

**BIOSYNTHESIS OF FUNGAL POLYKETIDE NATURAL
PRODUCTS: STRUCTURAL AND BIOCHEMICAL STUDIES
TOWARD UNDERSTANDING PROGRAMMING AND
SUBSTRATE SPECIFICITY**

by
Callie R. Huitt-Roehl

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy in Chemical Biology

Baltimore, Maryland
October 2017

© 2017 Callie R. Huitt-Roehl
All Rights Reserved

Abstract

Polyketides are a large and diverse family of natural products encompassing some of the most effective and valuable pharmaceuticals of all time, as well as some of the deadliest toxins. These fascinating molecules are biosynthesized by some of nature's largest and most complex enzymes, the polyketide synthases. Because of their diverse biological activities, polyketide-based natural products are an attractive target for biosynthetic engineering. In order to effectively engineer polyketide synthases to produce novel compounds, we must understand the native enzymology of these molecular machines. Since they consist of only a single set of catalytic domains, the fungal, iterative polyketide synthases are particularly challenging to decipher and engineer, relative to the assembly-line-like modular polyketide synthases. In this work, biochemical and structural studies are performed to improve understanding of iterative polyketide synthase enzymatic programming. By performing a screen of thirty unnatural starter units, the substrate specificity of the iterative polyketide synthase PksA was evaluated. Characterization of enzymatic derailment products allowed for interrogation of the tolerance of specific catalytic domains to the unnatural substrates. Through mechanism-based crosslinking of a closely related iterative polyketide synthase, CTB1, the structure of a polyketide synthase locked in a loading conformation was observed by cryo-

electron microscopy. This first-ever multidomain structure of a non-reducing polyketide synthase may help to guide future engineering efforts, by beginning to establish what conformational changes polyketide synthases undergo during biosynthesis and how programming is achieved across multiple catalytic domains.

Thesis Advisor:

Prof. Craig A. Townsend, Department of Chemistry, Johns Hopkins University

Additional Readers:

Prof. Caren L. Freel Meyers, Department of Pharmacology and Molecular
Sciences, Johns Hopkins University

Prof. Scott Bailey, Department of Biochemistry and Molecular Biology, Johns
Hopkins University

Dedicated to my parents, Jan Roehl and Robert Huitt,
who instilled in me a life-long love of learning.

Acknowledgements

First and foremost, I must thank my mentor, Prof. Craig A. Townsend. Prof. Townsend's love for science and passion for ensuring his students succeed is without compare. I am incredibly grateful to have studied under him, and have grown immeasurably as a scientist thanks to his tutelage.

I have been incredibly fortunate to work with fantastic collaborators during my time in the lab, to whom I owe a great deal of my success. Dr. Dominik Herbst and Dr. Timm Maier at the University of Basel are incredible structural biologists, and have drastically transformed the direction in which we are taking the study of PKSs. Dr. Melvin Bolton at the United States Department of Agriculture has greatly expanded our knowledge of cercosporin biosynthesis, and been an absolute pleasure to collaborate with.

Next, none of the work in this dissertation would be possible without the help of my fellow lab-mates, particularly the venerable Team PKS. Dr. Anna Vagstad and my rotation mentor Dr. Adam Newman taught me a great deal about both PKSs and how to be a great scientist. Anna's incredible work on PksA paved the way for many of my own experiments, and her meticulous record-keeping taught me proper lab organization and notebook-maintenance. Adam is one of the smartest graduate students I have met in my time at Johns

Hopkins, and taught me a great deal about cercosporin, as well as giving effective presentations. Mr. Douglas Cohen, despite working on the dark (highly-reducing) side of Team PKS, has been an incredible desk-mate and friend, always there to bounce ideas off or just commiserate with failed experiments. Mr. Jacob Kravetz, in addition to being the life of the party, is the synthetic backbone of Team PKS, providing crosslinkers for structural studies. The newest members of the team, Mrs. Jamie Alley and Ms. Paramita Pal, have been wonderful additions to the team, and I look forward to following their work in the future.

Dr. Phil Storm, another member of Team PKS, has been the world's best partner, both in science and in life. He has been instrumental in my success in graduate school, from my early days as a rotation student in the Townsend lab to these final weeks of dissertation writing. Without his support I would not have survived graduate school. I am so proud of the work he has accomplished in his time in the lab, taking the PKS project in exciting new directions, and look forward to what the next adventure brings us.

The Townsend lab is filled with incredibly talented scientists and wonderful friends. Dr. Darcie Long was the first friend I made in graduate school, helping me survive those first few terrifying weeks as a fellow California transplant. Nothing could be accomplished in the lab without Dr. Rongfeng Li,

whose biological expertise is irreplaceable. Mr. Ryan Oliver has been another great friend in the lab these past six years, and an incredibly knowledgeable resource, particularly in organic chemistry. Mr. Felipe d'Andrea is the most talented undergraduate I have ever met, and I cannot wait to see where his MD/PhD venture takes him. To the newest lab members - Mr. Michael Wheadon, Ms. Erica Sinner, Mr. Hunter Batchelder, Mr. Trevor Zandi, and Mr. Michael Lichstrahl, I thank you for continuing the tradition of excellent science in the Townsend lab, and look forward to following your future successes.

I also have to thank a number of previous Townsend lab members for taking me under their wing and making me feel like a part of the family when I first joined – Dr. Jesse Li, Dr. Victor Outlaw, Dr. Courtney Hastings, Dr. Andy Buller, Dr. Daniel Marous, Dr. Evan Lloyd, and Dr. Nicole Guadelli. For both your scientific expertise and your friendship I am eternally grateful. I would particularly like to thank Dr. Eric Hill, who has not only taught me a great deal about science and surviving graduate school, but has also been an incredible confidante and one of my closest friends.

Being a part of the Chemistry-Biology Interface (CBI) program at Johns Hopkins has been a wonderful experience. I have greatly appreciated the community our small department has fostered, and am very grateful for our director Prof. Steve Rokita, and his predecessor Prof. Marc Greenberg. I'd also

like to thank all of the support staff that keeps the department and facilities running smoothly, particularly Lauren McGhee, Boris Steinberg, and Dr. Phil Mortimer.

To my thesis committee, Prof. Caren Meyers, Prof. Phil Cole, and Prof. Scott Bailey, I am thankful for your patience in scheduling my yearly meetings, which was always difficult, and more importantly all of your helpful comments and guidance.

Finally, I would like to thank my family. Mike Collins and Kathe Fox took me under their wing when I moved to Baltimore, and have been wonderful “East coast parents” over the past six years. I am so thankful for all of the wonderful dinners at their house, and generous support. My brothers, Mike and Eric Huitt, and their mother Karen Loepp, have been a great source of support and love, and it has been wonderful to be able to spend more time with them. Most importantly, I have to thank my parents, Dr. Jan Roehl and Dr. Robert Huitt. Without their influence I would have never pursued a PhD. I could not ask for more supportive parents, and am eternally grateful for everything they have done for me. It has been difficult living a country away for the past six years, but they have helped ease the homesickness with regular trips home to paradise. Thank you for everything Mom and Dad, I love you.

Publications

Huitt-Roehl, C.R., Hill, E.A., Adams, M.M., Vagstad, A.L., Li, J.W., Townsend, C.A. Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA. *ACS Chem. Biol.*, **2015**, *10*, 1443–9

Herbst, D.A., **Huitt-Roehl, C.R.**, Jakob, R.P., Kravetz, J.M., Storm, P.A., Townsend, C.A., Maier, T. The structural organization of substrate loading in iterative polyketide synthases. *Nat. Chem. Biol.*, *Submitted*.

de Jonge, R.*, Ebert, M.K.*, **Huitt-Roehl, C.R.***, Pal, P., Suttle, J.C., Neubauer, J.D., Jurick II, W.M., Secor, G.A., Thomma, B.P.H.J., Van de Peer, Y., Townsend, C.A., Bolton, M.D. Ancient duplication and horizontal transfer of a toxin gene cluster reveals novel mechanisms in the cercosporin biosynthesis pathway. *Proc. Natl. Acad. Sci.*, *Submitted*. *Indicates co-first-authorship

Table of Contents

Front Matter	i
Title Page.....	i
Abstract	ii
Dedication.....	iv
Acknowledgements.....	v
Publications	ix
Table of Contents.....	x
List of Tables and Figures.....	xiv
Chapter 1: Introduction to polyketide natural products.....	1
1.1. Introduction.....	1
1.2. Polyketide synthases.....	7
1.2.1. The polyketide synthase catalytic cycle.....	7
1.2.2. Classification of PKSs – Early lessons from modular PKS.....	10
1.3. Tools for understanding polyketide biosynthesis	12
1.3.1. Domain dissection for <i>in vitro</i> analysis	12
1.3.2. Structures of NR-PKS monodomains.....	17
1.3.3. Full-length PKS structures and the rise of cryo-electron microscopy	22
1.4. Outlook	26
1.5. References	27
Chapter 2: Starter unit flexibility for engineered product synthesis by the non-reducing polyketide synthase PksA	34
2.1. Introduction.....	34
2.2. Results	38
2.2.1. Alternative starter unit <i>in vitro</i> assays	38
2.2.2. Linear acyl-SNAC starter units.....	38
2.2.3. Halogenated acyl-SNAC starter units.....	42
2.2.4. Methylated acyl-SNAC starter units	44

2.2.5. Unsaturated acyl-SNAC starter units	45
2.2.6. Heteroatom-substituted acyl-SNAC starter units	47
2.2.7. Sterically constrained acyl-SNAC starter units	50
2.2.8. SAT active site mutants	51
2.3. Discussion	53
2.4. Experimental Methods	57
2.4.1. Cloning	57
2.4.2. Protein expression and purification	57
2.4.3. <i>In vitro</i> assays	58
2.4.4. HPLC and UPLC-ESI-MS analysis of enzymatic products	58
2.4.5. Synthesis of acyl-SNAC substrates	59
2.4.6. Ellman's Reagent assay for acyl-SNAC hydrolysis	59
2.5. References	60

Chapter 3: The structural organization of substrate loading in iterative polyketide synthases.....63

3.1. Introduction	63
3.2. Results	69
3.2.1. Mechanism-based crosslinking of CTB1	69
3.2.2. Structure of CTB1 SAT-KS-MAT	78
3.2.3. Structure of CTB1 SAT ^o -KS-MAT ^o =ACP2	81
3.2.4. Mutagenic validation of KS-ACP binding region	84
3.3. Discussion	86
3.4. Experimental Methods	89
3.4.1. Preparation of expression constructs	89
3.4.2. Protein expression and purification for crosslinking and interface validation	90
3.4.3. Expression and purification of CTB1 SAT-KS-MAT for crystallization	91
3.4.4. Proteolysis of tag from CTB1 ACP2-His	93
3.4.5. Synthesis of α -bromopropionyl aminopantetheine	94
3.4.6. Phosphorylation of α -bromopropionyl aminopantetheine	100
3.4.7. Loading of ACP2 with α -bromoacyl crosslinker	101
3.4.8. Crosslinking of CTB1 SAT ^o -KS-MAT ^o to ACP2	102
3.4.9. Crosslinking of CTB1 SAT-KS ^o -MAT ^o to ACP2	103
3.4.10. <i>In vitro</i> reactions of CTB1 for interface validation and comparison of ACP1 and ACP2	103
3.4.11. Crystallization, data collection and structure determination of CTB1 SAT-KS-MAT	105

3.4.12. EM sample preparation and data collection	106
3.4.13. EM data processing and analysis.....	107
3.4.14. Cryo-EM structure refinement and modeling	110
3.4.15. Structure analysis and visualization	112
3.4.16. Sequence analysis.....	112
3.5. References	113

Chapter 4: Cercosporin biosynthesis: Expansion of the gene cluster and scope of production by plant pathogens.....120

4.1. Introduction.....	120
4.2. Results	126
4.2.1. Identification of new cercosporin biosynthetic genes	126
4.2.2. Characterization of pre-cercosporin.....	128
4.2.3. Identification of cercosporin production by <i>Colletotrichum fioriniae</i>	132
4.3. Discussion.....	133
4.4. Experimental Methods.....	138
4.4.1. Pre-cercosporin isolation and characterization	138
4.4.2. Pre-cercosporin purification by Sephadex LH20	140
4.4.3. <i>Colletotrichum</i> spp. cercosporin assay.....	141
4.5. References	142

Appendix A: General experimental methods146

A.1. Cloning.....	146
A.1.1. Polymerase chain reaction (PCR)	147
A.1.2. Plasmid assembly by restriction digestion and ligation.....	148
A.1.3. Plasmid assembly by Gibson assembly	148
A.1.4. Transformation of <i>E. coli</i> by electroporation.....	149
A.2. Heterologous protein expression	150
A.3. Protein purification	150
A.4. Standard medium recipes	151
A.4.1. Luria-Bertani Broth (LB)	151
A.4.2. Terrific Broth (TB)	151

Appendix B: Supplementary Material to Chapter 2153

B.1. Cloning and mutagenesis	153
B.2. Synthesis of acyl-SNAC substrates.....	155
B.3. Supplementary figures.....	174

B.4. References.....214

Appendix C: Supplementary Material to Chapter 3.....215
C.1. Supplementary figures.....215
C.2. Supplementary tables.....223

Appendix D: Supplementary Material to Chapter 4228
D.1. CTB gene sequences228
D.2. CTB protein sequences235

Curriculum Vitae238

List of Figures and Tables

Figure 1.1. Polyketide-based natural products.....	2
Figure 1.2. Collie's polyketide hypothesis	3
Figure 1.3. Site-specific radiolabeling of 6-MSA	4
Figure 1.4. Biosynthesis of actinorhodin	6
Figure 1.5. General polyketide synthase catalytic cycle.....	8
Figure 1.6. PksA biosynthetic scheme determined by domain deconstruction ...	15
Figure 1.7. Structure of PksA PT.....	18
Figure 1.8. Structure of PksA TE	19
Figure 1.9. Structure of PksA ACP	20
Figure 1.10. Structure of CazM SAT.....	22
Figure 1.11. Structure of porcine FAS	23
Figure 1.12. Structure of DEBS module 5 KS-AT	24
Figure 2.1. PksA biosynthetic scheme and experimental design.....	37
Figure 2.2. Relative production of naphthopyrones from linear acyl-SNAC starter units	40
Figure 2.3. Core structures of observed derailment products.....	42
Figure 2.4. Relative production of naphthopyrones from halogenated acyl-SNAC starter units	43
Figure 2.5. Relative production of naphthopyrones from methylated acyl-SNAC starter units	45
Figure 2.6. Relative production of naphthopyrones from unsaturated acyl-SNAC starter units	47
Figure 2.7. Relative production of naphthopyrones from heteroatom-substitutes acyl-SNAC starter units	49
Figure 2.8. Relative production of naphthopyrones from sterically constrained acyl-SNAC starter units	51
Figure 2.9. Comparison of naphthopyrone production by PksA SAT active site mutants	52
Figure 2.10. Ellman's reagent assay for acyl-SNAC hydrolysis by PksA SAT	52
Figure 2.11. Comparison of intact pentadomain and dissected two-part reactions.....	55
Figure 3.1. CTB1 biosynthetic scheme	68
Figure 3.2. Comparison of biosynthetic capabilities of CTB1 ACP1 and ACP2 ...	70
Figure 3.3. Loading of CTB1 ACP2	72
Figure 3.4. Enzymatic preparation of crosslinker-loaded ACP	74
Figure 3.5. Crosslinking time course.....	76

Figure 3.6. [1- ¹⁴ C]-acetyl-CoA competition assay.....	77
Figure 3.7. Crystal structure of CTB1 SAT-KS-MAT	80
Figure 3.8. Asymmetric cryo-EM structure of CTB1 SAT ^o -KS-MAT ^o =ACP2	83
Figure 3.9. Mutagenesis of ACP2	86
Figure 3.10. Schematic illustration of suggested modes of conformational coupling in CTB1.....	88
Figure 3.11. Synthesis of α -bromopropionyl crosslinker	94
Figure 4.1. Reduced and oxidized cercosporin	122
Figure 4.2. Early proposed biosynthesis of cercosporin.....	124
Figure 4.3. Revised proposal for cercosporin biosynthesis	125
Figure 4.4. Revised cercosporin biosynthetic gene cluster	127
Figure 4.5. HPLC analysis of <i>C. beticola</i> mutants.....	127
Figure 4.6. UV-Vis spectra of pre-cercosporin and cercosporin	129
Figure 4.7. Large-scale HPLC purification of pre-cercosporin	131
Figure 4.8. Analysis of <i>Colletotrichum fioriniae</i> extracts.....	133
Figure 4.9. Proposed role of CTB9 and CTB10 in cercosporin biosynthesis	135
Figure 4.10. Role of HypE in aflatoxin biosynthesis.....	135
Appendix Table B.1. Plasmids used in Chapter 2.....	155
Appendix Table B.2. Primers used in Chapter 2.....	155
Appendix Figure B.1. Structures of all starter units assayed in Chapter 2 and the resulting enzymatic products	174
Appendix Figure B.2. Product profile for starter unit 2-5	178
Appendix Figure B.3. Product profile for starter unit 2-6	178
Appendix Figure B.4. Product profile for starter unit 2-7	179
Appendix Figure B.5. Product profile for starter unit 2-8	179
Appendix Figure B.6. Product profile for starter unit 2-9	180
Appendix Figure B.7. Product profile for starter unit 2-10	180
Appendix Figure B.8. Product profile for starter unit 2-11	181
Appendix Figure B.9. Product profile for starter unit 2-12	181
Appendix Figure B.10. Product profile for starter unit 2-13	182
Appendix Figure B.11. Product profile for starter unit 2-14	182
Appendix Figure B.12. Product profile for starter unit 2-15	183
Appendix Figure B.13. Product profile for starter unit 2-16	183
Appendix Figure B.14. Product profile for starter unit 2-17	184
Appendix Figure B.15. Product profile for starter unit 2-18	184
Appendix Figure B.16. Product profile for starter unit 2-19	185
Appendix Figure B.17. Product profile for starter unit 2-20	185
Appendix Figure B.18. Product profile for starter unit 2-21	186
Appendix Figure B.19. Product profile for starter unit 2-22	186
Appendix Figure B.20. Product profile for starter unit 2-23	187

Appendix Figure B.21.	Product profile for starter unit 2-24	187
Appendix Figure B.22.	Product profile for starter unit 2-25	188
Appendix Figure B.23.	Product profile for starter unit 2-26	188
Appendix Figure B.24.	Product profile for starter unit 2-27	189
Appendix Figure B.25.	Product profile for starter unit 2-28	189
Appendix Figure B.26.	Product profile for starter unit 2-29	190
Appendix Figure B.27.	Product profile for starter unit 2-30	190
Appendix Figure B.28.	Product profile for starter unit 2-31	191
Appendix Figure B.29.	Product profile for starter unit 2-32	191
Appendix Figure B.30.	Product profile for starter unit 2-33	192
Appendix Figure B.31.	Product profile for starter unit 2-34	192
Appendix Figure B.32.	Product profile for control reactions containing only malonyl-SNAC	192
Appendix Figure B.33.	SDS-PAGE of purified proteins used in Chapter 2	193
Appendix Figure B.34.	NMR spectra of starter unit 2-11	194
Appendix Figure B.35.	NMR spectra of starter unit 2-12	195
Appendix Figure B.36.	NMR spectra of starter unit 2-13	196
Appendix Figure B.37.	NMR spectra of starter unit 2-14	197
Appendix Figure B.38.	NMR spectra of starter unit 2-15	198
Appendix Figure B.39.	NMR spectra of starter unit 2-16	199
Appendix Figure B.40.	NMR spectra of starter unit 2-17	200
Appendix Figure B.41.	NMR spectra of starter unit 2-18	201
Appendix Figure B.42.	NMR spectra of starter unit 2-19	202
Appendix Figure B.43.	NMR spectra of starter unit 2-22	203
Appendix Figure B.44.	NMR spectra of starter unit 2-23	204
Appendix Figure B.45.	NMR spectra of starter unit 2-24	205
Appendix Figure B.46.	NMR spectra of starter unit 2-25	206
Appendix Figure B.47.	NMR spectra of starter unit 2-26	207
Appendix Figure B.48.	NMR spectra of starter unit 2-27	208
Appendix Figure B.49.	NMR spectra of starter unit 2-28	209
Appendix Figure B.50.	NMR spectra of starter unit 2-29	210
Appendix Figure B.51.	NMR spectra of starter unit 2-30	211
Appendix Figure B.52.	NMR spectra of starter unit 2-32	212
Appendix Figure B.53.	NMR spectra of starter unit 2-33	213
Appendix Figure C.1.	Integration of loading domains in PKS.....	214
Appendix Figure C.2.	Cryo-EM data processing scheme	216
Appendix Figure C.3.	Crosslinking and cryo-EM sample preparation	217
Appendix Figure C.4.	Conformational variability in the cryo-EM structure of CTB1 SAT ^o -KS-MAT ^o =ACP2.....	218
Appendix Figure C.5.	Structural comparison between CTB1 SAT-KS-MAT and	

CTB1 SAT ^o -KS-MAT ^o =ACP2.....	219
Appendix Figure C.6. CTB1 SAT-KS-MAT alignment of regions interfacing ACP2 and phylogeny	220
Appendix Figure C.7. Comparison of KS-ACP interactions in PKS and FAS....	221
Appendix Table C.1. X-ray crystallographic and cryo-EM data collection and refinement statistics	223
Appendix Table C.2. Structural comparison and interface analysis of CTB1 SAT-KS-MAT	224
Appendix Table C.3. Plasmids used in Chapter 3.....	225
Appendix Table C.4. Primers used in Chapter 3.....	226

Chapter 1: Introduction to polyketide natural products

1.1. Introduction

The polyketide class of natural products encompasses tens of thousands of diverse compounds with equally diverse biological activities (Figure 1.1). Many polyketide-based natural products have been transformed into life-saving drugs, while others are deadly toxins. One of the earliest and most well-studied is the antibiotic erythromycin **1**, which is FDA approved in its native form as well as several synthetic derivatives (e.g. clarithromycin, azithromycin), and is on the World Health Organization's List of Essential Medicines¹⁻². Some of the most successful drugs of the 21st century, including the best-selling pharmaceutical of all time, atorvastatin (Lipitor), are the cholesterol-lowering, polyketide-based statins³. The first statin to reach market was lovastatin **2**, isolated from the fungus *Aspergillus terreus*⁴⁻⁵. A fascinating example of the diversity of biological activities of polyketide-derived pharmaceuticals is the immunosuppressant rapamycin **3**, which has also been shown to have anticancer and possibly lifespan-extension effects.

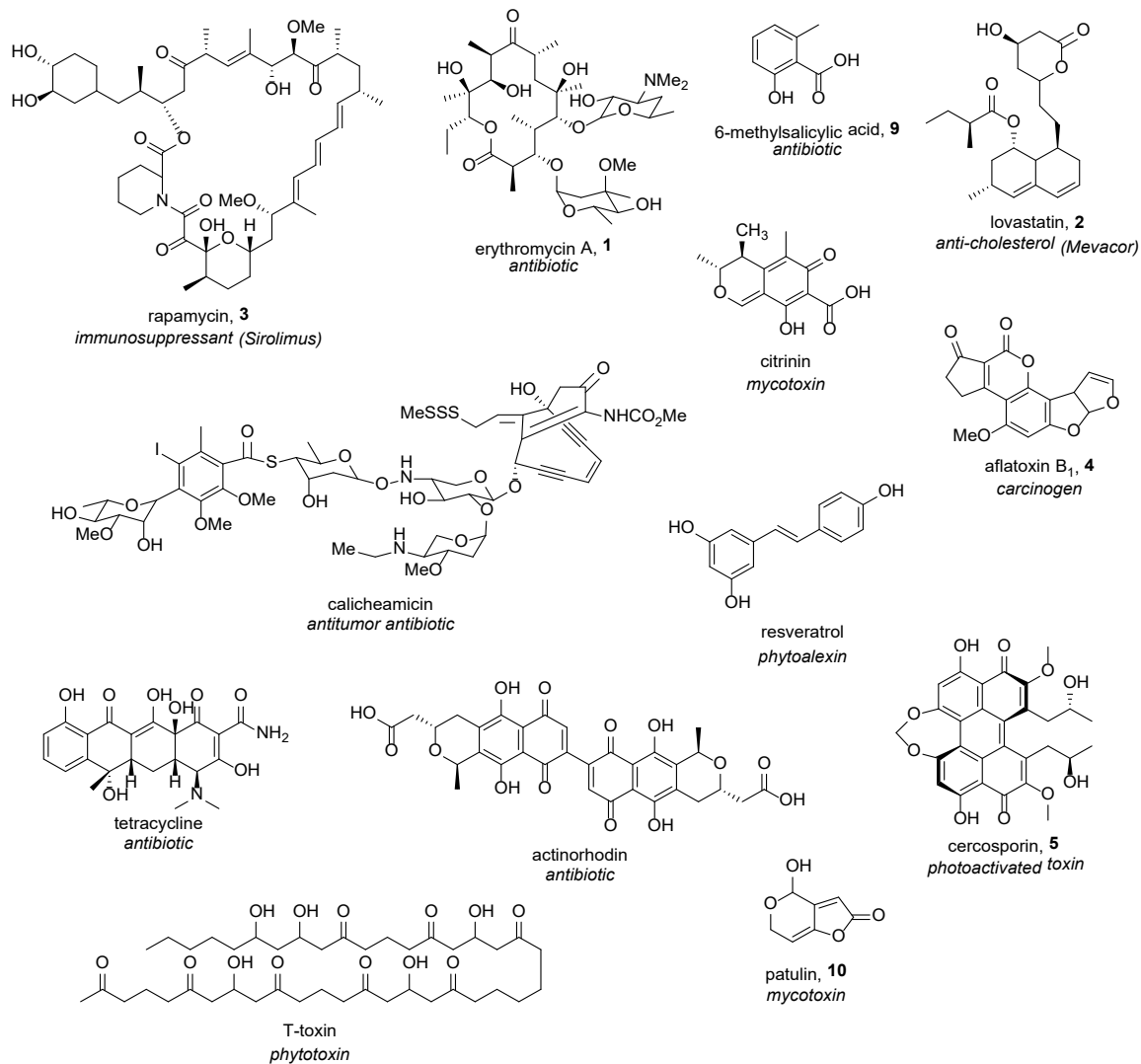


Figure 1.1. Examples of polyketide-based natural products.

While many polyketide-based natural products have beneficial effects on human health, others are potent toxins of significant concern to the agricultural industry. The hepatotoxin aflatoxin B₁ **4** is produced by *Aspergillus parasiticus*, a fungal pathogen of staple crops such as corn, wheat, and rice. *A. parasiticus* frequently colonizes improperly stored crops, and aflatoxin B₁ can enter the food

supply directly through contaminated crops or through accumulation in livestock from contaminated feed⁶. Since aflatoxin B₁ is so potent, for example, causing carcinomas in all rats fed with 15 ppb, it is of significant concern to human health⁷⁻⁸. Another potent polyketide-based toxin, cercosporin **5**, is produced by the plant pathogens *Cercospora* spp.⁹. *Cercospora* also infects a wide variety of agriculturally important crops, including sugar beets, corn, and rice, and causes millions of dollars in losses each year¹⁰.

Since the first characterization of a polyketide-like molecule by James Collie in 1893, the biosynthesis of this remarkable class of molecules has been of particular interest¹¹. After isolating orsinol **6** from a degradation reaction of dehydroacetic acid **7**, Collie proposed it could be formed by way of a triketone intermediate **8** (Figure 1.2). He further suggested that these polyketone intermediates may be involved in production of these molecules by living cells. Unfortunately, it was not until the 1950s that the true biosynthetic basis of polyketide synthesis was established, when Arthur Birch began researching 6-methyl-salicylic acid (6-MSA) **9**, the precursor to patulin **10**, a fungal polyketide produced by *Penicillium patulum*.

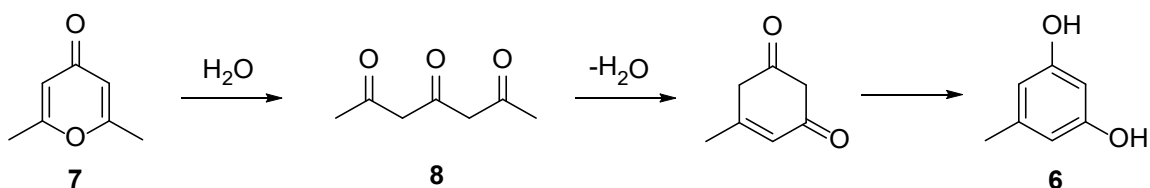


Figure 1.2. Collie's proposed formation of orsinol **6** through the triketide intermediate **8**.

Birch hypothesized that 6-MSA is synthesized by the 'head-to-tail' assembly of four acetate units. Subsequent cyclization and aromatization steps could then yield the final product. To prove this hypothesis, he fed radiolabeled acetate to the producing organism, and analyzed the resulting natural product for a specific labelling pattern¹². If his hypothesis was correct, the isolated 6-MSA would be labelled uniformly only at specific positions due to incorporation of the [1-¹⁴C]-acetate (Figure 1.3). Through specific degradation of the labelled molecule, followed by measurement of radioactivity of each of the fragments, Birch was able to demonstrate isotopic enrichment at the predicted sites, validating his hypothesis.

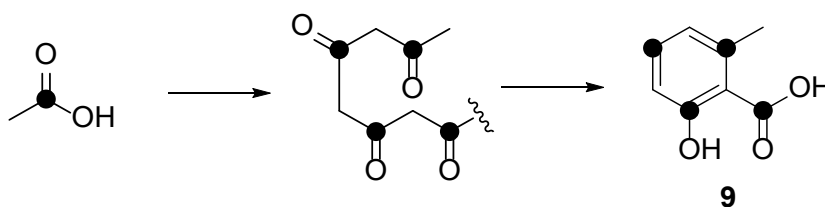


Figure 1.3. Birch's demonstration of the polyketide basis of 6-MSA by incorporation of [1-¹⁴C]-acetate (indicated by dots).

The success of Birch's labelling experiments triggered a burst of similar studies to determine the biosynthetic origin of a number of natural products. A further boon to the natural products field came in the 1970s, when Fourier-transform nuclear magnetic resonance (NMR) spectroscopy became commonly available and was applied to polyketide biosynthesis. This technique became especially useful when stable isotopic labels, such as ¹³C and ²H, became

commercially available. Previously, labelling patterns were determined by months of labor-intensive degradation studies. However, incorporation of these stable isotopes into polyketides by their producers allowed for determination of structures and labelling patterns in a single NMR experiment; therefore, much more complex molecules could be studied.

Compared to early, simple polyketides such as orsinol and 6-MSA, the known scope of polyketide-based natural products is incredibly broad. While the core reaction of polyketide biosynthesis - condensation of acetate units - yields a simple poly- β -ketone chain, a huge variety of tailoring reactions produce seemingly endless variations on the theme. While some of these modifications are installed by domains within the polyketide synthase (PKS) itself (e.g. reduction, dehydration, methylation), many more are catalyzed by distinct tailoring enzymes. Generally, the PKS and its associated tailoring genes are organized within a biosynthetic gene cluster¹³. Biosynthetic gene clusters contain all of the genes required for biosynthesis of a given natural product, as well as genes involved in regulation of pathway expression or resistance. With the advent of the genomic era of natural product biosynthesis, their fortunate organization into gene clusters has greatly increased the rate at which new natural product pathways have been identified.

Genetic techniques developed in the 1980s allowed for the characterization of the first PKS, a type II system responsible for the biosynthesis of the antibiotic actinorhodin **11**. The blue color of **11** provided an obvious screening method to identify genes involved in biosynthesis: mutants that abolish production lost their characteristic color and antibiotic activity¹⁴. By sequencing the region of the genome identified by these mutational studies, the biosynthetic enzymes were localized and characterized. Comparison of these enzymes to fatty acid synthase (FAS) sequences showed that they included FAS-like domains: ketosynthase (KS), acyl-transferase (AT), acyl-carrier protein (ACP), and ketoreductase (KR). Further genetic advances allowed specific removal of enzymes within the pathway and complementation of those genes through an added plasmid. These techniques led to interrogation of the pathway in full by analyzing intermediates accumulated *in vivo* by various blocked mutants (Figure 1.4)¹⁵⁻¹⁶.

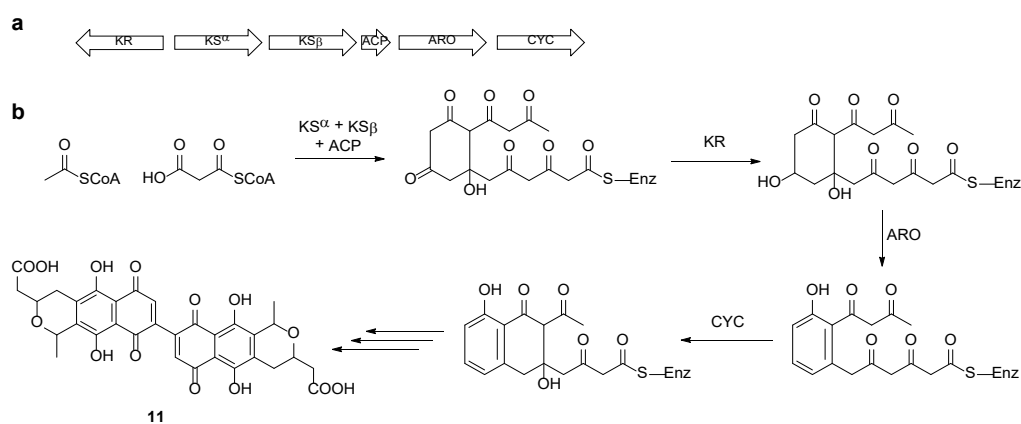


Figure 1.4. Biosynthesis of actinorhodin **11** by a type II PKS. (a) Organization of *act* genes involved in biosynthesis of **11** in *Streptomyces coelicolor*. (b) Role of each *act* gene in **11** biosynthesis, established by gene knockouts.

1.2. Polyketide Synthases

1.2.1. The Polyketide Synthase Catalytic Cycle

Polyketides are synthesized by some of the world's largest and most complex enzymes, the polyketide synthases (PKSs). For example, MlsA1, which contains the first ten modules involved in biosynthesis of mycolactone, weighs in at 1.8 MDa¹⁷. Since their discovery, understanding of PKSs has been based on the more well understood fatty acid synthase (FAS) from primary metabolism. Biosynthesis of fatty acids by FAS and polyketides by PKSs share a common core reaction: homologation of two-carbon units to form poly- β -ketone chains. This core reaction is accomplished by three domains: an acyl-transferase (AT), a ketosynthase (KS), and an acyl-carrier protein (ACP) (Figure 1.5). The ACP is post-translationally modified with a phosphopantetheine arm, which is an approximately 20 Å long flexible tether that allows for shuttling of substrates between the catalytic domains, while keeping the reactive poly- β -ketone intermediates covalently bound to the enzyme. An acyl starter unit, typically acetyl-CoA or decarboxylated malonyl-CoA, is selected by the AT to initiate catalysis. This starter unit is then loaded onto the KS, by way of the ACP. The AT is then free to load an extender unit, typically malonyl-CoA, which is subsequently transferred to the ACP. Finally, the KS catalyzes a decarboxylative

Claisen condensation between the starter unit and the ACP-bound extender unit, resulting in a diketide bound to the ACP, at which point the extension cycle repeats.

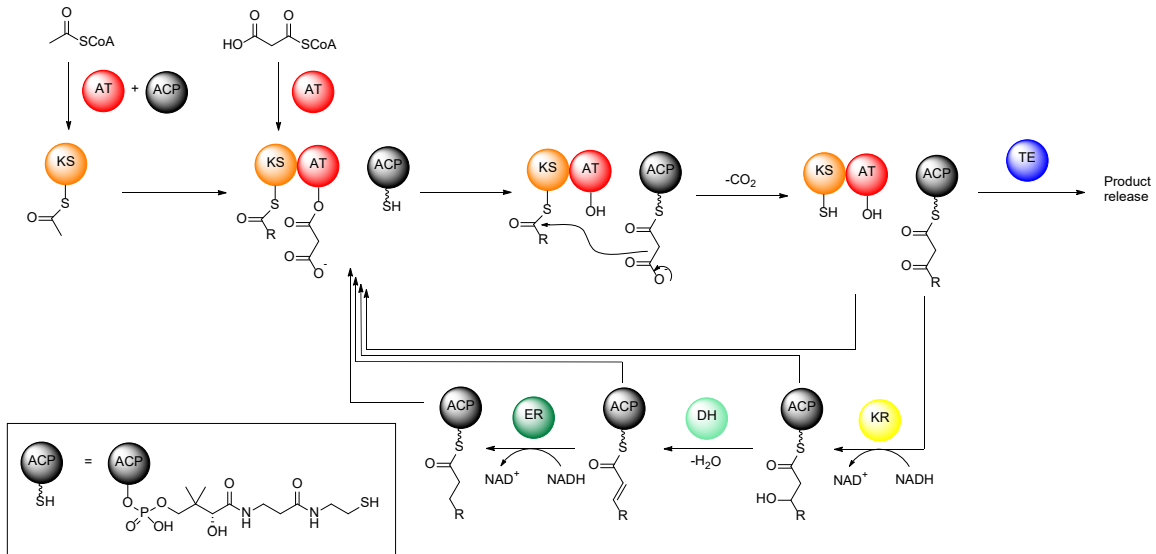


Figure 1.5. General polyketide synthase catalytic cycle. The core reaction of PKS biosynthesis is elongation *via* Claisen condensation, catalyzed by the KS. The newly-installed β -ketone can then be optionally reduced by KR, DH, and ER domains. All substrates and intermediates are covalently anchored to the enzyme by the ACP. The TE catalyzes release of the final product.

In FAS, the β -ketone is fully reduced to a methylene after each round of elongation by a full suite of tailoring domains. First, the keto group is reduced to a hydroxyl by the ketoreductase (KR). This hydroxyl group is then eliminated as water, forming an alkene, by the dehydratase (DH), followed by a final reduction to the methylene by the enoyl reductase (ER). PKSs may contain some, all, or none of these reductive domains, and the ketone installed by each elongation step can be optionally modified to any of these reduced states. PKSs may also

contain a C-methyltransferase (MT) domain for installation of a *S*-adenosyl-methionine (SAM) derived methyl group at the α -position following each round of extension. Finally, the fully extended polyketide or fatty acid is specifically released from the enzyme by a thioesterase (TE). This release is most commonly accomplished by hydrolysis, in the case of FAS, or macrolactonization, in the case of many modPKSs. However, TEs also occasionally catalyze lactonization¹⁸ or Claisen cyclization¹⁹⁻²⁰.

The non-reducing (NR)-PKSs contain two unique domains: a starter unit AT (SAT) and a product template (PT) domain. In these systems, the SAT domain is responsible for specific selection of the starter unit, which can be more complex than the typical acetyl-CoA starter unit, and in some cases directly transferred from an upstream FAS or HR-PKS. For example, the PksA SAT accepts a hexanoyl starter unit from a dedicated pair of yeast-like FASs²¹⁻²³. Similarly, in asperfuranone biosynthesis, the NR-PKS AfoE accepts a dimethyloctadienone starter unit from the highly-reducing (HR)-PKS AfoG²⁴. Hybrid systems in which the PKS accepts a starter unit from a non-ribosomal peptide synthetase further increase the diversity of polyketide-based natural products²⁵⁻²⁷. In systems with a dedicated SAT, the extender unit loading AT is termed the malonyl-CoA AT (MAT). The PT domain is responsible for

regiospecific cyclization of the fully oxidized, extended poly- β -ketone in NR-PKSs²⁸.

1.2.2. Classification of PKSs – Early lessons from modular PKSs

FASs and PKSs are closely related biosynthetic enzymes, both accomplish the condensation of acetate units into carbon chains, and share many biosynthetic domains. FASs are broadly categorized based on their structural organization. Type I FASs (fungal, mammalian) are multienzymes, where all domains are covalently linked in a single polypeptide. Conversely, in type II FASs (bacterial), each domain is contained on its own discrete protein. PKSs are also categorized based on this system, but include a third category – type III PKSs, which act on CoA- (rather than ACP-) bound substrates.

Type I PKSs exist as both modular (modPKS) and iterative (iPKS) systems. ModPKSs, where each set of domains (module) canonically performs only a single elongation/modification reaction to extend the growing polyketide chain, are the most well understood class of PKSs. Modules are organized in an assembly-line fashion, where the domains present in each module (generally) dictate the chemical modification at that stage of extension. The first characterized type I PKS was the modPKS involved in biosynthesis of the antibiotic erythromycin by *Saccharopolyspora erythraea*, 6-deoxyerythronolide B

synthase (DEBS)²⁹. DEBS was independently identified by two labs using two distinct genetic approaches. Peter Leadley's group at Cambridge took advantage of the clustering of biosynthetic genes with those that confer self-resistance, and identified PKS genes in close proximity to the erythromycin resistance gene *ermE*³⁰⁻³². Conversely, Leonard Katz's group at Abbott Laboratories identified mutations that abolished production of 6-deoxyerythronolide B, and found them to be located within PKS genes³³⁻³⁵.

DEBS consists of six biosynthetic modules housed in three polypeptides. Each module contains, at minimum, the domains required to catalyze the condensation of propionate units into a polyketide chain – KS, AT, and ACP. Additionally, a variable selection of tailoring domains is present in each module, which determines the reductive modification of each β -ketone during elongation. For example, module 4 contains the whole suite of reductive domains – DH, ER, and KR – and thus the ketone installed by module 4 is fully reduced to the saturated thioester. Conversely, module 1 contains only a KR, so the ketone should be reduced to a hydroxy group. As expected, the PKS product 6-deoxyerythronolide B has a structure fully consistent with the domain architecture of each module³⁵. Therefore, DEBS is the paradigm example of co-linearity in modPKS biosynthesis, where products can be predicted based on which domains are present in each module, and domain architecture can be

predicted based on the structure of the polyketide product. This concept of co-linearity and the assembly-line nature of modPKSs make them much simpler to analyze and engineer than iPKSs.

In contrast to the assembly-line architecture of modPKSs, where each module performs only a single round of extension, iPKSs consist of only a single set of domains, which act repeatedly during each catalytic cycle. Therefore, while the enzymes themselves are smaller and simpler, the programming and prediction of iPKS products is considerably more difficult. iPKSs are further categorized by the extent to which the polyketide is reduced, from highly-reducing (HR)-PKSs with up to a full suite of reductive domains, to non-reducing (NR)-PKSs which lack any reductive domains. A prototypical example of NR-PKSs is PksA, responsible for the first step of aflatoxin biosynthesis in *Aspergillus parasiticus*.

1.3. Tools for understanding polyketide biosynthesis

1.3.2. Domain dissection for in vitro analysis

The iterative NR-PKSs are typically fungal and notoriously difficult to express in traditional heterologous bacterial hosts (e.g. *E. coli*), so traditional biochemical *in vitro* experiments on this class of PKS have been in short supply.

Due to their iterative nature, traditional *in vivo* mutagenesis experiments do not allow for detailed interrogation of the biosynthesis of iPKSs. Inactivating a given catalytic domain through mutagenesis will completely abolish production *in vivo*, since each domain is used repeatedly throughout the catalytic cycle.

Conversely, the activity of individual domains, or even full modules, of modPKS can be interrogated by inactivation and subsequent analysis of the accumulated product. For example, when the KR of module 5 of DEBS is removed, an erythromycin analog is produced in which the keto group has been maintained at the expected position, rather than being reduced to a hydroxyl group³⁵.

Limited proteolysis is commonly used to dissect multidomain enzymes into their individual domains³⁶. The relatively unstructured linker regions between the globular domains is proposed to be more accessible to proteases. In early studies of FAS isolated from chicken, fortuitous proteolysis from native co-purified proteases was observed under specific conditions³⁷. The susceptibility of the multidomain FAS to limited proteolysis could be used as a biochemical tool; for example, through controlled trypsinization, the TE was isolated and characterized³⁸. This technique was later used to dissect the modPKSs DEBS1, 2, and 3, and identify their linker regions by limited proteolysis followed by *N*-terminal sequencing³⁹⁻⁴⁰. Together, these two methods allowed for cloning and heterologous expression of isolated PKS and FAS domains, which can be

analyzed *in vitro* to determine catalytic activity or for structural studies.

However, if native, intact protein cannot be expressed (as in the case of most NR-PKSs), this technique is of little utility. Instead, linker regions can be predicted based on primary structure, as they exhibit less sequence conservation than the catalytic domains. Udvary and Merski improved upon this simple method of linker region prediction with their UMA algorithm, which also takes predicted local secondary structure and hydrophobicity into account⁴¹. Their algorithm was validated with enzymes of known structure, such as methionine synthase (MetH) and DNA polymerase I. It has since been used to successfully dissect a number of NR-PKSs, including PksA¹⁹ and PksCT⁴².

Dissection of iPKS catalytic domains has allowed for detailed characterization of fungal polyketide biosynthesis. For instance, despite herculean efforts with a variety of heterologous expression systems, the NR-PKS PksA was not amenable to expression as an intact construct. It was not until the UMA algorithm allowed dissection of PksA into smaller mono-, di- and tri-domain constructs that its activity could be studied *in vitro*¹⁹. By reconstituting various PksA domain combinations, the role of each domain in biosynthesis was defined (Figure 1.6). The minimal PKS (SAT-KS-MAT + ACP), and the minimal PKS plus TE, produced only minute quantities of products. However, addition of the PT domain to the minimal PKS resulted in significant production of the

regiospecifically cyclized product naphopyrone **12**, demonstrating the PT's role in cyclization. Addition of the TE to this reaction nearly eliminated production of naphopyrone, instead producing the on-pathway product norsolorinic acid anthrone **13** and its oxidation product norsolorinic acid **14**. Therefore, domain dissection allows not only for *in vitro* analysis of PKSs that cannot be expressed in their entirety, but also allows for interrogation of activities of specific domains.

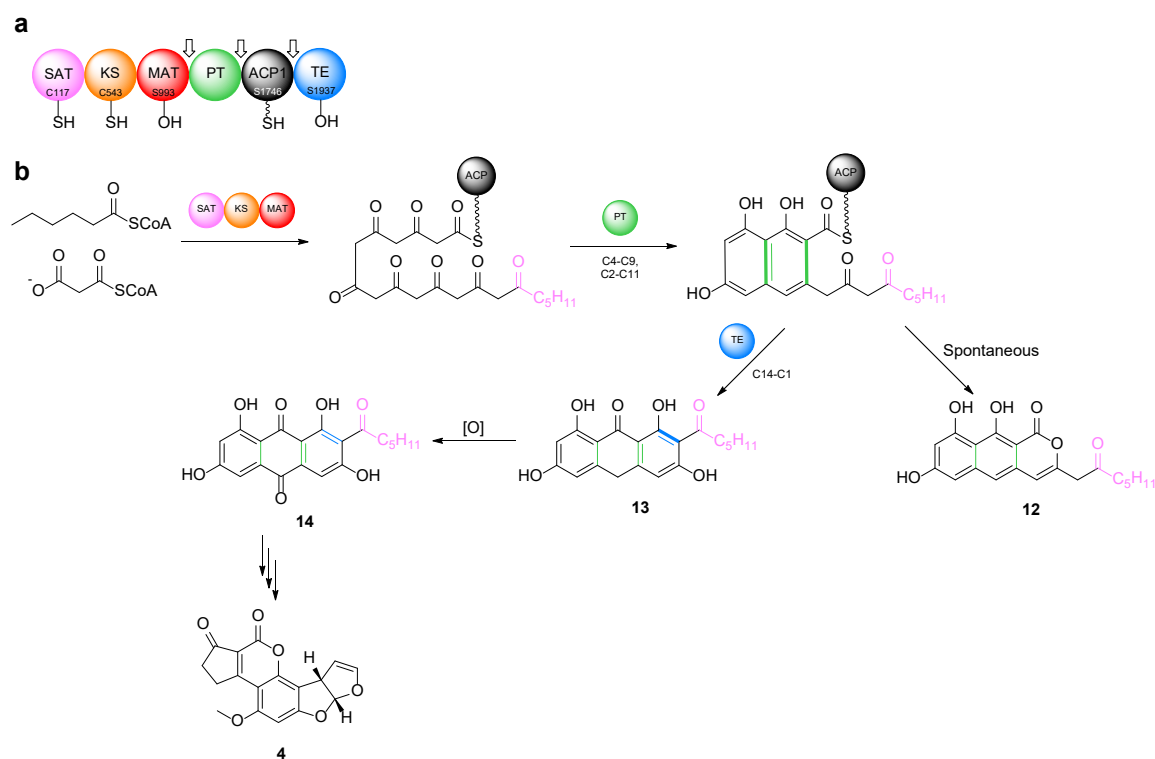


Figure 1.6. PksA biosynthetic scheme, determined by domain deconstruction. (a) Domain organization of PksA. Cut sites for domain deconstruction are indicated by block arrows. (b) Role of PksA domains in biosynthesis of **13**, determined by *in vitro* reconstitution of dissected domains, as indicated.

A particularly powerful application of domain dissection is the ability to swap domains of different PKSs to further decipher their role in catalytic

programming, as well as synthesize novel polyketides. By combinatorial reconstitution of domains from PKSs with products of defined chain length, cyclization pattern, and release mechanism, novel polyketides can be rationally designed and produced⁴³⁻⁴⁴. Chain length was found to be strictly controlled by the KS: in all combinations of CTB1 minimal PKS and alternative PT and TE domains, a C₁₄ polyketide was produced, regardless of the chain length specificity of the PT and TE domains⁴⁴. Cyclization register was faithfully determined by the PT domain, as long as the chain length was within one acetate unit of the PT's native substrate. For example, the ACAS PT combined with the CTB1 minimal PKS produced the C₆-C₁₁ cyclized pannorin. Finally, the TE is crucial for product release and enzymatic turnover, and tends to be less tolerant to alternative substrates than the PT. This observation is consistent with previous observations of TE-mediated editing of off-target or stalled enzymatic products⁴⁵. Productive biosynthesis of non-native products by combinatorial PKS reconstitution relies upon enzymatic cyclization and release occurring fast enough to overcome spontaneous cyclization and release, or TE editing. These limitations can be partially overcome by fusing domain combinations into a single polypeptide⁴³. In both native and combinatorial PKSs, deconstruction imposes a biosynthetic penalty since domains must come together

intermolecularly to accomplish chemistry, as opposed to intramolecular reactions in the native system.

1.3.2. Structures of NR-PKS monodomains

Other than the DEBS KS-AT didomain, structures of multidomain (or full-length) PKS constructs have been difficult to obtain, likely due to their structural flexibility. However, structures of each PKS domain have been elucidated as dissected monodomain constructs. While many structures have been solved of modPKS domains (including reductive domains), this summary will focus on domains isolated from the NR-PKSs.

The first structure of an isolated NR-PKS domain, PksA PT, was reported fairly quickly after the UMA algorithm allowed successful deconstruction of PksA²⁸. The PT domain, a domain unique to NR-PKSs, catalyzes aldol cyclization and aromatization of the reactive poly- β -ketone intermediate. In the case of PksA, cyclization occurs regiospecifically, first between C₄ and C₉, followed by C₂-C₁₁. The mechanism by which this regiospecific cyclization is accomplished was an interesting enzymological problem. The 1.8 Å resolution PT structure reveals a double hot dog fold, similar to that of the DH domains from mFAS and modPKSs (Figure 1.7). Notably, the PT domain is located just downstream of the MAT domain in NR-PKS, analogous to the location of the DH

in mFAS and modPKS. Palmitate (from *E. coli* heterologous expression) was found bound within a 30 Å deep pocket that can be divided into three regions. The deepest region of the pocket is lined with hydrophobic residues consistent with binding of the hexyl starter unit of the poly-β-ketone. Next is a slightly wider cyclization chamber, which contains the His/Asp catalytic dyad. Finally, a 14 Å long tunnel leading to the cyclization chamber is proposed to bind the phosphopantetheine arm bearing the poly-β-ketone substrate. The organization of the catalytic dyad and a corresponding oxyanion hole within the cyclization chamber are proposed to dictate the regioselectivity of PT cyclization.

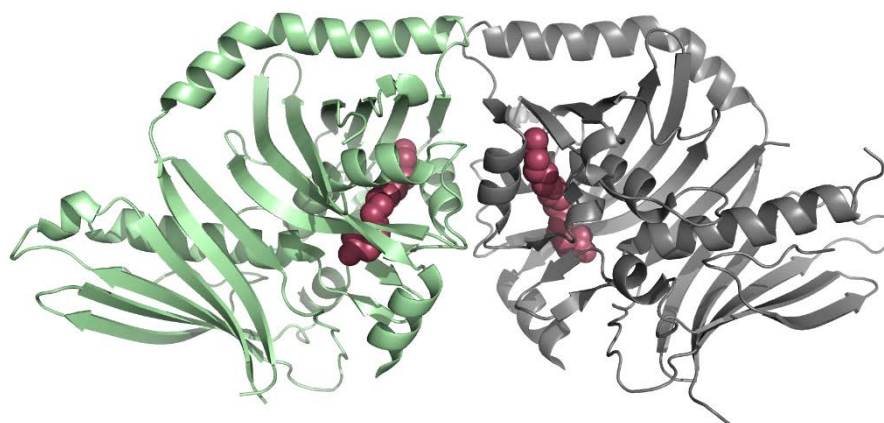


Figure 1.7. 1.8 Å crystal structure of PksA PT. PksA PT forms a double hot dog dimer. The two monomers are shown as ribbon diagrams in green and gray. Palmitate visualized in the deep active site pocket is shown as magenta spheres.

Shortly after the PksA PT structure was solved, the X-ray crystal structure of PksA TE at 1.7 Å resolution was published²⁰. The PksA TE is distinct from mFAS and modPKS TEs, in that it catalyzes a C-C bond formation by way of Claisen cyclization, instead of thioester hydrolysis, to release the final PKS

product norsolorinic acid anthrone¹⁹. Like other related TEs, monomeric PksA TE exhibits an α/β hydrolase fold and a helical lid region (Figure 1.8)⁴⁶⁻⁵⁰. Additionally, PksA TE contains the canonical Ser/His/Asp catalytic triad typical of TEs. However, whereas modPKS TEs contain an open, hydrophilic, substrate-binding channel, the substrate-binding region of PksA TE is deep and highly aromatic. As this feature is conserved amongst TEs with Claisen cyclase activity, the hydrophobic nature of the PksA TE substrate-binding chamber appears to be characteristic of this class of TE domains.

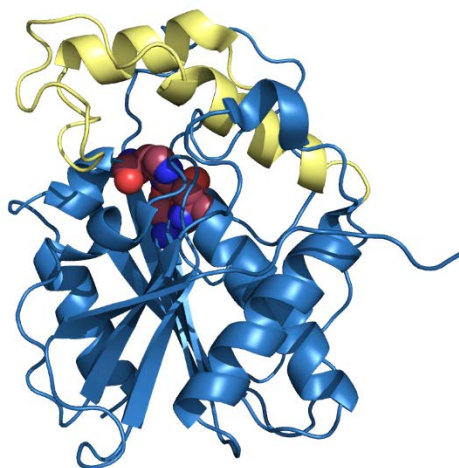


Figure 1.8. 1.7 Å crystal structure of PksA TE, shown as a ribbon diagram. The α/β hydrolase core is shown in blue and the lid region is shown in yellow. The catalytic triad residues (Ser1937/His2088/D1964) are shown as magenta spheres.

To round out the modifying region of PksA, the solution structure of PksA ACP was solved in 2010⁵¹. While type II FAS and PKS ACP structures have been extensively studied⁵²⁻⁵⁶, only a single other type I PKS ACP structure has been reported⁵⁷. PksA ACP was the first structure of an iPKS ACP, which is required

to shuttle a wide variety of intermediates throughout the catalytic cycle. Like other type I and type II ACPs, PksA ACP consists of a four-helix bundle (Figure 1.9). However, helix III lies perpendicular to the other three parallel helices (I, II and IV), in contrast to the other structures. This unusual conformation of helix III was attributed to its unusually hydrophobic nature, and resulting tight packing with the other helices. In FAS ACPs, the more flexible helix III undergoes a conformational shift to allow for acyl chain binding^{55, 58-60}. While a slight conformational change was observed between *holo* PksA ACP and acylated (hexanoyl-loaded) *holo* PksA ACP, no analogous sequestration of the hexanoyl starter unit was observed. ACP sequences are highly homologous amongst the NR-PKSs, suggesting these structural features may be conserved amongst this class of ACPs.

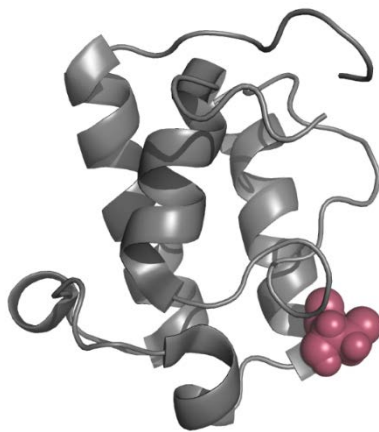


Figure 1.9. Solution structure of PksA ACP, shown as a ribbon diagram. The phosphopantetheine attachment point (Ser1746) is shown as magenta spheres.

The most recent structure of a NR-PKS domain is that of the CazM SAT domain, involved in chaetoviridin **15** biosynthesis in *Chaetomium globosum*⁶¹. NR-

PKS SAT domains have the unique capability of accepting an acyl starter unit from an upstream FAS or PKS, as opposed to the more typical acetyl-CoA. For example, PksA accepts its hexanoyl starter unit from a dedicated pair of yeast-like FASs²³. Similarly, CazM utilizes a 4-methyl-hex-2-enoyltriketide starter unit **16**, synthesized by the HR-PKS CazF, and transferred by the CazM SAT domain (Figure 1.10a)⁶². The structure of CazM SAT resembles other FAS and PKS AT domains, containing an α/β -hydrolase core and a small ferredoxin-like subdomain (Figure 1.10b). A primarily hydrophobic active site cavity was observed in the interface between the α/β -hydrolase and ferredoxin subdomains. These hydrophobic residues likely contribute to CazM's selectivity for its triketide substrate, in contrast to a more oxidized, tetraketide product of CazF. Comparison of crucial binding residues within CazM to other NR-PKS SAT domains may allow for rational engineering towards production of novel polyketides with non-native starter units.

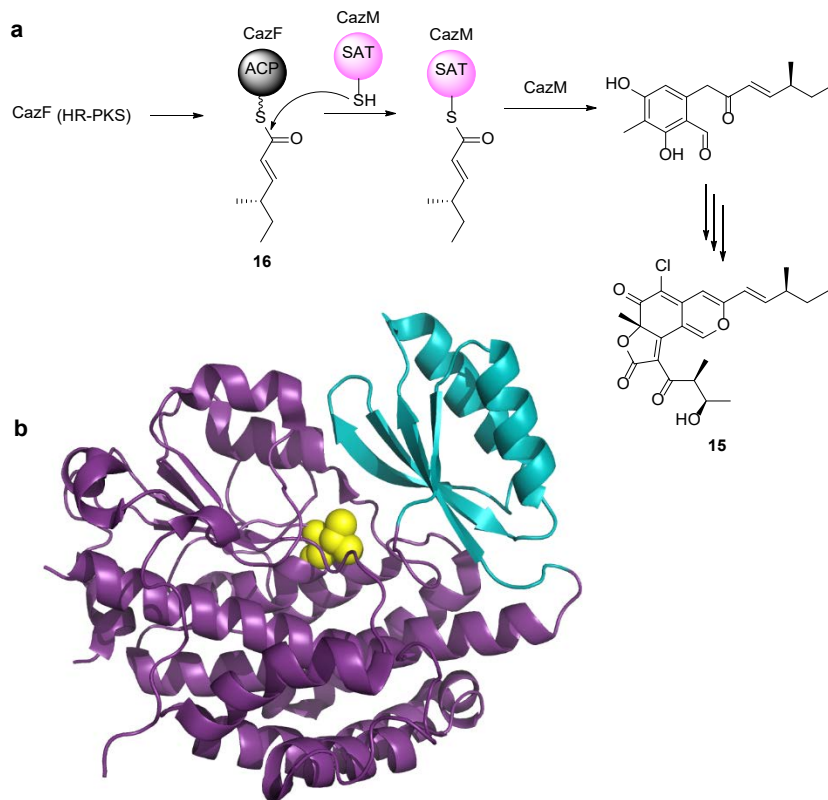


Figure 1.10. (a) Role of CazM SAT in biosynthesis of chaetoviridin **15** by the HR-PKS CazF and the NR-PKS CazM. CazM accepts a starter unit synthesized by CazF to initiate NR-PKS biosynthesis. (b) 1.6 Å crystal structure of CazM SAT. The α/β hydrolase core is shown in purple, and the ferredoxin subdomain is shown in cyan. The active site cysteine is shown as yellow spheres.

1.3.3. Full-length PKS structures and the rise of cryo-electron microscopy

As discussed previously, type I FAS has long been used as a basis for understanding type I PKS biochemistry. This tactic is also applicable for understanding the structure of type I PKSs, since they share many of the same biosynthetic domains, organized in similar architectures. The X-ray crystal structure of intact, porcine FAS (mFAS) was first reported in 2005 at 4.5 Å resolution⁶³, and was quickly refined to 3.2 Å resolution two years later⁶⁴. The

dimeric FAS resembles a gingerbread man, with “legs” consisting of the KS-AT condensing region and “arms” containing the modifying region (DH, KR, ER, ACP, TE) (Figure 1.11). The dimer is primarily mediated by the KS in the “leg” region, and the ER and DH in the “arm” region. Presumably due to their inherent flexibility, neither the ACP nor the TE domains were resolved in either structure. The approximate location of the TE at the end of each of the “arms” was inferred from EM images of guinea pig FAS labelled with anti-TE antibodies⁶⁵.

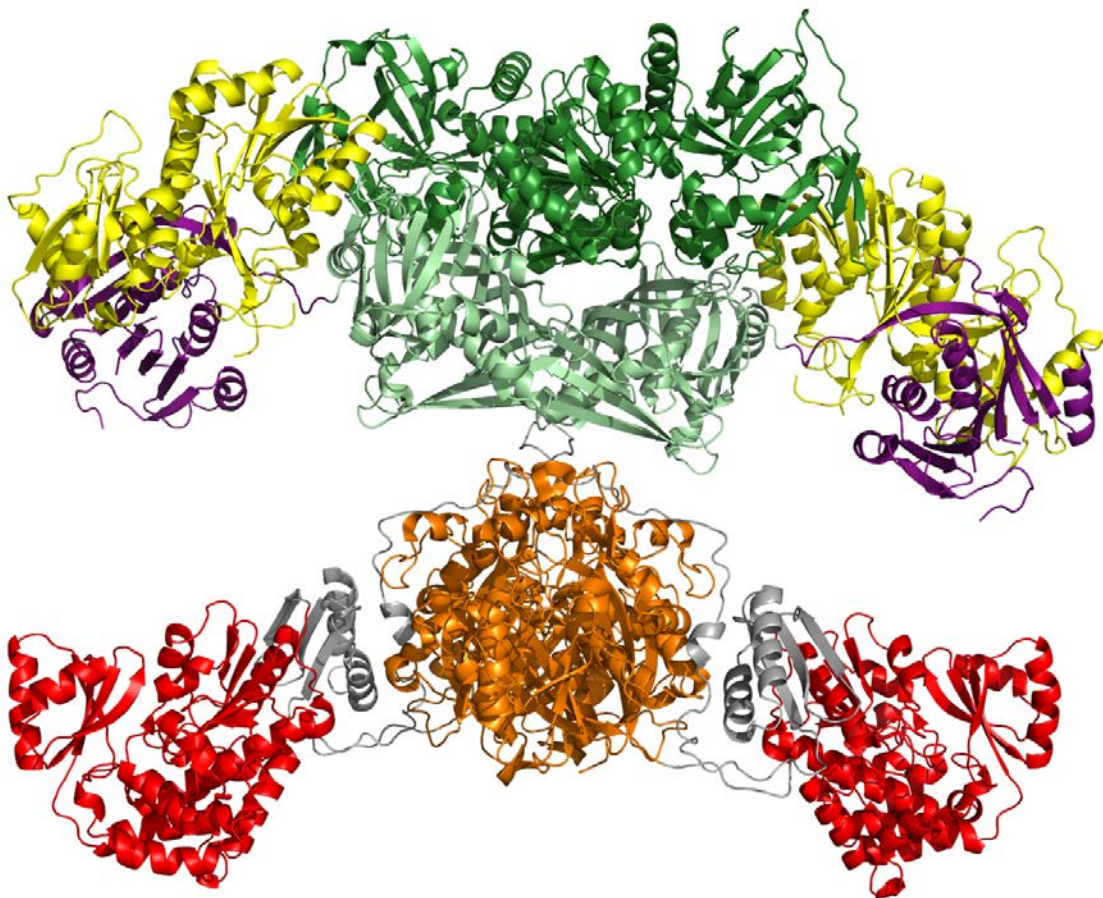


Figure 1.11. 3.2 Å crystal structure of porcine FAS. Catalytic domains are color coded (from N- to C-terminus): KS – orange, AT – orange, DH – light green, ψ MT – purple, ER - green, KR – yellow. Linker regions are shown in gray.

Shortly after the first porcine FAS structure was published, the first hint towards PKS domain architecture emerged from the Khosla lab, which published the 2.7 Å crystal structure of the KS-AT didomain from DEBS module 5⁶⁶. This structure closely resembled the corresponding domains in mFAS, in both overall topology and individual domain structures. The KS domains form a tight dimer around the 2-fold axis, and are flanked on each side by the AT domains (Figure 1.12). Due to this similarity in structure and domain architecture, porcine FAS was validated as a model for type I PKS, both biochemically and structurally, and type I PKSs, both modular and iterative, were proposed to have a structure similar to that of mammalian FAS.

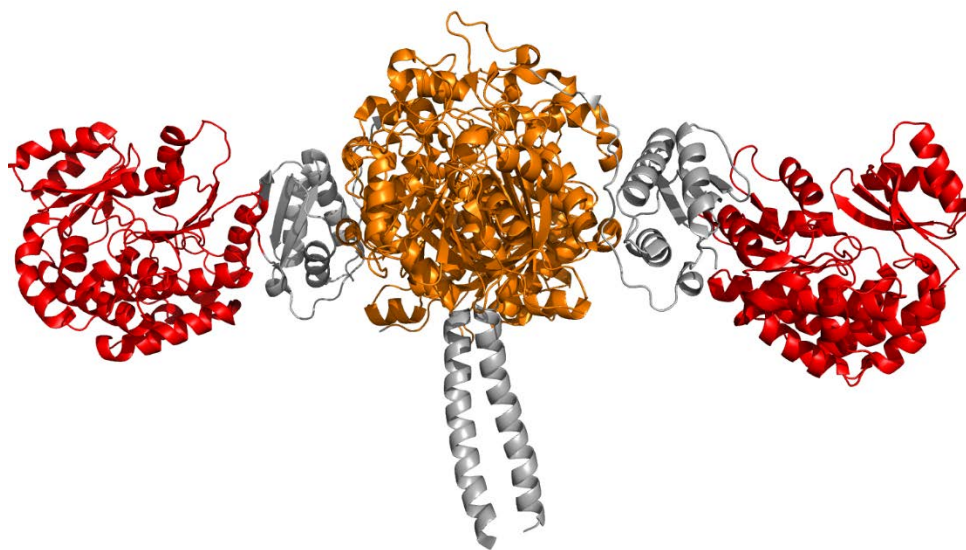


Figure 1.12. 2.7 Å crystal structure of the KS-AT didomain from DEBS module 5. Catalytic domains are color-coded as in Figure 1.11 (KS – orange, AT – red). Linker regions are shown in gray.

In 2016, the relevancy of type I FAS as a model for type I PKS structures was further supported by the multidomain structure of an iterative HR-PKS,

mycoseric acid synthase (MSA)⁶⁷. To achieve a hybrid crystal structure of the full-length enzyme, Maier and co-workers split MSA into a condensing region (KS-AT) and modifying region (DH-ER-KR). The condensing region structure was solved to 2.2 Å resolution, and closely resembles that of DEBS KS-AT and mFAS. The 3.75 Å resolution x-ray crystal of the modifying region also resembles that of mFAS, consisting of an overall dimeric organization mediated by the DH and ER domains. When combined *in silico*, these structures yield an overall iPKS structure resembling that of mFAS, with modifying “arm” and condensing “leg” regions connected by a flexible “waist”.

Recently, advances in cryo-electron microscopy (EM) have provided new promise for elucidation of full-length PKS structures previously inaccessible by X-ray crystallography. Previously, structure determination by cryo-EM was limited to relatively large and rigid proteins and biological particles, such as the ribosome and viral particles⁶⁸⁻⁷². Improvements in sample preparation and development of direct-electron detectors have pushed cryo-EM into the realm of high-resolution structure determination methods⁷³⁻⁷⁷. These advances have allowed for the determination of a cryo-EM structure of an intact PKS module involved in pikromycin biosynthesis, PikAIII⁷⁸. This structure was surprising, in that its overall architecture deviates from the “gingerbread man” model developed based on the mFAS structure, which has thus far been consistent with

other modPKS and iPKS structures. The PikAIII module is arch shaped, with a single reaction chamber at its center, rather than the dual reaction chambers observed in mFAS and HR-PKS. This architecture represents a 120° rotation of the AT relative to the KS compared to other modPKS structures, and a completely distinct KS-AT interface. Cryo-EM analysis of PikAIII also resulted in observation of substrate-dependent localization of the ACP domain⁷⁹.

1.4. Outlook

Through advances in PKS deconstruction and structural elucidation over the past ten years, the “black box” of NR-PKS biosynthesis is beginning to be unpacked. Activities of individual domains, and their role in programming biosynthesis, have been well studied. Similarly, many monodomain structures have been determined, illuminating the structural basis for starter unit selection, cyclization, and product release. However, the question remains as to how overall biosynthetic programming is achieved. Engineering of PKSs to produce novel compounds with valuable bioactivity is an important goal of the PKS field. However, success in achieving this goal has been limited. In order to successfully engineer PKSs, understanding of substrate tolerance and enzymatic programming must be improved. In this thesis, studies to interrogate the

tolerance of PKS domains to non-native starter units demonstrate some capabilities and limitations of iPKS engineering. Structures of full-length iPKSs will help decipher how individual domains interact to accomplish the complex series of reactions required by iterative polyketide biosynthesis. In an important step toward this goal, we present a multidomain structure of an iPKS locked in its loading state by mechanism-based crosslinking.

1.5. References

1. Staunton, J., Combinatorial biosynthesis of erythromycin and complex polyketides. *Curr Opin Chem Biol* **1998**, 2 (3), 339-45.
2. Staunton, J.; Wilkinson, B., Biosynthesis of Erythromycin and Rapamycin. *Chemical reviews* **1997**, 97 (7), 2611-2630.
3. Tobert, J. A., Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nature reviews. Drug discovery* **2003**, 2 (7), 517-26.
4. McKenney, J. M., Lovastatin: a new cholesterol-lowering agent. *Clin Pharm* **1988**, 7 (1), 21-36.
5. Sutherland, A.; Auclair, K.; Vederas, J. C., Recent advances in the biosynthetic studies of lovastatin. *Curr Opin Drug Discov Devel* **2001**, 4 (2), 229-36.
6. Jelinek, C. F.; Pohland, A. E.; Wood, G. E., Worldwide occurrence of mycotoxins in foods and feeds--an update. *J Assoc Off Anal Chem* **1989**, 72 (2), 223-30.
7. Newberne, P. M.; Butler, W. H., Acute and chronic effects of aflatoxin on the liver of domestic and laboratory animals: a review. *Cancer research* **1969**, 29 (1), 236-50.
8. Wogan, G. N.; Newberne, P. M., Dose-response characteristics of aflatoxin B1 carcinogenesis in the rat. *Cancer research* **1967**, 27 (12), 2370-6.
9. Daub, M. E.; Ehrenshaft, M., THE PHOTOACTIVATED CERCOSPORA TOXIN CERCOSPORIN: Contributions to Plant Disease and Fundamental Biology. *Annu Rev Phytopathol* **2000**, 38, 461-490.
10. Farr, D. F.; Bills, G. F.; Chamuris, G. P.; Rossman, A. Y., *Fungi on plants and plant products in the United States*. APS Press: Minnesota, 1989.
11. Collie, J. N.; Myers, W. S., VII. - The formation of orcinol and other condensation products from dehydracetic acid. *J Chem Soc Trans* **1893**, 63, 122-128.

12. Birch, A. J.; Massy-Westropp, R. A.; Moye, C. J., Studies in relation to biosynthesis. VII. 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* *Australian Journal of Chemistry* **1955**, *8* (4), 539-44.
13. Keller, N. P.; Hohn, T. M., Metabolic Pathway Gene Clusters in Filamentous Fungi. *Fungal genetics and biology : FG & B* **1997**, *21* (1), 17-29.
14. Malpartida, F.; Hopwood, D. A., Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **1984**, *309* (5967), 462-4.
15. McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C., Engineered biosynthesis of novel polyketides. *Science* **1993**, *262* (5139), 1546-50.
16. McDaniel, R.; Ebert-Khosla, S.; Fu, H.; Hopwood, D. A.; Khosla, C., Engineered biosynthesis of novel polyketides: influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. *Proc Natl Acad Sci U S A* **1994**, *91* (24), 11542-6.
17. Stinear, T. P.; Mve-Obiang, A.; Small, P. L.; Frigui, W.; Pryor, M. J.; Brosch, R.; Jenkin, G. A.; Johnson, P. D.; Davies, J. K.; Lee, R. E.; Adusumilli, S.; Garnier, T.; Haydock, S. F.; Leadlay, P. F.; Cole, S. T., Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A* **2004**, *101* (5), 1345-9.
18. Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A., Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun (Camb)* **2012**, *48* (96), 11772-4.
19. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008**, *320* (5873), 243-6.
20. Korman, T. P.; Crawford, J. M.; Labonte, J. W.; Newman, A. G.; Wong, J.; Townsend, C. A.; Tsai, S. C., Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci* **2010**, *107* (14), 6246-51.
21. Hitchman, T. S.; Schmidt, E. W.; Trail, F.; Rarick, M. D.; Linz, J. E.; Townsend, C. A., Hexanoate synthase, a specialized type I fatty acid synthase in aflatoxin B1 biosynthesis. *Bioorganic chemistry* **2001**, *29* (5), 293-307.
22. Watanabe, C. M.; Townsend, C. A., Initial characterization of a type I fatty acid synthase and polyketide synthase multienzyme complex NorS in the biosynthesis of aflatoxin B(1). *Chemistry & biology* **2002**, *9* (9), 981-8.
23. Foulke-Abel, J.; Townsend, C. A., Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase. *Chembiochem : a European journal of chemical biology* **2012**, *13* (13), 1880-4.
24. Chiang, Y. M.; Szewczyk, E.; Davidson, A. D.; Keller, N.; Oakley, B. R.; Wang, C. C., A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. *Journal of the American Chemical Society* **2009**, *131* (8), 2965-70.

25. Yoo, Y. J.; Kim, H.; Park, S. R.; Yoon, Y. J., An overview of rapamycin: from discovery to future perspectives. *Journal of industrial microbiology & biotechnology* **2016**.
26. Gulder, T. A.; Moore, B. S., Salinosporamide natural products: Potent 20 S proteasome inhibitors as promising cancer chemotherapeutics. *Angewandte Chemie* **2010**, *49* (49), 9346-67.
27. Du, L.; Sanchez, C.; Shen, B., Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic engineering* **2001**, *3* (1), 78-95.
28. Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpheh, O.; Tsai, S. C.; Townsend, C. A., Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **2009**, *461* (7267), 1139-43.
29. Caffrey, P.; Bevitt, D. J.; Staunton, J.; Leadlay, P. F., Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from *Saccharopolyspora erythraea*. *FEBS letters* **1992**, *304* (2-3), 225-8.
30. Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F., An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **1990**, *348* (6297), 176-8.
31. Malpartida, F.; Hopwood, D. A., Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. 1984. *Biotechnology* **1992**, *24*, 342-3.
32. Bevitt, D. J.; Cortes, J.; Haydock, S. F.; Leadlay, P. F., 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur J Biochem* **1992**, *204* (1), 39-49.
33. Weber, J. M.; Losick, R., The use of a chromosome integration vector to map erythromycin resistance and production genes in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *Gene* **1988**, *68* (2), 173-80.
34. Tuan, J. S.; Weber, J. M.; Staver, M. J.; Leung, J. O.; Donadio, S.; Katz, L., Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene* **1990**, *90* (1), 21-9.
35. Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L., Modular organization of genes required for complex polyketide biosynthesis. *Science* **1991**, *252* (5006), 675-9.
36. Klenow, H.; Overgaard-Hansen, K., Proteolytic cleavage of DNA polymerase from *Escherichia Coli* B into an exonuclease unit and a polymerase unit. *FEBS letters* **1970**, *6* (1), 25-27.
37. Stoops, J. K.; Arslanian, M. J.; Oh, Y. H.; Aune, K. C.; Vanaman, T. C.; Wakil, S. J., Presence of two polypeptide chains comprising fatty acid synthetase. *Proc Natl Acad Sci U S A* **1975**, *72* (5), 1940-4.

38. Smith, S.; Agradi, E.; Libertini, L.; Dileepan, K. N., Specific release of the thioesterase component of the fatty acid synthetase multienzyme complex by limited trypsinization. *Proc Natl Acad Sci U S A* **1976**, *73* (4), 1184-8.
39. Aparicio, J. F.; Caffrey, P.; Marsden, A. F.; Staunton, J.; Leadlay, P. F., Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. *The Journal of biological chemistry* **1994**, *269* (11), 8524-8.
40. Staunton, J.; Caffrey, P.; Aparicio, J. F.; Roberts, G. A.; Bethell, S. S.; Leadlay, P. F., Evidence for a double-helical structure for modular polyketide synthases. *Nat Struct Biol* **1996**, *3* (2), 188-92.
41. Udvary, D. W.; Merski, M.; Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *Journal of Molecular Biology* **2002**, *323* (3), 585-598.
42. Storm, P. A.; Herbst, D. A.; Maier, T.; Townsend, C. A., Functional and Structural Analysis of Programmed C-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin. *Cell Chem Biol* **2017**, *24* (3), 316-325.
43. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *Journal of the American Chemical Society* **2014**, *136* (20), 7348-62.
44. Vagstad, A. L.; Newman, A. G.; Storm, P. A.; Belecki, K.; Crawford, J. M.; Townsend, C. A., Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angewandte Chemie* **2013**, *52* (6), 1718-21.
45. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *Journal of the American Chemical Society* **2012**, *134* (15), 6865-77.
46. Tsai, S. C.; Lu, H.; Cane, D. E.; Khosla, C.; Stroud, R. M., Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **2002**, *41* (42), 12598-606.
47. Tsai, S. C.; Miercke, L. J.; Krucinski, J.; Gokhale, R.; Chen, J. C.; Foster, P. G.; Cane, D. E.; Khosla, C.; Stroud, R. M., Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc Natl Acad Sci U S A* **2001**, *98* (26), 14808-13.
48. Bruner, S. D.; Weber, T.; Kohli, R. M.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T.; Stubbs, M. T., Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **2002**, *10* (3), 301-10.
49. Samel, S. A.; Wagner, B.; Marahiel, M. A.; Essen, L. O., The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol* **2006**, *359* (4), 876-89.

50. Frueh, D. P.; Arthanari, H.; Koglin, A.; Vosburg, D. A.; Bennett, A. E.; Walsh, C. T.; Wagner, G., Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **2008**, *454* (7206), 903-6.
51. Wattana-amorn, P.; Williams, C.; Ploskon, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P., Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry* **2010**, *49* (10), 2186-93.
52. Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J., Solution structure of the actinorhodin polyketide synthase acyl carrier protein from *Streptomyces coelicolor* A3(2). *Biochemistry* **1997**, *36* (20), 6000-8.
53. Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W., Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein. *Biochemistry* **2003**, *42* (16), 4648-57.
54. Findlow, S. C.; Winsor, C.; Simpson, T. J.; Crosby, J.; Crump, M. P., Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* **2003**, *42* (28), 8423-33.
55. Zornetzer, G. A.; Fox, B. G.; Markley, J. L., Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* **2006**, *45* (16), 5217-5227.
56. Sharma, A. K.; Sharma, S. K.; Surolia, A.; Surolia, N.; Sarma, S. P., Solution structures of conformationally equilibrium forms of holo-acyl carrier protein (PfACP) from *Plasmodium falciparum* provides insight into the mechanism of activation of ACPs. *Biochemistry* **2006**, *45* (22), 6904-16.
57. Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C., Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein science : a publication of the Protein Society* **2007**, *16* (10), 2093-107.
58. Roujeinikova, A.; Baldock, C.; Simon, W. J.; Gilroy, J.; Baker, P. J.; Stuitje, A. R.; Rice, D. W.; Slabas, A. R.; Rafferty, J. B., X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure* **2002**, *10* (6), 825-35.
59. Roujeinikova, A.; Simon, W. J.; Gilroy, J.; Rice, D. W.; Rafferty, J. B.; Slabas, A. R., Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates. *J Mol Biol* **2007**, *365* (1), 135-45.
60. Upadhyay, S. K.; Misra, A.; Srivastava, R.; Surolia, N.; Surolia, A.; Sundd, M., Structural insights into the acyl intermediates of the *Plasmodium falciparum* fatty acid synthesis pathway: the mechanism of expansion of the acyl carrier protein core. *The Journal of biological chemistry* **2009**, *284* (33), 22390-400.
61. Winter, J. M.; Cascio, D.; Dietrich, D.; Sato, M.; Watanabe, K.; Sawaya, M. R.; Vederas, J. C.; Tang, Y., Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *Journal of the American Chemical Society* **2015**, *137* (31), 9885-93.

