spun down at 2000 x g

at 4 ^{o}C for 2 minutes. The resin was washed twice with the buffer before all the supernatant was removed, leaving 20 μL of resin. To the equilibrated resin, the pre-binding solution was added, followed by further incubation at 4 ^{o}C in end-over-end mixer for 1 ½ hours. After the incubation, the resin and supernatant (flow-through (FT)) were separated by centrifugation at 2000 x g at 4 ^{o}C for 2 minutes. The resin was washed twice with 40 μL of the corresponding buffer while the supernatant (Wash) was collected as previously stated above. The display-molecule bound to protein was eluted by adding 40 μL of 1x reverse transcription (RT), 1st strand buffer (50 mM Tris-HCl pH= 8.3, 75 mM KCl, 3 mM $MgCl_2$ and heating the resin and solution at 95 ^{o}C for 5 minutes. The supernatant (Elution) was collected as previously stated above.

PaaA Activity Assays

PaaA activity and modification assays were adapted from Ghodge *et al*. The buffer used in the activity assay was 50 mM HEPES at pH= 7~8 tested by pH paper, 0.01% Triton X-100, 0.5 mM TCEP, 10 mM $MgCl_2$, 4 mM ATP, and 0.5 U/μL of murine RNase inhibitor. 20 μM of PaaA was incubated with 25 μL of purified display molecule in a total volume of 50 μL. The reaction was vortex briefly to mix and then allowed to incubate at room temperature without further mixing for 16-18 hours overnight. Following the modification, 0.5 μL of endoproteinase GluC (500 U/mL) was added to the previous reaction and incubated at 37 ^{o}C, 500 rpm, for 2 hours. After the modification and digestion reactions, Ni-NTA selection was achieved as stated above with PaaA binding assays. The differences are 50 μL volumes, 25 μL of resin, the wash buffer was the original binding buffer with 50 mM imidazole while the elution buffer was the original binding buffer with 500 mM imidazole. A single incubation time of one hour was used to bind the 6xHis tag to the Nuvia^{TM} IMAC resin.

Reverse Transcription (RT) /PCR amplification of selection

The following procedures were used with all samples after selection. SuperScript™ III First-Strand Synthesis System from Invirtogen™ was used to preform reverse transcription following the procedures laid out in the manufacturer's manual. All reagents were nuclease free (both RNA and DNA). Briefly the following reactions were made to carry out the reverse transcription:

Part I	
Universal RT Primer (100 μM)	0.5 μL
Water	0 μL
10 mM dNTP's	1 μL
Template (from selection)	11.5 μL
Total:	13 μL

Part II	
5x 1st Strand Buffer	4 μL
Water	1 μL
100 mM DTT	1 μL
Superscript III (200 U/μL)	1 μL
Total (once added to Part I)	20 μL

Part I was heated at 65 ^{o}C for 5 minutes then incubated on ice for 1 minute. Next Part II was added to part I and then incubated at the following temperatures using a thermocycler: 55 ^{o}C for 1 hour, 75 ^{o}C for 15 minutes, and 4 ^{o}C for 5 minutes. This was followed by the addition of 0.4 μL of 5 U/μL of RNase H. This was incubated at 37 ^{o}C for 20 minutes. This gave back the complementary piece of DNA corresponding to the RNA selected. Amplification of the DNA was achieved using *Taq* DNA Polymerase with Standard *Ta*q Buffer following the manufacturer's manual using the universal forward and reverse primers in **Table 4.2**. The DNA was purified by 2% agarose gel and then gel extracted following the manufacturer's manual (Thermo Scientific GeneJET Gel Extraction Kit). The template DNA was eluted in nuclease free water and used as is.

4.6.6 Protein Expression and purification of PaaA and TbtF

PaaA

This procedure was adapted from Ghodge *et al*.[52] Buffers were steri-filterd and made with DEPC treated water while the Ni^{2+} IMAC column was stripped and regenerated before use. PaaA (from *paaA*-pET28b, C-terminal 6xHis tag) was expressed in BL21 (DE3) *E. coli* cells in LB medium containing 40 μg/mL kanamycin. Cells were grown at 37 °C until an OD_{600} of 0.5 was attained, at which point the temperature was lowered to 18°C. At an OD_{600} of 0.8, protein expression was induced with the addition of 0.1 mM isopropyl-1-thio-D-galactopyranoside (IPTG) and incubated overnight at 18 °C. Cells were collected and pelleted by centrifugation. Cell pellets were suspended in Buffer A (20 mM Tris, pH=7.5, 30 mM imidazole, 300 mM NaCl, 5% Glycerol, 1 mM DTT) supplemented with 0.5 mL of 25 mg/mL T4 lysozyme, 0.5 mL of 150 mM PMSF, 80 μl of DNaseI (1u/μl), one tablet of Pierce™ Protease Inhibitor Tablets (EDTA Free from Thermo Scientific), and then incubated on ice for 10 minutes. The cell pellet was then sonicated. The lysate was clarified by centrifugation at 15,000 rpm at 4 ^{o}C for 15 minutes. The supernatant was collected and filtered through a 0.45 μm syringe filter. The flow through from the filter was then passed over a Ni^{2+} IMAC column (HISTrap™ HP 5mL GE Healthcare) coupled to an FPLC (NGC-Quest-10 Bio-Rad). The Ni^{2+} IMAC column was washed with 5 column volumes (CV) of Buffer A. Protein was eluted with Buffer B (20 mM Tris, pH=7.5, 500 mM imidazole, 300 mM NaCl, 5% glycerol, 1 mM DTT) in a gradient of 0-14% B over 2 CV, 14-14% B over 2 CV, and then 14-100% B over 10 CV. Purity of eluted fractions was assessed by SDS-PAGE. Purest fractions were combined and concentrated in a 30,000 MWCO filter in an Amicon stirred cell. The concentrated protein was then exchanged into 50 mM HEPES, pH=7.5, 150 mM NaCl, 5% glycerol, and 1 mM DTT by passage over a

Sephadex™ PD-10 column (GE Healthcare). The protein concentration was estimated by A_{280} using an extinction coefficient of 0.988 $mg \bullet mL^{-1} \bullet AU^{-1}$.

TbtF

This procedure was adapted from Hudson *et al.*[53] Buffers were steri-filterd and made with DEPC treated water while the Ni^{2+} IMAC column was stripped and regenerated before use. TbtF (from pMCSG9-*tbtF*, N-terminal 6xHis-MBP tag) was expressed in BL21 (DE3) *E. coli* cells in LB medium containing 100 μg/mL ampicillin. Cells were grown at 37 °C until an OD_{600} of 0.6 was attained, at which point the temperature was lowered to 18°C and protein expression was induced with the addition of 0.2 mM IPTG and incubated overnight at 18 °C. Cells were collected and pelleted by centrifugation. Cell pellets were suspended in Buffer A (50 mM Tris, pH=7.5, 150 mM NaCl, 2.5% Glycerol, 0.1% Trition X-100) supplemented with 0.5 mL of 25 mg/mL T4 lysozyme, 0.5 mL of 150 mM PMSF, 80 μl of DNaseI (1u/μl), one tablet of Pierce™ Protease Inhibitor Tablets (EDTA Free from Thermo Scientific), and then incubated on ice for 10 minutes. The cell pellet was then sonicated. The lysate was clarified by centrifugation at 15,000 rpm at 4 ^{o}C for 15 minutes. The supernatant was collected and filtered through a 0.45 μm syringe filter. The flow through from the filter was then passed over a Ni^{2+} IMAC column (HISTrap™ HP 5mL GE Healthcare) coupled to an FPLC (NGC-Quest-10 Bio-Rad). The Ni^{2+} IMAC column was washed with 5 column volumes (CV) of Buffer B (50 mM Tris, pH=7.5, 400 mM NaCl, 2.5% Glycerol, 50 mM imidazole). Protein was eluted with Buffer C (50 mM Tris, pH=7.5, 300 mM NaCl, 2.5% Glycerol, 500 mM imidazole) in a gradient of 0-9% B over 2 CV, 9-9% B over 2 CV, and then 9-100% B over 10 CV. Purity of eluted fractions was assessed by SDS-PAGE. Purest fractions were combined and concentrated in a 30,000 MWCO filter in an Amicon stirred cell. The concentrated protein was then exchanged into 50 mM Tris, pH=7.5,

300 mM NaCl, and 2.5% Glycerol by passage over a SephadexTM PD-10 column (GE Healthcare). The protein concentration was estimated by A_{280} using an extinction coefficient of 0.96 $mg \bullet mL^{-1} \bullet AU^{-1}$.

4.7 REFERENCES

1. Roberts, R.W.; Szostak, J.W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297

2. Liu, R.; Barrick, J.E.; Szostak, J.W.; Roberts, R.W. *Methods Enzymol*. **2000**, *318*, 268

3. Nemoto, N.; Miyamoto-Sato, E.; Husimi, Y.; Yanagawa, H. *FEBS Lett.* **1997** *414*, 405

4. Cotton, S.W.; Zou, J.; Valencia, C.A.; Liu, R. *Nat. Protoc*. **2011**, *6*, 1163

5. Ju, W.; Valencia, C.A.; Pang, H.; Ke, Y.; Gao, W.; Dong, B.; Liu, R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14294

6. Shen, X.; Valencia, C.A.; Szostak, J.W.; Dong, B.; Liu, R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5969

7. Shen, X.; Valencia, C.A.; Gao, W.; Cotton, S.W.; Dong, B.; Huang; B.C.; Liu, R. *Cell Calcium*. **2008**, *43*, 444

8. Tateyama, S.; Horisawa, K.; Takashima, H.; Miyamoto-Sato, E.; Doi, N.; Yanagawa, H. *Nucleic Acids. Res*. **2006**, *34*, e27

9. Hammond, P.W.; Alpin, J.; Rise, C.E.; Wright, M.; Kreider, B.L. *J. Biol. Chem*. **2001**, *276*, 20898

10. Cujec, T.P.; Medeiros, P.F.; Hammond, P.; Rise, C.; Kreider, B.L. *Chem. Biol*. **2002**, *9*, 253

11. McPherson, M.; Yang, Y.; Hammond, P.W.; Kreider, B.L. *Chem. Biol*. **2002**, *9*, 691

12. Horisawa, K.; Tateyama. S.; Ishizaka, M.; Matsumura, N.; Takashima, H.; Miyamoto-Sato, E.; Doi, N.; Yanagawa; H. *Nucleic Acids. Res*. **2004**, *32*, e169

13. Fukuda, I.; Kojoh, K.; Tabata, N.; Doi, N.; Takashima, H.; Miyamoto-Sato, E.; Yanagawa, H. *Nucleic Acids. Res*. **2006**, *34*, e127

14. Josephson, K.; Hartman, M.C.T.; Szostak, J.W. *J. Am. Chem. Soc.* **2005**, *127*, 11727

15. Seelig, B.; Szostak, J.W. *Nature*. **2007**, *448*, 828

16. Bashiruddin, N.K.; Suga, H.; *Curr. Opin. Chem. Biol.* **2015**, *24*, 131

17. Takahashi, T.T.; Austin, R.J.; Roberts, R.W. *Trends Biochem. Sci.* **2003**, *28*, 159

18. Barendt, P.A.; Ng, D.T.W.; McQuade, C.N.; Sarkar, C.A.; *ACS Comb. Sci.* **2013**, *15*, 77

19. Guillen Schlippe, Y.V.; Hartman, M.C.T.; Josephson, K.; Szostak, J.W.; *J. Am. Chem. Soc.* **2012**, *134*, 10469

20. Horiya, S.; Bailey, J.K.; Temme, J.S.; Guillen Schlippe, Y.V.; Krauss, I.J. *J. Am. Chem. Soc.* **2014**, *136*, 5407

21. Millward, S.W.; Takahasi, T.T.; Roberts, R.W. *J. Am. Chem. Soc.* **2005**, *127*, 14142

22. Hayashi, Y.; Morimoto, J.; Suga, H. *ACS Chem. Biol.* **2012**, *7*, 607

23. Jalai-Yazdi, F.; Lai, L.H.; Takahasi, T.T.; Roberts, R.W. *Angew. Chem. Int. Ed.* **2016**, *55*, 4007

24. Maini, R.; Umemoto, S.; Suga, H. *Curr. Opin. Chem. Biol.* **2016**, *34*, 44

25. Conrad, U.; Scheller, J. *Comb. Chem. High Throughput Screen*. **2005**, *8*, 117

26. Huang, H.; Jedynak, B.M.; Bader, J.S. *PLoS Comput. Biol*. **2007**, *3*, e214

27. Huang, B.; Liu, R. *Biochem*. **2007**, *46*, 10102

28. Valencia, C.A.; Cotten, S.W.; Dong, B.; Liu, R. *Biotechnol. Prog*. **2008**, *24*, 561

29. 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.; Muller, 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*. **2012**, *30*, 108

30. Ortega, M. A.; van der Donk, W. A. *Cell Chem. Biol.* **2016**, *23*, 31

31. Dunbar, K. L.; Mitchell, D. A. *ACS Chem. Biol.* **2013**, *8*, 473

32. McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26* , 537

33. Bowers, A. A.; Acker, M. G.; Koglin, A.; Walsh, C. T. *J. Am. Chem. Soc*. **2010**, 132, 7519

34. Menzella, H. G. H.; Reeves, C. D. C. *Curr. Opin. Microbiol*. **2007**, *10*, 8

35. Sardar, D.; Schmidt, E. W. *Curr. Opin. Chem. Biol.* **2015**, *31*, 15

36. Kim, E.; Moore, B. S.; Yoon, Y. J. *Nat. Chem. Biol.* **2015**, *11*, 649

37. Ruffner, D. E.; Schmidt, E. W.; Heemstra, J. R. *ACS Synth. Biol*. **2015**, *4*, 482

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

39. Ortega, M. A.; Hao, Y.; Zhang, Q.; Walker, M. C.; van der Donk, W. A.; Nair, S. K. *Nature* **2015**, *517*, 509

40. Koehnke, J.; Mann, G.; Bent, A. F.; Ludewig, H.; Shirran, S.; Botting, C.; Lebl, T.; Houssen, W. E.; Jaspars, M.; Naismith, J. H. *Nat. Chem. Biol.* **2015**, *11*, 558

41. Ghodge, S. V.; Biernat, K. A.; Bassett, S. J.; Redinbo, M. R.; Bowers, A. A. *J. Am. Chem. Soc.* **2016**, *138*, 5487

42. Ortega, M. A.; Hao, Y.; Walker, M. C.; Donadio, S.; Sosio, M.; Nair, S. K.; van der Donk, W. A. *Cell Chem. Biol.* **2016**, *23*, 370
43. Tsai, T.-Y.; Yang, C.-Y.; Shih, H.-L.; Wang, A. H. J.; Chou, S.-H. *Proteins*. **2009**, *76*, 1042

44. Regni, C. A.; Roush, R. F.; Miller, D. J.; Nourse, A.; Walsh, C. T.; Schulman, B. A. *EMBO J.* **2009**, *28*, 1953

45. Cheung, W. L.; Chen, M. Y.; Maksimov, M. O.; Link, A. J. *ACS Cent. Sci.* **2016**, *2*, 702

46. Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729

47. Himes, P. M.; Allen, S. E.; Hwang, S.; Bowers, A. A. *ACS Chem. Biol.* **2016**, *11*, 1737

48. Melby, J. O.; Dunbar, K. L.; Trinh, N. Q.; Mitchell, D. A. *J. Am. Chem. Soc.* **2012**, *134*, 5309

49. Melby, J. O.; Nard, N. J.; Mitchell, D. A. *Curr. Opin. Chem. Biol.* **2011**, *15*, 369

50. Tianero, M. D.; Donia, M. S.; Young, T. S.; Schultz, P. G.; Schmidt, E. W. *J. Am. Chem. Soc.* **2012**, *134*, 418

51. Velasquez, J. E.; van der Donk, W. A. *Curr. Opin. Chem. Biol.* **2011**, *15*,11

52. Ghodge, S.V.; Biernat, K.A.; Bassett, S.J.; Redinbo, M.R.; Bowers, A.A. *J. Am. Chem. Soc.* **2016**, *138*, 5487

53. Hudson, G.A; Zhang, Z.; Tietz, J.I.; Mitchell, D.A.; van der Donk, W.A. *J. Am. Chem. Soc.* **2015**, *137*, 16012

54. Zhang. Z.; Hudson, G.A.; Mahanta, N.; Tietz, J.I.; van der Donk, W.A.; Mitchell, D.A. *J. Am. Chem. Soc.* **2016**, *138*, 15511

55. Burkhart, B.J.; Kakkar, N.; Hudson, G.A.; van der Donk, W.A.; Mitchell, D.A.; *ACS Cent. Sci*. **2017**, *Article ASAP*, doi: 10.1021/acscentsci.7b00141

56. Valencia, C.A.; Zou, J.; Liu, R. *Methods*. **2013**, *60*, 55

57. Goodwin, S.; McPherson, J.D.; McCombie, W.R. *Nat. Rev. Genet*. **2016**, *17*, 333

58. An Introduction to Next-Generation Sequencing Technology. Illumina ®. https://www.illumina.com/technology/next-generation-sequencing.htm

CHAPTER 5

CONCLUSION

By studying RiPPs biosynthetic enzymes and their innate tendency toward accepting changes to the core peptide, we have opened the door to creating, studying, and designing peptide natural product libraries for the creation of novel therapeutics based in nature. By focusing on the RiPPs class of natural products and sactipeptides, we have biochemically and structurally characterized a sactionine synthase and reported the first crystal structure of its kind, developed a heterologous platform for the production of sactipeptides to study the promiscuity of sactionine synthases outside their native producers, and developed an *in vitro* system for studying RiPPs biosynthetic promiscuity on a large scale using mRNA display.

In the presented work, we detailed the structure of a sactionine synthase, CteB from *C. thermocellum* ATCC 27405 and biochemically characterize its activity on its thermocellin precursor peptide. CteB installs a single sactionine bridge between Cys32 and Thr37 within the precursor peptide. The CteB structure represents the first structure of a sactionine synthase and PqqE-like enzyme. It houses three distinct domains: a wHTH domain that corresponds to the RRE that binds to the leader peptide portion of the precursor peptide, a conserved SAM activating domain, and a new SPASM domain motif that houses two auxiliary [4Fe-4S] clusters, one of which displays a single open coordination site (Aux I). The crystal structure provide insights into the enzymatic mechanism of sactionine bridge formation and evidence for the role

of SPASM auxiliary clusters in direct substrate ligation and potential activation required to facilitate product formation.

We have also developed a heterologous expression system for the production of sactipeptide derivatives of subtilosin A in *E. coli*. Using this system we demonstrated unexpected biosynthetic promiscuity of the sactionine synthase AlbA. We found there was great promiscuity especially in the loop region and at unmodified positions on the solvent-exposed exterior of the peptide macrocycle and we hypothesize that this could be exploited for the grafting of peptide epitopes (i.e. integrin binding motifs) as has been seen in lasso peptides and conotoxins. We have also introduced nonproteinogenic amino acids into sactipeptides using stop-codon suppression technology, specifically at a bridging partner position. This demonstrates the robustness of UAA incorporation in our system and may open up new chemistry and new applications for unnatural sactipeptides.

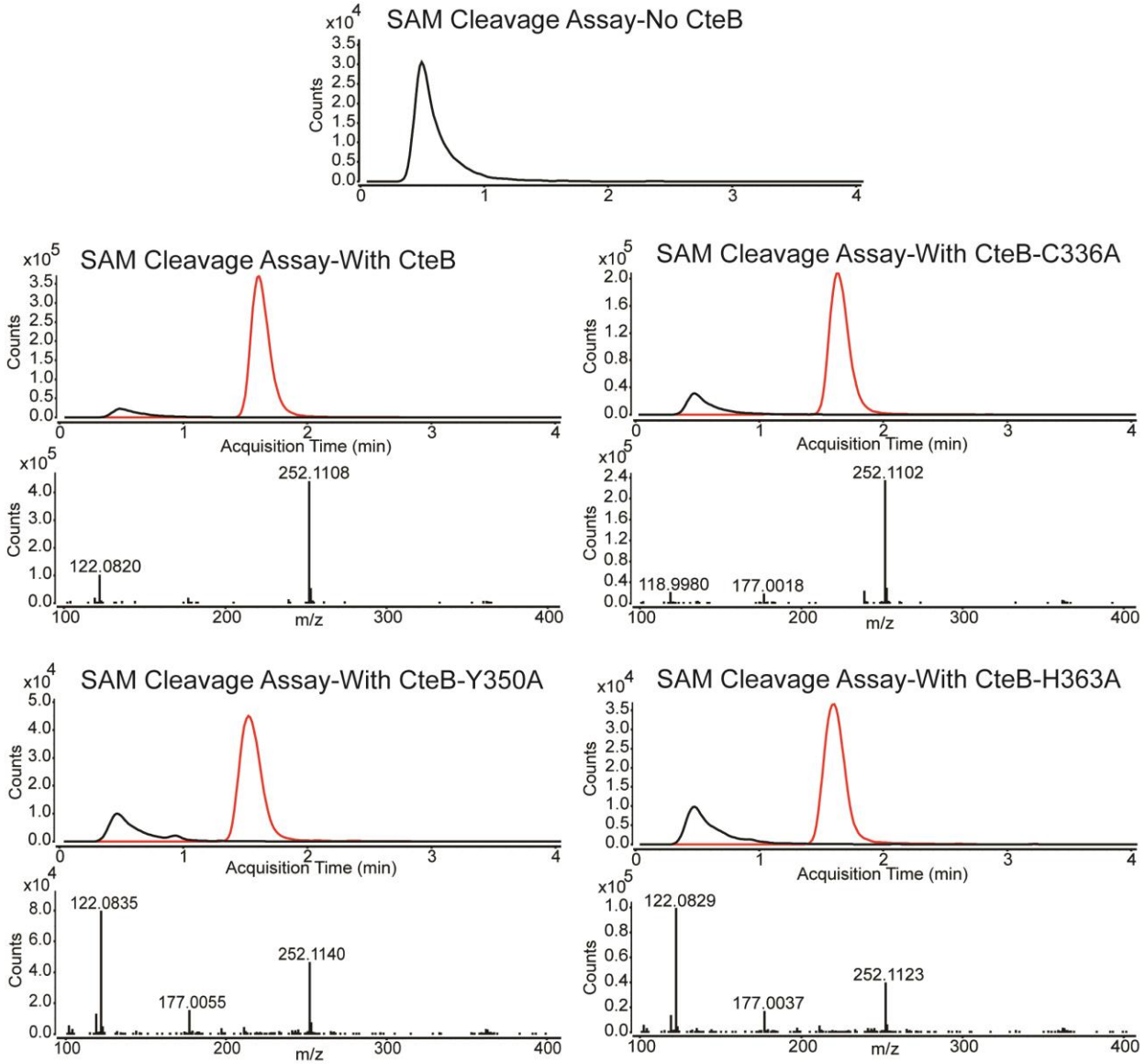
Furthermore, we have developed a system that combines RiPPs natural products and mRNA display to test their respective biosynthetic promiscuity on a large scale and in a high-throughput manner. Specifically, we used the pantocin A precursor peptide and modifying enzyme, PaaP and PaaA respectively, to test the biosynthetic promiscuity of PaaA in regards to six different randomized positions. We also used this system as well as the thiomuracin precursor peptide and binding protein, TbtA and TbtF respectively, to test binding efficacy and to select for the best possible binders with respect to binding to the RRE domain. This system can revolutionize the way RiPPs promiscuity is studied and lead the way to making natural product peptide libraries which can be selected for different properties and activities.

With the leaps and advances of next-generation sequencing, the pool for unknown and attractive natural products grows. Natural products are attractive lead molecules for novel

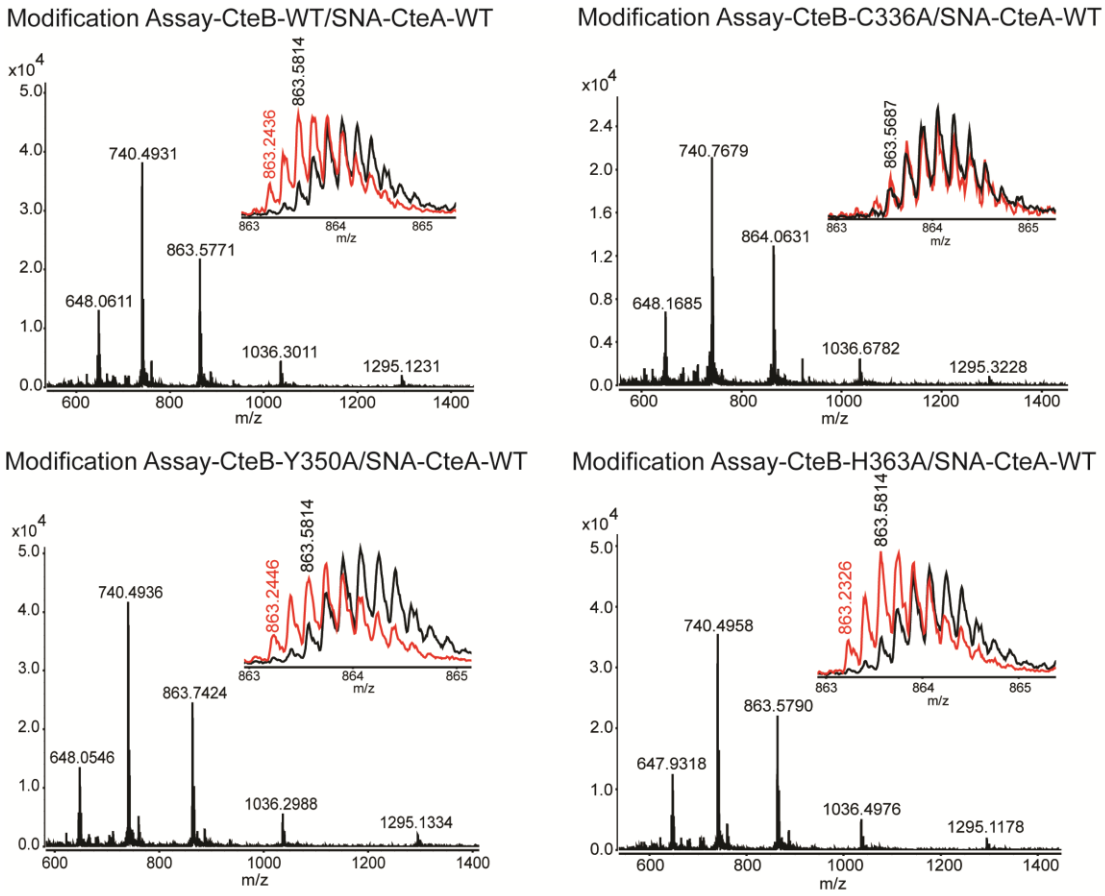
therapeutics due to their complex structures and their inherent biological activities. RiPPs, with their unique access to modified chemistries and privileged scaffolds can be mined for even more lead molecules and modified by their promiscuous enzymes allowing further optimization of activity and properties. By studying these enzymes and combining RiPPs with mRNA display, testing and optimizing these peptide products can lead to novel therapeutics for today's most pressing and troubling diseases.

APPENDIX A. SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 2

Figure A.1. Radical cleavage of SAM by CteB and variants. In black is the extracted ion chromatogram (EIC) of SAM (399.1445) and in red is the EIC of the 5'-deoxyadensosine (252.1091) product formed by radical cleavage of SAM.



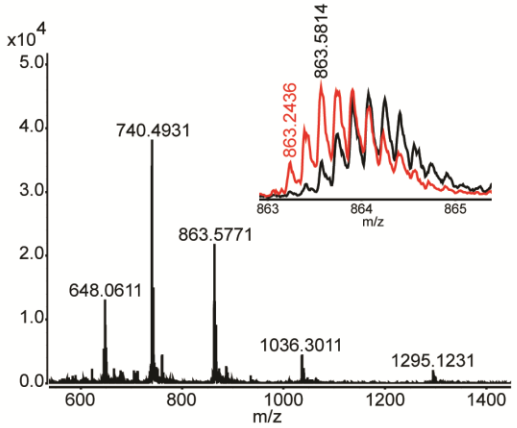
Figures A.2. Peptide Modification Assays-CteB Variants. In black is the mass spec chromatogram of the SNA-CteA construct without CteB. In red is the mass spec chromatogram of the SNA-CteA construct treated with CteB or its variants.



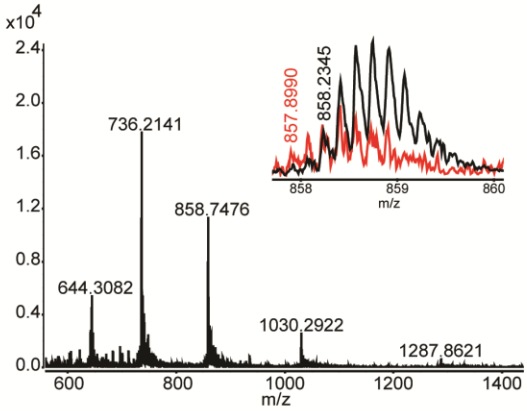
SNA-CteA-WT		
	Exact Mass	**Thioether Bridge**
	5175.4484	5173.4328
z	**m/Z**	**m/Z**
1	5176.4562	5174.4406
2	2588.7320	2587.7242
3	1726.1573	1725.4854
4	1294.8699	1294.3660
5	1036.0975	1035.6944
6	863.5826	863.2466
7	740.3576	740.0696
8	647.9389	647.6869
9	576.0576	575.8337
10	518.5527	518.3511

Figures A.3. Peptide Modification Assays-CteA Variants. In black is the mass spec chromatogram of the SNA-CteA or its variants without CteB. In red is the mass spec chromatogram of the SNA-CteA or its variants treated with CteB.

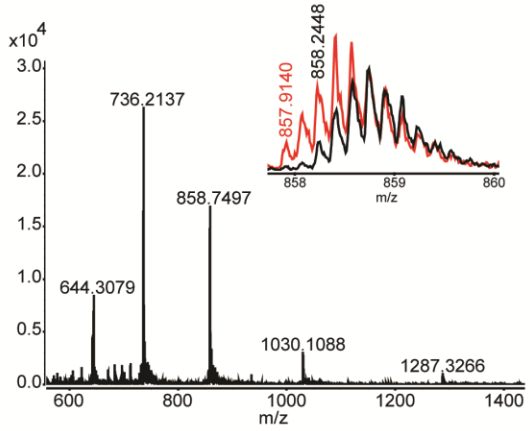
Modification Assay-CteB-WT/SNA-CteA-WT



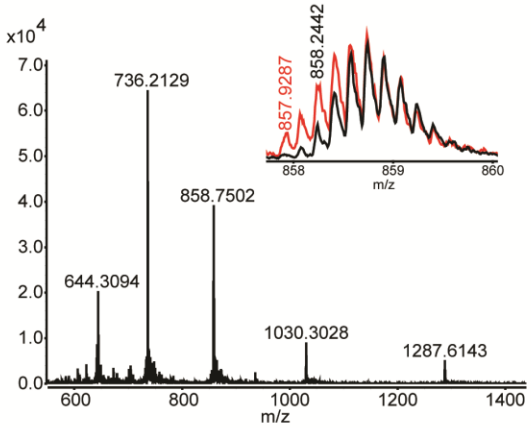
Modification Assay-CteB-WT/SNA-CteA-C21A



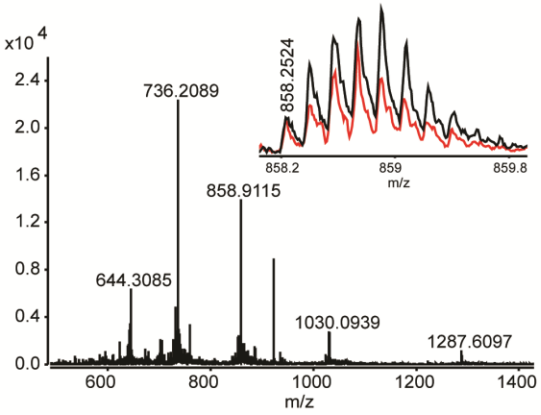
Modification Assay-CteB-WT/SNA-CteA-C24A



Modification Assay-CteB-WT/SNA-CteA-C28A



Modification Assay-CteB-WT/SNA-CteA-C32A



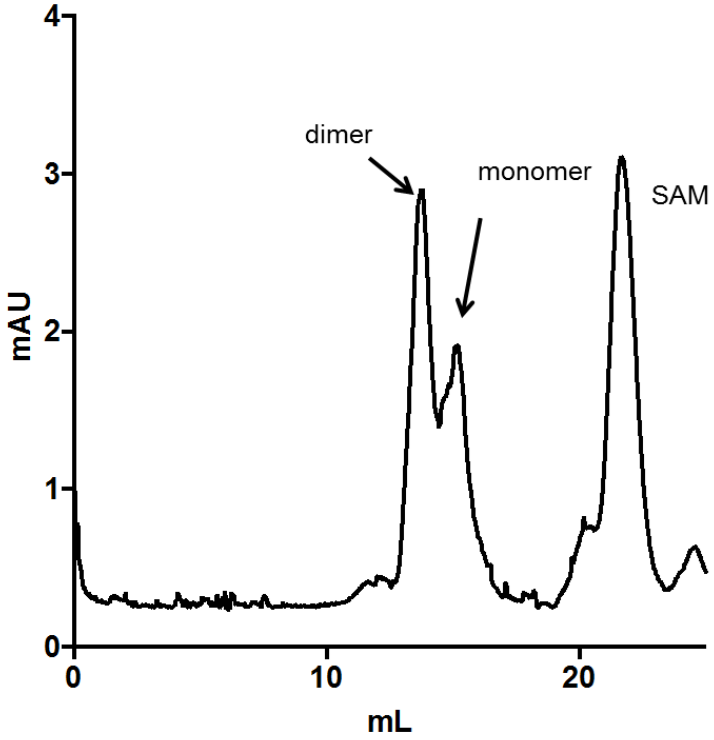
	SNA-CteA-WT			SNA-CteA-CtoA	
	Exact Mass	Thioether Bridge		Exact Mass	Thioether Bridge
	5175.4484	5173.4328		5143.4763	5141.4607
z	m/Z	m/Z	z	m/Z	m/Z
1	5176.4562	5174.4406	1	5144.4841	5142.4685
2	2588.7320	2587.7242	2	2572.7460	2571.7382
3	1726.1573	1725.4854	3	1715.4999	1714.8280
4	1294.8699	1294.3660	4	1286.8769	1286.3730
5	1036.0975	1035.6944	5	1029.7031	1029.3000
6	863.5826	863.2466	6	858.2539	857.9179
7	740.3576	740.0696	7	735.7902	735.5022
8	647.9389	647.6869	8	643.9424	643.6904
9	576.0576	575.8337	9	572.5052	572.2812
10	518.5527	518.3511	10	515.3555	515.1539

Table A.1. Crystallography Table: Data collection and refinement statistics

	CteB+SAM+CteA (M_1-C_{21}) anomalous[1]	**CteB+SAM+CteA (M_1-C_{21}) [2]**	**CteB+SAM[3]**
Data collection			
Wavelength (Å)	1.3776	1.0781	1.0333
Space group	$P2_12_12$	$P2_12_12$	$P1$
Cell dimensions			
α, β, γ (°)	90, 90, 90	90, 90, 90	83.09, 73.31, 66.63
a, *b*, *c* (Å)	69.74, 154.02, 51.93	68.92, 153.77, 51.82	51.93, 50.36, 81.36
Resolution (Å)	30.00-2.21 (2.33-2.21)	28.6-2.04 (2.11-2.04)	30.0-2.7 (2.79 – 2.70)
Redundancy	14.0 (14.2)	4.5 (3.9)	1.8 (1.7)
Completeness (%)	99.9 (100.0)	98.0 (89.4)	97.0 (90.0)
<I /σI>	21.0 (3.7)	10.47 (2.16)	6.56 (1.73)
Wilson *B*-factor (Å^2)		38.29	
R_{merge}	0.088 (0.708)	0.06768 (0.506)	0.060 (0.423)
$R_{measure}$	0.103 (0.758)	0.07639 (0.585)	0.0855 (0.604)
$CC_{1/2}$	0.999 (0.970)	0.997 (0.817)	0.994 (0.910)
Refinement			
Resolution (Å)		30.0-2.04	30.0-2.70
No. reflections		158251 (12163)	23100 (2125)
R_{work} / R_{free}		0.2047/0.2343	0.2400/0.2481
No. atoms		3794	6762
Protein		3622	6698
Ligand/ion		26	53
Water		146	11
B-factors (Å^2)			
Protein		57.4	78.6
Ligand/ion		44.8	70.9
Water		52.4	60.8
R.m.s. deviations			
Bond lengths (Å)		0.006	0.021
Bond angles (°)		1.580	1.970
Ramachandran analysis			
Favored (%)		98	98
Allowed (%)		2	2.1
Outliers (%)		0	0.12

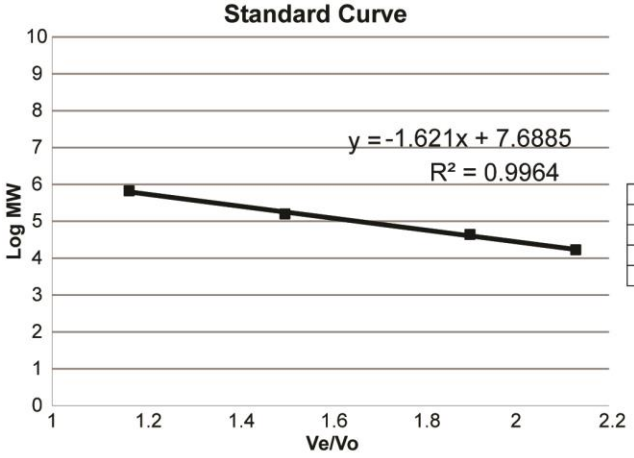
[1]Data collected at LRL-CAT beamline 31-ID-G
[2]Data collected at LS-CAT beamline 21-ID-D
[3]Data collected at GM/CA-CAT beamline 23-ID-D

Figure A.4. Size Exclusion Chromatography of TEV cleaved CteB crystals. For this SEC, crystals were combined and passed over the size exclusion column in the same manner as before. This shows that CteB can form an intermolecular disulfide bond in the crystal packing.

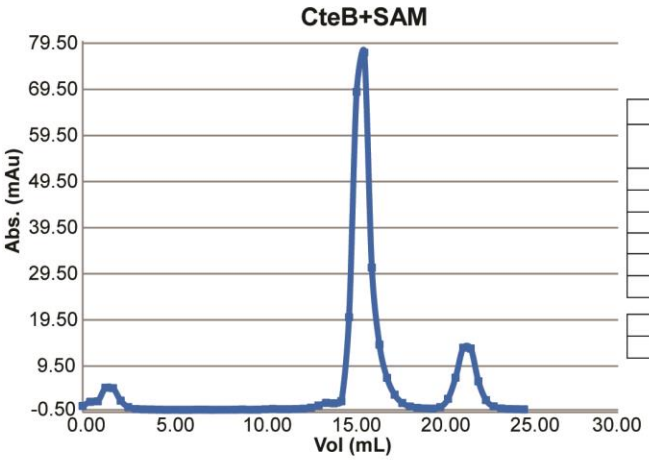


Elutions Vol	Ve/Vo	LogMW	CthE dimer	103800
13.8	1.67	4.98	CthE peak1	94844
Elutions Vol	**Ve/Vo**	**LogMW**	**CthE + pep cal.**	**577079**
15.28	1.85	4.69	CthE peak 2	48552

Figure A.5. Size Exclusion Chromatography of TEV cleaved CteB. Whether the substrate is present or not, the data suggests that CteB is a monomer. The data is in good agreeance with the standards as well.

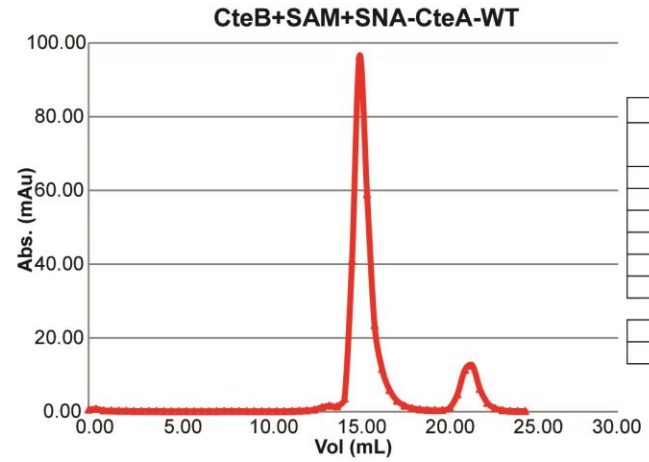


	MW (Da)	LogMW	Vol Elution (mL)	Ve/Vo
Myoglobin (horse)	17,000	4.2	17.56	2.1
Ovalbumin (chicken)	44,000	4.6	15.68	1.9
gama-globulin (bovine)	158,000	5.2	12.39	1.5
Thyroglobulin (bovine)	670,000	5.8	9.62	1.2
Void	N/A	N/A	8.25	1.0



CteB+SAM				
Reagent	[I]	units	[F]	Vol to add (uL)
TEV cleaved CteB	2.2	mM	0.1	5.0
HEPES pH 7.5	1000	mM	10	1.1
SAM	27	mM	0.5	2.0
SNA-CteA-WT	2.45	mM	0	0.0
water				101.9
			Total Vol	110

Elutions Vol	Ve/Vo	LogMW	CteB cal.	51900
15.6	1.89	4.62	CteB Wt exp.	42008



CteB+SAM+SNA-CteA-WT				
Reagent	[I]	units	[F]	Vol to add (uL)
TEV cleaved CteB	2.2	mM	0.1	5.0
HEPES pH 7.5	1000	mM	10	1.1
SAM	27	mM	0.5	2.0
SNA-CteA-WT	2.45	mM	0.15	6.7
water				95.1
			Total Vol	110

Elutions Vol	Ve/Vo	LogMW	CteB + pep cal.	57079
15.4	1.87	4.66	CteB Wt exp.	45987

Figure A.6. Non-reducing, denaturing SDS-Page Gel of CteB. CteB (after size-exclusion without DTT) at 35 μM in mixed with varying ratios of 10 mM GSH (reduced glutathione):GSSG (oxidized glutathione). CteB cannot form a disulfide-linked dimer in solution. CteB monomer is ~54.68 kDa.

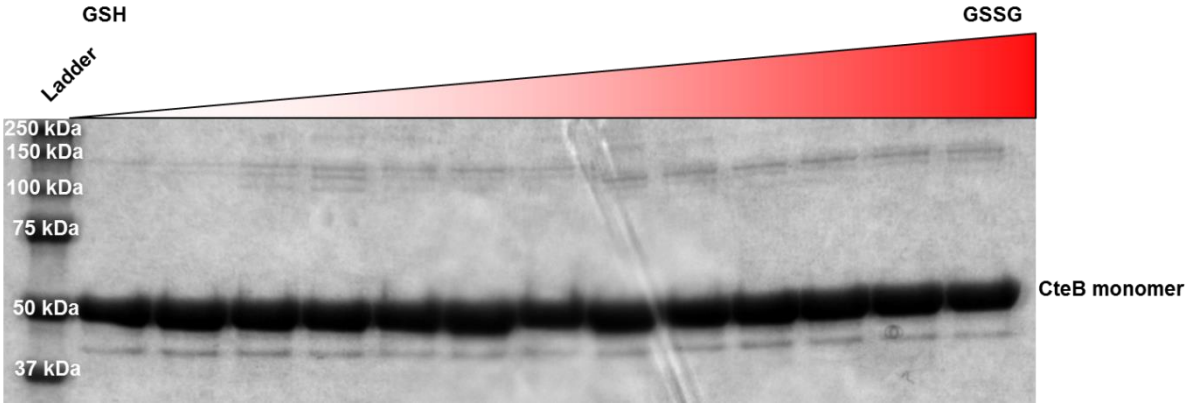


Figure A.7. Non-reducing, denaturing SDS-Page Gel of CteA-CteB Complex. Incubating CteA and CteB in solutions with varying ratios of GSH (reduced glutathione): GSSG (oxidized glutathione) shows that a CteA-CteB complex is forming and not a dimer of CteB. CteB monomer is ~54.68 kDa and the CteA-CteB complex is ~59.86 kDa.