62. Winter, J. M.; Sato, M.; Sugimoto, S.; Chiou, G.; Garg, N. K.; Tang, Y.; Watanabe, K., Identification and characterization of the chaetoviridin and chaetomugilin gene cluster in *Chaetomium globosum* reveal dual functions of an iterative highly-reducing polyketide synthase. *Journal of the American Chemical Society* **2012**, *134* (43), 17900-3.
63. Maier, T.; Jenni, S.; Ban, N., Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **2006**, *311* (5765), 1258-62.
64. Maier, T.; Leibundgut, M.; Ban, N., The crystal structure of a mammalian fatty acid synthase. *Science* **2008**, *321* (5894), 1315-22.
65. Kitamoto, T.; Nishigai, M.; Sasaki, T.; Ikai, A., Structure of fatty acid synthetase from the Harderian gland of guinea pig. Proteolytic dissection and electron microscopic studies. *J Mol Biol* **1988**, *203* (1), 183-95.
66. Tang, Y.; Kim, C. Y.; Mathews, II; Cane, D. E.; Khosla, C., The 2.7-Ångstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* **2006**, *103* (30), 11124-9.
67. Herbst, D. A.; Jakob, R. P.; Zahringer, F.; Maier, T., Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **2016**, *531* (7595), 533-7.
68. Armache, J. P.; Jarasch, A.; Anger, A. M.; Villa, E.; Becker, T.; Bhushan, S.; Jossinet, F.; Habeck, M.; Dindar, G.; Franckenberg, S.; Marquez, V.; Mielke, T.; Thomm, M.; Berninghausen, O.; Beatrix, B.; Soding, J.; Westhof, E.; Wilson, D. N.; Beckmann, R., Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc Natl Acad Sci U S A* **2010**, *107* (46), 19748-53.
69. Gilbert, R. J.; Fucini, P.; Connell, S.; Fuller, S. D.; Nierhaus, K. H.; Robinson, C. V.; Dobson, C. M.; Stuart, D. I., Three-dimensional structures of translating ribosomes by Cryo-EM. *Mol Cell* **2004**, *14* (1), 57-66.
70. Baker, M. L.; Zhang, J.; Ludtke, S. J.; Chiu, W., Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat Protoc* **2010**, *5* (10), 1697-708.
71. Sirohi, D.; Chen, Z.; Sun, L.; Klose, T.; Pierson, T. C.; Rossmann, M. G.; Kuhn, R. J., The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* **2016**, *352* (6284), 467-70.
72. Zhang, X.; Jin, L.; Fang, Q.; Hui, W. H.; Zhou, Z. H., 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **2010**, *141* (3), 472-82.
73. Campbell, M. G.; Cheng, A.; Brilot, A. F.; Moeller, A.; Lyumkis, D.; Veessler, D.; Pan, J.; Harrison, S. C.; Potter, C. S.; Carragher, B.; Grigorieff, N., Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **2012**, *20* (11), 1823-8.
74. Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A.; Cheng, Y., Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* **2013**, *10* (6), 584-90.

75. Henderson, R., Structural biology: Ion channel seen by electron microscopy. *Nature* **2013**, *504* (7478), 93-4.
76. Kuhlbrandt, W., Biochemistry. The resolution revolution. *Science* **2014**, *343* (6178), 1443-4.
77. Kuhlbrandt, W., Cryo-EM enters a new era. *Elife* **2014**, *3*, e03678.
78. Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structure of a modular polyketide synthase. *Nature* **2014**, *510* (7506), 512-7.
79. Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014**, *510* (7506), 560-+.

Chapter 2: Starter unit flexibility for engineered product synthesis by the non-reducing polyketide synthase PksA

This chapter was adapted with permission from Huitt-Roehl, C.R., Hill, E.A., Adams, M.M., Vagstad, A.L., Li, J.W. and Townsend, C.A., *ACS Chem Biol*, **2015**, *10*(6), 1443-9

2.1. Introduction

Fatty acid synthases (FASs) and all but one clan of the related polyketide synthases (PKSs) exemplify the rare catalytic strategy of iteration; that is, the catalytic domains of these large proteins are re-used, typically a fixed number of times, and can accommodate substrates as small as two carbons to as large as 20 or more¹⁻³. Thus, unlike the acute substrate discrimination associated with highly evolved primary metabolic enzymes, the potential synthetic hazards of substrate promiscuity are overcome by deeply buried active sites accessible to intermediates borne on carrier proteins and intraprotein transfers that confer high effective concentrations. Such intrinsic plasticity in iterative systems raises the fundamental question of whether these active sites can more readily accommodate alternative building blocks to synthesize non-native products. Previous studies of modular PKSs, where each module canonically catalyzes a single set of elongation and tailoring steps during biosynthesis, have shown limited tolerance to a narrow range of both non-native starter⁴⁻⁸ and extender

units⁹⁻¹⁰. The fungal non-reducing PKSs (NR-PKSs) mediate the simplest chain extension chemistry to a classical poly- β -ketone intermediate with no processing at the β -carbon during elongation, which distinguishes the other partially and highly reducing PKSs and all FASs¹¹⁻¹⁴.

NR-PKSs contain two domains unique among PKSs: a starter unit acyl-transferase (SAT) that brings the chain initiating component onto the enzyme¹⁵, and a product template (PT) domain that controls the cyclization of the fully elongated, highly reactive poly- β -ketone intermediate to specific aromatic and fused aromatic products^{3, 16-17}. Combinatorial assemblies of heterodomains from a selection of NR-PKSs have shown that polyketide chain length is largely controlled by the ketosynthase (KS) domain, and the cyclization pattern is determined by the PT domain¹⁸⁻¹⁹. The enzyme responsible for the biosynthesis of norsolorinic acid anthrone **2** in *Aspergillus parasiticus*, the precursor to aflatoxin B₁ **1**, PksA²⁰, accepts an unusual hexanoyl starter unit from a dedicated pair of yeast-like FASs (HexA/B, Figure 2.1a,b)²¹⁻²³. The SAT domain of most NR-PKSs prefers acetyl-CoA from common cellular metabolism and shows strong selection against even propionyl-CoA²⁴. However, there are similar enzyme pairs such as Hpm8 and Hpm3, where the highly-reducing PKS Hpm8 synthesizes a relatively large partially-reduced starter unit, which is then transferred to the NR-PKS Hpm3 for further elongation²⁵.

In this work we combine the SAT-KS-MAT (malonyl-CoA acyl-transferase) tridomain and the PT-ACP (acyl-carrier protein) didomain of PksA to examine the ability not only of the SAT to accept non-native starter units but also the KS and the PT to perform their central tasks in product synthesis as a more complete measure of both starter unit tolerance and NR-PKS synthetic function. To simplify the analysis, the thioesterase (TE) domain, which catalyzes a further (Claisen) cyclization to anthrone **2**, is not present and we rely on the facile self-enol lactonization of the PT-cyclized product **3** to release pyrone **4** or its starter unit-dependent structural variant from the ACP (Figure 2.1c).

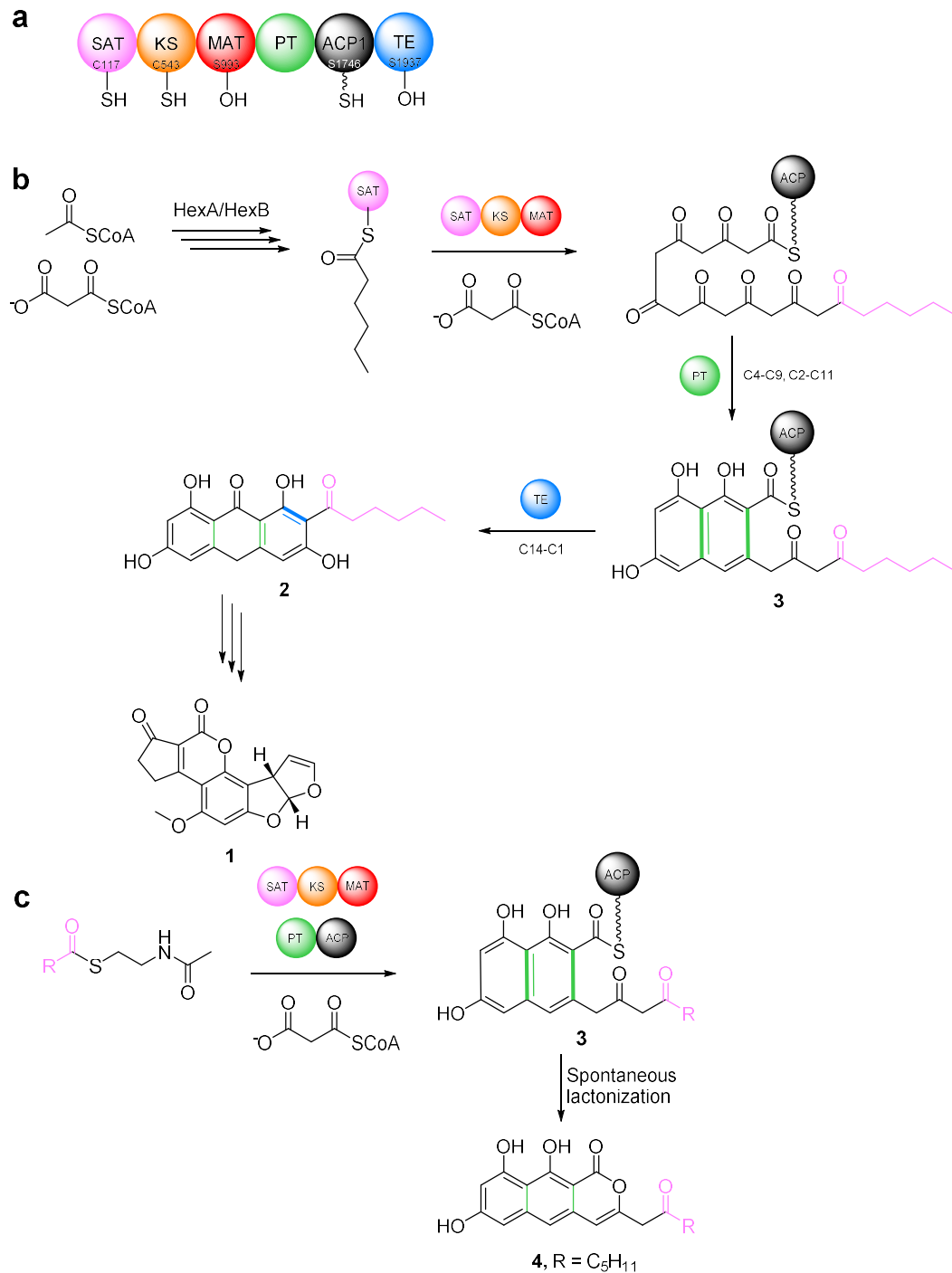


Figure 2.1. PksA biosynthetic scheme. (a) Domain organization of PksA. Active site residues are indicated. (b) PksA synthesizes the first intermediate in aflatoxin 1 biosynthesis – norsolorinic acid anthrone **2**. Bonds/moieties installed by each catalytic domain are color-coded and bolded for emphasis. (c) *In vitro* assay of native and alternative starter-units used in this study.

2.2. Results

2.2.1. Alternative starter unit *in vitro* assays.

The C-terminal His₆-tagged PksA SAT-KS-MAT tridomain and PT-ACP didomain were over-produced in *E. coli* and purified by Co²⁺-affinity chromatography as previously described¹⁸. Potential starter units were presented as *N*-acetylcysteamine (SNAC) thioesters (Figure 2.1c), which were prepared by coupling of *N*-acetylcysteamine to commercially-available or readily synthesized carboxylic acids (see Appendix B). Each substrate (1 mM) was assayed under uniform conditions in phosphate buffer at pH 7 containing glycerol (10%), malonyl-SNAC (1 mM), TCEP (1 mM), PksA SAT-KS-MAT and PT-ACP (10 μM), and incubated at room temperature for 4 hours. Products were extracted into ethyl acetate, separated by HPLC, and the UV-Vis spectrum of each was recorded (200-800 nm). Exact masses were determined by UPLC-ESI-MS. In all instances the naphthopyrone products gave a characteristic absorbance signature.

2.2.2. Linear acyl-SNAC starter units.

Early examination of PksA starter unit preferences of the SAT domain indicated modest selectivity for hexanoyl over butanoyl or octanoyl¹⁵. In a first set of experiments, systematic examination of the homologous series of linear C₄–C₈ SNAC thioesters (5–9) was undertaken in the synthesis of norpyrone 4 and its

structural variants. The non-native C₅ fatty acid thioester was incorporated with comparable efficiency to the native hexanoyl starter (Figure 2.2, Appendix Figures B.2, B.4), but reactions of C₄, C₇ and C₈ chain lengths were still evident. To facilitate comparison among the C₄–C₈ thioesters, the naphthopyrone product peak intensities were plotted relative to that of hexanoyl in Figure 2.2. Starter units longer than C₈ were not tested owing to their limited solubility under the assay conditions. Reactions of **6**, **8**, and **9** were less efficient in naphthopyrone product formation and correspondingly gave increased shunt and truncation products appearing at early retention times in the HPLC chromatogram (Appendix Figures B.3, B.5, B.6). Recent deconstruction and domain “swapping” or heterodomain combinatorial experiments have revealed tight control of overall polyketide chain length to largely \pm one ketide (C₂) extension, while the PT domain exerts strict programming of ring formation despite receiving chain lengths both shorter and longer than wild-type¹⁸⁻¹⁹. Reactions utilizing the C₈ starter unit **9** exemplify this tight chain length control, as one fewer ketide extension was performed to give a naphthopyrone product with the native C₂₀ chain length. Similarly, one extra ketide extension occurred in the reaction of the C₄ starter unit **6** to yield a C₂₀ product. In the case of the C₇ starter unit **8**, both C₂₁ and C₁₉ products were observed, whereas only the C₁₉ product was produced from the C₅ starter unit **7**. This overall chain-length constraint in PksA likely

amplifies the apparent narrow range of well-accepted linear starter units.

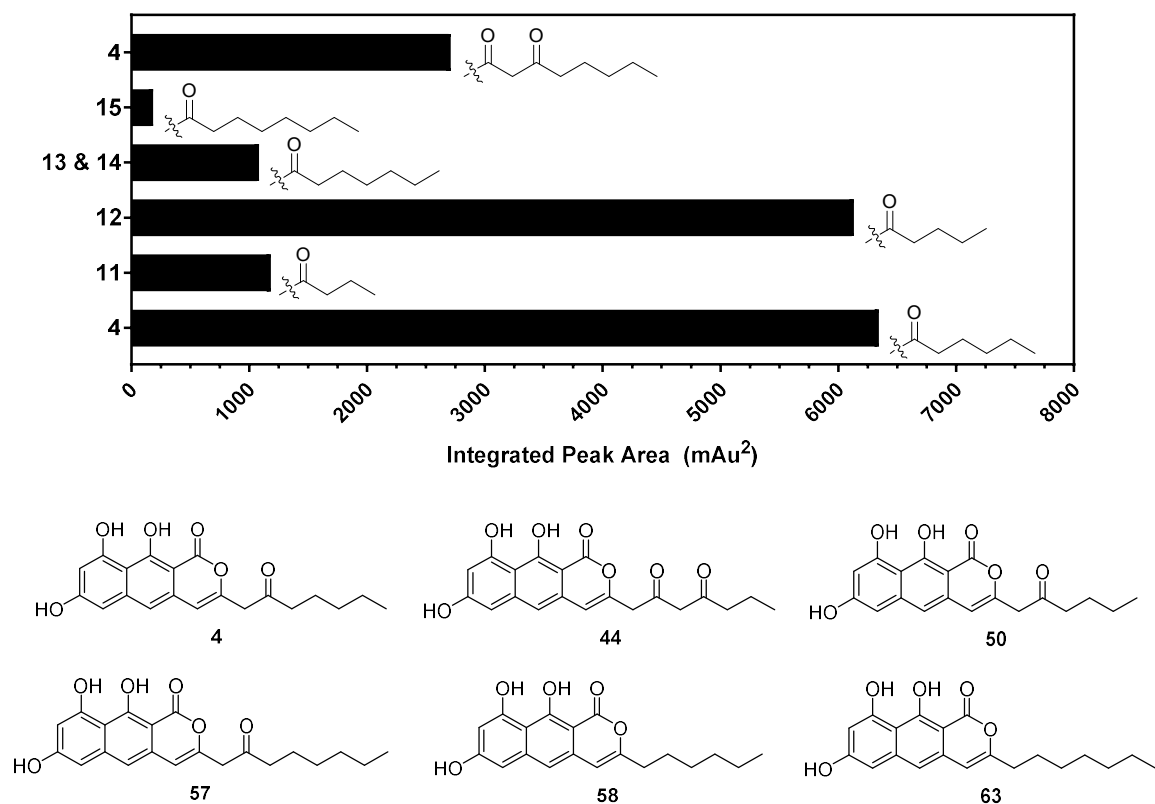


Figure 2.2. Relative production of on-target naphthopyrones from linear acyl-SNAC starter units. Products are indicated on the y-axis and starter units are indicated on the right side of each bar. In the case of starter unit 8, where two naphthopyrone products of different chain lengths were produced, the peak areas of both products have been combined for this chart.

To analyze the pattern of truncation and derailment products observed in all reactions, these off-target products have been categorized by basic core structures corresponding to known chromophores (Figure 2.3a, see Appendix Figure B.1 for full structures of all characterized products)^{19,26}. Control reactions containing malonyl-SNAC as the sole substrate produced primarily the naphthopyrone 42 (Figure 2.3b) in which decarboxylated malonyl-SNAC was

used as the starter unit. Monitoring the steady-state population of intermediates bound to the PksA ACP by mass spectrometry revealed acetyl in addition to the expected hexanoyl and malonyl²⁷. One can visualize that the appearance and inefficient utilization of an acetyl starter could arise by direct transfer from acetyl-SNAC, but more likely by transfer from malonyl-SNAC to the MAT domain and then to the ACP, followed by decarboxylation. In the case of the linear C₄–C₈ SNAC thioesters **5–9**, the primary derailment products observed were the tri- and tetra-ketide lactones having core structure C, while only small amounts of improperly cyclized but fully extended and cyclodehydrated SEK4 and SEK4b analogs were produced (cores D and E). Therefore, extension by the KS domain may be the limiting factor in productive biosynthesis with linear starter units shorter or longer than the native hexanoyl primer, rather than cyclization in the proper register by the PT domain. As expected, the product profile for the reaction of C₈ β-ketothioester **10** is similar to that of the native hexanoyl starter unit. While the observed products were nearly identical for the two reactions, efficiency was lower in the non-native case (Appendix Figure B.7). As this starter unit represents the first intermediate product of elongation, the reason for decreased productivity presumably owes to reduced binding efficiency to the SAT (or KS domain, see below) domain before transfer to the ACP.

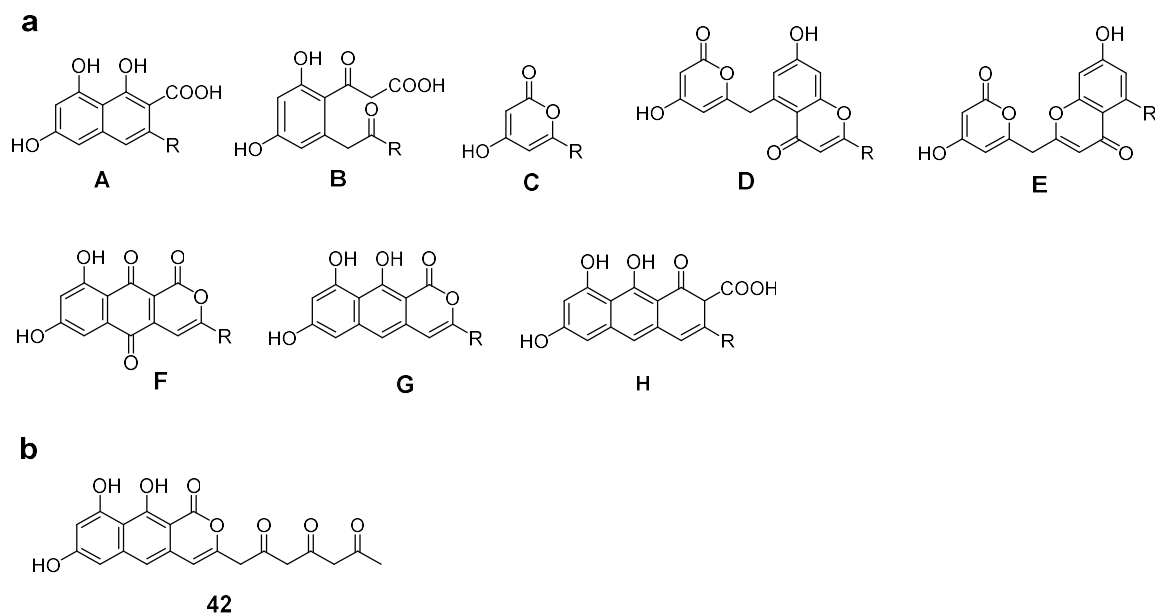


Figure 2.3. Core structures of observed derailment products. (a) Derailment product cores were determined based on exact mass and known chromophores. (b) Major product observed in control reactions containing malonyl-SNAC as the sole substrate.

2.2.3. Halogenated acyl-SNAC starter units.

Experiments conducted to evaluate the efficiency of naphthopyrone production by PksA utilizing substituted acyl-SNAC thioesters demonstrated that modifications at the distal end of the thioester were well tolerated by the enzyme, particularly halogenation and methylation (Figure 2.4). Starter units halogenated at the remote end of the acyl chain (**11–14**) were relatively well accepted, resulting in high product formation, albeit accompanied by significant amounts of derailment products, primarily the triacetic acid lactone analog (core C). The product profiles among the four tested halogenated starter units were similar, although relative and absolute amounts of on-target and derailment products varied depending on chain length and identity of the halogen

substituent. Similar to the linear, un-substituted starter units, reactions of the longer 7-atom halogenated substrates resulted in higher derailment production and lower naphthopyrone production. Additionally, brominated products were produced in smaller amounts than chlorinated products, presumably owing to the relative steric bulk of the halogen substituent (Figure 2.4, Appendix Figures B.8–B.11).

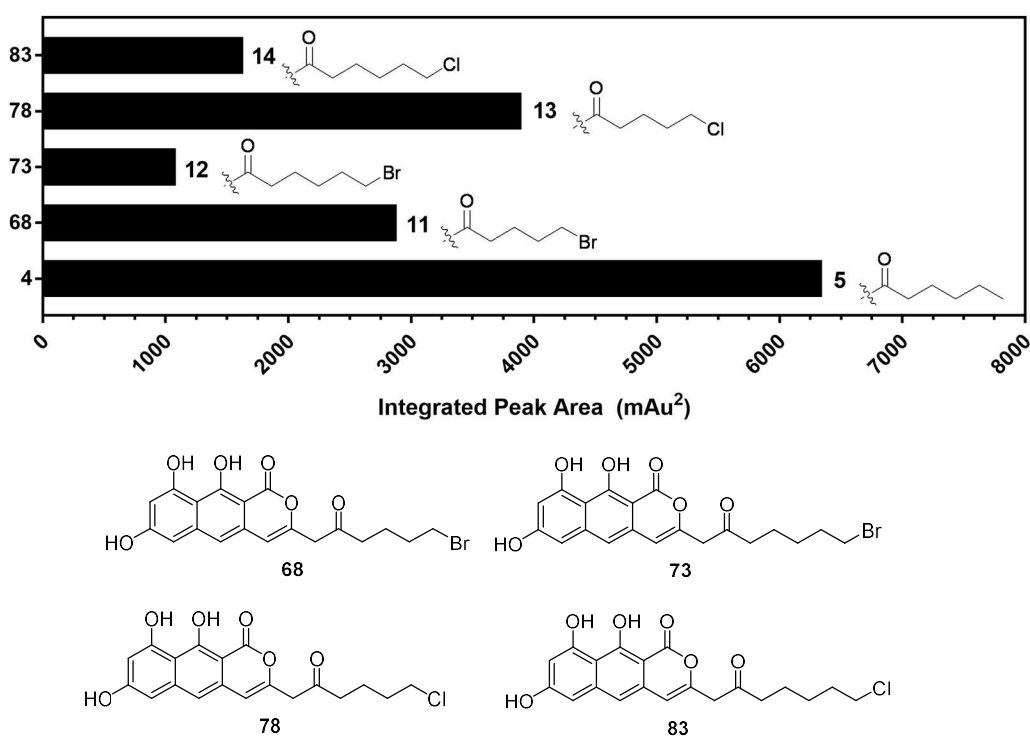


Figure 2.4. Relative production of on-target naphthopyrones from halogenated acyl-SNAC starter units. Products are indicated on the y-axis and starter units are indicated on the right side of each bar.

2.2.4. Methylated acyl-SNAC starter units.

Biosynthetic production and catalytic efficiency decrease with chain branching close to the thioester, as exemplified with reactions utilizing methylated starter units **15**, **16**, **17**, and **18** (Figure 2.5). Starter units with substituents at the 2- and 3- position were particularly disfavored, while starter units substituted at the 4- and 5- position were significantly better tolerated. In the case of 2-branched **15**, the major product was **42** (Figure 2.3b, Appendix Figure B.12), suggesting this starter unit is so poorly tolerated that inefficient priming and extension using decarboxylated malonyl-SNAC took place. Similarly, the reaction utilizing **16** also produced **42**, albeit in a smaller proportion. The fully extended but improperly cyclized SEK4/SEK4b analogs **86a,b** (cores D and E) were the major derailment products in reactions with **16**, suggesting the KS domain is significantly more tolerant of substitution at the C-3 position than the PT (Appendix Figure B.13). Although only a small difference in naphthopyrone production was observed between reactions utilizing **17** and **18** (Figure 2.5, Appendix Figures B.14, B.15), there was a marked difference in the ratio of naphthopyrone to derailment products. Thioester **17** gave a significant amount of the SEK4/SEK4b analogs **91a,b** (Appendix Figure B.14). While it was expected that a five-carbon long starter unit branched near its terminus, **19**, would give an outcome comparable to **17** based on the location of the methyl

group, production of naphthopyrone product was significantly lower (Figure 2.5, Appendix Figure B.16). These results indicate that branched starter units, especially if proximal to the thioester, can impede the ability of the PT domain to properly cyclize the mature polyketide chain.

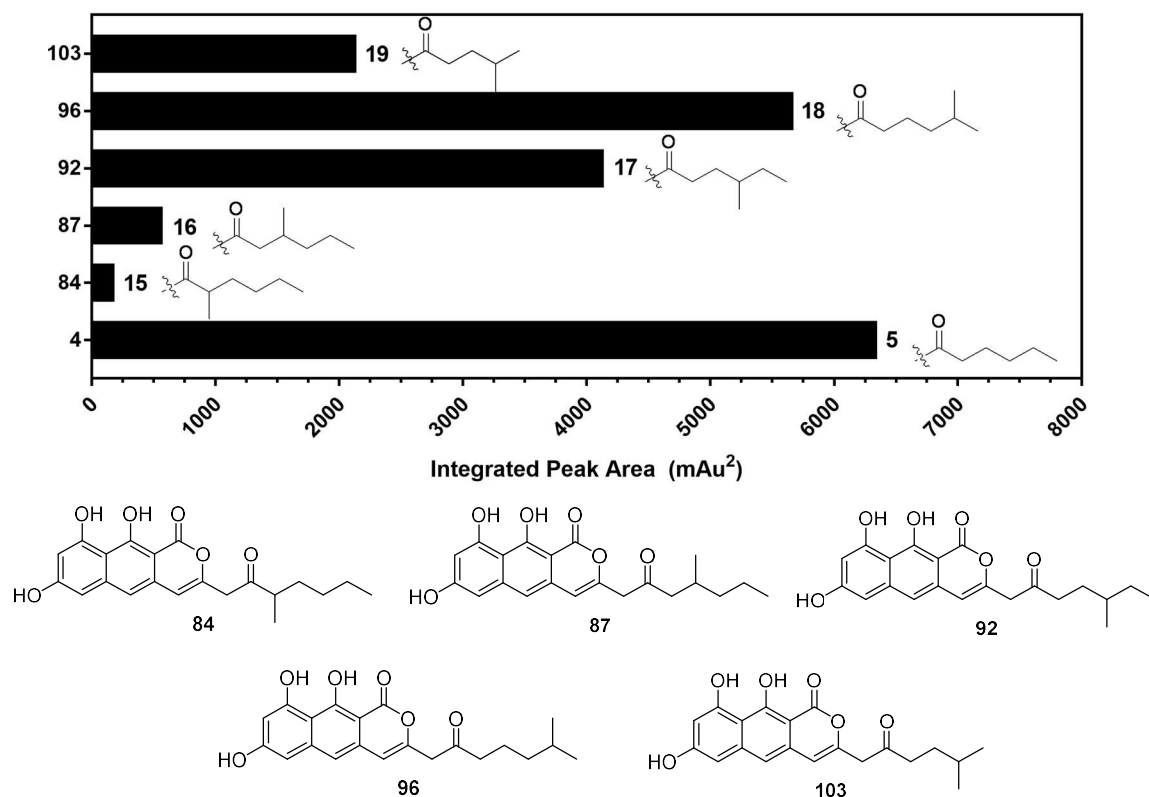


Figure 2.5. Relative production of on-target naphthopyrones from methylated acyl-SNAC starter units. Products are indicated on the y-axis and starter units are indicated on the right side of each bar.

2.2.5. Unsaturated acyl-SNAC starter units.

Analogous to the methyl-branched starter units, desaturation was relatively well tolerated at the alkyl terminus of the acyl chain (starter unit **22**), while the α,β -unsaturated SNAC thioester **20** reacted poorly (Figure 2.6). This

reduced reactivity may owe to the intrinsically greater stability of a conjugated thioester slowing each acyl transfer step, or enzyme inhibition by this electrophilic species. Three unsaturated starter units were tested for incorporation into naphthopyrone products. The reaction utilizing **22** resulted in moderate production of the on-target final product and few derailment products (Figure 2.6, Appendix Figure B.19). Conversely, reactions of **20** produced in the main the spontaneously cyclized SEK4/SEK4b analogs **104a,b** (Appendix Figure B.17). No significant peaks representing truncation products were observed in this reaction, suggesting the α,β -alkene was problematic for processing by the PT domain, but not the KS domain. Unexpectedly, production of the naphthopyrone in the reaction of **21** was approximately equal to that of **20**, but a more diverse profile of derailment products was observed with **20** (Figure 2.6, Appendix Figure B.18). These included both improperly or partially cyclized, but fully extended products as well as truncation products. As was the case of the methyl-branched starter units, the relative rigidity afforded by the alkene interferes more with proper biosynthetic processing when located closer to the thioester.

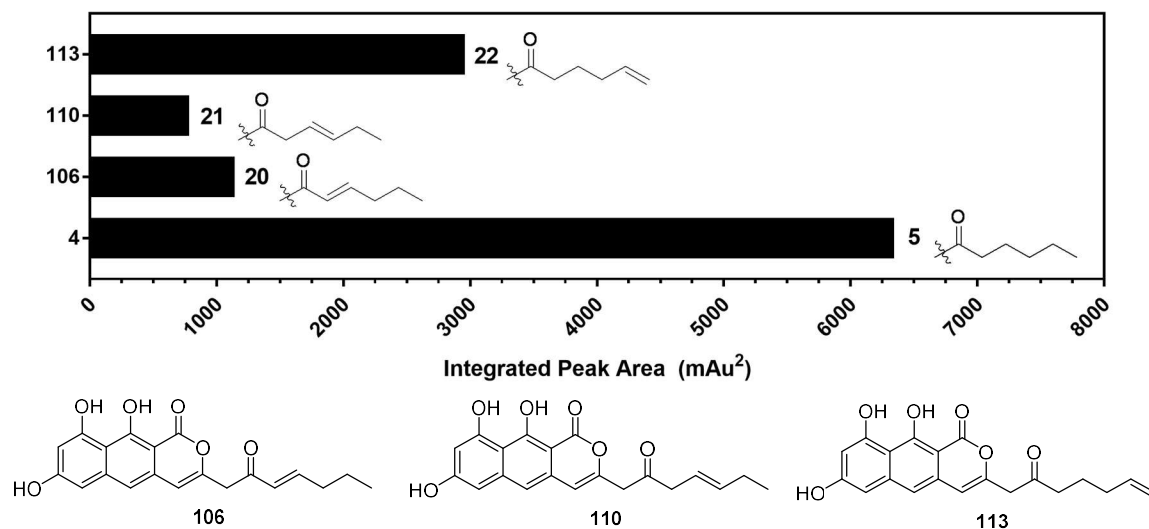


Figure 2.6. Relative production of on-target naphthopyrones from unsaturated acyl-SNAC starter units. Products are indicated on the y-axis and starter units are indicated on the right side of each bar.

2.2.6. Heteroatom-substituted acyl-SNAC starter units.

PksA is capable of incorporating oxyether- and thioether-containing starter units into naphthopyrone products (*e.g.* **23–28**), but significant decreases in production and efficiency were observed (Figure 2.7). Reactions utilizing methylthioether **23** resulted in product formation comparable to the analogous six-carbon starter unit **5**, with only a slight increase in derailment products (Figure 2.7, Appendix Figure B.20). The major derailment product observed for this reaction was the triketide lactone **114** (core C), while no significant peak intensities correlating to the SEK4/SEK4b analogs were observed. Processing of **23** into the naphthopyrone product **116** (Figure 2.7, Appendix Figure B.20) is, therefore, as efficient as that of the native starter unit. Reactions of five- and

four-atom thioethers (**24**, **25**, **26**) resulted in lower product formation than the comparable carbon-only starter units (Figure 2.7, Appendix Figures B.21-B.23). This effect is particularly evident with starter unit **26**, which produced significantly less product than its analogous carbon-only starter unit **6** (Figures 2.2, 2.7, Appendix Figure B.23). Derailment product types C, D and E were prevalent in these reactions, suggesting interference with the activity of both the KS and the PT domains. As in the case of starter unit **6**, reactions utilizing four-atom thioethers underwent one extra extension to maintain the native C₂₀ chain length, providing further evidence for tight chain length control by the KS domain. Additionally, an eight-atom oxyether starter unit **27** produced a C₂₀ chain length product, analogous to the C₈ alkane starter unit **9**. Surprisingly, production of the naphthopyrone product **129** containing this starter unit was significantly greater than that from starter unit **9**, with correspondingly lower levels of derailment products (Figures 2.2, 2.7, Appendix Figure B.24). The seven-atom oxyether starter unit **28** produced much less naphthopyrone product than **27**, and increased amounts of the spontaneously cyclized derailment product **130a,b** (cores D and E) (Appendix Figure B.25).

Starter units amenable to downstream coupling strategies, such as those containing terminal cyano- and azido- groups (*e.g.* **29** and **30**), were accepted by the enzyme, albeit with relatively low targeted product formation. In both cases,

significant amounts of triketide and tetraketide truncation products were produced (Appendix Figures B.26, B.27). In the case of starter unit **29**, fully extended but improperly cyclized products **134a,b** (Appendix Figure B.26) were observed as well.

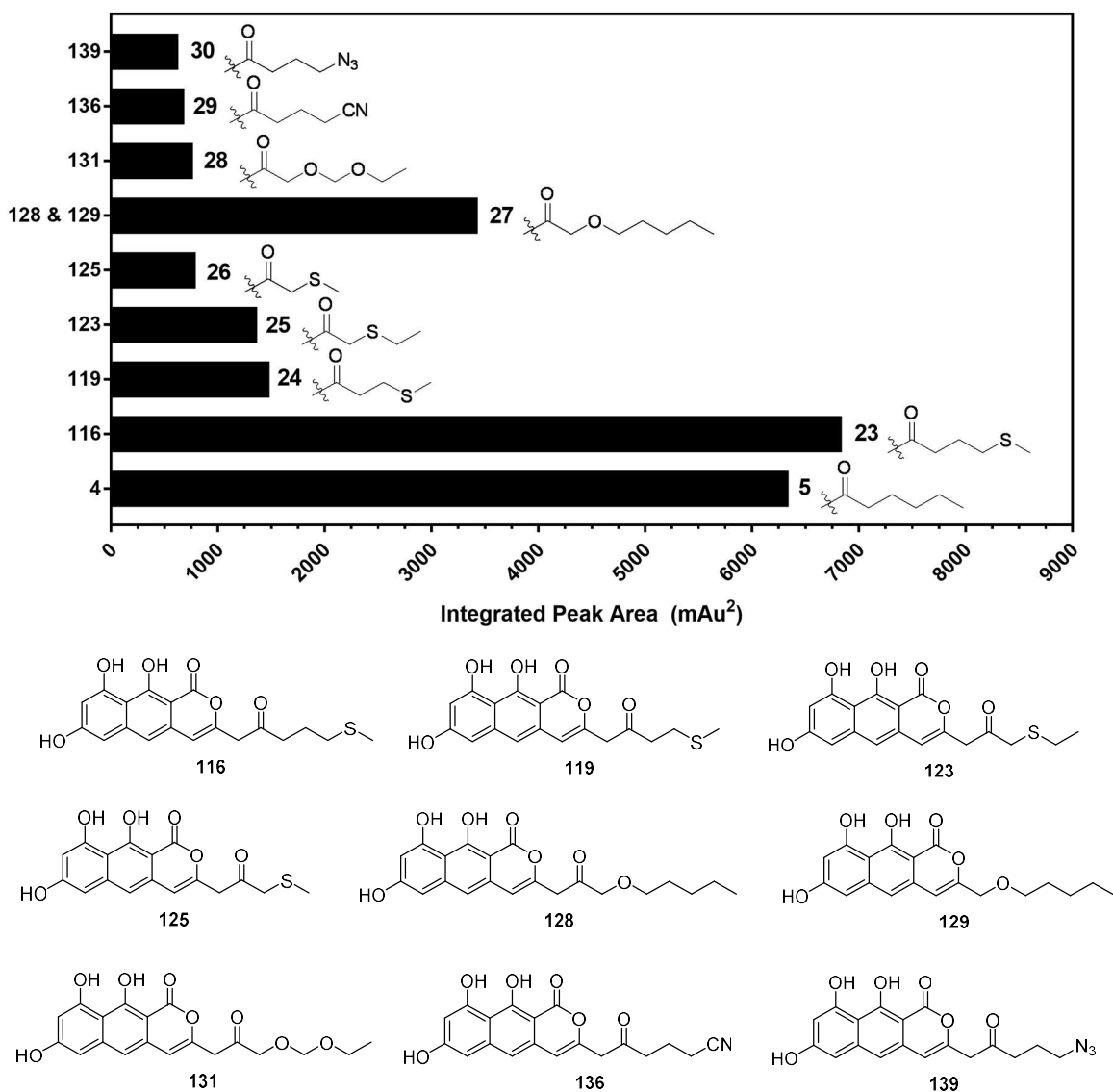


Figure 2.7. Relative production of on-target naphthopyrones from heteroatom-substituted acyl-SNAC starter units. Products are indicated on the y-axis and starter units are indicated on the right side of each bar. In the case of starter unit **27**, where two naphthopyrone products of different chain lengths were produced, the peak areas of both products have been combined for this chart.

2.2.7. Sterically constrained acyl-SNAC starter units.

To test the tolerance of PksA to bulkier, more rigid substrates, SNAC-thioesters containing cyclopropyl groups (**31**, **32**), an alkyne (**33**) and a benzyl group (**34**) were assayed. In the case of **31** and **32**, a small amount of naphthopyrone product containing the cyclopropyl group was detected in both reactions (Figure 2.8). Production with these starter units was extremely low, and all domains of the enzyme presumably had difficulty accommodating the cyclopropyl group within their active sites. Compound **42** was the dominant product of these reactions, suggesting decarboxylated malonyl-SNAC is a more favorable primer for these reactions than **31** and **32** (Appendix Figures B.28, B.29). This product was also observed in other reactions where production is poor, including starter units **6** and **15**, as noted above, and 2-butynoyl-SNAC (**33**) (Appendix Figure B.30). Control reactions where malonyl-SNAC alone was provided to the enzyme in the absence of a candidate acyl-SNAC starter unit gave similar results (Appendix Figure B.32). Conversely, benzoyl-SNAC reactions did not produce naphthopyrone product (Appendix Figure B.31). However, a very small amount of the triketide lactone (core C) was observed in the reaction of benzoyl-SNAC, suggesting some tolerance of this starter unit by the KS domain.

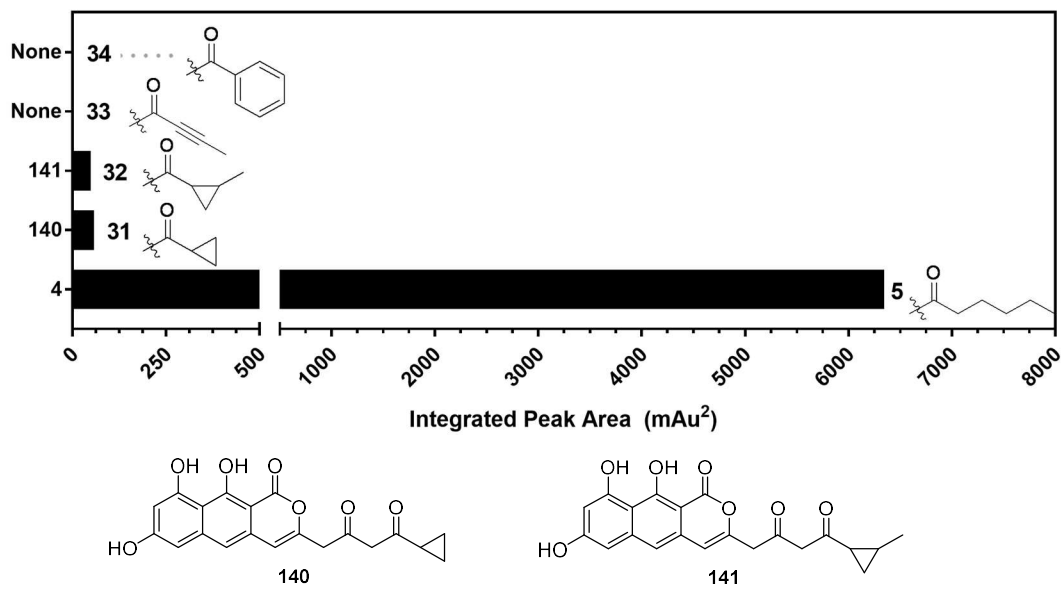


Figure 2.8. Relative production of on-target naphthopyrones from sterically constrained acyl-SNAC starter units. Products are indicated on the y-axis and starter units are indicated on the right side of each bar. The x-axis has been split for clarity due to exceptionally low production of 140 and 141.

2.2.8. SAT active site mutants.

A series of SAT active site (Cys117) mutants was tested for product formation to further evaluate the role of the SAT domain in starter unit selection. Mutants were initially selected to mimic the other two naturally-occurring SAT active site motifs (GXSXG as in *Giberella zeae* Pks13 and GXGXXG as in *Aspergillus terreus* ACAS)²⁸⁻²⁹ as well as another non-nucleophilic mutant (C117A). Against expectation, presumably inactive C117A formed comparable levels of product to wild-type enzyme, as depicted in Figure 2.9. Furthermore, the C117G mutant produced notably more product than wild-type. Reactions with C117S, however, resulted in extremely low product formation. This pattern was consistent among

all starter units tested and owes to hydrolysis of SNAC thioesters before inefficient acyl transfer to the ACP can take place (Figure 2.10).

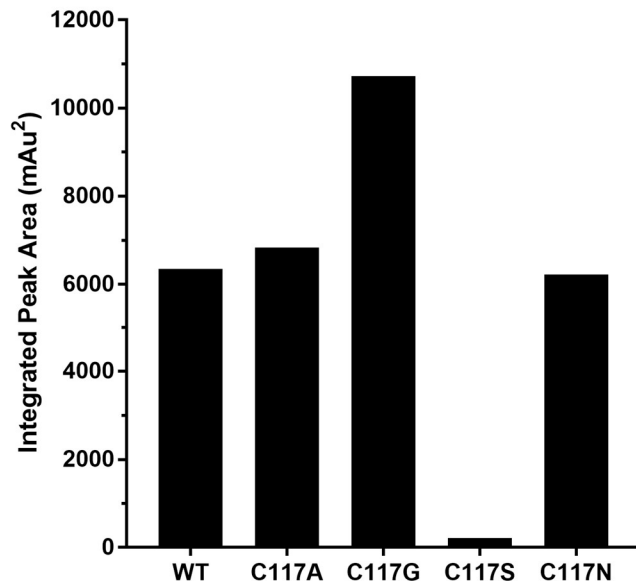


Figure 2.9. Comparison of hexyl-naphthopyrone (**4**) production by PksA-SAT active site mutants.

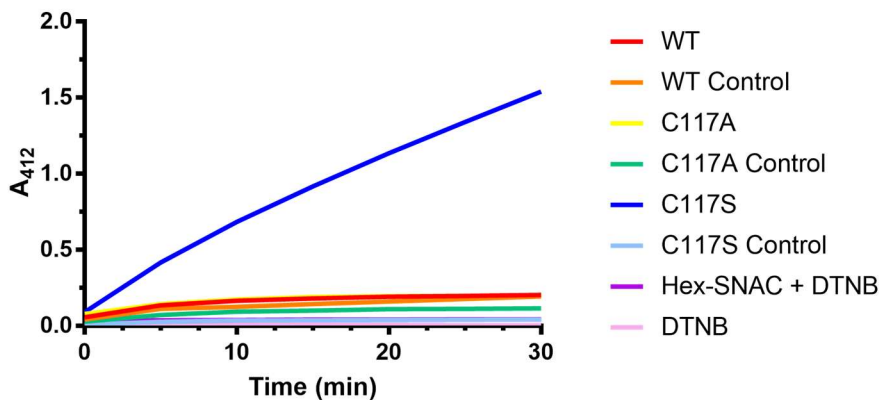


Figure 2.10. Ellman's reagent (DTNB) assay for detection of SNAC hydrolysis by PksA SAT active site mutants. PksA SAT C117S hydrolyzes hexanoyl-SNAC to the free thioester (detected by Ellman's reagent at 412 nm). Control reactions lack hexanoyl-SNAC substrate.

Given the largely unaltered reaction profiles with C117A and C117G variants, other SAT active site mutants were constructed to probe whether non-covalent binding was responsible for increased production in C117A and C117G

mutants. C117L, C117Q and C117N mutations were attempted in an effort to partially occlude the active site. Soluble SAT-C117N-KS-MAT was obtained and evaluated for naphthopyrone product formation. Production by this mutant was comparable to that of the wild-type enzyme and the C117A mutant, suggesting any non-covalent binding to the active site likely plays a small role, if any (Figure 2.9). It is worth noting, however, that production from C117G was still higher than the other variants, consistent with the possibility that non-covalent recognition and transfer may yet play a role in bringing starters onto the enzyme.

2.3. Discussion

PksA can accept and utilize a circumscribed selection of acyl-SNAC substrates as starter units to support naphthopyrone biosynthesis. The enzyme is tolerant to minor modifications in chain length and constitution of the thioester starter unit, especially at the distal end of the acyl chain, including substituents that allow subsequent modification of the resulting tricyclic product; for example “Click” chemistry³⁰. However, significant losses in production efficiency are observed with longer chain lengths or substitutions proximal to the thioester.

Recent experimental advances have allowed production of full-length NR-PKSs, albeit in much lower yields than smaller multi-domain constructs¹⁹. As a control to measure the synthetic penalty, if any, exacted by dissection into a 2-

part combination of SAT-KS-MAT and PT-ACP tri- and didomains, the corresponding fully intact 5-domain construct of PksA was prepared. Owing to low protein yields, *in vitro* reactions were performed with 1 μ M PksA SAT-KS-MAT-PT-ACP, but under otherwise identical conditions to the previous experiments. Three starter units (hexanoyl-SNAC [5], pentanoyl-SNAC [7], and benzoyl-SNAC [34]) were selected to compare production from both efficient and inefficient substrates. For the former two efficient substrates, presuming reasonably that product formation tracks linearly with pentadomain enzyme concentration, production by the two-part combination lagged the intact construct by only a factor of 1.5 in the case of 5 and 1.2 in the case of 7 (Figure 2.11). As seen above in experiments with the two-part system, no naphthopyrone product containing the benzyl ring was observed in reactions of 34. This finding was in complete accord with similar comparisons of intact *vs.* domain combinations determined recently, and confirms the validity of the deconstruction method for the purposes of the experiments described herein¹⁹.

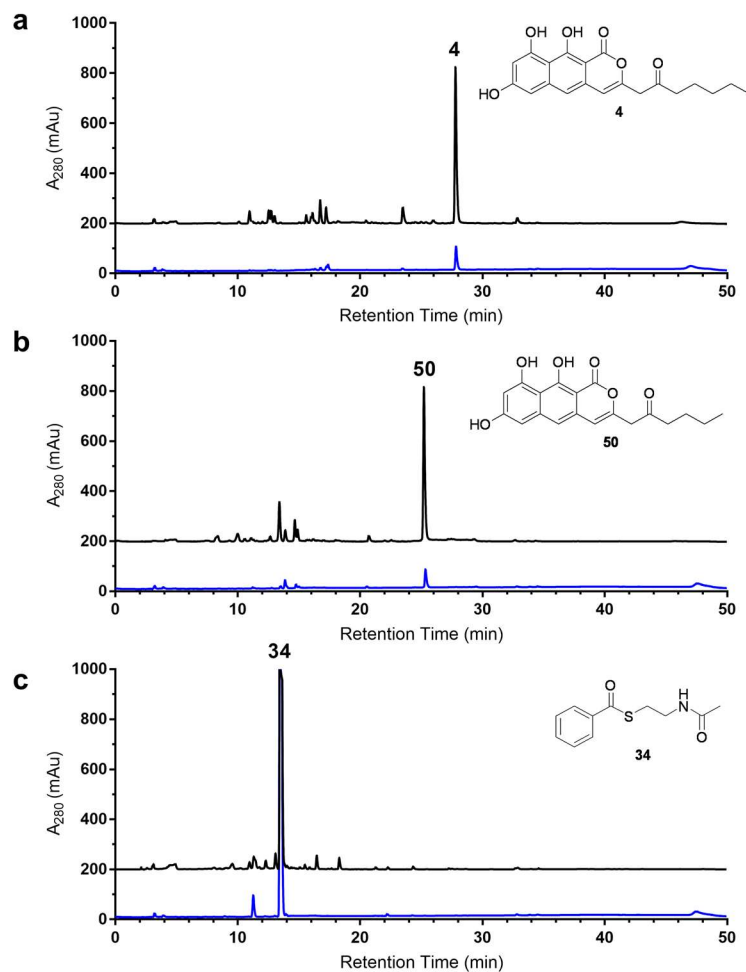


Figure 2.11. Comparison of intact pentadomain (blue) and dissected two-part (black) reactions. (a) Reaction of starter unit **5** (*S*-(2-acetamidoethyl) hexanethioate). (b) Reaction of starter unit **7** (*S*-(2-acetamidoethyl) pentanethioate). (c) Reaction of starter unit **34** (*S*-(2-acetamidoethyl) benzothioate).

The SAT active site mutants tested in this study gave unanticipated results. While some native SAT domains contain a glycine substitution of the active site cysteine, these have previously been presumed to be inactive and compensated for by the activity(ies) of another domain(s). Our experiments, however, suggest that at sufficiently high concentrations of starter unit, the KS can recognize the SNAC thioester as an adequate mimic of the normal ACP-

delivered starter unit to be loaded directly. Additionally, the possibility remains that even in the absence of a covalent bond between the SAT domain and a starter unit, non-covalent binding may increase effective concentrations sufficiently for efficient biosynthesis. Although an active SAT domain is required for starter unit transfer within the HexA/HexB•PksA complex²², under the super-physiological conditions utilized in this *in vitro* study, the SAT appears to be bypassed by some starter units and directly loaded on to the KS. Such relatively facile utilization of non-native starter units is fortuitous but useful from the perspective of synthetic biology and rationally-directed biosynthesis.

Unlike other SAT domains that scavenge acetyl-CoA from their cellular environment, PksA receives its hexanoyl starter unit directly in complex with HexA/B and thus may have less pressure to be discriminating in its starter unit selection. Analogous observations have been made in the case of the HR-PKS•NR-PKS pair Hpm3 and Hpm8, where a SNAC mimic of the starter unit can rescue production in the absence of an active SAT domain²⁵. Based on the limited data from these two examples, it can be anticipated that when SAT domains exist in FAS•NR-PKS or HR-PKS•NR-PKS complexes, their primary function is interprotein transfer of a relatively large and/or complex starter unit and they are likely to be additionally flexible in the modified primers they accept and/or can be bypassed to directly load the KS. Diversity of production is then

determined by the tolerance of the KS and PT to provide full polyketide chain elongation and correct cyclization register. Recently, an X-ray structure of a SAT domain has been elucidated, which may facilitate to mutagenesis experiments to remodel a SAT active site with the goal of efficient uptake of altered starter units and transfer to sustain directed biosynthesis in a physiological context³¹.

2.4. Experimental Methods

2.4.1. Cloning

Cloning details for all constructs used in this study are provided in Appendix B. All plasmids used are summarized in Appendix Table B.1, and primers are given in Appendix Table B.2.

2.4.2. Protein Expression and Purification

All proteins were heterologously expressed in *E. coli* BL21(DE3) and isolated as described previously²⁷, with the exception that Co²⁺-TALON resin (GoldBio) was used in place of Ni²⁺-NTA resin. Expression of soluble pentadomain PksA SAT-KS-MAT-PT-ACP was achieved by an additional one-hour cold shock in ice water prior to induction with 1 mM IPTG. Following purification, proteins were dialyzed against assay buffer (100 mM potassium phosphate pH 7, 10% glycerol) and optionally frozen at -80 °C prior to use. If

necessary, proteins were concentrated using 10K Molecular Weight Amicon Ultra Centrifugation Filters (Millipore). Protein concentration was measured in triplicate using the Bradford assay (Bio-Rad) and bovine serum albumin as a standard (New England Biolabs).

2.4.3. *In vitro* assays

Purified SAT-KS-MAT tridomain (10 μ M) and PT-ACP didomain (10 μ M) were combined with 1 mM malonyl-SNAC in assay buffer (100 mM potassium phosphate, pH 7.0, 10% glycerol, 1 mM TCEP). Reactions were initiated by addition of 1 mM acyl-SNAC starter unit and allowed to run for 4 h at room temperature. The 500 μ L reactions were quenched with 10 μ L hydrochloric acid and extracted three times with 500 μ L ethyl acetate. Extractions were pooled and evaporated to dryness using a Speed-Vac. Enzymatic products were resuspended in 300 μ L 40% aqueous acetonitrile for HPLC analysis.

2.4.4. HPLC and UPLC-ESI-MS analysis of enzymatic products

100 μ L injections of enzymatic products were analyzed over a Prodigy ODS3 analytical column (4.6 x 250 mm, 5 μ , Phenomenex) by a gradient HPLC method using an Agilent 1200 instrument equipped with an autosampler.

Bisolute method: linear gradient of 20% A/80% B to 90% A over 35 min, hold

90% A for 10 min, followed by re-equilibration to 20% A over 5 minutes with a 10 minute hold, where solvent A is acetonitrile + 0.1% formic acid and solvent B is 0.1% formic acid. UV-vis spectra were recorded over a range of 200-800 nm, and chromatograms recorded at 280 nm. High-resolution mass data were obtained by UPLC-ESI-MS on a Waters Acquity/Xeno-G2 in positive ion mode.

2.4.5. Synthesis of acyl-SNAC substrates

Malonyl-SNAC was produced using the malonyl-CoA synthetase MatB from *Rhizobium leguminosarum* and HPLC purified as described previously^{26, 32}. The synthesis of all acyl-SNAC starter units used in this study is described in Appendix B.

2.4.6. Ellman's reagent assay for acyl-SNAC hydrolysis

Ellman's reagent (5,5-dithio-bis-(2-nitrobenzoic acid), DNTB) was used to detect the free thiol of SNAC released by SAT-catalyzed hydrolysis. 5 μ M SAT, 0.2 mM DTNB, and 1 mM hexanoyl-SNAC were combined in 100 mM potassium phosphate buffer, pH 7. Absorbance at 412 nm was monitored for 1 h at 25 °C.

2.5. References

1. Cronan, J. E.; Thomas, J., Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods Enzymol* **2009**, *459*, 395-433.
2. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A., Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **2007**, *24* (1), 162-90.
3. Crawford, J. M.; Townsend, C. A., New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* **2010**, *8* (12), 879-89.
4. Kirschning, A.; Hahn, F., Merging chemical synthesis and biosynthesis: a new chapter in the total synthesis of natural products and natural product libraries. *Angew Chem Int Ed Engl* **2012**, *51* (17), 4012-22.
5. Moore, B. S.; Hertweck, C., Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat Prod Rep* **2002**, *19* (1), 70-99.
6. Goss, R. J.; Shankar, S.; Fayad, A. A., The generation of "unnatural" products: synthetic biology meets synthetic chemistry. *Nat Prod Rep* **2012**, *29* (8), 870-89.
7. Weissman, K. J., Mutasynthesis - uniting chemistry and genetics for drug discovery. *Trends Biotechnol* **2007**, *25* (4), 139-42.
8. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **2001**, *18* (4), 380-416.
9. Bravo-Rodriguez, K.; Ismail-Ali, A. F.; Klopries, S.; Kushnir, S.; Ismail, S.; Fansa, E. K.; Wittinghofer, A.; Schulz, F.; Sanchez-Garcia, E., Predicted incorporation of non-native substrates by a polyketide synthase yields bioactive natural product derivatives. *Chembiochem* **2014**, *15* (13), 1991-7.
10. Walker, M. C.; Thuronyi, B. W.; Charkoudian, L. K.; Lowry, B.; Khosla, C.; Chang, M. C., Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science* **2013**, *341* (6150), 1089-94.
11. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **2009**, *48* (26), 4688-716.
12. Udvary, D. W.; Merski, M.; Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *Journal of Molecular Biology* **2002**, *323* (3), 585-598.
13. Schumann, J.; Hertweck, C., Advances in cloning, functional analysis and heterologous expression of fungal polyketide synthase genes. *J Biotechnol* **2006**, *124* (4), 690-703.
14. Cox, R. J., Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org Biomol Chem* **2007**, *5* (13), 2010-26.
15. Crawford, J. M.; Dancy, B. C.; Hill, E. A.; Udvary, D. W.; Townsend, C. A., Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc Natl Acad Sci U S A* **2006**, *103* (45), 16728-33.
16. Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpheh, O.; Tsai, S. C.; Townsend, C. A., Structural basis for biosynthetic

- programming of fungal aromatic polyketide cyclization. *Nature* **2009**, *461* (7267), 1139-43.
17. Li, Y.; Image, II; Xu, W.; Image, I.; Tang, Y.; Image, I., Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J Biol Chem* **2010**, *285* (30), 22764-73.
 18. Vagstad, A. L.; Newman, A. G.; Storm, P. A.; Belecki, K.; Crawford, J. M.; Townsend, C. A., Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angew Chem Int Ed Engl* **2013**, *52* (6), 1718-21.
 19. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J Am Chem Soc* **2014**, *136* (20), 7348-62.
 20. Chang, P. K.; Cary, J. W.; Yu, J.; Bhatnagar, D.; Cleveland, T. E., The *Aspergillus parasiticus* polyketide synthase gene *pksA*, a homolog of *Aspergillus nidulans* *wA*, is required for aflatoxin B1 biosynthesis. *Mol Gen Genet* **1995**, *248* (3), 270-7.
 21. Watanabe, C. M.; Townsend, C. A., Initial characterization of a type I fatty acid synthase and polyketide synthase multienzyme complex NorS in the biosynthesis of aflatoxin B(1). *Chem Biol* **2002**, *9* (9), 981-8.
 22. Foulke-Abel, J.; Townsend, C. A., Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase. *Chembiochem* **2012**, *13* (13), 1880-4.
 23. Ma, Y.; Smith, L. H.; Cox, R. J.; Beltran-Alvarez, P.; Arthur, C. J.; Simpson, F. R. S. T., Catalytic relationships between type I and type II iterative polyketide synthases: The *Aspergillus parasiticus* norsolorinic acid synthase. *Chembiochem* **2006**, *7* (12), 1951-8.
 24. Crawford, J. M.; Vagstad, A. L.; Ehrlich, K. C.; Townsend, C. A., Starter unit specificity directs genome mining of polyketide synthase pathways in fungi. *Bioorg Chem* **2008**, *36* (1), 16-22.
 25. Zhou, H.; Qiao, K.; Gao, Z.; Meehan, M. J.; Li, J. W.; Zhao, X.; Dorrestein, P. C.; Vederas, J. C.; Tang, Y., Enzymatic synthesis of resorcylic acid lactones by cooperation of fungal iterative polyketide synthases involved in hypothemycin biosynthesis. *J Am Chem Soc* **2010**, *132* (13), 4530-1.
 26. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J Am Chem Soc* **2012**, *134* (15), 6865-77.
 27. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008**, *320* (5873), 243-6.
 28. Zhou, H.; Zhan, J.; Watanabe, K.; Xie, X.; Tang, Y., A polyketide macrolactone synthase from the filamentous fungus *Gibberella zeae*. *Proc Natl Acad Sci U S A* **2008**, *105* (17), 6249-54.

29. Awakawa, T.; Yokota, K.; Funa, N.; Doi, F.; Mori, N.; Watanabe, H.; Horinouchi, S., Physically discrete beta-lactamase-type thioesterase catalyzes product release in atrochryson synthesis by iterative type I polyketide synthase. *Chem Biol* **2009**, *16* (6), 613-23.
30. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **2001**, *40* (11), 2004-2021.
31. Winter, J. M.; Cascio, D.; Dietrich, D.; Sato, M.; Watanabe, K.; Sawaya, M. R.; Vederas, J. C.; Tang, Y., Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, *137* (31), 9885-93.
32. An, J. H.; Kim, Y. S., A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in *Rhizobium trifolii*--cloning, sequencing, and expression of the enzymes in *Escherichia coli*. *Eur J Biochem* **1998**, *257* (2), 395-402.

Chapter 3: The structural organization of substrate loading in iterative polyketide synthases

This chapter was adapted with permission from Herbst, D.A., Huitt-Roehl, C.R., Jakob R.P., Kravetz, J.M., Storm, P.A., Alley, J.R., Townsend, C.A., and Maier, T. *Nat Chem Biol*, submitted.

3.1. Introduction

Polyketide synthases (PKS) are large, multi-domain enzymes involved in the biosynthesis of a wide variety of pharmaceutically and agriculturally important natural products¹. Because of the value of many polyketide-based natural products, including lovastatin, erythromycin, and rapamycin, polyketide biosynthesis has been studied in great detail over the past several decades². Of particular interest to the Townsend lab are the fungal type I iterative non-reducing polyketide synthases (NR-PKS)³. In contrast to modular PKS (modPKS), in which multiple sets of catalytic domains, called modules, act sequentially in an assembly-line fashion⁴, the iterative PKSs (iPKSs) consist of a single set of domains. This single polypeptide iteratively catalyzes substrate loading and elongation, with each domain catalyzing multiple reactions to form the polyketide. iPKSs can be most closely compared to mammalian fatty acid synthase (FAS), which also acts in an iterative fashion to condense two carbon units into an acyl chain⁵. In both cases, a ketosynthase (KS) domain catalyzes a

series of decarboxylative Claisen condensations to homologate 2-carbon units, provided as malonyl-CoA by an acyltransferase (AT, termed the malonyl-CoA AT (MAT) in iPKS). All intermediates are covalently tethered to the enzyme through the phosphopantetheine arm of an acyl-carrier protein (ACP), which is post-translationally installed⁶. In FAS, the polyketide is fully reduced after each extension by the reductive domains within the modifying region: a ketoreductase (KR), dehydratase (DH), enoyl-reductase (ER). However, NR-PKSs lack any reductive domains, instead the reactive, poly- β -ketone intermediate is regio-specifically cyclized by a product template (PT) domain⁷. In addition to the PT, NR-PKSs contain another unique domain, the starter unit acyl-transferase (SAT), which selects an acyl-thioester starter unit to initiate biosynthesis⁸⁻¹⁰. Finally, in both FAS and PKS, the thioesterase (TE) releases the fatty acid or cyclized polyketide, most commonly through hydrolysis¹¹.

These multidomain enzymes are incredibly challenging to study. Classical mechanistic enzymology cannot be employed, as PKSs contain multiple active sites, catalyze multiple discrete chemical reactions, and act intramolecularly on covalently-tethered substrates. Furthermore, due to their large size and inherent structural flexibility, they are difficult to express in traditional heterologous hosts (e.g. *E. coli*) as intact constructs. In order to circumvent these limitations, much of the biochemical and structural work to

date on PKS has been accomplished by utilizing a dissection approach, where domains are excised at the genetic level and expressed as smaller constructs consisting of 1-3 domains¹². The primary hurdle in successful dissection of PKS domains is determination of *N*- and *C*-terminal cut sites that will yield stable, active protein. Early dissection attempts relied on limited proteolysis, where the relatively unstructured linker regions between domains are more readily proteolyzed¹³. The cut sites of the resulting fragments can then be identified by *N*-terminal sequencing¹⁴. More recently, linker regions were approximated by multiple sequence alignments, as their sequences are less conserved than those of the catalytic domains. Ultimately, an algorithm was developed by Udwy and Merski to accurately predict linker regions based on sequence conservation, secondary structure prediction, and hydrophobicity¹⁵.

This domain dissection approach has allowed the functions of individual PKS domains to be deduced by *in vitro* reconstitution. Furthermore, the structures of isolated monodomains have been solved for nearly all NR-PKS domains (SAT¹⁶, PT⁷, ACP¹⁷, and TE¹⁸). Although structures of KS-AT didomains from modPKSs have been determined¹⁹⁻²⁰, NR-PKS KS and MAT have never been successfully expressed as excised mono- or di-domains, despite extensive efforts. While isolated monodomain structures provide valuable insight into the mechanism of individual domains, the question remains how the domains are

integrated into the larger multienzyme, and how they coordinate their actions for the complex programming required for polyketide synthesis.

The X-ray crystal structure of animal FAS was first reported at 4.5 Å resolution in 2006²¹, and refined to 3.2 Å in 2008²². Animal FAS forms a gingerbread man shaped dimer, with the 'arms' and 'legs' connected by a flexible linker region. The condensing region (KS-AT) dimer, mediated by the KS, forms the 'legs', while the modifying region (DH-ER-KR-ACP-TE) dimer, mediated by the DH, forms the 'arms'. As the domain order and catalytic mechanisms are very similar, this breakthrough allowed the development of models of PKS architecture²³. Structures of isolated PKS domains, such as the DEBS KS-AT domain¹⁹, resemble their FAS counterparts, further supporting animal FAS as a structural model for type I PKS.

Recently, advances in cryo-electron microscopy (EM) have pushed the capabilities of the technique closer to atomic resolution. Cryo-EM, as well as another solution-based technique, small-angle X-ray scattering (SAXS), are well suited to elucidation of large PKS structures. The inherent structural flexibility of PKSs has thus far precluded successful crystallization, prohibiting structure determination by X-ray crystallography. However, cryo-EM and SAXS can accommodate more structural heterogeneity, and can be used to visualize enzymes in distinct conformational states²⁴⁻²⁵. Previously, high-resolution cryo-

EM structures were limited to large, comparatively rigid macromolecular complexes, such as the ribosome²⁶⁻²⁷. However, development of direct-electron detectors has allowed for high-resolution structures of smaller enzymes by limiting noise and blurring in image collection²⁸⁻²⁹. Recently, these advances allowed for the determination of a cryo-EM reconstruction of an entire PKS module, PikAIII³⁰⁻³¹.

The inherent flexibility of NR-PKSs makes crystallization or even high-resolution structural elucidation by cryo-EM very challenging. The resolution of cryo-EM is limited by the “biological resolution” of the molecule, including its inherent dynamics and conformational heterogeneity. To limit structural flexibility, we utilized a mechanism-based crosslinking approach to lock the enzyme into a single conformation, as well as to observe specific ACP-client domain interactions³²⁻³⁵. Using a selectively reactive substrate mimic, the ACP can be covalently crosslinked to the active site of select client domains, without off-target reactions. This technique has been used previously to observe the interaction between an AT and an ACP involved in vicenistatin biosynthesis at 2.3 Å resolution³⁶. Similarly, a mechanism-based crosslinker was used to interrogate residues involved in PT-ACP binding in an NR-PKS³⁷.

CTB1 is the NR-PKS involved in the first step of cercosporin **1** biosynthesis in the fungal plant pathogens *Cercospora* spp., which cause

hundreds of millions of dollars in agricultural loss each year³⁸⁻³⁹. Cercosporin is a perylenequinone natural product that facilitates *Cercospora* infection by acting as a photosensitizer, producing reactive oxygen species upon activation with visual light⁴⁰⁻⁴². The domain architecture and chemistry of CTB1 is prototypical of a NR-PKS, as described above, with the exception of an additional ACP domain and unusual TE activity (Figure 3.1a). The CTB1 TE catalyzes lactonization, rather than the more typical hydrolysis, to release the pyrone *nor*-toralactone **2** (Figure 3.1b)⁴³.

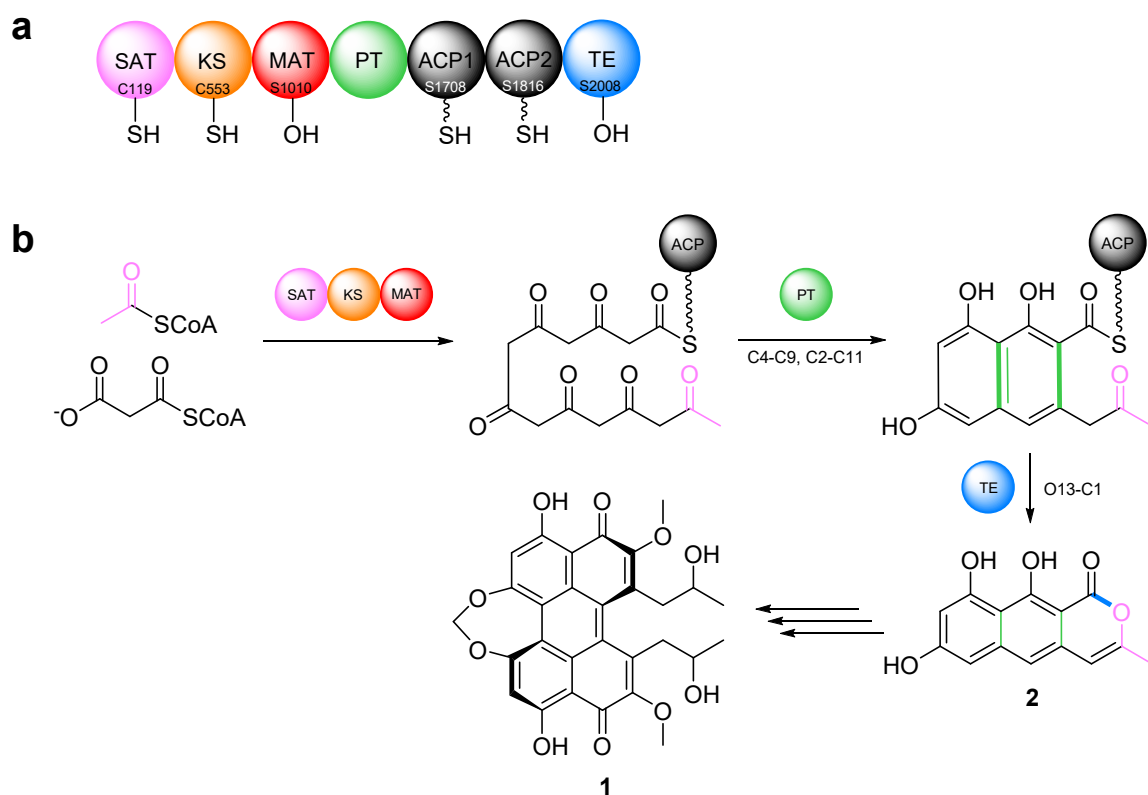


Figure 3.1. CTB1 biosynthesis. (a) Domain organization of CTB1. Active site residues are indicated within each domain. Wavy bonds on ACPs indicate the post-translationally installed phosphopantetheine arm. (b) Synthesis of *nor*-toralactone by CTB1. Bonds formed by specific domains are color coded and bolded for emphasis.

3.2. Results

3.2.1. Mechanism-based crosslinking of CTB1

In this study, we chose to crosslink the CTB1 SAT-KS-MAT to its cognate ACP at the KS active site. The SAT-KS-MAT, or loading and condensing region, is catalytically active as the isolated tridomain in cognate and non-native reactions⁴⁴⁻⁴⁵. ACP2 was selected over ACP1 as it shares a higher sequence identity to PksA ACP (40% for ACP2, 30% for ACP1), for which a solution structure has been determined¹⁷. Additionally, the ACP1 monodomain shows poor expression and stability, even following cut-site optimization. ACP1 and ACP2 were shown to be biosynthetically equivalent *in vitro* in minimal PKS reactions (SAT-KS-MAT + ACP) (Figure 3.2). Both ACPs produce identical product profiles, and although ACP1's overall production is slightly lower (likely due to protein instability), there is no observed cooperativity when the two ACPs are combined.

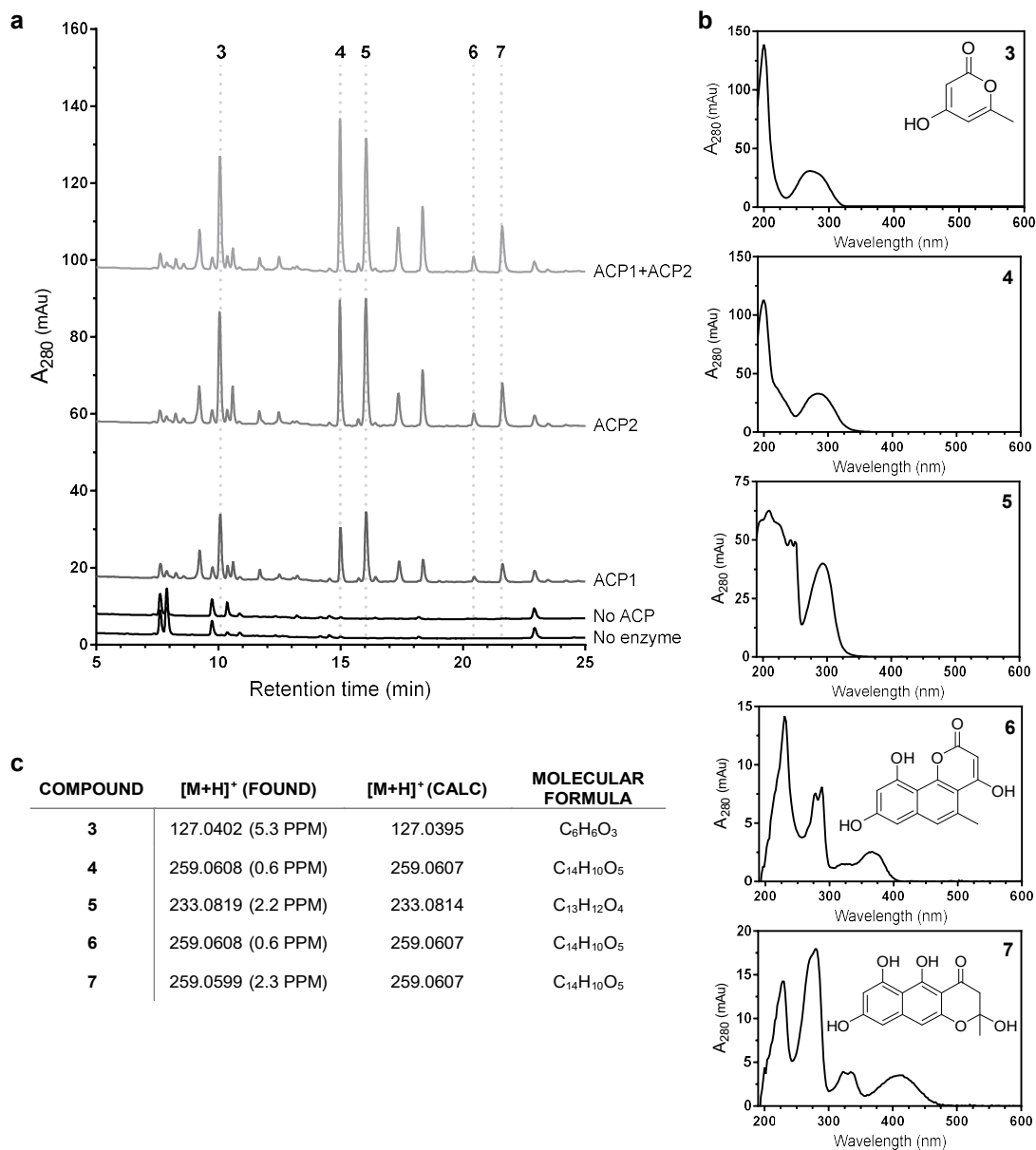


Figure 3.2. Comparison of biosynthetic capabilities of CTB1 ACP1 and ACP2. (a) HPLC chromatograms (280 nm) of product profiles of minimal CTB1 reactions with ACP1 (10 μ M), ACP2 (10 μ M), or both (5 μ M each). (b) UV-vis spectra of select derailment products. When possible, products were characterized on the basis of exact mass and UV-vis chromophore. (c) Exact masses of select derailment products determined by UPLC-ESI-MS.

To covalently attach CTB1 ACP2 to its client domain(s), a mechanism-based crosslinker mimicking the native acetyl-phosphopantetheine substrate was

prepared. In the native system, an acetyl starter unit is selected from the pool of primary metabolites as acetyl-CoA by CTB1 SAT and loaded onto the post-translationally installed phosphopantetheine arm of the ACP (*holo*-ACP) (Figure 3.3a). For crosslinking, we prepared an α -bromoacetyl aminodephospho-CoA substrate mimic **8**, where the native thioester has been substituted by an amide for stability, and an α -bromo group has been installed to facilitate irreversible attack by the active site cysteine, covalently linking the ACP to its client domain. Unfortunately, the acetyl substrate mimic proved to be too reactive, likely due to the primary halogen, and could not be loaded onto CTB1 ACP2 (Figure 3.3b). A less reactive α -bromopropionyl aminodephospho-CoA **9** was used in place of **8**, as CTB1 SAT has been shown to transfer propionyl groups, albeit at a lower rate than acetyl (Figure 3.3c)¹⁰.

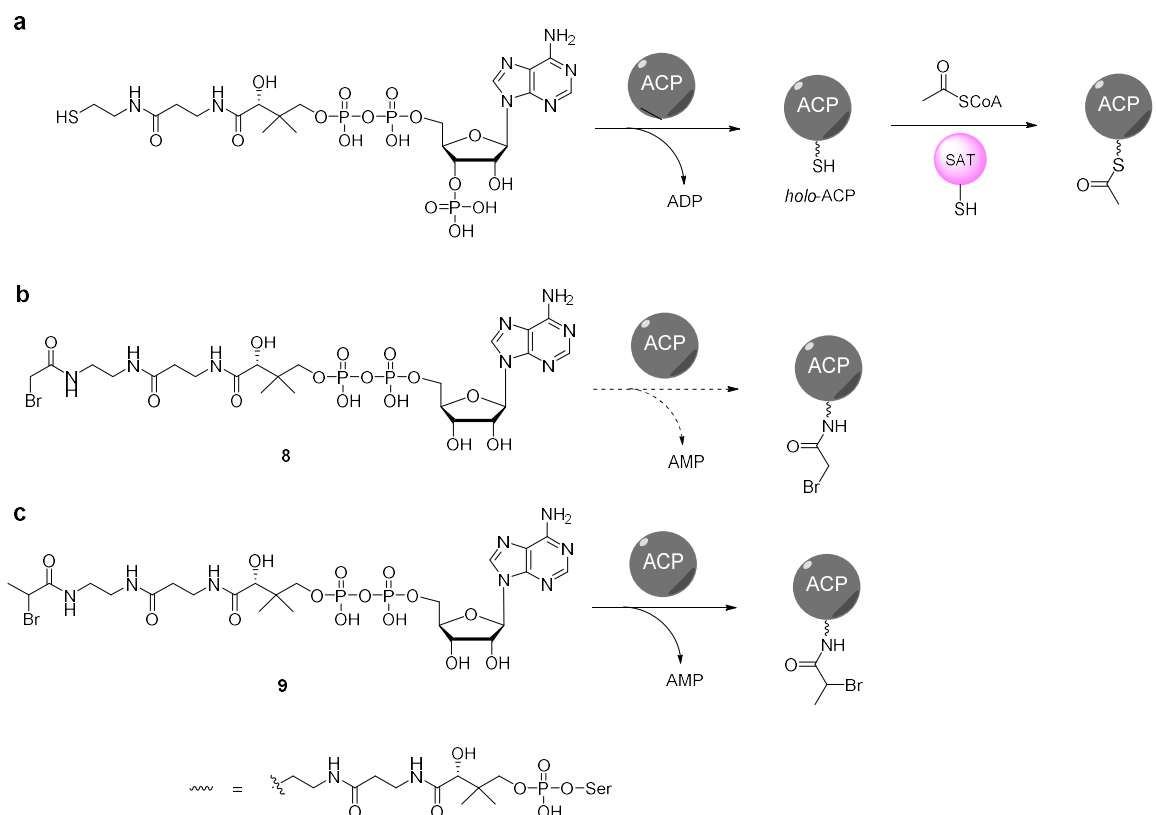


Figure 3.3. Loading of ACP with its native acetyl-CoA substrate (a), an α -bromo-acetyl substrate mimic (b) and an α -bromo-propionyl substrate mimic (c).

All crosslinkers were prepared synthetically to the acyl aminopantetheine stage **10**, followed by enzymatic transformation to the dephospho-CoA **9**, by way of an aminophosphopantetheine intermediate **11** (Figure 3.4). This enzymatic transformation is accomplished by two enzymes involved in CoA biosynthesis: pantetheine kinase (PanK)⁴⁶ and phosphopantetheine adenylyltransferase (PPAT)⁴⁷. A promiscuous phosphopantetheinyl transferase involved in surfactin biosynthesis (Sfp) is then used to transfer the crosslinker onto *apo* ACP⁴⁸. The α -bromopropionyl crosslinker **9** was successfully loaded

onto *apo* CTB1 ACP2, however, a significant amount of optimization of the process was required. Intermediates, particularly the phosphopantetheine mimic **11**, were difficult to isolate. High-performance liquid chromatography (HPLC) purification of **11** resulted in <10% recovery, likely because the phosphates cannot be fully protonated under typical HPLC conditions (0.1% TFA). Therefore, a one-pot enzymatic process was developed. First, the synthetically prepared α -bromopropionyl aminopantetheine **10** was phosphorylated with PanK. Once this reaction went to completion (as determined by UPLC-MS), PPAT, Sfp, and CTB1 ACP2 were added to adenylate the phosphopantetheine mimic **11** and load it onto the ACP. Phosphorylation of the 3' position of the ribose ring, to fully mimic native CoA, was eliminated. The required enzyme (dephospho-CoA kinase, DPCK) was found to be difficult to work with, and the ACP loading reaction with Sfp does not require the 3' phosphate.

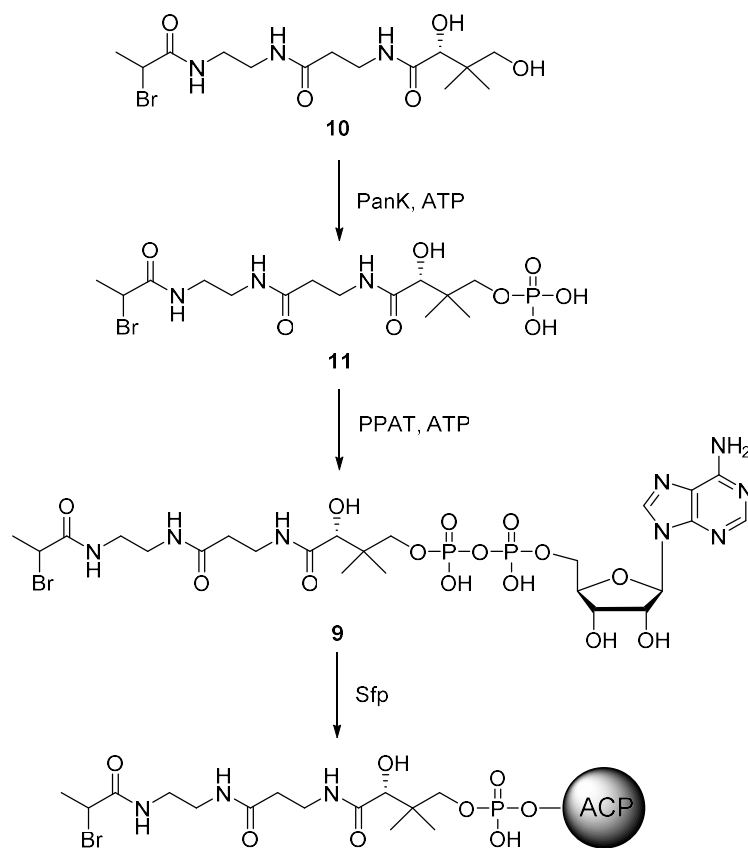


Figure 3.4. Enzymatic preparation of crosslinker-loaded ACP.

Crosslinker-loaded CTB1 ACP2 was successfully crosslinked to two of its client domains within the loading and condensing region: SAT and KS. However, no crosslinking to the MAT was observed, likely due to the lesser nucleophilicity of the active site serine in comparison to cysteine. In order to ensure selective crosslinking to a single client domain, double active-site mutants were prepared (C119A-S1010A (SAT[°]-KS-MAT[°]) for KS crosslinking, and H688A-S1010A (SAT-KS[°]-MAT[°]) for SAT crosslinking). The cysteine to alanine

mutation in the KS yielded insoluble protein, so alanine and more conservative glutamine mutations to the other two members of the catalytic triad (H688 and H733) were attempted. Of these mutations, one (H688A) was found to block crosslinking to the KS domain. Selective crosslinking with this mutant was confirmed by attempted crosslinking to the C119A-H688A-S1010A mutant, which was completely blocked.

Crosslinking to the SAT domain is markedly slower than to the KS domain, likely due to both the inherent reactivity of the domains, and the non-native propionyl substrate mimic. Consequently, preparation of 100% crosslinked sample was more straight forward in the case of KS crosslinking, and the KS crosslinked sample (SAT°-KS-MAT°=ACP2, ° denotes crosslinking) was chosen to move forward with structural elucidation. To achieve 100% crosslinking, extended (16 h) treatment of the client domains with reductant and a large excess of ACP2 was required. When CTB1 SAT-KS-MAT is in its native dimeric state (in the presence of 250 mM NaCl), 50% crosslinking is favored (1:2 ACP:SKM stoichiometry) (Figure 3.5).

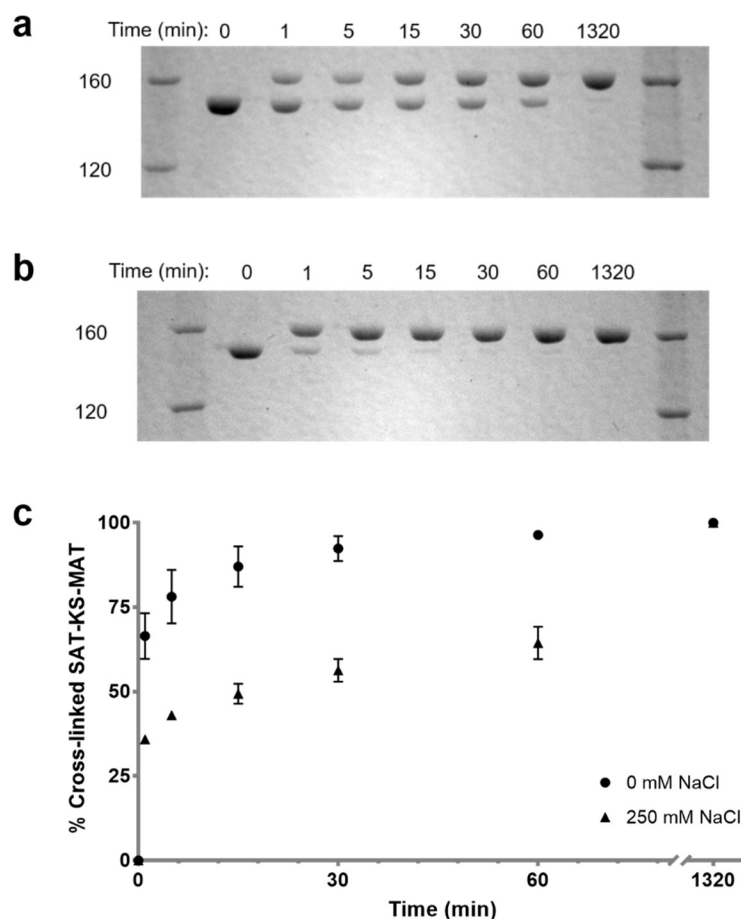


Figure 3.5. Crosslinking time course. Crosslinking of ACP2 to SAT^o-KS-MAT^o in presence (a) or absence (b) of 250 mM NaCl. (c) Quantification of extent of crosslinking by densitometry measurements. Data presented as mean \pm s.e.m.

Selective crosslinking to the KS active site, rather than non-specific crosslinking to a surface residue, was confirmed by a competition assay with radiolabeled acetyl-CoA. CTB1 SAT^o-KS-MAT^o was incubated with crosslinker-loaded ACP (or α -bromopropionyl aminophosphopantetheine **11**) for one hour prior to treatment with [1-¹⁴C]-acetyl-CoA. Pre-incubation with **11** or crosslinker-loaded ACP blocked radiolabeling of the protein, suggesting specific crosslinking to the KS active site (Figure 3.6). Additionally, only a single crosslinking event

was observed by SDS-PAGE (Figure 3.5), confirming a single ACP covalently crosslinked specifically to the KS active site of each SAT-KS-MAT monomer.

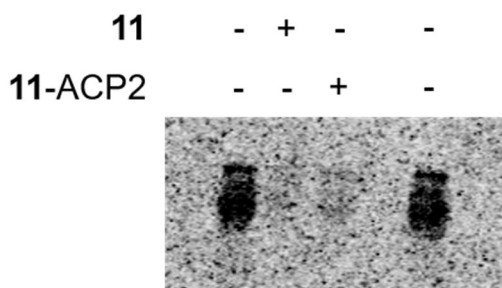


Figure 3.6. [^{14}C]-acetyl-CoA competition assay. Incubation with crosslinker **11** or crosslinker-loaded ACP (indicated above gel) blocked labelling of CTB1 SAT $^{\circ}$ -KS-MAT $^{\circ}$ with [^{14}C]-acetyl-CoA.

To obtain a more complete picture of the loading and condensing region's interaction with ACP2, CTB1 SAT-KS $^{\circ}$ -MAT $^{\circ}$ crosslinked at the SAT domain was prepared for structural elucidation. Based on the prior observation of preferred single-site ACP binding, we sought to prepare a sample in which only a single ACP2 is bound to each SAT-KS-MAT dimer. To accomplish this task, crosslinking was performed in the absence of salt, at a slightly lower concentration of ACP2, for one hour. At this point, approximately 50% of the SAT-KS-MAT monomer had been crosslinked (as determined by SDS-PAGE), and 250 mM NaCl was added to the reaction to significantly slow the rate of crosslinking. Rather than the multi-step purification performed on the KS-crosslinked sample, a streamlined, one step purification was utilized for the SAT-crosslinked sample, since removal of free SAT-KS-MAT was unnecessary (see Methods for details). The crosslinking reaction was purified by size-exclusion

chromatography, removing all free ACP2 or other contaminants. SDS-PAGE of the final crosslinked protein following purification confirms the desired stoichiometry of 1:2 ACP2:SAT-KS-MAT. This optimization of the preparation of crosslinked material increased final protein yield from 4% to 50%.

3.2.2. Structure of CTB1 SAT-KS-MAT

The structure of CTB1 SAT-KS-MAT was solved by X-ray crystallography to 2.8 Å (Figure 3.7, Appendix Table C.1). The dimeric protein is compact and of rhomboid shape. Dimerization is primarily mediated by the KS domains, each of which is connected to its MAT domain through a small linker domain (LD) (Figure 3.7a). The architecture of the KS-LD-MAT domains, as well as their individual structures, is consistent with related structures solved by X-ray crystallography: mammalian FAS (mFAS)^{22, 49}, fully-reducing iPKS⁵⁰, and several modPKSs^{19-20, 51}. However, the arrangement of these domains is markedly different in the structure of the PikAIII modular PKS, determined by cryo-EM at intermediate resolution³⁰⁻³¹.

The SAT resembles other ATs, containing an α/β hydrolase core fold and a ferredoxin-like subdomain, between which lies a cleft containing the active site. The SAT domains form a pseudo-dimer, arranged at 35° relative to the KS-MAT dimer (Figure 3.7b), where the direct interface of 324 Å² is hydrophobic, lacking any hydrogen bonds or salt bridges (Appendix Table C.2). However,

interactions with surrounding domains considerably stabilize this pseudo-dimer. The 26 amino acid (aa) linker between the C-terminus of the SAT and N-terminus of the KS binds a surface groove on the opposite SAT, adding 493 Å² to the interface (Figure 3.7c). While 9 aa of this linker is disordered and thus not visualized in the crystal structure, its connectivity is unambiguous based on relative distances to the resolved portions. An additional 957 Å² interface is formed between the SAT and the MAT of the opposite protomer, containing at least two salt bridges and five hydrogen bonds. The SAT also interfaces with the opposite KS domain (184 Å²), however, it does not contact the opposite LD. Notably, the SAT domain is located within the cleft formed by the KS and MAT of the opposite protomer, in a position analogous to proposed interactions of upstream ACPs with KSs in modPKSs (Figure 3.7d)⁵². The entrance to each SAT active site points into this cleft, at a distance of 56 Å from the opposite KS active site.

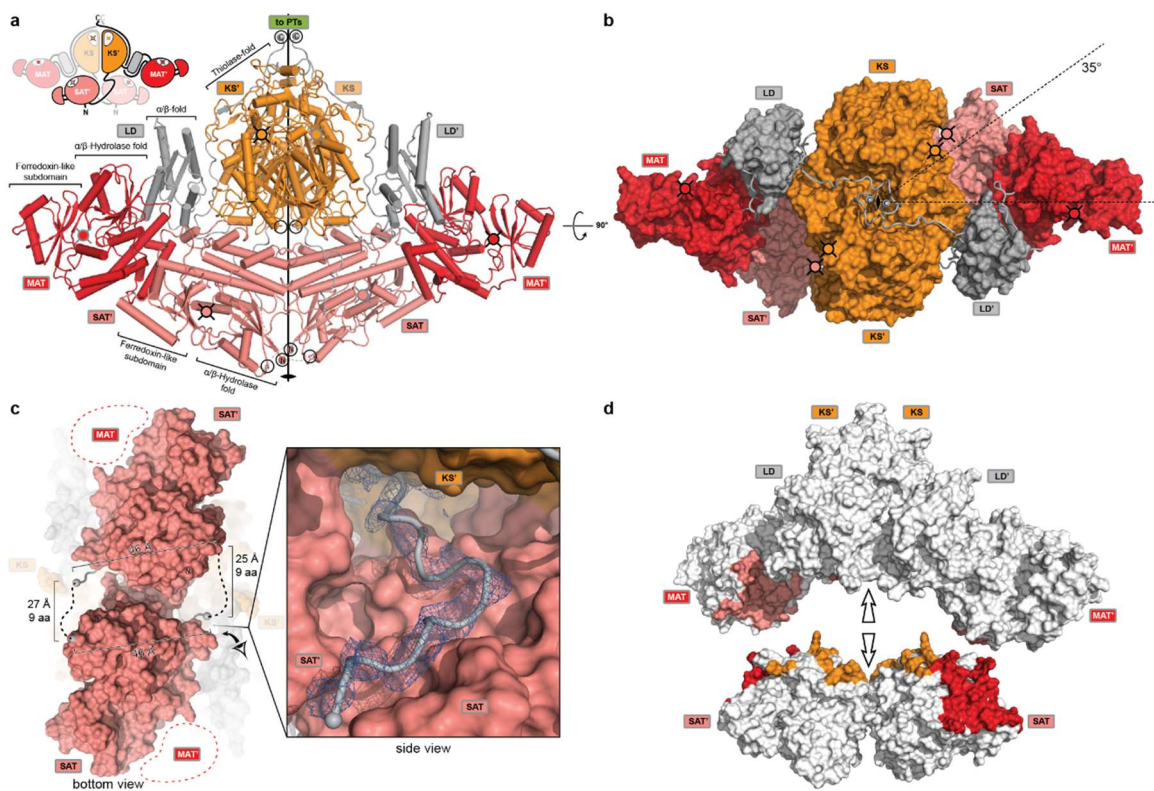


Figure 3.7. Crystal structure of CTB1 SAT-KS-MAT. (a) Cartoon representation of the SAT-KS-MAT crystal structure (front view); the intertwined domain connectivity is indicated in the inset. Active sites and *N*- and *C*-termini are indicated by circles. (b) Top view of SAT-KS-MAT, domains are shown in surface representation, linkers as cartoon. Active sites, termini, and twofold axis are indicated as in (a), the angle between the longest axes of the KS-LD-MAT and SAT (pseudo-) dimers is indicated. (c) The SAT-KS linker. Bottom view of the crystal structure as in (b) highlighting the connectivity of the SAT-KS linker and Euclidean distances. The inset provides a close-up view from the side of the *C*-terminal region of the SAT-KS linker involved in contacts to KS' and SAT. An Fo-Fc shaded omit difference map is shown at 2.5σ . (d) Interdomain interfaces between the SAT domains and the KS-LD-MAT region of CTB1 are mapped onto a split-surface representation of the SAT-KS-MAT structure by coloring according to the interaction partner. The SAT domains have been separated from the KS-LD-MAT for representation.

The *N*-terminus of the SAT and *C*-terminus of the MAT are located close to the twofold axis on opposite ends of the molecule (Figure 3.7a). Therefore, while CTB1 functions as an isolated iPKS, its domain arrangement is consistent with that required for connectivity in an assembly line. The only other known structure of a NR-PKS SAT, CazM SAT, shows similar domain connectivity to

that of CTB1 SAT¹⁶. The structure of CazM SAT, determined by X-ray crystallography as an isolated monodomain, shows its C-terminal linker terminating close to the position of CTB1 SAT's, despite a lack of sequence conservation (Appendix Figure C.1). This linker architecture is notably distinct from that of the isolated loading AT domain from the avermectin modPKS (AVES1)⁵³, which has a distinct domain organization where the loading AT connects to an ACP, rather than a KS. In fact, the AVES1 loading AT linker architecture more closely resembles the post-MAT linker in PKS and FAS condensing regions. These domain arrangements and distinct linker architectures support a specific, conserved organization among NR-PKS loading domains.

3.2.3. Structure of CTB1 SAT^o-KS-MAT^o=ACP2

Through mechanism-based crosslinking, CTB1 SAT^o-KS-MAT^o was covalently linked with CTB1 ACP2 to observe specific KS-ACP interactions and observe the impact of this interaction on the overall domain architecture. Crosslinked CTB1 SAT^o-KS-MAT^o=ACP2 was visualized by cryo-EM at 7.1 Å resolution (Figure 3.8, Appendix Table C.1). This map was interpreted by rigid body and real space refinement using an all-atom model based on the crystal structure of CTB1 SAT-KS-MAT and a homology model of CTB1 ACP2 based on

the PksA ACP solution structure¹⁷. Due to the limited resolution, and the absence of an authentic high-resolution structure, the analysis for ACP2 is restricted to conserved structural elements (Appendix Figure C.2, see Methods for details). Single-particle processing including reference-free maximum likelihood classification provided a single, predominant conformation of SAT^o-KS-MAT^o=ACP2 (for details see Methods). Surprisingly, while the SAT^o-KS-MAT^o was approximately 95% crosslinked to ACP2 (as determined by SDS-PAGE), only a single ACP was resolved in the structure (Figure 3.8a-c, Appendix Figure C.3). Additionally, the overall structure is asymmetric, with changes in domain architecture varied between the apparent crosslinked- and non-crosslinked protomers (Figure 3.8a-c). ACP2 is bound in a cleft between the LD and the KS/KS' dimer, with the DSL motif bearing the aminophosphopantetheine substrate mimic (located at the *N*-terminal end of helix II) in close proximity to the KS active site entrance (Figure 3.8d). While no ACP was resolved at the second KS active site entrance, the binding surface residues of the second KS are more disordered (Appendix Figure C.4), consistent with disordered binding of the second ACP.

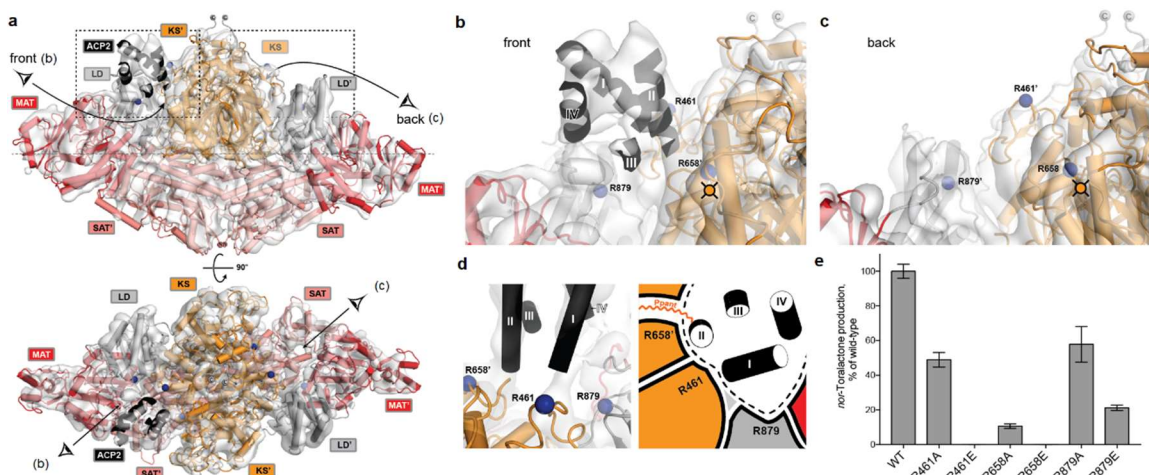


Figure 3.8. Asymmetric cryo-EM structure of CTB1 SAT°-KS-MAT°=ACP2. (a) Reconstructed volume and model of SAT°-KS-MAT°=ACP2 in front and top view. The canonical four helix bundle of CTB1 ACP2 is shown. All maps are contoured at 7.1σ . (b,c) Close-up of the active site (indicated in (a)) regions of the two KS domains with (b) or without (c) resolved ACP2. (d) Location of residues on SAT°-KS-MAT° interacting with ACP2 as structure and schematic representation. In the scheme, the approximate position of the KS active site tunnel and the crosslinker-modified serine in ACP2 are indicated. (e) Functional analysis of CTB1 SAT-KS-MAT mutants of ACP2 interacting residues. *Nor*-toralactone production was quantified by integrated peak areas of HPLC chromatograms at 280 nm (mean \pm s.e.m.).

The structure of SAT°-KS-MAT°=ACP2 is considerably more asymmetric than the SAT-KS-MAT crystal structure, strongly suggesting ACP2 binding causes an asymmetrical conformational shift across the SAT-KS-MAT dimer. In the crosslinked structure, the individual SAT domains are translocated towards the KS domains (Appendix Figure C.5). The proximal SAT moves towards the KS by 6.9 Å, while the opposite SAT moves 4.5 Å towards the opposite KS-MAT region. Additionally, the LD and MAT closest to the resolved bound ACP move upward by 4.4 Å and 8.4 Å, respectively (Figure 3.10). Together, these movements open the interface between the proximal SAT and MAT domains and enlarge the SAT-KS interface (Appendix Figure C.5). Since only one ACP is

resolved in the cryo-EM structure of the 95% crosslinked sample, these asymmetric conformational changes upon the first ACP binding likely disfavor ordered and competent binding of the second ACP. This negative feedback is consistent with the observed kinetic preference for 50% crosslinking (Figure 3.5), as well as activity assays of crosslinked samples of a related NR-PKS. When 50% crosslinked, Pks4 exhibits only 20% activity instead of the expected 50%, relative to a non-crosslinked sample³².

3.2.4. Mutagenic validation of KS-ACP binding region

Using the higher resolution crystal structure of SAT-KS-MAT, amino acids involved in ACP binding can be identified on the KS and LD. However, the limited resolution of the crosslinked EM structure makes assignment of interface amino acids of ACP2 less well defined. Three positively charged residues were identified on the SAT-KS-MAT surface and proposed to interact with acidic patches on ACP2 (Figure 3.8d). Residue R879, located within the LD, interacts with the loop connecting helix I and II in ACP2. Consistent with this proposal, production of *nor*-toralactone in *in vitro* reconstitution experiments with R879A and R879E is reduced, but not completely abolished (Figure 3.8e). Conversely, the KS-ACP interaction mediated by two arginines at the KS/KS' active site entrance appears to be more crucial for productive biosynthesis. R461

is strictly conserved in NR-PKSs (Appendix Figure C.6), and may interact with the acidic patch consisting of E1794 and E1795. Mutation of R461 to alanine reduces *nor*-toralactone production, while mutation to glutamate completely abolishes it (Figure 3.8e). R658' is proposed to interact with D1815 within the DSL motif and/or the phosphate of the phosphopantetheine arm, and is conserved as a positively charged residue in the CTB1 clade (Appendix Figure C.6). This residue appears to be most critical in mediating ACP binding, as the R658A mutation drastically reduces *nor*-toralactone production, while production in reactions with R658E is abolished (Figure 3.8e). Analysis of mutations to proposed binding residues within ACP2 is more complicated, as the ACP must interact with multiple client domains, even in a minimal PKS reaction (SAT-KS-MAT + ACP). However, we can infer perturbation of KS-ACP interaction if the minimal reconstituted PKS is capable of synthesizing short-chain (*e.g.* triacetic acid lactone **3**), but not fully extended derailment products (*e.g.* **4-7**). Mutagenesis of proposed binding residues of ACP2 demonstrated only two residues that perturb KS-ACP binding when individually mutated, E1794 and D1815 (Figure 3.9). However, these residues exist in acidic patches on the ACP2 surface, and multiple ACP residues are likely involved in interactions with a single residue on the KS/LD.

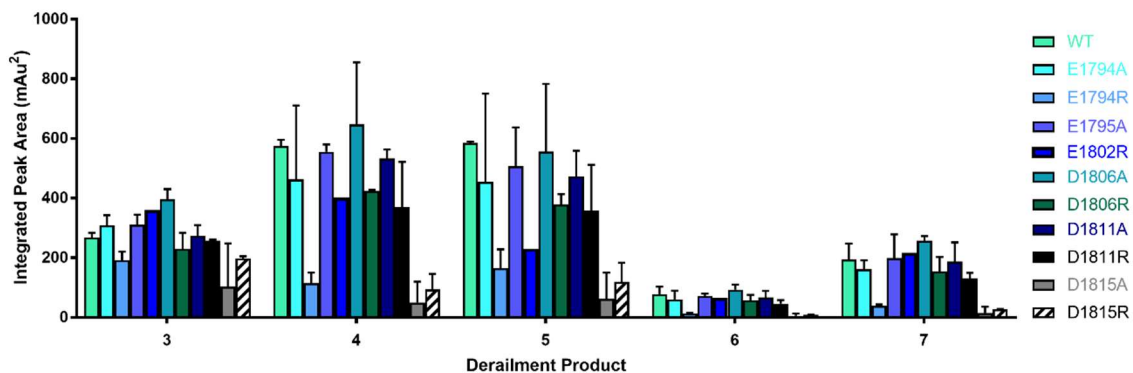


Figure 3.9. Quantification of derailment product production by ACP2 mutants in a minimal PKS reaction. Integrated peak areas were calculated from HPLC chromatograms at 280 nm in triplicate, and are shown as mean \pm s.e.m. See Figure 3.2 for derailment product characterization.

3.3. Discussion

Two previous examples of KS-ACP interaction within a multienzyme have been visualized: a crystal structure of yeast FAS (yFAS) at 3.1 Å resolution⁵⁴ and a cryo-EM reconstruction of a modPKS, PikAIII, at 8.6 Å and 7.3 Å³⁰. In both structures, the observed ACP interactions were strictly symmetrical, in contrast to the asymmetric interactions observed in CTB1 SAT^o-KS-MAT^o=ACP2. Only the heterododecameric yFAS fulfills the distance restraints for symmetric ACP interaction, with solvent accessible surface (SAS) distance from the ACP serine bearing the phosphopantetheine arm to the KS active site cysteine of approximately 27 Å, similar to the corresponding distance in SAT^o-KS-MAT^o=ACP2 of 28 Å (Appendix Figure C.7c, f-i). Conversely, the SAS distances between ACP and KS in PikAIII are non-productive (45.2 Å and 43.3 Å) (Appendix Figure C.7h, i). Consistent with this observation, the position and

orientation of ACP relative to KS is conserved between yFAS and CTB1 (Appendix Figure C.7b), but is considerably different between PikAIII and CTB1 (Appendix Figure C.7d,e). The ACP docking surface observed in CTB1, and verified by mutant activity assays in reconstitution experiments, is partially blocked in PikAIII due to the divergent location of the LD domain (Appendix Figure C.7d,e). The PikAIII structure also reveals symmetric large scale conformational transitions with the requirement for partial refolding of conserved domains, in contrast to asymmetric conformational changes derived by local hinge-bending or shear motions observed in CTB1 SAT^o-KS-MAT^o=ACP2, as well as a highly-reducing iPKS and modPKS⁵⁰. These latter conformational changes maintain the individual domain structures, key conserved interfaces, and the overall interdomain architecture.

Given the asymmetric conformational changes observed upon ACP binding, we propose that the loading and condensing region acts analogously to a V-twin engine, where conformational coupling between the two sets of active sites in the dimer is mediated by ACP binding (Figure 3.10). ACP binding to the KS of one protomer within the SAT-KS-MAT dimer results in an asymmetric conformational change which disfavors binding to the KS of the second protomer. This conformational coupling is consistent with biochemical demonstration of preferential single-site ACP loading in CTB1 and Pks4³².

Furthermore, a related seesaw-type conformational coupling has been observed in FabZ, a hexameric dehydratase involved in fatty acid biosynthesis⁵⁵. In this system, the dehydratase binds ACPs in a 6:3 ratio. We propose that this conformational change favors ACP binding to one of the other client domains, to allow for controlled programming of loading and condensing, where substrate (malonyl-CoA) loading by the MAT alternates with elongation by the KS on either side of the PKS dimer. Since domains (including SAT-KS-MAT tridomains) can be productively swapped between NR-PKS in *in vitro* reconstitution experiments, this V-twin engine-like mechanism of ACP-mediated conformational coupling may be representative of many NR-PKS⁴⁴⁻⁴⁵.

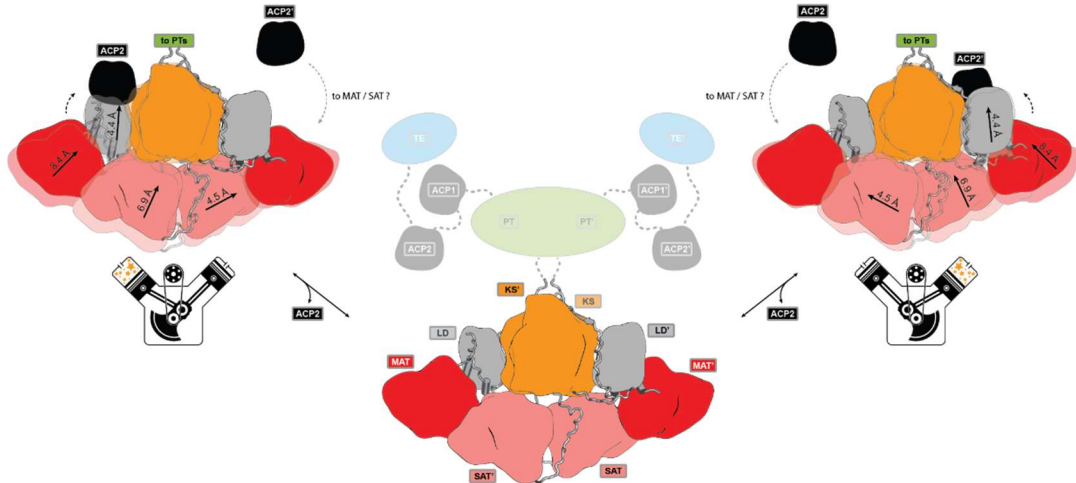


Figure 3.10. Schematic illustration of suggested modes of conformational coupling in CTB1. ACP2 preferentially binds to a single KS active site. ACP2 binding results in coupled conformational changes across the dimeric CTB1, disfavoring productive KS-ACP2 interaction in the other protomer and re-aligning active sites for ACP2 interaction and substrate shuttling. Such conformational coupling resembles a V-twin engine mechanism for alternating KS-ACP2 interactions. Additional regions of complete CTB1 and other NR-PKS, in particular the dimeric PT domains, to which ACPs are linked, could contribute to such a coupling mechanism.

3.4. Experimental Methods

3.4.1. Preparation of expression constructs

All plasmids used in this study are summarized in Appendix Table C.3. The plasmids for expression of wild-type CTB1 SAT-KS-MAT (pECTB1-NKA6), PT (pECTB1-PT), and TE (pECTB1-TE) have been previously described⁴³. Primers used for assembly of new plasmids in this study are detailed in Appendix Table C.4. All expression constructs were prepared and maintained in *E. coli* BL21(DE3). Cut sites for CTB1 ACP2 deconstruction were chosen based on the PksA ACP solution structure¹⁷. CTB1 ACP2 was cloned as an *N*-terminal Thrombin-cleavable His₆-tagged construct to allow for differential purification throughout the crosslinking process. The desired gene sequence was amplified by PCR from pECTB1-ACP (tandem ACP, previously described)⁴³ using CTB1-ACP2-5 and CTB1-ACP2-stop-3, and ligated into pET-28a at NdeI and NotI sites using T4 DNA ligase. Active-site mutations for selective crosslinking were introduced into pECTB1-NKA6 by Gibson assembly⁵⁶ of PCR-amplified fragments using standard protocol, giving pECTB1-SKM-C119A-S1010A and pECTB1-SKM-H688A-S1010A. Mutations to CTB1 SAT-KS-MAT and ACP2 for interface validation were introduced via Gibson assembly in the same manner. Assembled plasmids were screened by restriction digestion and sequences

confirmed by the Johns Hopkins Synthesis and Sequencing Facility (Baltimore, MD). Sequencing revealed the point mutation T321A (compared to Uniprot Q6DQW3) in all SAT containing constructs.

3.4.2. Protein expression and purification for crosslinking and interface validation

All proteins were expressed in *E. coli* BL21(DE3). One liter Terrific Broth cultures supplemented with 25 µg/mL kanamycin (GoldBio) were inoculated with overnight starter cultures, and grown at 37 °C with shaking at 250 rpm until OD₆₀₀ reached 0.7. Cultures were cold-shocked in ice water for 1 h prior to induction with 0.5 mM IPTG (GoldBio). Expression was carried out at 19 °C with shaking at 250 rpm for approximately 16 h. Cells were harvested by centrifugation at 4,000 x g for 15 min and flash frozen in liquid nitrogen for storage at -80 °C.

Cell pellets were thawed in lysis buffer A (50 mM potassium phosphate, pH 8, 300 mM NaCl, 10% (v/v) glycerol) and lysed by sonication. Lysate was cleared by centrifugation at 27,000 x g for 25 min. The resulting supernatant was batch bound to Co²⁺-TALON resin (Clontech) at 4 °C, typically for 1 h. The protein-bound resin was applied to a gravity-flow column and washed and eluted with lysis buffer containing increasing concentrations of imidazole (0-100 mM), as directed by the manufacturer. Fractions containing the protein of

interest were pooled and dialyzed against the appropriate buffer (50 mM Tris pH 7.5, 5% (v/v) glycerol unless otherwise indicated) at 4 °C. For dialysis of CTB1 SAT°-KS-MAT°, 1 mM DTT was included to maintain the KS active-site cysteine in its reduced form. If necessary, the protein was concentrated in Amicon Ultra centrifuge filters of appropriate molecular weight cut-off (Millipore). Protein concentration was quantified based on absorbance at 280 nm on a Cary 50 UV-Vis spectrophotometer and extinction coefficients calculated using ExPASy ProtParam (Appendix Table C.3).

3.4.3. Expression and purification of CTB1-SAT-KS-MAT for crystallization

C-terminal hexa-histidine (His₆) tagged CTB1 SAT-KS-MAT (pECTB1-NKA6, aa 1-1293) was overexpressed with *Streptomyces* chaperonins (pETcoco-2A-L1SL2 plasmid)⁵⁷ in *E. coli* BL21(DE3) pRIL. Cells were cultured in 2xYT medium, supplemented with 0.5 % (v/v) glycerol, NPS (25 mM (NH₄)₂SO₄, 50 mM KH₂PO₄, 50 mM Na₂HPO₄), kanamycin (100 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL). An expression culture (1.5 L) was inoculated (1:20), grown at 37 °C for 2 h, cooled to 20 °C, and induced with IPTG (0.1 mM) at an OD₆₀₀ of 1.0. Cells were harvested after 12 h by centrifugation (7,000 × g) and resuspended in lysis buffer B (50 mM HEPES pH 7.4, 200 mM NaCl, 2.5mM MgCl₂, 40 mM imidazole, 10% (v/v) glycerol, 5 mM β-

mercaptoethanol), supplemented with protease inhibitors (200 μ M phenylmethylsulfonyl fluoride, 20 μ M bestatin, 4 μ M E64, 2 μ M pepstatin A, 20 μ M phenantrolin, 2 μ M phosphoramidon) as well as DNase I, RNase A, and lysozyme. Cells were placed on ice and lysed by sonication. The lysate was cleared by centrifugation (100,000 \times g, 30 min) and immobilized by metal-affinity chromatography on a 25 mL Ni-affinity column (GenScript) pre-equilibrated with lysis buffer. Unbound protein was eluted with five column volumes (CV) of HisA buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 2.5mM MgCl₂, 40 mM imidazole, 10% (v/v) glycerol, 5 mM β -mercaptoethanol, protease inhibitors). The sample was eluted with a linear gradient to 100% HisB buffer (50 mM HEPES pH 7.4, 500 mM imidazole, 50 mM NaCl, 10%(v/v) glycerol, 2.5 mM β -mercaptoethanol, and protease inhibitors) and directly loaded on a strong anion exchange column (PL-SAX 4,000 Å, 10 μ m, Agilent). The protein was eluted with a linear gradient from 0% AIC-A (50 mM HEPES pH 7.4, 50 mM NaCl, 10% (v/v) glycerol, 2.5 mM β -mercaptoethanol) to 100% AIC-B (50 mM HEPES pH 7.4, 1 M NaCl, 10% (v/v) glycerol, 2.5 mM β -mercaptoethanol). The protein was concentrated using Amicon Ultra centrifuge filters (10K MWCO, Millipore) and subjected to size exclusion chromatography (SEC) (Superdex 200 16/60, GE Healthcare) using 20 mM HEPES pH 7.4, 250 mM NaCl, 5% glycerol (v/v), 1 mM dithiothreitol. Protein-containing fractions were pooled and concentrated in

Amicon Ultra centrifuge filters (10K MWCO, Millipore) to 10-15 mg/mL and frozen in liquid nitrogen.

3.4.4. Proteolysis of tag from CTB1-ACP2-His

To facilitate separation of ACP2 from both crosslinker loading reactions and crosslinking reactions, the *N*-terminal His₆-tag was removed by thrombin proteolysis using a Thrombin CleanCleave Kit (Sigma Aldrich). The thrombin agarose resin was washed with reaction buffer (50 mM Tris pH 8, 10 mM CaCl₂) and conditions were optimized as recommended by the manufacturer. Complete cleavage was achieved using approximately 100 µL thrombin agarose resin per 50 mg CTB1-ACP2 at 2 mg/mL CTB1-ACP2 in reaction buffer. Reactions were rotated for 3 h at 25 °C. Cleavage was monitored by SDS-PAGE and confirmed by MALDI-TOF (Bruker Auto Flex III). Cleaved CTB1-ACP2 was isolated by brief centrifugation at 500 × *g* to pellet the thrombin resin and dialyzed in 50 mM Tris pH 7.5, 5% (v/v) glycerol at 4 °C.

3.4.5. Synthesis of α -bromopropionyl aminopantetheine (10)

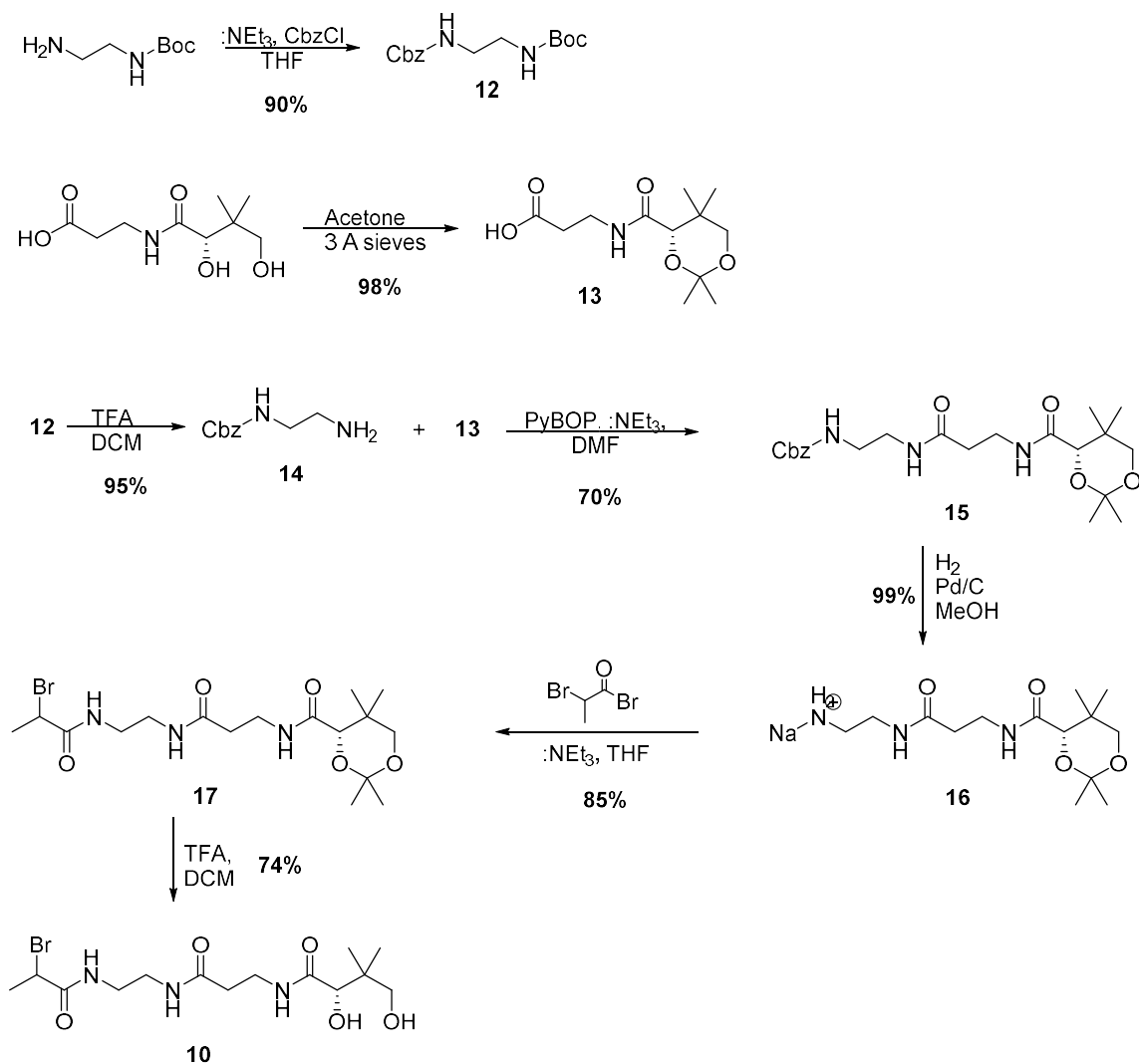


Figure 3.11: Synthesis of crosslinker 10.

All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated.

Benzyl *tert*-butyl ethane-1,2-diylidicarbamate (12): In a 500 mL round-bottomed flask, *N*-Boc-ethylenediamine (14 mL, 88.5 mmol) was stirred in anhyd. DCM (200 mL) under argon. Triethylamine (16 mL, 113.9 mmol) was added, followed

by slow addition of Cbz-Cl (15 mL, 105 mmol). DMAP (1.1 g, 8.9 mmol) was added, and the solution was stirred at room temperature overnight. The reaction was quenched by the addition of a saturated aqueous solution of NH₄Cl (100 mL). The aqueous layer was washed with EtOAc (3×50 mL). The combined organic extracts were dried over anhyd. Na₂SO₄. Crystallization proceeded spontaneously upon concentration on a rotary evaporator and the resulting solid was filtered on a glass frit to give *N*-Boc,*N'*-Cbz-ethylenediamine **12** (23.1 g, 79.7 mmol, 90%) as a white solid. ¹H NMR (400 MHz, CDCl₃): δ 7.35-7.33 p.p.m. (m, 5H), 5.09 (s, 2H), 3.29-3.24 (m, 4H), 1.42 (s, 9H). ¹³C{¹H} NMR (101 MHz, CDCl₃): δ 136.41, 128.44, 128.04, 128.01, 79.46, 66.65, 45.71, 41.38, 40.52, 28.29, 8.51. UPLC-HRMS (ESI) calc'd for C₁₅H₂₂N₂O₄Na⁺ 317.1472; found 317.1471 [M+Na]⁺.

(S)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid (13): *D*-pantothenic acid hemicalcium salt (11 g, 46.2 mmol) and *p*-toluene sulfonic acid hydrate (17 g, 100 mmol) were added to acetone (200 mL) in a flame-dried, round-bottomed flask under argon. Molecular sieves (4 Å, 200 g) were added, and the reaction mixture was stirred vigorously enough to break up the sieves. The reaction was run overnight at room temperature. The sieves were filtered off through a bed of Celite. The solution was concentrated by rotary evaporation and the resulting syrup was redissolved in EtOAc (100 mL). The organic solution was washed with brine (2×50 mL) and dried with anhyd. Na₂SO₄. The

solution was concentrated partially *in vacuo*, and hexanes were added. The ketal-protected pantothenic acid **13** (7.64 g, 29.4, 64%) precipitated as a white solid upon concentration to dryness. ¹H NMR (400 MHz, CDCl₃): δ 7.06 p.p.m. (t, *J* = 5.6 Hz, 1H), 4.10 (s, 1H), 3.67 (d, *J* = 11.6 Hz, A of AB_q, 1H), 3.61-3.56 (m, 1H), 3.50-3.44 (m, 1H), 3.27 (d, *J* = 11.6 Hz, B of AB_q, 1H), 2.59 (t, *J* = 6.0 Hz, 2H), 1.39 (s, 3H), 1.37 (s, 3H), 0.96 (s, 3H), 0.91 (s, 3H). ¹³C{¹H} NMR (101 MHz, CDCl₃): δ 176.32, 170.22, 99.04, 77.00, 71.35, 34.08, 33.77, 32.92, 29.34, 21.96, 18.78, 18.66.

UPLC-HRMS (ESI) calc'd for C₁₂H₂₁NO₅Na⁺ 282.1312; found 282.1312 [M+Na]⁺.

Benzyl (2-aminoethyl)carbamate (14): To a 250 mL round-bottomed flask containing DCM:TFA (30 mL:15 mL) was added **12** (10 g, 34 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with toluene, and concentrated by rotary evaporation three times to give Cbz-ethylenediamine **14** (6.27 g, 32.3 mmol, 95%) as a yellow oil, which was taken directly to the next reaction. The observed spectral data matched those data reported previously⁵⁸.

Benzyl (S)-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl)carbamate (15): In a flame-dried, round-bottomed flask, **13** (8 g, 30.6 mmol), **14** (6.6 g, 34 mmol) and triethylamine (9.48 mL, 68 mmol) were dissolved in DCM (300 mL). The solution was cooled to 0 °C in an ice bath. PyBOP (19.45 g, 37.4 mmol) was added in one portion. The reaction mixture was stirred for 3

h, and allowed to warm to room temperature. The reaction was quenched by the addition of a saturated aqueous solution of NH_4Cl (150 mL). The aqueous layer was extracted with EtOAc (2×100 mL) and the combined organic extracts were dried over anhyd. Na_2SO_4 , followed by concentration by rotary evaporation. A white solid precipitated from the solution and was filtered off. The remaining oil was fractionated by flash silica chromatography (100% EtOAc) to afford pure, diprotected Cbz-aminopantetheine **15** (9.32 g, 21.4 mmol, 70%) as a colorless solid, which was recrystallized from EtOAc/Hexanes. $^1\text{H NMR}$ (400 MHz, CDCl_3): δ 7.25 p.p.m. (m, 5H), 7.05-7.02 (m, 2H), 5.92 (s, 1H), 5.00 (s, 2H), 3.98 (s, 1H), 3.57 (d, $J = 11.7$ Hz, A of AB_q , 1H), 3.42 (septet, $J = 6.6$ Hz 2H), 3.30-3.21 (m, 4H), 3.17 (d, $J = 11.7$ Hz, B of AB_q , 1H), 2.33 (t, $J = 6.4$ Hz, 2H), 1.38 (s, 3H), 1.32 (s, 3H), 0.92 (s, 3H), 0.88 (s, 3H). $^{13}\text{C}\{^1\text{H}\}$ NMR (101 MHz, CDCl_3): δ 171.69, 170.39, 157.04, 136.37, 128.49, 128.14, 128.10, 99.09, 77.09, 71.32, 66.81, 40.81, 40.13, 36.12, 34.79, 32.90, 29.40, 22.05, 18.81, 18.64. UPLC-HRMS (ESI) calc'd for $\text{C}_{22}\text{H}_{34}\text{N}_3\text{O}_6^+$ 436.2442; found 436.2443 $[\text{M}+\text{H}]^+$.

(S)-N-(3-((2-aminoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (16): Compound **15** (7 g, 18.4 mmol) was added to MeOH (10 mL) in a 100 mL Parr bomb flask. The reaction mixture and flask were purged with argon. 10% Pd/C (700 mg) was added. The reaction vessel was placed under a H_2 atmosphere (50 bar), and shaken for 3 h. The reaction mixture was filtered

over Celite to remove the catalyst and the Celite washed with MeOH (3×10 mL).

The filtrate was concentrated by rotary evaporation to afford ketal-protected aminopantetheine **16** (5.5 g, 18.3 mmol, 99%) as a pale yellow oil, which was taken directly to the next step. ¹H NMR (400 MHz, CDCl₃): δ 7.07 p.p.m. (t, *J* = 8.0 Hz, 1H), 6.88 (s, 1H), 4.04 (s, 1H), 3.65 (d, *J* = 15.6 Hz, A of AB_q, 1H), 3.52-3.46 (m, 3H), 3.42 (s, 1H), 3.38-3.37 (m, 2H), 3.24 (d, *J* = 15.6 Hz, B of AB_q, 1H), 2.86 (t, *J* = 7.5 Hz, 2H), 2.42 (t, *J* = 8.7 Hz, 2H), 1.44 (s, 3H), 1.40 (s, 3H), 0.98 (s, 3H), 0.93 (s, 3H). ¹³C{¹H} NMR (101 MHz, CDCl₃): δ 171.51, 170.05, 98.90, 76.96, 71.19, 41.54, 41.02, 35.81, 34.86, 32.75, 29.26, 21.95, 18.70, 18.51. UPLC-HRMS (ESI) calc'd for C₁₄H₂₈N₃O₄⁺ 302.2074; found 302.2079 [M+H]⁺.

(4S)-N-(3-((2-(2-bromopropanamido)ethyl)amino)-3-oxopropyl)-2,2,5,5-

tetramethyl-1,3-dioxane-4-carboxamide (17): In a round-bottomed flask under argon, **16** (1.77 g, 5.87 mmol) was added to DCM (20 mL). To this solution was added triethylamine (1.82 mL, 12.9 mmol). The temperature of the solution was lowered to 0 °C in an ice bath. Bromopropionyl bromide (0.645 mL, 6.16 mmol) was added dropwise. The initially milky solution was stirred at 0 °C for 1 h and became clear. The reaction was quenched by the addition of a saturated aqueous solution of NH₄Cl (30 mL), and the organic layer was washed with saturated aqueous Na₂CO₃ (15 mL) and brine (15 mL). The organic layer was concentrated by rotary evaporation. The resulting off-white solid was purified by flash silica

chromatography (1:9 MeOH:EtOAc) to give 2-bromopropionyl aminopantetheine **17** (2.13 g, 4.87 mmol, 83%) as a colorless solid. The isolated mixture of diastereomers gave the following analytical data: $^1\text{H NMR}$ (400 MHz, CDCl_3): δ 7.11 p.p.m. (s, 1H), 7.02 (t, $J = 5.6$ Hz, 1H), 6.57 (br s, 1H), 4.39/4.38 (2 \times q, $J = 7.0$ Hz, 1H), 4.09/4.07 (2 \times s, 1H), 3.68 (d, $J = 12.0$ Hz, A of AB $_q$, 1H), 3.59-3.52 (m, 2H), 3.43-3.38 (m, 4H), 3.28 (d, $J = 12.0$ Hz, B of AB $_q$, 1H), 2.46 (t, $J = 6.2$ Hz, 2H), 1.85/1.84 (2 \times d, $J = 7.0$ Hz, 3H), 1.46 (s, 3H), 1.42 (s, 3H), 1.02 (s, 3H), 0.97/0.96 (2 \times s, 3H). $^{13}\text{C}\{^1\text{H}\}$ NMR (101 MHz, CDCl_3): δ 172.10/172.07, 170.42, 170.10/170.07, 98.90/98.90, 76.95, 71.12, 43.81/43.80, 40.12/40.05, 38.93, 35.81/35.75, 34.85, 32.74/32.73, 29.26, 22.31/22.82, 21.96/21.95, 18.72, 18.53. UPLC-HRMS (ESI) calc'd for $\text{C}_{17}\text{H}_{31}\text{BrN}_3\text{O}_5^+$ 436.1442; found 436.1433 $[\text{M}+\text{H}]^+$.

(2S)-N-(3-((2-(2-bromopropanamido)ethyl)amino)-3-oxopropyl)-2,4-dihydroxy-3,3-dimethylbutanamide (10): To DCM (2 mL) stirring in a 25 mL round-bottomed flask, **17** (25 mg, 0.06 mmol) was added. 4 mL of deionized water, and 0.1 mL of TFA were added to the reaction mixture. The solution was allowed to stir at room temperature for 60 min. The reaction mixture was diluted with 1:1 H_2O :toluene and concentrated by rotary evaporation three times to remove residual TFA. The resulting pale yellow oil was purified by reverse-phase (C_{18}) flash silica chromatography (gradient: 5-95% ACN in H_2O) to afford 2-bromopropionyl aminopantetheine **10** (15 mg, 1.2 mmol, 63%) as a white solid

after lyophilization. ^1H NMR (400 MHz, MeOD): δ 4.47 p.p.m. (q, $J = 6.8$ Hz, 1H), 3.91 (s, 1H), 3.50-3.47 (m, 2H), 3.48 (d, $J = 10.8$ Hz, A of AB_q, 1H), 3.40 (d, $J = 10.8$ Hz, B of AB_q, 1H), 3.36-3.32 (m, 4H), 2.43 (t, $J = 6.8$ Hz, 2H), 1.77 (d, $J = 6.8$ Hz, 3H), 0.93 (s, 6H). $^{13}\text{C}\{^1\text{H}\}$ NMR (101 MHz, MeOD): δ 176.00, 174.17, 172.76, 77.30, 70.28, 43.97, 40.32, 40.30, 39.70, 36.58, 36.38, 22.41, 21.35, 20.88. **UPLC-HRMS** (ESI) calc'd for C₁₄H₂₇BrN₃O₅⁺ 396.1129; found 396.1120 [M+H]⁺.

3.4.6. Phosphorylation of α -bromopropionyl aminopantetheine (**10**)

α -Bromopropionyl aminopantetheine **10** was phosphorylated enzymatically under the following conditions: 0.25 mg/mL PanK, 2 mM **10**, 5 mM ATP, 20 mM KCl, 10 mM MgCl₂, 50 mM Tris pH 7.5 at room temperature for approximately 20 h. Complete phosphorylation to yield **11** was confirmed by UPLC-ESI-MS (Waters Acquity/Xevo-G2, negative ion mode). All protein was removed from the reaction with a 3K MWCO Amicon Ultra centrifuge filter (Millipore). The flow-through was flash frozen in liquid nitrogen and lyophilized to dryness. The resulting white powder was resuspended in 5% aqueous acetonitrile. **11** was HPLC purified on an Agilent 1100 equipped with a Kinetex C18 semi-prep column (250 x 10 mm, 5 μ , Phenomenex) using the following method at 4 mL/min: hold 5% solvent A/95% solvent B for 3 minutes, 5-35% solvent A over 17 minutes, 35-95% solvent A over 5 minutes, followed by

re-equilibration to 5% solvent A, where solvent A is acetonitrile + 0.1% trifluoroacetic acid and solvent B is 0.1% trifluoroacetic acid. Purity and identity of the isolated compound was confirmed by UPLC-ESI-MS as above, and the pooled fractions were lyophilized to dryness.

Due to poor recovery of **11** from HPLC purification, preparation of **11** for crosslinking to CTB1 SAT-KS^o-MAT^o was optimized. The phosphorylation of **10** was performed as before, and used directly in the next step (adenylation and ACP loading), once the reaction was complete.

3.4.7. Loading of ACP2 with α -bromoacyl crosslinker

α -Bromopropionyl aminophosphopantetheine **11** was adenylated to yield **9** and subsequently loaded onto CTB1 ACP2 in a one-pot enzymatic reaction under the following conditions: 0.5 mM **11**, 0.5 mg/mL PPAT, 5 μ M Sfp, 0.5 mM CTB1 ACP2, 5 mM ATP, 10 mM MgCl₂, 20 mM KCl, 50 mM Tris pH 7.5, 10% glycerol. Reactions were run at room temperature and monitored by MALDI-TOF (Bruker AutoFlex III). Once loading was complete, the His-tagged Sfp and PPAT were separated from untagged CTB1 ACP2 by passing the reaction solution through an equilibrated Co²⁺-TALON column (Clontech) and recovering the flow-through. Additionally, any unreacted crosslinker was removed by repeated concentration and dilution with a 3K MWCO Amicon Ultra centrifuge

filter (Millipore). Complete removal of free crosslinker was confirmed by UPLC-ESI-MS analysis.

3.4.8. Crosslinking of CTB1 SAT^o-KS-MAT^o to ACP2

CTB1 SAT^o-KS-MAT^o was combined with 20-fold excess CTB1 ACP2 bearing the α -bromoacyl crosslinker in 50 mM Tris pH 7.5, 10% glycerol (v/v) at room temperature for 1 h. Crosslinking was monitored by SDS-PAGE, and found to be approximately 95% complete after 1 h. Crosslinked CTB1 SAT^o-KS-MAT^o=ACP2 was separated from remaining free CTB1 SAT^o-KS-MAT^o by anion exchange chromatography. Q-Sepharose Fast Flow resin (GE Healthcare) was equilibrated with 50 mM Tris pH 7.5, 10% glycerol, and the crosslinking reaction mixture was applied to the column. CTB1 SAT^o-KS-MAT^o and CTB1 SAT^o-KS-MAT^o=ACP2 were eluted using a gradient of potassium chloride from 0-500 mM. To remove free CTB1 ACP2 from the crosslinked protein, fractions containing CTB1 SAT^o-KS-MAT^o=ACP2 were pooled and applied to an equilibrated Co²⁺-TALON column. CTB1 ACP2 was removed in the flow-through, and pure CTB1 SAT^o-KS-MAT^o=ACP2 was eluted with a gradient of imidazole from 2-100 mM.

3.4.9. Crosslinking of CTB1 SAT-KS^o-MAT^o to ACP2

CTB1 SAT-KS^o-MAT^o was combined with 10-fold excess CTB1 ACP2 bearing the α -bromoacyl crosslinker in 50 mM Tris pH 7.5, 10% glycerol (v/v) at room temperature for 1 h to reach 50% crosslinking. NaCl was added to a final concentration of 250 mM to impede further crosslinking. CTB1 SAT-KS^o-MAT^o=ACP2 was isolated from free ACP2 by size-exclusion chromatography using a Bio-Rad NGC fitted with an ENrich SEC650 column (10 x 300 mm, Bio-Rad). Fractions containing CTB1 SAT-KS^o-MAT^o=ACP2 were pooled and concentrated using an Amicon Ultra centrifuge filter (10K MWCO, Millipore).

3.4.10. In vitro reactions of CTB1 for interface validation and comparison of ACP1 and ACP2

CTB1 SAT-KS-MAT (wild-type and select mutants of proposed interface residues), PT, ACP2 (wild-type and select mutants of propose interface residues), and TE were purified as described above, and dialyzed into 100 mM potassium phosphate pH 7, 10% glycerol. ACP2 mutants were additionally purified by size exclusion chromatography (ENrich SEC 70, 10 x 300 mm, Bio-Rad) using 100 mM potassium phosphate pH 7, 10% glycerol. ACP2 was activated enzymatically with CoA under the following conditions: 2 μ M Sfp, 0.5 mM CoA, 10 mM MgCl₂, 200 μ M ACP2, 100 mM potassium phosphate pH 7, 10% glycerol. Activation reactions were run at 25 °C for 1 h, and complete conversion to *holo*-ACP2 was

confirmed by UPLC-ESI-MS (Waters Acquity/Xevo-G2, positive ion mode) and MALDI-TOF (Bruker AutoFlex III). For full PKS reconstitution, 10 μ M CTB1-SAT-KS-MAT, PT, ACP2 and TE were combined in assay buffer (100 mM potassium phosphate pH 7, 10% glycerol, 1 mM TCEP). Minimal PKS reactions excluded PT and TE. Reactions were initiated by addition of 0.5 mM acetyl-S-N-acetylcysteamine (SNAC) and 2 mM malonyl-SNAC, and incubated for 4 h at 25 °C. The 250 μ L reactions were quenched with 10 μ L hydrochloric acid and extracted thrice with ethyl acetate. Extracts were pooled, dried, and resuspended in 250 μ L 20% aqueous acetonitrile for HPLC analysis.

Product profiles of each enzyme mutant were analyzed on an Agilent 1200 equipped with a Prodigy ODS3 analytical column (4.6 x 250 mm, 5 μ , Phenomenex). 100 μ L injections were separated by a bisolvent method at 1 mL/min: 5-85% solvent A over 40 min, 85-95% solvent A over 10 min, re-equilibrate to 5% solvent A over 10 min, where solvent A is acetonitrile + 0.1% formic acid, and solvent B is 0.1% formic acid. Chromatograms were recorded at 280 nm, and UV-vis spectra were recorded over a range of 200-800 nm. Exact masses of derailment products were obtained by UPLC-MS (Waters Acquity/Xevo-G2, positive ion mode).

3.4.11. Crystallization, data collection and structure determination of CTB1 SAT-KS-MAT

CTB1 SAT-KS-MAT plate-like crystals were obtained by the sitting drop vapor diffusion method in 0.2 M MgCl₂, 0.1 M Bis-Tris Propane pH 6.5, 18% (v/v) PEG3350 at 4 °C, a drop ratio of 1.0 µL + 0.5 µL and further optimized by seeding. Crystals grew to a final size of 0.4×0.04×0.04 mm³ within one week. The crystals were dehydrated and cryo-preserved by successively increasing the concentration to 0.5 M MgCl₂, 0.1 M Bis-Tris Propane pH 6.5, 25% PEG3350, 22% ethylene glycol. Diffraction data of crystals in space group C222₁ (a= 108.1 Å, b= 230.2 Å, c= 253.8 Å) and 56% solvent content were collected at the Swiss Light Source (SLS, Villigen, Switzerland) at beamline X06SA, a temperature of 100 K and a wavelength of 1.000 Å. Data reduction was performed using XDS and XSCALE⁵⁹ and data were analyzed using phenix.xtriage⁶⁰. The structure was solved with PHASER⁶¹ using molecular replacement and the KS and AT domains of DEBS module five¹⁹ as well as the AT domain of CurL⁵¹ as search models. An initial model was obtained in iterative cycles of rebuilding with BUCCANEER⁶² and density modification using solvent flattening and histogram matching in PARROT⁶³. Iterative cycles of manual model building and real space refinement in Coot⁶⁴ and TLS refinement in Phenix⁶⁰ yielded a high quality model with

$R_{\text{work}}/R_{\text{free}}$ values of 0.21/0.24 at 2.8 Å resolution with excellent geometry (Ramachandran favored/outliers: 97.07%/0.00%) (Appendix Table C.1a).

3.4.12. EM sample preparation and data collection.

Initial sample quality was evaluated by conventional negative-stain EM. Initial cryo-EM analysis revealed a monomer-dimer equilibrium, which has previously been observed for PKS and FAS fragments^{30, 50-51, 65-66}. The oligomeric state of the CTB1 SAT-KS-MAT was analyzed by analytical size-exclusion chromatography (SEC) (4.6/250 Superose 6 Increase, GE Healthcare), using buffers with different salt concentrations and a flow rate of 0.1 mL/min. The results indicated a fast equilibrium and a mostly dimeric species in low salt buffer (20 mM Tris pH 7.4, 50 mM NaCl, 2.5 mM tris(2-carboxyethyl)phosphine (TCEP)) (Appendix Figure C.3c). For cryo-EM, crosslinked sample was concentrated to 15-16 mg/mL, applied to analytical SEC and collected in fractions of 50 µL. For grid preparation 4 µL of individual fractions were applied to glow discharged (30 s) lacey carbon grids (300 mesh, copper), immediately blotted for 2 s and plunge-frozen in liquid ethane using a Vitrobot (FEI, Vitrobot III). An incubation step was omitted due to fast adsorption of the particles to the carbon. The integrity of the crosslinked sample was analyzed by denaturing

polyacrylamide gel electrophoresis (PAGE) of the remaining sample after grid preparation (Appendix Figure C.3d).

Grid quality was analyzed using a Philips CM200 FEG cryo-transmission electron microscope, operated at an acceleration voltage of 200 kV. Optimal conditions were identified for peak fractions diluted to 270 $\mu\text{g/mL}$, in a trade-off between low particle density (approx. 70 particles per micrograph), monomer formation and particle adsorption to carbon. Final data were collected using a Titan Krios electron microscope (FEI), operated at 300 kV, a GIF Quantum LS imaging filter (Gatan) and a K2 Summit (Gatan) operating in counting mode. Images were acquired at 0.8-4.5 μm defocus and a nominal magnification of 105,000x, corresponding to a pixel size of 1.326 \AA (Appendix Table C.1b). Movies were collected with a total dose of approx. 90 $\text{e}^-/\text{\AA}^2$ per 18 s exposure, fractionated over 60 frames.

3.4.13. EM data processing and analysis

Drift correction of dose fractionated frames²⁸ was performed using Zorro⁶⁷ and global contrast transfer function (CTF) parameters were determined using CTFFIND4.1⁶⁸. Particles were picked from aligned dose-filtered averages using the swarm semi-automated procedure as implemented in e2boxer.py⁶⁹, followed by extraction from unfiltered averages using Relion-2⁷⁰ and local CTF parameter

refinement using Gctf v1.06⁷¹. Single particle analysis included several rounds of reference-free maximum *a posteriori*-2D classification⁷⁰ to remove junk particles and monomers, which revealed a clearly distinguishable shape and could be reconstructed at low resolution (Appendix Figures C.2a, C.3b). For reconstruction of the dimeric, crosslinked CTB1 SAT^o-KS-MAT^o=ACP, an initial volume was generated using e2initialmodel.py⁶⁹ and low-pass filtered to 50 Å. Exhaustive global followed by local reference-free 3D classification and alignment without symmetry were applied, yielding highly similar classes that were combined in a consensus refinement prior to movie refinement and particle polishing as implemented in Relion-2⁷⁰. Parameters for B-factor fitting as function of movie frames have been carefully examined as well as the different number of frames in the particle movie stacks. At this stage all 60 frames were included in the weighted particle movie stacks as no significant difference between including different amounts of frames per movie could be identified after applying weighting. Another round of 2D and 3D classification yielded three almost identical volumes that were refined as individual classes and in all possible combinations of classes. The combination of two classes (Appendix Figure C.2a) resulted in a slightly improved model at 7.3 Å resolution (25,107 particles) as compared to all three classes at 7.5 Å (44,859 particles). For the refinement of the final map at 7.1 Å resolution weighted particle averages of 27

movie frames were used, corresponding to a total dose of 41 e-/Å². All reported resolution estimates are based on the 0.143 threshold criterion of the Fourier Shell Correlation (FSC)⁷² between two halves of the dataset refined independently in RELION-2, after accounting for correlations introduced by masking⁷³. Throughout all refinements in Relion-2⁷⁰, particles were grouped by k-mean clustering of their CTF parameters with a minimum of 50 particles per group. The alignment was not significantly influenced by a preferred orientation⁷⁴ (Appendix Figure C.2b, c), which was assessed by reducing the number of particles per angular sampling group to the mean of all groups by random subset deletion in four independent refinements. Later refinements were carried out using a soft mask and solvent corrected FSC during the refinement⁷³. Final maps were sharpened with a sharpening B-factor of -350 Å². Local resolution was calculated with ResMap⁷⁵ and indicated higher resolution for secondary structure elements in the center of the particle (Appendix Figure C.2d, e). Therefore, masked 3D classification and refinement after signal subtraction⁷⁶ was carried out, resulting in slightly improved resolution, but without significant differences in the electron density map. Map generation from coordinates as well as filtering and file type conversion was done using EMAN2 tools⁶⁹. For analysis, map transformations were applied using the CCP4 tool

maprot⁷⁷. Maps were resampled using UCSF Chimera⁷⁸ and difference maps were calculated Python Numpy and MRCZ⁷⁹.

3.4.14. Cryo-EM structure refinement and modelling

Modelling of the cryo-EM structure was started by rigid body fitting of individual domains of the CTB1 SAT-KS-MAT crystal as well as of an ACP2 homology model in Coot⁶⁴. The model for CTB1 ACP2 was generated using SwissModel⁸⁰ based on PksA ACP¹⁷, which shares 40% sequence identity and can be functionally replaced with the CTB1 tandem ACPs in *in vitro* assays⁴⁴⁻⁴⁵. Initial B-factors of the model were scaled relative to B-factors of atomic displacement parameter (ADP) refined crystal structures in the PDB at similar resolution. Prior to map-based real-space refinement, the model was protonated using phenix.reduce and geometry minimization with secondary structure restraints and reference coordinate restraints was applied using phenix.geometry_minimization⁶⁰. Afterwards global minimization and ADP refinement with reference structure restraint was carried out against the unsharpened cryo-EM map in phenix.real_space_refine⁶⁰. Reference structure and nonbonded distance restraint weights were carefully monitored and optimized. As the local resolution around the resolved ACP2 is not sufficient to unambiguously interpret possible conformational changes of loop regions

relative to the isolated PksA ACP NMR structure, the loop regions were excluded from the final model after all atom refinement, without affecting model statistics (Appendix Table C.1b). The resulting model contained the regions corresponding to the canonical ACP four helix bundle (residues: 1782-1796, 1816-1828, 1839-1842, 1846-1854) and was validated using phenix.molprobity⁸¹ with Ramachandran plot statistics of “favored/outliers”: 96.47%/0.12%. A position for a second crosslinked ACP2 could not be undoubtedly resolved in any intermediate or final EM map. The sample has been analyzed after grid preparation and is approximately 95% crosslinked (Appendix Figure C.3d). Due to residual noise around the putative second ACP2 binding site the presence of a second ACP2 in a mostly disordered state without a solid interface, tethered only by covalent crosslinking (Appendix Figure C.4a,b) cannot be strictly excluded. However, we cannot exclude the possibility that effects of selective adsorption of partially or fully crosslinked CTB1 SAT^o-KS-MAT^o=ACP2 or of selective destabilization under grid preparation also contribute to selective imaging of a single-crosslinked state with low population in the sample used for grid preparation.

3.4.15. Structure analysis and visualization

Related structures were identified using PDBeFold (20% query / 10% target)⁸² and interfaces were analyzed using QtPISA v2.1.0⁸³. Transformations and coordinate manipulations were carried out using CCP4⁸⁴ tools and MOLEMAN⁸⁵. Bias-removal for crystallographic $F_{\text{obs}}-F_{\text{calc}}$ omit maps was achieved by applying a random perturbation to coordinates ($\Delta 0.2 \text{ \AA}$) and B-factors ($\Delta 20\%$ of the mean overall B-factor) with MOLEMAN2⁸⁵ prior to refinement. $C\alpha$ -r.m.s.d. between domains of the crystal and cryo-EM structure were calculated using LSQMAN⁸⁶. Structures of PikAIII were modeled according to Dutta *et al.*³⁰ using Chimera⁷⁸. Solvent accessible surface distances (SAS) between $C\alpha$ residues of ACP phosphopantetheinylation and KS active site residues were calculated using Xwalk⁸⁷. Figures, movies and plots were generated using Pymol⁸⁸ and Python Matplotlib.

3.4.16. Sequence analysis

Sequence alignments were generated using Clustal Omega⁸⁹. A Phylogenetic tree was generated using the neighboring joining algorithm in Geneious v8.1.6⁹⁰.

3.5. References

1. Hertweck, C., The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **2009**, *48* (26), 4688-716.
2. Staunton, J.; Weissman, K. J., Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **2001**, *18* (4), 380-416.
3. Crawford, J. M.; Townsend, C. A., New insights into the formation of fungal aromatic polyketides. *Nat Rev Microbiol* **2010**, *8* (12), 879-89.
4. Robbins, T.; Liu, Y. C.; Cane, D. E.; Khosla, C., Structure and mechanism of assembly line polyketide synthases. *Curr Opin Struct Biol* **2016**, *41*, 10-18.
5. Smith, S., The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J* **1994**, *8* (15), 1248-59.
6. Carreras, C. W.; Gehring, A. M.; Walsh, C. T.; Khosla, C., Utilization of enzymatically phosphopantetheinylated acyl carrier proteins and acetyl-acyl carrier proteins by the actinorhodin polyketide synthase. *Biochemistry* **1997**, *36* (39), 11757-11761.
7. Crawford, J. M.; Korman, T. P.; Labonte, J. W.; Vagstad, A. L.; Hill, E. A.; Kamari-Bidkorpeh, O.; Tsai, S. C.; Townsend, C. A., Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **2009**, *461* (7267), 1139-43.
8. Crawford, J. M.; Dancy, B. C.; Hill, E. A.; Udworthy, D. W.; Townsend, C. A., Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc Natl Acad Sci U S A* **2006**, *103* (45), 16728-33.
9. Crawford, J. M.; Vagstad, A. L.; Ehrlich, K. C.; Townsend, C. A., Starter unit specificity directs genome mining of polyketide synthase pathways in fungi. *Bioorg Chem* **2008**, *36* (1), 16-22.
10. Crawford, J. M.; Vagstad, A. L.; Whitworth, K. P.; Ehrlich, K. C.; Townsend, C. A., Synthetic strategy of nonreducing iterative polyketide synthases and the origin of the classical "starter-unit effect". *Chembiochem* **2008**, *9* (7), 1019-23.
11. Du, L.; Lou, L., PKS and NRPS release mechanisms. *Nat Prod Rep* **2010**, *27* (2), 255-78.
12. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008**, *320* (5873), 243-6.
13. Klenow, H.; Overgaard-Hansen, K., Proteolytic cleavage of DNA polymerase from Escherichia Coli B into an exonuclease unit and a polymerase unit. *FEBS Lett* **1970**, *6* (1), 25-27.
14. Mattick, J. S.; Tsukamoto, Y.; Nickless, J.; Wakil, S. J., The architecture of the animal fatty acid synthetase. I. Proteolytic dissection and peptide mapping. *J Biol Chem* **1983**, *258* (24), 15291-9.

15. Udworthy, D. W.; Merski, M.; Townsend, C. A., A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *Journal of Molecular Biology* **2002**, 323 (3), 585-598.
16. Winter, J. M.; Cascio, D.; Dietrich, D.; Sato, M.; Watanabe, K.; Sawaya, M. R.; Vederas, J. C.; Tang, Y., Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J Am Chem Soc* **2015**, 137 (31), 9885-93.
17. Wattana-amorn, P.; Williams, C.; Ploskon, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P., Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry* **2010**, 49 (10), 2186-93.
18. Korman, T. P.; Crawford, J. M.; Labonte, J. W.; Newman, A. G.; Wong, J.; Townsend, C. A.; Tsai, S. C., Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc Natl Acad Sci* **2010**, 107 (14), 6246-51.
19. Tang, Y.; Kim, C. Y.; Mathews, II; Cane, D. E.; Khosla, C., The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* **2006**, 103 (30), 11124-9.
20. Tang, Y.; Chen, A. Y.; Kim, C. Y.; Cane, D. E.; Khosla, C., Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem Biol* **2007**, 14 (8), 931-43.
21. Maier, T.; Jenni, S.; Ban, N., Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **2006**, 311 (5765), 1258-62.
22. Maier, T.; Leibundgut, M.; Ban, N., The crystal structure of a mammalian fatty acid synthase. *Science* **2008**, 321 (5894), 1315-22.
23. Weissman, K. J., Uncovering the structures of modular polyketide synthases. *Nat Prod Rep* **2015**, 32 (3), 436-53.
24. Brignole, E. J.; Smith, S.; Asturias, F. J., Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat Struct Mol Biol* **2009**, 16 (2), 190-7.
25. Brignole, E. J.; Asturias, F., Single-particle electron microscopy of animal fatty acid synthase describing macromolecular rearrangements that enable catalysis. *Methods Enzymol* **2010**, 483, 179-202.
26. Armache, J. P.; Jarasch, A.; Anger, A. M.; Villa, E.; Becker, T.; Bhushan, S.; Jossinet, F.; Habeck, M.; Dindar, G.; Franckenberg, S.; Marquez, V.; Mielke, T.; Thomm, M.; Berninghausen, O.; Beatrix, B.; Soding, J.; Westhof, E.; Wilson, D. N.; Beckmann, R., Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc Natl Acad Sci U S A* **2010**, 107 (46), 19748-53.
27. Gilbert, R. J.; Fucini, P.; Connell, S.; Fuller, S. D.; Nierhaus, K. H.; Robinson, C. V.; Dobson, C. M.; Stuart, D. I., Three-dimensional structures of translating ribosomes by Cryo-EM. *Mol Cell* **2004**, 14 (1), 57-66.
28. Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A.; Cheng, Y., Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **2013**, 10 (6), 584-90.