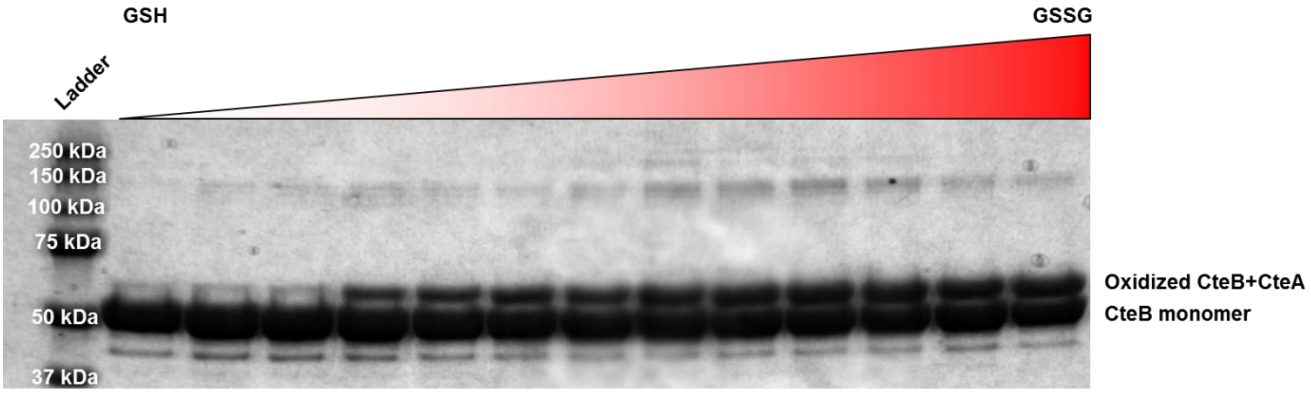


Figure A.8. Location of possible disulfide bond between symmetry mates of CteB. a) Overall structure of CteB+SAM+CteA 21mer showing the 2-fold symmetry axis (red ellipse) between two molecules of CteB from adjacent asymmetric units. The red circle denotes an area of weak electron density that connects the two symmetry mates. b) Zoomed in view of the red circle. The electron density ($2F_o$-F_c is shown in green) is weak in this area with a break (dashed lines) between residue Q333 and I337. An area of density connects the symmetry mates where C336 should be located.

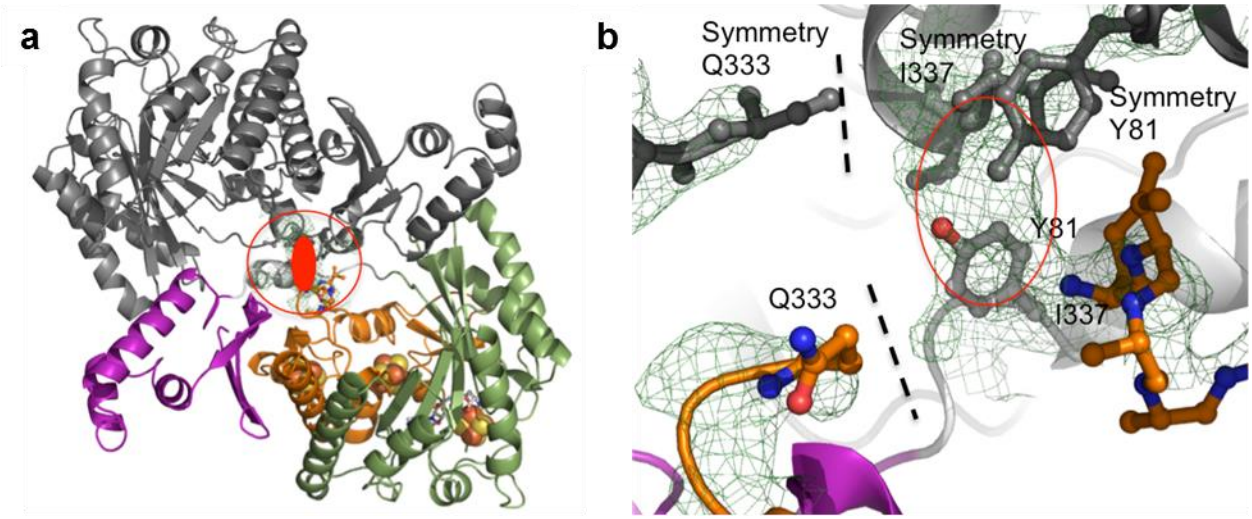


Figure A.9. Structural comparison CteB to anSME. a) Residues 90-449 CteB (colored as in Figure 2.3) overlaid with anSME (gray, 4K38). Despite sharing only 20 % sequence identity, CteB and anSME overlay with and R.M.S.D of 2.52 Å. b) Active site of CteB showing SAM (grey sticks) bound to the RS cluster along with absolutely conserved amino acids (shown as sticks and colored by domain location as in a) found in the active site based on proximity to the overlaid peptide substrate (grey, Kp18Cys) of anSME from the pdb 4K38. His-363 occupies the same location as the active site base Asp-277 (blue) of anSME. The role of the remaining residues are likely hydrogen bonding interactions for proper orientation of the peptide substrate. c) SAM binding interactions with CteB. d) Overlay of SPASM domains from CteB (orange) and anSME (gray, 4K38). The overall secondary structural elements (R.M.S.D. of 2.3 Å) are highly conserved between these domains despite the difference in cluster coordination and substrate sequences.

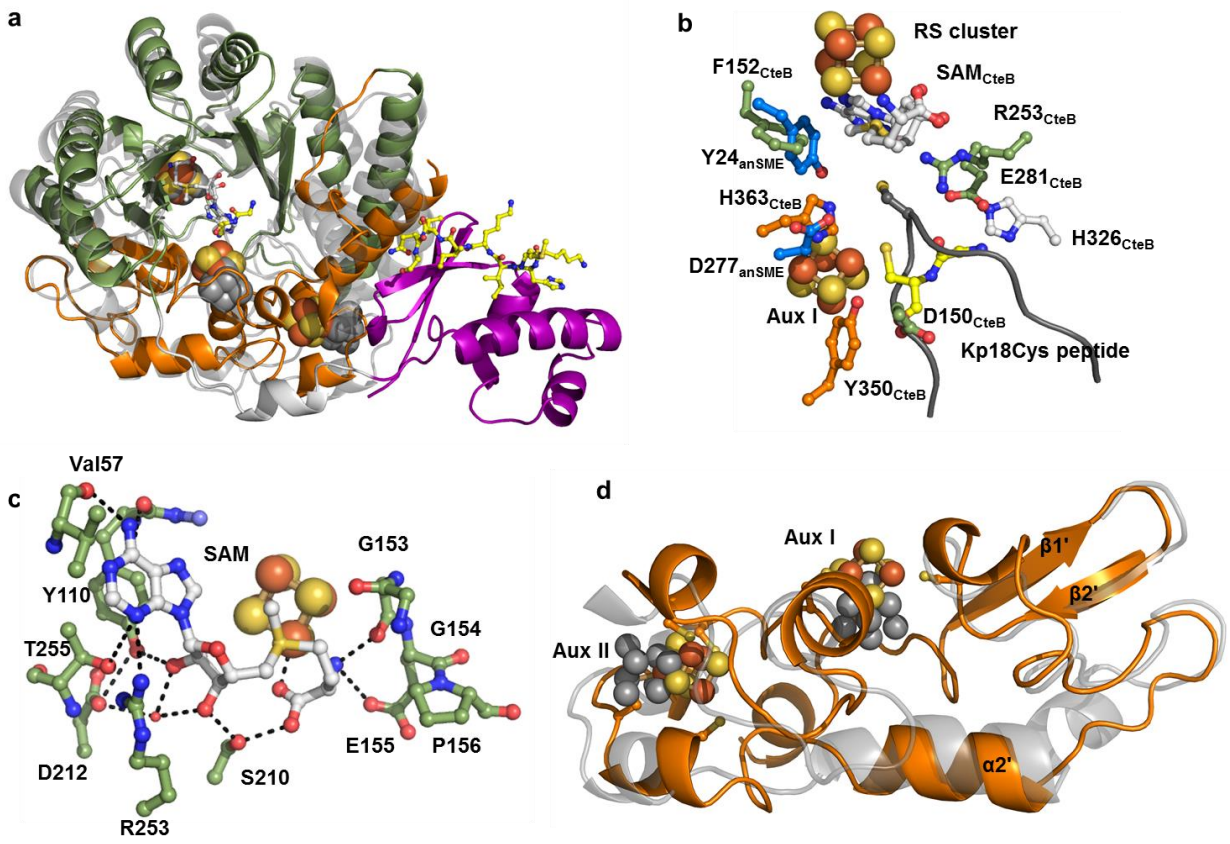


Figure A.10. Fe edge anomalous difference electron density map. The Fe anomalous difference electron density map (brown mesh) contoured at 4.0σ. SAM is shown in grey sticks with a simulated annealing omit composite map ($2F_o$-F_c) contoured to 1.5 σ.

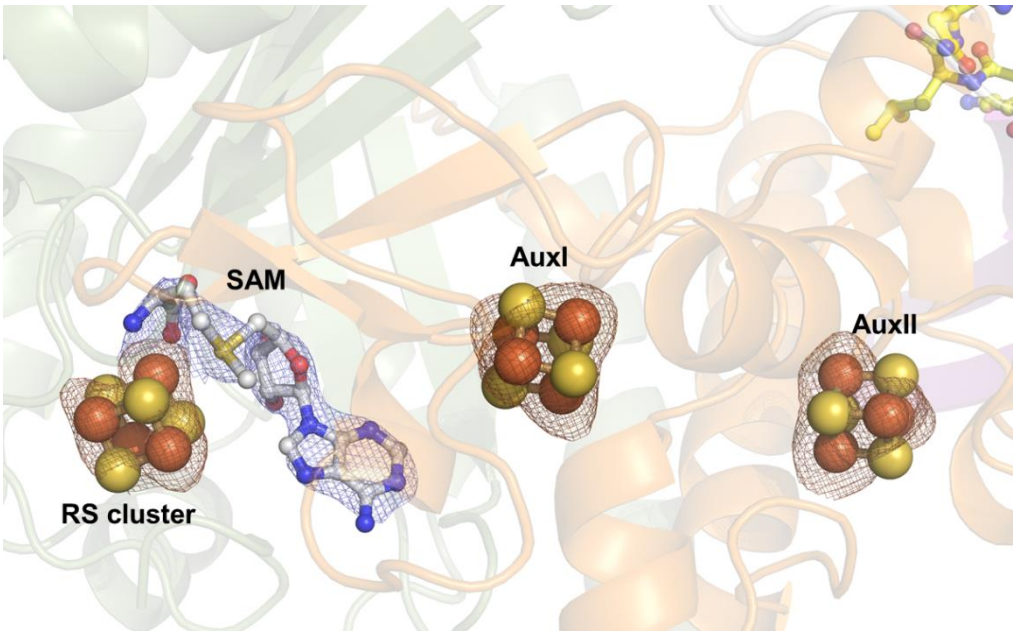


Figure A.11. Structural Alignment of CteB+SAM+CteA and Apo-CteB. Using Pymol, structural alignments were created between CteB+SAM+CteA and apo-CteB monomers. In color is the CteA bound-CteB monomer while in gray is the apo-CteB monomer. An overall alignment (residues 1-450) created a C_α RMSD of 1.318 $Å^2$ while an active site alignment (residues 76-450) created a C_α RMSD of 1.279 $Å^2$.

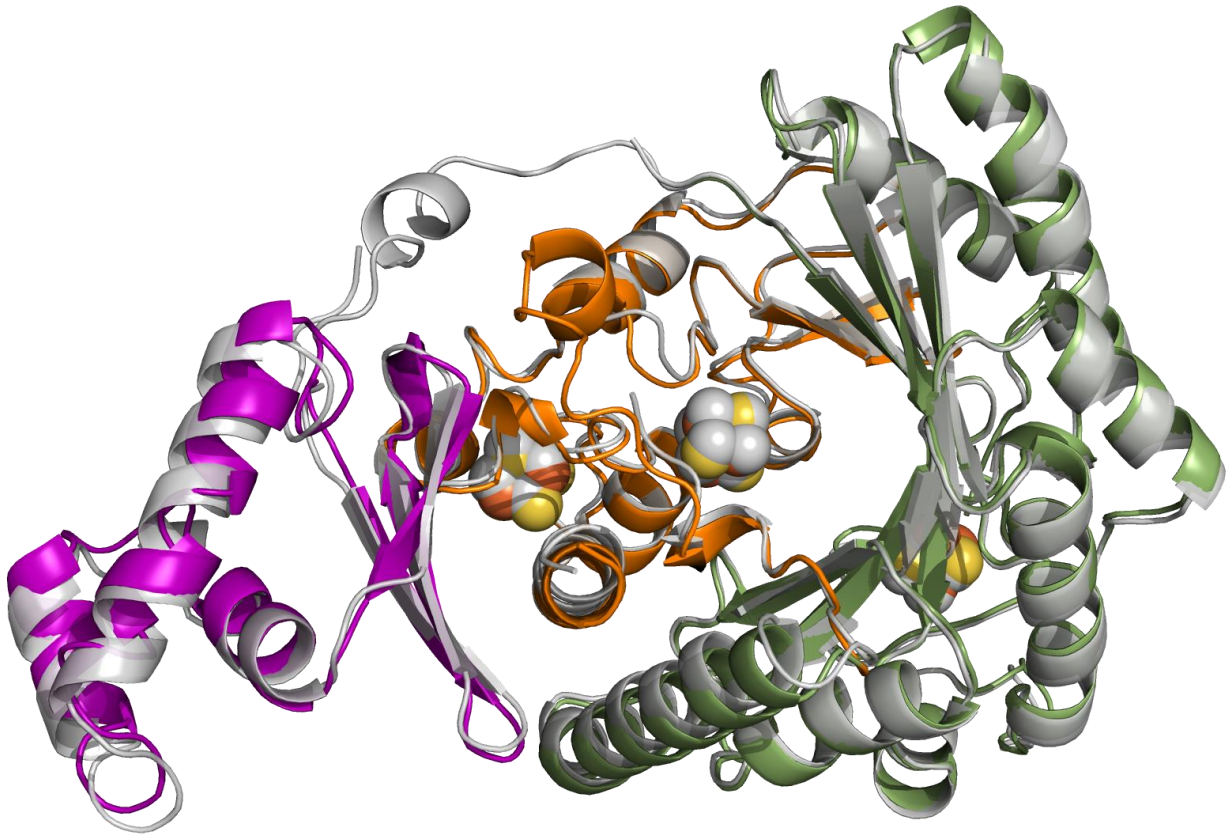


Figure A.12. Sequence Similarity Network of rSAM-Pfam (PF04055). Acquired by using the online EFI-Enzyme Similarity Tool (http://efi.igb.illinois.edu/efi-est/stepa.php).[1] Used the following parameters: E-value: -5, Alignment score of 40, 200-500 amino acids in length, 40% identity. Important rSAM enzymes marked by a big red square and labeled.

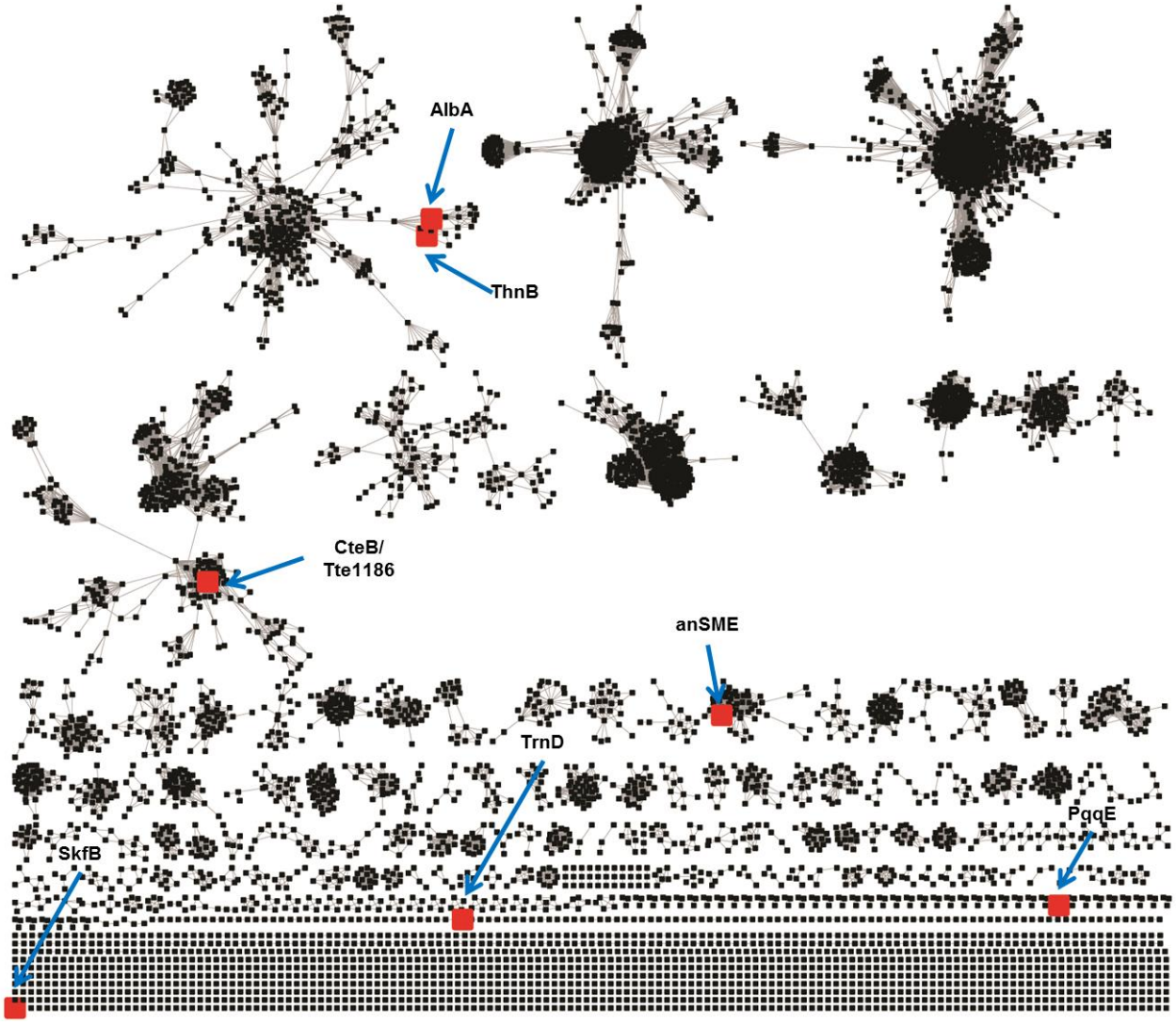


Figure A.13. Alignment of characterized radical SAM enzymes. Alignment was generated using MUSCLE with default parameters. 100% similarity = red, 75% similarity = blue, 50% similarity = gray. Black arrows point to cysteines that ligate the SAM 4Fe-4S center. Red arrows point to seven cysteine motif that corresponds to the SPASM domain in CteB.

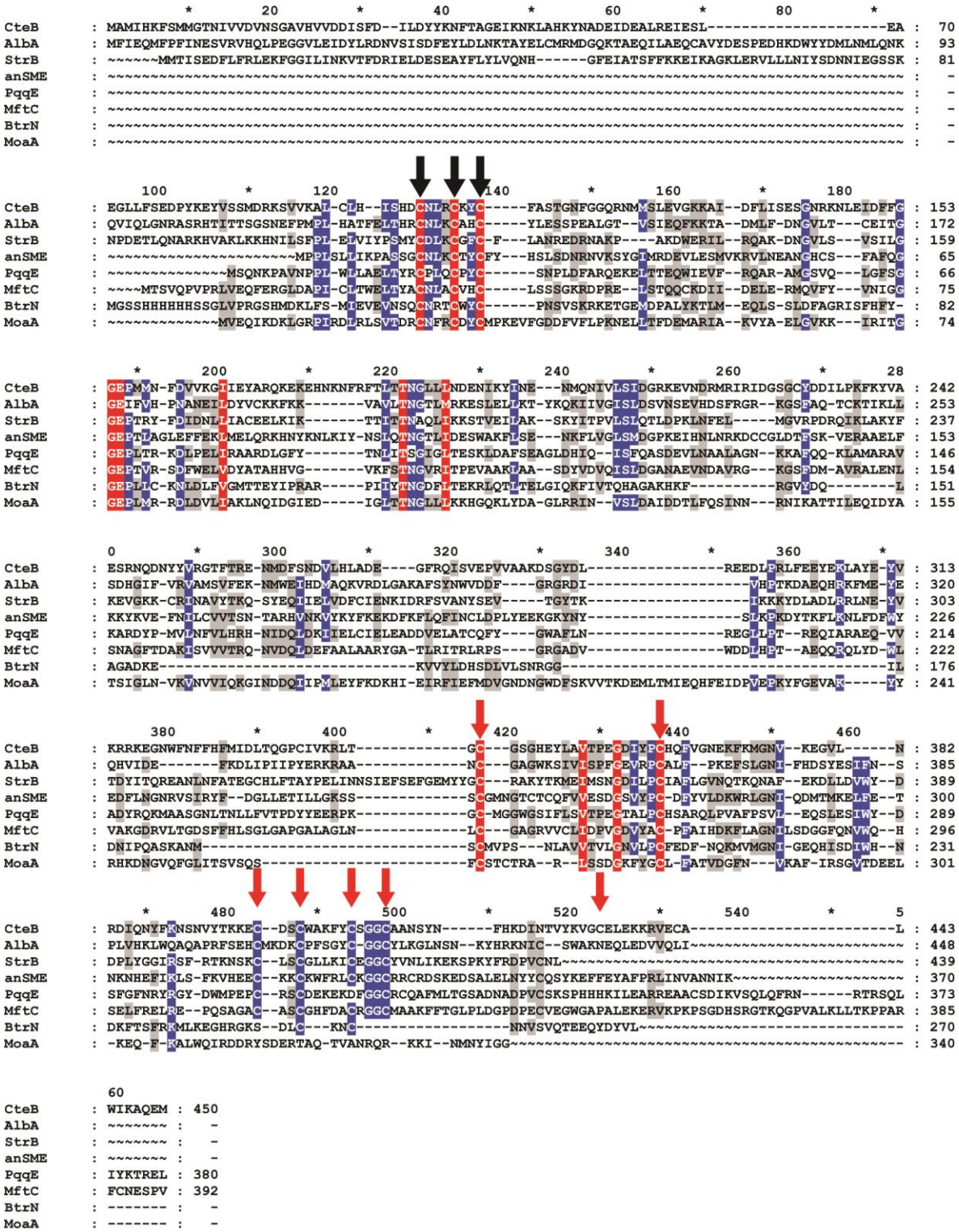


Figure A.14. Alignment of characterized sactionine synthases. Alignment was generated using MUSCLE with default parameters. 100% similarity = red, 75% similarity = blue, 50% similarity = gray. Red arrows point to seven cysteine motif that corresponds to the SPASM domain.