29. Campbell, M. G.; Cheng, A.; Brilot, A. F.; Moeller, A.; Lyumkis, D.; Veessler, D.; Pan, J.; Harrison, S. C.; Potter, C. S.; Carragher, B.; Grigorieff, N., Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **2012**, *20* (11), 1823-8.
30. Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structure of a modular polyketide synthase. *Nature* **2014**, *510* (7506), 512-7.
31. Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G., Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014**, *510* (7506), 560-+.
32. Bruegger, J.; Haushalter, R. W.; Vagstad, A. L.; Shakya, G.; Mih, N.; Townsend, C. A.; Burkart, M. D.; Tsai, S. C., Probing the selectivity and protein-protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers. *Chem Biol* **2013**, *20* (9), 1135-46.
33. Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B.; McCammon, J. A.; Opella, S. J.; Tsai, S. C.; Burkart, M. D., Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **2014**, *505* (7483), 427-31.
34. Meier, J. L.; Haushalter, R. W.; Burkart, M. D., A mechanism based protein crosslinker for acyl carrier protein dehydratases. *Bioorg Med Chem Lett* **2010**, *20* (16), 4936-9.
35. Worthington, A. S.; Rivera, H.; Torpey, J. W.; Alexander, M. D.; Burkart, M. D., Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. *ACS Chem Biol* **2006**, *1* (11), 687-91.
36. Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T., Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc Natl Acad Sci U S A* **2016**, *113* (7), 1802-7.
37. Barajas, J. F.; Finzel, K.; Valentic, T. R.; Shakya, G.; Gamarra, N.; Martinez, D.; Meier, J. L.; Vagstad, A. L.; Newman, A. G.; Townsend, C. A.; Burkart, M. D.; Tsai, S. C., Structural and Biochemical Analysis of Protein-Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew Chem Int Ed Engl* **2016**, *55* (42), 13005-13009.
38. Chen, H.; Lee, M. H.; Daub, M. E.; Chung, K. R., Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. *Mol Microbiol* **2007**, *64* (3), 755-70.
39. Choquer, M.; Dekkers, K. L.; Chen, H. Q.; Cao, L.; Ueng, P. P.; Daub, M. E.; Chung, K. R., The CTB1 gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of *Cercospora nicotianae*. *Mol Plant Microbe Interact* **2005**, *18* (5), 468-76.
40. Cavallini, L.; Bindoli, A.; Macri, F.; Vianello, A., Lipid peroxidation induced by cercosporin as a possible determinant of its toxicity. *Chem Biol Interact* **1979**, *28* (2-3), 139-46.

41. Daub, M. E., Peroxidation of tobacco membrane lipids by the photosensitizing toxin, cercosporin. *Plant Physiol* **1982**, 69 (6), 1361-4.
42. Daub, M. E.; Ehrenshaft, M., THE PHOTOACTIVATED CERCOSPORA TOXIN CERCOSPORIN: Contributions to Plant Disease and Fundamental Biology. *Annu Rev Phytopathol* **2000**, 38, 461-490.
43. Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A., Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun (Camb)* **2012**, 48 (96), 11772-4.
44. Newman, A. G.; Vagstad, A. L.; Storm, P. A.; Townsend, C. A., Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J Am Chem Soc* **2014**, 136 (20), 7348-62.
45. Vagstad, A. L.; Newman, A. G.; Storm, P. A.; Belecki, K.; Crawford, J. M.; Townsend, C. A., Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angew Chem Int Ed Engl* **2013**, 52 (6), 1718-21.
46. Levintow, L.; Novelli, G., The synthesis of coenzyme A from panthetheine: preparation and properties of panthetheine kinase. *J Biol Chem* **1954**, 207 (2), 761-5.
47. Geerlof, A.; Lewendon, A.; Shaw, W. V., Purification and characterization of phosphopantetheine adenylyltransferase from Escherichia coli. *J Biol Chem* **1999**, 274 (38), 27105-11.
48. Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T., Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **1998**, 37 (6), 1585-95.
49. Pappenberger, G.; Benz, J.; Gsell, B.; Hennig, M.; Ruf, A.; Stihle, M.; Thoma, R.; Rudolph, M. G., Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design. *J Mol Biol* **2010**, 397 (2), 508-19.
50. Herbst, D. A.; Jakob, R. P.; Zahringer, F.; Maier, T., Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **2016**, 531 (7595), 533-7.
51. Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L., Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem Biol* **2013**, 20 (11), 1340-51.
52. Kapur, S.; Lowry, B.; Yuzawa, S.; Kenthirapalan, S.; Chen, A. Y.; Cane, D. E.; Khosla, C., Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc Natl Acad Sci U S A* **2012**, 109 (11), 4110-5.
53. Wang, F.; Wang, Y.; Ji, J.; Zhou, Z.; Yu, J.; Zhu, H.; Su, Z.; Zhang, L.; Zheng, J., Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase. *ACS Chem Biol* **2015**, 10 (4), 1017-25.

54. Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N., Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **2007**, *316* (5822), 288-90.
55. Zhang, L.; Xiao, J.; Xu, J.; Fu, T.; Cao, Z.; Zhu, L.; Chen, H. Z.; Shen, X.; Jiang, H.; Zhang, L., Crystal structure of FabZ-ACP complex reveals a dynamic seesaw-like catalytic mechanism of dehydratase in fatty acid biosynthesis. *Cell Res* **2016**, *26* (12), 1330-1344.
56. Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **2009**, *6* (5), 343-5.
57. Betancor, L.; Fernandez, M. J.; Weissman, K. J.; Leadlay, P. F., Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with *Streptomyces* chaperonins in *Escherichia coli*. *Chembiochem: A European Journal of Chemical Biology* **2008**, *9* (18), 2962-6.
58. Barker, P. L.; Gendler, P. L.; Rapoport, H., Acylation of dibasic compounds containing amino amidine and aminoguanidine functions. *The Journal of Organic Chemistry* **1981**, *46* (12), 2455-2465.
59. Kabsch, W., Xds. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 2), 125-32.
60. 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., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 2), 213-21.
61. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J Appl Crystallogr* **2007**, *40* (Pt 4), 658-674.
62. Cowtan, K., The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D* **2006**, *62* (9), 1002-1011.
63. Cowtan, K., Recent developments in classical density modification. *Acta Crystallographica. Section D: Biological Crystallography* **2010**, *66* (Pt 4), 470-478.
64. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D: Biological Crystallography* **2004**, *60* (Pt 12 Pt 1), 2126-32.
65. Edwards, A. L.; Matsui, T.; Weiss, T. M.; Khosla, C., Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J Mol Biol* **2014**, *426* (11), 2229-45.
66. Asturias, F. J.; Chadick, J. Z.; Cheung, I. K.; Stark, H.; Witkowski, A.; Joshi, A. K.; Smith, S., Structure and molecular organization of mammalian fatty acid synthase. *Nat Struct Mol Biol* **2005**, *12* (3), 225-32.
67. McLeod, R. A.; Kowal, J.; Ringler, P.; Stahlberg, H., Robust image alignment for cryogenic transmission electron microscopy. *J Struct Biol* **2017**, *197* (3), 279-293.
68. Rohou, A.; Grigorieff, N., CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **2015**, *192* (2), 216-21.

69. Tang, G.; Peng, L.; Baldwin, P. R.; Mann, D. S.; Jiang, W.; Rees, I.; Ludtke, S. J., EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **2007**, *157* (1), 38-46.
70. Kimanius, D.; Forsberg, B. O.; Scheres, S. H.; Lindahl, E., Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **2016**, *5*.
71. Zhang, K., Gctf: Real-time CTF determination and correction. *J Struct Biol* **2016**, *193* (1), 1-12.
72. Rosenthal, P. B.; Henderson, R., Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology* **2003**, *333* (4), 721-45.
73. Chen, S.; McMullan, G.; Faruqi, A. R.; Murshudov, G. N.; Short, J. M.; Scheres, S. H.; Henderson, R., High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **2013**, *135*, 24-35.
74. Urnavicius, L.; Zhang, K.; Diamant, A. G.; Motz, C.; Schlager, M. A.; Yu, M.; Patel, N. A.; Robinson, C. V.; Carter, A. P., The structure of the dynactin complex and its interaction with dynein. *Science* **2015**, *347* (6229), 1441-1446.
75. Kucukelbir, A.; Sigworth, F. J.; Tagare, H. D., Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **2014**, *11* (1), 63-5.
76. Bai, X. C.; Rajendra, E.; Yang, G.; Shi, Y.; Scheres, S. H., Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **2015**, *4*.
77. Stein, P. E.; Boodhoo, A.; Armstrong, G. D.; Cockle, S. A.; Klein, M. H.; Read, R. J., The crystal structure of pertussis toxin. *Structure* **1994**, *2* (1), 45-57.
78. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25* (13), 1605-12.
79. McLeod, R. A.; Diogo Righetto, R.; Stewart, A.; Stahlberg, H., MRCZ - A proposed fast compressed MRC file format and direct detector normalization strategies. *bioRxiv* **2017**.
80. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C., SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **2003**, *31* (13), 3381-5.
81. Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **2010**, *66* (1), 12-21.
82. Krissinel, E.; Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **2004**, *60* (Pt 12 Pt 1), 2256-68.
83. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007**, *372* (3), 774-97.

84. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **1994**, 50 (5), 760-763.
85. Kleywegt, G. J., Validation of protein models from Calpha coordinates alone. *J Mol Biol* **1997**, 273 (2), 371-6.
86. Kleywegt, G., Use of Non-crystallographic Symmetry in Protein Structure Refinement. *Acta Crystallographica Section D* **1996**, 52 (4), 842-857.
87. Kahraman, A.; Malmstrom, L.; Aebersold, R., Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **2011**, 27 (15), 2163-4.
88. Schrodinger, L., The PyMOL Molecular Graphics System, Version 1.7.0.3. 2010.
89. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **2011**, 7, 539.
90. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, 28 (12), 1647-9.

Chapter 4: Cercosporin biosynthesis: Expansion of the biosynthetic gene cluster and scope of production by plant pathogens

Adapted with permission from de Jonge, R.*, Ebert, M.K.*, Huitt-Roehl, C.R.*, Pal, P., Suttle, J.C., Neubauer, J.D., Jurick II, W.M., Secor, G.A., Thomma, B.P.H.J., Van de Peer, Y., Townsend, C.A., Bolton, M.D. Ancient duplication and horizontal transfer of a toxin gene cluster reveals novel mechanisms in the cercosporin biosynthesis pathway. *Manuscript in preparation*. *Indicates co-first-authorship

4.1. Introduction

The fungal plant pathogens *Cercospora spp.* infect hundreds of agriculturally and economically important crops¹, including soy bean, sugar beet, tobacco, maize, and rice, causing hundreds of millions of dollars in damage each year. Cercosporin **1**, a perylenequinone toxin produced by *Cercospora spp.*, has been shown to be the primary virulence factor involved in infectivity.

Cercosporin was first isolated from *C. kikuchii*, a pathogen of Japanese soy bean, in 1957. It was shown to be a crystalline red pigment with a quinone core, and proposed to be symmetrical, but with optical activity²⁻³. This interesting finding led to extensive investigation into the stereochemistry of cercosporin, ultimately revealing that it is indeed an atropisomer⁴. The helical configuration of this atropisomer was unambiguously determined by X-ray crystallography⁵. Cercosporin's perylenequinone core acts as a potent photosensitizer, with

quantum yields >0.8⁶. Following activation to a long-lived triplet state by visible light, cercosporin reacts with molecular oxygen, releasing reactive oxygen species (ROS), specifically singlet oxygen and superoxide radical⁷. The activity of cercosporin is photodynamic, and the relative production of singlet oxygen and superoxide is modulated by the presence of various reducing agents, allowing for a wider scope of reactivity⁸. Therefore, cercosporin does not have a specific cellular target, but rather causes non-specific damage to DNA, lipids, proteins, and other targets. For example, cercosporin-induced ROS have been shown to cause epoxidation of lipids⁹, resulting in cell membrane damage within minutes of exposure¹⁰ as well as changes in membrane composition and structure¹¹. Damage to the plant's cell membrane causes leakage of nutrients, feeding the fungus and allowing for further penetration of the leaves. Cercosporin is also toxic to mammals and bacteria, in a light- and oxygen-dependent manner¹².

Cercospora spp. resist the non-specific toxic effects of the cercosporin they produce through several mechanisms, primarily maintaining cercosporin in an inactive, reduced state (**2**) prior to active secretion (Figure 4.1). The quantum yield of **2** is dramatically lower than that of cercosporin (0.02-0.04 in water), significantly diminishing its ability to form ROS¹³. Cercosporin-resistant species of *Cercospora* and *Alternaria* were shown to contain primarily reduced cercosporin within their hyphae, while cercosporin-sensitive fungi were not

capable of reducing cercosporin¹⁴⁻¹⁵. To date, two cercosporin-specific resistance genes have been identified in cercosporin producers. Cercosporin facilitator protein (CFP), a major facilitator superfamily (MFS) transporter, is required for virulence and cercosporin secretion, as well as cercosporin resistance in *C. kikuchii*¹⁶⁻¹⁷. Heterologous expression of CFP in a cercosporin-sensitive fungus can induce cercosporin resistance¹⁸. Cercosporin resistance gene 1 (*crg1*), is a zinc-finger transcription factor involved in cercosporin resistance and biosynthesis¹⁹⁻²⁰. Introduction of these cercosporin resistance genes into crops could be a viable strategy to prevent or mitigate *Cercospora* infections without the use of harmful pesticides²¹⁻²⁴.

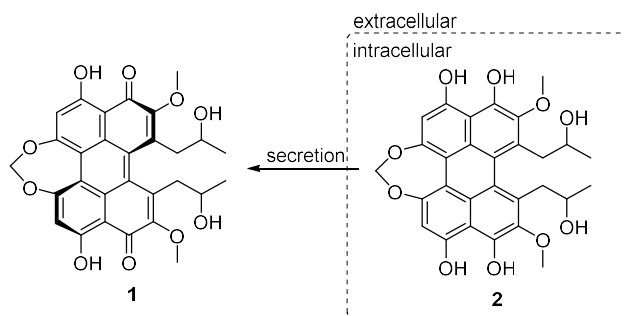


Figure 4.1. Cercosporin is maintained in a reduced state by producers until it is actively exported.

Despite extensive research on cercosporin resistance and mechanism of action, little is known about the chemical details of its biosynthesis. Cercosporin is a polyketide-derived natural product²⁵⁻²⁶, however, the complete biosynthetic pathway has yet to be conclusively elucidated. A series of oxidations, reductions, methylations, and a dimerization transform the polyketide synthase

(PKS) product into the final perylenequinone product. Many of the intermediates of the pathway are unstable and prone to oxidation, impeding elucidation of the pathway by traditional biological methods, such as analyzing accumulated intermediates in gene knockout strains²⁷.

The gene cluster was first identified in *C. nicotianae*, a pathogen of tobacco, in 2007²⁸. The cluster was found to contain six biosynthetic genes, a zinc-finger transcription factor (CTB8), and a MFS transporter involved in cercosporin secretion (CTB4)²⁹. These two regulatory and resistance genes are distinct from *crg1* and *CFP*, and are pathway specific. As expected, the previously identified non-reducing (NR)-PKS, CTB1, is included in the cluster²⁶. Also included in the cluster are CTB2, an *O*-methyltransferase³⁰; CTB3, a didomain *O*-methyltransferase/FAD-dependent monooxygenase³¹; and three oxidoreductases, CTB5-7³². All eight genes are regulated by light, and feedback inhibition of transcription of other genes in the cluster was shown in response to knockouts of each biosynthetic gene^{28, 32}. A tentative biosynthetic pathway was proposed based on putative functions of the six identified biosynthetic genes (Figure 4.2).

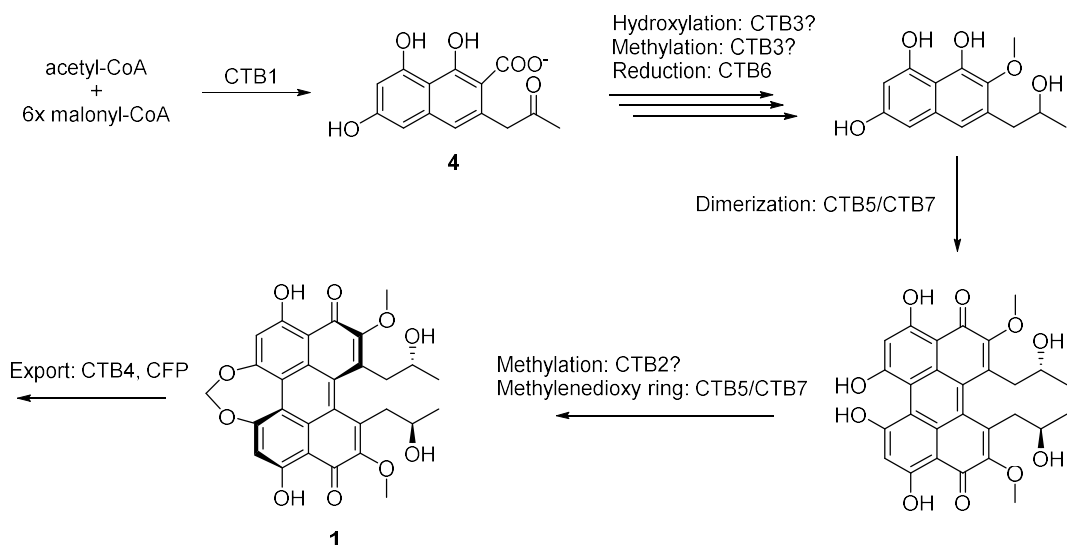


Figure 4.2. Proposed biosynthesis of cercosporin based on putative enzyme functions of cluster identified in *C. nicotianae* by Chen *et al* (2007).

However, the biosynthetic pathway had to be reconsidered when the on-pathway product of CTB1 was found to be *nor*-toralactone **3**, not the carboxylic acid **4**³³. *In vitro* characterization of reconstituted CTB1 showed that the NR-PKS homologates an acetyl starter unit and six malonyl extender units to form a heptaketide. The heptaketide is regiospecifically cyclized by the product template (PT) domain (C₄-C₉ and C₂-C₁₁) and finally released by the thioesterase (TE). Rather than more traditional TE chemistry (e.g. hydrolysis), the CTB1 TE catalyzes lactonization to form *nor*-toralactone. This lactonization mechanism was confirmed by ¹⁸O-acetyl-CoA labelling, where retention of the labelled acetyl carboxyl group was observed in the final product³³. The observed product of CTB1 is not consistent with the previously proposed biosynthetic pathway, requiring an updated proposal (Figure 4.3).

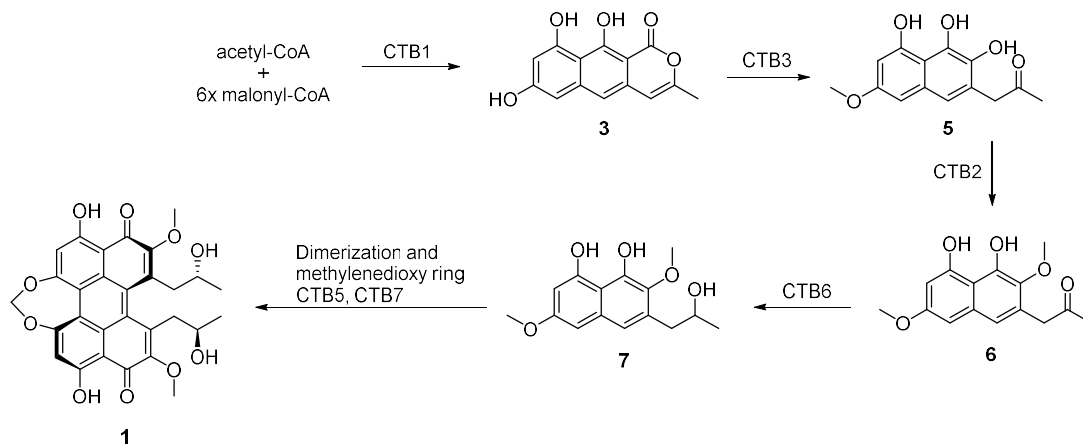


Figure 4.3. Revised biosynthetic proposal based on characterization of CTB1 and CTB3 by Newman *et al* (2016).

Analysis of intermediates accumulated by the knockouts of each biosynthetic gene in *C. nicotianae* allowed for a revised proposal of the biosynthetic pathway²⁷. Given the unexpected tricyclic structure of *nor*-toralactone 3, an enzyme is needed to open the pyrone ring. CTB3 was proposed to be the second enzyme in the pathway responsible for this ring opening step. CTB3 is an unusual didomain, bifunctional enzyme, containing a *N*-terminal *O*-methyltransferase and a *C*-terminal flavin-dependent monooxygenase. *In vitro* analysis confirmed CTB3 catalyzes two modifications of *nor*-toralactone, *O*-methylation and oxidative ring opening. A selection of unusual, off-pathway quinones were isolated from knockouts of CTB3, CTB5, and CTB6, while no soluble intermediates could be isolated from Δ CTB2 or Δ CTB7. These quinones were rationalized as spontaneous oxidation products of the true on-pathway intermediates, and a revised biosynthetic pathway was proposed (Figure 4.3),

where CTB2 methylates the product of CTB3 (5) to form the doubly-methylated (6), which is reduced by CTB6 to form the hydroxyl (7)²⁷. Finally, dimerization and methylenedioxy ring formation is proposed to be catalyzed by CTB5 and CTB7.

4.2. Results

4.2.1. Identification of new cercosporin biosynthetic genes

Following publication of this revised biosynthetic pathway, five new genes were discovered in the cercosporin gene cluster in *C. beticola*, a pathogen of sugar beets, using an evolutionary genomics approach (Figure 4.4). These genes were not identified with the original eight-gene cluster, as they are separated by two open reading frames that are not regulated by light or the transcription factor CTB8, and whose knockouts still produce cercosporin²⁸. One of the newly identified genes is cercosporin facilitator protein (CFP), a MFS transporter that has previously been identified in *C. kikuchii*, but thought to lie outside the cluster¹⁶. Additionally, four new biosynthetic genes were identified: an α -ketoglutarate dependent oxygenase (CTB9), an EthD-domain containing protein (CTB10), a β -IG-h3 fasciclin (CTB11), and a laccase (CTB12). All five genes abolish cercosporin production when knocked out in *C. beticola* (Figure 4.5).

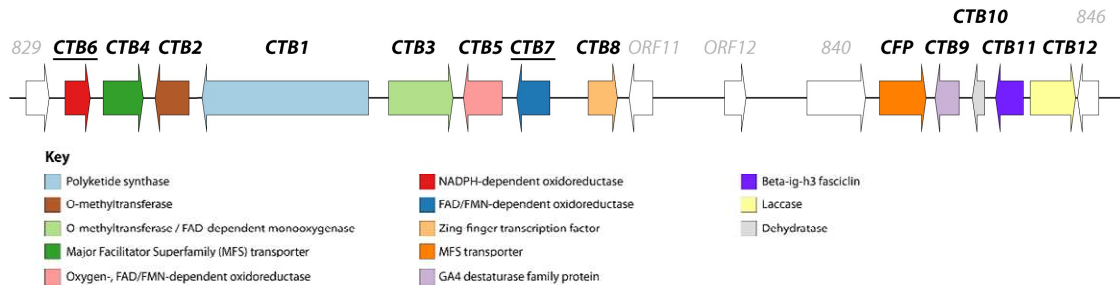


Figure 4.4. Revised cercosporin biosynthetic gene cluster in *C. beticola*

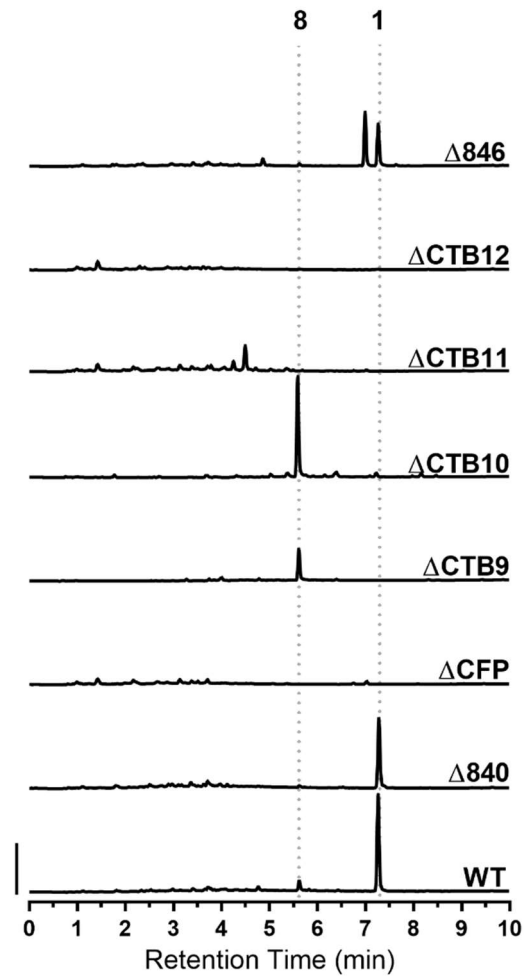


Figure 4.5. HPLC analysis of *C. beticola* mutants. 280 nm chromatograms are shown. Scale bar indicates 250 mAu.

4.2.2. Characterization of pre-cercosporin (8)

Two knockouts of the new biosynthetic genes (CTB9 and CTB10) accumulated a red, cercosporin-like pigment. However, this pigment had a different retention time by HPLC and migrated differently in potato dextrose agar (PDA). To characterize the red metabolite that accumulated in Δ CTB9 and Δ CTB10 mutants, an ethyl acetate extract of the collected mycelia was analyzed by reverse-phase HPLC. At 280 nm, a single peak was observed in both mutant extracts with identical retention times (Figure 4.5) and UV-Vis spectra (Figure 4.6b,c). This peak was compared to a reference sample of cercosporin produced by wild-type *C. beticola* (Figure 4.6a). The retention time of this peak was shorter than that of cercosporin suggesting a more polar metabolite. Comparison of the UV-vis spectra (Figure 4.6a-c) of the unknown compound and cercosporin revealed nearly identical chromophores, suggesting close structural relation. The exact mass of the metabolite from the mutants was determined (Δ CTB9: $m/z = 537.1762$, Δ CTB10: $m/z = 537.1757$, $[M+H^+]$), consistent with the elemental composition $C_{29}H_{28}O_{10}$. This mass is 2 Da greater than that of cercosporin (+2 hydrogens), which led to a proposed structure for pre-cercosporin (Figure 4.6d). Alternative hydroquinones of cercosporin could be excluded simply on the basis of the UV-vis spectral information and chemical instability. The presence of a free phenol in pre-cercosporin in place of the unusual 7-membered

methylenedioxy of cercosporin is consonant with the red shift of the long wavelength λ_{\max} and the shorter HPLC retention time.

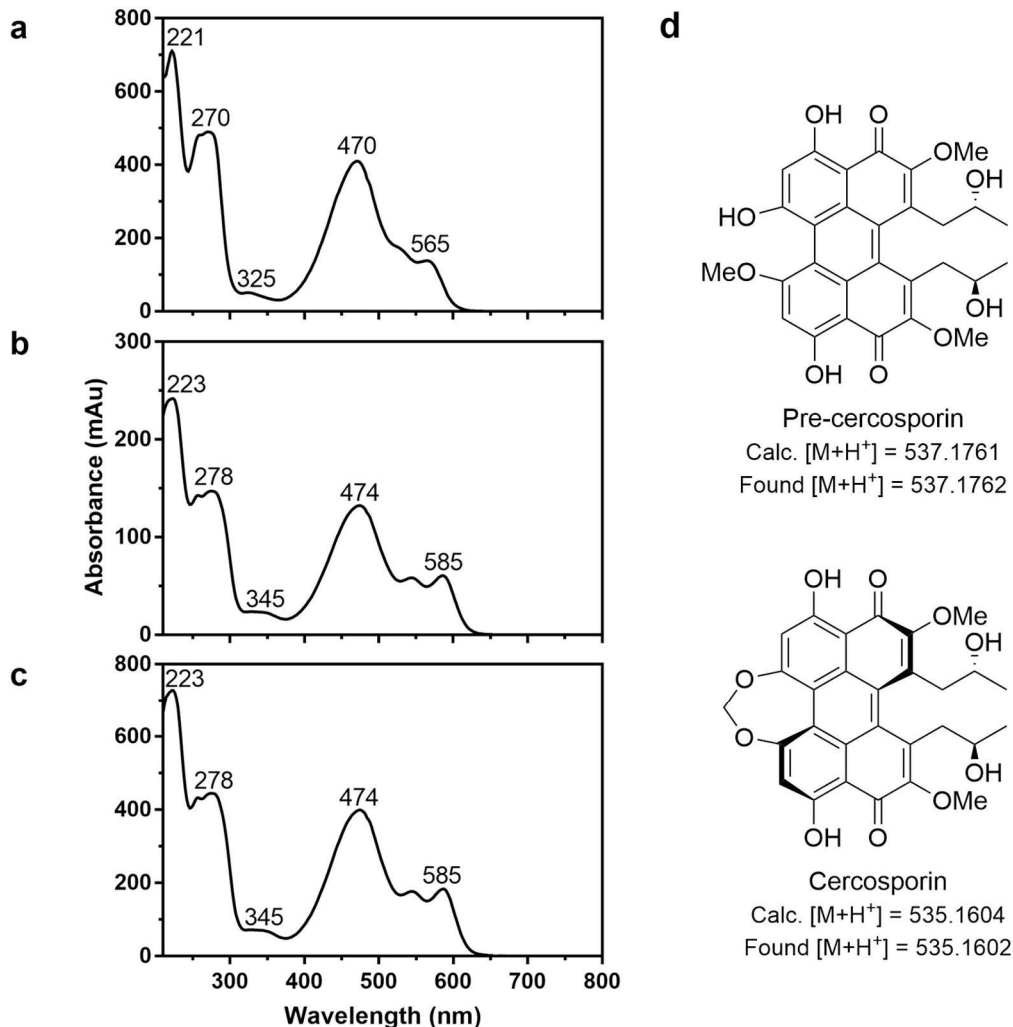


Figure 4.6. UV-vis spectra of cercosporin (a), pre-cercosporin produced by *C. beticola* Δ CTB9 (b), and pre-cercosporin produced by *C. beticola* Δ CTB10. Structures and mass spectrometry data for pre-cercosporin and cercosporin are shown in (d).

To firmly support the tentative structure of pre-cercosporin, the crude extract of Δ CTB9 was further purified by reverse-phase HPLC. Isolation of sufficient amounts of pre-cercosporin to characterize by NMR proved difficult

due to its instability and tendency to polymerize. The relative instability of pre-cercosporin compared to cercosporin suggests a possible role for the methylenedioxy bridge in overall stability. Several large-scale purifications were attempted by HPLC, but in all cases the small amount of pre-cercosporin obtained (< 1 mg in all cases) was too unstable to characterize by NMR. Attempts to acetylate the five free hydroxyls on pre-cercosporin were primarily unsuccessful. Complete acetylation of all positions was not achieved, and a mixture of pre-cercosporin with varied numbers of acetyl groups was isolated (Figure 4.7b). This mixture also appeared to be unstable, and could not be characterized by NMR. Extracts of liquid cultures of *C. beticola* Δ 842 were relatively complex, in contrast to extracts of solid cultures which consist of pre-cercosporin and little else, as determined by absorbance at 210 nm. However, liquid cultures can be grown at significantly larger scales than solid culture. Therefore, attempts to clean up the liquid culture extracts using size exclusion chromatography (Sephadex LH20) prior to HPLC were attempted. Unfortunately, this step did not result in more effective HPLC purification, likely due to pre-cercosporin degrading and/or polymerizing on the Sephadex LH20 column or during evaporation of the collected fractions (Figure 4.7c). Finally, we opted to isolate pre-cercosporin from solid cultures, and perform the extractions as quickly as possible in the dark (previous extractions were run for 24 h). This

method yielded a very clean crude extract (Figure 4.7a), and allowed isolation of enough pre-cercosporin for $^1\text{H-NMR}$ analysis.

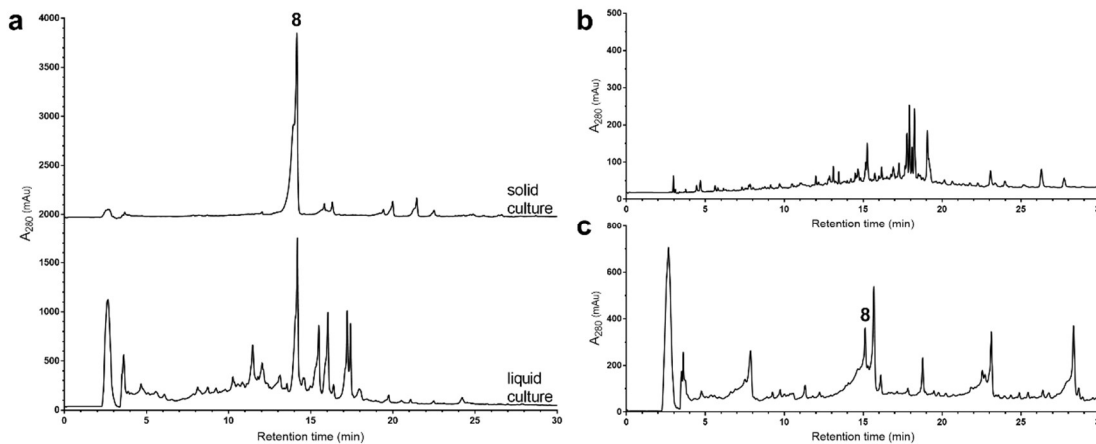


Figure 4.7. Large-scale HPLC purification of pre-cercosporin. Only extracts of solid cultures were clean enough to isolate pure pre-cercosporin. a) Solid (top) and liquid (bottom) cultures of *C. beticola* $\Delta 842$, extracted with ethyl acetate, quickly and in the dark. b) Crude *C. beticola* $\Delta 842$ extract acetylated with acetic anhydride. A mixture of acetylated pre-cercosporin products was obtained. c) *C. beticola* $\Delta 842$ extract purified by Sephadex LH20.

$^1\text{H-NMR}$ analysis confirmed the proposed structure of pre-cercosporin based on the earlier HPLC and UPLC-MS data. The spectrum lacked a methylenedioxy singlet at $\delta 5.74$, diagnostic of cercosporin, and instead contained a new methoxyl signal at $\delta 4.28$ and a phenol at $\delta 9.25$. While we did not isolate enough material to definitively determine the stereochemistry of the atropisomer by X-ray crystallography, the $^1\text{H-NMR}$ spectrum suggests a single atropisomer. Presumably the helical configuration of pre-cercosporin is identical to cercosporin, however we cannot rule out the possibility that the final stereochemistry is set by CTB9 or CTB10.

4.2.3. Identification of cercosporin production by *Colletotrichum fioriniae*

To date, cercosporin production has only been observed by *Cercospora spp.* and the closely related pathogen *Pseudocercospora capsellae*³⁴⁻³⁵. However, evolutionary genomics revealed the presence of cercosporin biosynthetic gene clusters in multiple species of *Colletotrichum*, another fungal genus with significant agricultural impacts, as well as *Sclerotinia sclerotiorum* and the rice pathogen *Magnaporthe oryzae*. *Colletotrichum*, like *Cercospora*, infects hundreds of agriculturally and economically important crops, including fruits such as apple, avocado, citrus, and banana. For example, *Colletotrichum fioriniae* is one of the top pre- and post-harvest pathogens of apples, causing significant losses in fresh and dried apple products³⁶.

Since our phylogenomic analyses suggested that several *Colletotrichum spp.* harbor cercosporin biosynthetic gene clusters, we questioned whether any *Colletotrichum spp.* can produce cercosporin. To initially assess this, two *Colletotrichum fioriniae* strains (HC89 and HC91) isolated from apple were assayed for cercosporin production using the KOH assay²⁶. No cercosporin-like pigment was observed in the media under the same conditions that stimulate cercosporin production in *C. beticola*. However, the addition of the epigenetic modifier trichostatin A³⁷, a histone deacetylase inhibitor, induced production of a red cercosporin-like compound into the medium. To characterize this red

metabolite, mycelia from both *Co. fiorinae* strains were extracted with ethyl acetate. Reverse-phase HPLC analysis as before revealed a peak with a retention time and UV-vis spectrum consistent with cercosporin in both extracts (Figure 4.8a, b). The presence of cercosporin was confirmed by UPLC-ESI-MS (Figure 4.8c).

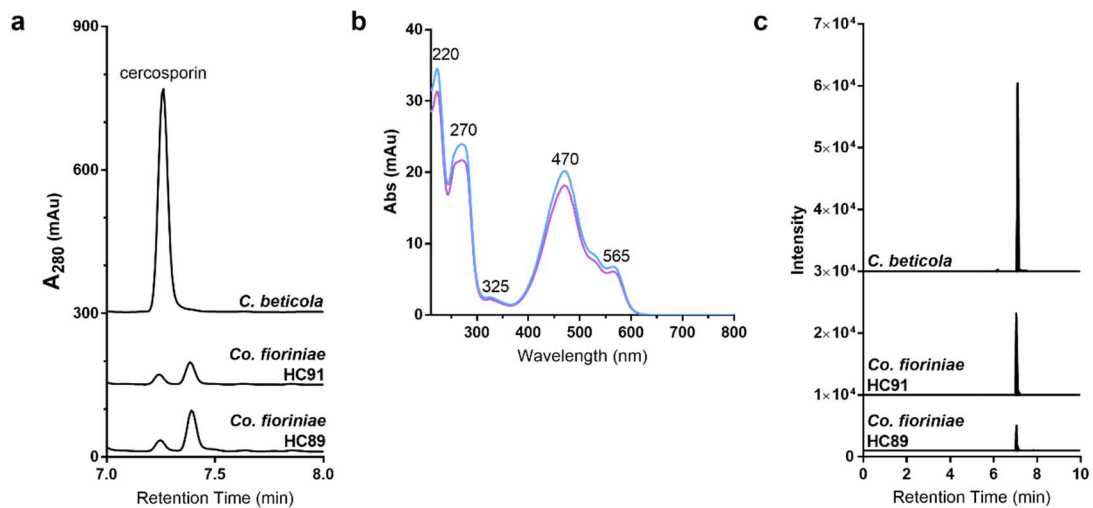


Figure 4.8. Analysis of *Colletotrichum fiorinae* extracts. a) HPLC chromatograms at 280 nm. b) UV-visual spectra of cercosporin produced by *Co. fiorinae* HC89 and HC91. c) Extracted ion chromatograms ($m/z = 535.1604$) of indicated extracts.

4.3. Discussion

Given the discovery of four new biosynthetic genes within the cercosporin gene cluster, the proposed biosynthetic pathway needs to be revised.

Unfortunately, instability of pathway intermediates and feedback inhibition of gene expression have impeded efforts to elucidate the middle steps of the

pathway, including the cryptic dimerization to form the unusual atropisomer. However, we have now identified the final step in the biosynthetic pathway, oxidative closure of the 7-membered methylenedioxy bridge. Isolation of precercosporin from both Δ CTB9 and Δ CTB10 shows that two genes are involved in this reaction, in contrast to 5-membered methylenedioxy bridge formation, which is catalyzed by a single P450 enzyme in plant alkaloid biosynthesis³⁸⁻³⁹.

CTB9 is an α -ketoglutarate dependent oxygenase, and presumably oxidizes the methoxyl position, analogous to the methylenedioxy forming P450s (Figure 4.9). Since in all known cases of methylenedioxy ring formation, only a single enzyme (a P450) is required, the role of CTB10 (an EthD-domain containing protein) is less clear. However, given the unfavorable 7-membered ring being formed, it may be required to facilitate closure of the ring. EthD-domain containing proteins are very small (~100 amino acids), poorly understood proteins, generally implicated as accessory proteins having no true catalytic function⁴⁰. For example, in aflatoxin biosynthesis, the EthD-domain containing HypE is required for the oxidation of **9** to **10** by a cytochrome P450 (Figure 4.10)⁴¹⁻⁴². Therefore, the possibility remains that CTB10's role is purely structural, to stabilize and/or activate CTB9 in some yet-to-be-elucidated mechanism.

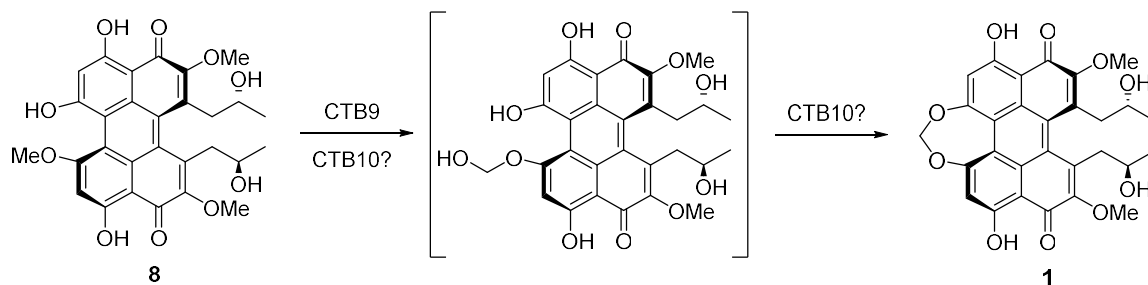


Figure 4.9. Proposed role of CTB9 and CTB10 in cercosporin biosynthesis.

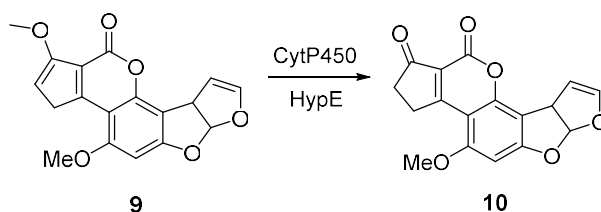


Figure 4.10. Role of an EthD-domain containing protein (HypE) in aflatoxin biosynthesis.

Two new additional biosynthetic genes have been identified in this study: CTB11, a putative β -IG-h3 fasciclin, and CTB12, a putative laccase. As no intermediates are accumulated in either knockout, the precise role of CTB11 and CTB12 remains unknown. Both enzymes are predicted to be extracellular, with small signal peptides, and, in the case of CTB11, a transmembrane domain. Fasciclin domain proteins are membrane-anchored, extracellular proteins involved in cell adhesion⁴³⁻⁴⁴. Laccases are multicopper oxidases involved in a variety of cellular processes, including fungal pigment biosynthesis, lignification, and delignification⁴⁵. One possible role for these extracellular enzymes is to facilitate adhesion of the fungus to the host plant, and subsequently degrade the plant's cell wall through delignification to allow cercosporin access to the cell

membrane. Alternatively, CTB12 may be involved in the dimerization step of cercosporin biosynthesis.

Characterization of the cercosporin biosynthetic pathway through identification of accumulated intermediates from gene knockouts has thus far proven to be problematic. With the exception of *nor*-toralactone **3** and pre-cercosporin **8**, all isolated intermediates have been apparent oxidation products of true on-pathway intermediates. Therefore, the true on-pathway intermediates have been logically inferred. Additionally, many knockouts of cercosporin biosynthetic genes do not accumulate isolable intermediates, likely due to feedback inhibition of expression of earlier genes in the pathway. Alternatively, the unstable, oxidatively-sensitive intermediates may be polymerizing, forming an insoluble compound(s) that is not readily extracted using traditional methods. Isolated pre-cercosporin appeared to dimerize over time (as detected by UPLC-MS), and earlier, monomeric intermediates in the pathway are likely to be even more susceptible to oxidative polymerization. Therefore, alternate methods, such as *in vitro* reconstitution of enzymes, need to be implemented to fully characterize cercosporin biosynthesis.

The identification of cercosporin production in two isolates of *Co. fioriniae* has significant implications for the apple packing, storage, and processing industries. Bitter rot, caused by *Colletotrichum spp.*, is one of the top pre- and

postharvest pathogens of apple³⁶. This disease is a major problem for the apple industry as it limits fresh fruit in the field and during storage, and has a quiescent stage allowing decay to occur on seemingly high quality apples, only to come out of storage rotten. Hence, contamination of processed apple products with cercosporin could be a significant health hazard. For example, other fungal-produced toxins (e.g. patulin, citrinin, penicillic acid) can contaminate processed apple products⁴⁶. Patulin, produced by *Penicillium spp.*, is the most troubling as it is carcinogenic and consequently the United States and Europe have strict patulin limits in fruit juices and processed pome fruit products⁴⁷. Future studies will focus on the role of cercosporin production during the *Colletotrichum*-apple fruit interactions in addition to assaying processed fruit products made from apples with bitter rot symptoms to determine levels of the toxin in fruit. Although only *Co. fioriniae* strains were analyzed for the ability to produce cercosporin, the identification of highly similar *CTB* clusters in other *Colletotrichum* species suggest that cercosporin production is wide-spread in this genus.

4.4. Experimental Methods

4.4.1. Pre-cercosporin isolation and characterization

Mycelial plugs of *C. beticola* $\Delta CTB9$ were placed on top of eight “thin” PDA (Difco) plates (3.0 mL PDA per 50 mm Petri plate). Cultures were incubated at 22 °C for one week under continuous light. Three separate methods were attempted to prepare crude secondary metabolite extractions. 1) PDA and mycelia were extracted with ethyl acetate for 4 min. The resulting supernatant was collected and frozen for further analysis. 2) PDA and mycelia were placed into a GenElute Maxiprep binding column (Sigma Aldrich) and centrifuged at 3500 $\times g$ for 10 min. The flow-through was collected and frozen for further analysis. 3) PDA and mycelia were placed into a GenElute Maxiprep binding column along with 15 mL ethyl acetate. After a 30 s incubation, the column was centrifuged at 2400 $\times g$ for 10 min. The flow-through was collected and frozen for further analysis. To obtain sufficient pre-cercosporin for isolation and NMR analysis, primary extracts from all three methods were combined.

The combined extracts were re-suspended with water and acidified with conc. HCl. Pre-cercosporin was extracted quickly from this aqueous solution by partitioning thrice with ethyl acetate in the dark, wrapping the glassware with aluminum foil. The combined ethyl acetate fractions were washed with brine, dried over anhydrous sodium sulfate and evaporated under vacuum at 30 °C.

The reddish-brown residue was resuspended in methanol and filtered through 0.2 μm PTFE filters. The methanol extracts were initially analyzed by reverse phase HPLC on an Agilent model 1200 fitted with a Kinetex XB-C18 column (4.6 mm x 75 mm, 2.6 μm , Phenomenex). Injections of 1 μl were run at 1.25 mL/min on a linear gradient of 5% solvent C/95% solvent D to 95% solvent C/5% solvent D over 10.8 min, where solvent C was 0.1% formic acid and solvent D was acetonitrile + 0.1% formic acid. Chromatograms were monitored at 436, 280, and 210 nm, and UV-vis spectra were recorded over a range of 210-800 nm. High-resolution mass data were obtained from a Waters Acquity/Xevo-G2 UPLC-ESI-MS in positive ion mode.

To isolate sufficient pre-cercosporin for $^1\text{H-NMR}$ analysis, the filtered methanol extract prepared above was purified by reverse-phase HPLC on an Agilent model 1100 fitted with a Kinetex XB-C18 semi-prep column (10 mm x 250 mm, 5 μm , Phenomenex). The crude extract (10 mg/mL in methanol) was injected (generally 500 μL) and run at 4 mL/min using the following method: 20% solvent C/80% solvent D for 3 min, 20-70% solvent C over 17 min, 70-95% solvent D over 5 min, where solvent C and D were as above. Chromatograms were recorded at 436, 280, and 210 nm. The metabolite of interest was collected from multiple injections, combined, and lyophilized to dryness. The purified

pre-cercosporin was analyzed by UPLC-ESI-MS as described above and ¹H-NMR.

4.4.2. Pre-cercosporin purification by Sephadex LH20

C. beticola Δ842 total culture mass (mycelia and agar) was ground under liquid nitrogen twice and lyophilized to dryness. The resulting powder was resuspended in acidified water (pH < 2) and extracted thrice with ethyl acetate. The ethyl acetate extractions were pooled, washed with brine, and evaporated to dryness. The crude extract was resuspended in methanol.

A Sephadex LH20 column (5.25 g) was prepared by swelling in excess methanol for five hours. The resin was poured into a glass fritted column (1.5 cm diameter) and packed at 3 mL/min until the volume stabilized. 300 μL aliquots of the crude extract were chromatographed over the column at 1.3 mL/min. Fractions were collected and analyzed for pre-cercosporin by HPLC on an Agilent 1200 as above. Fractions containing pre-cercosporin were pooled, evaporated to dryness, and resuspended in methanol for HPLC purification on an Agilent 1100 as above.

4.4.3. *Colletotrichum* spp. cercosporin assay

To determine whether *Colletotrichum* species were able to produce cercosporin, two monoconidial isolates *C. fioriniae* (HC89 and HC91) were grown on 9 cm Petri plates containing 15 mL of PDA as described above to replicate conditions that were conducive for cercosporin production *in vitro*. Seven-day old cultures of each isolate were grown in a temperature controlled incubator at 25 °C with natural light. A pinkish to dark red color was visible in the media for all isolates except HC75, which had a yellow-colored pigment. Using a #2 cork borer, three plugs were removed from each isolate from the edge, middle and center of each colony and placed in small screw cap glass vials. Three plugs were also removed from an uncolonized PDA plate and included as a negative control. Cercosporin (Sigma-Aldrich) was dissolved in acetone to 100 mM and used as a positive control. 5N KOH was added to each vial to cover the surface of the plugs and incubated on a shaking incubator at room temperature for 4 h. Supernatants were examined for cercosporin spectrophotometrically. To induce cercosporin production, we followed the procedures described by Shwab et al. except 10 µM trichostatin A (TSA; Sigma) was used³⁷. Cercosporin production by *C. fioriniae* HC89 and HC91 was confirmed by HPLC and UPLC-ESI-MS analysis as described above for pre-cercosporin.

4.5. References

1. Farr, D. F.; Bills, G. F.; Chamuris, G. P.; Rossman, A. Y., *Fungi on plants and plant products in the United States*. APS Press: Minnesota, 1989.
2. Kuyama, S., Tamura, T., Cercosporin. A pigment of *Cercosporina Kikuchii* Matsumoto et Tomoyasu, I. Cultivation of fungus, isolation and purification of pigment. *J Am Chem Soc* **1957**, 79 (21), 5725-6.
3. Kuyama, S., Tamura, T., Cercosporin. A pigment of *Cercosporina Kikuchii* Matsumoto et Tomoyasu. II. Physical and chemical properties of cercosporin and its derivatives. *J Am Chem Soc* **1957**, 79 (21), 5726-9.
4. Yamazaki, S., Ogawa, T., The chemistry and stereochemistry of cercosporin. *Agr Biol Chem* **1972**, 36 (10), 1707-18.
5. Nasini, G.; Merlini, L.; Andreetti, G. D.; Bocelli, G.; Sgarabotto, P., Stereochemistry of cercosporin. *Tetrahedron* **1982**, 38 (18), 2787-96.
6. Dobrowolski, D. C., Foote, C. S., Cercosporin, a singlet oxygen generator. *Angew Chem Int Ed Engl* **1983**, 22 (9), 720-1.
7. Daub, M. E.; Hangarter, R. P., Light-induced production of singlet oxygen and superoxide by the fungal toxin, cercosporin. *Plant Physiol* **1983**, 73 (3), 855-7.
8. Hartman, P. E.; Dixon, W. J.; Dahl, T. A.; Daub, M. E., Multiple modes of photodynamic action by cercosporin. *Photochem Photobiol* **1988**, 47 (5), 699-703.
9. Cavallini, L.; Bindoli, A.; Macri, F.; Vianello, A., Lipid peroxidation induced by cercosporin as a possible determinant of its toxicity. *Chem Biol Interact* **1979**, 28 (2-3), 139-46.
10. Daub, M. E., Peroxidation of tobacco membrane lipids by the photosensitizing toxin, cercosporin. *Plant Physiol* **1982**, 69 (6), 1361-4.
11. Daub, M. E.; Briggs, S. P., Changes in tobacco cell membrane composition and structure caused by cercosporin. *Plant Physiol* **1983**, 71 (4), 763-6.
12. Yamazaki, S. O., Akiyama, Y., Fuwa, K., Cercosporin, a novel photodynamic pigment isolated from *Cercospora kikuchii*. *Agric Biol Chem* **1975**, 39 (1), 287-8.
13. Daub, M. E.; Li, M.; Bilski, P.; Chignell, C. F., Dihydrocercosporin singlet oxygen production and subcellular localization: a possible defense against cercosporin phototoxicity in *Cercospora*. *Photochem Photobiol* **2000**, 71 (2), 135-40.
14. Daub, M. E.; Leisman, G. B.; Clark, R. A.; Bowden, E. F., Reductive detoxification as a mechanism of fungal resistance to singlet oxygen-generating photosensitizers. *Proc Natl Acad Sci U S A* **1992**, 89 (20), 9588-92.
15. Sollod, C. C.; Jenns, A. E.; Daub, M. E., Cell surface redox potential as a mechanism of defense against photosensitizers in fungi. *Appl Environ Microbiol* **1992**, 58 (2), 444-9.
16. Callahan, T. M.; Rose, M. S.; Meade, M. J.; Ehrenshaft, M.; Upchurch, R. G., CFP, the putative cercosporin transporter of *Cercospora kikuchii*, is required for wild type cercosporin production, resistance, and virulence on soybean. *Mol Plant Microbe Interact* **1999**, 12 (10), 901-10.

17. Upchurch, R. G.; Rose, M. S.; Eweida, M., Over-expression of the cercosporin facilitator protein, CFP, in *Cercospora kikuchii* up-regulates production and secretion of cercosporin. *FEMS Microbiol Lett* **2001**, *204* (1), 89-93.
18. Upchurch, R. G.; Rose, M. S.; Eweida, M.; Callahan, T. M., Transgenic assessment of CFP-mediated cercosporin export and resistance in a cercosporin-sensitive fungus. *Curr Genet* **2002**, *41* (1), 25-30.
19. Chung, K. R.; Jenns, A. E.; Ehrenshaft, M.; Daub, M. E., A novel gene required for cercosporin toxin resistance in the fungus *Cercospora nicotianae*. *Mol Gen Genet* **1999**, *262* (2), 382-9.
20. Chung, K. R.; Daub, M. E.; Kuchler, K.; Schuller, C., The CRG1 gene required for resistance to the singlet oxygen-generating cercosporin toxin in *Cercospora nicotianae* encodes a putative fungal transcription factor. *Biochem Biophys Res Commun* **2003**, *302* (2), 302-10.
21. Kuykendall, L. D.; Stockett, T. M.; Saunders, J. W., Rhizobium radiobacter conjugation and callus-independent shoot regeneration used to introduce the cercosporin export gene *cfp* from *Cercospora* into sugar beet (*Beta vulgaris* L.). *Biotechnol Lett* **2003**, *25* (9), 739-44.
22. Kuykendall, L. D.; Upchurch, R. G., Expression in sugar beet of the introduced cercosporin toxin export (CFP) gene from *Cercospora kikuchii*, the causative organism of purple seed stain in soybean. *Biotechnol Lett* **2004**, *26* (9), 723-7.
23. Upchurch, R. G.; Rose, M. S.; Eweida, M.; Zuo, W., Expression of the cercosporin transporter, CFP, in tobacco reduces frog-eye lesion size. *Biotechnol Lett* **2005**, *27* (20), 1543-50.
24. Panagiotis, M.; Kritonas, K.; Irini, N. O.; Kiriaki, C.; Nicolaos, P.; Athanasios, T., Expression of the yeast *cpd1* gene in tobacco confers resistance to the fungal toxin cercosporin. *Biomol Eng* **2007**, *24* (2), 245-51.
25. Okubo, A.; Yamazaki, S.; Fuwa, K., Biosynthesis of Cercosporin. *Agric Biol Chem* **1975**, *39* (5), 1173-5.
26. Choquer, M.; Dekkers, K. L.; Chen, H. Q.; Cao, L.; Ueng, P. P.; Daub, M. E.; Chung, K. R., The CTB1 gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of *Cercospora nicotianae*. *Mol Plant Microbe Interact* **2005**, *18* (5), 468-76.
27. Newman, A. G.; Townsend, C. A., Molecular Characterization of the Cercosporin Biosynthetic Pathway in the Fungal Plant Pathogen *Cercospora nicotianae*. *J Am Chem Soc* **2016**, *138* (12), 4219-4228.
28. Chen, H.; Lee, M. H.; Daub, M. E.; Chung, K. R., Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. *Mol Microbiol* **2007**, *64* (3), 755-70.
29. Choquer, M.; Lee, M. H.; Bau, H. J.; Chung, K. R., Deletion of a MFS transporter-like gene in *Cercospora nicotianae* reduces cercosporin toxin accumulation and fungal virulence. *FEBS Lett* **2007**, *581* (3), 489-94.
30. Staerckel, C.; Boenisch, M. J.; Kroger, C.; Bormann, J.; Schafer, W.; Stahl, D., CbCTB2, an O-methyltransferase is essential for biosynthesis of the phytotoxin

- cercosporin and infection of sugar beet by *Cercospora beticola*. *BMC Plant Biol* **2013**, *13*, 50.
31. Dekkers, K. L.; You, B. J.; Gowda, V. S.; Liao, H. L.; Lee, M. H.; Bau, H. J.; Ueng, P. P.; Chung, K. R., The *Cercospora nicotianae* gene encoding dual O-methyltransferase and FAD-dependent monooxygenase domains mediates cercosporin toxin biosynthesis. *Fungal Genet Biol* **2007**, *44* (5), 444-54.
 32. Chen, H. Q.; Lee, M. H.; Chung, K. R., Functional characterization of three genes encoding putative oxidoreductases required for cercosporin toxin biosynthesis in the fungus *Cercospora nicotianae*. *Microbiology* **2007**, *153* (Pt 8), 2781-90.
 33. Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A., Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun (Camb)* **2012**, *48* (96), 11772-4.
 34. Gunasinghe, N.; You, M. P.; Cawthray, G. R.; Barbetti, M. J., Cercosporin from *Pseudocercospora capsellae* and its critical role in white leaf spot development. *Plant Dis* **2016**, *100* (8), 1521-1531.
 35. Crous, P. W.; Braun, U.; Hunter, G. C.; Wingfield, M. J.; Verkley, G. J. M.; Shin, H.-D.; Nakashima, C.; Groenewald, J. Z., Phylogenetic lineages in *Pseudocercospora*. *Stud Mycol* **2013**, *75* (1), 37-114.
 36. Munir, M.; Amseden, B.; Dixon, E.; Vaillancourt, L.; Ward Gauthier, N. A., Characterization of *Colletotrichum* species causing bitter rot of apple in Kentucky orchards. *Plant Dis* **2016**, *100* (11), 2194-2203.
 37. Shwab, E. K.; Bok, J. W.; Tribus, M.; Galehr, J.; Graessle, S.; Keller, N. P., Histone deacetylase activity regulates chemical diversity in *Aspergillus*. *Eukaryot Cell* **2007**, *6* (9), 1656-64.
 38. Ikezawa, N.; Iwasa, K.; Sato, F., Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis in *Eschscholzia californica*. *FEBS J* **2007**, *274* (4), 1019-35.
 39. Diaz Chavez, M. L.; Rolf, M.; Gesell, A.; Kutchan, T. M., Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy *Argemone mexicana*. *Arch Biochem Biophys* **2011**, *507* (1), 186-93.
 40. Chauvaux, S.; Chevalier, F.; Le Dantec, C.; Fayolle, F.; Miras, I.; Kunst, F.; Beguin, P., Cloning of a genetically unstable cytochrome P-450 gene cluster involved in degradation of the pollutant ethyl tert-butyl ether by *Rhodococcus ruber*. *J Bacteriol* **2001**, *183* (22), 6551-7.
 41. Ehrlich, K. C., Predicted roles of the uncharacterized clustered genes in aflatoxin biosynthesis. *Toxins (Basel)* **2009**, *1* (1), 37-58.
 42. Ehrlich, K. C.; Chang, P. K.; Scharfenstein, L. L., Jr.; Cary, J. W.; Crawford, J. M.; Townsend, C. A., Absence of the aflatoxin biosynthesis gene, *norA*, allows accumulation of deoxyaflatoxin B1 in *Aspergillus flavus* cultures. *FEMS Microbiol Lett* **2010**, *305* (1), 65-70.

43. Elkins, T.; Hortsch, M.; Bieber, A. J.; Snow, P. M.; Goodman, C. S., *Drosophila fasciclin I is a novel homophilic adhesion molecule that along with fasciclin III can mediate cell sorting. J Cell Biol* **1990**, *110* (5), 1825-32.
44. Elkins, T.; Zinn, K.; McAllister, L.; Hoffmann, F. M.; Goodman, C. S., Genetic analysis of a *Drosophila* neural cell adhesion molecule: interaction of fasciclin I and Abelson tyrosine kinase mutations. *Cell* **1990**, *60* (4), 565-75.
45. Mayer, A. M.; Staples, R. C., Laccase: new functions for an old enzyme. *Phytochemistry* **2002**, *60* (6), 551-65.
46. Wright, S. A. I., Patulin in food. *Current Opinion in Food Science* **2015**, *5*, 105-109.
47. Puel, O.; Galtier, P.; Oswald, I. P., Biosynthesis and toxicological effects of patulin. *Toxins (Basel)* **2010**, *2* (4), 613-31.

Appendix A: General Experimental Methods

A.1. Cloning

When necessary, genes are isolated from genomic DNA by PCR using primers lacking any restriction enzyme cut sites. This allows for more efficient binding of the primers to the genomic DNA. Introns are identified based on NCBI gene prediction, as well as Softberry's FGENESH gene prediction algorithm. When available, cDNA sequences greatly simplify this task. cDNA constructs are then put together using overlap extension PCR or Gibson assembly. PCR products are gel purified by running on standard agarose gels and extracted.

New primers containing appropriate restriction enzyme cut sites, or with overlap to the target plasmid are then used to PCR amplify the constructed gene or domains for cloning into expression vectors. Both the gene to be inserted and the vector are digested with the appropriate restriction enzymes – generally NdeI and NotI for pET-24a and pET-28a – according to standard protocol. Digested vector must then be gel-purified to avoid self-ligation, while the insert can be gel- or PCR-purified. Insert and vector are then ligated together using T4 DNA ligase overnight at 16°C. Following purification of the ligation reaction by ethanol precipitation, the newly constructed vector is transformed into BL21(DE3) via

electroporation and plated on LB agar plates with antibiotic for selection (25 ug/mL kanamycin for pET vectors in BL21). Alternatively, Gibson assembly can be performed in lieu of restriction digestion and ligation¹⁰. Constructed plasmids are first screened using restriction digestion assays to ascertain whether the insert has been successfully ligated into the vector. Constructs are finally verified by sequencing at the Johns Hopkins University Sequencing and Synthesis Facility or Genewiz.

A.1.1. Polymerase Chain Reaction (PCR)

Typical PCRs were run under the following conditions: 0.2 mM each dNTP, 0.4 μ M forward primer, 0.4 μ M reverse primer, 3% DMSO, and 1 μ L template (miniprep plasmid) in Phusion HF Buffer (New England Biolabs). The PCR was performed in an Eppendorf thermocycler, starting with a 2 min hold at 94° C. 1 μ L Phusion polymerase (New England Biolabs) was added midway through this initial hold. 30 cycles were performed, with a denaturing step of 15 s at 94 °C, an annealing step of X s at 63 °C (where X = 30 s/kb of desired product), and an elongation step of 30 s at 72 °C, followed by a 5 min hold at 72 °C. If necessary, annealing temperatures were optimized from 50-65 °C. PCR products were visualized by agarose gel (typically 1.1%) and purified with a GeneJET Gel Extraction Kit (ThermoScientific).

A.1.2. Plasmid assembly by restriction digestion and ligation

Inserts (purified PCR products) and vectors (plasmids miniprep using a GeneJET Plasmid Miniprep Kit (Thermo Scientific)) were digested using the appropriate restriction enzyme (New England Biolabs) as directed. Complete digestion was confirmed by agarose gel, and digested fragments were purified with a GeneJET Gel Extraction Kit (Thermo Scientific). Digested fragments were ligated using T4 DNA Ligase (New England BioLabs) as directed. The ligated plasmid was purified by ethanol precipitation. 2 μ L pellet paint (Millipore), 10 μ L 3M sodium acetate, and 20 μ L ligation reaction were added to 200 μ L ice-cold ethanol. The mixture was vortexed thoroughly and centrifuged at 14,000 rpm for 2 min. The supernatant was removed and the pellet resuspended in 200 μ L ice cold ethanol. The DNA was re-pelleted by another 2 min centrifugation. The supernatant was removed and the pellet thoroughly dried in a Speedvac. The purified DNA was resuspended in 20 μ L sterile ddH₂O and transformed into *E. coli* BL21(DE3) by electroporation.

A.1.3. Plasmid assembly by Gibson assembly

Primers to amplify fragments for Gibson assembly were designed using NEBuilder or SnapGene. Resulting PCR products were purified using a GeneJET PCR Purification Kit (Thermo Scientific). Similarly, the target vector

was digested using appropriate restriction enzymes (generally NdeI and NotI-HF), as described above, and PCR purified. DNA concentrations were measured by A_{260} using a Cary 50 UV-Vis spectrophotometer. Equimolar amounts of each insert and vector were combined with Gibson Master Mix to reach the following final concentrations: 0.08 U T4 exonuclease (New England Biolabs), 0.5 U Phusion polymerase (New England Biolabs), 80 U Taq DNA ligase (New England Biolabs), 5% PEG-8000, 100 mM Tris HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 0.2 mM each dNTP, 1 mM NAD. Gibson reactions were incubated at 50 °C for 1 h, diluted 5-fold, and transformed into *E. coli* BL21(DE3) by electroporation.

A.1.4. Transformation of *E. coli* by electroporation

Plasmid (2 μ L) was added to 18 μ L aliquots of electro-competent *E. coli* BL21(DE3) on ice. This mixture was applied to chilled Potter cuvettes and pulsed with a 10 kV/cm electric field for 10 ms at 4 °C. The cell mixture was then incubated in 180 μ L LB at 37 °C for 1 h, with shaking. The resulting inoculum was applied to LB agar plates with appropriate antibiotic, and incubated overnight at 37 °C. When transforming purified plasmid, the inoculum was streaked using a sterile loop, while 50-100 μ L of transformations of ligation or Gibson reactions were spread on the plate with a sterile cell-spreader.

A.2. Heterologous protein expression

E. coli BL21 (DE3) cells transformed with the gene of interest were grown at 37 °C while shaking in LB supplemented with 25 µg/mL kanamycin. Once the OD₆₀₀ reached 0.6, flasks were cold shocked in ice water for one hour, until cold to the touch. Protein expression was then induced by addition of 0.5 mM IPTG. Cells were induced overnight while shaking at 17 °C. Cells were then harvested by centrifuging at 4000xg for 20 minutes and typically flash frozen in liquid nitrogen for storage at -80 °C.

A.3. Protein purification

Cell pellets were thawed and resuspended in a lysis buffer consisting of 50 mM K/PO₄, pH 8, 300 mM NaCl, and 10% glycerol (generally 4 mL lysis buffer per gram of cell pellet). Cells were lysed by sonication, and the resulting lysate was cleared by centrifuging at 27,000xg for 30 min. Cleared lysate was batch bound to Ni²⁺-NTA or Co²⁺-TALON resin (generally 1 mL resin/1 L culture, depending on efficiency of expression) for one hour at 4 °C. The resin was collected by centrifugation at 50xg, washed with lysis buffer and applied to a fritted gravity flow column. The resin was washed and then protein eluted by a series of washes of increasing imidazole concentration. Purified protein was dialyzed overnight at 4 °C in appropriate buffer (*e.g.* 100 mM K/PO₄ pH 7 with

10% glycerol for *in vitro* reactions). A minimum of 100X total protein volume of dialysis buffer was used, and proteins were switched into fresh buffer at least once. Protein concentration was measured by absorbance at 280 nm using a Cary 50 UV-Vis spectrophotometer. Extinction coefficients were calculated using ExPasy ProtParam. Purity was observed by running all purification fractions on SDS-PAGE gels. If necessary, protein was concentrated using Amicon 3K or 10K MWCO Ultracentrifuge filters (Millipore). Most proteins can be flash frozen in liquid nitrogen for storage at -80 °C.

A.4. Standard medium recipes

A.4.1. Luria-Bertani Broth (LB)

10 g/L tryptone, 5 g/L yeast extract, and 5 g/L sodium chloride were combined in ddH₂O and autoclaved for 25 min. For agar plates, 7.5 g/L agar was added prior to autoclaving. Autoclaved media was allowed to cool to approximately 50 °C prior to addition of antibiotic. Plates were then poured and allowed to set thoroughly prior to storage at 4 °C.

A.4.2. Terrific Broth (TB)

For each liter of culture 24 g yeast extract, 12 g tryptone, and 8 mL 50% glycerol were combined in 900 mL ddH₂O and autoclaved for 25 min. Potassium phosphate buffer was prepared and autoclaved separately: 125.4 g/L K₂HPO₄ and

23.1 g/L KH_2PO_4 . Once sufficiently cooled (approximately 37 °C), 100 mL potassium phosphate buffer was added to each 900 mL of media using a sterile graduated cylinder.

Appendix B: Supplementary Material to Chapter 2

B.1. Cloning and mutagenesis

All plasmids used in this study are summarized in Table B.1. Cloning of plasmids to express PksA-SAT-KS-MAT (pENKA4) and PT-ACP (pEPTACP), as well as a series of active site mutants of PksA-SAT-KS-MAT (see table) has been described previously¹⁻². Additional mutants and PksA-SAT-KS-MAT-PT-ACP were cloned for this study using standard methods. DNA manipulations were carried out in *E. coli* BL21(DE3). PksA-SAT-KS-MAT-PT-ACP was amplified from a pET-28a vector containing full-length PksA cDNA using pksA5-NheI and ACP-3-NotI primers. The resulting PCR product was ligated into pET28a at NheI and NotI sites to give pEPksA-SKMPA. For the SAT active site mutants, overlap extension PCR was used to introduce specific point mutations at Cys117. These fragments were amplified from pENKA4 using the primers listed in Table B.2. PCR products were ligated into pET-28a at NcoI and NotI sites to give the plasmids listed in Table B.1. All expression constructs were confirmed by automated sequencing (Johns Hopkins University Synthesis and Sequencing Facility, Baltimore MD).

Appendix Table B.1. Plasmids used in Chapter 2

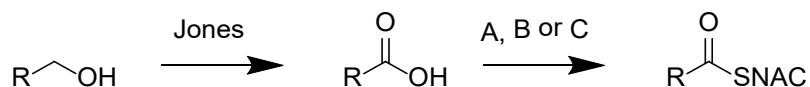
Plasmid	Protein	Tag	Mol. Wt (kDa)	Reference
pENKA4	SAT-KS-MAT	C-His ₆	144	1
pEPT-ACP	PT-ACP	C-His ₆	59	1
pENKA4-C117A	SAT ⁰ -KS-MAT	C-His ₆	144	2
pENKA4-C117S	SAT ⁰ -KS-MAT	C-His ₆	144	This study
pENKA4-C117G	SAT ⁰ -KS-MAT	C-His ₆	144	This study
pENKA4-C117N	SAT ⁰ -KS-MAT	C-His ₆	144	This study
pEPksA-SKMPA	SAT-KS-MAT-PT-ACP	N-His ₆ & C-His ₆	201	This study

Appendix Table B.2. Primers used in Chapter 2

Primer	Sequence 5' to 3'
NT5.1	gatccatggctcaatcaaggcaactc
MAT3.4	gagtgcggccgcggatggacggt
C117S-5'	gtgctgggattcagcatgggtccttg
C117S-3'	caaggaacctatgctgaatcccagcac
C117G-5'	gtgctgggattcggcaggggtccttg
C117G-3'	caaggaacctatgccgaatcccagcac
C117N-5'	cgtgctgggattcaacatgggtccttgcc
C117N-3'	caaggaacctatgtgaatcccagcac
pksA5-NheI	gctagcatggctcaatcaaggc
ACP-3-NotI	gtaagcggccgcctagccagcatccccgcttc

B.2. Synthesis of acyl-SNAC substrates

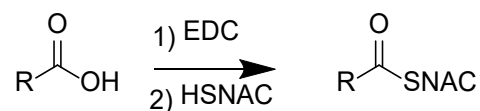
All reactions, unless otherwise stated, were performed under a positive N₂ or Ar atmosphere, in anhydrous, freshly distilled solvents. Commercially available compounds were purchased from Sigma-Aldrich or Alfa-Aesar and were used without further purification. All ¹H and ¹³C-NMR spectra were acquired on a 300 MHz or 400 MHz Bruker spectrometer and are reported in parts per million (δ) referenced against a TMS standard or residual solvent peak. All NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. Column chromatography was carried out on Silica Gel 60 (Sorbent Technologies, 200 x 400 mesh). Exact masses were determined on an Analytical VG-70SE Magnetic Sector Mass Spectrometer at the Johns Hopkins University Chemistry Department Mass Spectrometry Facility. ESI-MS spectra were measured at the Mass spectrometry facility of Old Dominion University in Norfolk, VA on a Bruker Apex-Qe Qh-FTMS instrument.



To a room temperature solution of the alcohol in acetone (200 mM), Jones reagent was added dropwise until the red color persisted. After 15 min, the

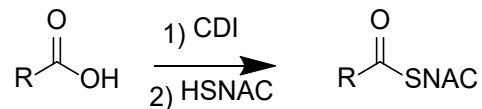
reaction was quenched with isopropanol, filtered, concentrated, dissolved (5% HCl, CH₂Cl₂), separated, dried and concentrated to afford the desired carboxylic acid (~90%).

Method A:



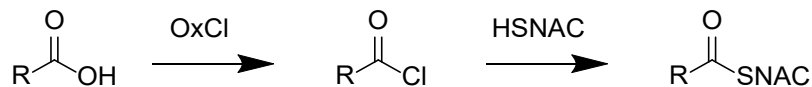
To a room temperature suspension of *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (1 eq.) in CH₂Cl₂, under N₂, Et₃N (1 eq.) was added. The carboxylic acid (1 eq.) was added dropwise followed by the addition of 4-(dimethylamino)pyridine (0.1 eq.). After 10 min. *N*-acetylcysteamine (1 eq.) was added dropwise. The resulting solution was allowed to react for 12-14 h then poured into a saturated solution of NH₄Cl and extracted with CH₂Cl₂ (3×). The combined organic phases were washed with H₂O and brine, dried (Na₂SO₄), concentrated and purified by column chromatography to afford the corresponding thioester.

Method B:



A room temperature solution of carboxylic acid (1 eq.) and CDI (1.4 eq.) in dry THF were allowed to react for 10 min. A solution of *N*-acetylcysteamine (1.1 eq.) was added and the reaction was allowed to stir at room temperature for 12-14 h. The solvent was removed under vacuum and the residue was redissolved in 10 mL of EtOAc, washed (10 ml of brine), dried (Na₂SO₄), concentrated and purified by column chromatography to furnish the respective SNAC-derivative.

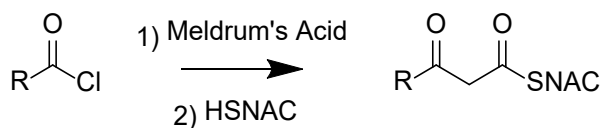
Method C:



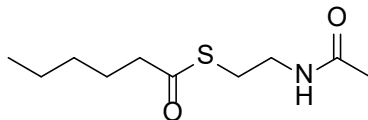
To a 0 °C solution of the carboxylic acid (1 equiv) and DMF (cat) in CH₂Cl₂ (150 mM), oxalyl chloride (1.1 equiv) was added slowly. The ice bath was removed and the mixture was allowed to warm to room temperature over 2 h, then concentrated *in vacuo* and used immediately without further purification. To a 0 °C solution of *N*-acetylcysteamine (1.0 equiv) and DIPEA (1.1 equiv) in CH₂Cl₂ (350 mM), a 0 °C solution of the acid chloride (1.0 equiv) in CH₂Cl₂ (500 mM) was added slowly, and the mixture was allowed to warm to room

temperature over 12 h. The resulting light orange solution was washed (NH₄Cl, sat, 1x, CuSO₄, sat, 1x, NH₄Cl, sat, 3x), filtered through a plug of silica gel, concentrated and recrystallized or purified by column chromatography.

Method D:

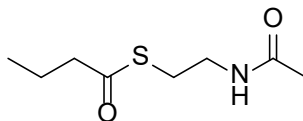


To a 0 °C solution of Meldrum's acid (1.0 equiv) and pyridine (1.0 equiv) in CH₂Cl₂ (350 mM), a 0 °C solution of the acid chloride (1.0 equiv) in CH₂Cl₂ (350 mM) was added. The resulting red solution was allowed to react at 0 °C for one hour, then warm to room temperature for 2 h. The mixture was concentrated *in vacuo*, then dissolved in toluene (200 mM). *N*-acetylcysteamine (1.0 equiv) was added and the red solution was heated to reflux for 2 h, cooled, washed (NH₄Cl, sat, 3x, CuSO₄, sat, 2x), dried, concentrated and recrystallized to afford the desired β-keto thioester.



S-(2-acetamidoethyl) hexanethioate (5)

Synthesis previously described³.



S-(2-acetamidoethyl) butanethioate (6)

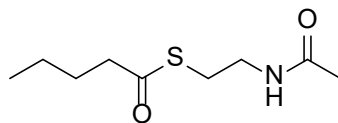
Method B

¹H NMR (400 MHz, CDCl₃): δ 6.17 (br s, 1H), 3.38 (app q, *J* = 6.3, 2H), 2.98 (t, *J* = 6.5, 2H), 2.51 (t, *J* = 7.4, 2H), 1.92 (s, 3H), 1.65 (sext, *J* = 7.4, 2H), 0.91 (t, *J* = 7.4, 3H).

¹³C NMR (101 MHz, CDCl₃): δ 200.0, 170.4, 46.0, 39.7, 28.4, 23.2, 19.2, 13.5. HR-

ESI-MS: 212.0706 ([*M* + Na]⁺, (C₈H₁₅NO₂S)Na⁺; calc. 212.0716). Spectral data

matched that reported by Patel *et. al.*⁴.

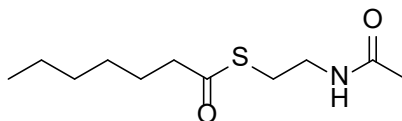


S-(2-acetamidoethyl) pentanethioate (7)

Method A

¹H NMR (300 MHz, CDCl₃): δ 6.48 (br s, 1H), 3.32 (app q, *J* = 6.3 Hz, 2H), 2.94 (t, *J* = 6.6 Hz, 2H), 2.49 (t, *J* = 7.2 Hz, 2H), 1.89 (s, 3H), 1.57 (pent, *J* = 7.2 Hz, 2H), 1.25 (hex, *J* = 7.2 Hz, 2H), 0.83 (t, *J* = 7.2 Hz, 3H). ¹³C NMR (75 MHz, CDCl₃): δ 200.1,

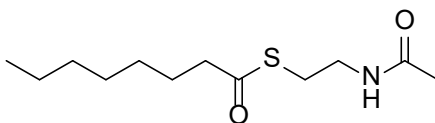
170.5, 43.8, 39.7, 28.4, 27.7, 23.1, 22.07, 13.7. HR-ESI-MS calculated exact mass for $C_9H_{18}NO_2S$: 204.1058, found: 204.1062 $[M+H]^+$. Spectral data matched that reported by Piasecki *et. al.*⁵.



S-(2-acetamidoethyl) heptanethioate (8)

Method A

1H -NMR (400 MHz, $CDCl_3$): δ 6.07 (br s, 1H), 3.66 (app q, $J = 6.8$ Hz, 2H), 2.98 (t, $J = 6.8$ Hz, 2H), 2.53 (t, $J = 7.6$ Hz, 2H), 1.93 (s, 3H), 1.61 (pent, $J = 7.6$ Hz, 2H), 1.25 (br m, 6H), 0.84 (t, $J = 7.6$ Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 200.4, 170.4, 44.3, 39.9, 31.5, 28.7, 28.6, 25.9, 23.3, 22.6, 14.1. HR-ESI-MS: 254.1181 ($[M + Na]^+$, $C_{11}H_{21}NO_2SNa^+$; calc. 254.1185). Spectral data matched that reported by Prasad *et. al.*⁶.

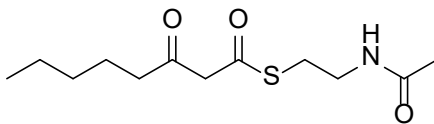


S-(2-acetamidoethyl) octanethioate (9)

Method C

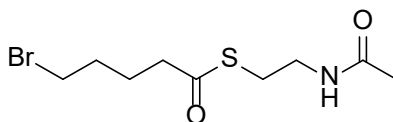
1H NMR (400 MHz, $CDCl_3$): δ 5.89 (br s, 1H), 3.46 (app q, $J = 6.2$ Hz, 2H), 3.09 (t, $J = 6.3$ Hz, 2H), 2.52 (t, $J = 7.4$ Hz, 2H), 1.97 (s, 3H), 1.59 (pent, $J = 7.3$ Hz, 2H), 1.34–1.27 (m, 8H), 0.89 (t, $J = 6.8$ Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 202.5, 192.6,

170.5, 57.3, 43.6, 39.4, 31.3, 29.4, 23.3, 23.3, 22.5, 14.01. HRFAB m/z calculated exact mass for $C_{12}H_{24}NO_2S$ 246.1527, found 246.1525 $[M+H]^+$. Spectral data matched that reported by Tse *et. al.*⁷.



S-(2-acetamidoethyl) 3-oxooctanethioate (10)

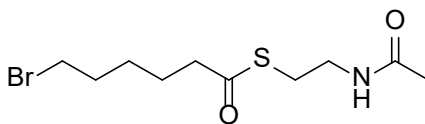
Synthesis previously described⁸.



S-(2-acetamidoethyl) 5-bromopentanethioate (11)

Method B: 76% of a colorless oil.

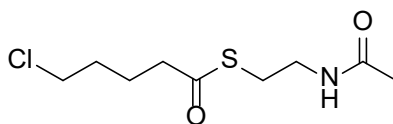
1H NMR (400 MHz, $CDCl_3$): δ 6.05 (br s, 1H), 3.36 (m, 4H), 2.98 (t, $J = 8.0$ Hz, 2H), 2.55 (t, $J = 8.0$ Hz, 2H), 1.92 (s, 3H), 1.82 (m, 4H). ^{13}C NMR (101 MHz, $CDCl_3$): δ 199.3, 170.4, 42.9, 39.6, 32.8, 31.7, 28.5, 24.1, 23.2. HRFAB m/z calculated exact mass for $C_9H_{17}NO_2S^{79}Br$ 282.0163, found 282.01662 $[M+H]^+$ (^{79}Br); calculated exact mass for $C_9H_{17}NO_2S^{81}Br$ 284.0142, found 284.0146 $[M+H]^+$ (^{81}Br). See Figure B.32 for NMR spectra.



S-(2-acetamidoethyl) 6-bromohexanethioate (12)

Method B: 38% of a colorless oil.

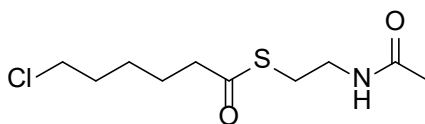
^1H NMR (400 MHz, CDCl_3): δ 5.88 (br s, 1H), 3.39 (app q, $J = 6.2$ Hz, 2H), 3.34 (t, $J = 6.7$ Hz, 2H), 2.99 (t, $J = 6.7$ Hz, 2H), 2.55 (t, $J = 6.7$ Hz, 2H), 1.92 (s, 3H), 1.83 (m, 2H), 1.67 (m, 2H), 1.44 (m, 2H). ^{13}C NMR (101 MHz, CDCl_3): δ 199.6, 170.2, 43.7, 39.6, 33.3, 32.3, 28.5, 27.4, 24.7, 23.2. HRFAB m/z calculated exact mass for $\text{C}_{10}\text{H}_{19}\text{NO}_2\text{S}^{79}\text{Br}$ 296.0319, found 296.0324 $[\text{M}+\text{H}]^+$ (^{79}Br); calculated exact mass for $\text{C}_{10}\text{H}_{19}\text{NO}_2\text{S}^{81}\text{Br}$ 298.0299, found 298.0305 $[\text{M}+\text{H}]^+$ (^{81}Br). See Figure B.33 for NMR spectra



S-(2-acetamidoethyl) 5-chloropentanethioate (13)

Method B: 99% of a colorless oil.

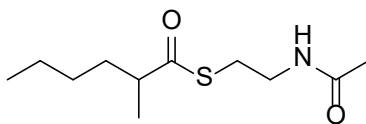
^1H NMR (400 MHz, CDCl_3): δ 5.91 (br s, 1H), 3.51 (m, 2H), 3.41 (app q, $J = 6.2$ Hz, 2H), 3.01 (t, $J = 6.6$ Hz, 2H), 2.58 (m, 2H), 1.92 (s, 3H), 1.80 (m, 4H). ^{13}C NMR (101 MHz, CDCl_3): δ 199.3, 170.2, 44.2, 43.1, 39.6, 31.6, 28.6, 23.2, 22.8. HRFAB m/z calculated exact mass for $\text{C}_9\text{H}_{17}\text{NO}_2\text{S}^{35}\text{Cl}$ 238.0668, found 238.0676 $[\text{M}+\text{H}]^+$ (^{35}Cl); calculated exact mass for $\text{C}_9\text{H}_{17}\text{NO}_2\text{S}^{37}\text{Cl}$ 240.0639, found 240.0639 $[\text{M}+\text{H}]^+$ (^{37}Cl). See Figure B.34 for NMR spectra



S-(2-acetamidoethyl) 6-chlorohexanethioate (14)

Method B: 97% of a colorless oil.

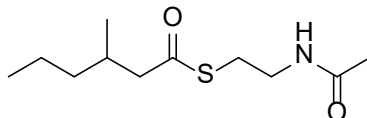
^1H NMR (400 MHz, CDCl_3): δ 5.91 (br s, 1H), 3.49 (t, $J = 6.6$ Hz, 2H), 3.39 (app q, $J = 6.2$ Hz, 2H), 3.00 (t, $J = 6.6$ Hz, 2H), 2.57 (t, $J = 7.4$ Hz, 2H), 1.93 (s, 3H), 1.80 – 1.60 (m, 4H), 1.45 (m, 2H). ^{13}C NMR (101 MHz, CDCl_3): δ 199.8, 170.4, 44.7, 43.8, 39.7, 32.1, 28.5, 26.1, 25.4, 22.9. HRFAB m/z calculated exact mass for $\text{C}_{10}\text{H}_{19}\text{NO}_2\text{S}^{35}\text{Cl}$ 252.0825, found 252.0823 $[\text{M}+\text{H}]^+$ (^{35}Cl); calculated exact mass for $\text{C}_{10}\text{H}_{19}\text{NO}_2\text{S}^{37}\text{Cl}$ 254.0796, found 254.0794 $[\text{M}+\text{H}]^+$ (^{37}Cl). See Figure B.35 for NMR spectra



S-(2-acetamidoethyl) 2-methylhexanethioate (15)

Method A: 87 % of a colorless oil

^1H NMR (400 MHz, CDCl_3): δ 5.84 (br s, 1H), 3.43 (app q, $J = 6.1$ Hz, 2H), 3.02 (t, $J = 6.4$ Hz, 2H), 2.64 (hex, $J = 6.8$ Hz, 1H), 1.96 (s, 3H), 1.70 (m, 1H), 1.42 (m, 1H), 1.29 (m, 4H), 1.11 (d, $J = 6.8$ Hz, 3H), 0.89 (t, $J = 6.8$ Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 204.9, 170.4, 48.9, 40.1, 34.0, 29.5, 28.3, 23.4, 22.8, 17.9, 14.1. HR-ESI-MS: 254.1182 ($[\text{M} + \text{Na}]^+$, $\text{C}_{11}\text{H}_{21}\text{NO}_2\text{SNa}^+$; calc. 254.1185). See Figure B.36 for NMR spectra

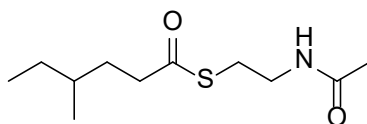


S-(2-acetamidoethyl) 3-methylhexanethioate (16)

Method C: 95% of a colorless oil.

^1H NMR (400 MHz, CDCl_3): δ 5.87 (br s, 1H), 3.44 (app q, $J = 6.1$ Hz, 2H), 3.02 (t, $J = 6.6$ Hz, 2H), 2.58 (m, 2H), 1.98 (s, 3H), 1.71 (m, 1H), 1.46 (m, 1H), 1.35 (m, 2H), 1.16 (m, 1H), 0.87 (t, $J = 7.2$ Hz, 3H), 0.87 (d, $J = 6.8$ Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 200.0, 170.4, 51.4, 39.9, 38.8, 30.9, 28.4, 23.2, 19.9, 19.5, 14.1. HRFAB m/z calculated exact mass for $\text{C}_{11}\text{H}_{21}\text{NO}_2\text{S}$ 232.1371, found 232.1369 $[\text{M}+\text{H}]^+$. See

Figure B.37 for NMR spectra

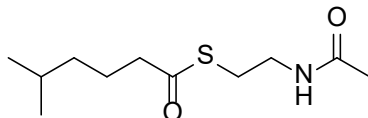


S-(2-acetamidoethyl) 4-methylhexanethioate (17)

Method C: 86% of a colorless oil.

^1H NMR (400 MHz, CDCl_3): δ 6.00 (br s, 1H), 3.41 (app q, $J = 6.0$ Hz, 2H), 3.00 (t, $J = 6.4$ Hz, 2H), 2.64 (q, $J = 6.8$ Hz, 1H), 2.00 (s, 3H), 1.67 (m, 1H), 1.42 (m, 1H), 1.25 (m, 3H), 1.11 (d, $J = 6.8$ Hz, 3H), 0.86 (t, $J = 6.8$ Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 200.5, 170.5, 42.0, 39.9, 33.9, 32.1, 29.1, 28.4, 23.1, 18.8, 11.2. HRFAB m/z calculated exact mass for $\text{C}_{11}\text{H}_{21}\text{NO}_2\text{S}$ 232.1371, found 232.1366 $[\text{M}+\text{H}]^+$. See

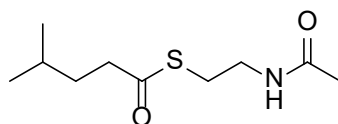
Figure B.38 for NMR spectra



S-(2-acetamidoethyl) 5-methylhexanethioate (18)

Method B: 100% of a colorless oil.

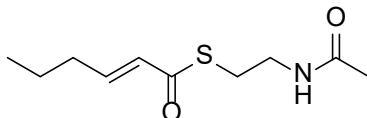
^1H NMR (400 MHz, CDCl_3): δ 5.94 (br s, 1H), 3.38 (app q, $J = 6.2$ Hz, 2H), 2.98 (t, $J = 6.6$ Hz, 2H), 2.53 (app.t, $J = 15.2, 7.5$ Hz, 2H), 1.93 (s, 3H), 1.62 (m, 2H), 1.51 (m, 1H), 1.17 (m, 2H), 0.84 (d, $J = 6.2$ Hz, 6H). ^{13}C NMR (101 MHz, CDCl_3): δ 200.2, 170.3, 44.3, 39.7, 38.1, 28.4, 23.5, 24.1, 23.1, 22.0. HRFAB m/z calculated exact mass for $\text{C}_{11}\text{H}_{22}\text{NO}_2\text{S}$ 232.1371, found 232.1377 $[\text{M}+\text{H}]^+$. See Figure B.39 for NMR spectra



S-(2-acetamidoethyl) 4-methylpentanethioate (19)

Method B: 100% of a colorless oil.

^1H NMR (400 MHz, CDCl_3): δ 5.94 (br s, 1H), 3.38 (app q, $J = 6.2$ Hz, 2H), 2.97 (t, $J = 6.6$ Hz, 2H), 2.53 (app t, $J = 7.4$ Hz, 2H), 1.91 (s, 3H), 1.51 (m, 3H), 0.85 (d, $J = 6.2$ Hz, 6H). ^{13}C NMR (101 MHz, CDCl_3): δ 200.4, 170.3, 42.3, 39.7, 34.4, 28.4, 27.6, 23.1, 22.2. HRFAB m/z calculated exact mass for $\text{C}_{10}\text{H}_{20}\text{NO}_2\text{S}$: 218.1215, found: 218.1215 $[\text{M}+\text{H}]^+$. See Figure B.40 for NMR spectra

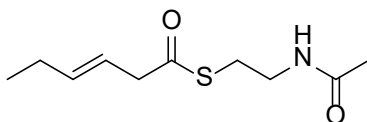


S-(2-acetamidoethyl) (E)-hex-2-enethioate (20)

Method A

^1H NMR (400 MHz, CDCl_3): δ 6.92 (dt, $J = 15.1, 7.3$ Hz, 1H), 6.12 (d, $J = 15.5$ Hz, 1H), 5.97 (br s, 1H), 3.44 (app q, $J = 6.4$ Hz, 2H), 3.08 (t, $J = 6.4$ Hz, 2H), 2.18 (app q, $J = 7.2$ Hz, 2H), 1.96 (s, 3H), 1.50 (pent, $J = 7.4$ Hz, 2H), 1.72 (t, $J = 7.4$ Hz, 3H).

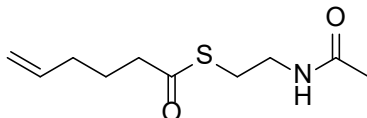
^{13}C NMR (101 MHz, CDCl_3): δ 190.6, 170.5, 146.7, 128.57, 47.7, 40.0, 34.4, 28.4, 23.4, 21.3. HR-ESI-MS: 238.0866 ($[M + \text{Na}]^+$, $\text{C}_{10}\text{H}_{17}\text{NO}_2\text{SNa}^+$; calc. 238.0872). Spectral data matched that reported by Tang *et. al.*⁹.



S-(2-acetamidoethyl) (E)-hex-3-enethioate (21)

Method C

^1H NMR (400 MHz, CDCl_3): δ 5.90 (br s, 1H), 5.65 (m, 1H), 5.48 (m, 1H), 3.43 (app q, $J = 6.3$ Hz, 2H), 3.24 (ddd, $J = 1.0, 4.4, 6.8$ Hz, 1H), 3.02 (t, $J = 6.4$ Hz, 2H), 2.07 (m, 2H), 1.96 (s, 3H), 1.00 (t, $J = 7.6$ Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 198.9, 170.4, 138.3, 119.8, 47.6, 39.7, 28.5, 25.6, 23.2, 13.3. HRFAB m/z calculated exact mass for $\text{C}_{10}\text{H}_{17}\text{NO}_2\text{S}$: 216.1058, found: 216.1062 $[\text{M}+\text{H}]^+$. Spectral data matched that reported by Gay *et. al.*¹⁰.

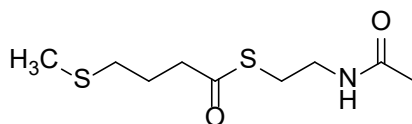


S-(2-acetamidoethyl) hex-5-enethioate (22)

Method A: 24% of a colorless oil

^1H NMR (300 MHz, CDCl_3): δ 6.13 (br s, 1H), 5.72 (ddt, J = 17.0, 10.2, 6.7 Hz, 1H), 5.02–4.95 (m, 2H), 3.38 (app q, J = 6.3 Hz, 2H), 2.99 (t, J = 6.3 Hz, 2H), 2.55 (t, J = 7.5 Hz, 2H), 2.05 (q, J = 6.9, CH_2), 1.93 (s, 3H), 1.72 (pent, J = 7.5 Hz, 2H). ^{13}C NMR (75 MHz, CDCl_3): δ 199.4, 170.5, 137.2, 115.5, 43.1, 39.4, 32.6, 28.3, 24.50, 23.0. HR-ESI-MS calculated exact mass for $\text{C}_{10}\text{H}_{18}\text{NO}_2\text{S}$: 216.1058, found: 216.1059 $[\text{M}+\text{H}]^+$.

See Figure B.41 for NMR spectra

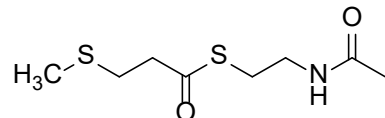


S-(2-acetamidoethyl) 4-(methylthio)butanethioate (23)

Method A: 30% of a colorless oil

^1H NMR (400 MHz, CDCl_3): δ 5.89 (br s, 1H), 3.42 (app q, J = 6.3 Hz, 2H), 3.04 (t, J = 6.3 Hz, 2H), 2.75 (t, J = 6.9 Hz, 2H), 2.44 (t, J = 6.9 Hz, 2H), 2.03 (s, 3H), 1.97 (m, 2H), 1.96 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 199.3, 170.4, 42.7, 39.6, 33.3, 28.6, 24.6, 23.4, 15.4. HR-ESI-MS: 258.0589 ($[\text{M} + \text{Na}]^+$, $\text{C}_9\text{H}_{17}\text{NO}_2\text{S}_2\text{Na}^+$; calc. 258.0593).

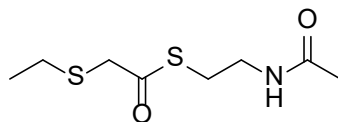
See Figure B.42 for NMR spectra



S-(2-acetamidoethyl) 3-(methylthio)propanethioate (24)

Method A: 27% of a colorless oil

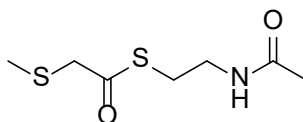
^1H NMR (400 MHz, CDCl_3): δ 6.29 (br s, 1H), 3.36 (app q, $J = 6.3$ Hz, 2H), 2.99 (t, $J = 6.5$ Hz, 2H), 2.81 (AA'BB', $J \approx 7.1, 1.9$ Hz, 2H), 2.73 (AA'BB', $J \approx 7.1, 1.9$ Hz, 2H), 2.06 (s, 3H), 1.91 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 197.9, 170.4, 43.4, 39.3, 29.3, 28.5, 23.1, 15.5. HR-ESI-MS: 244.0431 ($[\text{M} + \text{Na}]^+$, $\text{C}_8\text{H}_{15}\text{NO}_2\text{S}_2\text{Na}^+$; calc. 244.0436). See Figure B.43 for NMR spectra



S-(2-acetamidoethyl) 2-(ethylthio)ethanethioate (25)

Method A: colorless oil

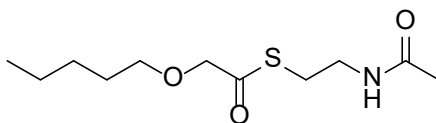
^1H NMR (400 MHz, CDCl_3): δ 5.89 (br s, 1H), 3.45 (app q, $J = 6.4$ Hz, 2H), 3.43 (s, 2H), 3.06 (t, $J = 6.4$ Hz, 2H), 2.64 (dq, $J = 0.8, 7.2$ Hz, 2H), 1.97 (s, 3H), 1.27 (dt, $J = 0.8, 7.2$ Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 197.4, 170.3, 41.9, 39.5, 29.1, 27.0, 23.2, 14.1. HR-ESI-MS calculated exact mass for $\text{C}_8\text{H}_{16}\text{NO}_2\text{S}_2$: 222.0623, found: 222.0622 $[\text{M}+\text{H}]^+$. See Figure B.44 for NMR spectra



S-(2-acetamidoethyl) 2-(methylthio)ethanethioate (26)

Method A: colorless oil

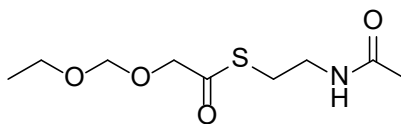
^1H NMR (400 MHz, CDCl_3): δ 5.98 (br s, 1H), 3.44 (app q, $J = 6.3$ Hz, 2H), 3.38 (s, 2H), 3.06 (t, $J = 6.3$ Hz, 2H), 2.19 (s, 3H), 1.96 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 197.0, 170.4, 44.2, 39.4, 29.1, 23.2, 16.5. HR-ESI-MS calculated exact mass for $\text{C}_7\text{H}_{14}\text{NO}_2\text{S}_2$: 208.0466, found: 208.0464 $[\text{M}+\text{H}]^+$. See Figure B.45 for NMR spectra



S-(2-acetamidoethyl) 2-(pentyloxy)ethanethioate (27)

Method A: 61% of a colorless oil

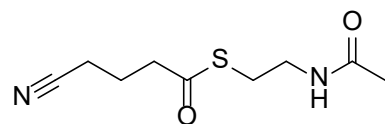
^1H NMR (300 MHz, CDCl_3): δ 6.04 (br s, 1H), 4.10 (s, 2H), 3.54 (t, $J = 6.9$ Hz, 2H), 3.41 (app q, $J = 6.4$ Hz, 2H), 3.03 (t, $J = 6.4$ Hz, 2H), 1.95 (s, 3H), 1.63 (p, $J = 6.9$ Hz, 2H), 1.35 (m, 4H), 0.90 (t, $J = 7.1$ Hz, 3H). ^{13}C NMR (75 MHz, CDCl_3): δ 201.0, 170.4, 75.8, 72.9, 39.5, 29.3, 28.1, 27.4, 23.2, 22.5, 14.0. HR-ESI-MS: 270.1132 ($[\text{M} + \text{Na}]^+$, $\text{C}_{11}\text{H}_{21}\text{NO}_3\text{SNa}^+$; calc. 270.1134). See Figure B.46 for NMR spectra



S-(2-acetamidoethyl) 2-(ethoxymethoxy)ethanethioate (28)

Method A colorless oil

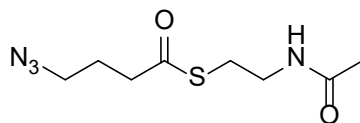
^1H NMR (400 MHz, CDCl_3): δ 6.15 (br s, 1H), 4.73 (s, 2H), 4.22 (s, 2H), 3.63 (q, J = 7.0 Hz, 2H), 3.40 (app q, J = 6.4 Hz, 2H), 3.03 (t, J = 6.4 Hz, 2H), 1.94 (s, 3H), 1.19 (t, J = 7.0 Hz, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 199.7, 170.5, 95.3, 71.9, 64.1, 39.4, 27.5, 23.1, 15.0. HR-ESI-MS calculated exact mass for $\text{C}_9\text{H}_{18}\text{NO}_4\text{S}$: 236.0957, found: 236.0966 $[\text{M}+\text{H}]^+$. See Figure B.47 for NMR spectra



S-(2-acetamidoethyl) 4-cyanobutanethioate (29)

Method A: 60% of a colorless oil

^1H NMR (300 MHz, CDCl_3): δ 5.95 (br s, 1H), 3.42 (app q, J = 6.0 Hz, 2H), 3.01 (t, J = 6.0 Hz, 2H), 2.71 (t, J = 7.2 Hz, 2H), 2.52 (t, J = 6.9 Hz, 2H), 1.95 (pent, J = 7.2 Hz, 2H), 1.95 (s, 3H). ^{13}C NMR (75 MHz, CDCl_3): δ 198.1, 170.4, 118.9, 42.0, 39.4, 28.9, 23.3, 21.2, 16.3. HR-ESI-MS calculated exact mass for $\text{C}_9\text{H}_{15}\text{N}_2\text{O}_2\text{S}$: 215.0854, found: 215.0856 $[\text{M}+\text{H}]^+$. See Figure B.48 for NMR spectra



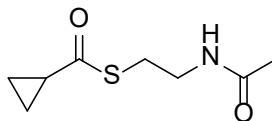
S-(2-acetamidoethyl) 4-azidobutanethioate (30)

Method A: 72% of a colorless oil

^1H NMR (300 MHz, CDCl_3): δ 6.17 (br s, 1H), 3.39 (app q, J = 6.3 Hz, 2H), 3.32 (t, J = 6.6 Hz, 2H), 3.00 (t, J = 6.3 Hz, 2H), 2.65 (t, J = 7.2 Hz, 2H), 1.93 (s, 3H), 1.88

(pent, $J = 6.6$ Hz, 3H). ^{13}C NMR (75 MHz, CDCl_3): δ 198.6, 170.5, 50.4, 40.8, 39.4, 28.6, 24.7, 23.1. HR-ESI-MS: 253.0735 ($[\text{M} + \text{Na}]^+$, $\text{C}_8\text{H}_{14}\text{N}_4\text{O}_2\text{SNa}^+$; calc. 253.0711).

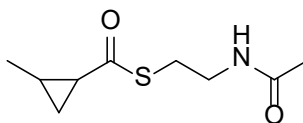
See Figure B.49 for NMR spectra



S-(2-acetamidoethyl) cyclopropanecarbothioate (31)

Method C

^1H NMR (400 MHz, CDCl_3): δ 6.08 (br s, 1H), 3.41 (app q, $J = 6.1$ Hz, 2H), 3.03 (t, $J = 6.4$ Hz, 2H), 2.02 (sym m, 1H), 1.95 (s, 3H), 1.15 (sym m, 2H), 1.35 (m, 2H), 0.97 (sym m, 2H). ^{13}C -NMR (101 MHz, CDCl_3): δ 199.9, 170.3, 39.8, 31.1, 28.5, 22.8, 23.2, 22.7, 11.1. HR-ESI-MS calculated exact mass for $\text{C}_8\text{H}_{14}\text{NO}_2\text{S}$: 188.0745, found: 188.0751 $[\text{M}+\text{H}]^+$. Spectral data matched that reported by Lee *et. al.*¹¹.



S-(2-acetamidoethyl) 2-methylcyclopropane-1-carbothioate (32)

Method C: 72% of a colorless oil

As a 3:1 mixture of diastereomers. Major: ^1H NMR (400 MHz, CDCl_3): δ 5.99 (br s, 1H), 3.42 (app q, $J = 6.0$ Hz, 2H), 3.02 (t, $J = 6.8$ Hz, 2H), 1.95 (s, 3H), 1.76 (sym

m, 1H), 1.55 (b m, 1H), 1.35 (m, 1H), 1.13 (d, $J = 6.0$ Hz, 3H), 0.82 (sym m, 1H).

Minor: ^1H NMR (400 MHz, CDCl_3): δ 5.99 (br s, 1H), 3.42 (app q, $J = 6.0$ Hz, 2H),

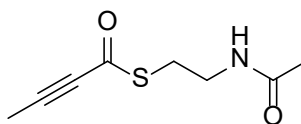
3.06 (m, 2H), 2.10 (b m, 1H) 1.95 (s, 3H), 1.48 (b m, 1H), 1.14 (d, $J = 6.0$ Hz, 3H),

1.11 (m, 1H), 1.10 (m, 1H). Major: ^{13}C -NMR (101 MHz, CDCl_3): δ 199.2, 170.3,

39.9, 31.4, 28.5, 23.2, 20.4, 19.6, 17.9. Minor: ^{13}C -NMR (101 MHz, CDCl_3): δ 197.7,

170.3, 40.1, 28.6, 28.3, 23.2, 19.6, 16.3, 11.9. HR-ESI-MS calculated exact mass for

$\text{C}_9\text{H}_{16}\text{NO}_2\text{S}$: 202.0902, found: 202.0910 $[\text{M}+\text{H}]^+$. See Figure B.50 for NMR spectra



S-(2-acetamidoethyl) but-2-ynethioate (33)

Method A: colorless prisms

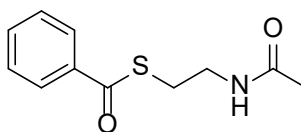
^1H NMR (400 MHz, CDCl_3): δ 6.01 (br s, 1H), 3.45 (app q, $J = 6.3$ Hz, 2H), 3.10 (t, J

= 6.3 Hz, 2H), 2.05 (s, 3H), 1.97 (s, 3H). ^{13}C NMR (101 MHz, CDCl_3): δ 176.6, 170.4,

92.5, 77.9, 39.3, 29.3, 23.2, 4.3. HR-ESI-MS calculated exact mass for $\text{C}_8\text{H}_{12}\text{NO}_2\text{S}$:

186.0589, found: 186.0585 $[\text{M}+\text{H}]^+$. Melting point: 59 °C. See Figure B.51 for NMR

spectra

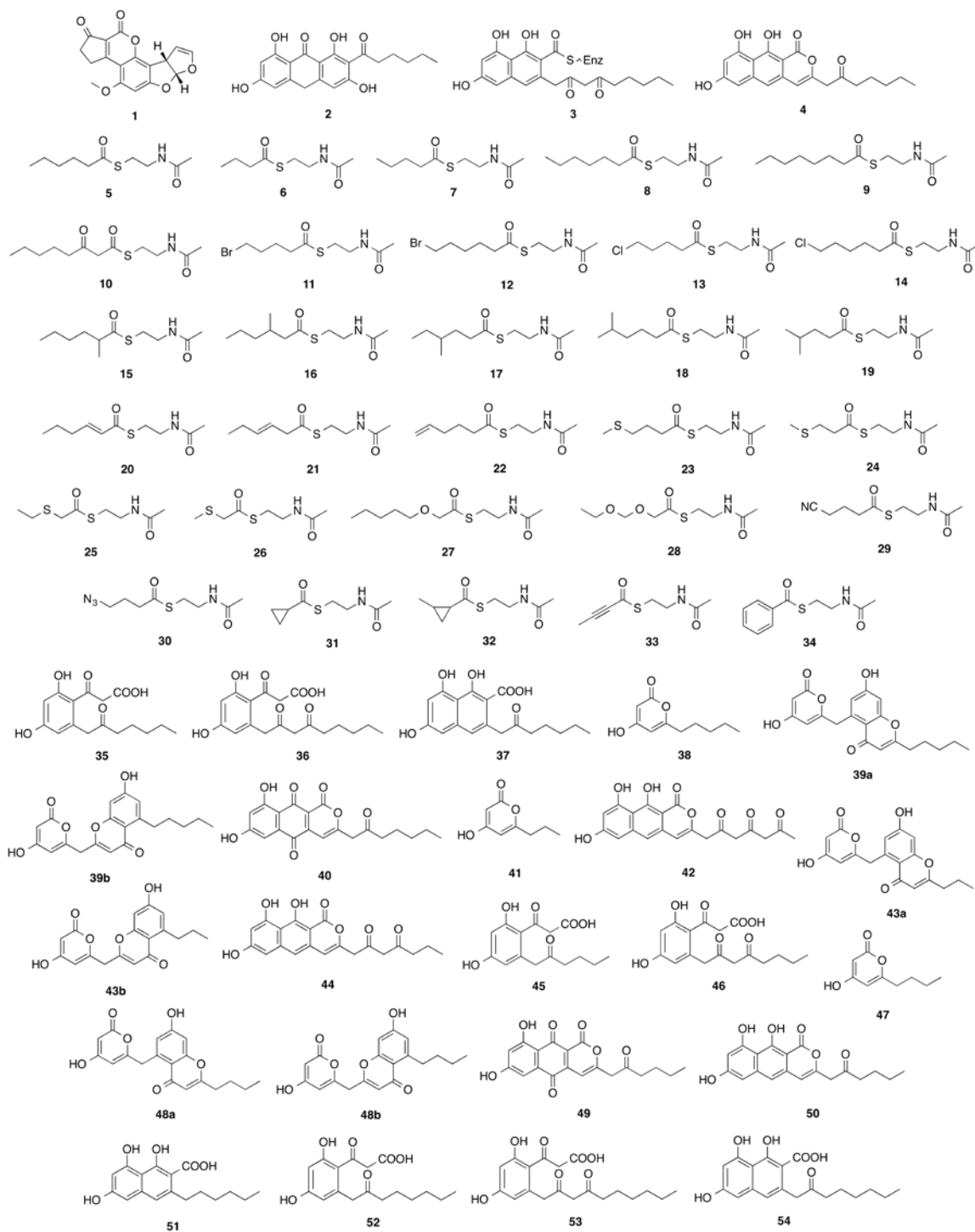


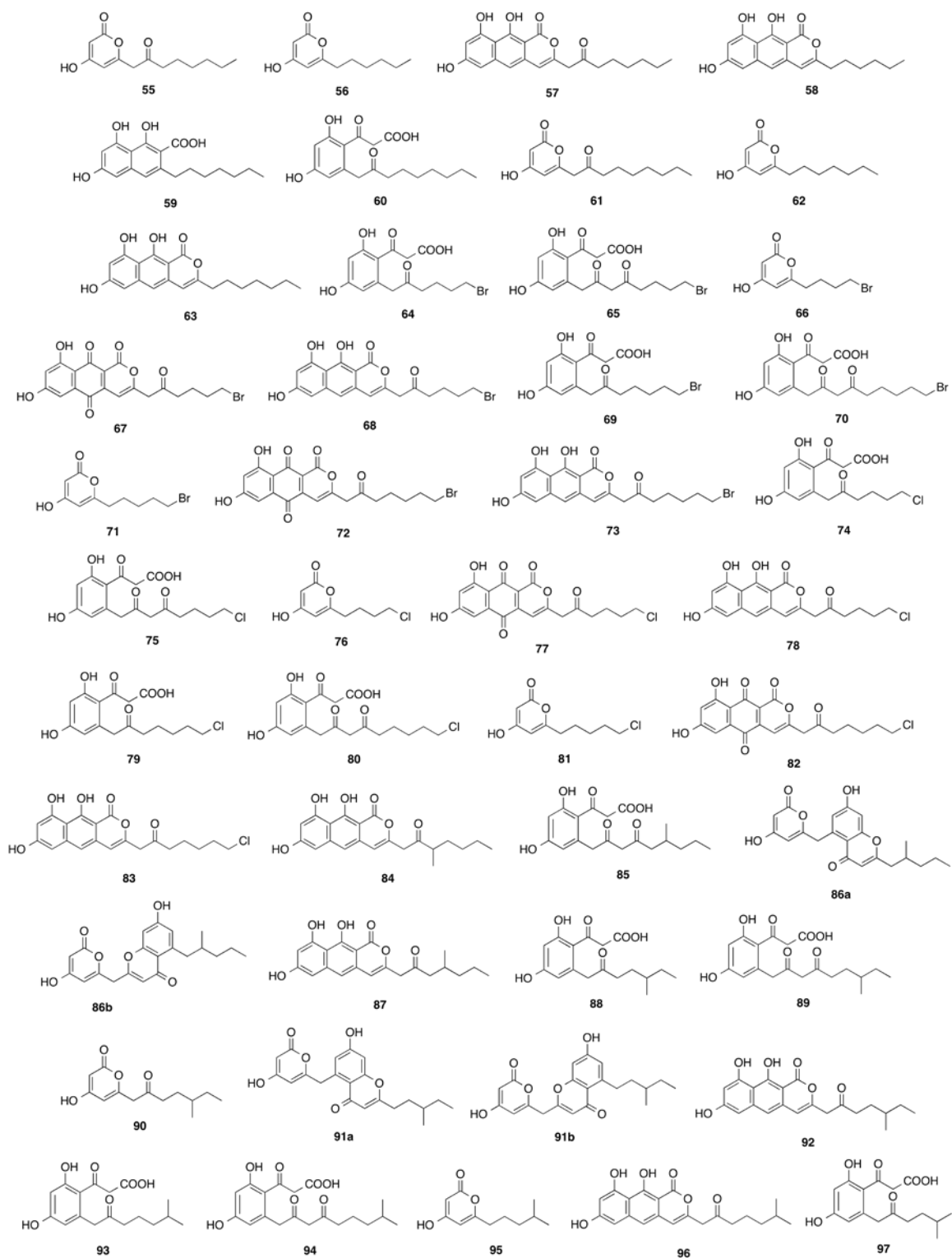
S-(2-acetamidoethyl) benzothioate (34)

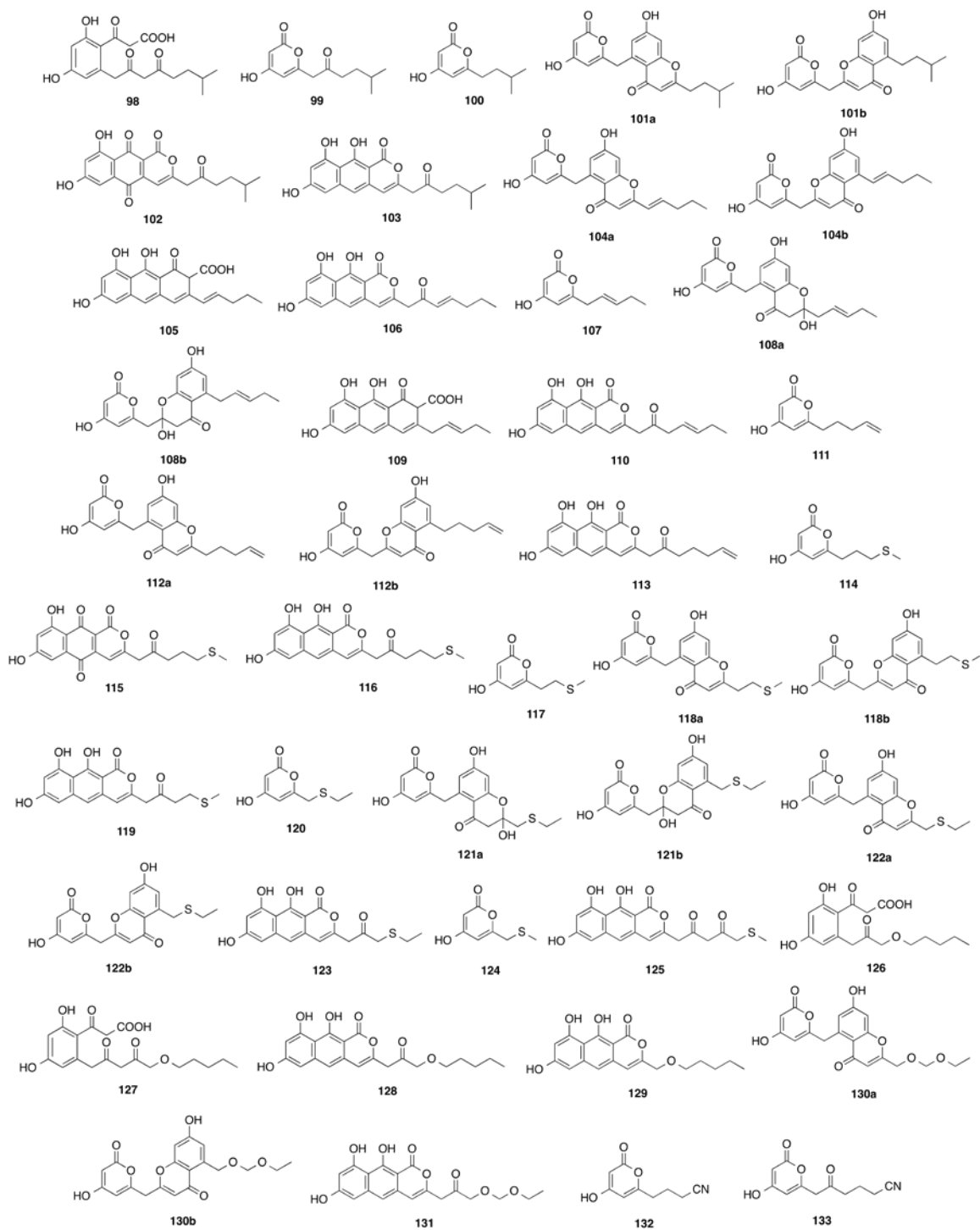
Method C

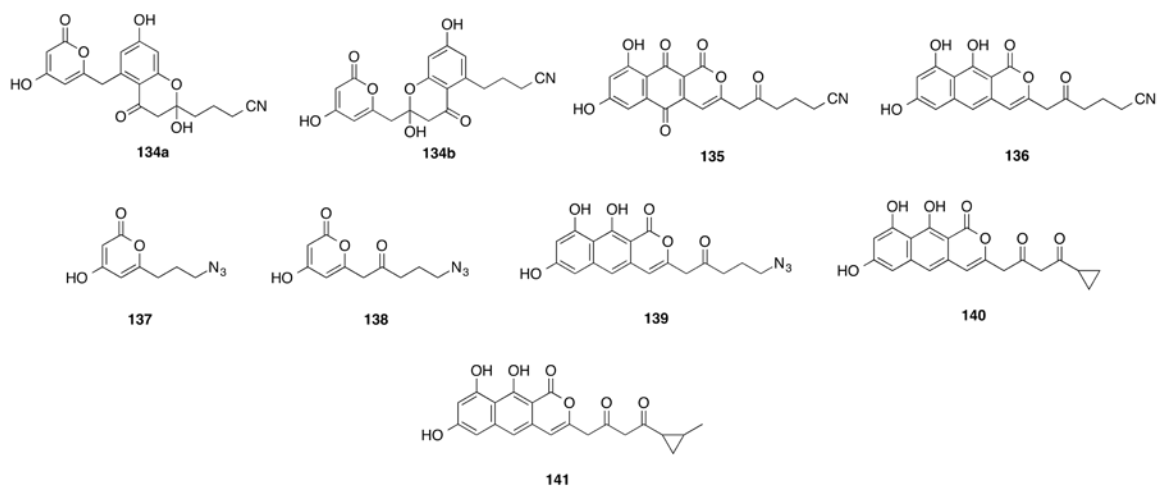
^1H NMR (300 MHz, CDCl_3): δ 7.96 (d, $J = 4.9$ Hz, 2H), 7.58 (d, $J = 3.8$ Hz, 1H), 7.46 (dd, $J = 4.7, 3.8$ Hz, 2H), 6.01 (br s, 1H), 3.51 (app q, $J \approx 6.0$ Hz, 2H), 3.23 (t, $J = 6.0$ Hz, 2H), 1.97 (s, 3H). ^{13}C NMR (75 MHz, CDCl_3): δ 192.4, 170.5, 136.8, 133.8, 128.8, 127.4, 39.8, 28.7, 23.4. HR-ESI-MS: 246.0554 ($[M + \text{Na}]^+$, $\text{C}_{11}\text{H}_{13}\text{NO}_2\text{SNa}^+$; calc. 246.0559). Spectral data matched that reported by Prasad *et. al.*⁶.

B.3. Supplementary figures

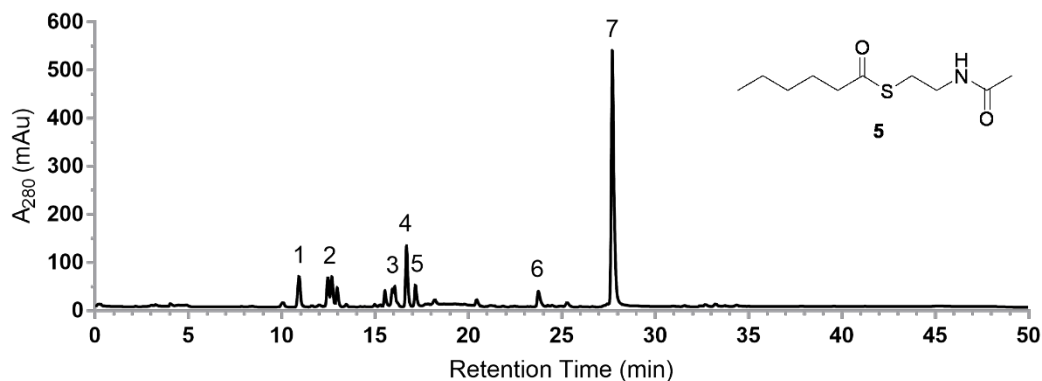






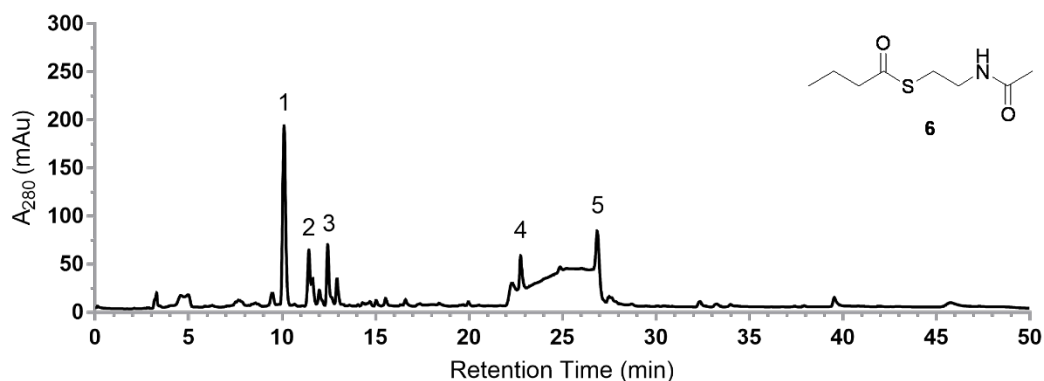


Appendix Figure B.1. Full structures of all starter units assayed in this paper and the resulting enzymatic products. See following figures (Appendix Figure B.2-B.32) for HPLC and HR-ESI-MS characterization of enzymatic products **35-141**.



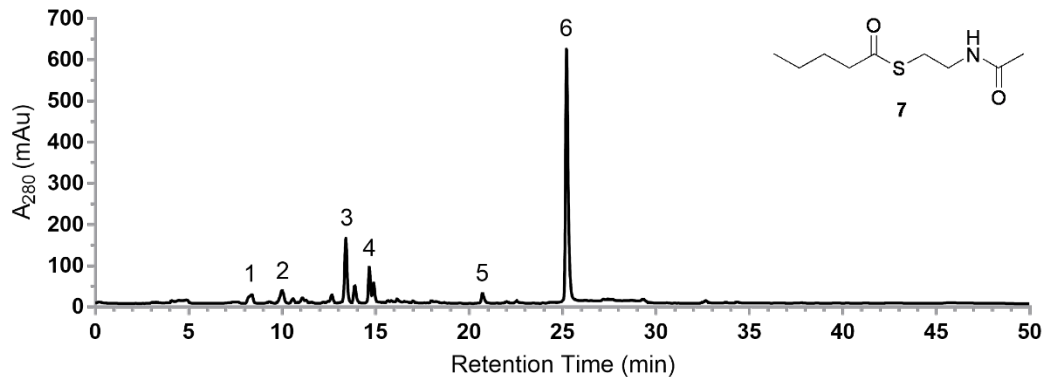
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	35	B	C ₅ H ₁₁	260, 288	309.1342	Hex+5Mal	-H ₂ O	+H ₂ O	C ₁₆ H ₂₀ O ₆
2	36	B	CH ₂ COC ₅ H ₁₁	262, 290	351.1434	Hex+6Mal	-H ₂ O	+H ₂ O	C ₁₈ H ₂₂ O ₇
3	37	A	CH ₂ COC ₅ H ₁₁	220, 288	333.1337	Hex+6Mal	-2H ₂ O	+H ₂ O	C ₁₈ H ₂₀ O ₆
4	38	C	C ₅ H ₁₁	220, 286	183.1027	Hex+2Mal	-	O-C	C ₁₀ H ₁₄ O ₃
5	39a,b	D/E	C ₅ H ₁₁	224, 278, 306	357.1342	Hex+7Mal	-H ₂ O	O-C	C ₂₀ H ₂₀ O ₆
6	40	F	CH ₂ COC ₅ H ₁₁	248, 270, 340, 468	371.1124	Hex+7Mal	-2H ₂ O	O-C, [O]	C ₂₀ H ₁₈ O ₇
7	4	G	CH ₂ COC ₅ H ₁₁	270, 280, 390	357.1332	Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₂₀ O ₆

Appendix Figure B.2. Product profile for starter unit **5** reaction (*S*-(2-acetamidoethyl) hexanethioate)



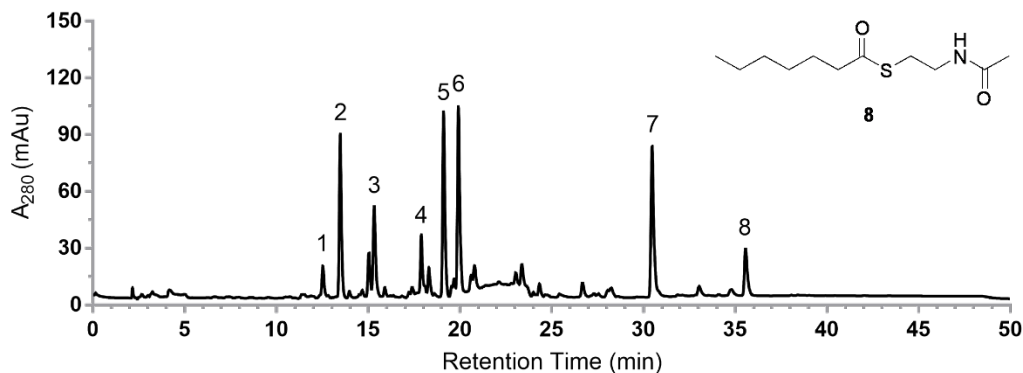
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release	Mol. Formula
1	41	C	C ₃ H ₇	232, 284	155.0712	But+2Mal	-	O-C	C ₈ H ₁₀ O ₅
2	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	270, 280, 390	385.0917	10Mal	-2H ₂ O	O-C	C ₂₀ H ₁₆ O ₈
3	43a,b	D/E	C ₃ H ₇	224, 278, 306	329.1027	But+7Mal	-2H ₂ O	O-C	C ₁₈ H ₁₆ O ₆
4	44	G	CH ₂ COCH ₂ COC ₃ H ₇	270, 280, 390	371.1130	But+8Mal	-2H ₂ O	O-C	C ₂₀ H ₁₈ O ₇
5	44	G	CH ₂ COCH ₂ COC ₃ H ₇	270, 280, 390	371.1128	But+8Mal	-2H ₂ O	O-C	C ₂₀ H ₁₈ O ₇

Appendix Figure B.3. Product profile for starter unit **6** reaction (*S*-(2-acetamidoethyl) butanethioate)



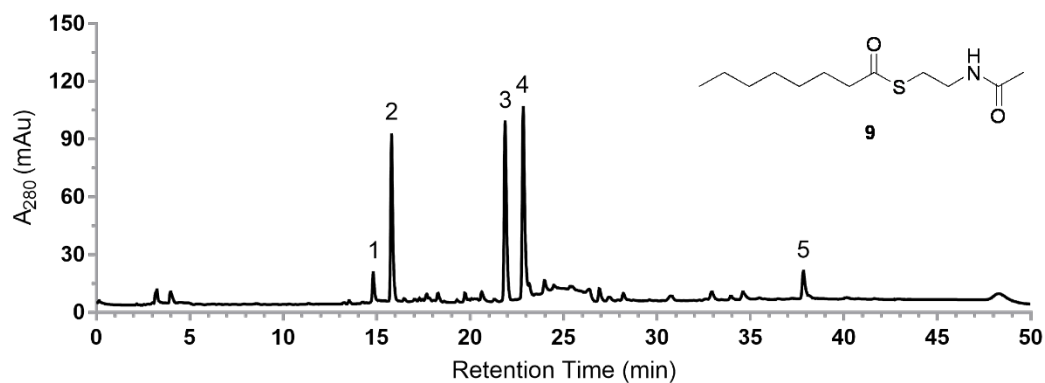
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	45	B	C ₄ H ₉	260, 288	295.1184	Pent+5Mal	-H ₂ O	+ H ₂ O	C ₁₅ H ₁₈ O ₆
2	46	B	CH ₂ COC ₄ H ₉	262, 288	337.1285	Pent+6Mal	-H ₂ O	+ H ₂ O	C ₁₇ H ₂₀ O ₇
3	47	C	C ₄ H ₉	220, 286	169.0872	Pent+2Mal	-	O-C	C ₉ H ₁₂ O ₃
4	48a,b	D/E	C ₄ H ₉	224, 278, 306	343.1178	Pent+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₈ O ₆
5	49	F	CH ₂ COC ₄ H ₉	246, 270, 336, 468	357.0974	Pent+7Mal	-2H ₂ O	O-C, [O]	C ₁₉ H ₁₆ O ₇
6	50	G	CH ₂ COC ₄ H ₉	272, 280, 390	343.1174	Pent+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₈ O ₆

Appendix Figure B.4. Product profile for starter unit 7 reaction (*S*-(2-acetamidoethyl) pentanethioate)



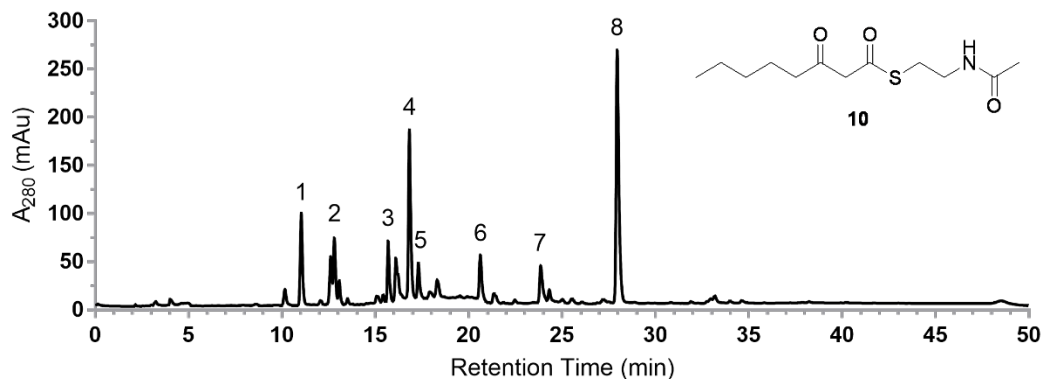
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	51	A	C ₆ H ₁₃	262, 288	305.1391	Hept+5Mal	-2H ₂ O	+H ₂ O	C ₁₇ H ₂₀ O ₅
2	52	B	C ₆ H ₁₃	260, 288	323.1494	Hept+5Mal	-H ₂ O	+H ₂ O	C ₁₇ H ₂₂ O ₆
3	53	B	CH ₂ COC ₆ H ₁₃	262, 288	365.1602	Hept+6Mal	-H ₂ O	+H ₂ O	C ₁₉ H ₂₄ O ₇
4	54	A	CH ₂ COC ₆ H ₁₃	222, 288	347.1497	Hept+6Mal	-2H ₂ O	+H ₂ O	C ₁₉ H ₂₂ O ₆
5	55	C	CH ₂ COC ₆ H ₁₃	224, 286	239.1288	Hept+3Mal	-	O-C	C ₁₃ H ₁₈ O ₄
6	56	C	C ₆ H ₁₃	220, 286	197.1183	Hept+2Mal	-	O-C	C ₁₁ H ₁₆ O ₃
7	57	G	CH ₂ COC ₆ H ₁₃	270, 280, 390	371.1493	Hept+7Mal	-2H ₂ O	O-C	C ₂₁ H ₂₂ O ₆
8	58	G	C ₆ H ₁₃	270, 280, 390	329.1403	Hept+6Mal	-2H ₂ O	O-C	C ₁₉ H ₂₀ O ₅

Appendix Figure B.5. Product profile for starter unit 8 reaction (*S*-(2-acetamidoethyl) heptanethioate)



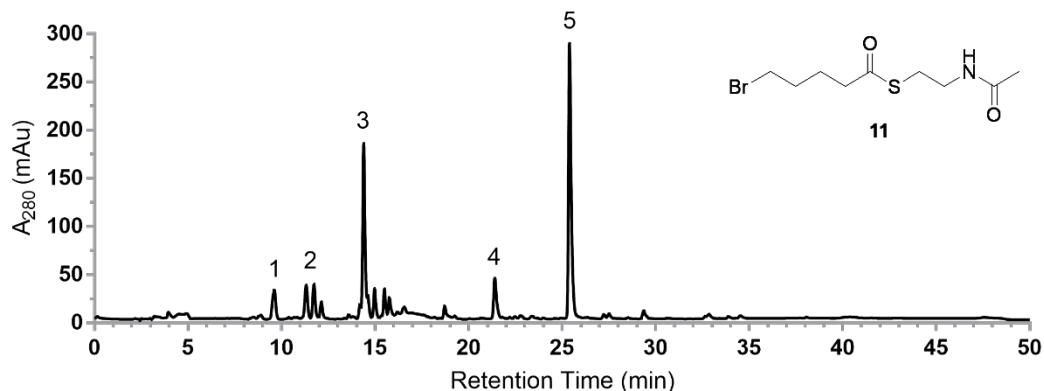
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	59	A	C ₇ H ₁₅	262, 288	319.1548	Oct+5Mal	-2H ₂ O	+H ₂ O	C ₁₈ H ₂₂ O ₅
2	60	B	C ₇ H ₁₅	260, 288	337.1651	Oct+5Mal	-H ₂ O	+H ₂ O	C ₁₈ H ₂₄ O ₆
3	61	C	CH ₂ COC ₇ H ₁₅	220, 286	253.1443	Oct+3Mal	-	O-C	C ₁₄ H ₂₀ O ₄
4	62	C	C ₇ H ₁₅	220, 286	211.1339	Oct+2Mal	-	O-C	C ₁₂ H ₁₈ O ₃
5	63	G	C ₇ H ₁₅	270, 280, 390	343.1538	Oct+6Mal	-H ₂ O	O-C	C ₂₀ H ₂₂ O ₅

Appendix Figure B.6. Product profile for starter unit **9** reaction (*S*-(2-acetamidoethyl) octanethioate)



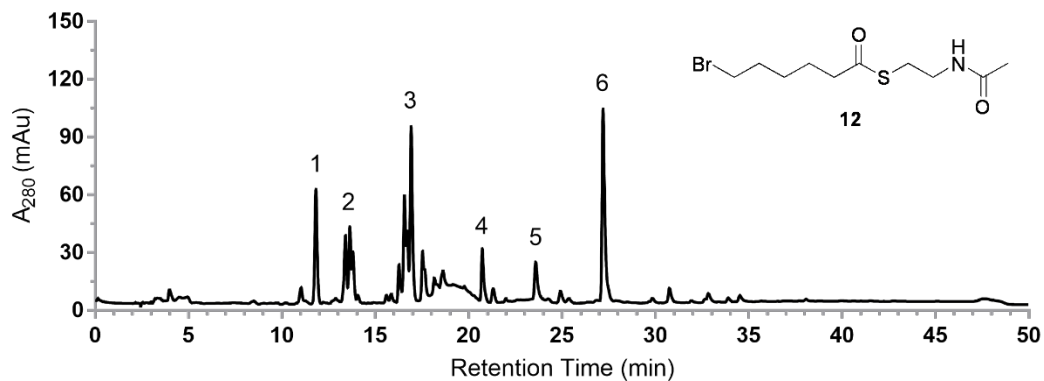
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	35	B	C ₅ H ₁₁	260, 288	309.1342	Oct+4Mal	-H ₂ O	+H ₂ O	C ₁₆ H ₂₀ O ₆
2	36	B	CH ₂ COC ₅ H ₁₁	262, 288	351.1435	Oct+5Mal	-H ₂ O	+H ₂ O	C ₁₈ H ₂₂ O ₇
3	37	A	CH ₂ COC ₅ H ₁₁	220, 288	333.1336	Oct+5Mal	-2H ₂ O	+H ₂ O	C ₁₈ H ₂₀ O ₆
4	38	C	C ₅ H ₁₁	218, 286	183.1025	Oct+Mal	-	O-C	C ₁₀ H ₁₄ O ₃
5	39a,b	D/E	C ₅ H ₁₁	224, 278, 310	357.1337	Oct+6Mal	-2H ₂ O	O-C	C ₂₀ H ₂₀ O ₆
6				234, 248, 280, 304	333.1341	Oct+5Mal			C ₁₈ H ₂₀ O ₆
7	40	F	CH ₂ COC ₅ H ₁₁	246, 264, 340, 468	371.1127	Oct+6Mal	-2H ₂ O	O-C, [O]	C ₂₀ H ₁₈ O ₇
8	4	G	CH ₂ COC ₅ H ₁₁	270, 280, 390	357.1332	Oct+6Mal	-2H ₂ O	O-C	C ₂₀ H ₂₀ O ₆

Appendix Figure B.7. Product profile for starter unit **10** reaction (*S*-(2-acetamidoethyl) 3-oxooctanethioate)



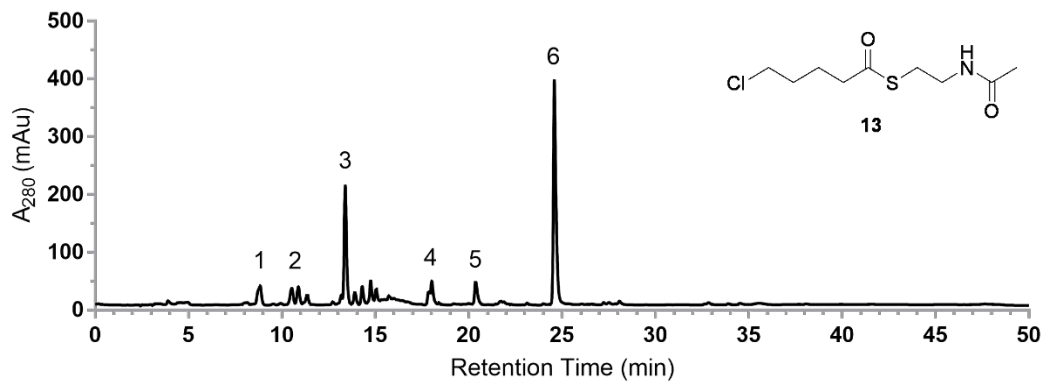
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	64	B	C ₄ H ₈ Br	262, 290	373.0296	Br-Pent+5Mal	-H ₂ O	+H ₂ O	C ₁₅ H ₁₇ BrO ₆
2	65	B	CH ₂ COC ₄ H ₈ Br	262, 290	415.0380	Br-Pent+6Mal	-H ₂ O	+H ₂ O	C ₁₇ H ₁₉ BrO ₇
3	66	C	C ₄ H ₈ Br	220, 286	246.9974	Br-Pent+2Mal	-	O-C	C ₉ H ₁₁ BrO ₃
4	67	F	CH ₂ COC ₄ H ₈ Br	246, 268, 338, 468	435.0079	Br-Pent+7Mal	-2H ₂ O	O-C, [O]	C ₁₉ H ₁₅ BrO ₇
5	68	G	CH ₂ COC ₄ H ₈ Br	272, 280, 388	421.0273	Br-Pent+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₇ BrO ₆

Appendix Figure B.8. Product profile for starter unit **11** reaction (*S*-(2-acetamidoethyl) 5-bromopentanethioate)



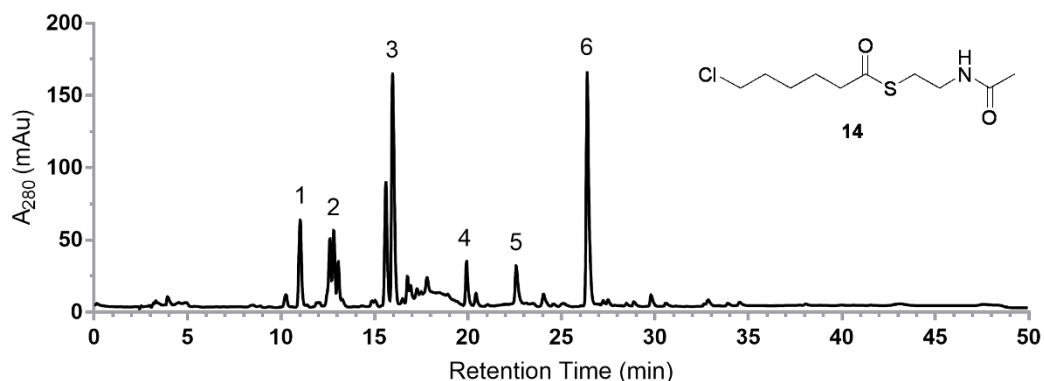
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	69	B	C ₅ H ₁₀ Br	260, 288	387.0440	Br-Hex+5Mal	-H ₂ O	+H ₂ O	C ₁₆ H ₁₉ BrO ₆
2	70	B	CH ₂ COC ₅ H ₁₀ Br	262, 288	429.0569	Br-Hex+6Mal	-H ₂ O	+H ₂ O	C ₁₈ H ₂₁ BrO ₇
3	71	C	C ₅ H ₁₀ Br	220, 286	261.0121	Br-Hex+2Mal	-	O-C	C ₁₀ H ₁₃ BrO ₃
4				212, 234, 278, 302	411.0431	Br-Hex+6Mal			C ₁₈ H ₁₉ BrO ₆
5	72	F	CH ₂ COC ₅ H ₁₀ Br	246, 264, 336, 468	449.0241	Br-Hex+7Mal	-2H ₂ O	O-C, [O]	C ₂₀ H ₁₇ BrO ₇
6	73	G	CH ₂ COC ₅ H ₁₀ Br	272, 280, 390	435.0427	Br-Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₁₉ BrO ₆

Appendix Figure B.9. Product profile for starter unit **12** reaction (*S*-(2-acetamidoethyl) 6-bromohexanethioate)



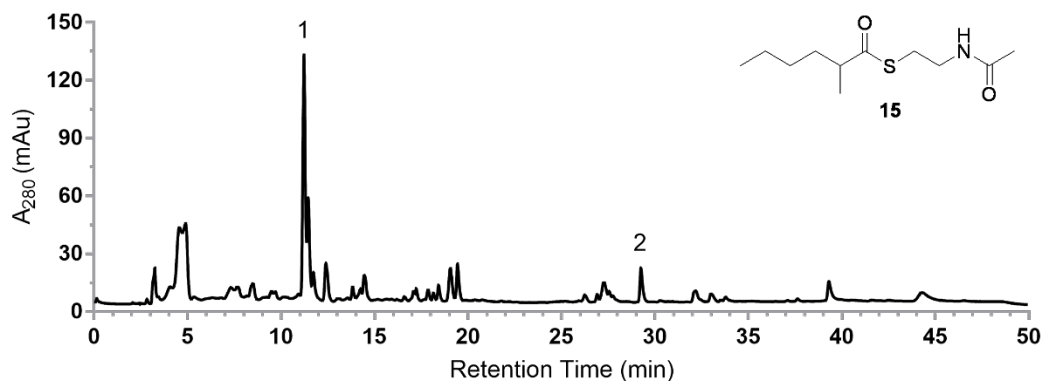
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	74	B	C ₄ H ₈ Cl	262, 288	329.0812	Cl-Pent+5Mal	-H ₂ O	+H ₂ O	C ₁₅ H ₁₇ ClO ₆
2	75	B	CH ₂ COC ₄ H ₈ Cl	262, 288	371.0872	Cl-Pent+6Mal	-H ₂ O	+H ₂ O	C ₁₇ H ₁₉ ClO ₇
3	76	C	C ₄ H ₈ Cl	220, 286	203.0479	Cl-Pent+2Mal	-	O-C	C ₉ H ₁₁ ClO ₃
4				218, 286	245.0481				C ₁₁ H ₁₃ ClO ₄
5	77	F	CH ₂ COC ₄ H ₈ Cl	246, 264, 338, 468	391.0569	Cl-Pent+7Mal	-2H ₂ O	O-C, [O]	C ₁₉ H ₁₅ ClO ₇
6	78	G	CH ₂ COC ₄ H ₈ Cl	272, 280, 390	377.0784	Cl-Pent+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₇ ClO ₆

Appendix Figure B.10. Product profile for starter unit **13** reaction (*S*-(2-acetamidoethyl) 5-chloropentanethioate)



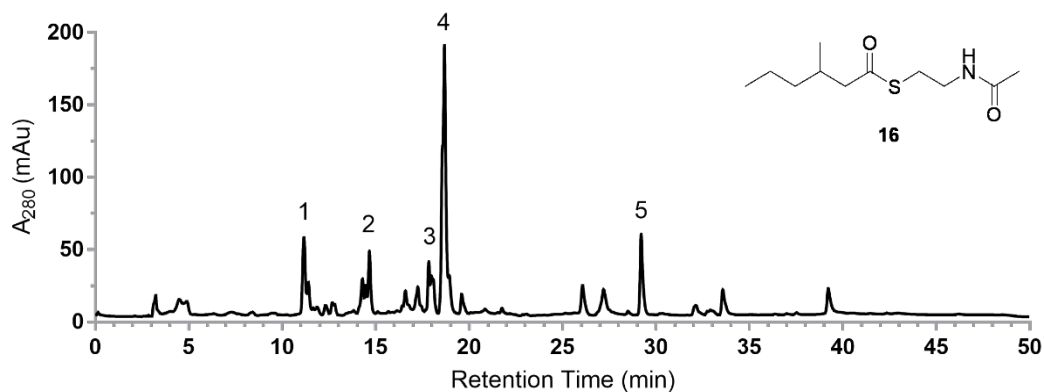
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	79	B	C ₅ H ₁₀ Cl	260, 288	343.0923	Cl-Hex+5Mal	-H ₂ O	+H ₂ O	C ₁₆ H ₁₉ ClO ₆
2	80	B	CH ₂ COC ₅ H ₁₀ Cl	262, 288	385.1041	Cl-Hex+6Mal	-H ₂ O	+H ₂ O	C ₁₈ H ₂₁ ClO ₇
3	81	C	C ₅ H ₁₀ Cl	220, 286	217.0627	Cl-Hex+2Mal	-	O-C	C ₁₀ H ₁₃ ClO ₃
4				218, 234, 278, 300	367.0952	Cl-Hex+6Mal			C ₁₈ H ₁₉ ClO ₆
5	82	F	CH ₂ COC ₅ H ₁₀ Cl	246, 270, 330, 468	405.0741	Cl-Hex+7Mal	-2H ₂ O	O-C, [O]	C ₂₀ H ₁₇ ClO ₇
6	83	G	CH ₂ COC ₅ H ₁₀ Cl	270, 280, 390	391.0914	Cl-Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₁₉ ClO ₆

Appendix Figure B.11. Product profile for starter unit **14** reaction (*S*-(2-acetamidoethyl) 6-chlorohexanethioate)



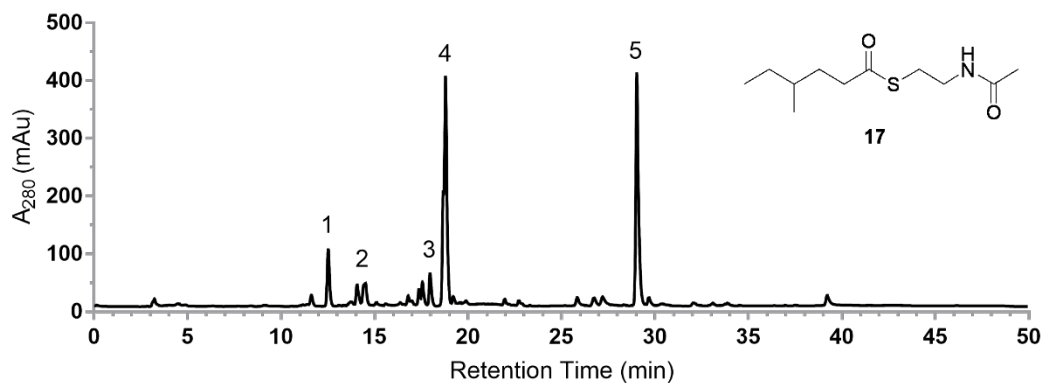
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration Release, Other	Mol. Formula
1	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	272, 280, 390	385.0911	10Mal	-2H ₂ O	O-C C ₂₀ H ₁₆ O ₈
2	84	G	CH ₂ COCH(CH ₃)C ₄ H ₉	270, 280, 390	371.1487	Me-Hex+7Mal	-2H ₂ O	O-C C ₂₁ H ₂₂ O ₆

Appendix Figure B.12. Product profile for starter unit **15** reaction (*S*-(2-acetamidoethyl) 2-methylhexanethioate)



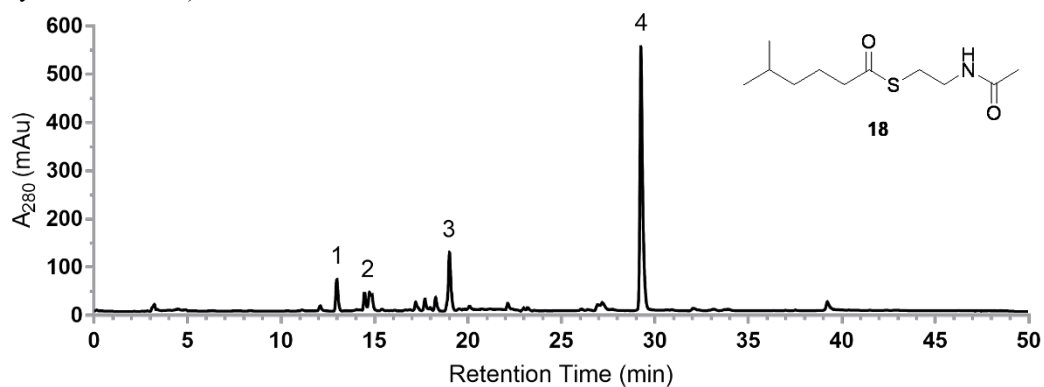
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration Release, Other	Mol. Formula
1	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	270, 280, 390	385.0905	10Mal	-2H ₂ O	O-C C ₂₀ H ₁₆ O ₈
2	85	B	CH ₂ COCH ₂ CH(CH ₃)C ₃ H ₇	266, 288	365.1596	Me-Hex+6Mal	-H ₂ O	+H ₂ O C ₁₉ H ₂₄ O ₇
3				240, 288, 336	197.1178			C ₁₁ H ₁₆ O ₃
4	86a,b	D/E	CH ₂ CH(CH ₃)C ₃ H ₇	224, 278, 310	371.1493	Me-Hex+7Mal	-2H ₂ O	O-C C ₂₁ H ₂₂ O ₆
5	87	G	CH ₂ CO CH ₂ CH(CH ₃)C ₃ H ₇	270, 280, 390	371.1478	Me-Hex+7Mal	-2H ₂ O	O-C C ₂₁ H ₂₂ O ₆

Appendix Figure B.13. Product profile for starter unit **16** reaction (*S*-(2-acetamidoethyl) 3-methylhexanethioate)



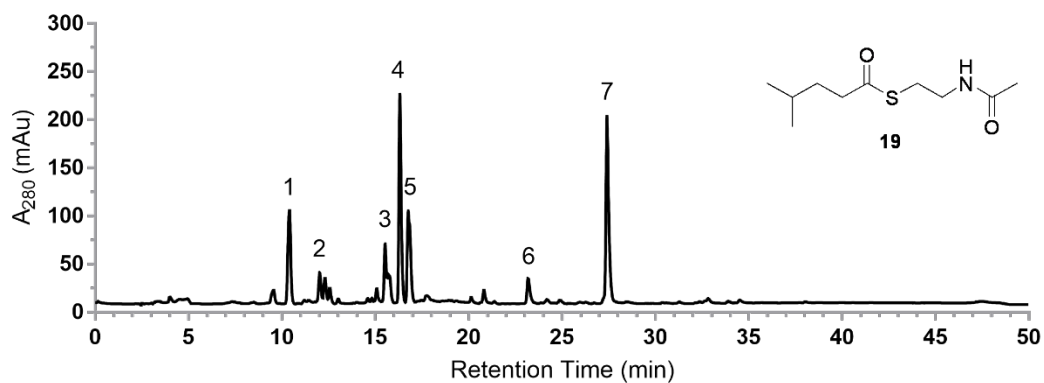
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	88	B	C ₂ H ₄ CH(CH ₃) C ₂ H ₅	260, 288	323.1493	Me-Hex +5Mal	-H ₂ O	+H ₂ O	C ₁₇ H ₂₂ O ₆
2	89	B	CH ₂ CO C ₂ H ₄ CH(CH ₃)C ₂ H ₅	262, 288	365.1594	Me-Hex +6Mal	-H ₂ O	+H ₂ O	C ₁₉ H ₂₄ O ₇
3	90	C	CH ₂ COC ₂ H ₄ CH(CH ₃) C ₂ H ₅	220, 284	239.1281	Me-Hex +3Mal	-	O-C	C ₁₃ H ₁₈ O ₄
4	91a,b	D/E	C ₂ H ₄ CH(CH ₃) C ₂ H ₅	224, 278, 310	371.1492	Me-Hex +7Mal	-2H ₂ O	O-C	C ₂₁ H ₂₂ O ₆
5	92	G	CH ₂ COC ₂ H ₄ CH(CH ₃) C ₂ H ₅	270, 280, 390	371.1486	Me-Hex +7Mal	-2H ₂ O	O-C	C ₂₁ H ₂₂ O ₆

Appendix Figure B.14. Product profile for starter unit **17** reaction (*S*-(2-acetamidoethyl) 4-methylhexanethioate)



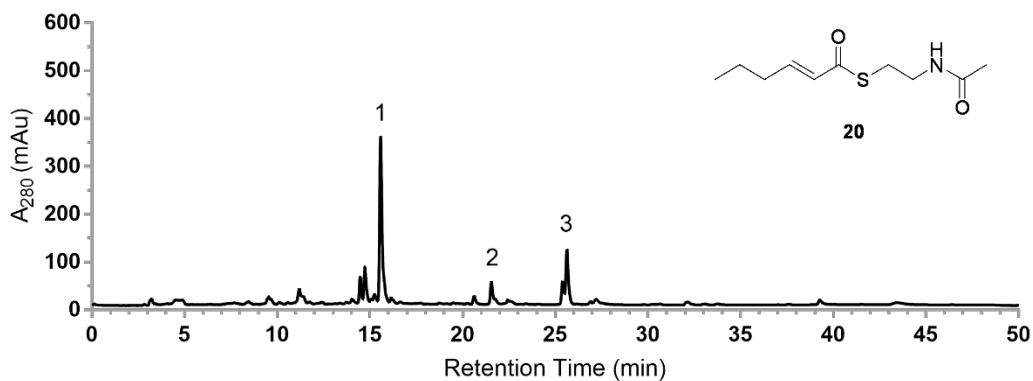
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	93	B	C ₃ H ₆ C(CH ₃) ₂	260, 288	323.1488	Me-Hex +5Mal	-H ₂ O	+H ₂ O	C ₁₇ H ₂₂ O ₆
2	94	B	CH ₂ COC ₃ H ₆ C(CH ₃) ₂	266, 288	365.1590	Me-Hex +6Mal	-H ₂ O	+H ₂ O	C ₁₉ H ₂₄ O ₇
4	95	C	C ₃ H ₆ C(CH ₃) ₂	224, 284	197.1179	Me-Hex +2Mal	-	O-C	C ₁₁ H ₁₆ O ₃
5	96	G	CH ₂ COC ₃ H ₆ C(CH ₃) ₂	270, 280, 390	371.1487	Me-Hex +7Mal	-2H ₂ O	O-C	C ₂₁ H ₂₂ O ₆

Appendix Figure B.15. Product profile for starter unit **18** reaction (*S*-(2-acetamidoethyl) 5-methylhexanethioate)



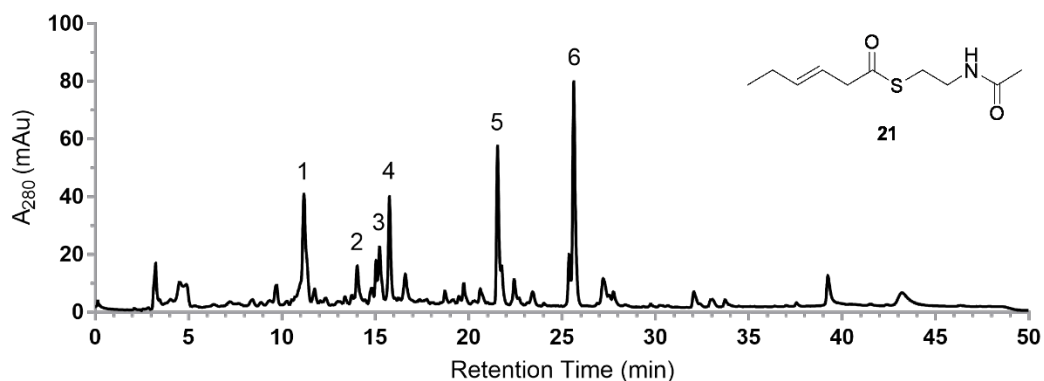
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	97	B	C ₂ H ₄ C(CH ₃) ₂	260, 288	309.1338	Me-Pent+5Mal	-H ₂ O	+H ₂ O	C ₁₆ H ₂₀ O ₆
2	98	B	CH ₂ COC ₂ H ₄ C(CH ₃) ₂	262, 288	351.1435	Me-Pent+6Mal	-H ₂ O	+H ₂ O	C ₁₈ H ₂₂ O ₇
3	99	C	CH ₂ COC ₂ H ₄ C(CH ₃) ₂	220, 284	225.1127	Me-Pent+3Mal	-	O-C	C ₁₂ H ₁₆ O ₄
4	100	C	C ₂ H ₄ C(CH ₃) ₂	220, 286	183.1024	Me-Pent+2Mal	-	O-C	C ₁₀ H ₁₄ O ₃
5	101a,b	D/E	C ₂ H ₄ C(CH ₃) ₂	232, 278, 304	357.1339	Me-Pent+7Mal	-2H ₂ O	O-C	C ₂₀ H ₂₀ O ₆
6	102	F	CH ₂ COC ₂ H ₄ C(CH ₃) ₂	246, 270, 330, 468	371.1121	Me-Pent+7Mal	-2H ₂ O	O-C, [O]	C ₂₀ H ₁₈ O ₇
7	103	G	CH ₂ COC ₂ H ₄ C(CH ₃) ₂	270, 280, 390	357.1349	Me-Pent+7Mal	-2H ₂ O	O-C	C ₂₀ H ₂₀ O ₆

Appendix Figure B.16. Product profile for starter unit **19** reaction (*S*-(2-acetamidoethyl) 4-methylpentanethioate)



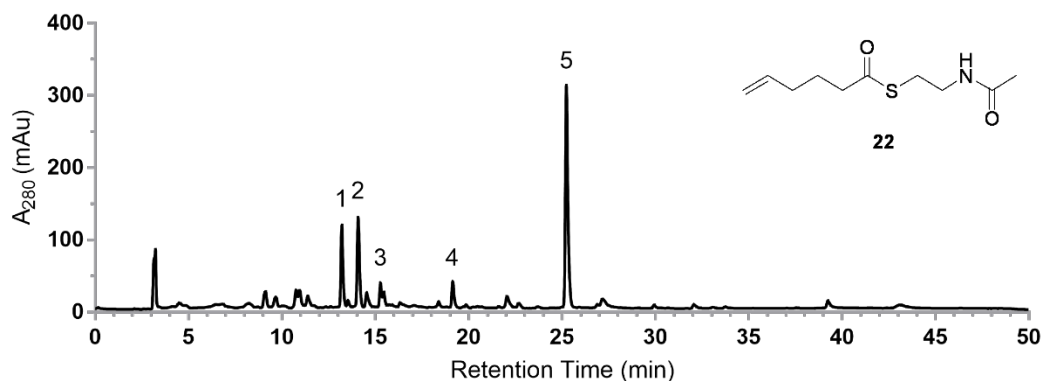
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	104a,b	D/E	CHCHC ₃ H ₇	226, 262	355.1176	Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₁₈ O ₆
2	105	H	CHCHC ₃ H ₇	222, 264, 304, 374	355.1179	Hex+7Mal	-2H ₂ O	+H ₂ O, C-C	C ₂₀ H ₁₈ O ₆
3	106	G	CH ₂ COCHCHC ₃ H ₇	270, 280, 390	355.1184	Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₁₈ O ₆

Appendix Figure B.17. Product profile for starter unit **20** reaction (*S*-(2-acetamidoethyl) (*E*)-hex-2-enethioate)



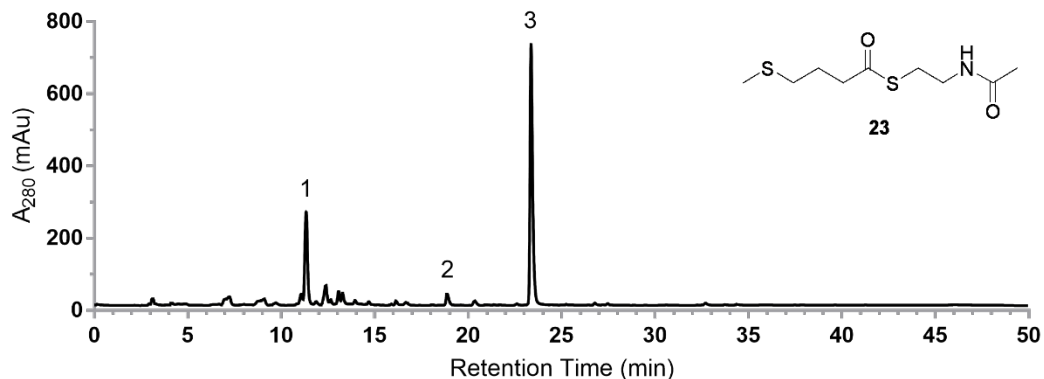
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1				218, 268, 382	349.1281	Hex+6Mal			C ₁₈ H ₂₀ O ₇
2				228, 284	331.1187	Hex+6Mal			C ₁₈ H ₁₈ O ₆
3	107	C	CH ₂ CHCHC ₂ H ₅	220, 288	181.0874	Hex+2Mal	-	O-C	C ₁₀ H ₁₂ O ₃
4	108a,b	D/E*	CH ₂ CHCHC ₂ H ₅	224, 280	373.1280	Hex+7Mal	-H ₂ O	O-C	C ₂₀ H ₂₀ O ₇
5	109	H	CH ₂ CHCHC ₂ H ₅	222, 264, 304, 374	355.1182	Hex+7Mal	-2H ₂ O	C-C	C ₂₀ H ₁₈ O ₆
6	110	G	CH ₂ COCH ₂ CHCHC ₂ H ₅	270, 280, 390	355.1183	Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₁₈ O ₆

Appendix Figure B.18. Product profile for starter unit **21** reaction (*S*-(2-acetamidoethyl) (*E*)-hex-3-enethioate). *108a,b consist of hydrated versions of cores D, E.