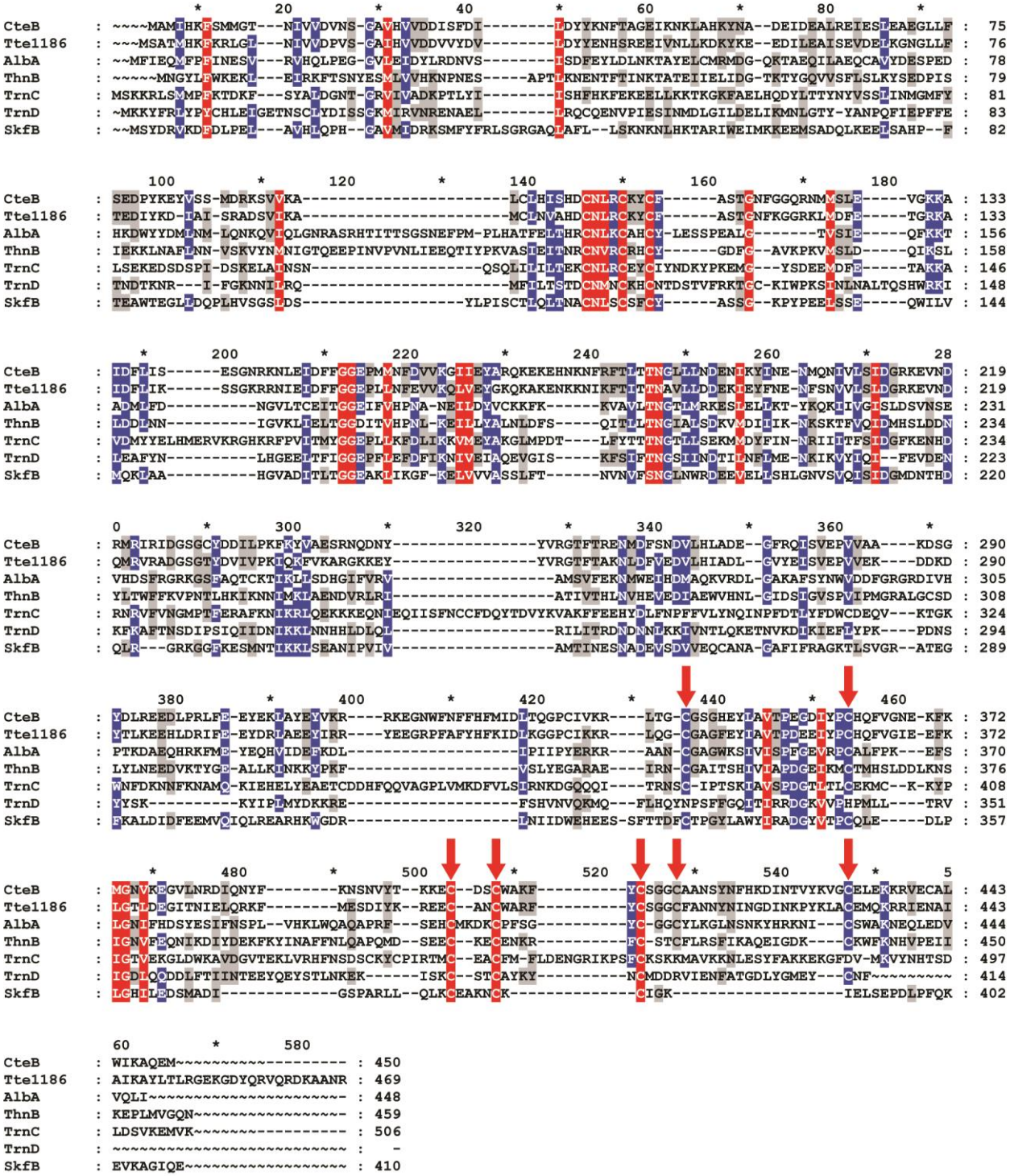


Figure A.15. SPASM Domain Logo Analysis of SCIFF radical SAM maturases (IPR024025). An analysis of the SPASM domain of SCIFF radical SAM maturases that contain relevant cysteines that ligate 4Fe-4S centers. An alignment was generated using MUSCLE with default parameters with all members of the protein family IPR024025 (952 sequences). From this alignment, the SPASM domain was located and the sequences were imported to WebLogo software (internet version 3)[2,3] to create the sequence logos below. The cysteines labeled represent the cysteines found in CteB and their respective locations in the CteB protein sequence. Cysteines 344, 362, and 413 ligate AuxI while cysteines 400, 403, 409, and 432 ligate AuxII in the CteB structure.

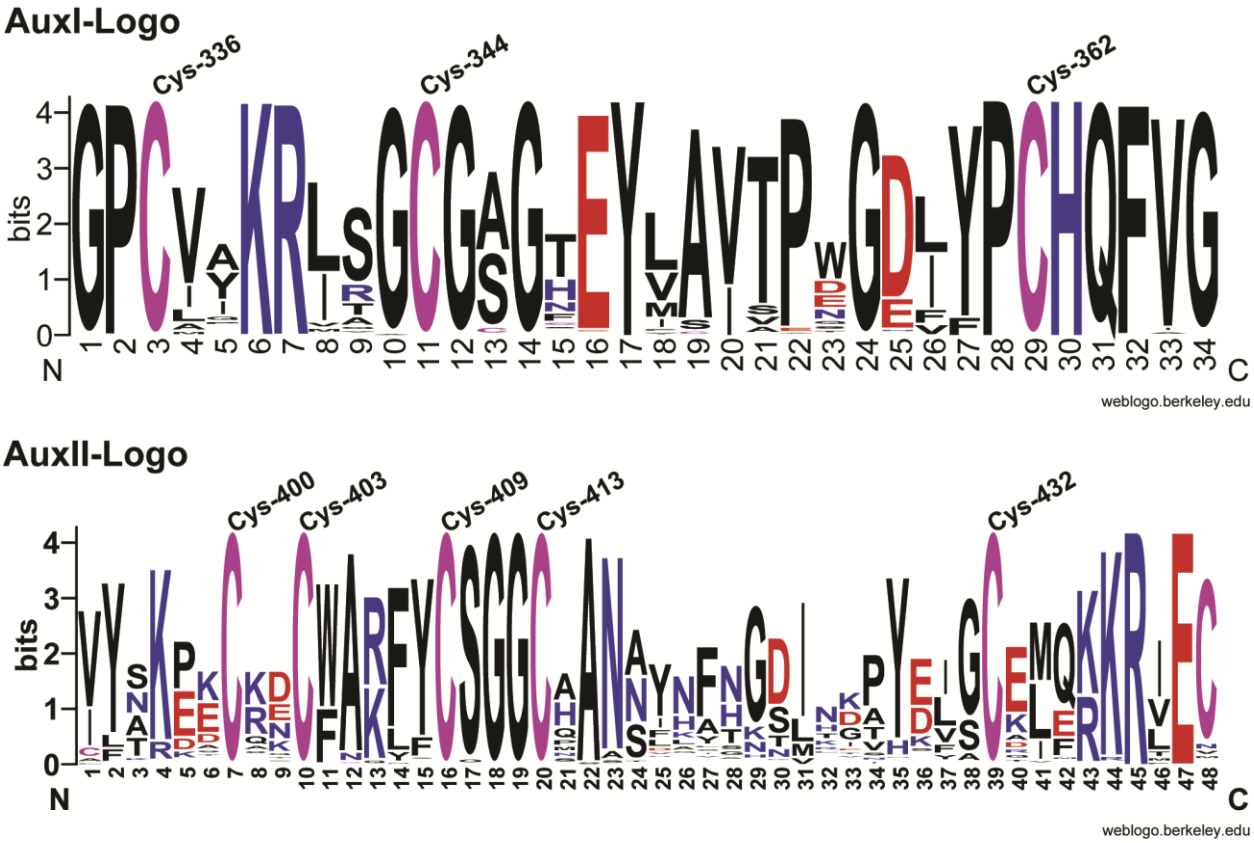
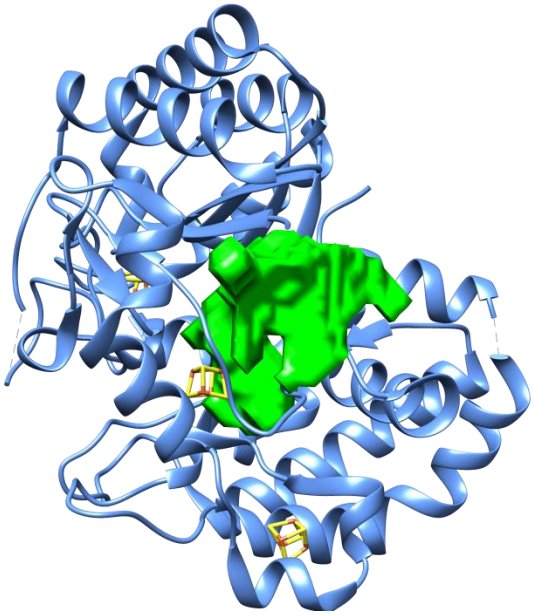


Figure A.16. Comparison of anSME and CteB active site volumes. (a and b) Calculated volumes for anSME (PDB 4K36) and CteB, respectively, were generated using the 3Vee, volume calculator (http://3vee.molmovdb.org).[4] The settings were: grid resolution-low, small probe-2.0, and big probe-9.0

a anSME- Calculated vol.= 1760Å^3



b CteB- Calculated vol.= 2997 Å^3

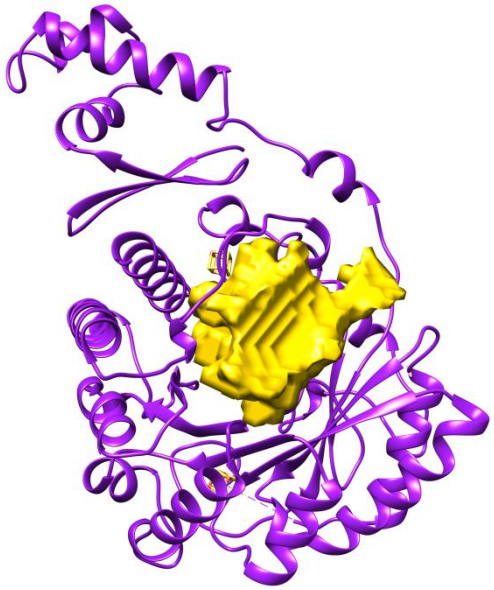


Figure A.17. Secondary Structure Prediction of CteA. Jpred was ran for a secondary structure prediction for CteA. It concurs with the presence of a helical region in the peptide present in the model. Green arrows indicate predicted β-strands while red tubes indicate predicted α-helices.

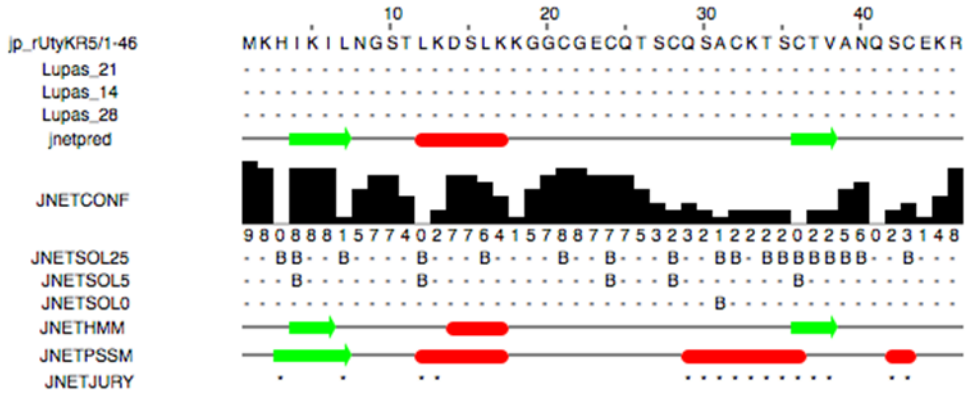
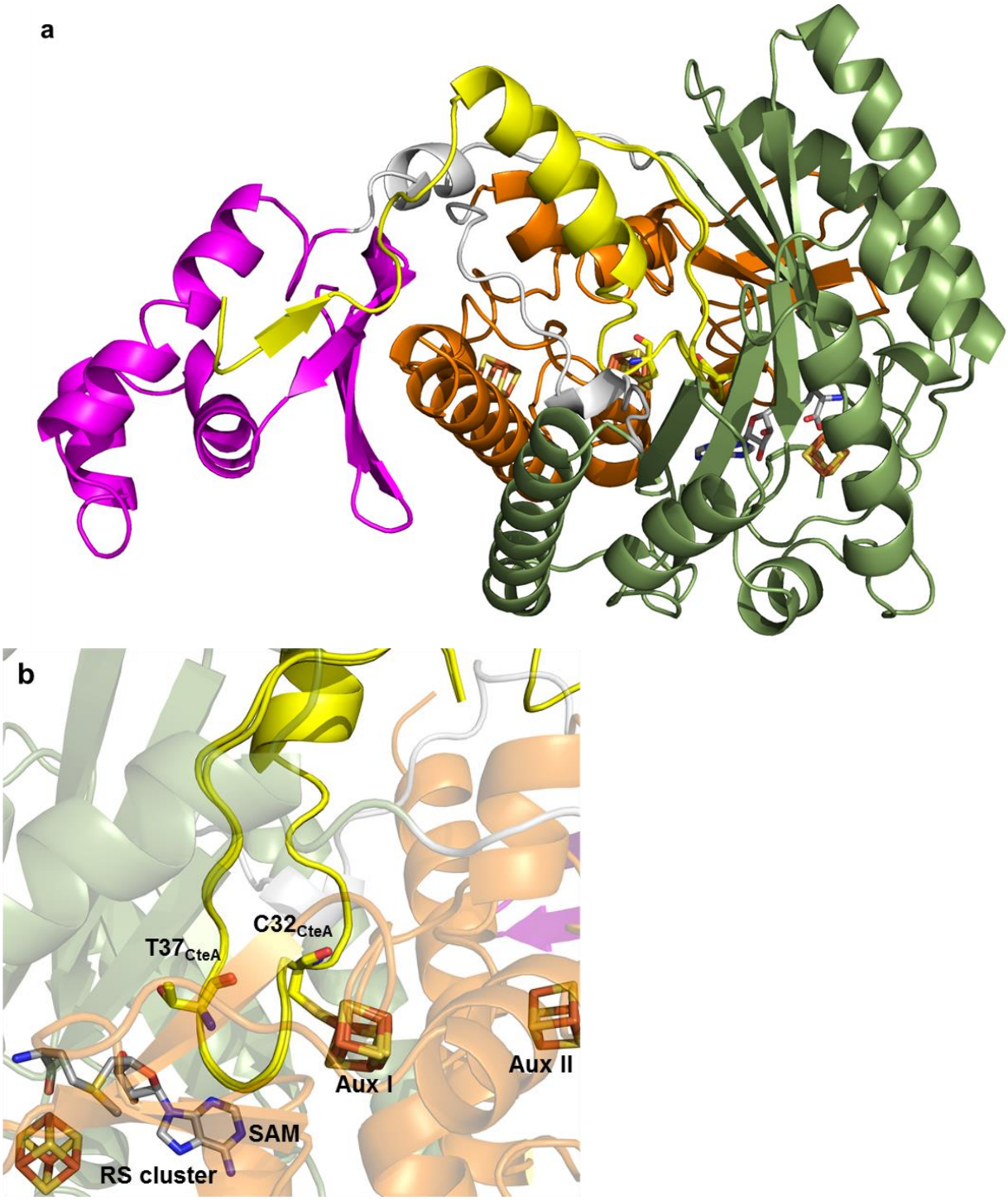


Figure A.18. Full Model of CteA bound to CteB. a) Model depicting a possible interaction of CteA with CteB using the parameters described in Experimental Section 2.7.8. CteB is colored as in Figure 2.3 while CteA is in yellow. Two lowest energy conformations are depicted here for CteA. b) Zoom in on the active site of the model. This shows CteA-C32 ligating the open coordination site in Aux I while CteA-T37 is in close proximity to SAM.



REFERENCES

1. Gerlt, J.A.; Bouvier, J.T.; Davidson, D.B.; Imker, H.J.; Sadkhin, B.; Slater, D.R.; Whalen, K.L. *Proteins Proteomics.* **2015**, *1854*, 1019

2. Crooks G.E.; Hon G.; Chandonia J.M.; Brenner S.E; *Genome Res.* **2004**, *1*4, 1188

3. Schneider TD; Stephens RM; *Nucleic Acids Res.* **1990**, *18*, 6097

4. Voss, N.R.; Gerstein, M. *Nucleic Acids Res.* **2010**, *38*, W555

APPENDIX B. SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 3

Figure B.1-B.27. Mass spectrometry (spec.) and tandem mass spec. analysis for the determination of bridging partner residues.

Figure B.1. SboA-WT

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3215	-3.99
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7395	4.28
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.5906	14.32
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.2095	-0.28
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2738	5.37
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4056	10.51
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	1432.7067	-11.94

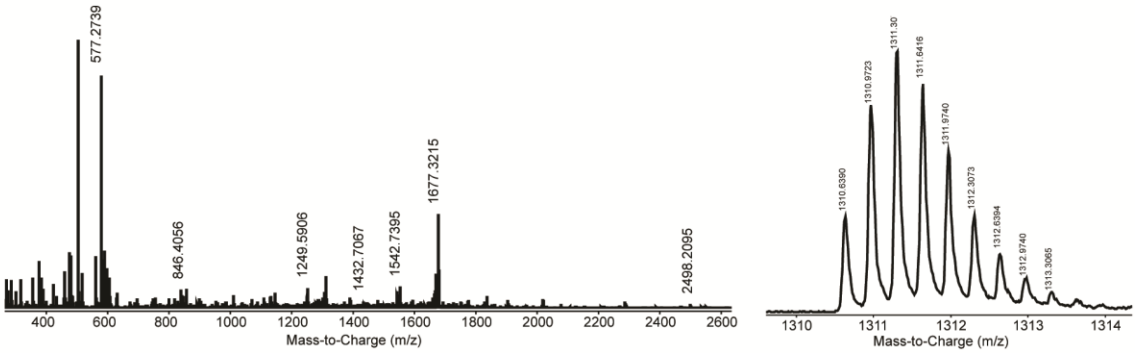


Figure B.2. SboA-F22A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{140}H_{230}N_{37}O_{47}S_3$	1639.2992	1639.2217	47.28
-5-27 (b_{27})	$C_{128}H_{211}N_{34}O_{43}S_3$	1504.7304	1504.7638	-22.20
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.5856	18.33
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.2966	-35.15
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2678	15.76
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4328	-21.62
22-35 (y_{22})	$C_{64}H_{90}N_{15}O_{18}$	1356.6583	1356.7243	-48.65

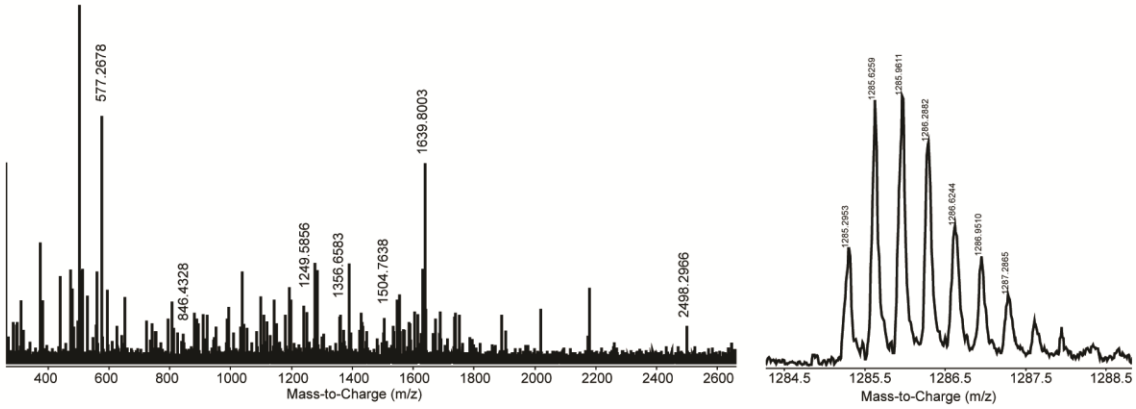


Figure B.3. SboA-F22M

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{142}H_{234}N_{37}O_{47}S_4$	1669.3009	1669.3079	-4.19
-5-27 (b_{27})	$C_{130}H_{215}N_{34}O_{43}S_4$	1534.7321	1534.7184	8.93
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.5921	13.12
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2784	-2.60
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	N/A	N/A
22-35 (y_{22})	$C_{66}H_{94}N_{15}O_{18}S$	1416.6616	N/A	N/A

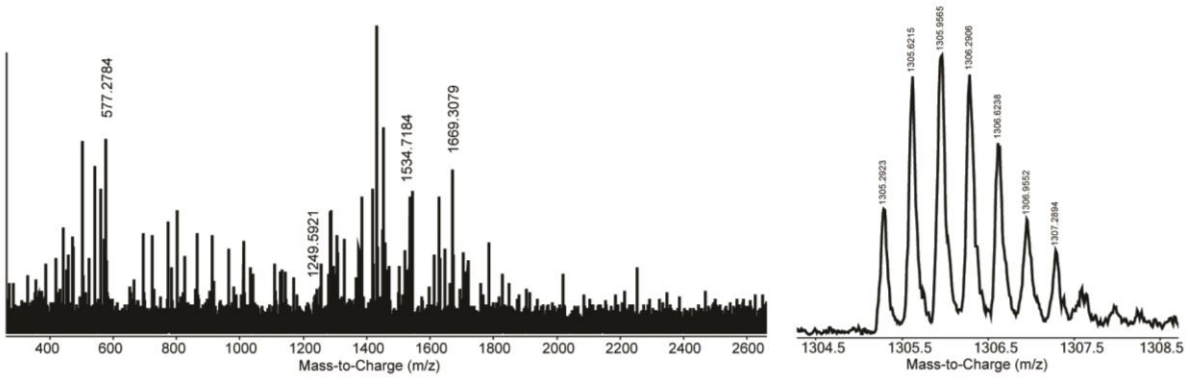


Figure B.4: SboA-T28A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{145}H_{232}N_{37}O_{46}S_3$	1662.3096	1662.3260	-9.87
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7264	12.77
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6199	9.13
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.1112	39.07
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2775	-1.04
28-35 (y_{28})	$C_{41}H_{54}N_9O_9$	816.4039	816.4012	3.31
22-35 (y_{22})	$C_{69}H_{92}N_{15}O_{17}$	1402.6790	N/A	N/A

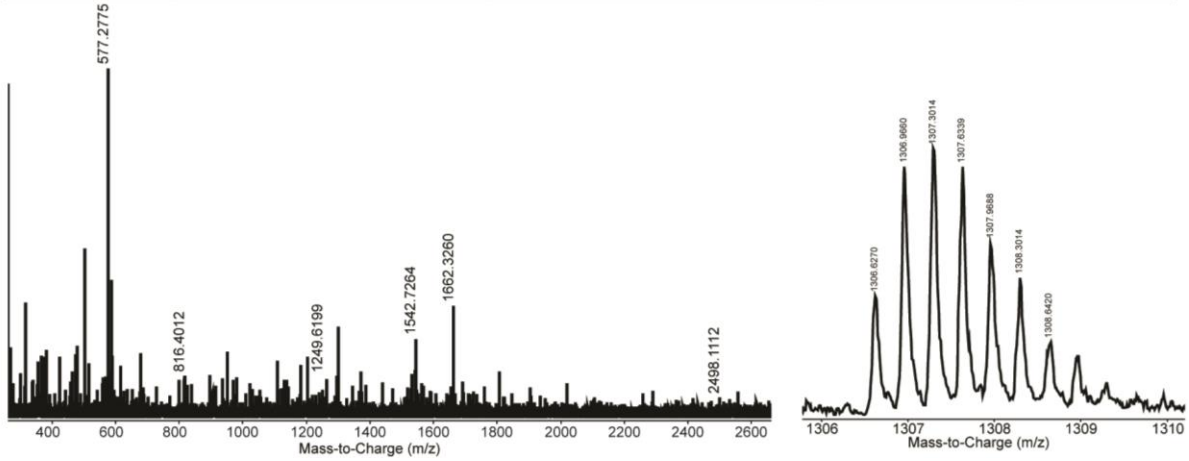


Figure B.5. SboA-T28F

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{151}H_{236}N_{37}O_{46}S_3$	1700.3252	1700.3276	-1.41
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7799	-21.91
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	N/A	N/A
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2742	4.68
28-35 (y_{28})	$C_{47}H_{58}N_9O_9$	892.4352	892.4416	-7.17
22-35 (y_{22})	$C_{75}H_{96}N_{15}O_{17}$	1478.7103	N/A	N/A

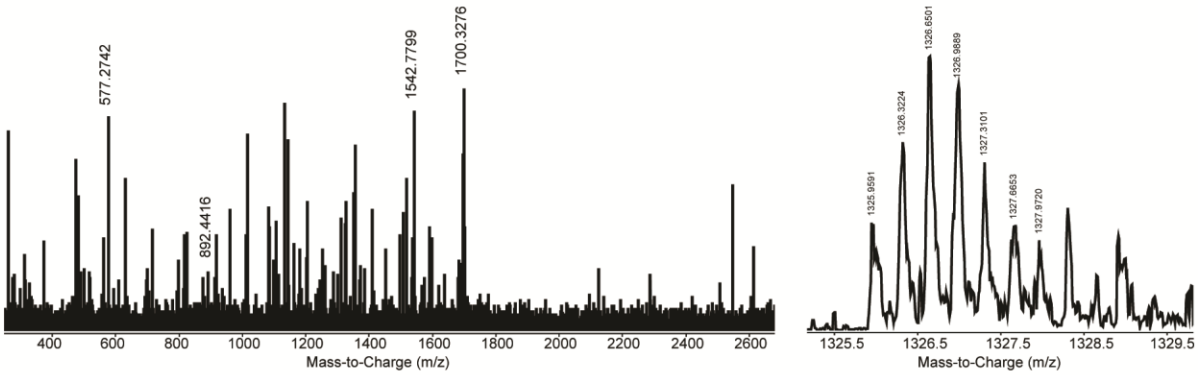


Figure B.6. SboA-T28N

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{232}N_{37}O_{48}S_3$	1684.3045	1684.2964	4.81
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7045	26.96
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	N/A	N/A
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2707	10.74
28-35 (y_{28})	$C_{42}H_{54}N_9O_{11}$	860.3937	N/A	N/A
22-35 (y_{22})	$C_{70}H_{92}N_{15}O_{19}$	1446.6688	1446.6406	19.49

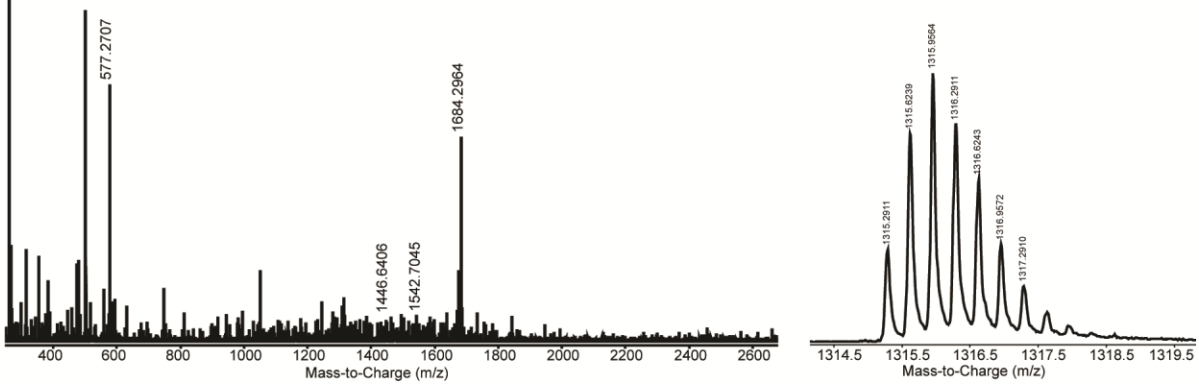


Figure B.7. SboA-T28S

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{145}H_{232}N_{37}O_{47}S_3$	1670.3070	1670.2938	7.90
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7234	14.71
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6305	17.61
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.0999	43.59
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2653	20.09
28-35 (y_{28})	$C_{41}H_{54}N_9O_{10}$	832.3988	N/A	N/A
22-35 (y_{22})	$C_{69}H_{92}N_{15}O_{18}$	1418.6739	N/A	N/A

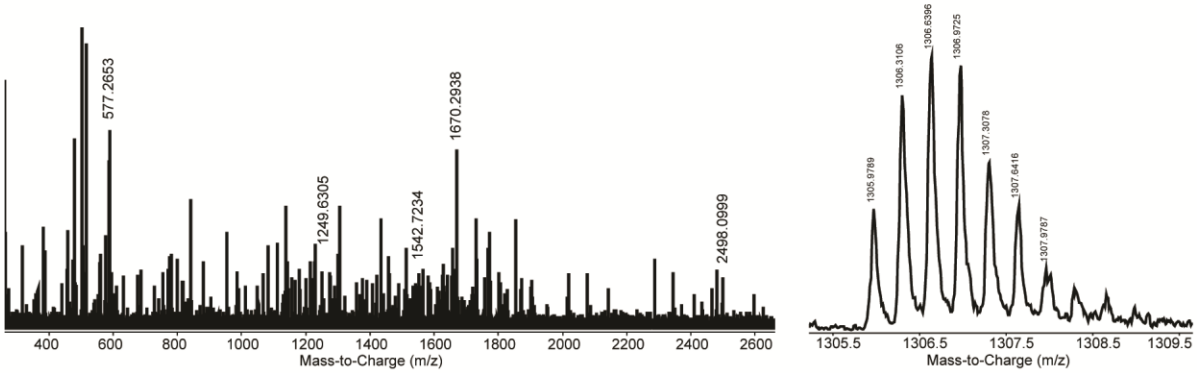


Figure B.8. SboA-T27/A28 Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3032	6.92
-5-27 (b_{27})	$C_{135}H_{217}N_{34}O_{44}S_3$	1557.7514	1557.7161	22.37
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.5992	7.44
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2717	9.01
28-35 (y_{28})	$C_{41}H_{54}N_9O_9$	816.4039	816.3901	16.90
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	N/A	N/A

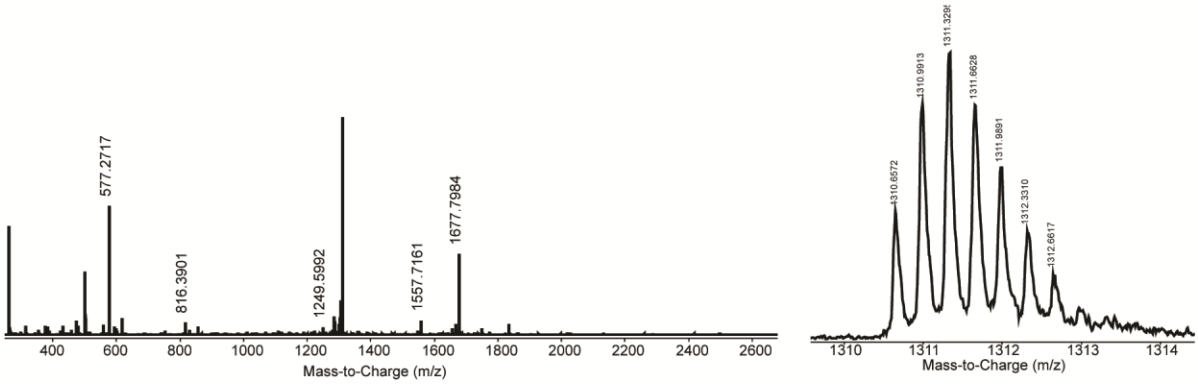


Figure B.9. SboA-F31A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3156	-0.48
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.8002	-35.07
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6515	-34.41
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{24}H_{33}N_6O_6$	501.2456	501.2506	-9.98
28-35 (y_{28})	$C_{36}H_{52}N_9O_{10}$	770.3832	770.3635	25.57
22-35 (y_{22})	$C_{64}H_{90}N_{15}O_{18}$	1356.6583	N/A	N/A

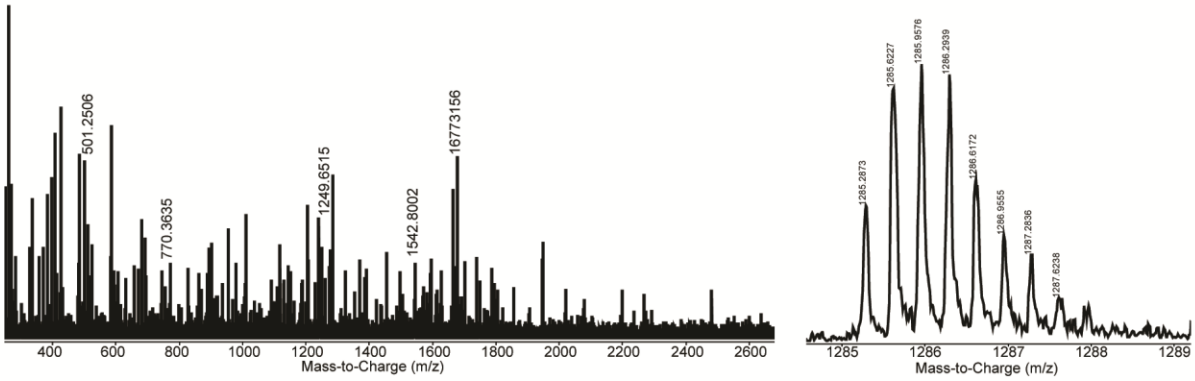


Figure B.10. SboA-F31S

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.2930	13.00
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.8559	-71.17
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6135	-4.00
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.2607	-20.77
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{24}H_{33}N_6O_7$	517.2405	517.2346	11.41
28-35 (y_{28})	$C_{36}H_{52}N_9O_{11}$	786.3781	786.3845	-8.14
22-35 (y_{22})	$C_{64}H_{90}N_{15}O_{19}$	1372.6532	1372.7016	-35.26

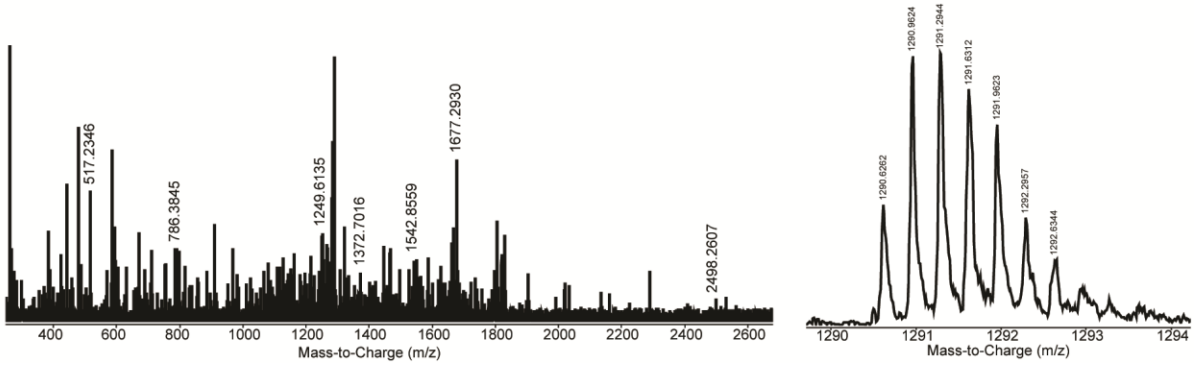


Figure B.11. SboA-G26A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{147}H_{236}N_{37}O_{47}S_3$	1684.3227	1684.3495	-15.91
-5-27 (b_{27})	$C_{135}H_{217}N_{34}O_{43}S_3$	1549.7539	N/A	N/A
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6306	-17.69
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2697	12.47
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4794	-76.68
22-35 (y_{22})	$C_{71}H_{96}N_{15}O_{18}$	1446.7052	N/A	N/A

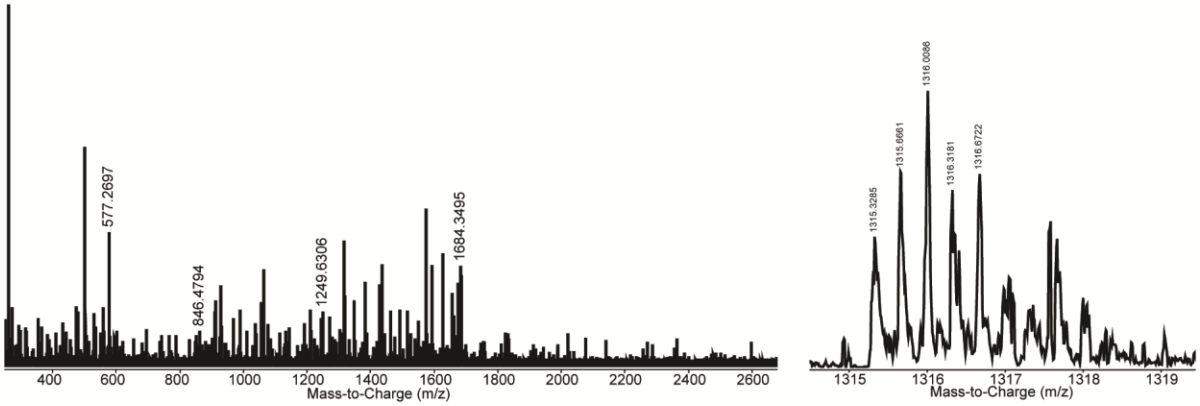


Figure B.12. SboA-G32A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3123	1.49
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7403	3.76
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	N/A	N/A
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.2042	1.84
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{31}H_{39}N_6O_6$	591.2926	591.2945	-3.21
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	N/A	N/A
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	N/A	N/A

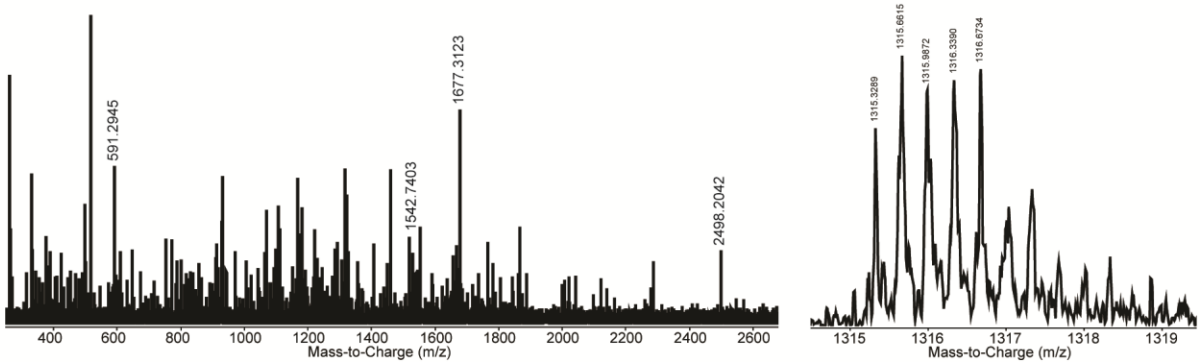


Figure B.13. SboA-G26A, G32A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{147}H_{236}N_{37}O_{47}S_3$	1684.3227	1684.3277	-2.97
-5-27 (b_{27})	$C_{135}H_{217}N_{34}O_{43}S_3$	1549.7539	1549.8105	-36.52
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.5972	9.04
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{31}H_{39}N_6O_6$	591.2926	591.2999	-12.28
28-35 (y_{28})	$C_{43}H_{58}N_9O_{10}$	860.4301	N/A	N/A
22-35 (y_{22})	$C_{72}H_{98}N_{15}O_{18}$	1460.7209	N/A	N/A

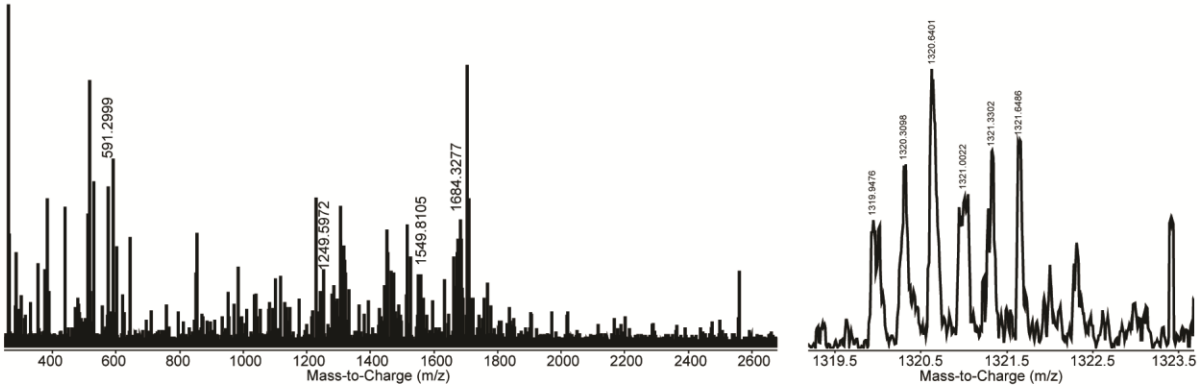


Figure B.14. SboA-G26A, G29A, G32A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{148}H_{238}N_{37}O_{47}S_3$	1691.3305	1691.3430	-7.39
-5-27 (b_{27})	$C_{135}H_{217}N_{34}O_{43}S_3$	1549.7539	1549.7429	7.10
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6099	-1.12
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.1837	10.05
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{31}H_{39}N_6O_6$	591.2926	591.2977	-8.63
28-35 (y_{28})	$C_{44}H_{60}N_9O_{10}$	874.4458	N/A	N/A
22-35 (y_{22})	$C_{73}H_{100}N_{15}O_{18}$	1474.7365	N/A	N/A

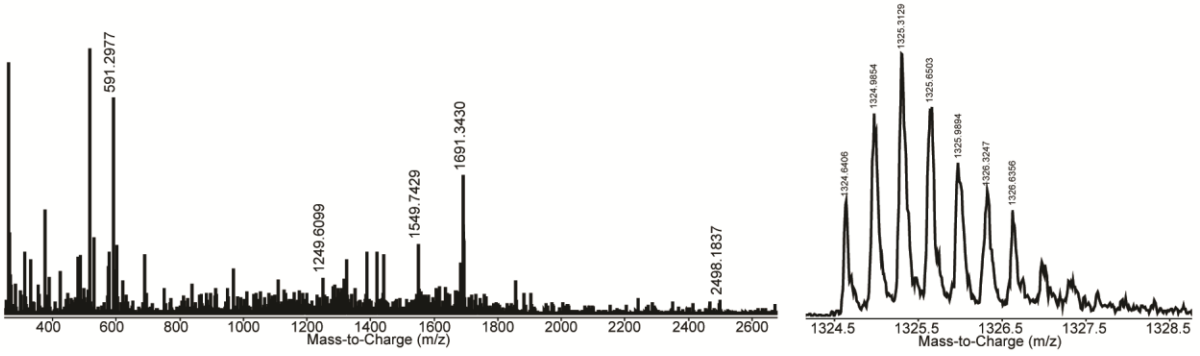


Figure B.15. SboA-P18A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{144}H_{232}N_{37}O_{47}S_3$	1664.3070	1664.2965	6.31
-5-27 (b_{27})	$C_{132}H_{213}N_{34}O_{43}S_3$	1529.7382	1529.6821	36.67
-5-21 (b_{21})	$C_{104}H_{175}N_{28}O_{35}S_3$	1236.3007	N/A	N/A
-5-21 (b_{21})	$C_{104}H_{175}N_{28}O_{35}S_3$	2472.1931 (M^+)	2472.2168	-9.59
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2684	14.72
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.3928	25.64
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	N/A	N/A