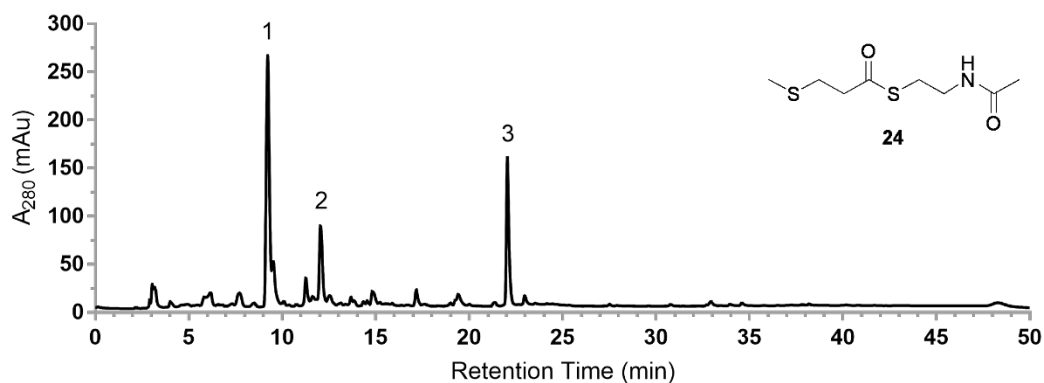
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1				272, 280, 390	490.1537				?
2	111	C	C ₃ H ₆ CHCH ₂	220, 286	181.0865	Hex+2Mal	-	O-C	C ₁₀ H ₁₂ O ₃
3	112a,b	D/E	C ₃ H ₆ CHCH ₂	224, 278, 310	355.1177	Hex+7Mal	-2H ₂ O	O-C	C ₂₀ H ₁₈ O ₆
4				270, 280, 390	331.1183	Hex+6Mal			C ₁₈ H ₁₈ O ₆
5	113	G	CH ₂ COC ₃ H ₆ CHCH ₂	270, 280, 390	355.1182	Hex+7Mal	-2H ₂ O	C-O	C ₂₀ H ₁₈ O ₆

Appendix Figure B.19. Product profile for starter unit **22** reaction (*S*-(2-acetamidoethyl) (*E*)-hex-5-enethioate)



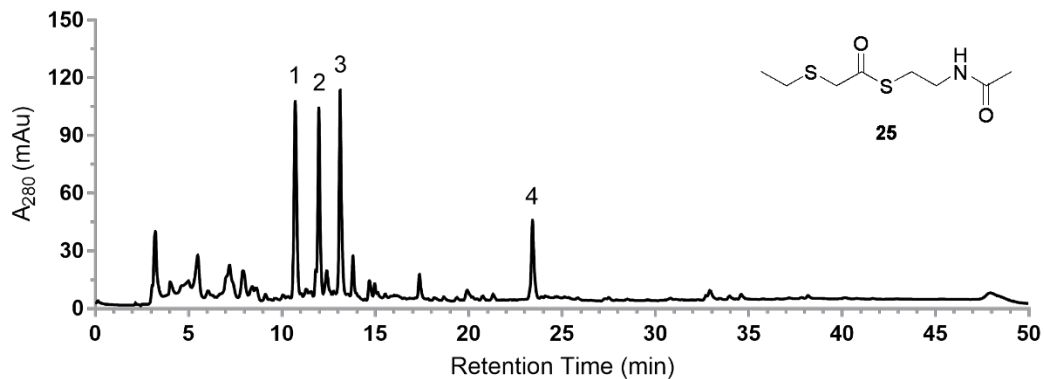
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	114	C	C ₃ H ₆ SCH ₃	220, 286	201.0591	'Hex'+2Mal	-	O-C	C ₉ H ₁₂ O ₃ S
2	115	F	CH ₂ COC ₃ H ₆ SCH ₃	246, 270, 330, 468	389.0688	'Hex'+7Mal	-2H ₂ O	O-C, [O]	C ₁₉ H ₁₆ O ₇ S
3	116	G	CH ₂ COC ₃ H ₆ SCH ₃	270, 280, 390	375.0899	'Hex'+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₈ O ₆ S

Appendix Figure B.20. Product profile for starter unit **23** reaction (*S*-(2-acetamidoethyl) 4-(methylthio)butanethioate)



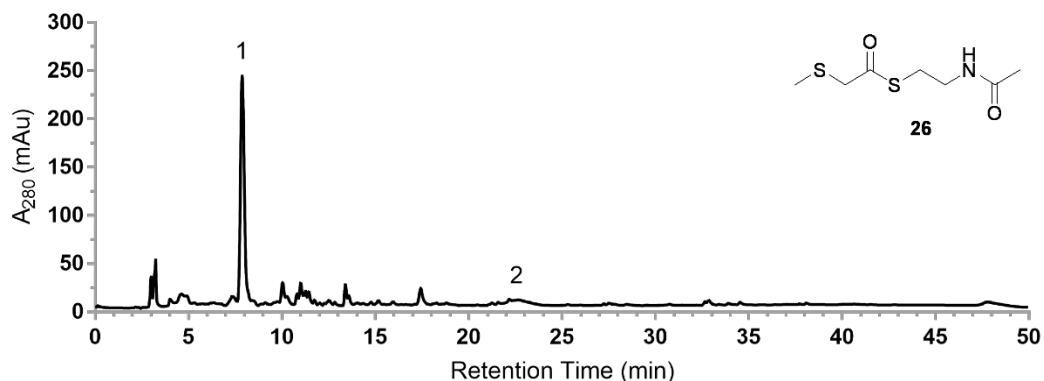
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	117	C	C ₂ H ₄ SCH ₃	220, 286	187.0428	'Pent'+2Mal	-	O-C	C ₈ H ₁₀ O ₃ S
2	118a,b	D/E	C ₂ H ₄ SCH ₃	224, 280, 388	361.0740	'Pent'+7Mal	-2H ₂ O	O-C	C ₁₈ H ₁₆ O ₆ S
3	119	G	CH ₂ COC ₂ H ₄ SCH ₃	270, 280, 390	361.0744	'Pent'+7Mal	-2H ₂ O	O-C	C ₁₈ H ₁₆ O ₆ S

Appendix Figure B.21. Product profile for starter unit **24** reaction (*S*-(2-acetamidoethyl) 3-(methylthio)propanethioate)



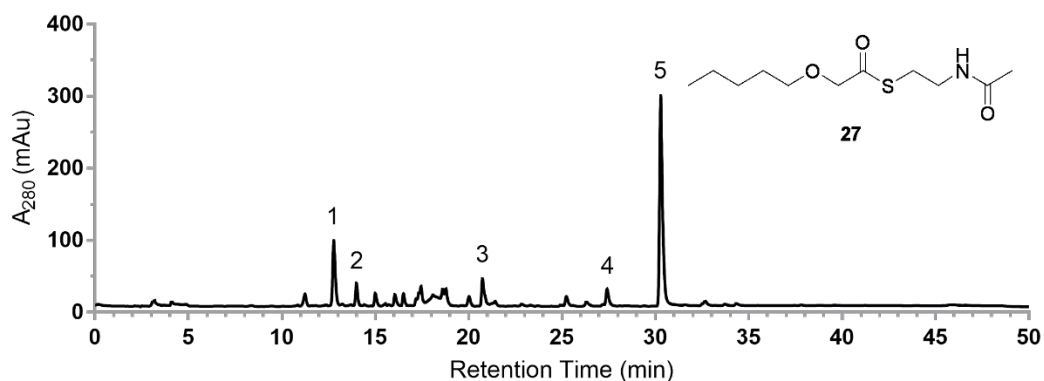
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration Release, Other	Mol. Formula
1	120	C	CH ₂ SC ₂ H ₅	222, 288	187.0433	'Pent'+2Mal	- O-C	C ₈ H ₁₀ O ₃ S
2	121a,b	D/E*	CH ₂ SC ₂ H ₅	224, 282	379.0844	'Pent'+7Mal	-H ₂ O O-C	C ₁₈ H ₁₈ O ₇ S
3	122a,b	D/E	CH ₂ SC ₂ H ₅	224, 278	361.0744	'Pent'+7Mal	-2H ₂ O O-C	C ₁₈ H ₁₆ O ₆ S
4	123	G	CH ₂ COCH ₂ SC ₂ H ₅	270, 280, 390	361.0744	'Pent'+7Mal	-2H ₂ O O-C	C ₁₈ H ₁₆ O ₆ S

Appendix Figure B.22. Product profile for starter unit **25** reaction (*S*-(2-acetamidoethyl) 2-(ethylthio)propanethioate). *129a,b are hydrated versions of 130 a,b



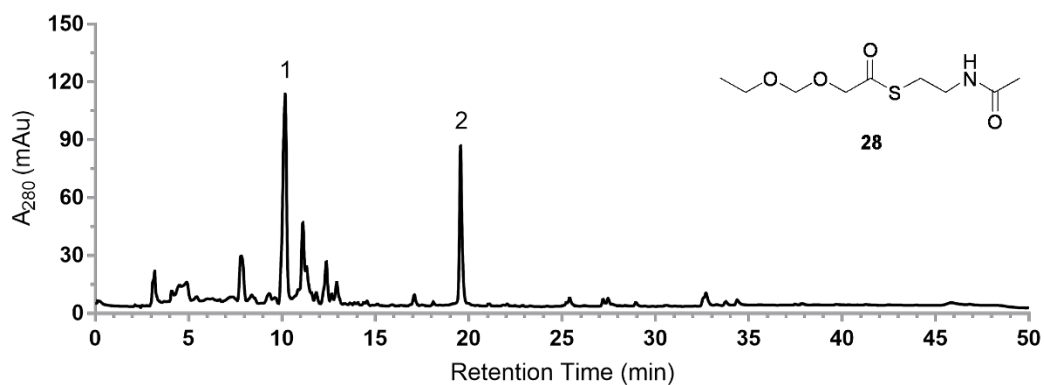
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration Release, Other	Mol. Formula
1	124	C	CH ₂ SCH ₃	236, 288	173.0273	'But'+2Mal	- O-C	C ₇ H ₈ O ₃ S
2	125	G	CH ₂ COCH ₂ COCH ₂ SCH ₃	272, 280, 390	389.0678	'But'+8Mal	-2H ₂ O O-C	C ₁₉ H ₁₆ O ₇ S

Appendix Figure B.23. Product profile for starter unit **26** reaction (*S*-(2-acetamidoethyl) 2-(methylthio)ethanethioate)



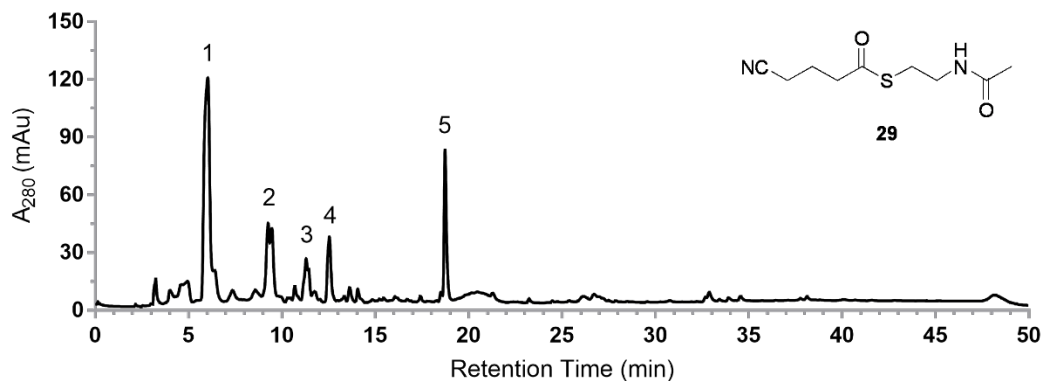
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	126	B	CH ₂ OC ₅ H ₁₁	262, 288	339.1443	'Oct'+5Mal	-H ₂ O	+H ₂ O	C ₁₇ H ₂₂ O ₇
2	127	B	CH ₂ COCH ₂ OC ₅ H ₁₁	262, 288	381.1544	'Oct'+6Mal	-H ₂ O	+H ₂ O	C ₁₉ H ₂₄ O ₈
3				234, 278, 300	363.1441	'Oct'+6Mal			C ₁₉ H ₂₂ O ₇
4	128	G	CH ₂ COCH ₂ OC ₅ H ₁₁	270, 280, 390	387.1476	'Oct'+7Mal	-2H ₂ O	O-C	C ₂₁ H ₂₂ O ₇
5	129	G	CH ₂ OC ₅ H ₁₁	270, 280, 390	345.1339	'Oct'+6Mal	-2H ₂ O	O-C	C ₁₉ H ₂₀ O ₆

Appendix Figure B.24. Product profile for starter unit **27** reaction (*S*-(2-acetamidoethyl) 2-(pentyloxy)ethanethioate)



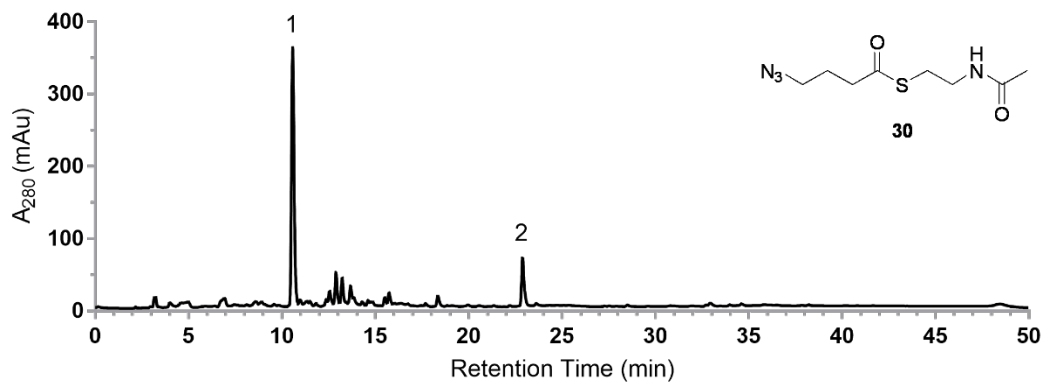
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	130a,b	D/E	CH ₂ OCH ₂ OC ₂ H ₅	232, 280	375.1074	'Hept'+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₈ O ₈
2	131	G	CH ₂ CO CH ₂ OCH ₂ OC ₂ H ₅	270, 280, 390	375.1076	'Hept'+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₈ O ₈

Appendix Figure B.25. Product profile for starter unit **28** reaction (*S*-(2-acetamidoethyl) 2-(ethoxymethoxy)ethanethioate)



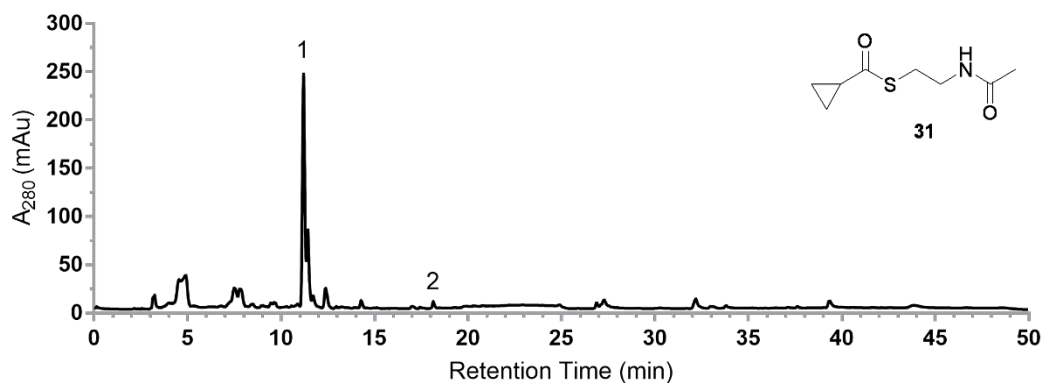
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	132	C	C ₃ H ₆ CN	220, 284	180.0661	Cyanobut+2Mal	-	O-C	C ₉ H ₉ NO ₃
	133	C	CH ₂ COC ₃ H ₆ CN	220, 284	222.0766	Cyanobut+3Mal	-	O-C	C ₁₁ H ₁₁ NO ₄
2	134a,b	D/E*	C ₃ H ₆ CN	224, 280	372.1071	Cyanobut+7Mal	-H ₂ O	O-C	C ₁₉ H ₁₇ NO ₇
3	135	F	CH ₂ COC ₃ H ₆ CN	226, 270, 390	368.0763	Cyanobut+7Mal	-2H ₂ O	O-C, [O]	C ₁₉ H ₁₃ NO ₇
4				234, 266, 350					
5	136	G	CH ₂ COC ₃ H ₆ CN	270, 280, 390	354.0964	Cyanobut+7Mal	-2H ₂ O	O-C	C ₁₉ H ₁₅ NO ₆

Appendix Figure B.26. Product profile for starter unit **29** reaction (*S*-(2-acetamidoethyl) 4-cyanobutanethioate). *142a,b consist of hydrated versions of cores D, E



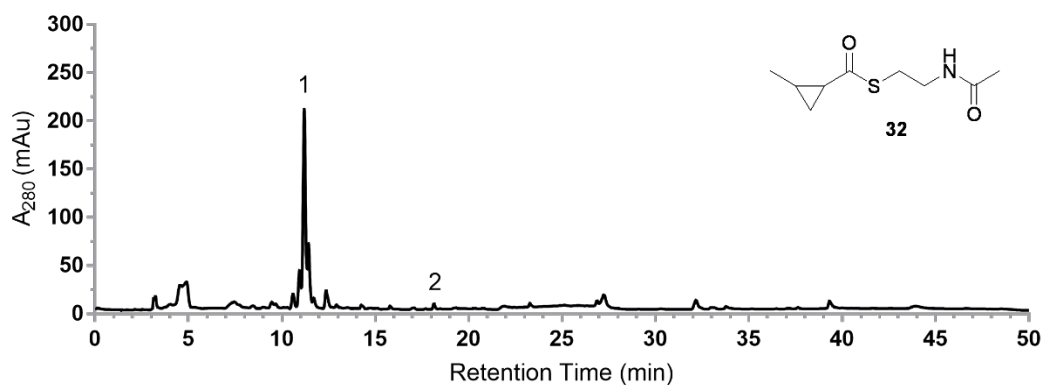
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	137	C	C ₃ H ₆ N ₃	220, 284	196.0718	Azidobut+2Mal	-	O-C	C ₈ H ₉ N ₃ O ₃
	138	C	CH ₂ COC ₃ H ₆ N ₃	220, 284	238.0417	Azidobut+3Mal	-	O-C	C ₁₀ H ₁₁ N ₃ O ₄
2	139	G	CH ₂ COC ₃ H ₆ N ₃	270, 280, 390	370.1022	Azidobut+7Mal	-2H ₂ O	O-C	C ₁₈ H ₁₅ N ₃ O ₆

Appendix Figure B.27. Product profile for starter unit **30** reaction (*S*-(2-acetamidoethyl) 4-azidobutanethioate)



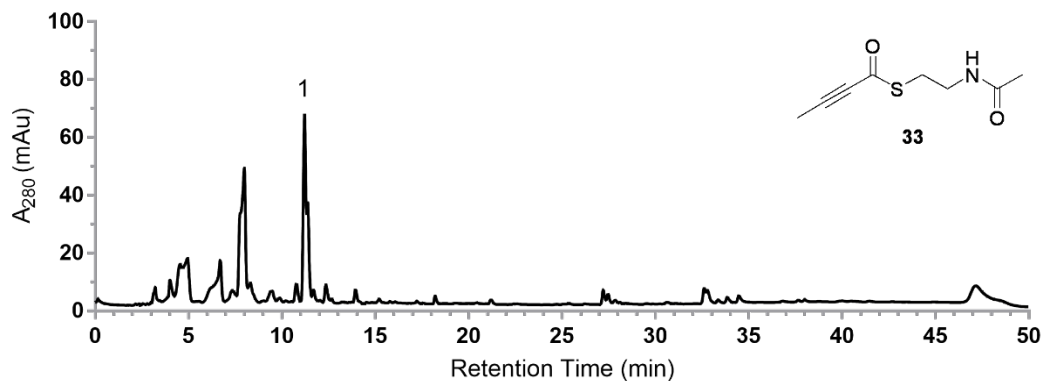
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	270, 280, 390	385.0923	10Mal	-2H ₂ O	O-C	C ₂₀ H ₁₆ O ₈
2	140	G	CH ₂ COCH ₂ COC ₃ H ₅	280, 390	369.0964	+8Mal	-2H ₂ O	O-C	C ₂₀ H ₁₆ O ₇

Appendix Figure B.28. Product profile for starter unit **31** reaction (*S*-(2-acetamidoethyl) cyclopropanecarbothioate)



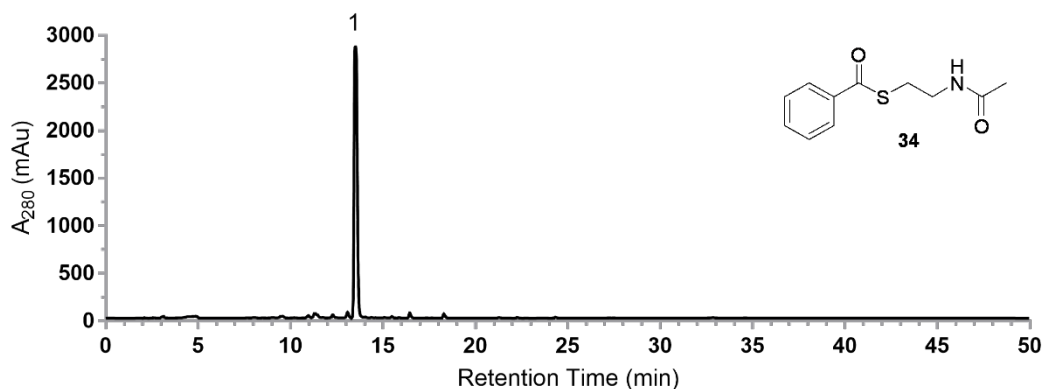
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	270, 280, 390	385.0921	10Mal	-2H ₂ O	O-C	C ₂₀ H ₁₆ O ₈
2	141	G	CH ₂ COCH ₂ COC ₃ H ₄ CH ₃	280, 390	383.1123	+8Mal	-2H ₂ O	O-C	C ₂₁ H ₁₈ O ₇

Appendix Figure B.29. Product profile for starter unit **32** reaction (*S*-(2-acetamidoethyl) 2-methylcyclopropane-1-carbothioate)



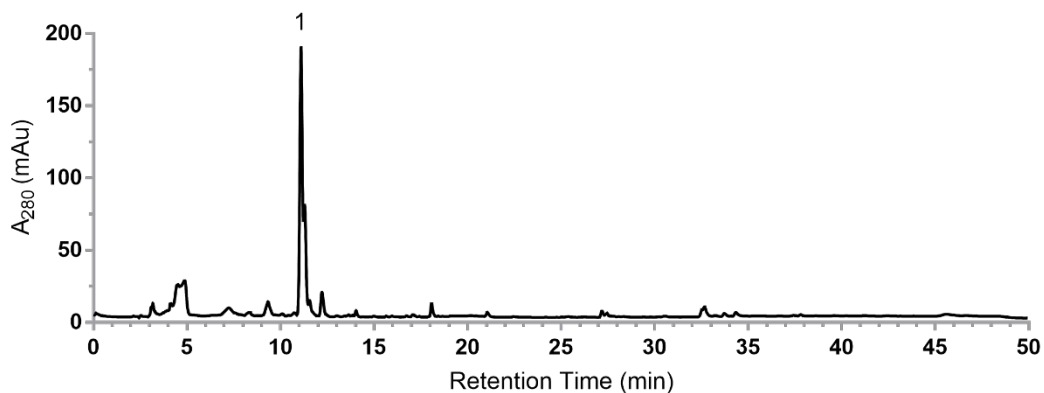
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	270, 280, 390	385.0914	10Mal	-2H ₂ O	O-C	C ₂₀ H ₁₆ O ₈

Appendix Figure B.30. Product profile for starter unit 33 reaction (*S*-(2-acetamidoethyl) but-2-ynethioate)



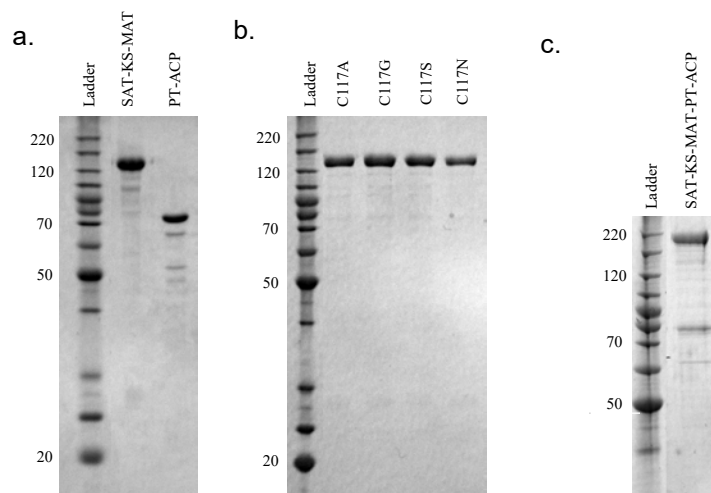
Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	34	N/A		245, 274	224.0749				C ₁₁ H ₁₃ NO ₂ S

Appendix Figure B.31. Product profile for starter unit 34 reaction (*S*-(2-acetamidoethyl) benzoethioate)

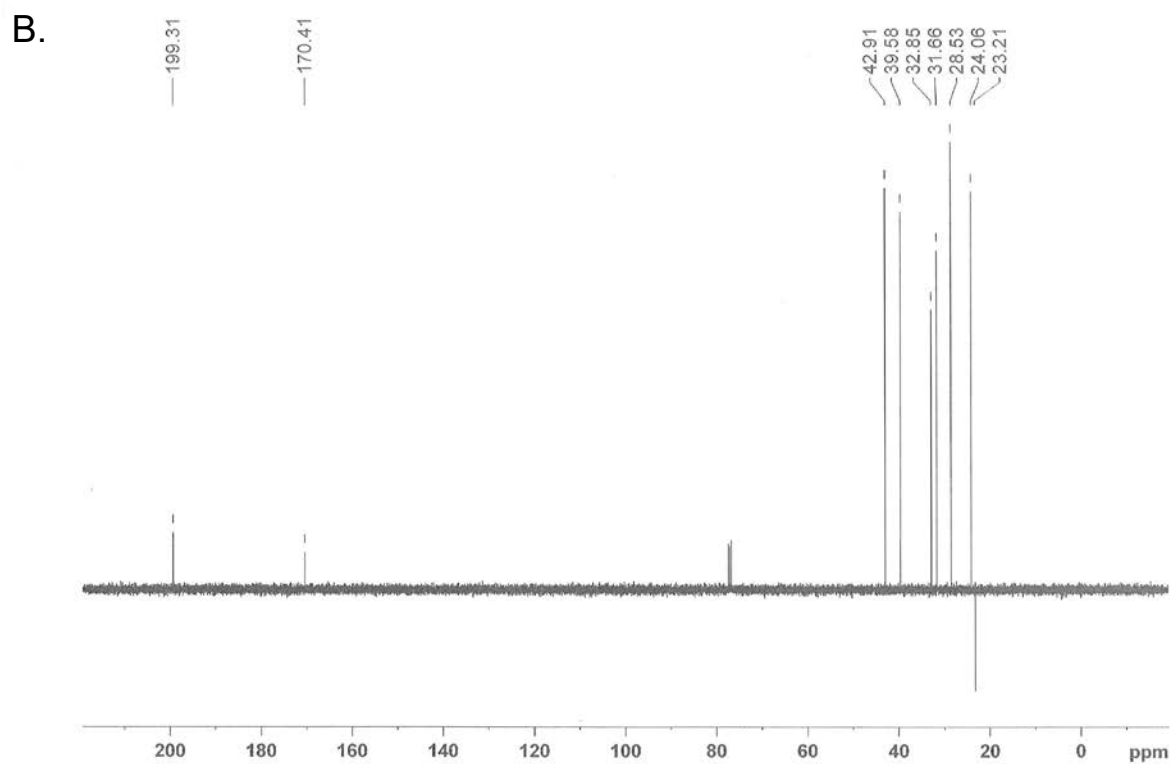
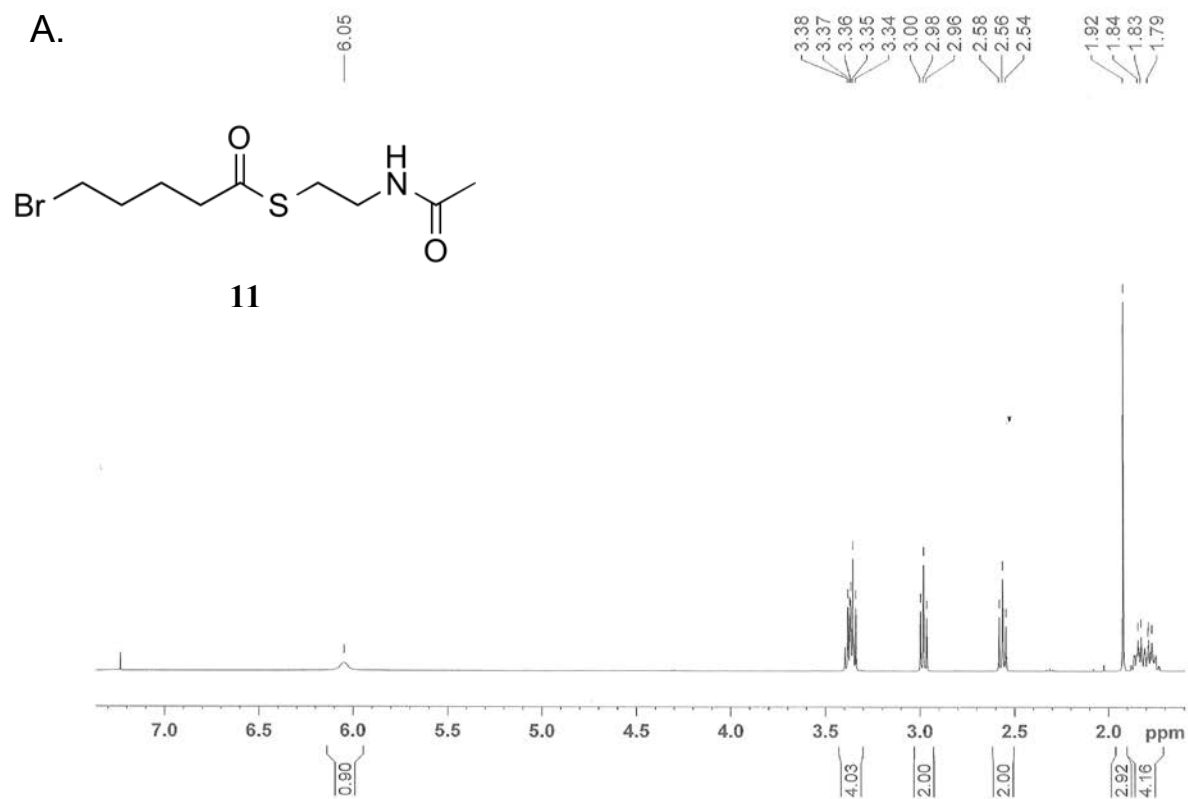


Peak	Compound	Core Structure	R	UV max	[M+H] ⁺	Chain Length	Dehydration	Release, Other	Mol. Formula
1	42	G	CH ₂ COCH ₂ COCH ₂ COCH ₃	270, 280, 390	385.0912	10Mal	-2H ₂ O	O-C	C ₂₀ H ₁₆ O ₈

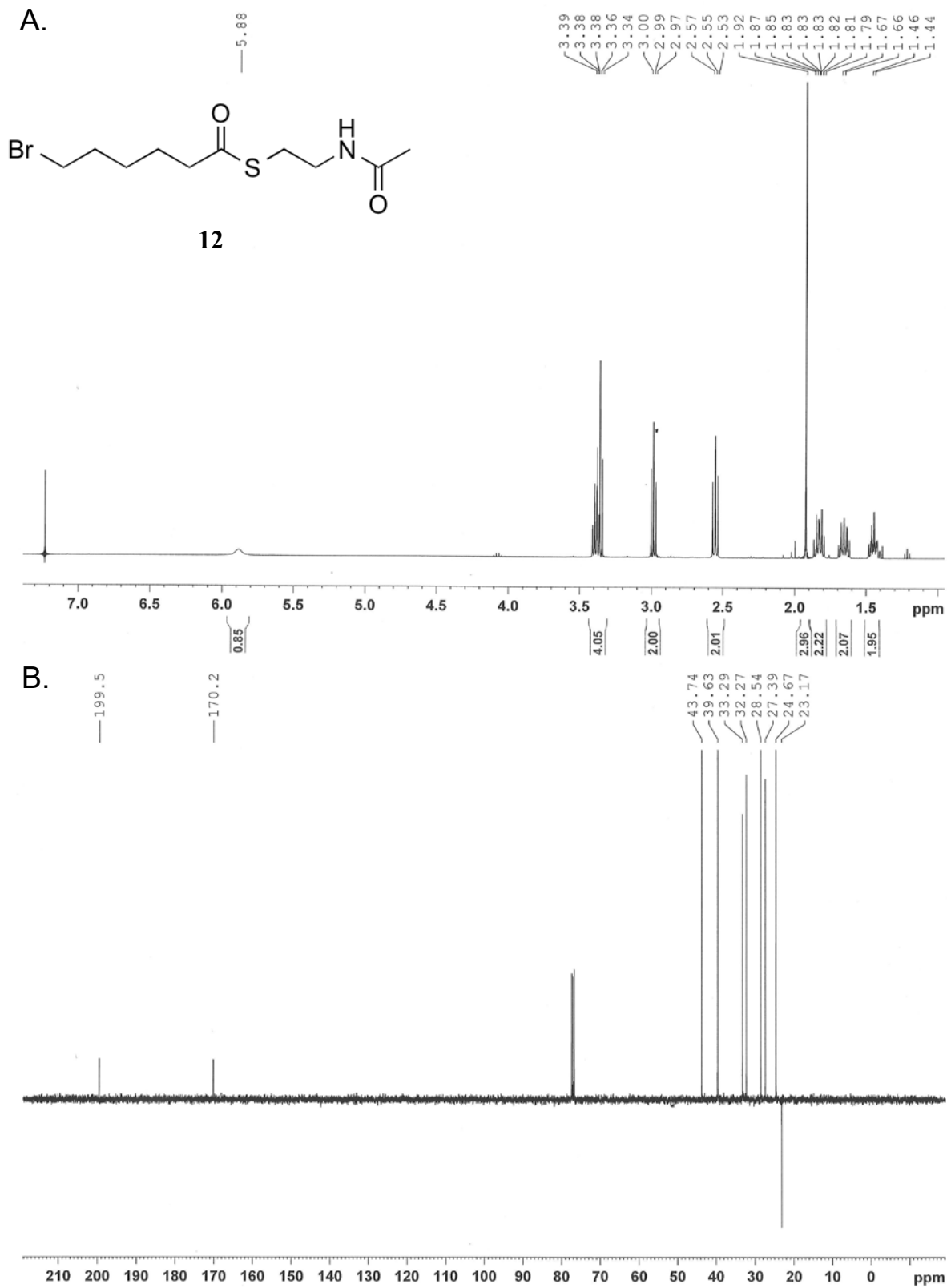
Appendix Figure B.32. Product profile for control reaction containing only malonyl-SNAC



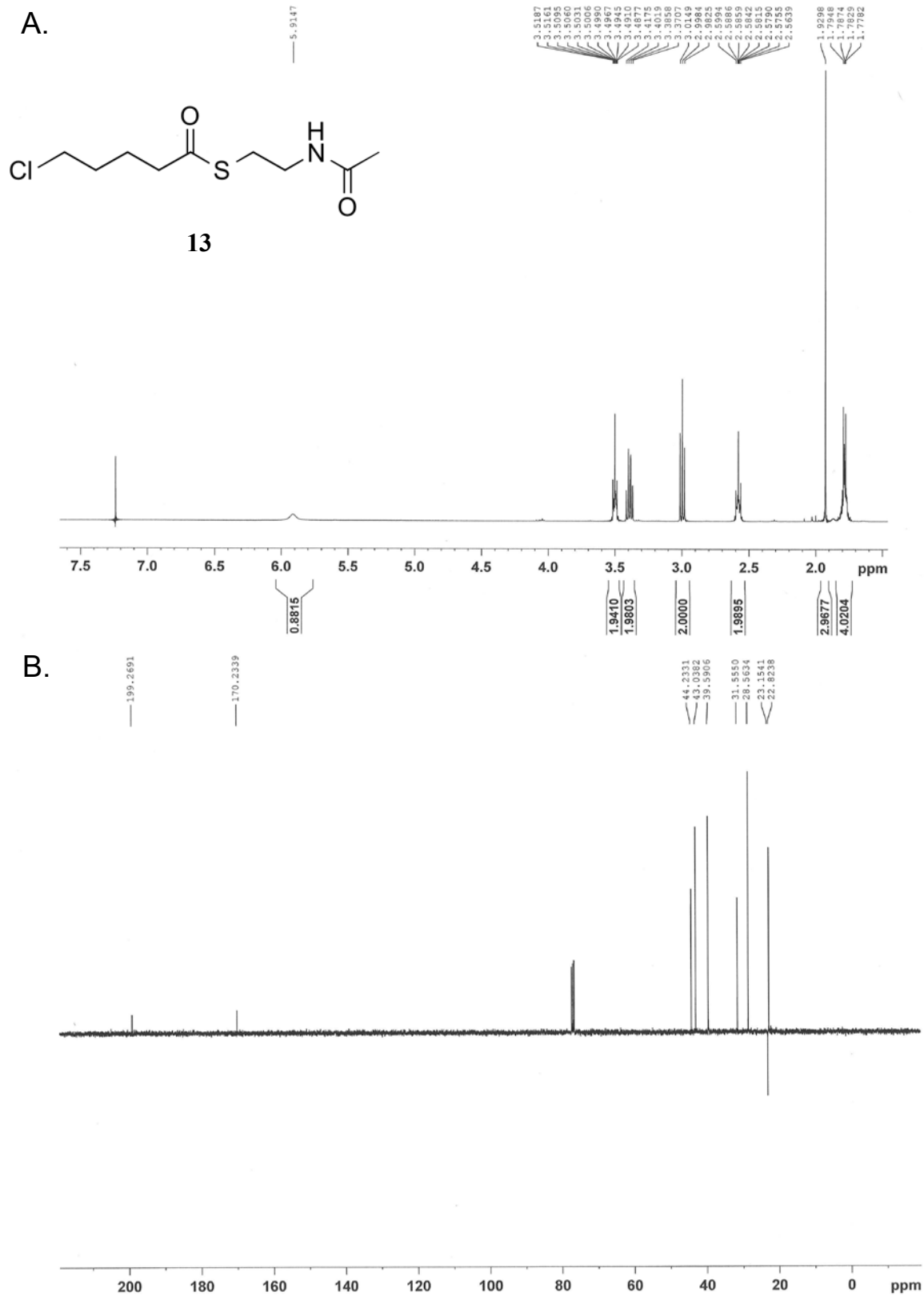
Appendix Figure B.33. Purified proteins used in this study separated by SDS-PAGE and stained with Coomassie Blue. BenchMark Protein Ladder (Invitrogen) was used as a molecular weight standard (indicated in kDa). (A) 12% SDS-PAGE of wild-type PksA SAT-KS-MAT and PT-ACP. (B) 12% SDS-PAGE of PksA SAT-KS-MAT mutants. (C) 12% SDS-PAGE of PksA-SAT-KS-MAT-PT-ACP



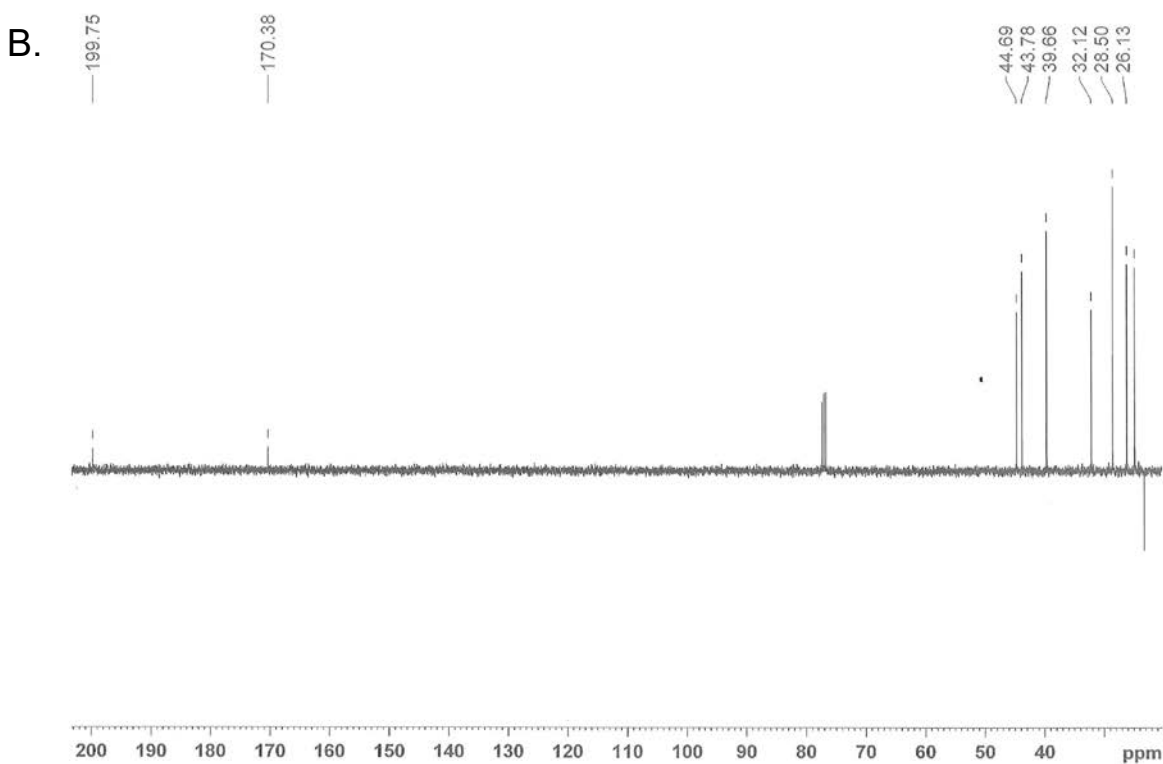
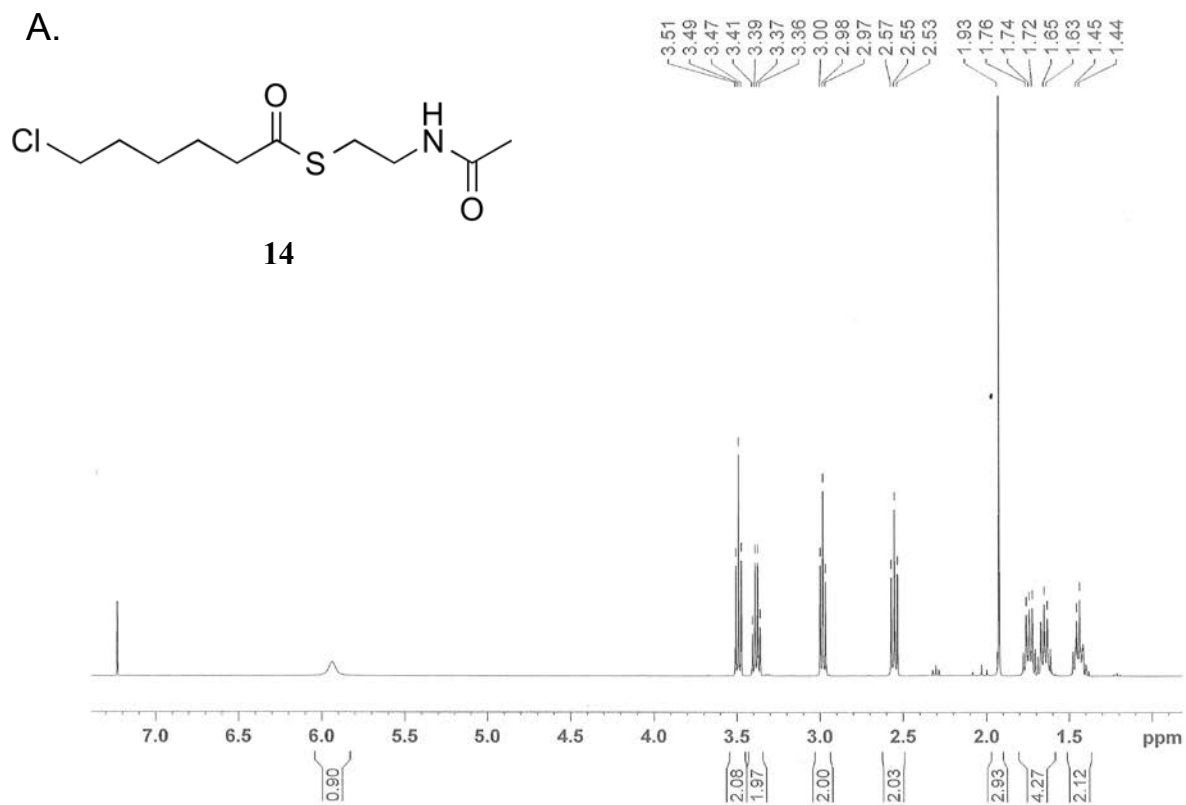
Appendix Figure B.34. NMR spectra of *S*-(2-acetamidoethyl) 5-bromopentanethioate (**11**)
 A) ^1H NMR spectrum B) ^{13}C NMR spectrum



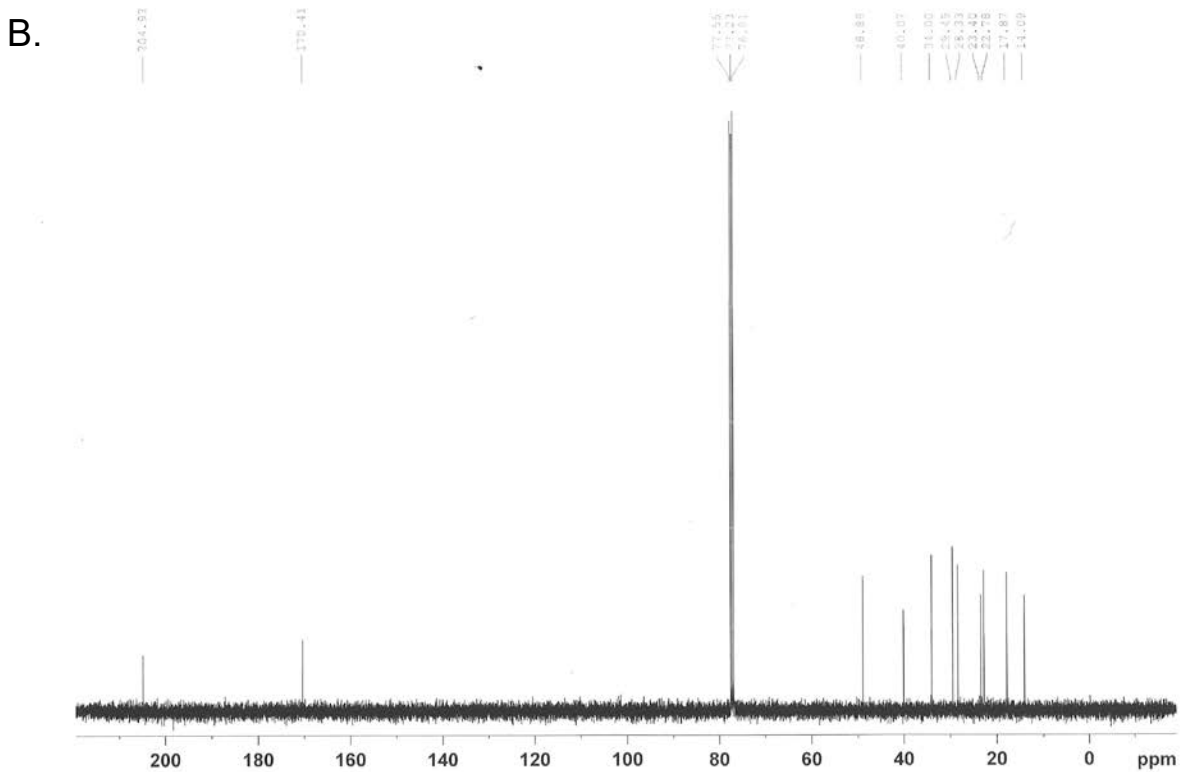
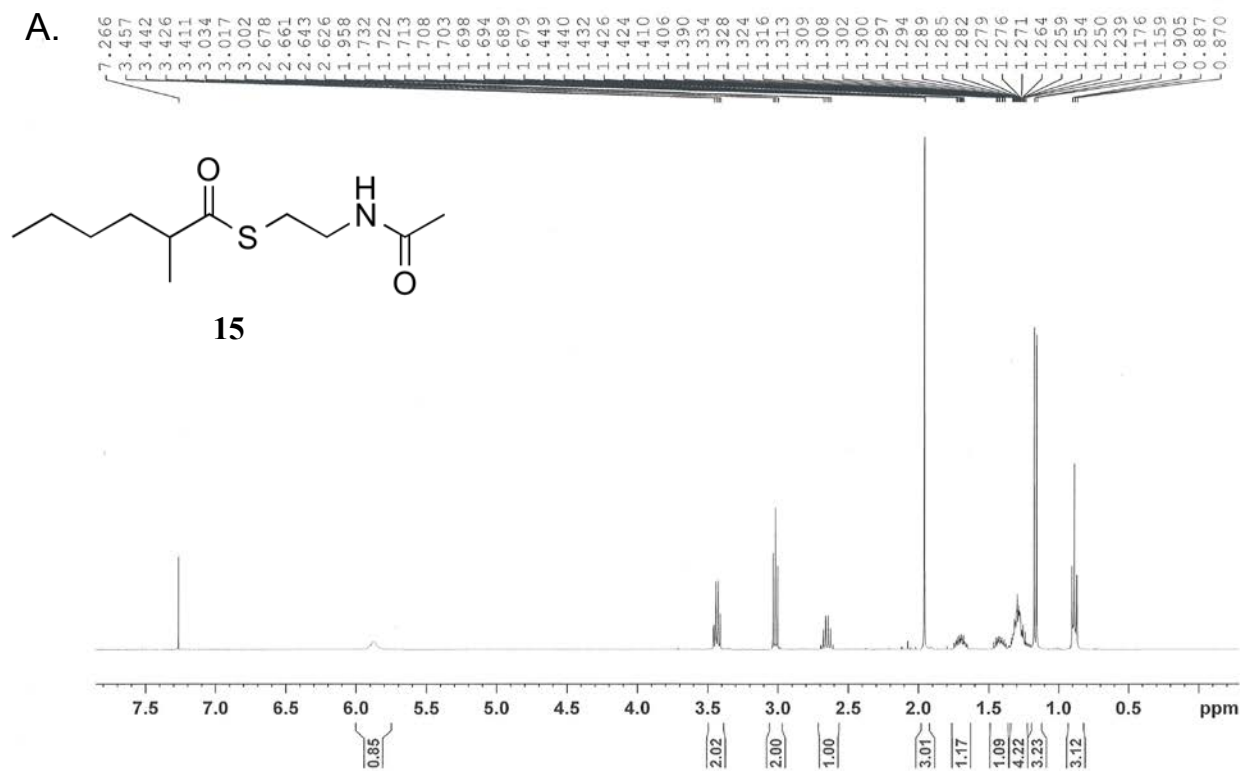
Appendix Figure B.35. NMR spectra of *S*-(2-acetamidoethyl) 6-bromohexanethioate (**12**) A) ^1H NMR spectrum B) ^{13}C NMR spectrum



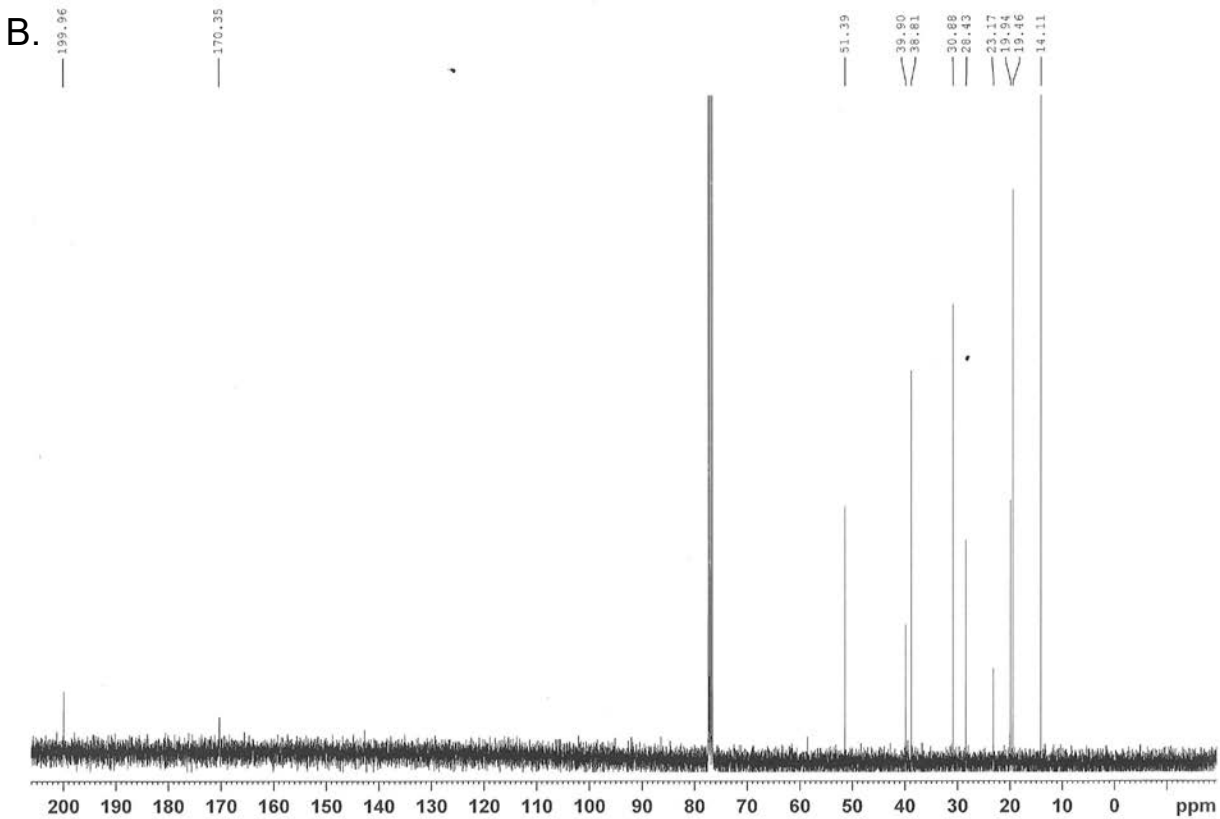
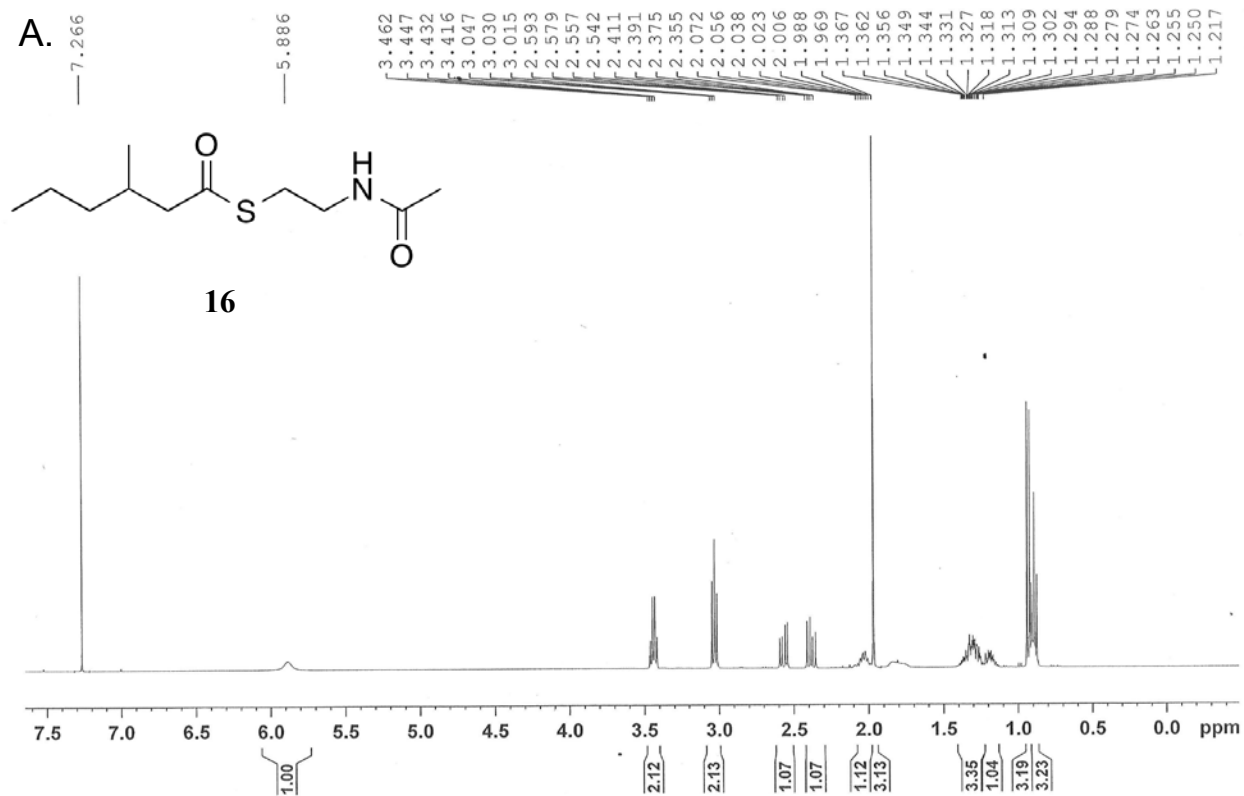
Appendix Figure B.36. NMR spectra of *S*-(2-acetamidoethyl) 5-chloropentanethioate (**13**)
 A) ^1H NMR spectrum B) ^{13}C NMR spectrum



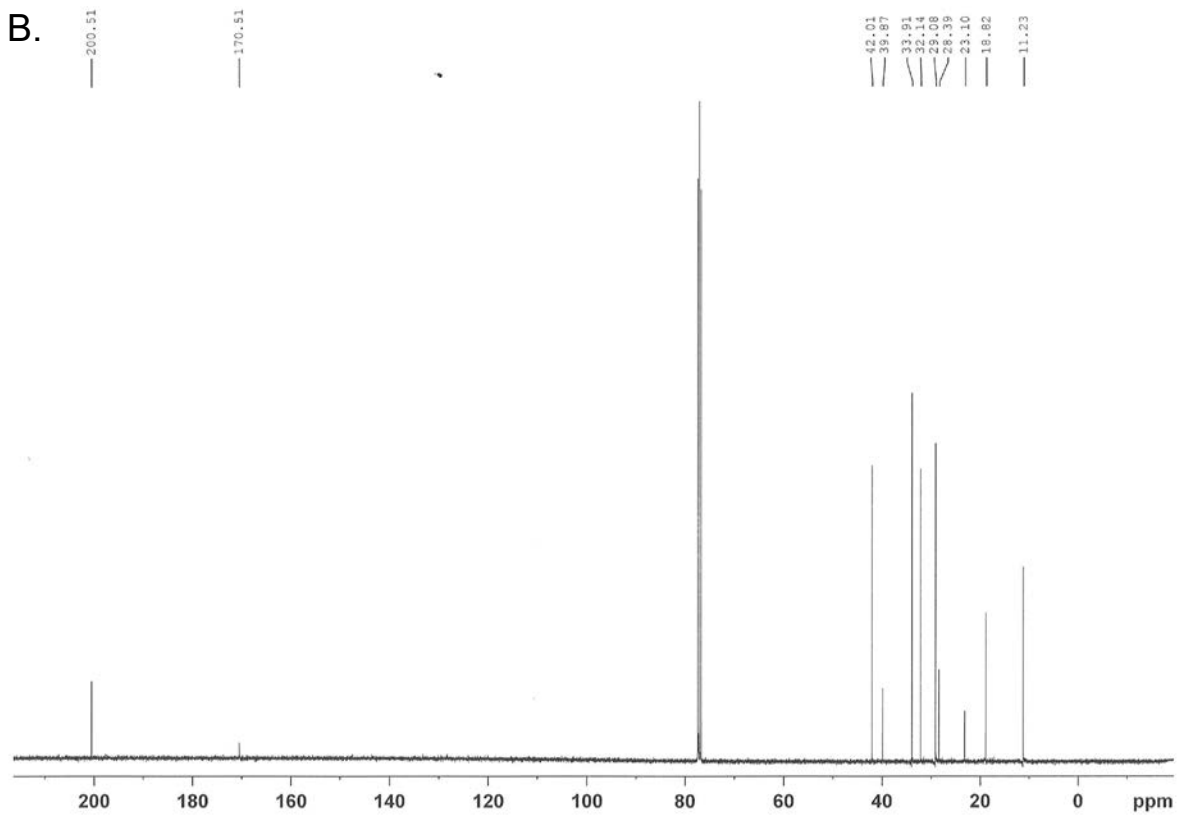
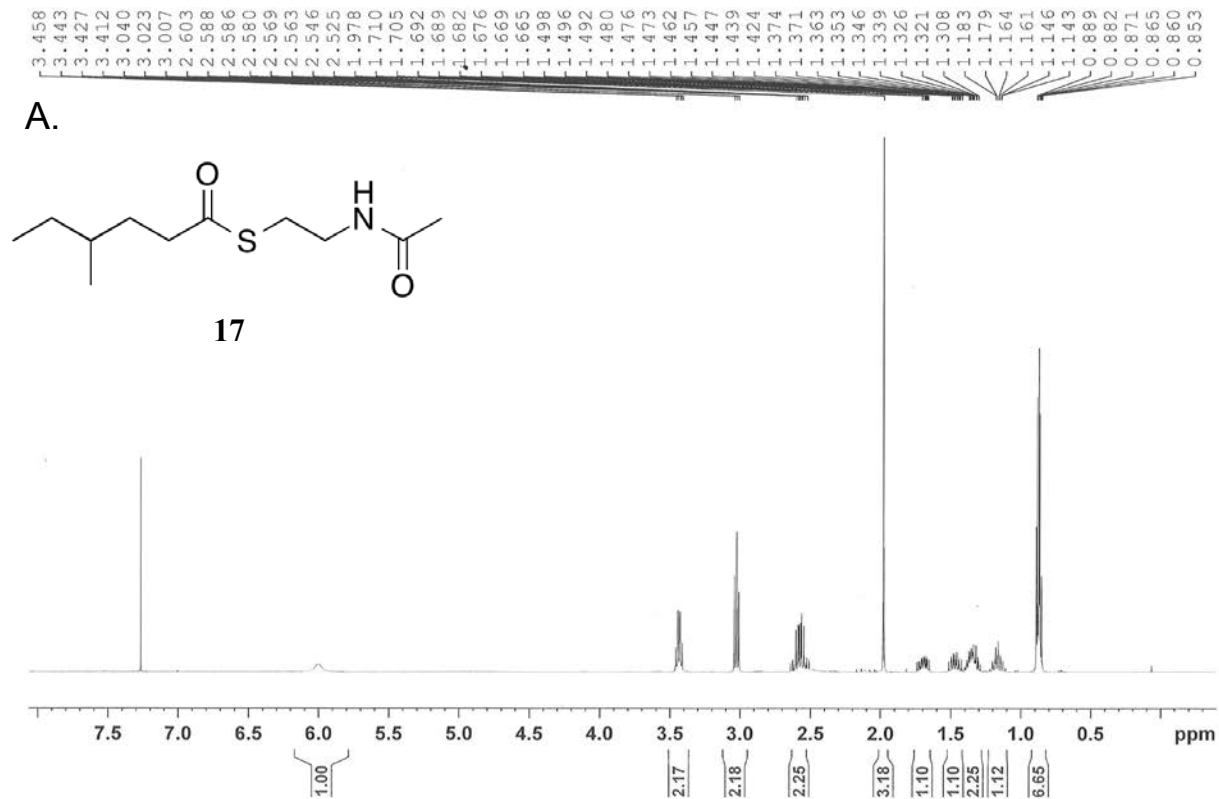
Appendix Figure B.37. NMR spectra of *S*-(2-acetamidoethyl) 6-chlorohexanethioate (**14**) A) ¹H NMR spectrum B) ¹³C NMR spectrum



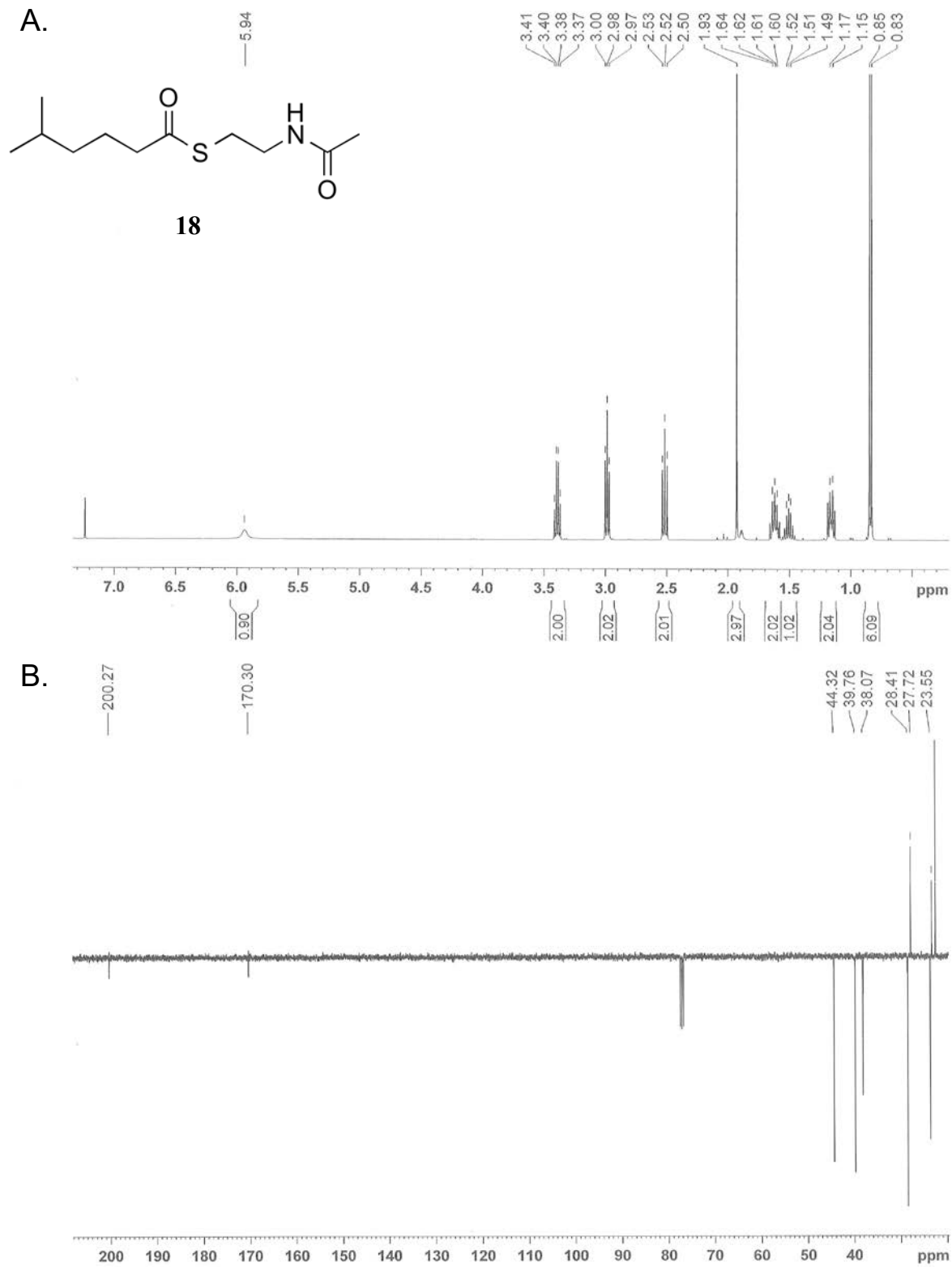
Appendix Figure B.38. NMR spectra of *S*-(2-acetamidoethyl) 2-methylhexanethioate (**15**)
 A) ¹H NMR spectrum B) ¹³C NMR spectrum



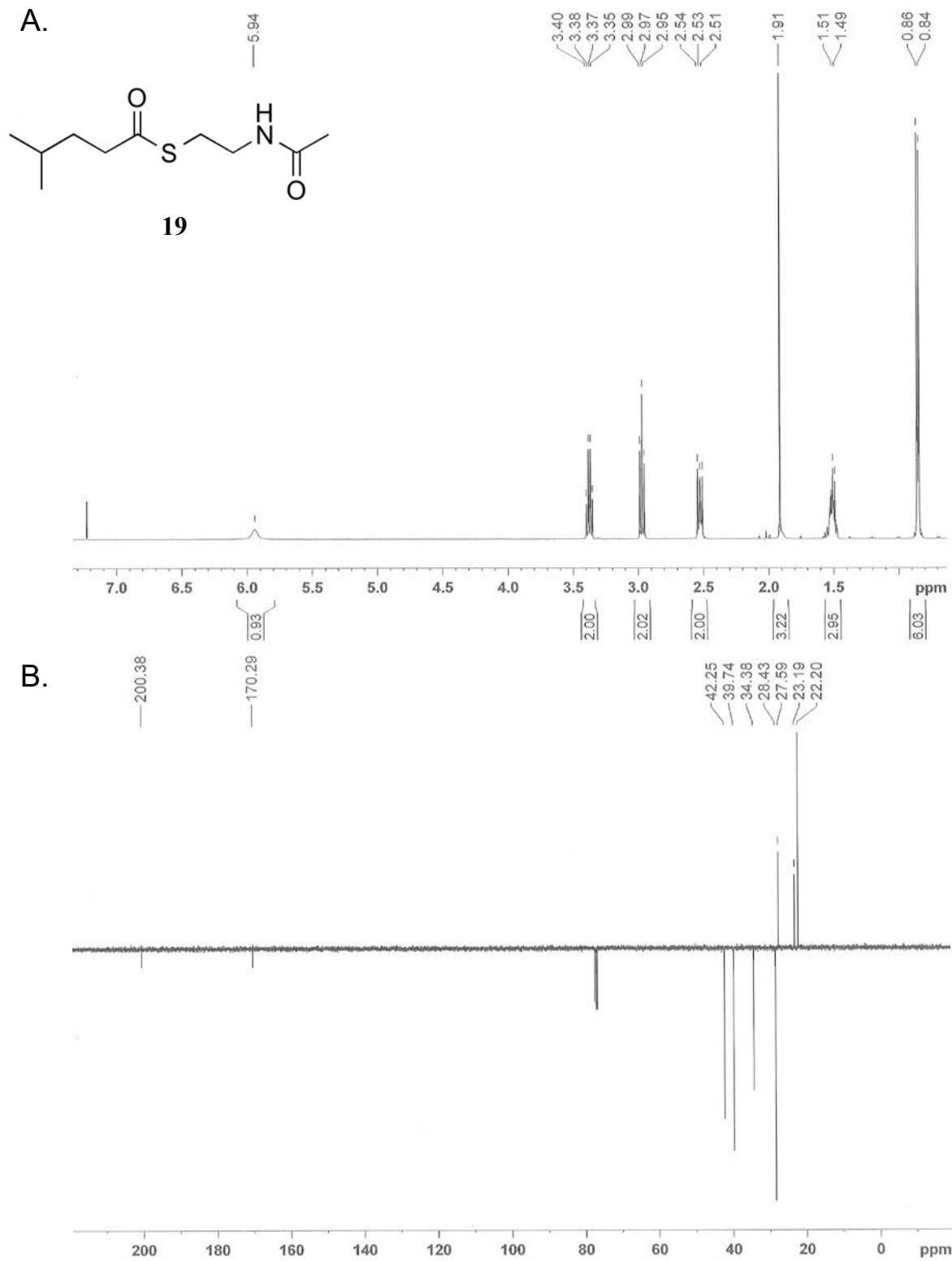
Appendix Figure B.39. NMR spectra of *S*-(2-acetamidoethyl) 3-methylhexanethioate (**16**)
 A) ^1H NMR spectrum B) ^{13}C NMR spectrum



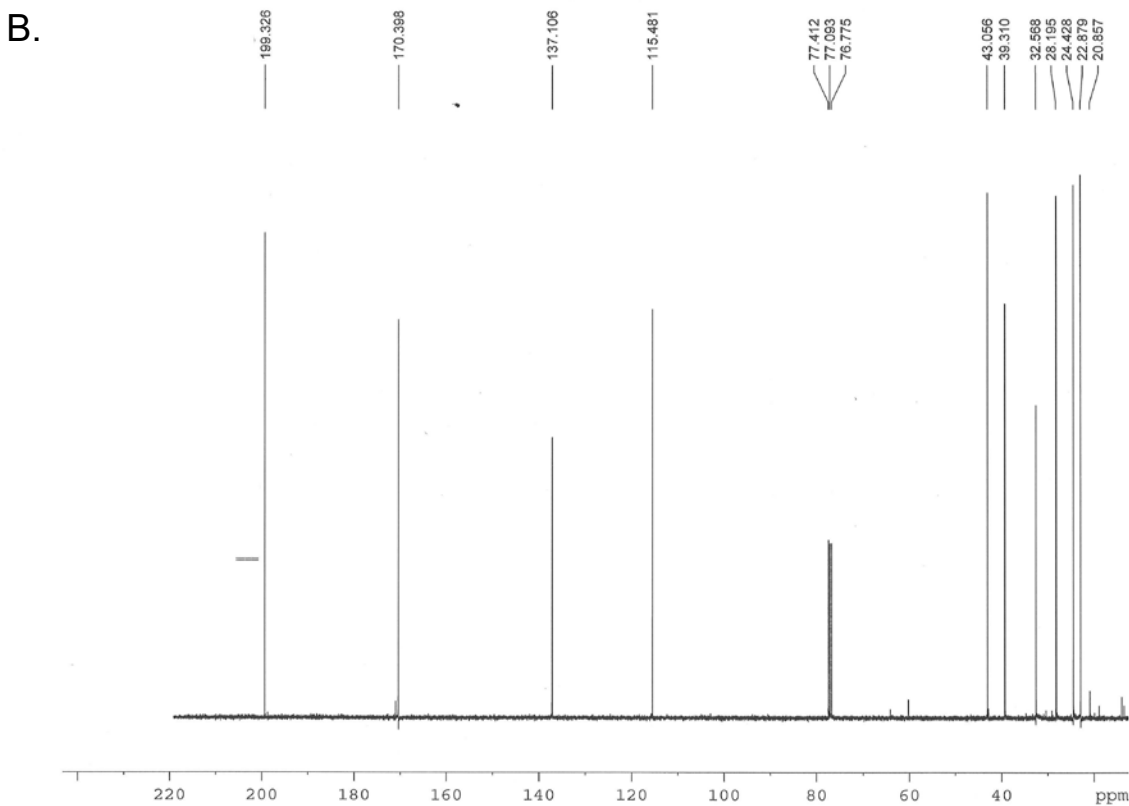
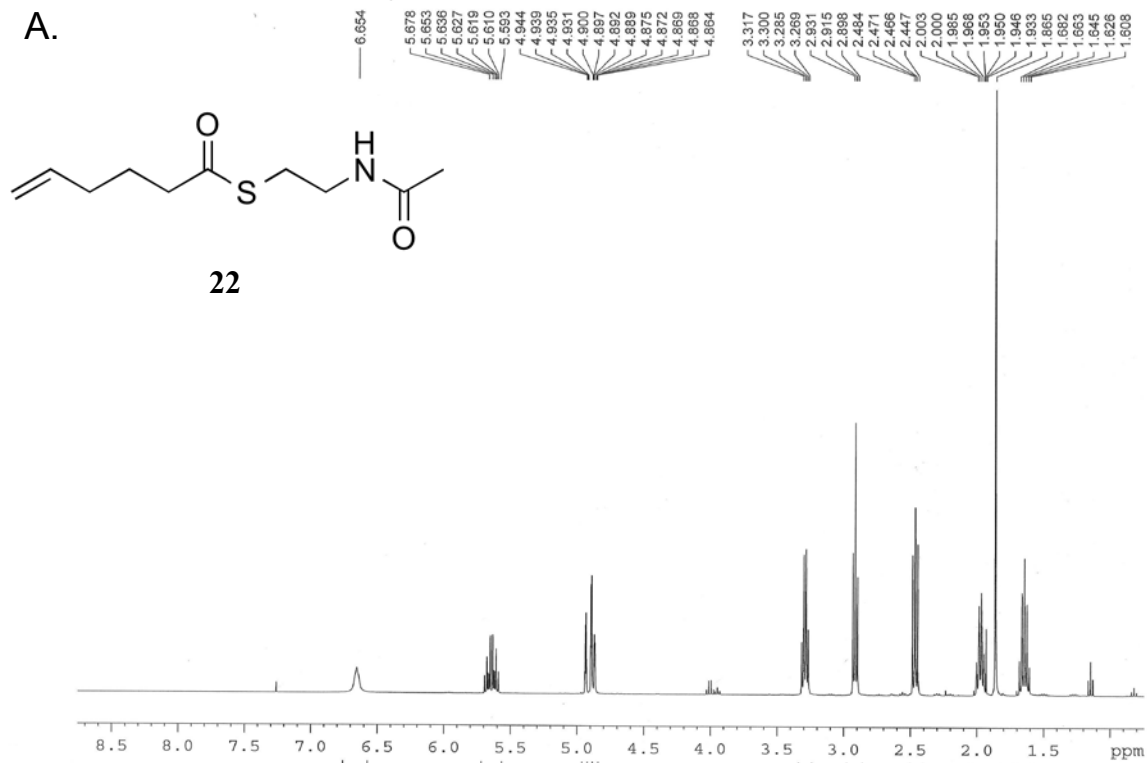
Appendix Figure B.40. NMR spectra of *S*-(2-acetamidoethyl) 4-methylhexanethioate (**17**) A) ¹H NMR spectrum B) ¹³C NMR spectrum