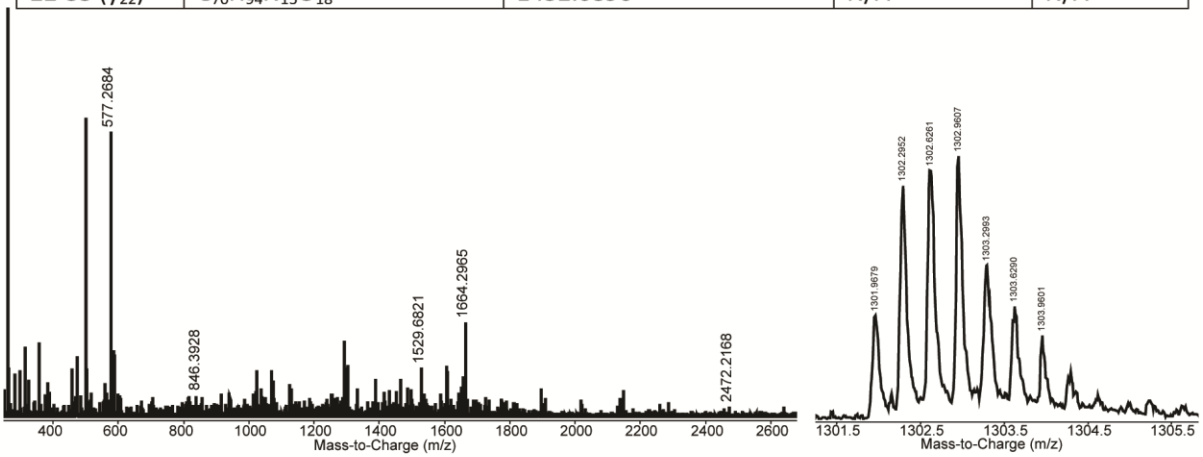


Figure B.16. SboA-P20A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{144}H_{232}N_{37}O_{47}S_3$	1664.3070	1664.3132	-3.73
-5-27 (b_{27})	$C_{132}H_{213}N_{34}O_{43}S_3$	1529.7382	1529.6809	37.46
-5-21 (b_{21})	$C_{104}H_{175}N_{28}O_{35}S_3$	1236.3007	N/A	N/A
-5-21 (b_{21})	$C_{104}H_{175}N_{28}O_{35}S_3$	2472.1931 (M^+)	2472.1816	4.65
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2846	-27.77
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4349	-24.10
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	N/A	N/A

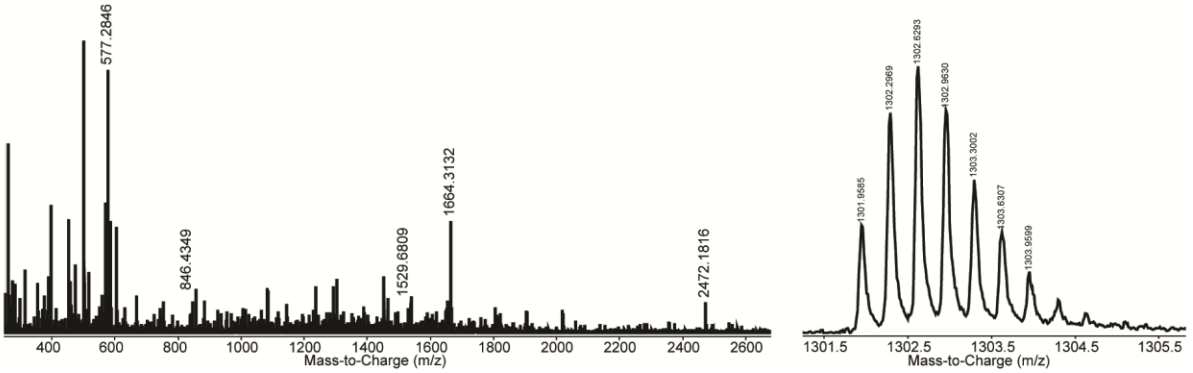


Figure B.17. SboA-P18A, P20A

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{142}H_{230}N_{37}O_{47}S_3$	1651.2992	1651.2857	8.18
-5-27 (b_{27})	$C_{130}H_{211}N_{34}O_{43}S_3$	1516.7304	1516.7611	-20.24
-5-21 (b_{21})	$C_{102}H_{173}N_{28}O_{35}S_3$	1223.5929	1223.6054	-10.22
-5-21 (b_{21})	$C_{102}H_{173}N_{28}O_{35}S_3$	2446.1775 (M^+)	2446.0967	33.03
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2796	-4.68
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4130	1.77
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	1432.7467	-39.86

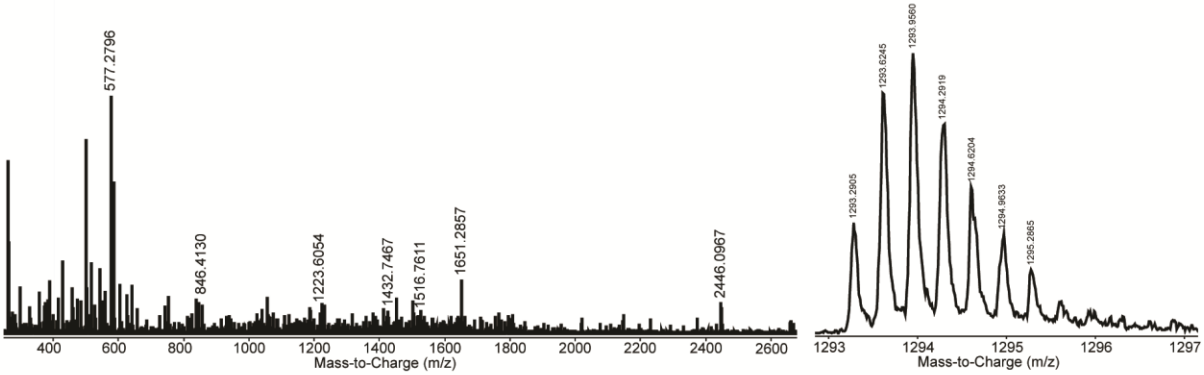


Figure B.18. SboA-ΔP18

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{29})	$C_{141}H_{227}N_{36}O_{46}S_3$	1628.7885	1628.8002	-7.18
-5-27 (b_{26})	$C_{129}H_{208}N_{33}O_{42}S_3$	1494.2197	1494.2194	0.20
-5-21 (b_{18})	$C_{92}H_{158}N_{25}O_{30}S_3$	1095.0423	1095.0474	-4.66
-5-21 (b_{18})	$C_{92}H_{158}N_{25}O_{30}S_3$	2189.0763(M^+)	2189.0961	-9.04
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{30})	$C_{30}H_{37}N_6O_6$	577.2769	577.2775	-1.04
28-35 (y_{27})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4162	-2.01
22-35 (y_{19})	$C_{79}H_{106}N_{17}O_{22}$	1644.7693	1644.7927	-14.23

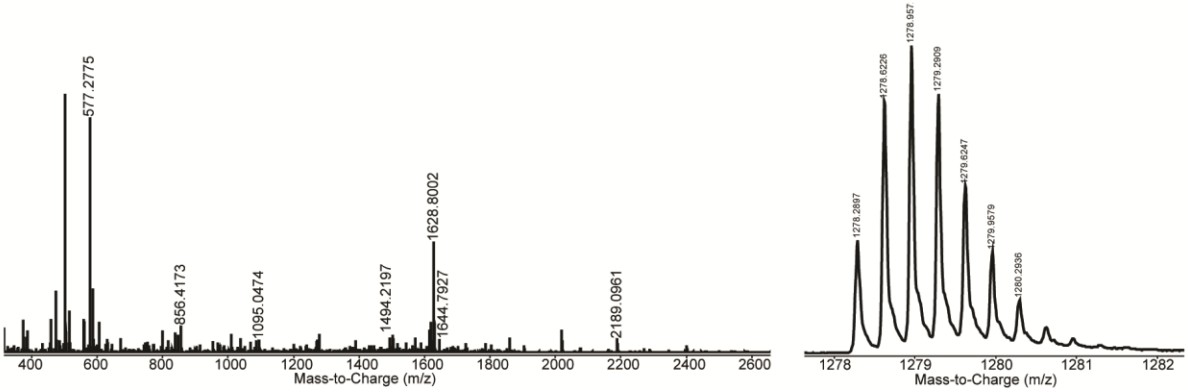


Figure B.19. SboA-ΔP20

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{29})	$C_{141}H_{227}N_{36}O_{46}S_3$	1628.7885	1628.7492	24.13
-5-27 (b_{26})	$C_{129}H_{208}N_{33}O_{42}S_3$	1494.2197	N/A	N/A
-5-21 (b_{20})	$C_{101}H_{170}N_{27}O_{34}S_3$	1201.0821	1201.0488	27.72
-5-21 (b_{20})	$C_{101}H_{170}N_{27}O_{34}S_3$	2401.1560(M^+)	2401.2187	-26.11
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{30})	$C_{30}H_{37}N_6O_6$	577.2769	577.2835	-11.43
28-35 (y_{27})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4216	-8.39
22-35 (y_{21})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	N/A	N/A

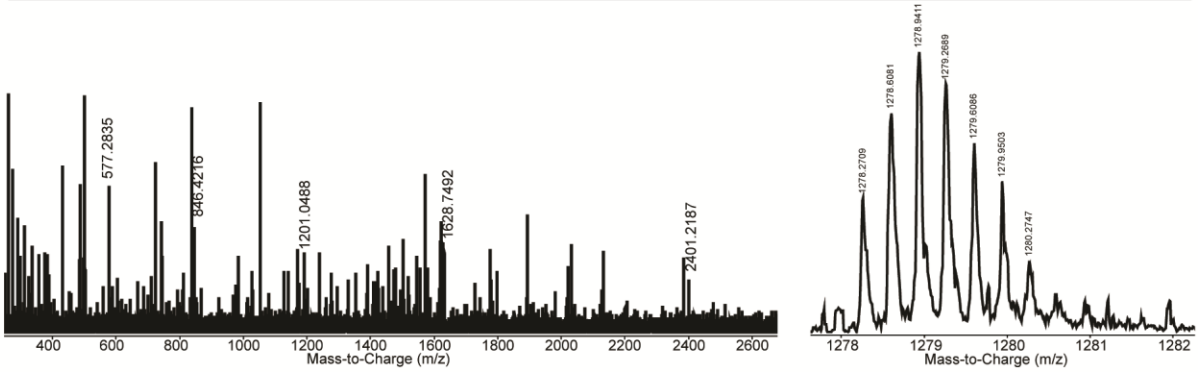


Figure B.20. SboA-C3/G4 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{158}H_{252}N_{39}O_{51}S_3$	1804.3780	1804.4137	-19.79
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2775	-1.04

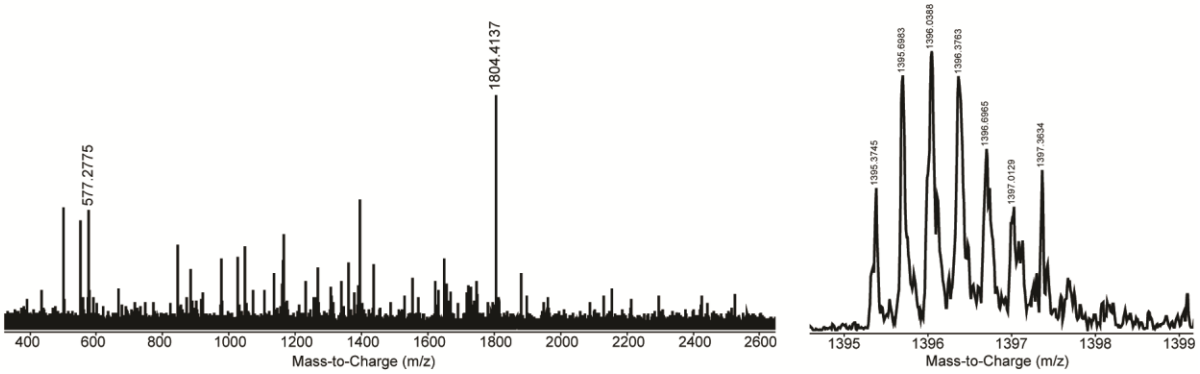


Figure B.21. SboA-S7/C8 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{158}H_{252}N_{39}O_{51}S_3$	1804.3780	1804.3694	4.77
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2839	-12.13

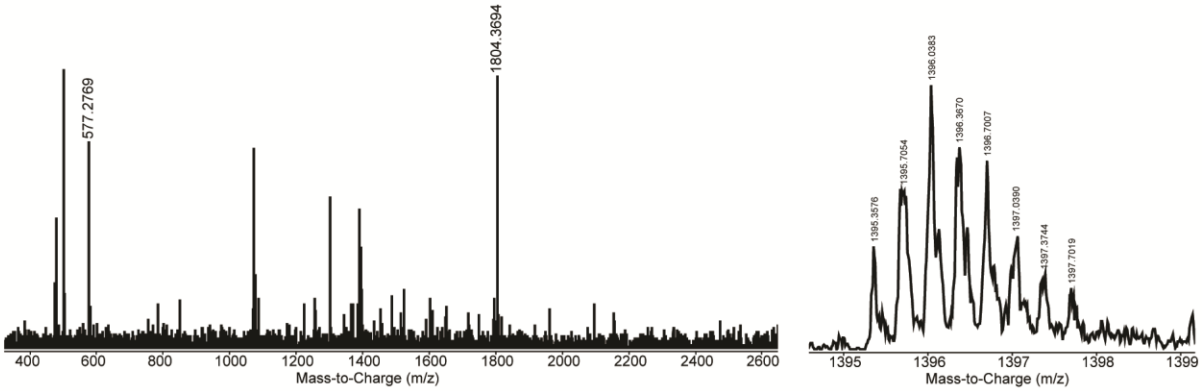


Figure B.22. SboA-G7/C10 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{158}H_{252}N_{39}O_{51}S_3$	1804.3780	1804.3865	-4.71
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2709	10.39

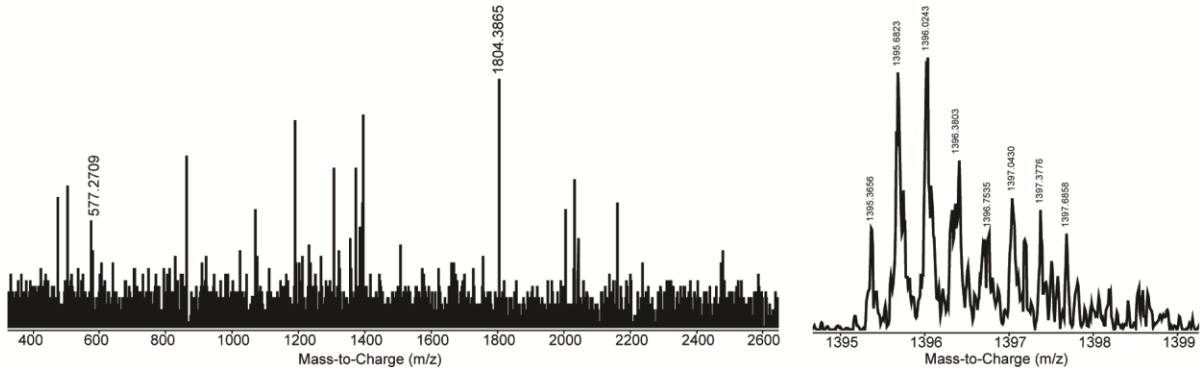


Figure B.23. SboA-C10/G13 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3156	-0.48
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7622	-10.44
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6794	-56.58
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2686	14.38
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4093	6.14
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	N/A	N/A

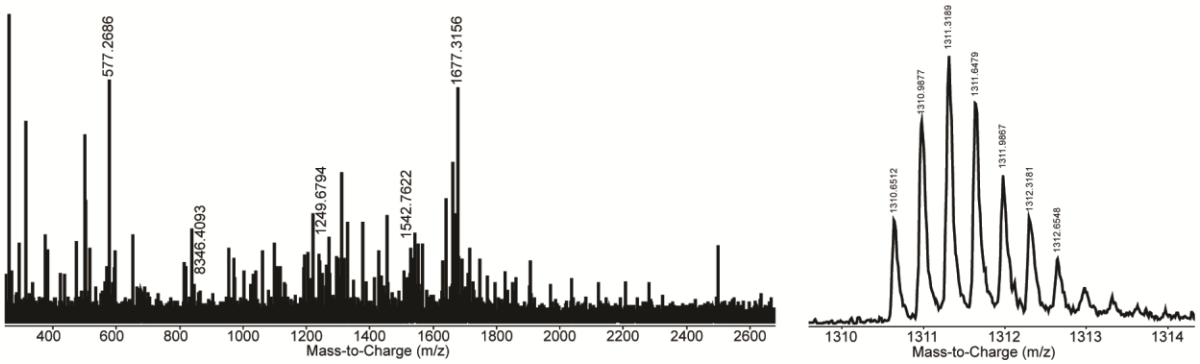


Figure B.24. SboA-C12/A13 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3663	-30.70
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7933	-30.59
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6431	-27.69
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.3927	-73.61
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2792	-3.98
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	N/A	N/A
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	1432.6528	25.69

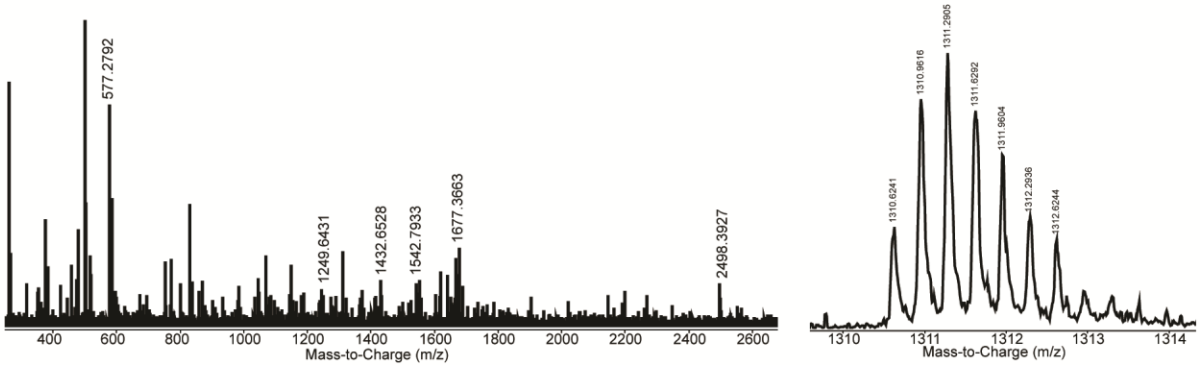


Figure B.25. SboA-L13/C14 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.2857	17.35
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7592	-8.49
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.5616	37.53
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.1805	11.33
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2670	17.15
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.4136	1.06
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	1432.6408	34.06

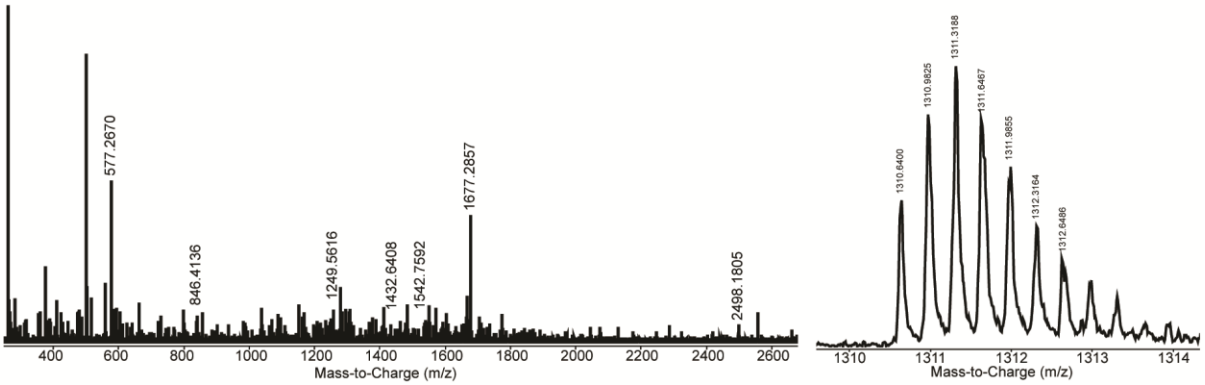


Figure B.26. SboA-V13/C15 Cysteine Swap

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.2100	62.48
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7339	7.91
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6122	-2.96
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	N/A	N/A
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{30}H_{37}N_6O_6$	577.2769	577.2631	23.91
28-35 (y_{28})	$C_{42}H_{56}N_9O_{10}$	846.4145	846.3429	84.59
22-35 (y_{22})	$C_{70}H_{94}N_{15}O_{18}$	1432.6896	1432.6904	-0.56

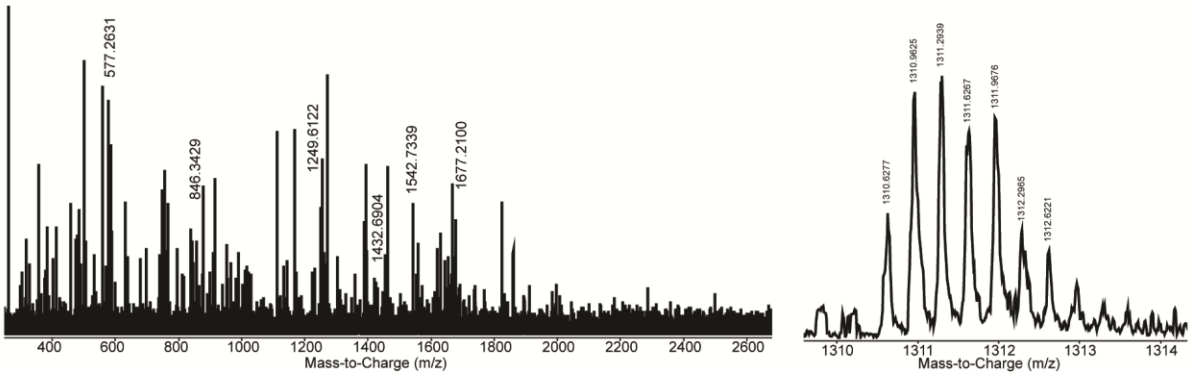


Figure B.27. SboA-F31TAG-*O*-Me-Tyr

Amino Acids	Proposed Formula (M^+)	Calculated Mass $(M+H)^{+2}$	Observed Mass	Error (ppm)
-5-30 (b_{30})	$C_{146}H_{234}N_{37}O_{47}S_3$	1677.3148	1677.3211	-3.76
-5-27 (b_{27})	$C_{134}H_{215}N_{34}O_{43}S_3$	1542.7461	1542.7790	-21.33
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	1249.6085	1249.6779	-55.54
-5-21 (b_{21})	$C_{106}H_{177}N_{28}O_{35}S_3$	2498.2088 (M^+)	2498.2178	-3.60
Amino Acids	Proposed Formula $(M+H)^+$	Calculated Mass $(M+H)^+$	Observed Mass	Error (ppm)
31-35 (y_{31})	$C_{31}H_{39}N_6O_7$	607.2875	607.2728	24.21
28-35 (y_{28})	$C_{43}H_{58}N_9O_{11}$	876.4250	N/A	N/A
22-35 (y_{22})	$C_{71}H_{96}N_{15}O_{19}$	1462.7001	N/A	N/A

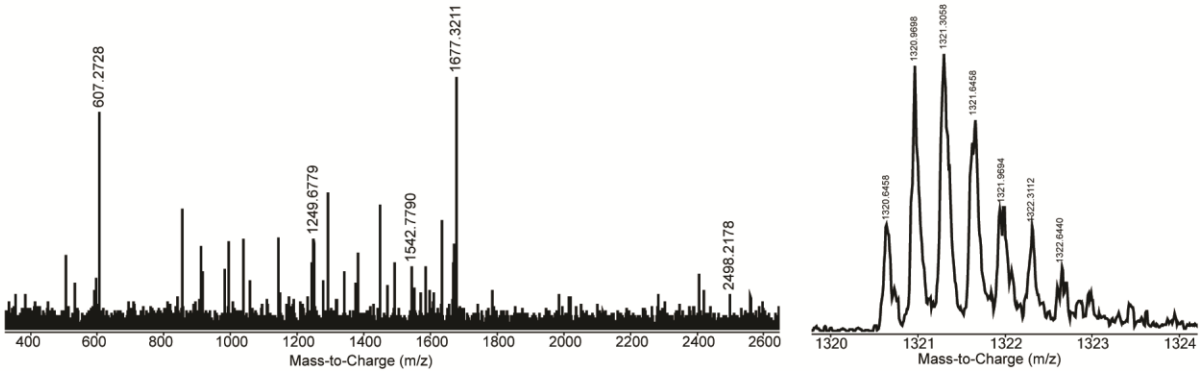


Figure B.28. 12%-SDS-PAGE gel showing increased expression of AlbA in vector pETDuet-SboA-AlbA with corrector plasmid pPH151 in BL21 cells. Time points shown at 0 and 20 hours of induction with 0.5 mM IPTG.

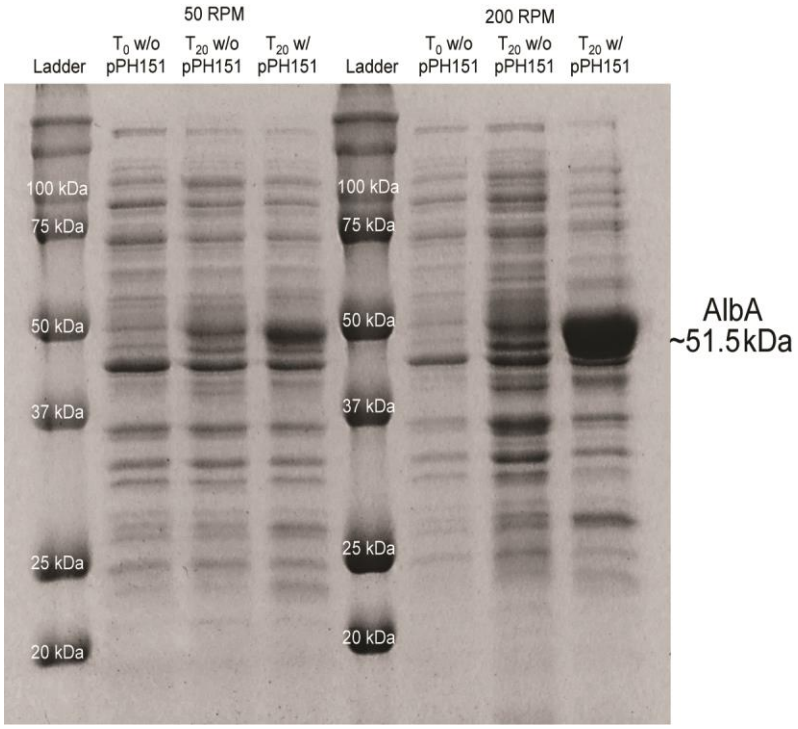


Figure B.29. UV comparison of the natural product subtilosin A vs. the wild-type (WT) and several mutants produced in the pETDuet system at 50 rpm. The UV was monitored at 220 nm.

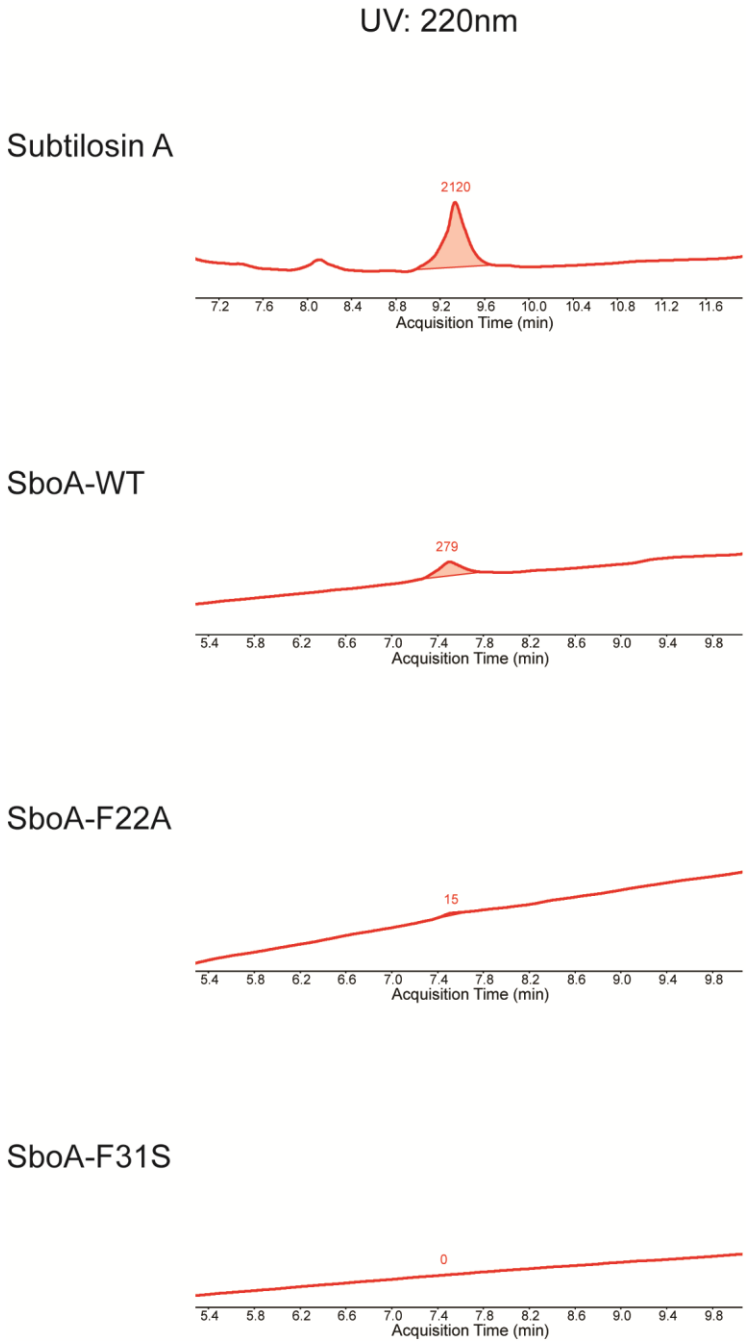


Table B.1. List of mutants made and produced in this study. Additional columns show the area under the curves of the EICs and their respected ratios with respect to (w.r.t.) the wild-type (WT) EIC. EICs shown in **Figures B.30-B.59**.

Mutant #	N-term changes	C-term changes	Bridges (#)	Bridging Partners	Area (Production	Ratio (w.r.t. WT)
1		F22A	3	A22, T28, F31	253817	0.11
2		F22M	3	M22, T28, F31	496203	0.21
3		F22S	no product	--		
4		F22T	no product	--		
5		T28A	3	F22, A28, F31	610516	0.26
6		T28F	3	F22, F28, F31	548025	0.23
7		T28N	3	F22, N28, F31	1159393	0.49
8		T28S	3	F22, S28, F31	347541	0.15
9		T27/A28	3	F22, A28, F31	3274059	1.39
10		F31A	3	F22, T28, A31	1103271	0.47
11		F31M	no product	--		
12		F31N	no product	--		
13		F31S	3	F22, T28, S31	228963	0.10
14		F31T	no product	--		
15		G26A	3	F22, T28, F31	448021	0.19
16		G29A	3	--	639210	0.27
17		G32A	3	F22, T28, F31	939142	0.40
18		G26A,G29A	3	--	519791	0.22
19		G26A,G32A	3	F22, T28, F31	596449	0.25
20		G29A,G32A	3	--	409722	0.17
21		G26A,G29A,G32A	3	F22, T28, F31	1004970	0.43
22		P18A	3	F22, T28, F31	1922485	0.82
23		P20A	3	F22, T28, F31	1924242	0.82
24		P18A, P20A	3	F22, T28, F31	1974363	0.84
25		ΔP18	3	P19, T27, F30	1636917	0.70
26		ΔP20	3	F21, T27, F30	203890	0.09
27		ΔP18, ΔP20	no product			
28	C1/N4		no product	--		
29	C2/K4		no product	--		
30	C3/G4		1	F31, 1 disulfide	1650293	0.70
31	A4/C5		no product	--		
32	C6/T7		no product	--		
33	S7/C8		1	F31, 1 disulfide	2724556	1.16
34	I7/C9		no product	--		
35	G7/C10		1	F31, 1 disulfide	2675077	1.14
36	C10/G13		3	F22, T28, F31	3934302	1.67
37	C11/A13		no product	--		
38	C12/A13		3	F22, T28, F31	4488382	1.91
39	L13/C14		3	F22, T28, F31	2653574	1.13
40	V13/C15		3	F22, T28, F31	7933484	3.37
41		F31TAG-Ome-Tyr	3	F22, T28, Ome-Tye-31	918557	0.39