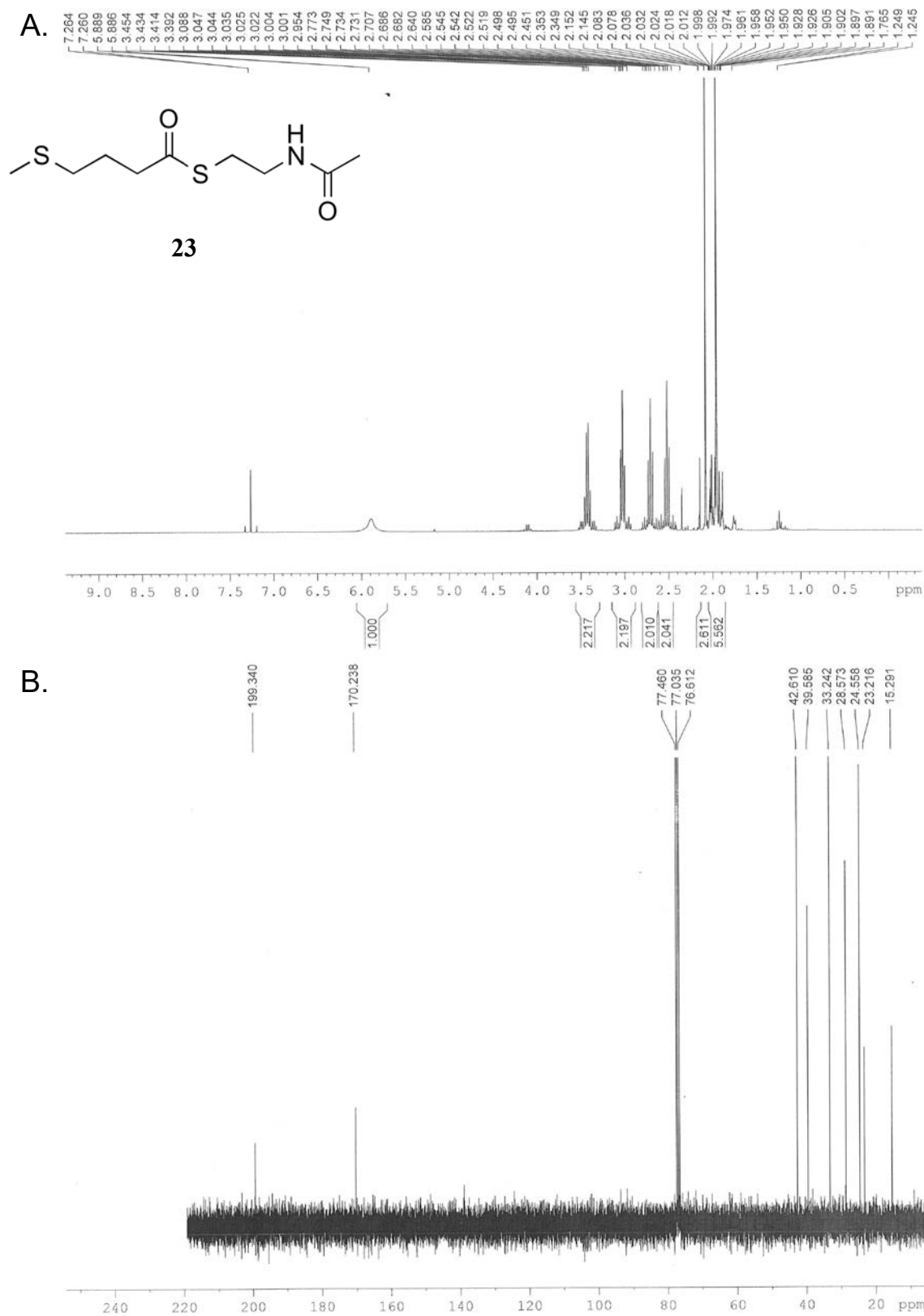
Appendix Figure B.41. NMR spectra of *S*-(2-acetamidoethyl) 5-methylhexanethioate (**18**)
 A) ^1H NMR spectrum B) ^{13}C NMR spectrum



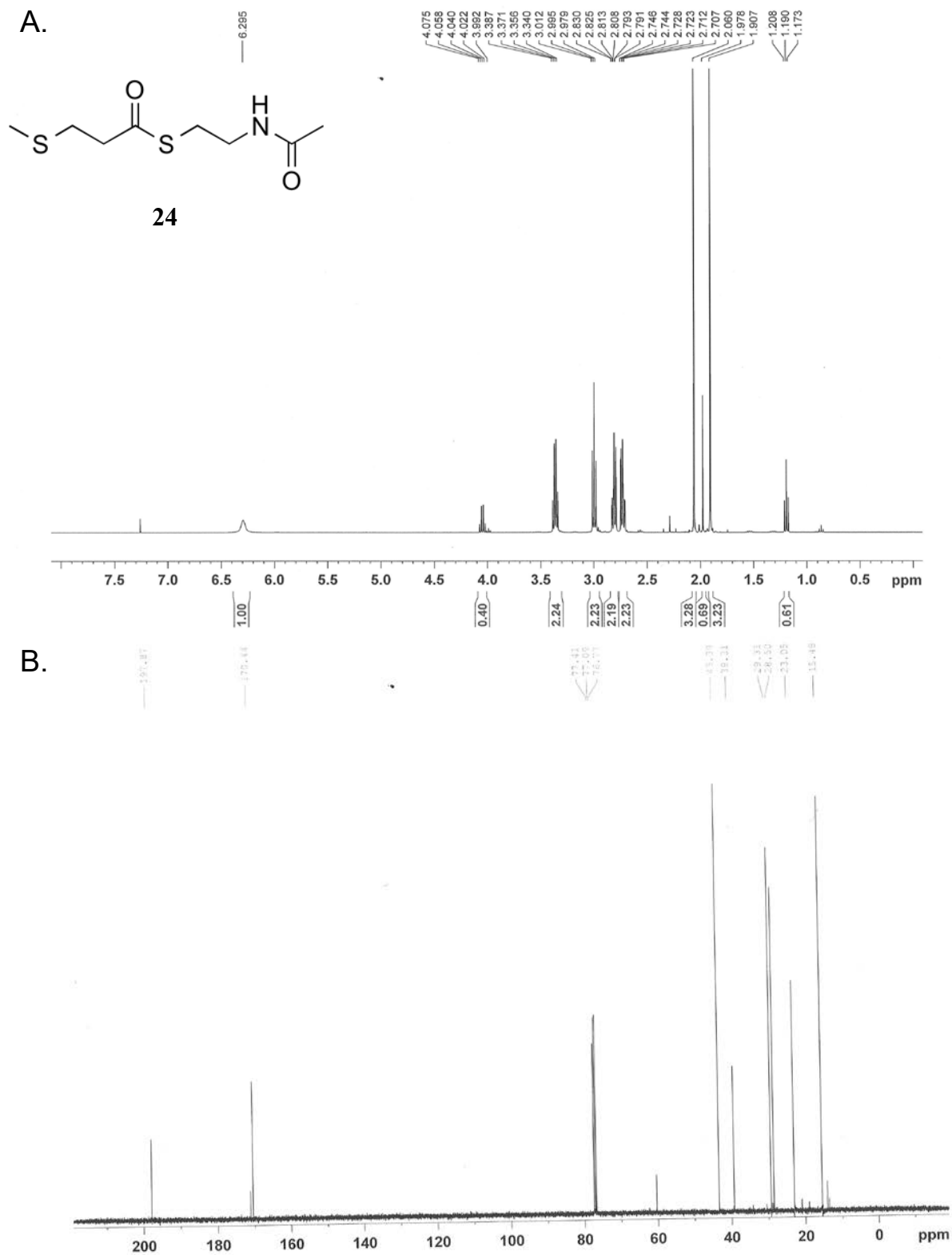
Appendix Figure B.42. NMR spectra of *S*-(2-acetamidoethyl) 4-methylpentanethioate (**19**)
 A) ^1H NMR spectrum B) ^{13}C NMR spectrum



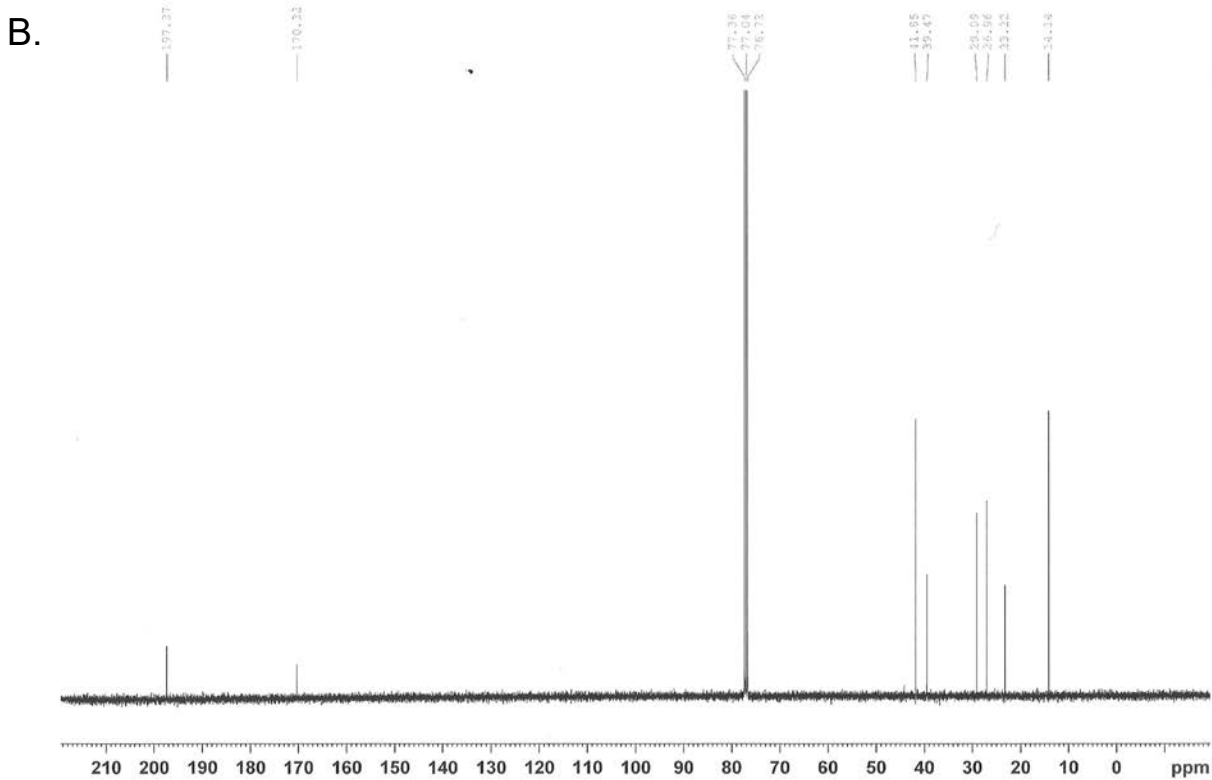
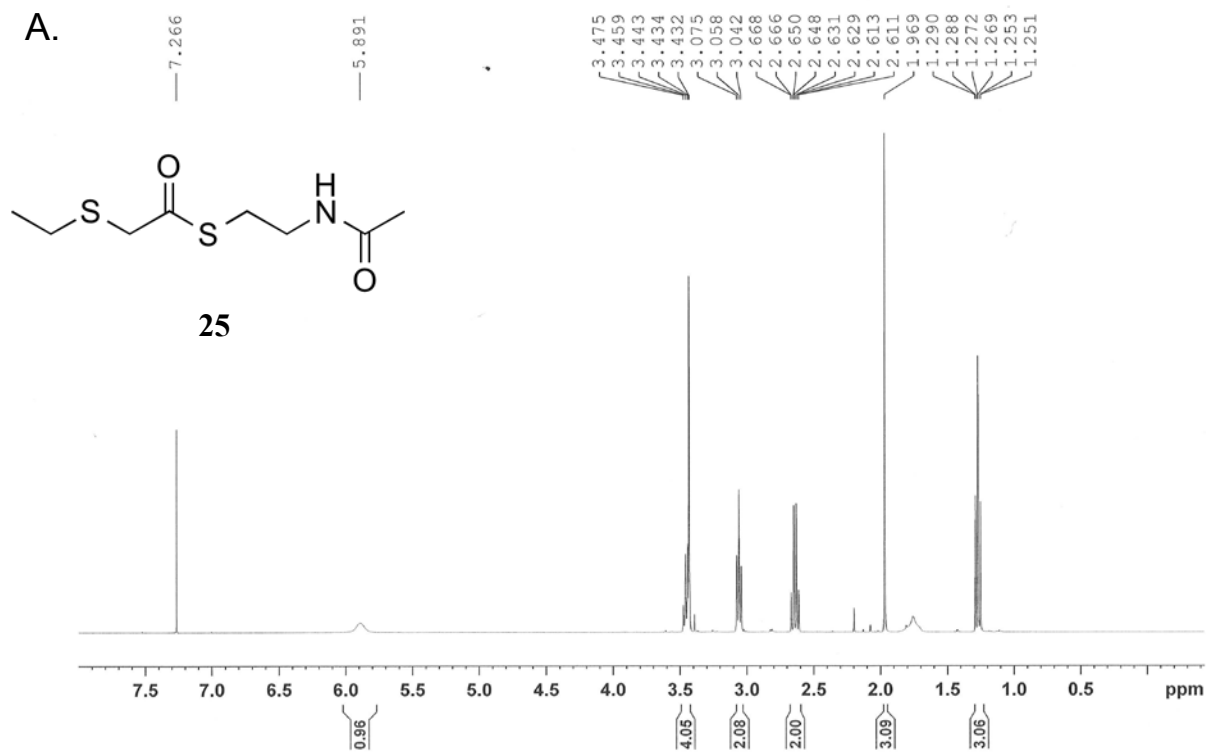
Appendix Figure B.43. NMR spectra of *S*-(2-acetamidoethyl) hex-5-enethioate (**22**) A) ¹H NMR spectrum B) ¹³C NMR spectrum



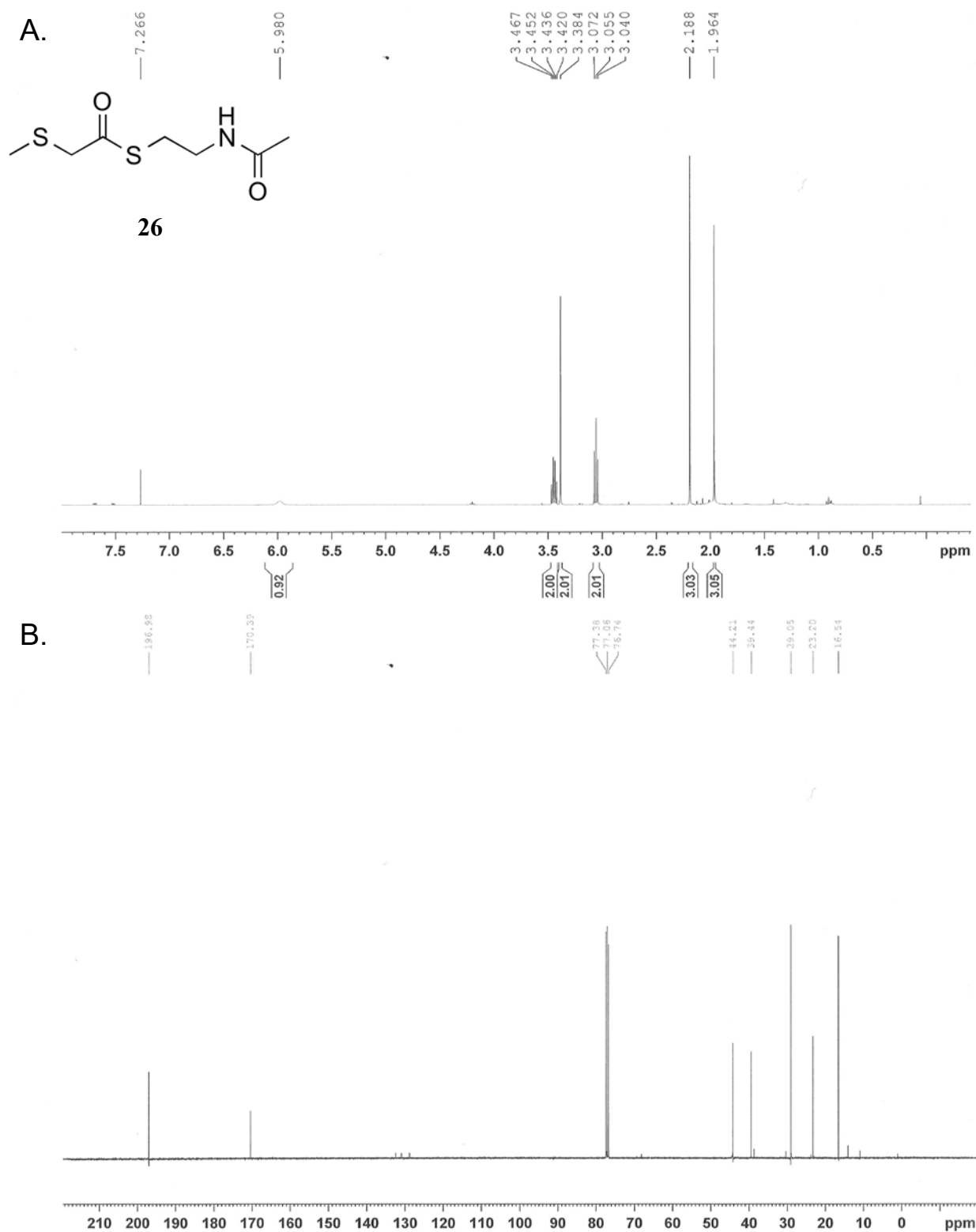
Appendix Figure B.44. NMR spectra of *S*-(2-acetamidoethyl) 4-(methylthiol)butanethioate (**23**) A) ^1H NMR spectrum B) ^{13}C NMR spectrum



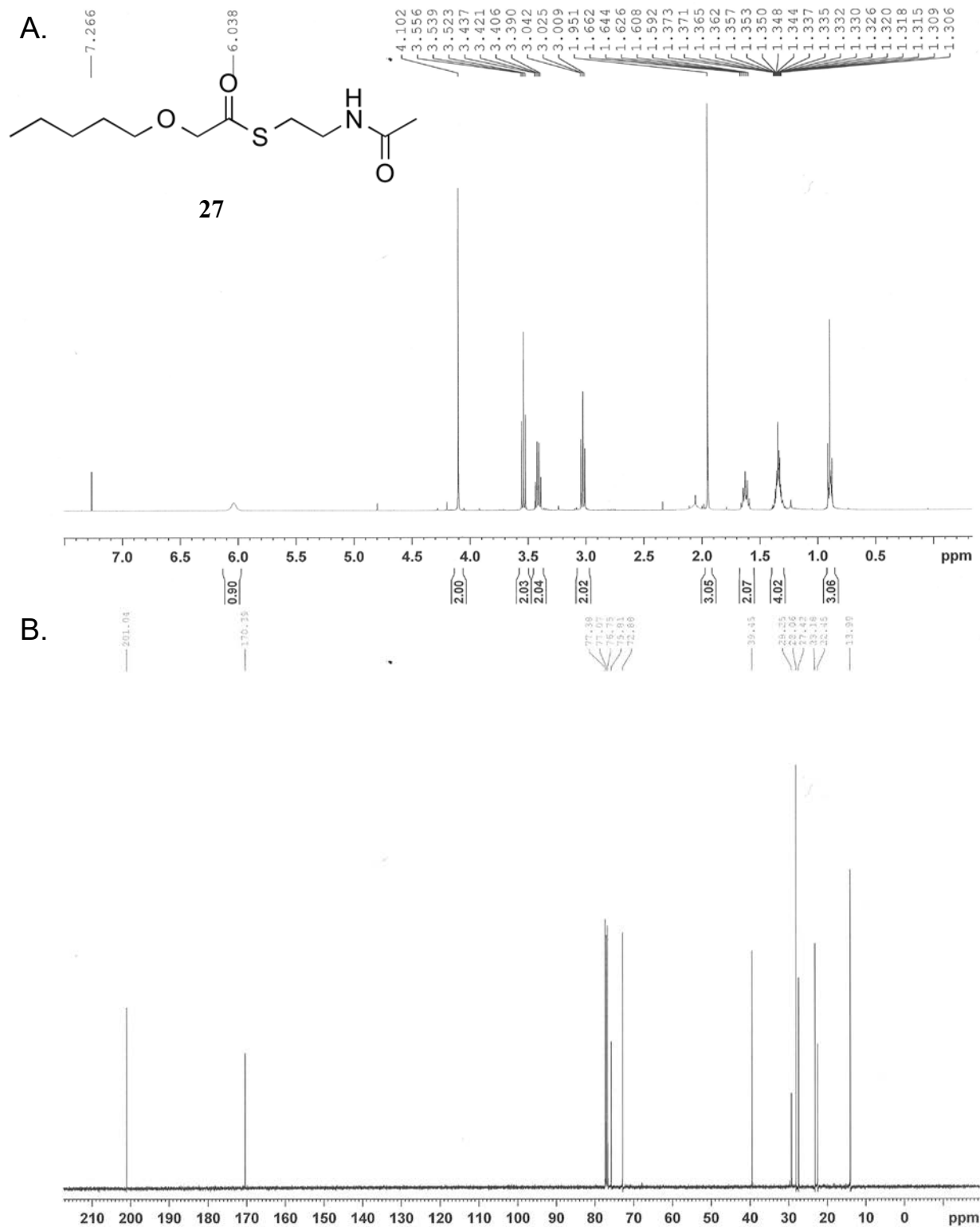
Appendix Figure B.45. NMR spectra of *S*-(2-acetamidoethyl) 3-(methylthio)propanethioate (**24**) A) ¹H NMR spectrum B) ¹³C NMR spectrum



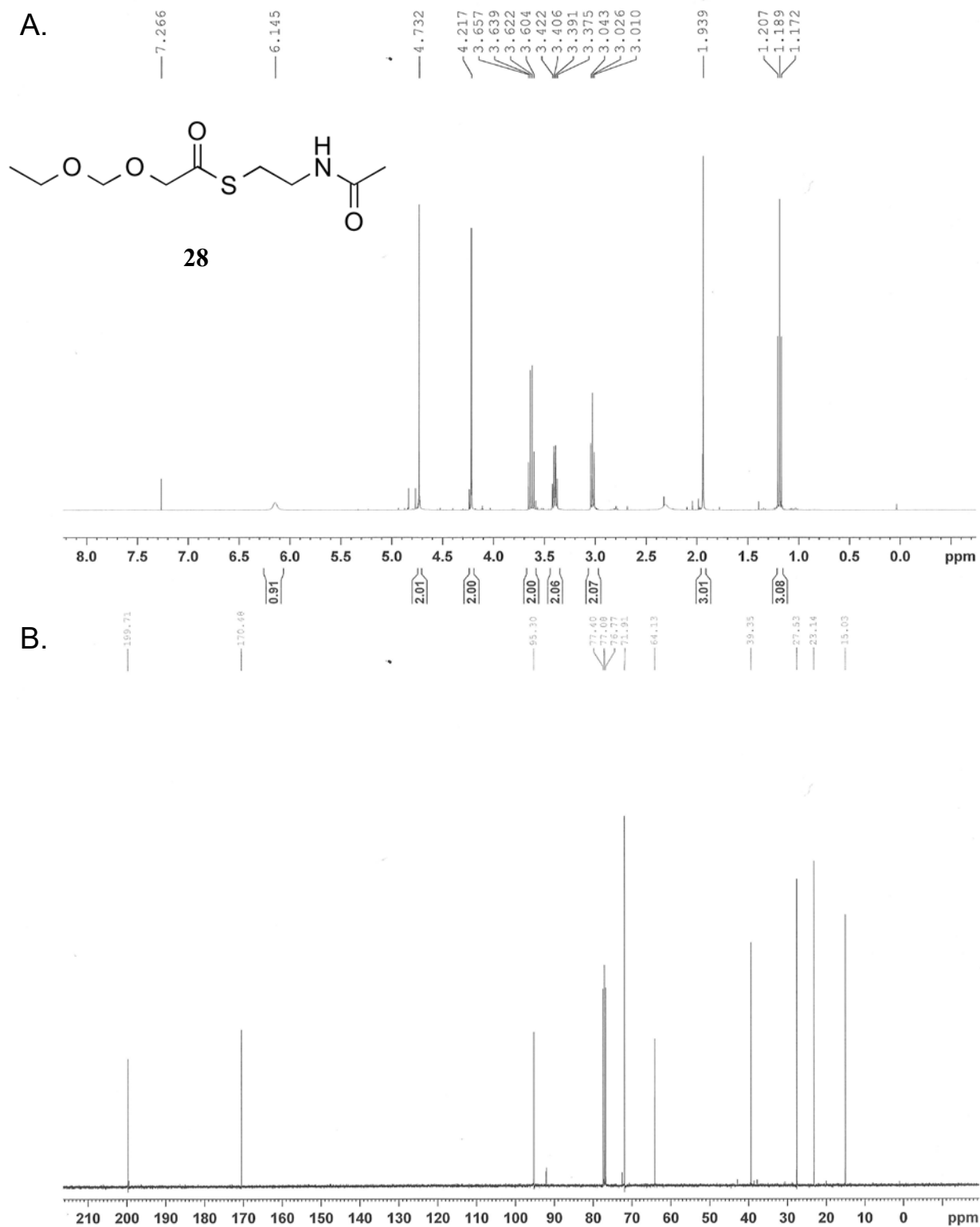
Appendix Figure B.46. NMR spectra of *S*-(2-acetamidoethyl) 2-(ethylthio)ethanethioate (**25**) A) ^1H NMR spectrum B) ^{13}C NMR spectrum



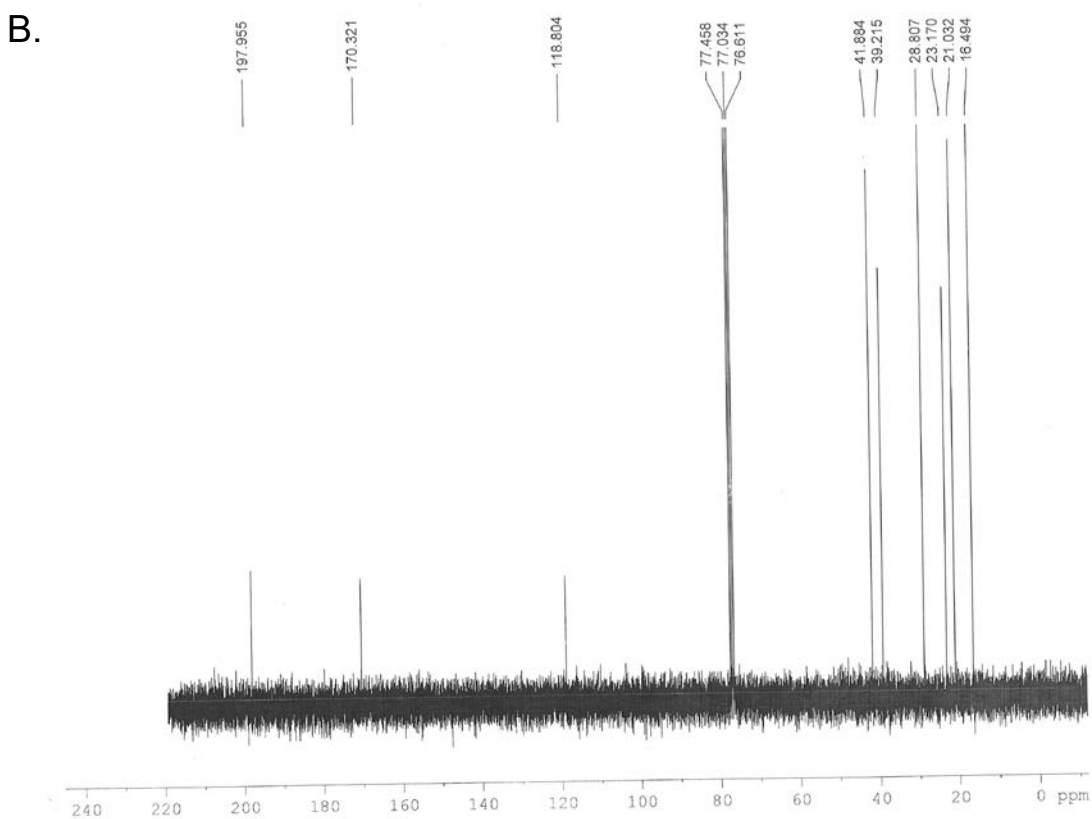
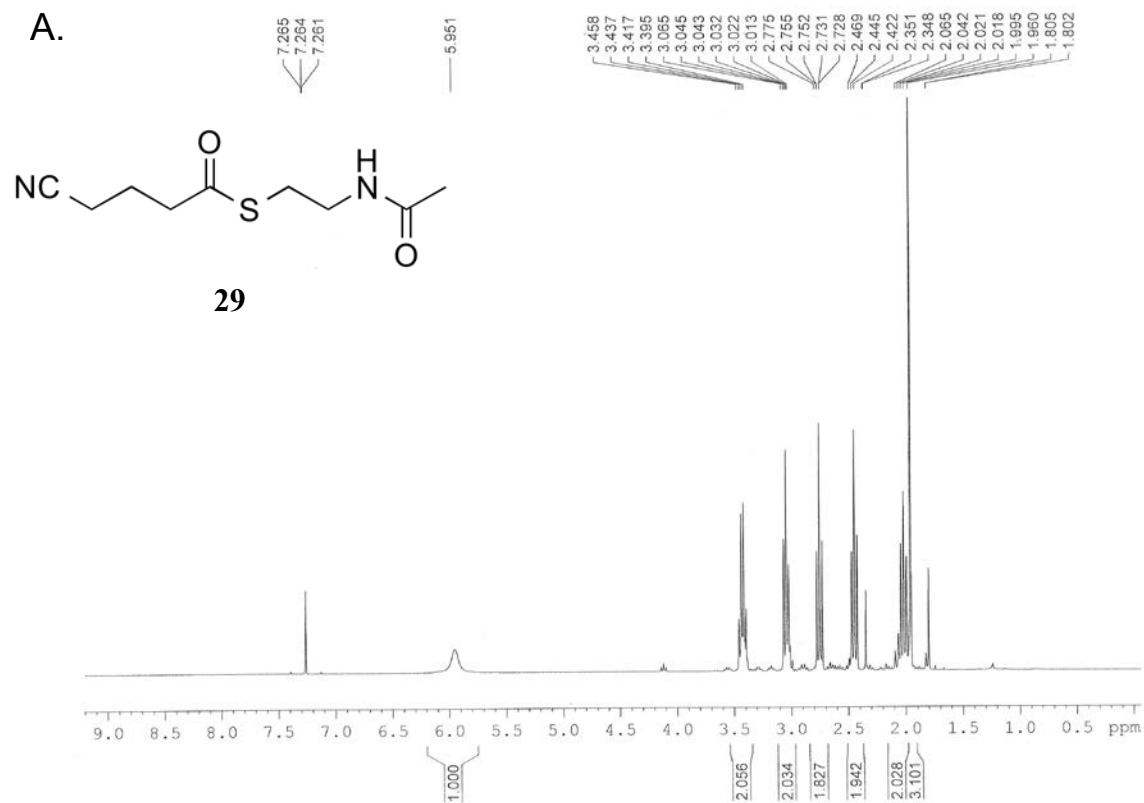
Appendix Figure B.47. NMR spectra of *S*-(2-acetamidoethyl) 2-(methylthio)ethanethioate (**26**) A) ^1H NMR spectrum B) ^{13}C NMR spectrum



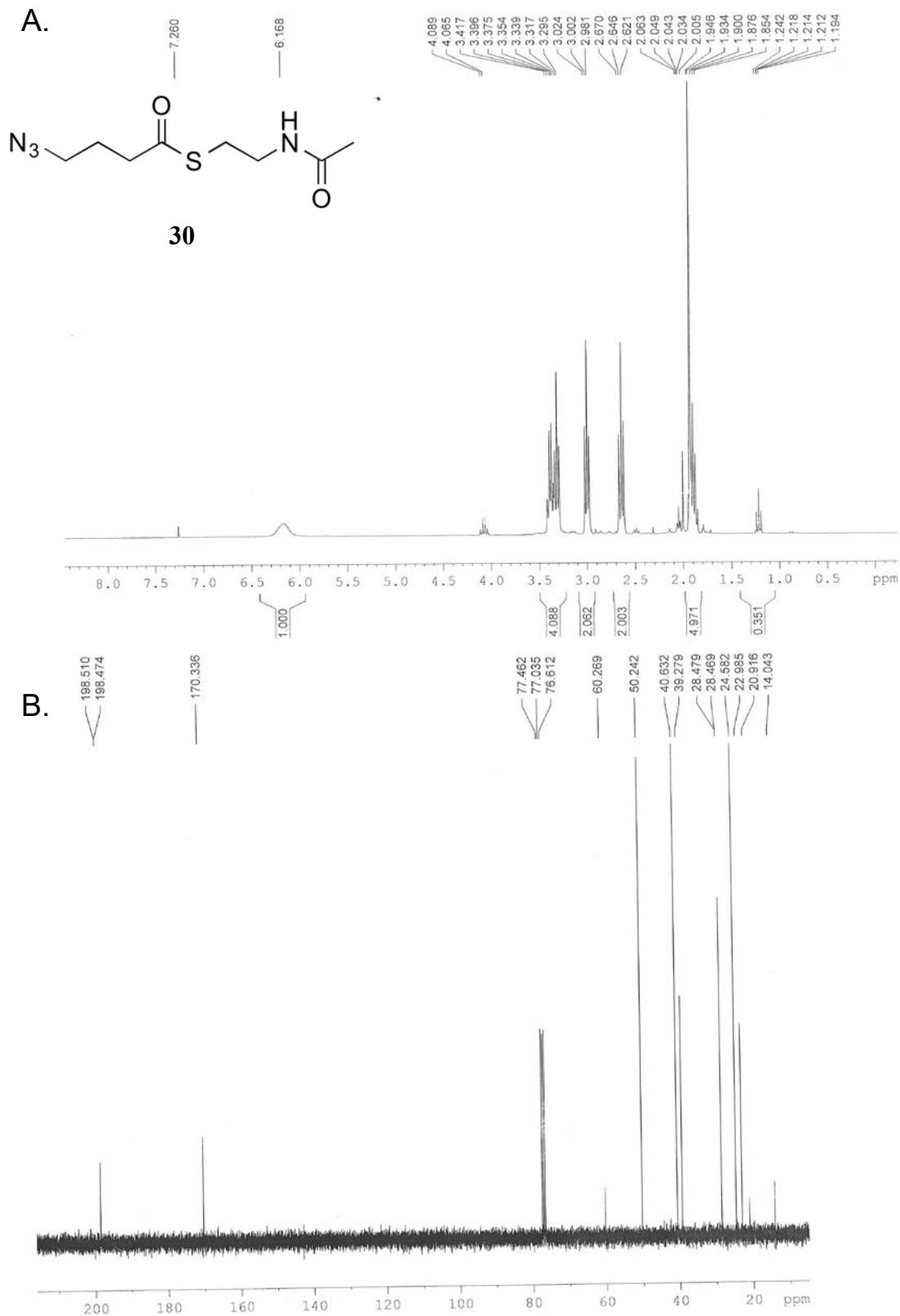
Appendix Figure B.48. NMR spectra of *S*-(2-acetamidoethyl) 2-(pentyloxy)ethanethioate (**27**) (A) ^1H NMR spectrum (B) ^{13}C NMR spectrum



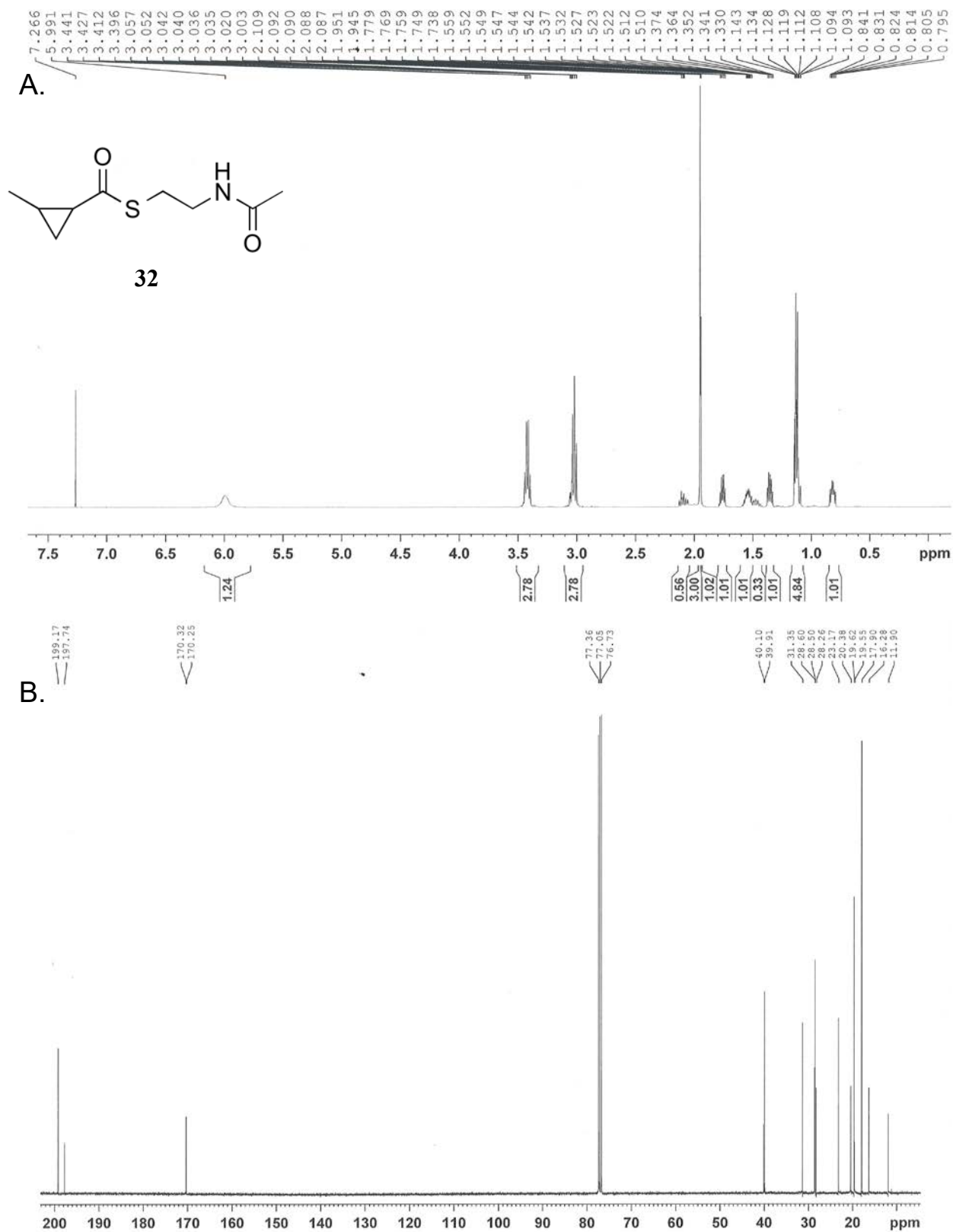
Appendix Figure B.49. NMR spectra of *S*-(2-acetamidoethyl) 2-(ethoxymethoxy)ethanethioate (**28**) (A) ¹H NMR spectrum B) ¹³C NMR spectrum



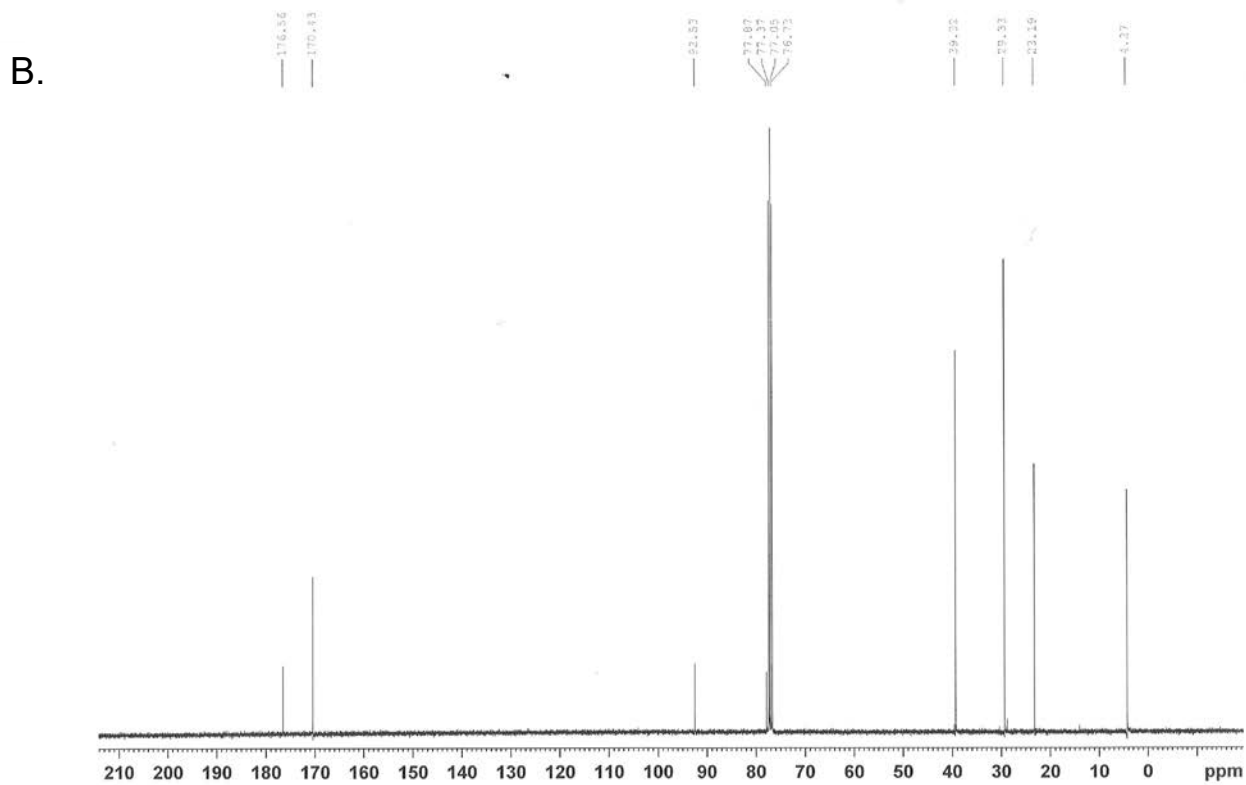
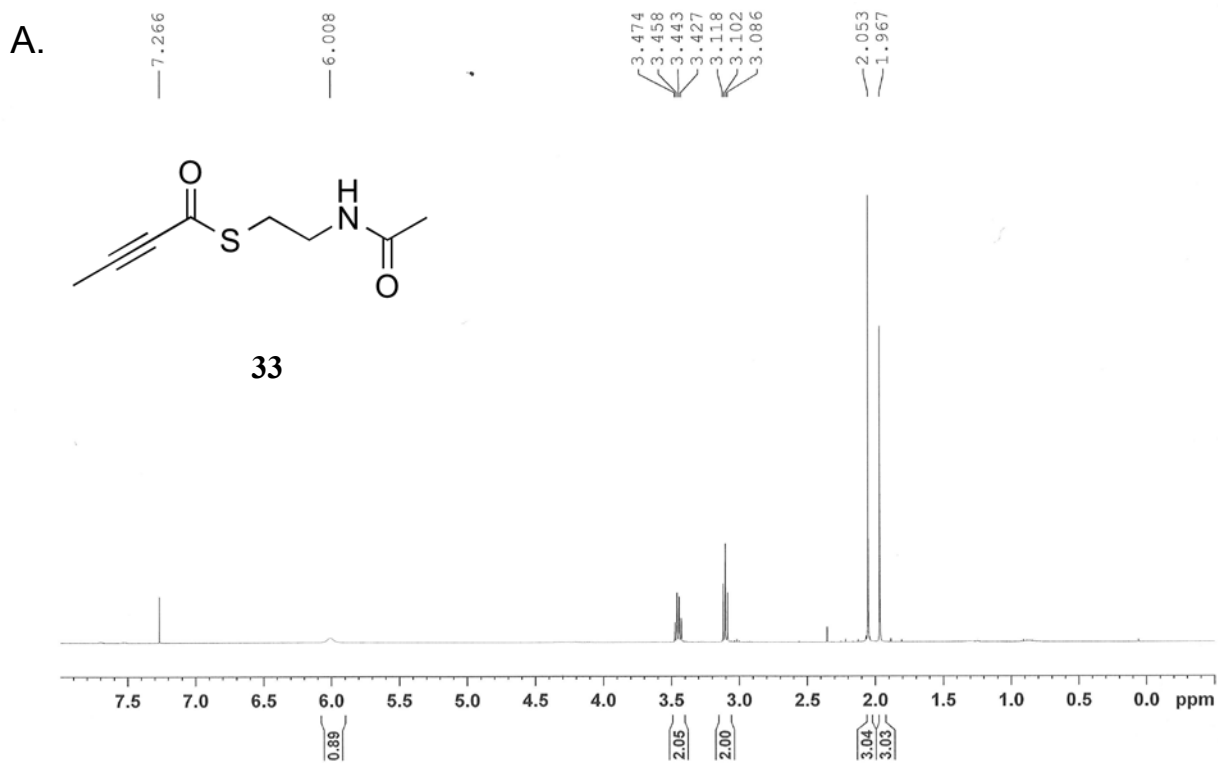
Appendix Figure B.50. NMR spectra of *S*-(2-acetamidoethyl) 4-cyanobutanethioate (**29**) (A) ^1H NMR spectrum B) ^{13}C NMR spectrum



Appendix Figure B.51. NMR spectra of *S*-(2-acetamidoethyl) 4-azidobutanethioate (**30**) (A) ^1H NMR spectrum B) ^{13}C NMR spectrum



Appendix Figure B.52. NMR spectra of *S*-(2-acetamidoethyl) 2-methylcyclopropane-1-carbothioate (**32**) A) ^1H NMR spectrum B) ^{13}C NMR spectrum



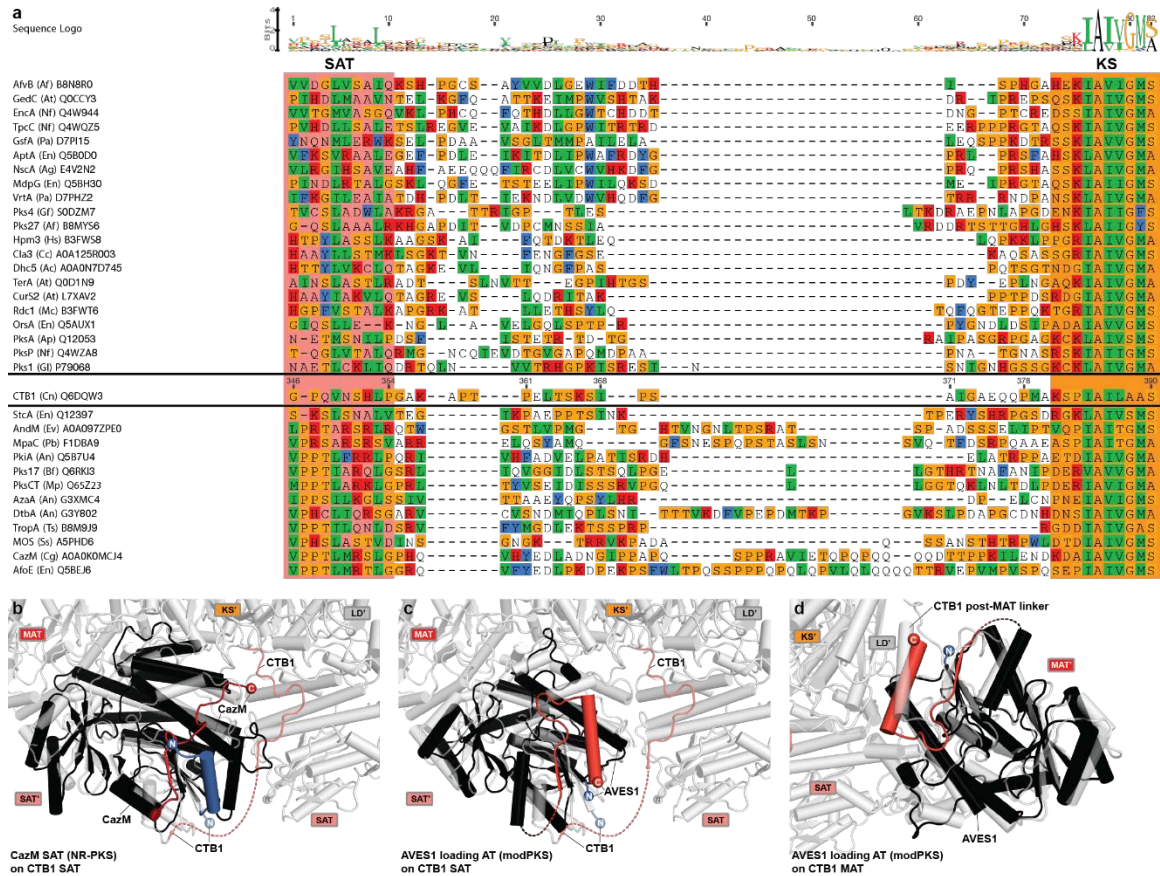
Appendix Figure B.53. NMR spectra of *S*-(2-acetamidoethyl) but-2-ynethioate (**33**) A) ^1H NMR spectrum B) ^{13}C NMR spectrum

B.4. References

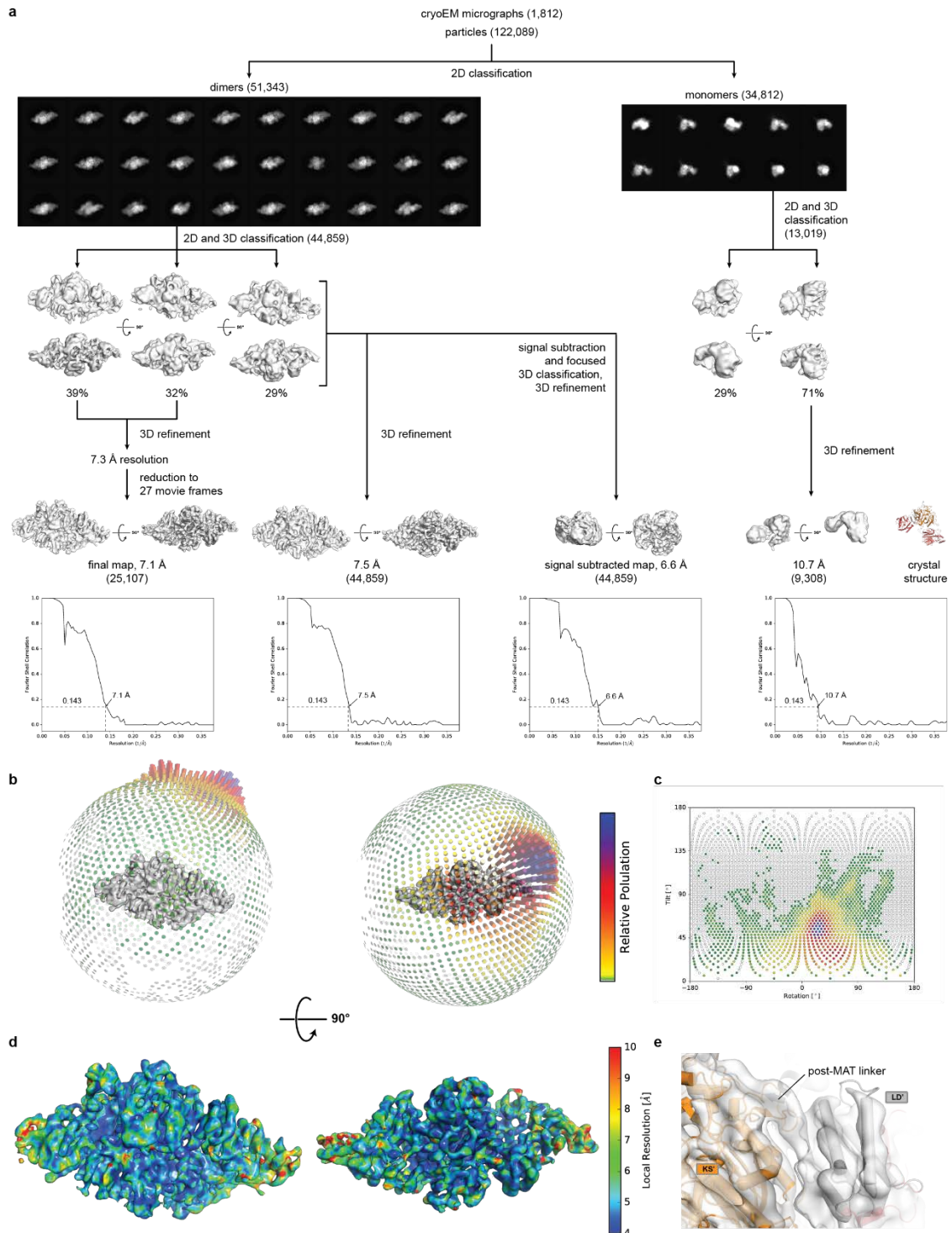
1. Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, N. L.; Townsend, C. A., Deconstruction of iterative multidomain polyketide synthase function. *Science* **2008**, *320* (5873), 243-6.
2. Vagstad, A. L.; Bumpus, S. B.; Belecki, K.; Kelleher, N. L.; Townsend, C. A., Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J Am Chem Soc* **2012**, *134* (15), 6865-77.
3. Watanabe, C. M.; Wilson, D.; Linz, J. E.; Townsend, C. A., Demonstration of the catalytic roles and evidence for the physical association of type I fatty acid synthases and a polyketide synthase in the biosynthesis of aflatoxin B1. *Chem Biol* **1996**, *3* (6), 463-9.
4. Patel, J.; Hoyt, J. C.; Parry, R. J., Investigations of coronatine biosynthesis. Overexpression and assay of CmaT, a thioesterase involved in coronamic acid biosynthesis. *Tetrahedron* **1998**, *54* (52), 15927-36.
5. Piasecki, S. K.; Taylor, C. A.; Detelich, J. F.; Liu, J.; Zheng, J.; Komsoukianants, A.; Siegel, D. R.; Keatinge-Clay, A. T., Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem Biol* **2011**, *18* (10), 1331-40.
6. Prasad, G.; Borketey, L. S.; Lin, T. Y.; Schnarr, N. A., A mechanism-based fluorescence transfer assay for examining ketosynthase selectivity. *Org Biomol Chem* **2012**, *10* (33), 6717-23.
7. Tse, M. L.; Watts, R. E.; Khosla, C., Substrate tolerance of module 6 of the epothilone synthetase. *Biochemistry* **2007**, *46* (11), 3385-93.
8. Brobst, S. W.; Townsend, C. A., The potential role of fatty acid initiation in the biosynthesis of the fungal aromatic polyketide aflatoxin B1. *Canadian Journal of Chemistry* **1994**, *72* (1), 200-7.
9. Tang, M. C.; He, H. Y.; Zhang, F.; Tang, G. L., Baeyer-Villiger oxidation of acyl carrier protein-tethered thioester to acyl carrier protein-linked thiocarbonate catalyzed by a monooxygenase domain in FR901464 biosynthesis. *ACS Catal* **2013**, *3* (3), 444-7.
10. Gay, D. C.; Spear, P. J.; Keatinge-Clay, A. T., A double-hotdog with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem Biol* **2014**, *9* (10), 2374-81.
11. Lee, M. S.; Qin, G.; Nakanishi, K.; Zagorski, M. G., Biosynthetic studies of brevetoxins, potent neurotoxins produced by the Dinoflagellate *Gymnodinium breve*. *J Am Chem Soc* **1989**, *111*, 6234-6241.

Appendix C: Supplementary Material for Chapter 3

C.1. Supplementary figures

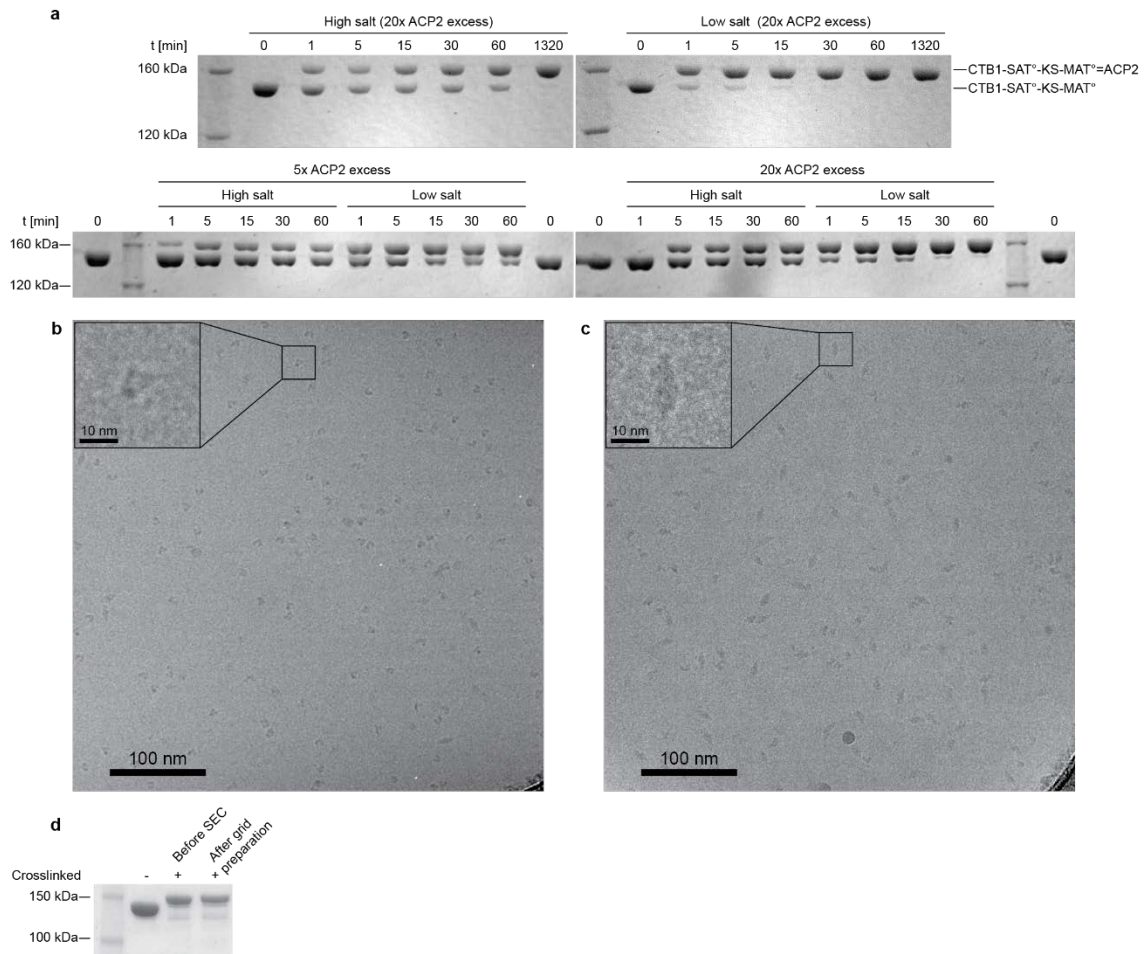


Appendix Figure C.1. Integration of loading domains in PKS. **(a)** Sequence alignment of 34 SAT-KS linker sequences of NR-PKS. All sequences are labelled as “protein name (organism abbreviation) Uniprot number”. **(b)** Superposition of the CazM SAT (black) onto the SAT domain of CTB1 SAT-KS-MAT (light grey) shows a similar location of the SAT C-terminal linker ending (dark red/light red for CazM/CTB1), which points towards the N-terminus of the KS in CTB1. **(c)** Superposition of the loading domain of the AVES1 avermectin modPKS (black) onto the SAT domain of CTB1 SAT-KS-MAT reveals differences in C-terminal linker organization. **(d)** Superposition of the AVES1 loading domain onto CTB1-MAT shows that the AVES1 post-loading domain linker contains an α -helix that matches the linker architecture of MAT in CTB1 (and AT domains in modPKS).

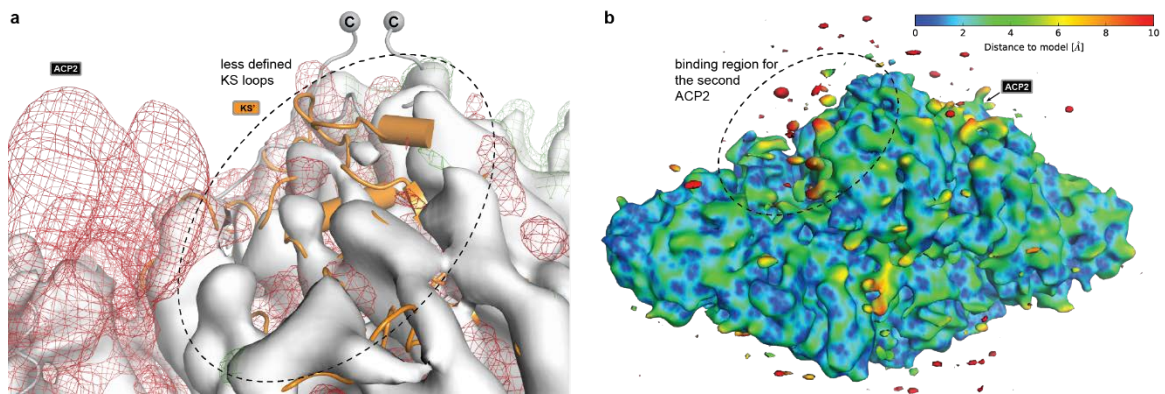


Appendix Figure C.2. Cryo-EM data processing scheme. **(a)** 2D- and 3D-classification and sorting scheme for reconstructions of CTB1-SAT^o-KS-MAT^o=ACP2 dimers as well as monomers. 3D-class distributions are indicated below the models. FSC plots (corrected for effects of the mask) used for determining resolution based on the 0.143 criterion³⁰ is shown below the final models. **(b,c)** 3D (b) and 2D (c) angular distribution plot based on the alignment used for the final reconstruction. **(d)** Local resolution map of the final reconstruction at 7.1 Å resolution shows a resolution below 5 Å in the central KS region, decreasing towards the lateral MAT domains to

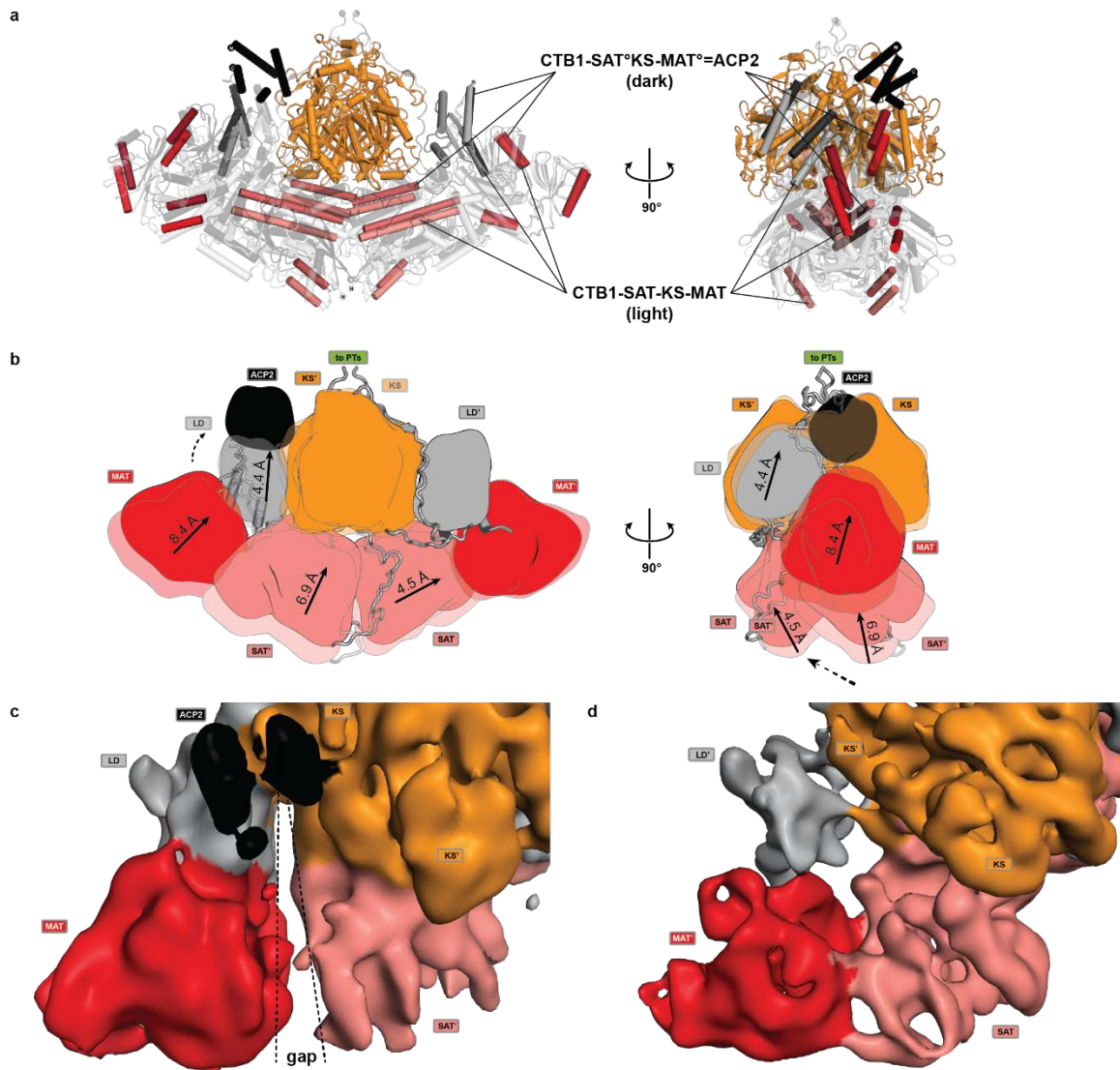
around 10 Å. **(e)** The final map reveals features of secondary structure elements as well as linkers. All maps are contoured at 7.1σ .



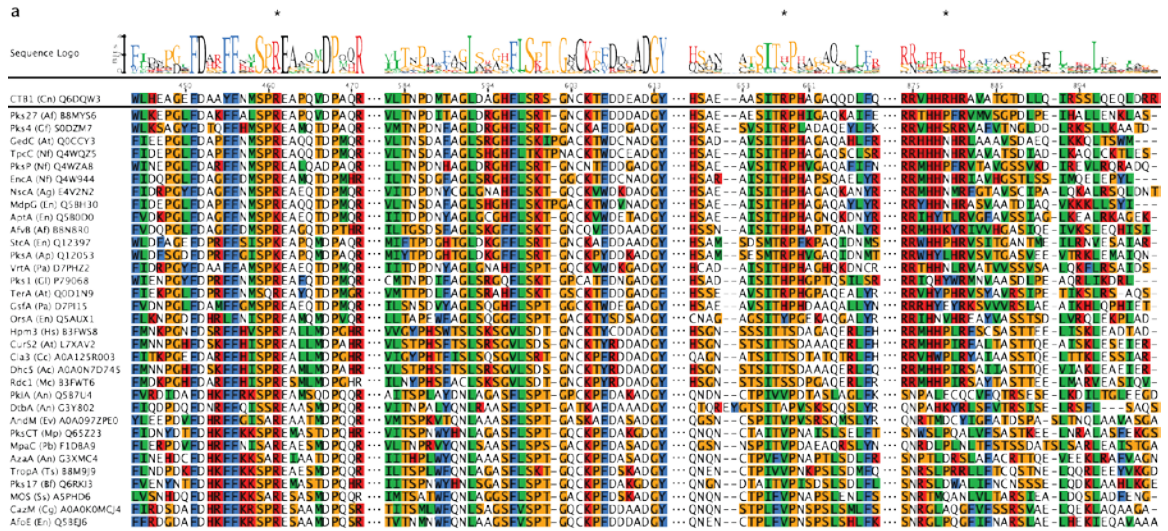
Appendix Figure C.3. Crosslinking and cryo-EM sample preparation. **(a)** Crosslinking time course in high (50 mM Tris pH 7.5, 250 mM NaCl) and low (50 mM Tris pH 7.5) salt buffer as well as for two different CTB1-SAT^o-KS-MAT^o:ACP2 ratios. Crosslinking proceeds faster in low salt buffer and at higher excess of ACP2. **(b,c)** Drift corrected and dose filtered cryo-EM micrographs reveal clearly recognizable shapes for monomeric (b) and dimeric (c) particles. **(d)** Denaturing SDS-PAGE analysis with Coomassie staining of the sample used for cryo-EM grid preparation before SEC and after grid preparation.



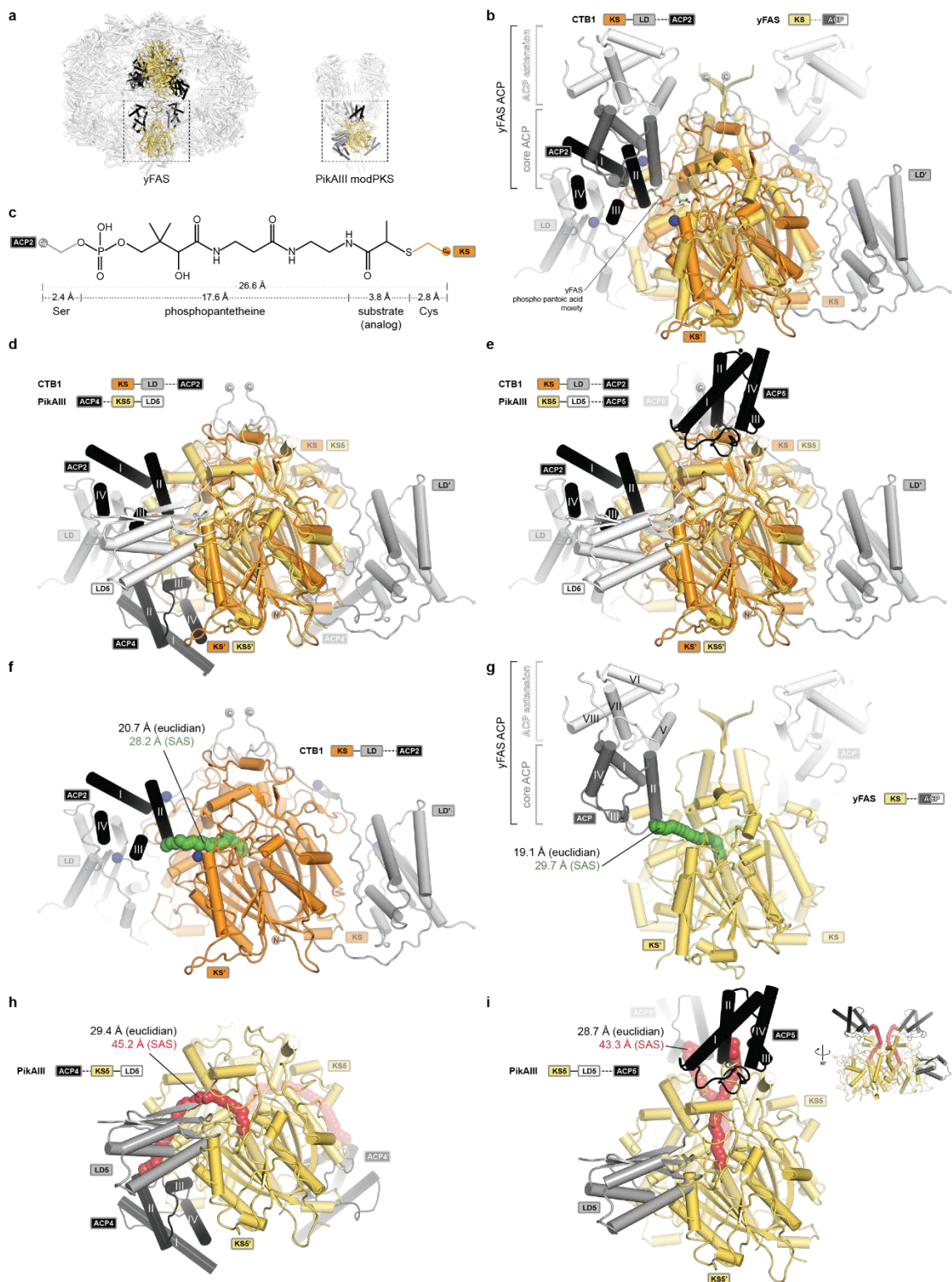
Appendix Figure C.4. Conformational variability in the cryo-EM structure of CTB1-SAT⁰-KS-MAT⁰=ACP2. **(a,b)** KS-based C2 symmetry differences map of the final reconstruction at 7.1 Å resolution, contoured at 7.1 σ (green: positive, red: negative). The potential ACP2 binding region on the KS surface is less ordered on the side lacking ACP2 density. **(b)** An unmasked map of the final reconstruction contoured at 3 σ and colored by distance to the atomic model shows enhanced noise around the second ACP binding site



Appendix Figure C.5. Structural comparison between CTB1-SAT-KS-MAT and CTB1-SAT°KS-MAT°=ACP2. **(a)** Superposition on the central KS dimer. Transparent cartoon representation of CTB1 is shown in white (CTB1 SAT-KS-MAT) and black (CTB1 SAT°KS-MAT°=ACP2). Selected helices are shown in light (CTB1 SAT-KS-MAT) and dark (CTB1 SAT°KS-MAT°=ACP2) domain colors. **(b)** Schematic representation of relative domain motions. Distances are derived by domain-wise C_{α} -r.m.s.d. calculations. **(c, d)** Structural differences at the MAT-SAT interface in the cryo-EM reconstruction are shown for the unsharpened map. While the side with visualized ACP2 (c) reveals a gap, the side with unresolved ACP2 shows a connected interface (d).



Appendix Figure C.6. CTB1 SAT-KS-MAT alignment of regions interfacing ACP2 and phylogeny. All sequences are labelled as “protein name (organism abbreviation) Uniprot number”. **(a)** Sequence alignment of regions relevant for ACP2 interaction. *indicates interface mutants that reduce activity. **(b)** Phylogenetic analysis of NR-PKS indicates three main clades with one clustering around CTB1. Distance unit is given as accepted amino-acid substitutions per site and indicated by scale bar.



Appendix Figure C.7. Comparison of KS-ACP interactions in PKS and FAS. **(a)** ACP interactions with multienzyme KS domains have been observed in the hetero-dodecameric yFAS and PikAIII modPKS. **(b)** Superposition of the architecturally distinct yFAS-KS with CTB1-KS-LD°=ACP2. The divergent ACP of yFAS consists of a C-terminal four-helix bundle extension (white) and the

conserved four-helix bundle at the N-terminus, which interacts with the KS in a similar position as observed for ACP2 in CTB1. ACP-interfacing residues in CTB1 are indicated by blue spheres (see Figure 3.8) **(c)** Derivation of the length of a fully extended crosslink between ACP2 and KS in CTB1. The defined length of the phosphopantetheine cofactor bridges a distance of approx. 27 Å between the ACP-Ser and the KS-Cys C_α atoms for functional ACP interaction with active sites. **(d,e)** Superpositions of PikAIII with ACP4 (d) and ACP5 (e) on CTB1-KS-LD=ACP2 show the relative locations of the ACP binding interfaces. The LD5 arrangement in PikAIII is invers to CTB1. **(f-i)** C_α-C_α solvent accessible surface (SAS) distances between the ACP anchor and the KS active site cysteine in CTB1 (f), yFAS (g), and PikAIII with ACP4 (h) and ACP5 (i) are indicated. Solvent accessible surface paths are indicated by tubes of spheres and colored according to the agreement (green) or disagreement (red) with the length of the Ppant cofactor. The path for ACP5 to KS5 in PikAIII (i) is only accessible in the absence of the post-(M)AT linker, which was not resolved in the cryo-EM structure. An additional miniaturized-sideview is indicated. Corresponding euclidian distances are given in the panels. Domain colors are indicated in the panels; only KS, LD, and ACP domains are shown (b-i).

C.2. Supplementary tables

a	
CTB1-SAT-KS-MAT	
Data collection	
Space group	C222 ₁
Cell dimensions	
a, b, c (Å)	108.05, 230.2, 253.8
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	126.9 – 2.77
R _{merge} (%) [*]	21.2 (193.8)
<i>I</i> / σ <i>I</i> [*]	10.13 (1.03)
CC _{1/2} [*]	99.5 (43.3)
Completeness (%) [*]	99.2 (91.9)
Redundancy [*]	8.5 (5.6)
Unique reflections [*]	79,860 (5,448)
Wilson B factor (Å ²)	60.9
Refinement	
Protomers	2
Resolution (Å)	63.45 – 2.77
R _{work} / R _{free}	0.21 / 0.24
No. atoms	39,201
Protein	38,682
Solvent	519
B-factors	70.17
Protein (Å ²)	70.31
Solvent (Å ²)	60.17
R.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.496
b	
CTB1-SAT ^o -KS-MAT ^o =ACP2	
Data collection and processing	
Microscope	Titan Krios
Voltage (keV)	300
Defocus range(μm)	0.8-4.5
Movies	1,728
Frames per movie	60
Exposure per movie (s)	18
Nominal magnification	105,000x
Pixel size (Å)	1.326
Dose rate (e ⁻ sec ⁻¹ Å ⁻¹)	5.0
Total dose (e ⁻ Å ⁻²) initial / final	90 / 41
Refinements	
Particles initial / final	122,089 / 25,107
Map resolution (Å)	7.1
C-beta deviations	0
Rotamer outliers (%)	0.00
All-atom clashscore	6.09
MolProbity score	1.57
No. atoms	39,058
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.290

Appendix Table C.1. X-ray crystallographic (a) and cryo-EM (b) data collection and refinement statistics. Resolution cutoff in (a) was determined by CC_{1/2} criterion¹. Highest resolution shell is shown in parentheses.

a

Structure 1	Structure 2	C α r.m.s.d. [Å]	Aligned residues	
CTB1 SAT	CTB1 SAT (2 nd chain)	0.34	350	
	Closest* in PDB (3G87) no multienzyme	2.56	262	
	human FAS MAT (3HHD)	2.59	249	
	CazM SAT (4RPM) loading domain	2.63	272	
	PksE AT (5DZ7) [†]	2.64	245	
	procine FAS MAT (2VZ9)	2.65	244	
	PKS13 AT (3TZX)	2.75	249	
	PksC AT (5DZ6) [†]	2.78	251	
	CTB1 MAT	2.81	255	
	DEBS AT ₅ (2HG4)	2.81	254	
	CurL AT (4MZ0)	2.83	251	
	DEBS AT ₃ , (2QO3)	2.86	261	
	DisD AT (3RGI) [†]	2.87	258	
	VinK (5CZD)	2.87	242	
	AVES1 AT (4RL1) loading domain	2.89	256	
	ZmaA (4QBU)	3.01	244	
	DYNE8 AT (4AMP)	3.05	217	
	MAS-like AT (5BP1)	3.18	259	
	OzmQ partial AT (4OQJ)	3.22	92	
	CTB1 KS	CTB1 KS (2 nd chain)	0.23	435
PksL KS (5ENY)		1.40	392	
C0ZGQ5 KS (4Z37)		1.40	398	
CurL KS (4MZ0)		1.41	400	
MgsF KS (4TKT) [†]		1.41	401	
PksJ KS (4NA3) [†]		1.42	404	
human FAS KS (3HHD)		1.44	389	
procine FAS KS (2VZ9)		1.48	388	
DEBS KS ₃ , (2QO3)		1.49	406	
DEBS KS ₅ , (2HG4)		1.49	392	
MgsE KS (4QYR) [†]		1.52	400	
OzmH KS0 (4OPF) [†]		1.52	339	
MAS-like KS (5BP1)		1.55	347	
OzmN KS (4WKY) [†]		1.60	374	
RhiE KS (4KC5) [†]		1.61	399	
Closest* - FabF KS (4JB6) no multienzyme		1.64	378	
OzmQ (4OQJ) [†]		1.75	375	
CTB1 MAT		CTB1 MAT (2 nd chain)	0.23	304
		CurL AT (4MZ0)	1.72	281
		MAS-like AT (5BP1)	1.82	285
	PKS13 AT (3TZY)	1.83	290	
	human FAS MAT (3HHD)	1.83	280	
	procine FAS MAT (2VZ9)	1.85	279	
	DYNE8 AT (4AMP)	1.86	231	
	DEBS AT ₃ , (2QO3)	1.87	285	
	DEBS AT ₅ , (2HG4)	1.89	290	
	Closest* in PDB: AVES1 AT (4RL1) loading domain	1.92	284	
	ZmaA (4QBU)	2.00	275	
	DisD AT (3RGI) [†]	2.02	269	
	PksC AT (5DZ6) [†]	2.06	266	
	PksE AT (5DZ7) [†]	2.09	264	
	VinK (5CZD)	2.26	261	
	OzmQ partial AT (4OQJ)	2.50	89	
	CazM SAT (4RPM) loading domain	2.68	252	

b

Interface 1	Interface 2	Min area [Å ²]	Max area [Å ²]	Hydrogen bonds	Salt bridges
SAT	KS-LD-MAT-post MAT linker	1197.2	1230.1	6-8	3
SAT [§]	KS [§]	183.8	184.9	1-2	0
SAT [§]	MAT [§]	946.1	966.8	5-6	2-3
SAT	SAT	323.7	323.7	0	0
SAT dimer	KS dimer	659.9	659.9	3	0
SAT dimer	KS-LD-MAT-post MAT linker dimer	2718.5	2718.5	14	7
SAT [§]	SAT-KS linker [§]	486.7	500.1	1-2	0
KS	KS	2607.1	2607.1	10	1
KS	LD	415.9	416.5	4	2
KS	post-MAT linker	1460.0	1465.8	12-13	3-4
LD	post-MAT linker	1262.5	1277.4	8	0-1
MAT	post-MAT linker	908.5	911.5	4	1

Appendix Table C.2. Structural comparison and interface analysis of CTB1 SAT-KS-MAT. (a) C α r.m.s. deviations obtained for structural comparison of crystallized CTB1 domains with their closest structural neighbors in the PDB and of multienzyme PKSs and FASs (to the best of our knowledge). For structures with several protomers, only the best matches are reported. *Q-score

based, [†]*trans*-AT PKS. (b) Interfaces in the crystal structures of CTB1 SAT-KS-MAT. [§]Interface with the second protomer.