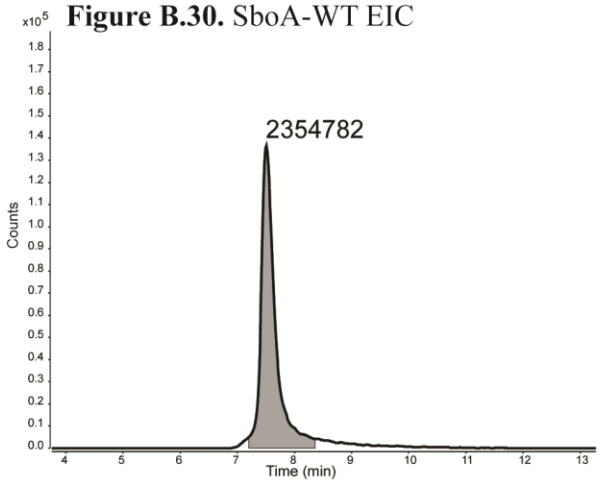


Figure B.30. SboA-WT EIC

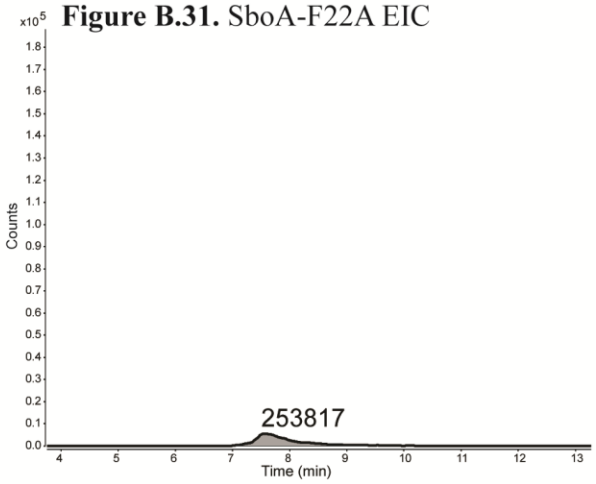


Figure B.31. SboA-F22A EIC

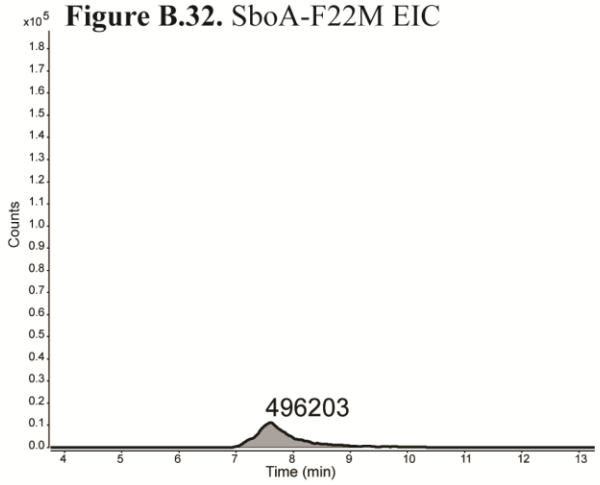


Figure B.32. SboA-F22M EIC

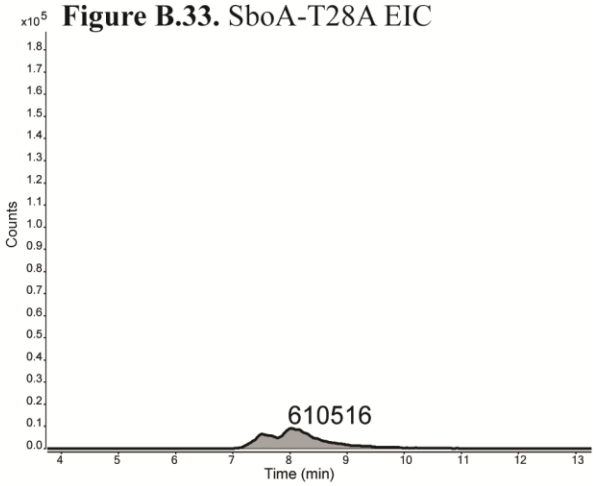


Figure B.33. SboA-T28A EIC

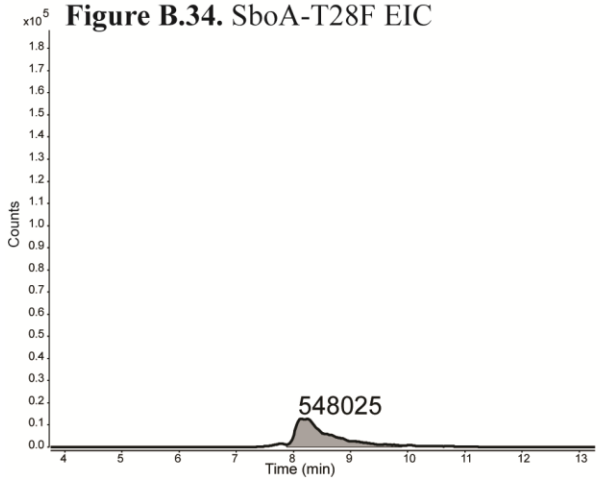


Figure B.34. SboA-T28F EIC

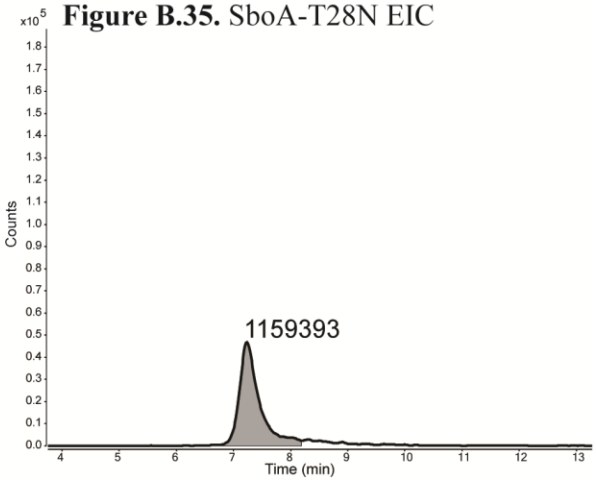


Figure B.35. SboA-T28N EIC

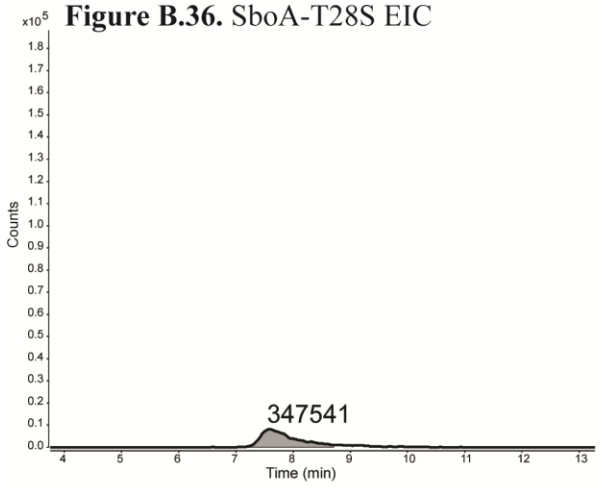


Figure B.36. SboA-T28S EIC

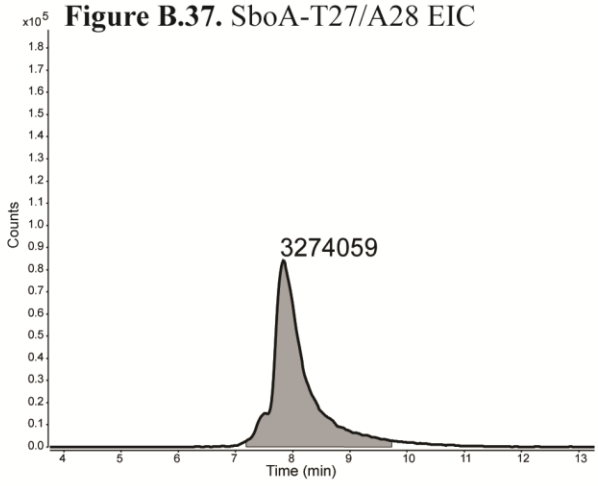


Figure B.37. SboA-T27/A28 EIC

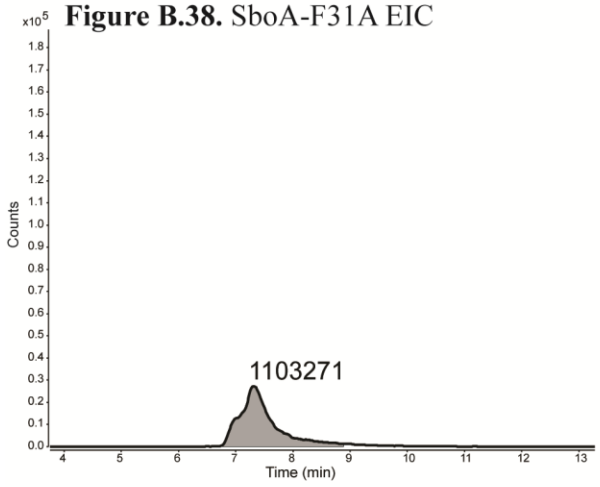


Figure B.38. SboA-F31A EIC

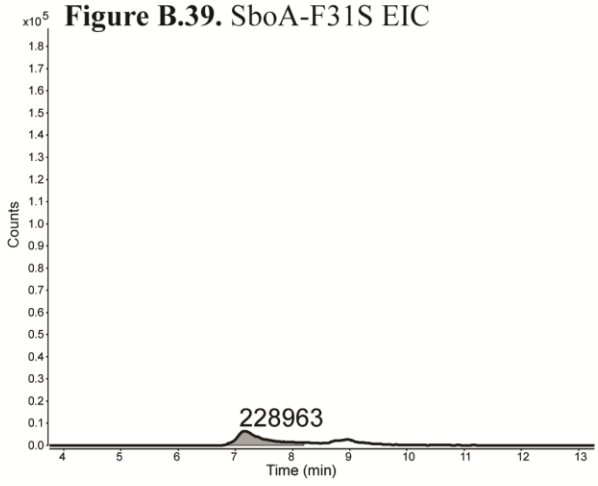


Figure B.39. SboA-F31S EIC

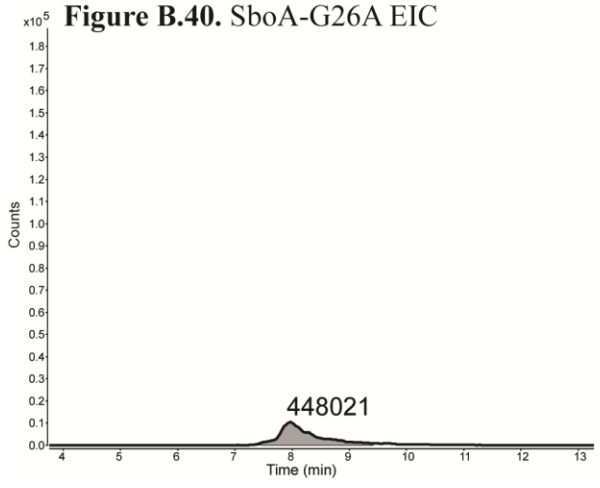


Figure B.40. SboA-G26A EIC

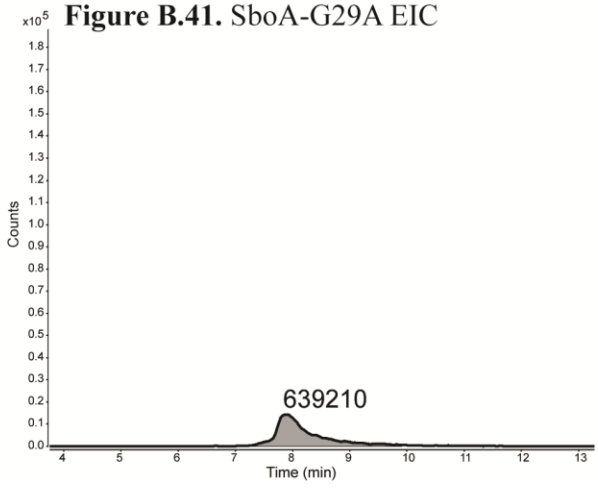


Figure B.41. SboA-G29A EIC

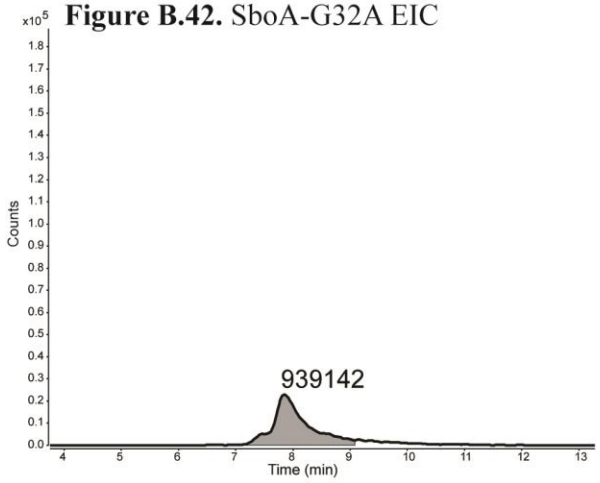


Figure B.42. SboA-G32A EIC

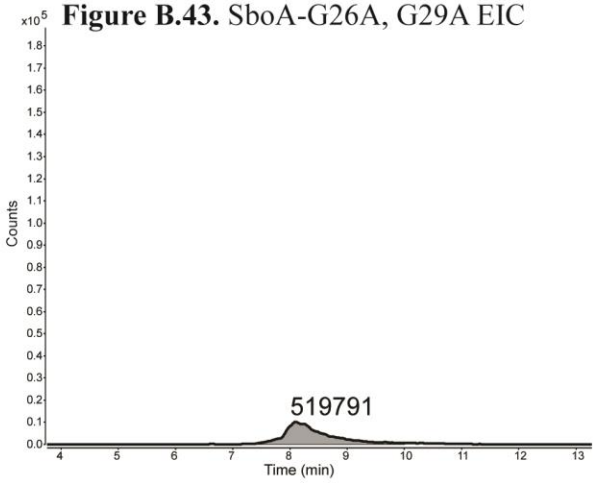


Figure B.43. SboA-G26A, G29A EIC

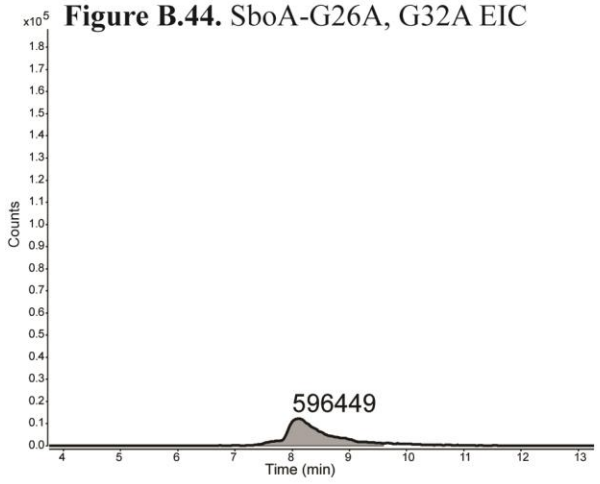


Figure B.44. SboA-G26A, G32A EIC

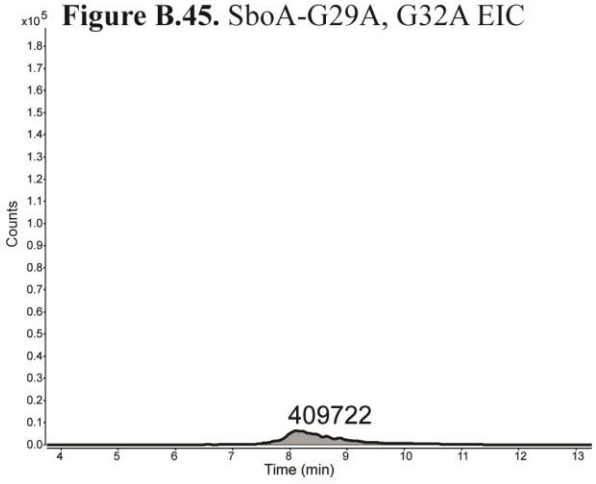


Figure B.45. SboA-G29A, G32A EIC

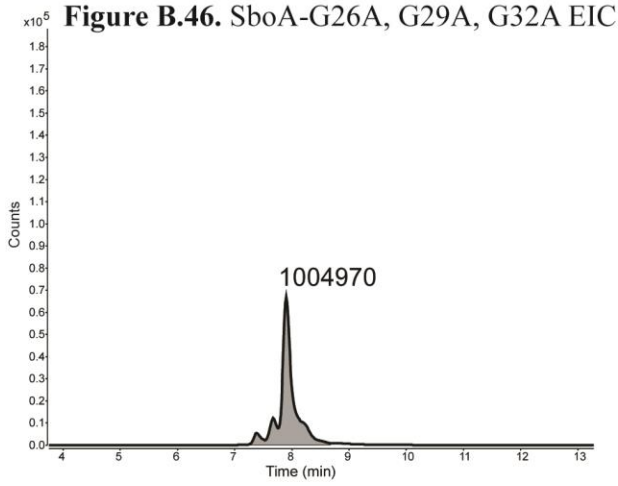


Figure B.46. SboA-G26A, G29A, G32A EIC

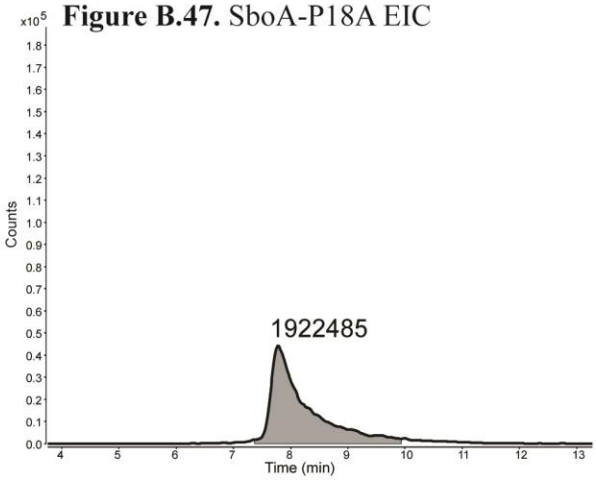


Figure B.47. SboA-P18A EIC

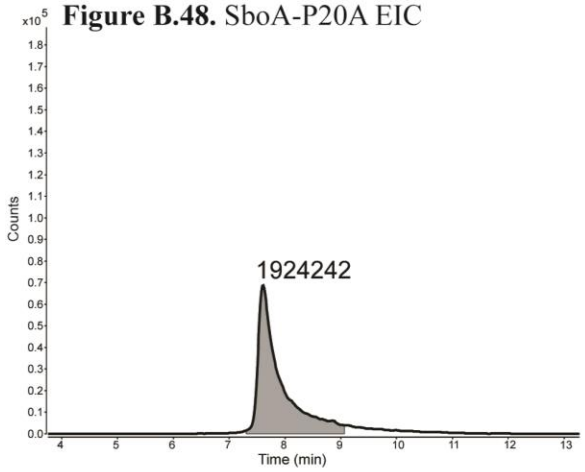


Figure B.48. SboA-P20A EIC

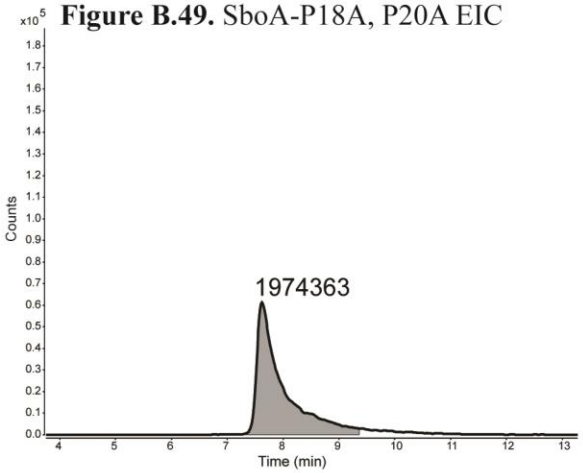


Figure B.49. SboA-P18A, P20A EIC

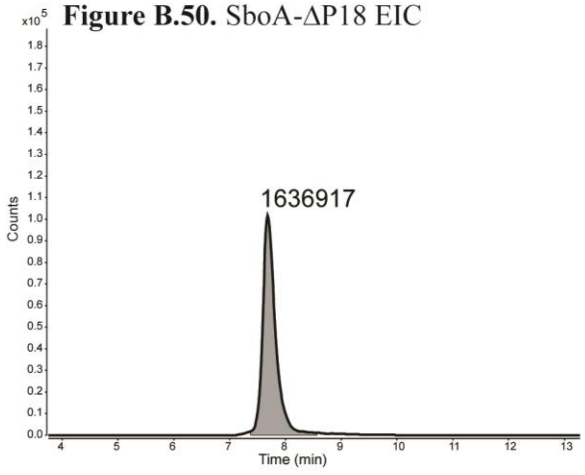


Figure B.50. SboA-ΔP18 EIC

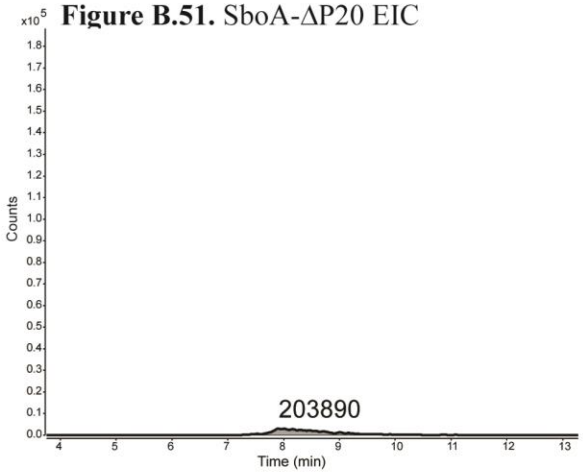


Figure B.51. SboA-ΔP20 EIC

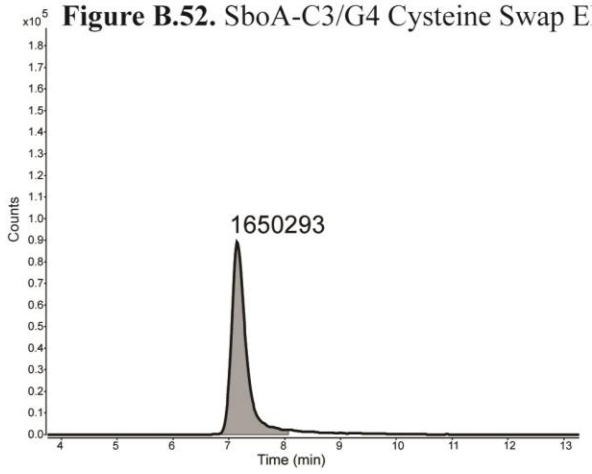


Figure B.52. SboA-C3/G4 Cysteine Swap EIC

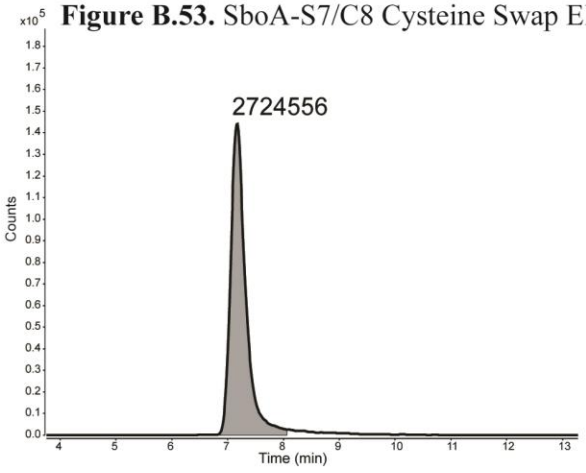


Figure B.53. SboA-S7/C8 Cysteine Swap EIC

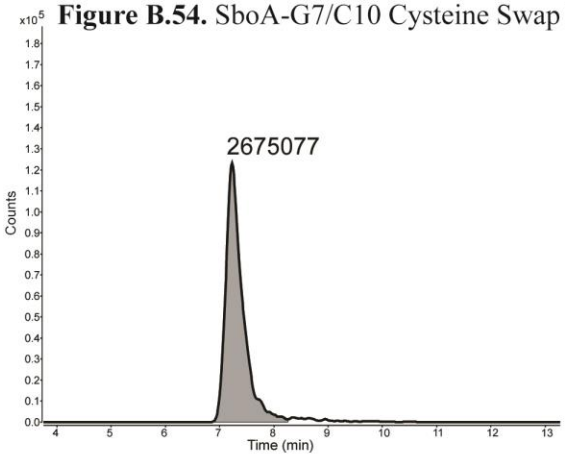


Figure B.54. SboA-G7/C10 Cysteine Swap EIC

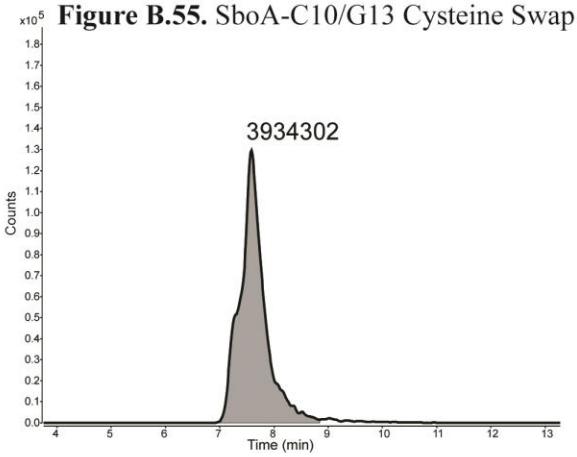


Figure B.55. SboA-C10/G13 Cysteine Swap EIC

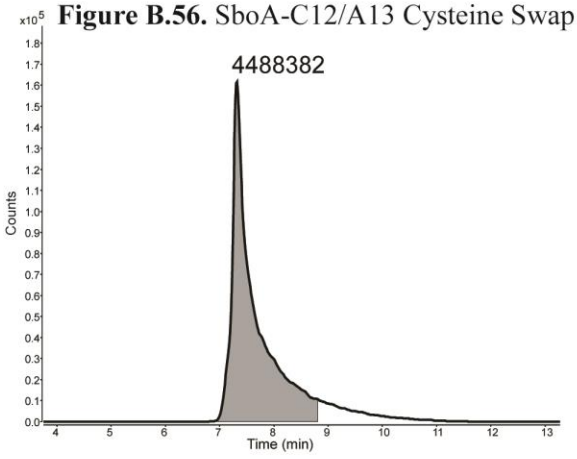


Figure B.56. SboA-C12/A13 Cysteine Swap EIC

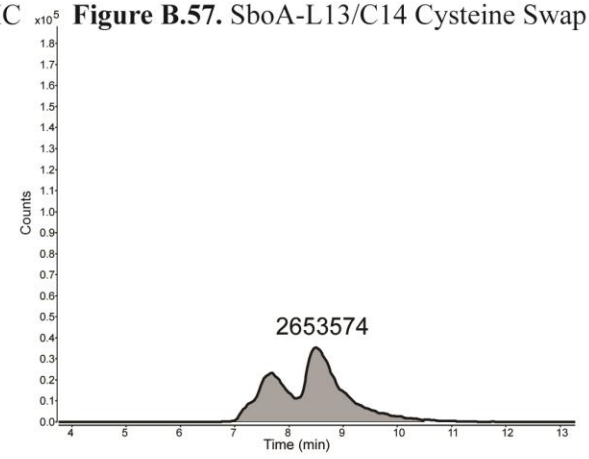


Figure B.57. SboA-L13/C14 Cysteine Swap EIC

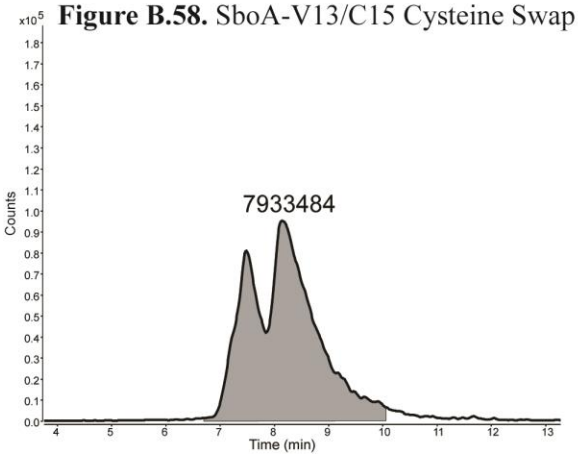


Figure B.58. SboA-V13/C15 Cysteine Swap EIC

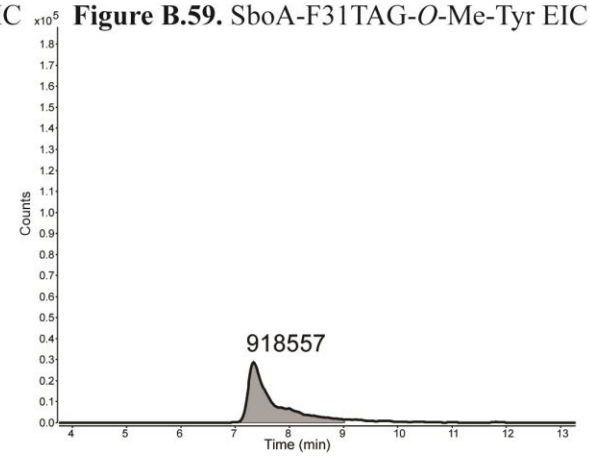


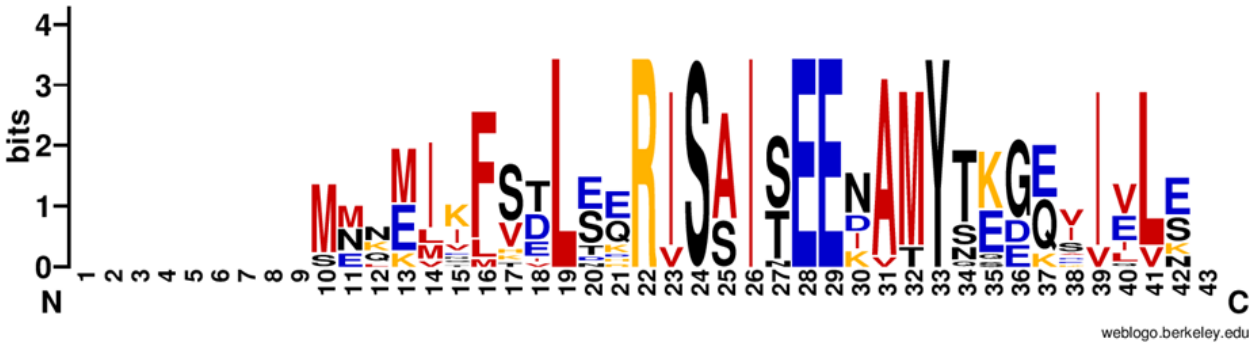
Figure B.59. SboA-F31TAG-*O*-Me-Tyr EIC

APPENDIX C. SUPPLEMENTARY FIGURES AND TABLES FOR CHAPTER 4

Figure C.1. Alignment of PaaP and its homologs. Alignment was generated using MUSCLE with default parameters. 100% similarity = red, 75% similarity = blue, 50% similarity = gray.

```
                                *         20         *         40
WP_0725972  : ----------MKKLVFSDLEKRISAISEEDATYSKGKIILV-- : 31
APG93320    : MARFRKKGTSMKKLVFSDLEKRISAISEEDATYSKGKIILV-- : 41
AOA71868    : ----------MLEITFVDLEERISAISEEDAMYQSGESIQLS- : 32
WP_0719921  : ------------MEFVILNERVSAINEENAMYNEDESIELE- : 29
WP_0715482  : ---------MENEIILHDLTERISSISEEIAMYNEDEEVILES : 34
WP_0720150  : ---------MENEIIFTDLEDRVSSISEEIAMYTEDQIIELE- : 33
WP_0744010  : ---------MNNEIIFSELDERISSISEEIAMYTEEEDIELE- : 33
WP_0715242  : ---------MNQEVIFVELEERISSISEEKAMYTEEEAIELE- : 33
WP_0718236  : ---------MNQEMSFVELEERISSISEEKAMYTEEESIELE- : 33
EAR53337    : ------------MIKMKVLTHRISAISEEKVMYSQGEKIVLK- : 30
WP_0751152  : ------------MINFSTLSRRISAITEENAMYTKGQIVILK- : 30
Paap        : ------------MIKFSTLSQRISAITEENAMYTKGQVIVLS- : 30
WP_0127703  : ------------MIKFSTLSQRISAITEENAMYTKGQVIVLS- : 30
ACT07482    : ------------MIKFSTLSQRISAITEENAMYTKGQVIVLS- : 30
WP_0719910  : ------------MIKLSTLSQRISAITEENAMYTKGQVIVLN- : 30
WP_0720051  : ------------MIKFSTLSQRISAITEENAMYTKGQVIVLN- : 30
```

Figure C.2. Logo Analysis of Alignment above. From the alignment, the sequences were imported to WebLogo software (internet version 3)[1,2] to create the sequence logos below.



REFERENCES

1. Crooks G.E.; Hon G.; Chandonia J.M.; Brenner S.E; *Genome Res.* **2004**, *1*4, 1188

2. Schneider TD; Stephens RM; *Nucleic Acids Res.* **1990**, *18*, 6097