Plasmid	Protein	Vector	Tag	MW (g/mol)	ϵ (M ⁻¹ cm ⁻¹)
pECTB1-NKA6 ²	SAT-KS-MAT	pET-24a	C-His ₆	140150	138200
pECTB1-SKM-C119A-S1010A	SAT ^o -KS-MAT ^o	pET-24a	C-His ₆	140150	138200
p28CTB1-ACP2	ACP2	pET-28a	N-His ₆	11970	5500
pECTB1-SKM-R461A	SAT-KS-MAT-R461A	pET-24a	C-His ₆	140150	138200
pECTB1-SKM-R461E	SAT-KS-MAT-R461E	pET-24a	C-His ₆	140150	138200
pECTB1-SKM-R658A	SAT-KS-MAT-R658A	pET-24a	C-His ₆	140150	138200
pECTB1-SKM-R658E	SAT-KS-MAT-R658E	pET-24a	C-His ₆	140150	138200
pECTB1-SKM-R879A	SAT-KS-MAT-R879A	pET-24a	C-His ₆	140150	138200
pECTB1-SKM-R879E	SAT-KS-MAT-R879E	pET-24a	C-His ₆	140150	138200
pECTB1-PT ²	PT	pET-24a	C-His ₆	41240	26150
pECTB1-TE ²	TE	pET-28a	N-His ₆	33630	38055
pECTB1-ACP2-E1794A	ACP2-E1794A	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-E1794R	ACP2-E1794R	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-E1795A	ACP2-E1795A	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-E1795R	ACP2-E1795R	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-E1802A	ACP2-E1802A	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-E1802R	ACP2-E1802R	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-D1806A	ACP2-D1806A	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-D1806R	ACP2-D1806R	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-D1811A	ACP2-D1811A	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-D1811R	ACP2-D1811R	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-D1815A	ACP2-D1815A	pET-28a	N-His ₆	11970	5500
pECTB1-ACP2-D1815R	ACP2-D1815R	pET-28a	N-His ₆	11970	5500
pECTB1-ACP1	ACP1	pET-28a	N-His ₆	11160	N/A

Appendix Table C.3. Plasmids used in this study

Primer	Sequence 5'-3'
CTB1-ACP2-5	ATTACATATGGATCCATCCCCTAACGAGAT
CTB1-ACP2-stop-3	TAATGCGGCCCGCTATTCCGTTGACCCAGAGAACC
T7	TAATACGACTCACTATAGGG
T7-term	GCTAGTTATTGCTCAGCGG
CTB1-C119A-5	TGCATTACCGGCGTTGCAACCGGCGCA
CTB1-C119A-3	CGTCAATGCGCCGTTGCAACGCCGGT
CTB1-S1010A-5	GGTAGTTGGCCACGCGTTGGGCGAGTATGC
CTB1-S1010A-3	GCATACTCGCCCAACGCGTGGCCAACTACC
CTB1-R461A-5	ACATGAGTCCGGCGGAAGCGCCGC
CTB1-R461A-3	GCGGCGCTTCCGCCGACTCATGT
CTB1-R461E-5	ACATGAGTCCGGAAGAAGCGCCGC
CTB1-R461E-3	GCGGCGCTTCTTCCGGACTCATGT
CTB1-R658A-5	GCCTCTATCACTGCGCCTCATGCCGGAG
CTB1-R658A-3	CTCCGGCATGAGGCGCAGTGATAGAGGC
CTB1-R658E-5	GCCTCTATCACTGAACCTCATGCCGGAG
CTB1-R658E-3	CTCCGGCATGAGGTTCAGTGATAGAGGC
CTB1-R879A-5	GAGTGCACCATGCGCACAGAGCCGTAGC
CTB1-R879A-3	GCTACGGCTCTGTGCGCATGGTGCACTC
CTB1-R879E-5	GAGTGCACCATGAACACAGAGCCGTAGC
CTB1-R879E-3	GCTACGGCTCTGTGTTTCATGGTGCACTC
CTB1-ACP2-E1794A-5	CCTCAAGATCCTGTCTGCGGAGAGTGGCCTCAC
CTB1-ACP2-E1794A-3	GTGAGGCCACTCTCCGCAGACAGGATCTTGAGG
CTB1-ACP2-E1794R-5	CCCTCAAGATCCTGTCTCGTGAGAGTGGCCTCAC
CTB1-ACP2-E1794R-3	GTGAGGCCACTCTCACGAGACAGGATCTTGAGGG
CTB1-ACP2-E1795A-5	GATCCTGTCTGAAGCGAGTGGCCTCAC
CTB1-ACP2-E1795A-3	GTGAGGCCACTCGCTTACAGACAGGATC
CTB1-ACP2-E1795R-5	CCCTCAAGATCCTGTCTGAACGTAAGTGGCCTCACTGA
CTB1-ACP2-E1795R-3	TCAGTGAGGCCACTACGTTACAGACAGGATCTTGAGGG
CTB1-ACP2-E1802A-5	CCTCACTGATGAGGCGTTGACTGATGACAC
CTB1-ACP2-E1802A-3	GTGTCATCAGTCAACGCCTCATCAGTGAGG
CTB1-ACP2-E1802R-5	GTGGCCTCACTGATGAGCGTTTACTGATGACACAAG
CTB1-ACP2-E1802R-3	CTTGTGTCATCAGTCAAACGCTCATCAGTGAGGCCAC
CTB1-ACP2-D1806A-5	GAGGAGTTGACTGATGCGACAAGTTTCGCCG
CTB1-ACP2-D1806A-3	CGGCGAAACTTGTGCGCATCAGTCAACTCCTC
CTB1-ACP2-D1806R-5	TGAGGAGTTGACTGATCGTACAAGTTTCGCCGACG
CTB1-ACP2-D1806R-3	CGTCGGCGAACTTGTACGATCAGTCAACTCCTCA
CTB1-ACP2-D1811A-5	CACAAGTTTCGCCGCGGTGGGCGTTG
CTB1-ACP2-D1811A-3	CAACGCCACCGCGGCGAAACTTGTG
CTB1-ACP2-D1811R-5	GACACAAGTTTCGCCCGTGTGGGCGTTGATAGC
CTB1-ACP2-D1811R-3	GCTATCAACGCCACACGGGCGAAACTTGTGTC
CTB1-ACP2-D1815A-5	CGTGGGCGTTGCGAGCCTCATGAGTCT
CTB1-ACP2-D1815A-3	AGACTCATGAGGCTCGCAACGCCACG
CTB1-ACP2-D1815R-5	CGACGTGGGCGTTCGTAGCCTCATGAGTCT
CTB1-ACP2-D1815R-3	AGACTCATGAGGCTACGAACGCCACGTCG
CTB1-ACP1-5-A	ATTACATATGGCAACCCAAGTACTCCGCAA
CTB1-ACP1-3-C	ATTAGCGGCCGCGATCTCGTTAGGGGATGGATCAGT

Appendix Table C.4. Primers used in this study

1. Karplus, P. A.; Diederichs, K., Linking crystallographic model and data quality. *Science* **2012**, 336 (6084), 1030-3.
2. Newman, A. G.; Vagstad, A. L.; Belecki, K.; Scheerer, J. R.; Townsend, C. A., Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem Commun (Camb)* **2012**, 48 (96), 11772-4.

Appendix D: Supplementary Material to Chapter 4

D.1. CTB gene sequences

mRNA sequences from *C. beticola* are shown.

D.1.1. CBET_00833 (CTB1)

ATGGAAGATGGAGCGCAGATGCGAGTGGTGGCATTTCGGGGATCAGACATATGACTGCAGTGAGGCAG
TCTCACAGCTCTTAAGGGTCCGCGACGATGCCATAGTCGTCGACTTCCTCGAACGGGCAACAGCCGTT
TTGAAGGCAGAGTTGGCCAGGCTAAGCAGCGAGCAACAAGAAGAGACGCCACGATTTCGCCACGCTCG
CTGAACCTTGTGCCACGCTACCGAGCTGGCACTTTAAATCCAGCAGTGTGCGCAGGCATTGACATGCATC
ACCCAGCTGGGGCTTTTCATCCGACAACACAGCAGCGGACAGGAAGCCTACCCACAGCAAACGATAG
TTGCATTACTGGCGTCTGCACCGGAGCGCTGACCGCCGTCGCTGTCGGCTCTGCTTCGAGTGTGACC
GCTTTGGTGGCGCTCGCATTGCACACTGTTGCCGTCGCCGTGCGATTGGGTGCTCGAGCCTGGGAGA
TTGGCAGCTGCCTTGCCGATGCCAGGAGAGGTGCGAATGGCCGGTATGCTTCTTGACATCGGCGGT
GGGCGGTATCAGTCCCAAGACCTTCAGGATCGAATTCAGCATAACATGACCGAGCAGGCTCTTGAT
CCGTGTCGGTCCATATCTCTCGCAGCGGTCGGCCCGGCCAATCAAGCGTGTGCGGCTGCTCCCGT
CATCTCGACGCGTTCCTCAGCACACTTCTCGACCGCTGACAACAACCGGACTTCCATTACCGCTC
CGTATCACGCACCTCACCTCTTCACTGCGAAAGATGTGCGAGCATGTCACGGACTGTCTGCCTCCAGC
GACGCATGGCCGACAGTGCAGCATCCCCATCATCAGCTTCTCCCGCATGAGGCGGTGTCGCGTGGTG
CAAGCTTCCCGGCCGCCATGAGTGAGGCGGTGCGAGACTGCTTGATCCGCCCCATTGCCCTCGACCG
CATGGCAGTGAGCATTGCCAATCATGCTCGAGATCTCGGCAAAGACTCTGTGCTTCCGTCACCCATCG
CTCTGTGCTTCAGCGACAAGCTGGGTCCACAAGTGAACAGCCATCTGCCAGGCACGAAAGCACCGAC
GCCTGAGCTGACGTCCACATCGTCTATCCCATCTGCTATAGGAGCGGAACAGCAGCCGATGGCAAAT
CCCCATTGCTATCCTTGCTGCCTCTGGTCGTTTCCACAATCTTCATCCATGGACCAGTTCTGGGATG
TACTTATTAACGGCGTCGACACGCACGAACCTTGACCTCCAACCCGTTGGAATGCAGCTACTCATGTCT
CCGAAGATCCTAAAGCAAAGAATGTGTCTGGCACAGGTTTCGGTTGCTGGCTCCACGAAGCCGGTGAA
TTTGATGCCGCTATTTTAACATGAGTCCGCGTGAAGCGCCGAGGTTGATCCAGCACAGCGTCTTGC
GCTCTTACTGCCACCGAAGTCTCGAGCAGGCCGGCATTGTCCCGAACCGAACGTCGTAACCTCAGA
AAAACCGAGTTGGTGTATGGTACGGCGCAACCAGCAATGATTGGATGGAGACCAACAGCGCTCAAAAC
GTTGATACGTAATTCATCCCGGTGGCAATCGCGCATTCATCCAGGAAGAGTGAACACTTCCACAAG
TTTAGTGGCCCATCTTACACGATCGATACAGCCTGCAGCTCCAGTTTGGCAGCATTGCACATGGCGTGC
AACGCACTTTGGCGAGGCGAAGTCGATACAGCCATCGTAGGTGGGACCAACGTCCTCACGAATCCAGA
CATGACGGCGGGACTCGACGCGGGTCACTTCTTGTCAAGGTCTGGAAATTGTAAGACGTTTCGACGACG
AAGCCGATGGCTATTGCAGAGGTGAGGCAGTGGTACCCTCATTCTCAAACGGCTGCCAGACGCGCA
AGCGGACAAAGATCCAATTCAGGCTTCAATTCTGGGAATTGCCACTAATCACTCAGCCGAGGCCGCT
CTACTACTAGGCTCATGCCGGAGCGCAGCAAGACTTGTCCAACAAGTTCTCACGGAGACAGGTCTT
ACCGCAACGACATTAGTGTGTGCGAGATGCATGGTACTGGCACCCAGGCTGGAGACAGTGGTGAAA
CAACGTCGTCGTTGGAGACCCTAGCGCCTTTGAACCGATCCGGCTCTGCTGTGCGAACAACACCTCTC
TACATTGGCGCAGTCAAGTCCAATGTGGGTGATGCTGAGTCCGCAGCTGGGGTCAGCAGTCTGGCCAA
GATCTTGCTTATGCTCAAGCATTCCAAGATCCCTCCTCATGTGGGCATCAAACGAAGCTGAATCACCG
GCTACCAGACCTAGCTGCACGAAATACACACATAGCGCGGACTGAGGTACCTTGGCCTCGGCCAAAGA
ATGGCAAGCGTCGTTCTGCTCAATAACTTCTCGGCCGCTGGAGGTAACACGTGCCTTGTGCTTGG
GATGCGCCCGAGCCCGAGGACTCTCAAGAAGTCGACCCTAGAGAACATCACATCGTTGCACTCTCTGC
TAAAACACCTGATTCAATGGTGAACAACCTCACGAACATGATAACCTGGATCGACAAGCACTCTGGAGA
CAGCATCGCCACCTTGCCTCAACTGTCTTACAGCAACTGCACGAAGAGTGCACCATAGACACAGAG
CCGTAGCTACCCGCACTGATCTGCTGCAAATCCGTTTCGTCGCTTCAAGAACAGCTTGGCCCGGGTG
TCCGGCGAGAGAAGTATCCCTCACCCGCCAATGGACCTAGCTTTTGTCTTGGCTTTCCTACTGGCCAAGG
CTCGGCGTTCGAAGGTATGGGTGTAGATCTCTACAAACGTTTCGCTCATTTCGGTTCAGACATTGCCCG
CTATGATCAGATCTGCGAGGGTATGAGCCTGCCCTCGATCAAAGCTATGTTTCGAGGACGAGAAAGTGT
TCTCCACAGCTTACCAACTTTGCAGCAGCTCACGCATGTCTGTTTTTCAGATGGCCCTGTACAGACTAT
GGAAGTCCCTCGGCGTACAAGCGAAGGCGGTAGTTGGCCACAGTTTGGGCGAGTATGCTGCACTCTA
CGCCGCTGGAGTGTATCGCAATCCGATACGCTCTACCTGGTGGGGCGGCGTGCACAGCTGATGGAG

AAACATCTTTGCAAGGCACACATGCAATGCTGGCCGTCCGTGCGAAAGAAGAAGCCATTGTGCGCAGC
GATTGACGGGCTCCAGGAGAGGCATACGAGTTCTTTGCCGCAATGGCGAACAGCGAAACGTTCTCG
GAGGCACCGTTGCTCAAATCCAGGCGGCGAAAGCCGCGCTTGAGGCTAAGAAGATTCCGATGCCAGTA
CTTGACACCCCGATGGCATTCCACACTGGCCAAGTTGATCCGATTCTGCCCGAGCTCTTGCAAGTCCG
CTGCAGCATGCTCCATCCAGGATCCCCAAATTCCTGTCACTCCCCAGCATATGGCAAAGTGATCAGGT
CTGCCAAGGACTTCCAACAGAGTACTTACCCATCACTGCCGCACTTCTGTCAACATGGTCGATGCTC
TCCAAAGTGCAGTCGAAGAAGGCTTACTTGACAAGAACATTATTGGCCTCGAGATTGGCCCTGGCCCG
GTCGTACGCAGTTCGTCAAGGAAGCTGTGGGCACAACCATGCAGACCTTTGCGTCCATCAATAAAGA
CAAAGACACATGGCAGCTCATAACGCAAGCGCTAGCTAAGTTCTATCTTGCAAGTCCAGTATCGAGT
GGTCGCGCTATCACGAAGACTTCCCTGGAGCTCAGAAAGTCCCTTGAGCTCCCGGCTTATGGCTGGACC
TTGAAAACTACTGGCTGCAATATGTCAACGACTGGTCTCTGAGGAAGGGCGATCCAGCCGTGTTGT
TGCCGCGTCAAATTTGAACTCTCTTCTGTCGATACACAAGGTCATAACAAACACAATCACCGCCAACAG
CGACGGCGAGCTTGTGCTGGACGCAGACCTCAGTCGAGAGGACCTGCATCCCATGGTTCAGGGGCAT
CAGGTCTACGGTGTCCACTGTGCACACCTTCCGTGTACGCGGACATCGCTCTGACACTCGGCGAGTA
TATTCGACAGGTCAAGCCAGGCGAGGTTGCACAGACATCCGTTGAAGTAGCAGAGATGAACATTC
AAAGCGCACTGGTGGCTAACAACACGGGCGAGAGTGCAACTCCTTCGCACGTATGCCAAGTTTGACCC
AAGGCCAGGTAGCGTCATGCACGTTCTCTAGTATCAAGGAGGACGGCAGTAGCGTAGTCGAGCAGC
ATGCTAATTGCAAGATCCGTTCCGGCAGTCTCGAGAAGGAGAAGACCGCGCTCGAGAGTGCTGCACTA
GCTGCCAAGCCAGAATGGCCGCTCTGAAAACACAAGTTGGCCAGGATGACAACACATACCGCTTCAG
CAAAGGCATGATTTACAAGATGATCGGCCAATTAGCTGACTTCGACGAGAAGTACCGCGGGCTCTGCG
CGATCACACTCGACAACGACGCCATGGAAGCCTCGGGCAAAGTATCATTCAAGGGCATTCCAAACGAG
GGCAAATCCACTAGCCCCGGCTTATCTCGACGCGTTGTGCAACTTGGCGGATTCGTCAAGCAGC
GAACGAGGGTGTGGATCTTGAGAAAGAAGTCTTTGCAATCACGGCTGGGGTCCATGCGCTTCTTCG
CCGCTCTGGATCCAGCAATGACTTACTACACTCATGTGAAGATGACCCAAGGCAAAGACAAATTTGGA
CTGGCGATGCTTGTCTTCGACGACAAGCAAGCATTGATCGGCATTGTTGGGGAGTGGCGTTGCAG
GGCGTGCCCAAGCGACTTATGCATTACATTGTTACAGCTGCCAACAAGAAAGCTTCCGGCCCCGCGAC
AGAGAAGAAGGGCTCTAGCCCCGAGTCGAAAAGAAAGCCAGCGCGCCAGTCGCGCCCACGAGGCC
AGCGATCCAGCGTAAGAATGCTTCGATTCCCTCCACCTGCAACCCAAGTACTCCGCAAACAAGACCAT
CAAGACGCCAAGTGTGTCGGCACTGATAGCCCCGGCCCTCGAGATTGTTAGCGAGGAGATCGGGATG
CCAATCGACGAGCTCAAGGATGATATCGACTTACCGATGCTGGTCTTGACAGTCTGCTCTCCTTGTA
ATTAGCAGTCGCATGCGGGACAGCTGGGCATCGAATTCGAGTCCGCGCAGTTCATGGAGATTGGATC
TATCGGTGGACTCAAGGAGTCTTGACCAGGCTCAGTCCCCCAGTAGCAGTCGCGGTTGCCACTGCCG
TGAAATTTGCAAGGAGGAAGCGCTCACTTCAATTGGAAGAGCTTACTGATCCATCCCCAAGCAGATCG
GCACTGTCTGGCGCGATGCCCTCAAGATCCTGTCTGAAGAGAGTGGGCTCACTGATGAGGAATTGACT
GATGACACAAGTTTCGCCGACGTGGGCGTTGATAGCCTCATGAGTCTTGTGATCACCAGTCGCCTACG
GGATGAATTGGACATCGACTTCCCCGACCGAGCATTATTCGAAGAATGCCAGACTATATTTGACCTTCG
CAAGAGGTTCTCTGGGTCAACGGAAAGCTTCGACTCGACGACGACCAAGCCAGCGCTGGTGATCGG
ACGCCACTTGCACCGGATCCAGCGCTCATCTCCGCCCTCCTCCGAGTTCGATGGCGAGACGCCGAT
GACTGATCTGACGAGGTGTTGATTCTCCCCAGCGCAGAAGAGGATACCATCCCCGCCAAAGGAC
GAATCCCCGCTGCATGGTCGATGATTTGCAAGGCTCACAGAAGCGGTGCAAGGAGATTCTTTTCTTGT
TTCCAGACGGCGCTGGCGCCGCAACTTCACTTGTCTTTACCTCGTTTGGGTGAAGACATTGGCGTAG
TCGCCTTCAATTCGCCTTTCATGAAGACACCGCACAAGTTTGTGATCATACTTACCGGACGTATCG
CGTCTATGTAGAAGGCATTCCGAGGCCGTAAGCGCAAGGCCCGTATCATCTGGGCGGCTGGTCTGCT
GGTGGTATTCTGGCCTATGCCGTTGCCAAGAACTCATCGCAGCTGGCGAGGAGGTTTCGACACTCCT
CCTCATCGACTCGCCTTCGCCAACCAGGCTTAGATCGCCTTCCAACACGATTCTTCGATCACTGCAC
GAACGTCGGACTCTTTGGAACGGAGCTCTCCAGAGGCAGTGGGGTCCCAACAAGACACCCGAATGG
CTGATGCCCTCACTTCAGAGCTAGTATTGAGCTGCTACATGACTACCAGCTCCTCCTATGAAGCTTGGC
AACAAAACGAAAGTCATGGTGATATGGGCAGGTGAATGTGCCTTTGATGGCGTTTCGCTATGCTCACATA
CCGCCCTCTGCAGGCGACACCGACGAAGACACCGAGGGTATGAAGTTCTTGACGGAGAAGCGGAAAG
ATTTTGGAGCCACAGAATGGGCAAGTTTGTCCCTGGCACTGATGTTGATGCGAGGGTGTGAGAGC
GAGCACCATTTCAGCATGATGCGTGATTCTGGTGCACAGATGCTTGTGAGCATATGCGAGACGGATT
GGGATTGTCTCGTCTGA

D.1.2. CBET_00832 (CTB2)

ATGGCTAACCGAATTGAAGCGGACAATCTCTTTGAGCTCACGGCAGAGCTGGTCTCAGCCTCCGCCAA
ATTGCACAAGTTTCTCGACCAGAAGAACCTCCCGCAGCCATCTTTTGTGCTCCAGCTCCATCGGTAGC
TCTCAACACCGCCAACAAGCCTTACTACGATGCGAGAAGCGCGATTGTAGAGGCTGCTGAGCAACTCA

TCCGCCTGGTCCGTGGACCTCGCGATACTCTTCTCGCTCTGTCTTCGAGCACTGTGCTACAGCATCG
ATGCAGGTCGTTTTCAAATACAAGTTTGCGAATCACATTCCGCTACATGGCTCGACAACATTATAGCAAGA
TTGCCGAAGCAGTCGGAGATGGTGTGACAACAGCTCTCGTTGAGCGCACGATACAACATTGCGCTTCC
TTTGGCCTATTCGAGACGATCCCTGGCGGCTATGTTACTCACAAATGCTACCTCGTCACTACTGGTCACC
GATCCAGATCTCGAAGCCTGGATGTATCTCTCAGCGGTGATCGCTTACCCAGCTGGCGCAGCTATCCC
CAAGGCCGTGGAGCAGTATGGCGTGAGCAGTGAGGCCACCGAAGCAGGGTACGGCGTGAGCATAGG
AAGAAAGATTGCACAATTCAGCGATTCCGCGAGCCCCGATGGGAAGAAGGACCACGAGATGTTTCGCAC
GAGCCATGCGTGGTATCGCAGCTGGTGGTGCATACGACTCCGCCACGCAGTGCAGCGGGATACCC
TTGGCACCTTCTCACAGAAGGGCGCAGGACACCTGGTAGTGGATGTCGGCGGAGGTCCGGGCCACGTC
GCCATGGCACTCGCGGAAAAGTACCCAAGCTTGCATTCCAAGTCCAAGATCTACCCGAGACTGTCCA
AGTGGGAGCGAAGAAGTGCCTGAACACCTTCGTAACACGTAACCTTCGTGGCACATGATTTTATGAC
CCCTCAACCTGCTCATGAAGTGCAAGACGGCGAAGGAATCGTATACTTTGCTCGATTTATCTGCACGA
CTGGAGTGACAAGTATGCCACGAAGATCGTGCAGGCGCTTGGCACTGGCTTGAGGCCACAAGACCGC
ATCATCTTGAACGAAGTGGTCGTTCCCTGAAGCCGGGCAAGTTGGCAGAGAGACGGAACGAAGAATGCA
CGATCGTGATCTGCTGATGTTGATGAACCTCAATGGACGTGAGCGGACACAGAGTGCATTTCGAGGCGA
TCTTCGCTTCAGTGACTCCCAAGCTGCGGCTGCAGAGGGTCAATCACCCAGAGCAAGGCCAATTGTGCG
CTCATCGAGGTGACTCTTGATGGCGTTGAGCTTCTGCCAGGCGAATGGTGTCAATGGCCATGCGAA
TGGGACTAATGGCGTGAATGGTCATTA

D.1.3. CBET_00834 (CTB3)

ATGATGCAGTTCACACGCGATCTTGAGGCGTCTTGAGGCGGTATCGGCCAACGCCAGGAGCTGC
TCAAATCTCTCAAGAGTCGCAAGGATGTTCAAGACCTCAACGCGTCGTTGCCGAAGGATCCTTTAGACA
ACTGCGATGCTCAAACCTCAAGCCGCTCGTGCAGCTGGCAGAGGCAGCGACAAGAATCTTGCAGTT
GTGATCCGACCTCAAGAGTATCTGGAGCATCTACAAAACGGCTATCAACATTTAACCTGTTTTCGCTG
GCTGGTGGAACTCAACATATTGGACCACCTTCCACATAGCGGAACGATCAGCTACACAGATCTTGCGA
GAAAAGCCAGCGTGCCGCTATGCAATTGAGAAGCATCTGTGCATGGCCATATGCAATGGATTCTG
GAAGAGCCCCGAGGCCAACCAAGTCCGCCACAGTCGATTTCCGCCTTGTTTCGCTCGCGATGAAAGCTA
TTTAGGTTGGCTAGATGGATGGTCAACTACTCTGTGCCAGCTGCATACAAGCTTAGCGACGCCACGC
GATCGTGGGGCAGACTGTCGCCAAAGATCAGACCGCTTCAATCTGGGAATGGATGTGAAAGTCCCA
TTCTTTGACCATCTCCGCCAGACGCCCGCAATGAAGGACGCCTTTGCAGCTTATATGCGTAATGTGACT
TCGAACGCAACTTGGGGCCTTCAGCACGCAGTACCCGGCTTCGACTGGGCTTCCCTTCCGCGGGGCG
CAAAAGTCGTGGATGTCCGGTGGCTCTCTTGGGCATGGTAGCATTGCCATTGCCAAGGAGCACACTCAC
CTTACCTTCGTCAATTCAAGATCTGCCAGAGACGGTTCGCTGGTGGCAGGAAAGAAATGGCCAAAATGA
CAAGATTGAAGCTTCTGTTAAATCTCGCATCACCTTTCAGGAACACGACTTCTTTGGTCTCAAACAGTG
AAGGATGCCGATGTTTACTTTCTTCGCATGATCTGTACGACTGGCCGACAACGAAGCCAAGGTCATC
CTCTCTCAGATTGCGCAGCACTGAAACCTGGGGCGCAAAATAGTCATGACACCAATTTCCCCCA
GCCCGGCACAACCTAGCGTTTTGCAAGAGCAAACTACGCATTCCGGATCAACAAATGATGGAAGCTT
CAATGCCAAGGAGCGTGAATTGGAGACTGGAGCTCATTGATGCAATCTGCCGGTCTCGAGATTTCTC
GCGTGAACCAGCCGCTCAACAGTGTGATGGGTCTGCTCACAGTCCGCTCAGCCGGCCAGACTGCCCT
CTCCGGAACGAATACACTCACGCCAGAGTTGGTGACGGCAGTCTCCGCAAGCACTGGCTCTGCTGATT
CGAGACCAGTCTCATTGCAGGCGCGGGTATTGCTGGGCTCTGTCTTGCACAGGCTTTGAAGAAGGCC
GGAATTGACTTTTCGCTCTTCGAAAGGGACTCCCATATCGATGCTCGGCCACAAGGATACCGACTCAA
ATTCGAAGCAGACGCCGCACAGTCTCTCAAGAACATCTGCCTGACGATGTTTATGAGGCTTTCGAAT
CTCAAATGCCGTCACCGCCGTAGGCGAGACAGACTTCAATCCCTTCAATGGCAACATCATCCACAGCC
GCACTGGTGGCGGCTGTCTGGCAAGAAGGGACTGTATGCGACATTCACTGTTGACCGCAAAGCATT
AGAACTCAGCTCATGACTGGCATTGAGGACAAGATCTCGTTCCGGAAAGGAAATCGCGTACTACAAGAC
TGATGACTCTGCATCTACGGTTACCGCTGAATTCAGGACGGCACTCACGTCACCGGAAGTTTCTGG
CCGGCACTGATGGCTTACACTCTGTCTGTTCCGAAGACATGTGTACCAGACCATCGTATTGTGGATACTG
GTGCTGCCTGCATCTACGGCAAGACTGTAATGACACCGGAATTCCTCGCGCGGTTCCCGGAGAAAGGC
TTGAGGTTCACTACTGTGGTACGCGACATCGCACCTATGCTACAATCTTGTCTCATCGGCGACAGCCCA
GTCACCTTACTACTGGAACCCATCCGATTCCAGCAAGCCCTCGCGTCCCGCTACCCAGAAGTGCCTG
AGACTACGTTACTGCGCACTCATCGGACCCAAAGGAACGCTTCGGATCGCAAGAGGTGACTTCCATGA
AGAACTTCGTCTCACTGGACCAAGCGGCAGAACAGGCTGCCAAGCTCAGTCTCGCAGTACCGGAGGAA
TGGCATCCAAGCCTCCGCGCGCTGTTGAGCTCCAAGACACAAAGCAAGCATCGCTCATTTCGCTTGC
ATCCACAATCCCGATATTCCCTCATGGGAGTCCCACTCCAATGTTACCGTTCTTGGCGATAGCATTCA
TCCAATGAGCCCTTGTGGTGGAGTCGGAGCGAACACCGCAATAGTCGATGCCGACGCCTTGGCTAAG
TGCTCGTTGAGCATGGCACGAAGCCACCGGTGAATGCAATCGCCGAATTTGAGGCCGCGATGAGAACA

AGAGCGAAGAGGAACATTTGGAGGAGTGAAGTTGGTAGTAAGAGAATGTTTGGGCAGAAGAATTTGGT
GGATTGTTTCAGAGTTCGTGTTTTGA

D.1.4. CBET_00831 (CTB4)

ATGGCTCTCCCTATCACTGACGATGACCTCGACGGACTGAAGCAGCCCTATGTGACTTTCTCGTCTGGT
TCAGCATCACCAACCGCAATCGACGACTGACGCTATGGATCTCGAAGAGCAGGTACTCGACGCAATCAA
GAGTGATGCTTTCTGGTGGACTGGGTTGGCGAGGATGACAAGGGAAATCCTCAGAATCTACCTTACT
GGCGAAGTGGGTGATCACAATGTCTTTGGCATTGTATGCTCTCTCGACGACGTTTTCTTCGTGCGTCT
TCGGGGCTGCAACGCACGTTCTCGCAGAAGAGTTTGCCCTTGCCCGGAGACGGTTCGTGCTCGGTTG
CACTTCCCTCTTCATGGTAGGATTGCTACCGGGCCCATCTTCTGGGGACCCCTTCAGCGAGGCTTTTTG
GCCGGACTCGACCTCTTCTGGCAGGCTACCTGGCCTTTGCGGTCTGCAACTGCCCATCGCCGACGC
AAGAAGTCTGACCTCCATATGTATCCTGCGCTTTCTGGGAGGATTCTTCGGCGCAGCGCCTTCGTCAAT
CCTGTGAGGCATCTTGGCAGATATCTGGTCTCCAGAGAACGAGGTTTCGCTATGCCTACAGTTGGCG
CGTTTCTGACTATCGGGCCAATCCTCGGACCGCTGATTGGATCAGTCTCGTCCAGAGCGTGTGCGGT
TGGCGGTGGATCGCGAATGTCGTTGCGATCGCCTCGTTCTTCATTGCTGTCTTTACCTTTCCATTCTC
CCCGAAACGTACACGCCATTGCTTTTGGCGCGTCTGTGTAACGCATGCGACATATGACACGCAACTG
GGCTTACCGTTCCAAATCTGAAGAAGCGCAGAGCAGTATTGGCGATTTTTCGGGAGCGTTACTTGCTTCG
GCCTGCTCGCATGCTGGCCCTTGAACCCATTCTTTAATGATGACACTCTATGTCAGTGTGTCTTTTCGGT
CTGTTGTACAACCTTTCTTCTGGCATAACCAACATCGTTTATTTCAGGAGCGCGGCTGGGACCAGACGACT
GCCAGTCTTCCACTCATCTCCATCCTCGTGGGCGTCATTATCGCAGGAGCACTACTTTTCGTTTACGACC
AACTCCAGATGGGCACCGAACGCCAAAGAGGGACGACCTCAGGAGACAAGACTTTTGTCTCATGATGGT
TGGCGCTGTCTCGCTCCAGCTGGAATGTTCTCTTCGCTTGGACAAGCTCCGCTACCATGAATCCTTG
GCCTCAAATCTTGTGCGGATTCTACTGGCTTCGGTATCCACTTGATCAATATGCAGGGCATGAACATA
CATCATCGACTCTTACAAAATTTATGCGAACTCGGCCATAGCCGCTAACACCTTCTTGAGATCCCTGTTT
GCTGCAGGATTTCCGATCTTGGCGACTTTCGATGTATGCTGCAATCGGAGTAAAATGGGGAACGACCAT
TCTTGCATTGCTTGCAGTTGCCATGATCCCGATCCCGATACTGTTCTACTACTTCGGTGCCAAAATCAGA
GCGAAGAGCAAGTGGCAGCCACCTTTGTAA

D.1.5. CBET_00835 (CTB5)

ATGTTGGGGCTTAATTTACAACAGGTCTCTCCAATGTTCTACGATATCCAGCATTGTCAGTGGTGTAG
GCTCATATCAACATGGGTGAGACAGCAGTGCATGGGCTTCTGTGGCGGCGTCAAATCTTGTGCGAT
GCGTTGACCAAAAGTTTGGGCAAGAATTCAGTGGTTTTCCCATATGATGCGGCCTACTCTCAATCAATG
GGCTCATACTTACGCTTGAAGAATAGCGACCTCCACCCAGTTGCATAGCACTTCCACGATCAGCAGAA
GATGTCTCGAAGGCGAGTGCACACATTGTCTCTCGGCGCACACAAATGGGAGGGGCAATGCCAATTCG
GCGTCCGAGGTGGTGGCCATACGCCCTTCAAAGGAGCCGCAAGCACCGACAACGGGATTGTGCTCGA
TCTTCTCCACATGCCGTCGCGGGCATATCGCCAGACTATGAAACCATTACAGTGTACCCAGTACGAC
ATGGGATCTGGTATACGAAGTCTCGATGCTCACAACCGAAGCACGCTTGGTACCAAAGTCCCGGTA
TCGGTGTGCGGAGGAGCGTCAACGAGCTGCGGAGTCTCCTATTTCTCGCCTCGCTATGGATATATTTGC
GATATGGTAGAGAACTGGGAAGTCGTTCTGGCCACCGGCGATATCGTCAATGCCAACGCGAATGAGAA
TCCCGATCTCTGAAAAGCGCTGCGAGGTGGAATCAACAACCTTCGGCATCGTACCCGCTGTTACATTGA
AGACCTTTGGACAAGGCCCTTCTGGGGCGGCCAGACCTTCCACTCCATCGACACCCGCCAAGAACAC
TTCAAGAATCATGAGAACTGGCCTCGGCGCATCCCTATGATCCATACGCACATTACATCAATACTCTT
GTCTGGGCAAACGGAGGCCATTGGTTCATCGGCAACAGCATCCAGTACACCAAGAGCGACCCGCCCG
TGGCAGAGCCAGAAGTCTTCAAACCGTTCTCAAGACCGAACGAACCCCGATTTTCCCTGGGCTGCCT
GAAGACTCTTCGAGTGGACAACGTCACCTCGTTCTCCCGCGAGTACGCAGCAAACACCTTATACCC
TCAACGCTGGCAATTCGCGTGCATCAGCTTCGCTCCGGATGCAGACTTCATGGAGACATTCTTCCAAAT
GGCCAACGACGCCATGCAGCAATACGTAATAATTGGCAGGCTTCAAGCTCATCCTCAACTACCAACCCG
CACCAACTGTTTACGCTGGAACGTAATGGAGCAGTGGACTCCCTGGGCCAATCCAAACTGAAGGCAAT
GTTGTCTTTGTACTGTTGGCTGTGCTGCTACGATGAGTCCGAGGCCAGTTCGACGATGCTATACCAA
GAGCGTCCAAGACTTGTCCACGCTGCAAAATGCCAAGGCCAAGAGTTGGGATCTACAGGATTTTTCAT
CCAACCTACTTATGCAGACAGCTGGCAGAGCCCTTTGATTACAGAAGCAAGACAGGATGAGGAAAC
TGTTGCCACGAGCAAGAAGTATGATCCTCTACAGGTTTTCCAGAAGCAGGTGCCGGGTGGCTTCAAG
CTGCCGCGAGATTTGA

D.1.6. CBET_00830 (CTB6)

ATGGCCGACTCTCTCGTTCTATTGACTGGTGCTACCGGCTTCATTGGCTTTCGTATCTTGATTGAGCTG
CTCCGTCAAGGATACAGCGTGGGAGCTGTGATCCGCAGTGCTGGCAAGGGACAATGGCTTGAGTCGC
GGCTGACTGCTGTGATGAAAGGCTCGGACTACAAGGACAGGTTCCGAAACTACCACTGTGGCGGATTC
GTGACAGATGGAGCTTTCGATCAAGCAGCTGAGAACACGTCTTACATTATCCACGTGGCCAGCCCGAT
AGTCAGCTCGGACAACCCAGACGACTGGGAACACGACTTCAAACGTGTTGCTGTGAAAGGTAGTATTG
GGGTCTGGAGGCAGCCAAACGAAGCGGAACTGTGCGACGAGTCGTCATCACATCTAGCATGGTTGG
ACTTTTCTCGCCGAAAGCTCTTTTTGCTGAGCCATCTGAGGTGCCTCTGAATGCAGAATCGCGCATTCC
GGAAATGGAGCCTCCTTATGCGCACAAAATGCTGGCATAACCAAGCTGGGAAGATCGCAAGCATCAACA
GCGCGGAAGCTTGGATCAAGAACGAGAAAACCGGCGTTCGATCTTGTTCACATGCACCCGTCCTTCGTG
ACTGGTCGGGATGACCTAGCGACGACCGGAGGATCTGCGGAAATTCTCGTGAATTGGCATTTCGAT
GCAGATTGTAAGCGCCACAAGAATCCTATCGGCAAGCCAATCCTGACATGTCAACAACGATGATGTCG
CTCGCTGTGATGATTGGCGCTGGATCCCAAGGTCACGGGCAATCAGAGTTTCTTATTTTCGTGCAGTC
CGGAGGATGGATCGGAATGGGATGATGTGAAGAAGTTTCGTGCAGAGGGAGTTTCTGAAGCGGTTGC
GCAGGGCGTGTGGCCAATGACGGTCACATGCCGACTGTCAATAAAGGCGTCAGATTTCGACGTGCGG
AAAACCTGAGGAGACCTTTGGCTTCAAGCACATTCCCTATGAGGCACAAGTCTTGTGTTGTTCAAGCAG
TACTTGGAGTTGCCGGAAGGATGAAGGGGTGGAGATCTCGACGACGGCGTAG

D.1.7. CBET_00836 (CTB7)

ATGGCATCATCTAACCGAAGAGTCCTGGTCAATGGCGGAGGACCCGCTGGCGCCGTGACGGCTTTCT
GGCTTGCCAAAGGGCGGCTTCGAGGTGGTTGTTACAGAGCGCTCAATGAGTCGGCCCTACGGACAAGG
GGTCGACGTCACGGGACGCGCTTCCGACATTATCAAGAAGATGGGATTGGAGCAACGCATTTCGAGACA
GCACTACCGGTGAAGCAGGTCGACCGTTGTGCGACGATCAAGGCGAAGATGTTGCTCCCCCGTGGG
TACTGCACCTATCGAGGGAGGAACGCGCAGCGTCACACAAGAAATCGAGATCATGAGACGGGACTTGA
CAAAAATCTTTGTAGACGCTGCAGAAGCATTGCCAAATGTCACGTTCCGATATGGCTGCACGGTCGACG
AAGTTCAACAGCATGAGAAGTCAATCACTGCCGTTTTATCCGACACAGGCAAACCAGAAGACTTCACTG
CGATCATTGGTGCAGACGGGCTTGGATCAGCTATACGCAAGCTCACATTTCGATCCGGAGATCAACAGA
CGTTCTGTGTCGCCGACCAATACCTACGTTGCGTTCTTCTCCATACCAGGTGATCCAAAGTACGATACC
CCAGTCGGCAAACCTCAACATGCCAACAAGGTCGCGGAATACTTGTCCGCCCCATTGATAAGAAGGG
CACTCAAAGGTCATGTTACCTGATGTGCGAGTCCGATAGCCAGGAGCTGGCACAGGTTGCACGAACTG
GATCGCAAGAAGACCAAAAGGCTCTTTTGGACAACAGATTTCAGAGAGTTACCGGGCCCACTGGGTAAA
CGAGCCGTGAAGGATGACACAGCGCTGACGACTTCTACTTCACCCGCATAGTGCAAATCAAATTGGA
CTCGTGGCATAAGTGGGCGAGCGGCTCTGGTGGGGGATGCTGGTTACTCTCCTTCTCCGCTTACAGGG
CAAGGCACAACCTTTGCTATCATCGGTGCCTATGTGCTCGCGGGTGAGATGGCCAAAAGTCCGGACGA
TCTAGAACGGGCTTTCACCTCGTATTACGCTATACTCAACAAGTTTGTGAGCGAGTCCCAGGAAATTCC
ATTTGGAGGTCAAGCTCCGAACTTATCTTGCCACAGTCAGACTGGGGCATATGGCTTCTTTCGCACCTT
CTACAAAATCATCTCCTGGACAGGCATCTGGCGCTTGCTCAATTCGGGAACGAAACCGTAAAGGTTGA
GCTTCTGAGTACGATTTTGGTGGGCTGGATTGA

D.1.8. CBET_00837 (CTB8)

ATGGCGAAGGGGAGCGCAGGCGATGCACCCAACACAAGAGACACATCTTTCAAGCGCCCGAAGATCA
GGGAATCGTGCCTACTGCTCTAGCCAGAAGATCCGCTGCACCAAAGAACGCCCGCTTGTGCCCGT
TGCGTCAATAAAGGACTCCTATGTGAGTACAACATCTCGCGCGTACCGGCACCAGGCGTCATTCTGTT
CGAGCAACGCCTGAGCCTGAAACAACAATATCCAATGCTCCTACCTCCAGCGTTGCCCGGACTCTGT
TAAGATCGATGGCAAGCGGTGCGCCGCCATGAGCGACTTTGCATTGCTAGATGGTTTGGAGACATTCC
ACAATAGCCTGTGGCACCAGCCCATCACGACAGACATCCAGGATATAGACATGCAGTACTTCGACTTTT
TTGACCCGGGTGGCTACCAAGCTGAGCCGGAACCCATTAATTCATTTCGACATAGACAGCAGCCTTTTGT
GTGGGACATCGACAGCAGGTTATCTTCCAGAGCTGGACGCAGAAGCGTCAACACGGCCTTCATCGTCA
TCATCGCCTCCTAGTCAACGGAGCGATGGCGGCAGGGCGACGACACATGGTGGTGGTGGATGCATAA
GTACGGCGCTCCAGATCTTCTCCGAACCTCATGTATCAAGTTTCAGCGTGCCCCATAGCCGCGGGCGCA
CCATCCCACAACATTCGCGAGTTTCGACCATGTGCTGGACAGCAATCGCGCAGCTTGGAAAAAATGTC
CAGCATTCTGACTGCCCGCGTGTGCCATGACTCAAGAGGTTCTCACGGCCTTGTTCCTGGCGGTTCC
AAAAAGCGCTGTGCTGGTACTCGGCAGCCTTGGATGTGGCGGGCGACGGCGAGCCACCAGCCCGT
CTCGCGGGTGAAGAGCCCGCCGGCATTCTGGGCAGCTACGCGCTTGGAGCCAGGCGCAGACGCT
GGCACGCGCGTATGTGGTATGGCACAGTTGCAACAGCACTTCCAGCCGCTTCTGGCAAAGCTGCAG
CGAATATCGTCCCTGTGCGCGCTGGGGGCGCGCTCGTCTGACGACCTCGCTGAGCTCCGTCTCGT

CGCTGCAGTCATCCACGTCCGGATCGGCGGTGATTGAGTGCCAAAAGCGCGCGCTGCAGGAGGCGCT
GGAGGATGTTGTGGCCAAGATTGAGGGCATCAAGCGCGGGTGA

D.1.9. CBET_00842 (CTB9)

ATGACAAGTACTATCACGACCACTGAGACTCTGCAGGATGCAGTACCATTCTGGCACCACCAAGCCC
ACCTGAAGACACAAGCAACAAGGAATTGCCAGAGAAGCCTTACTACGATGTTGAATTCACCTACCGTCT
TGACCCACGTGATGGAGGCGACGAAGTCATTTGGGGCGGTACTGTTGGTTTGATGCCAGGAAGTATG
AGACCCGACGGTACGAATCAACAATGAACGTGGTAATGAACACAACCTTCAACCTCGACACTCATGGCT
TTGCTTGGGTCAAGCACAAAACCTCCGTGACTGAATTTGCGGACTACTTGGCAATTCGTCAGGGACCGT
ACTTTGGCGAAGTTGCTGAGATGTTGAAGAGGGTCACTGGAGCGACTAAGGTTTCATGTGATCGGACAT
CTCCATCGTTCTGTTGAATTACAACGATACGACTGAGGAGGAGAAGAATGCCCCAGACATGACAATGAC
CAAGGGGCGAGACTCCAGGACGTTTCGTGCACGTGCATCAATCCTACCAAGGCGCAGTTTCGTAGACTCT
ACCTGGATCTTCCCAAGAAGAGGGCCCGCAGGCTGGAGAAAACCAGATGGGCCATCATCAACGTCTG
GCGTCCGTTTCGAAAGTCACCAATGAGCCGCTAGCAGTCTGCGATGCTCGATCAGTCCGAGAGGAC
GAGTTGTTCAACACTCTGCATCTTGTACCAATGAGATGGCCGACGCCGCACCTCAGGAGAACCAGAT
GTGGGCTGTTGCTCCTCCCAAGACTCCAACACAGCACAAGTGGCATTATGTATCTGGTATGACGGAGG
ATGAGGCGTTGTTGATCAAGATGTTTGATTCTAAGAAGGATGGCACTGCAAGACGTGTTCCACATTCTT
CTTCCCTACTCCAGATGATTTTCGGAGAGCCGAGAGCGAGTACTGAAACGAGATGCTTTGTGTTCTGGG
AAGATCAGGAGGCTGAGTAG

D.1.10. CBET_00843 (CTB10)

ATGGGATCCATCGGAGAGCCAAACCGTCTGCTCTGCTGGAGCATCTACGTCACCAAGAAGCCAGACCA
ATCTGAGGAGGATCACCACAACCATGTCTCCAAGTCAATGCTCCCATGATGATAACATTTCTGAAGAA
ATATGGCATTGTTCTGTTACTGTGAAGCACAACGATGCTCACTCTAAGCCCAAGCAAGCGGCTCTGAT
GGCCGGCCAGCCAGAAGAGAATGTCCTCGTTATGACACTGTTTTTGGATGATTGTCAAAGACATTGA
GAGTATTCAGACTATGCAAAAGGATGAGGAGTTTTTTCGTAATACTATTCCAGATCACTTCAACTTTGCG
GATATGACTCGCAGCAAGGGATCGTTGACTTGGATTGAGGAGTTCACATTTCTAG

D.1.11. CBET_00844 (CTB11)

ATGCATTTCCAGCGCTCGCTGTGGCGGGCTGCCTGCTCTCCCGTGCCACTGCGCAATCACTGGATCA
AGTGCTTGCAAAGCGGGACTCATTCTCACTGCTACGCGATCTTCTTCATCAGCATGGCCTGGTCCGATGA
ACTGGAATTCACGGCCAATGCCACGTTTTTTCGCGATGACGAATCAGGCACTGCGCAGCCTGGCCGACT
TTGGCATCAACCTTACGACCGCTGACCCCAACATCGTCTCGAGCGATCTTCAAGTATGCACAACCTGGATG
CCATCTACACAACCGACACGGTCAAGGCATTGCACCACGAAGCAAAGGTTGTGCAGACAGCTCTGCAG
CCACACCTGTTCAACAATTTGACGAGAGGTCAAGCTCGAAGCTGAGAAGCAATCGAACAGGCGGAGC
GAATGGTATCCTGGTCAATCCGGATTGGGTGCTTTCGCGCCGTCGTGGAAGCAGACATTCCTTACG
ATCACGGCGTCAATCAGCTATCGACGCGAACATGGTCTGCTCACAACATTTCTGAGACTGCAAGG
CTCGGCGGCATGACTGAATTTGAACTCTGTTGGAACGATCCGACAGCGTTGCTAGACTGGAGAGCTT
GTCGGATGTCACGATCTTCAATCCGCAAGACGAAGCACTCGCCAAGCTGCAGCCTATTCTGGATATGCT
TACATCGGAGCAGCTGAAGTCTGTGGTGGCTCAGCATGCCGTGCCGAATCGAGTGTGATCAGAGTC
TCTTTGATGGCGTGGAGACTTTGGAACTTTGGATGGCAGCACGTTGCGTATTTCGTCGTGAAAGAGA
GGCGAGATCTTTGTTAATGGTGTGAAGTCGTTTCGCACTGATCTGTTGCTATATGGGGGAGTGGCACAT
TTGATTGATGGGGCTCTTTTGCCTGAGAAAGATGCGTCTGCTCGACCGGTTTATTCGCCGAGCAGC
TGGTGGATCGAGTCTGAGTCTGGAAAATCCTCGCGAGCCACCAGCTTACGCTGCTGGCAGTGTGCGAA
TGGCTCTGGTGTCCATACTGTACAAGGCTTACCAGTCACGGAAACAGAGTCATCAGCTGATCAGACCG
AGCGATGGTCTTGGCAATTATGAGAAGGTGTAG

D.1.12. CBET_00845 (CTB12)

ATGTGGTCTGTTGACTCTACCCCTTGGCGTTGACCTACTGTTTCAATGCGTCAGCCAGCCGACGCT
CGTCCATCGGGATGCGTGGACGATGTCGAGGTCGTTCAAGAGATTGGAAGCAAAGAAATACAGGCTCC
AGTTGTTGCTTCGAAATCTCCCTAACGACACAGTCAATCGACGAGCTGGAATAGGTTTTTCGGGAAGC
GATCTTCATCAACGATGCCTTTATCGGACCTACCCTCTACGCCAAGCAAGGCGATAGGATCGAGTTCGT
CGTTCACAACACTACATGCAACAAGACACAAGCATCCATTTCCATGGCATCGATCAGCGATCTACGCCATG
GTCGGACGGAGTGCCAGGCCTGACACAAAGCCAAATCAGGCCTGGGGCGTCAATCCTGTACAACCTGG

ACTGCACACGACGCCGGAACATACTTCTACCATTCTCACGCCAAATCTCAAATGATGGACGGCCTTTAT
GGCGCGGTGCGTATTGCACCAGATGACGAGGCACCGCGTCCATTTTCATCTTATCAGCAGCGGCGAGG
CCGATCAGGCGGCAATGCTTGCAGCGGAGAAGTTGATGCGACCCATCTTCGTCAGTGACTGGTCCCAG
TACACGTCTGCAGAATACCACGGAATCCAGCATGCTGCGAATATTGATTTCTCGTGCATGGACTCCATC
CTCGTCCAAGGGGTTGGCTCGCAGTACTGCCTGTCCGAGGAAGAGCTTGACGATATGACGAATCCCAT
TGTTTTGCAGCTGTTGAAAGAGCTCGCAGGCGGACATATGACGCCGAAGGGTTGCATCCCTCCGTTGC
AAATGTTCAATGGAGACTTTGAATTGCACCTGGAGAATGTGCCGGAGCTCGCTTACAACAAGTGCAAAG
GAGGACAGAGCTCGAAGGGAAACTATACGATTGACGTTGACACCTCGATCGGCTGGGCGGCGTTGAC
ATTCGTCAATCCTGGAGGGTTGTACCCTCTGCAGCTGTCGATTGACTCGCACGAATTGTACGTGTACGC
TG TAGATGGGCAGTACGTGTACCCAATAGTCGCGAGATCGTGTCTGGTCAACACTGGGAGTCAATAT
CCGTGATGATCAAGCTCGATCAAGAGAAGGCCAGACACGTTGTGCGAGTTGCCAACGATTACCTCAAC
CAGATCTTAGGCGGCTTCGCCGAGCTAGCGTACGATGGCGCGACAAATGCTCCCAAACACCCGCGATCC
GAAGACGAACTATGGTGGCAAGCTGATTAGCAGCGAGATGGTGTGTTTTGTCCCGAGGACAGCAGC
CCATACCCAGCCTTGCAGCTGCGCAGAGCGCCGACTCCACTTTCAAGCTTCGGCTGAAAAAACTAGG
CCAACCTTATAGAGCGTACGAGTGGACGCAGACCGGCAGCCTGGGCTACAACATTAGCCATGAGCAC
GACGACCCTCCGTTGCTGCTGCAGAATGTAGAAGATGTTCCCGCCACGGAACCTACGCTGAAAACGCA
GATCGGTGATTGGGTGACCTCGTCCCTCGTAACAGCAGGTCCATTGCCCCAAGCGCATCCAATGCACA
AACACGGCAAACAAAGTGTTCCCTCATCGGCTCGGGTTCAGGGAGCTTTCCGTGGGAGAGCGTGGAAGA
GGCGATCCCACATCTGCCCGAAGGCACTTTCAACTTCCAAGATCCTCCATACCTGGATACATTCAACAC
AGTGGAGATGGAGGGGCAAGCCAACGATACCTGGACTGCAGTCAGGTACAAGGCTGAATATGCCGGT
GCATGGTTATTCCACTGCCATGTGCAAACGCATTTGTCGGGCGGGATGGGCATGGTGTCTCTGGATGG
CGTGGATGCTTGGCCAGAAGTGCCCTTAGCTTATCAGGAGTGGAAATGGATTTCGAGCCGCTGCTTTGT
CGTGA

D.2.13. CBET_00841 (CFP)

ATGGCAAGCCCAGCGCGATCAACGCATACTGACACAGAATCTCACGATGTTGTGAAGAGTGACTCCGA
ATCGAAACTGGAAGTGGAGCACAGCGATTCTGATATTCAAGATGAGAAGCCAGCCTCGAAATCTGCGG
AGACTCCTGAAGCCAAGTCAGAAGAAGATGAAGAACTCAACGATCAAGGCCAAAAGTATATCTCCGGC
TGGCCTTTGGTATTCTCTTGTAGCCATGGTCTCAACAGTCTTCATTGTGCTTTGAGCAACACCATCA
TCAGCACAGCCATCCCGCCATCACAAACAGCATTCAATAGTACGCGAGATATTGGTTGGTACAACCTCC
GCGAGGCTCTTGCGGCCACAGCATTCCAACACTACCTTTCCGACGAGCGTATCTGTTGATGGACCTGAAG
TGGACTTTTCTCGTGTCAATTGACCCTATACCTGATTGGCAGTCTGATCTGCGGAGTGGCAAACCTTTCT
GTGCTTCTCATTCTGGGCAGAGCAATTGCAGGTGTCCGGCAACGCAGGAGTCTTCGCTGGCGTCTTCAT
CATCATCGCTCGAAACGTTCCCTCTGCGAAAACGAGCACTTTATGCTGGGCTGGTTGGAGCGACATTTGC
CATTGCTGCTGTGCTGGGACCTGTTCTGGGTGGTATCTTACTGACCGTATTAGCTGGAGATGGTGCTT
CTACATCAATCTGCCTATTGGAGCTGTAGCGGTGGCTATCATAGTGTTCCTCTTGCCATCGAGACCAG
TGAAAAAGCAGCAGAGGTGAAGGACCTTTCTGGTGGCAGTTCTTCTAAAGCTCAATCCTTTTGGGTC
GGCCCTTCTACTGGGATCCCTGGCATGTTTCTTTCTTCCCTTCAGTGGGGCGGCGGAATATCCCT
GGAGTGCTGGTTCGCTGGTTCGCGGTCCTTGTGGTCTTCGCTGTCAGCTTCATTGGATGGCTGGTTCTG
CAATATTTCCAAGGTGAAGAGGCCACACTACCATAACAACGTTGCGAAACAGCGTACTGTGGGTGGCGC
CTCCATCTATAACCCTGCTTCTGAGCGCCGCAATTTGGACTCGTGATATACTATCTGCCTCTTTGGTTCAA
GCAGTACGATCTGACAGTGCCGAAGCTGCTGGTCTCAAGCAACTTGGTATCGTCATCTCGCTCACTCTC
TCATCAATTGCAGCTGGCGGTGCTGTTGTA AAAAATAGGATATTACTATCCTTTTCAATTTACGCCGGA
TCTTATGCAGCATCGGCGCTGGCTTGTCTTACACGATCACACTCGACACACCGCAATGGGATATTATCG
GTTATTCGATCGTATTGCCATTGGTATCGGCGTCACTCTCGAGCAATCCAACGTTGCTGTTCCAGACTG
TCCTGCCCGATGCTCAGATTCCAGCAGGCACAAGCTTGGTACTCTTCGTCGATTACTTGGATCAGCAA
TCCCCGGACCCATCGGGCAAAGTGTACTCCAGACAACACTTGCCAGTAGGCTAGGGACTGAGGTGCGT
GAGCAAGCTTATGGTGGCACCGGAGCAACTGAGATCCGCTCAAAGCTCGACAACATCTTTGGAGCTGG
CACACCTGAAGCTCGAGATGCCCTTGACGCTTTCAATGATTCTGTGACGAAGATCTTCATGGTGC
CATAGTCTCTTGTCTGAGTGTGTTGCCTTCCACTCATCGAGCTCAAGAGCGTTAAGCGTGAGAAACG
AGACAACGAAGACGCCAAAGAAGGCAAGAAAATAATGGGACGACGGGCAAGACAGAAGATCCAGAG
AAAGGGCAGAGTGCAGAGATCGCGGGCAAAGCAGTGTGA

D.2. CTB protein sequences

D.2.1. CTB1

MEDGAQMRVVAFGDQTYDCSEAVSQQLLRVRDDAIVVDFLERATAVLKAEARLSSEQQEETPRFATLAELV
PRYRAGTLNPAVSQALTCITQLGLFIRQHSSGQEAYPTANDSCITGVCTGALTAVAVGSASSVTALVPLALHT
VAVAVRLGARAWEIGSCLADARRGANGRYASWTSAVGGISPQDLQDRISAYMTEQALASVSVPYLSAAVGP
GQSSVSAAPVILDAFLSTLLRPLTTTRLPITAPYHAPHLFTAQDVQHVTDCLPPSDAWPTVRIPISFSRDEAVS
RGASFPAA MSEAVRDCLIRPIALDRMAVSIANHARDLGKDSVLPSPIALSFSDKLGQVNSHLPGTKAPTPEL
TSTSSIPSAIGAEQQPMAKSPIAILAASGRFPQSSSMDQFWDVVLINGVDTHELVPPTRWNAATHVSEDPAKAK
NVSGTGFGCWLHEAGEFDAAYFNMSPREAPQVDPAQRLALLTATEVLEQAGIVPNRTSSTQKNRVGVWYG
ATSNDW METNSAQNVDTYFIPGGNRAFIPGRVNYFHKFSGPSYITIDTACSSSLAALHMACNALWRGEVDTA
IVGGTNVL TNPDMTAGLDAGHFLSRSGNCKTFDDEADGYCRGEAVVTLILKRLPDAQADKDP IQASILGIATN
HSAEAASITRPHAGAQQDLFQQVLTETGLTANDISVCEMHGTGTQAGDSGETTSVVETLAPLNRSGSAVRT
TPLYIGAVKSNVGHAE SAAGVSSLAKILLMLKHSKIPPHVGIKTKLNHRLPDLAARNTHIARTEVPWPRPKNG
KRRVLLNNSAAGGNTCLVLEDAPEPEDSQEVDPREHHIVALSAKTPDSMVNNTNMITWIDKHSGDSIATL
PQLSYTTTARRVHHRRAVATGTDLLQIRSSLQEQLDRRVSGERSIPHPNGSPFVLAFTGQGSFAFEGMGV
DLYKRFASFRSDIARYDQICEGMSLPSIKAMFEDEKVFSTASPTLQQLTHVCFQMALYRLWKS LGVQAKAVV
GHSLGEYAALYAAGVLSQSDTYLVGRRRAQLMEKHL SQGTHAMLAVRAKEEAIVAAIDGPPGEAYEFSCRN
GEQRNVLGGTYAQA AKAAL EAKKIRCQYLDTPMAFHTGQVDPILPELLQVAAACSIQDPQIPVISPAYGKVI
RSAKDFQPEYFTHHCRSSVNMVDALQSAVEEGLLDKNIIGLEIGPGPVVTQFVKEAVGTTMQTFASINKDKD
TWQLITQALAKFYLAGASIEWSRYPHEDFPGAQKVLELPAYGWTLKKNYWLQYVNDWSLRKGDPAVVVAASN
LELSSSIHKVITNTITANS DGELVVDADLSREDLHPMVQGHQVYGVPLCTPSVYADIALTLGEYIRQVIKPGEV
AQTSVEVAEMNIQSALVANNTGRVQLLRTYAKFDPKAQVASCTFSSIKEDGSSVVEQHANC KIRFGSLEKEK
TALESAALAAQARMAALKTQVGQDDNTYRFSKGMIMIGQLADFDEKYRGLCAITLNDNAMEASGKVSFK
GIPNEGKFHSSPAYLDALSQ LGGFVMNANEGVDLEKEVFNHGWGSMRFFAALDPAMTYTHVKMTQ GK
DKLWTGDVLI FDDKQALIGIVGGVALQGVPKRMLHYIVTAANKKASGPPTTEKKGSSPPVEKKASAPVAPTRP
AIQRKNASIPPPATQVTPQNKTIKTPSVSALIAPALEIVSEEIGMPIDELKDDIDFTDAGLDSLLSLVISSMRDQ
LGIEFESAQFMEIGSIGGLKEFLTRLSPPVAVAVATAVEIVKEEALTSLEELTDPSPNEIGTVWRDALKILSEES
GLTDEELTDDTSFADVGVDSLMSLVITSRLRDELIDFPDRALFEECQITIFDLRKRFSGSTESF DSTTTKPSA
GDATPPLTDSSASSPPSSEFDGETPMTDLDEVFDSPPAQKRIPSPPKGRIPPAWSMYLQGSQKRSKEILFLF
PDGAGAATSYLSL PRLGEDIGVAVFNPPFMKTPHKFVDHTLPDVIASVVEGIRGRQAQGPYHLGGWSAGGIL
AYAVAQELIAAGEEVSTLLLIDSPSPTKGLDRLPTRFFDHCTNVGLFGTEL SRGSGGPNKTPEWLMPHFRASI
ELLHDYHAPPMKLGNKTKVMVIWAGECAFDGVRYAHIPPSAGD TDEDTEGMKFLTEKRKDFGATEWASLFP
GTDVDARVVESEHHFSMMRDSGAQMLVEHMRDGLGIVSS*

D.2.2. CTB2

MANRIEADNLFELTAE LVSASAKLHKFLDQKNLPQPSFDAPAPSVALNTANKPYYDARSAIVEAAEQ LIRLVR
GPRD TLLALSFEHCATASMQVVFYKXFANHIPLHGSTTYSKIAEAVGDGVTTALVERTIQHCASFGLFETIPG
GYVTHNATSSLLV TDPDLEAWMYLSAVIAYPAGAAIPKAVEQYGVSSSEATEAGYGVSIGRKIAQFQRFREPD
GKKDHMFARAMRGIAAGGAYDFRHAVDGGYPWHLLTEGAGHLVVDVGGGPGHVAMALAEKYP SLRFQV
QDLPETVQVGAKNCPEHLRKHVTFVAHDFMTPQPAHEVQDGEGIVYFARFILHDWSDKYATKIVQALATGL
RPQDRILNEVVVPEAGQVGRETERRMHDRLLMLMNLNGRERTQSAFEAIFASVTPKRLRLQRVIHPEQGE L
SLIEVTLDGVELPAQANGVNGHANGTNGVNGH*

D.2.3. CTB3

MMQFQRDLEASLEAVSANAQELLKSLKSRKDVQDLNASLPKDPLDNCAQTQAARAQLAEAAATRILRLSIRP
QEYLEHLQNGYQHLTCFRWLVELNILDHLP HSGTISYTDLARKASVPPMQLRSICRMAICNGFLEEPEANQV
RHSRISALFARDES YLGWARWMVNYSVPAAYKLSDATRSWGETVAKDQTA FN LGMDVKVPPFDHLRQTPA
MKDAFAAYMRNVTSNATWGLQHAVTGFDWASLPRGAKVVDVGGSLGHGSAIAKEHHTLTFVIQDLPETVA
GARKEMAQNDKIEASVKSRTFQEHDFFGPQTVKDADVFLRMICHDPDNEAKVILSQIRAALKPGAQIVIM
DTILPQPGTTSVLQEQLRIRDLTMMEVFNAKERELEDDWSSLMQSAGLEISRNVQPLNSVMGLLTVRSAGQ
TALSGTNTLTPELVTA VSASTGSADSRPVL IAGAGIAGLCLAQALKKAGIDFRVFERDSHIDARPQGYRLKFE
ADAAQSLKNILPDDVYEA FELSNVAVGETDFNPFNGNIHSRTGGGLSGKKGLYATFTVDRKAFRTQLMT
GIEDKISFGKEIAYYKTDSSASTVTAEFKDGTHVTGSFLAGTDGLHSVVRKTCVPDHRIVD TGAACIYGKTVM

TPEFLARFPEKGLRFMTVVSADIAPMLQSCCLIGDSPVTLLEPIRFSEASRARYPELPADYVYWALIGPKERFG
SQEVTSMKNFVSLDQAAEQAAKLSLAVTEEWHPSLRALFELQDTKQASLIRVASTIPDIPSWESHSNVTVLG
DSIHPMSPCGGVGANTAIVDADALAKVLVEHGTPPVNAIAEFEAAMRTRAKRNIWRSEVGSKRMFQGNL
VDCSEVF*

D.2.4. CTB4

MALPITDDDLGLKQPYVTFSSGSASPPQSTTDAMDLEEQVLDAIKSDAFLVDWVGEDDKGNPQNLPYWR
KWVITMSLALYALSTTFSSSVFGAATHVLAEEFALPAETVVLGCTSLFMVGFATGPIFWGPFSEAFGRTRPLL
AGYLAFAVLQPLIADARSLTSICILRFLGGFFGAAPSSILSGILADIWSPRERGFAMPTVGAFLTIGPILGPLIGS
VLVQSVLGRWRIANVVAIASFFIAVFTFPFLPETYTPLLLARRAERMHRMTRNWAYRSKSEEAAQSSIGFAE
RYLLRPARMLALEPILLMMTLVSVSFLLYNFFLAYPTSFQERGWDTTASLPLISILVGVIIAGALLSFTTNS
RWAPNAKEGRPQETRLLLMMVGAVSLPAGMFLFAWTSSATMNPWPQILSGIPTGFGIHLINMQGMNYIIDSY
KIYANSAIAANTFLRSLFAAGFPILATSMYAAIGVKWGTTILALLAVAMIPILFYFYGAKIRAKSKWQPPL*

D.2.5. CTB5

MLGLNLQQVLSNVPTISSIVSGVGSYQHGSDDSSAWASVAASKSCCDALTKSLGKNSVFPYDAAYSQSMGS
YFSLKNSDLHPSICALPRSAEDVSKAVRTLSLGAHKWEGQCQFQVRRGGGHTPFKGAASTDNGIVDLLHMP
SAGISPDYETITVSPSTTWDLVYEVLDANRSTLGTAVGIGVGGASTSCGVSYFSPRYGYICDMVENWEV
LATGDIVNANANENPDLWKALRGGINNFIVTAVTLKTFGQGFVGGQTFHSIDTRQEHFKNHEKLASAHPY
DPYAHYINTLVWANGGHWFIGNSIQYTKSDPPVAEPEVFKPFLKERTPIFPLPEDTLRVDNVTFSFSREYAA
NTLYPQRWQFACISFAPDADFMETFFQMANDAMQQYVVKLAGFKLILNYQPAPTQQLERNGAVDSLGPQTE
GNVVFVHWAVSYDESEAQFDDAITKSVQDLFHAANAKAKELGIYRHFQPTYADSWQSPFDYRSKSTIEELV
ATSKKYDPLQVFQKQVPGGFKLPQI*

D.2.6. CTB6

MADSLVLLTGATGFIGFRILIELLRQGYSVRAVIRSAGKQWLESRLTAVMKGSDYKDRFETTTVADFVTDGA
FDQAAENTSYIIHVASPIVSSDNPDDWEHDFKRVAVKGSIGVLEAAKRSRGTVRRVITSSMVGFLFSPKALFAE
PSEVPLNAESRIPEMEPPYAHKMLAYQAGKIASINSAEAWIKNEKPAFDLVHMHPSFVTGRDDLATTREDLR
KFSSNWHSMQIVLGHKNPIGKPILTCHNDDVARCHVLALDPKVTGNQSFSLISCSPEDGSEWDDVKKFVQRE
FPEAVAQGVLPNDGHPMTVNKGVRFVDRKTEETFGFKHIPYEAQVLDVVKQYLELPEKDEGVEISTTA*

D.2.7. CTB7

MASSNRRVLVNGGGPAGAVTAFWLAKGGFEVWVTERSMSRPYQGGVDVTGRASDIKKMGLEQRIRDSTT
GEAGLTVVDDQGEDVAPPLGTAPIEGGTASVTQEIEIMRRDLTKIFVDAEALPNVTFRYGCTVDEVQQHEK
SITAVLSDTGKPEDFTAIIGADGLGSAIRKLTDFPEINRRSVSPTNTYVAFFSIPGDPKYDTPVGKLQHANKGR
GILVRPIDKKGTRQSCYLSQSDSQELAQVARTGSQEDQKALLDNRFREFTGPLGKRAVEGMHSADDFYFT
RIVQIKLDSWHSGRAALVGDAGYSPSPLTGQGTTLAIIGAYVLGEMAKSPDDLERAFTSYAYAILNKVSESQ
EIPFGGQAPKLILPQSDWGIWLLRTFYKIISWTGIWRLLNFGNETVKVELPEYDFGGGLD*

D.2.8. CTB8

MAKGSAGDAPNTRDTSFKRPKIRESCHTHCSSQKIRCTKERPACARCVNKGLLCQYNISRRTGTRRHSVRAT
PEPETTISNAPTSSVAPDSVKIDGKRSPAMSDFALLDGLETFDNLWHQPITTDIQDIDMQYFDFDPGGYQA
EPEPINSFDIDSTLLCGTSTAGYLPELDAEASTRPSSSSSPPSQRSDGGRATTHGGGGCISTALQIFSELHVS
SSACPIAAGAPSHNIREFDHVLDSNRAALEKLSSILDCPPCCHDQEVLTALFLAVQKALSWYSAALDVAGDG
EPTSPSSRVKSPPAFLGSYALGAQAQTLARAYVMAQLQQHFQPLLAKLQRISSLSALGARSSSTTSLSSVS
SLQSSTSGSAVIECQKRALQEALEDVVAKIEGKRG*

D.2.9. CTB9

MTSTITTTETLQDAVPFVAPPSPPEDTSNKEPEKPYDVEFNRYRLDPRDGGDEVIWGGTVGLMRRKYETR
TVRINNERGNEHNFNLDTHGFAWVKHKTSVTEFADYLAIQGPYFGEVAEMLKRVGTGATKVHVHIGHLHRS
NYNDTTEEEKNAPDMTMTKGQTPGRFVHVDQSYQGAVRRLYLDLPQEEARRLEKTRWAIINVWRPVRKVT

NEPLAVCDARSVREDELFNTHLHVPMPRPDAAPQENQMWAVAPPKTPTQHKWHYVSGMTEDEALLIKMF
DSKKDGTARRVPHSSFPDPPDFGEPRASTETRCFVFWEDQEAE*

D.2.10. CTB10

MGSIGEPNRLLCWSIYVTKKPDQSEEDHHNHVSKVNAPMMIPFLKKYGIVRYTVKHNDASHKPKQAALMAG
QPEENVLAYDTVFEMIVKDIESIQTMQKDEEFLRTTIPDHFNFADMTRSKGSLTWIEEFTF*

D.2.11. CTB11

MHFPALAVAGCLLSRATAQSLDQVLAKRDSFSLLRDLLHQHGLVDELEFTANATFFAMTNQALRSLADFGIN
LTTADPNIAIRAIFKYAQLDAIYTTDTVKALHHEAKVVQTALQPHLFNNLTRGQAAKLRNRTGGANGILVESG
LGVLTPVVEADIPYDHGVIHAIDANMVLPHNISESTARLGGMTEFLNLLERSDSVARLESLSDVTFIPQDEALAK
LQPILDMLTSEQLKSVVAQHAVPNRVLYQSLFDGVETLETLDGSTLRIRRGKRGEIFVNGAEVVRTDLLLYGG
VAHLIDGALLPEKDASCSTGLFAAAAGSSRVWKILASHQLTLLAVLAMALVSILYKAYQSRKQSHQLIRPSD
GLGNYEKV*

D.2.12. CTB12

MWSVRLYPLALTLFQCVSPAAARPSGCVDDVEVVQEIGSKEIQAPVVRFEISLTTQSIDAAGIGFREAFIND
AFIGPTLYAKQGDRIEFVVHNMQQDTSIHFHGIDQRSTPWSGVPGLTQSQIRPGASFLYNWTAHDAGTY
FYHSHAKSQMMDGLYGAVVIAPDDEAPRPFHLISSGEADQAAMLAEEKLMRPIFVSDWSQYTSAEYHGIQH
AANIDFSCMDSILVQGVGSQYCLSEEELDDMTNPIVLQLLKELAGGHMTPKGCIPPLQMFNGDFELHLENVP
ELAYNKCKGGQSSKGNVTIDVDTSIGWAALTFVNPGLYPLQLSIDSHELIVYAVDGGYVYPIVADRVLVNT
GSRISVMIKLDQEKARHVVRVANDYLNQILGGFAELAYDGATNAPKHPKHPKNTNYGGKLISSEMVSFVPEDSS
PYPALRPAQSADSTFKLRLKLGQPYRAYEWTQTGSLGYNISHEHDDPPLLLQNVEDVPATELTLKTQIGDW
VDLVLVTAGPFAQAHPMHKHGNKVFLIGSGSGSFPWESVEEAIPHLPEGTFNFQDPPYLDTFNTVEMEGQA
NDTWTAVRYKAEYAGAWLFHCHVQTHLSGGMGMVVLGDVDAWPEVPLAYQEWNGFEPALAS*

D.2.13. CFP

MASPARSTHTDTESHDVVKSDESKELEHSDSDIQDEKPASKSAETPEAKSEEDEELNDQGEKYISGWPL
VFLLLAMVSTVFIVALSNIIISTAIPAITTAFNSTRDIGWYNSGEALAATAFQLPFGRAYLLMDLKWTFVLSLTY
LIGSLICGVANSSVLLILGRAIAGVGNAGVFAGVFIIARNVPLRKRALYAGLVGATFAIAAVLGPVGGIFTDRIS
WRWCFYINLPIGAVAVAIIVFLLPSRPGEKAAEVKDLSSWWQFFLKLNPFGSALLLGSACFFLALQWGGGEY
PWSAGRVAVLVFAVSFIGWLVLQYFQGEATLPYNVAKQRTVGGASIYTLTLLSAAFGLVIYYLPLWFQAV
RSDSAEAAGLKQLGIVISLTLSSIAAGGAVVKIGYYPFIYAGTILCSIGAGLLYTITLDTQWDIIGYSIVFAIGIG
VSLEQSNVAVQTVLPDAQIPAGTSLVLFVRLLSAIPGPIGQSVLQTTLASRLGTEVAEQAYGGTGATEIRSK
LDNIFGAGTPEARDALDAFNDSVTKIFMVAIVSCLSVLPLPLIELKSVKREKRDNEAKEGKKTNGTTGKTED
PEKGQSAEIAKAV*

Curriculum Vitae

Callie Rose Elizabeth Huitt-Roehl was born to Jan A. Roehl and Robert E. Huitt on July 28, 1988 in Falls Church, Virginia. Shortly thereafter she moved to Pacific Grove, CA, where she grew up and attended elementary through high school. She attended the University of California Santa Cruz, where she obtained a Bachelor of Science in Biochemistry and Molecular Biology in 2010. In 2011, she returned to the east coast to attend graduate school at Johns Hopkins University in Baltimore, Maryland. At Johns Hopkins, she joined the Chemistry-Biology Interface program, where she studied fungal polyketide biosynthesis in the lab of Craig A. Townsend. As a graduate student, she served as a teaching assistant for several organic chemistry courses, both lecture- and laboratory-based. She also gave several guest lectures in Bioorganic Chemistry and mentored a number of young graduate students from the Chemistry and Chemistry-Biology Interface programs.