UNIVERSITY OF CALIFORNIA, SAN DIEGO

Synthetic Active Site Probes for PKS and NRPS Biosynthetic Enzymes

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Jordan Leslie Meier

Committee in Charge:

Professor Michael Burkart, Chair
Professor Steve Briggs
Professor Simpson Joseph
Professor Tadeusz Molinski
Professor Jerry Yang

2009

The Dissertation of Jordan Leslie Meier is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2009

DEDICATION

To Mark Kearley, for talking a 20-year old kid out of pharmacy school.

To Mom and Dad, for allowing me to dwell in possibility
(and putting up with it all).

To Aimee, the light at the end of the tunnel.

EPIGRAPH

For a long time I was reporter to a journal, of no very wide circulation, whose editor has never yet seen fit to print the bulk of my contributions, and, as is too often common with writers, I got only my labor for my pains. However in this case my pains were their own rewards.

-- *Henry David Thoreau*

You ain't gotta lie to kick it!

-- *Tupac Shakur*

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

4’-PPant	4’-phosophopantetheine
A	adenylation domain
ABPP	activity-based protein profiling
ACP	acyl carrier protein
AMP	adenosine monophosphate
ArCP	aryl carrier protein
AT	acyltransferase domain
ATP	adenosine triphosphate
B-Ala	beta alanine
C	condensation domain
C	Celsius
CC	click chemistry
CoA	coenzyme A
CP	carrier protein
DBP	1,3-dibromopropane
DEBS	6-deoxyerythronolide
DH	dehydratase domain
DMACA	7-dimethylamineocoumarin-4-acetic acid
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DPCK	dephospho-CoA kinase (CoaE)
E	epimerization domain
ER	enoyl reductase domain
ESI	electrospray ionization
FA	fatty acid
FAS	fatty acid synthase
FP	fluorophosponate
FRET	fluorescence resonance energy transfer
FTMS	ion cyclotron Fourier-transform mass spectrometry
h	hour
HMW	high molecular weight
IPTG	isopropyl-β-D-thiogalactopyranoside
kcat	catalytic rate
Km	substrate concentration required for ½ Vmax
KR	ketoreductase domain
KS	ketosynthase domain
LC	liquid chromatography
LB	Luria-Bertani
M	molar
MALDI	matrix-assisted laser desorption/ionization
mg	milligram
MIC	minimum inhibitory concentration
min	minute

mL	milliliter
MLSA	mycolactone synthase A
mM	millimolar
MS	mass spectrometry
MS-MS	tandem-mass spectrometry
MudPIT	multidimensional protein identification technology
N5-Pan	N5-pentyl pantothenamide
NAC	N-acetyl-cystamine
NaCl	sodium chloride
nM	nanomolar
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthetase
OASIS	orthogonal active site identification system
PAGE	polyacrylamide gel electrophoresis
Pan	pantothenate
PanK	pantothenate kinase (CoaA)
PCP	peptidyl carrier protein
PEG	polyethylene glycol
Pik	pikromycin
PK	polyketide
PKS	polyketide synthase
PMSF	phenylmethylsulfonyl fluoride
pNPP	p-nitrophenyl propionate
PPAT	4’-phosphopantetheine adenyltransferase (CoaD)
PPTase	phosphopantetheinyl transferase
s	second
SDS	sodium dodecyl sulfate
Sfp	phosphopantetheinyl-transferase from B. subtilis
SKBR3	human breast cancer cell line
TE	thioesterase domain
TOF	time-of-flight
Tyc	tyrocidine
µg	microgram
µL	microliter
µM	micromolar
Vi	initial velocity
Vmax	maximum velocity

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGMENTS

This is by far the most important part of my thesis. None of this work would have been possible without the strong foundation of my family. Dad, you always encouraged me to make an interesting life for myself, and I think I've found one. Mom, you are a reservoir of strength and courage that inspires me every time we talk or see each other. I promise I'll be home soon. Matt and Nathan, your little brother now has a higher degree than either of you. Just though I'd remind you of that. Nyah nyah! Joking aside, by far the biggest sacrifice I have made for these past five years has been spending my time so far from each of you – I love you all.

Being far from home, I've been fortunate enough to find a surrogate family here in California. Foremost among these is my love and future wife, Aimee Fourmont. In addition to your beauty, intelligence, and sense of humor, you are the most honest and truest friend I have ever known. Thank God for cow suits! I love you babe and look forward to our journey through life. Now that I'm graduated, it's PEANUT BUTTER JELLY TIME!! In addition to Aimee and the Burkart lab (I'll get to them later) some other good friends who have helped support me these past five years include Ted Sherer, Darren and Alycia Austen, Jason Karp, RJ Scott, Morgan Lawrenz, Joy Bala, Krystal Farthing, Matt Schenauer, and Mark Dante. David Dobberpuhl and Mark Kearley were former mentors at Creighton who were kind enough to continue that role from afar, and provided me with many encouraging words and much wise advice in a difficult situation. I also must thank Wendy Strangman, who coaxed me through a dark time I had following a bicycle accident. To those last three in particular, you each helped far more than you know.

The research contained in this thesis could not have been performed without the help of the absolutely incredible scientists I was blessed to work with as coollaborators. Deserving special mention in this category is Dr. Sherry Niessen at the TSRI Center for Physiological Proteomics, one of those scientists so proficient and productive that just watching her staggering output makes you almost want to give up. Luckily not every scientist has to be a Sherry, just like not every ballplayer has to be Michael Jordan, but it sure is fun to play on their team. In addition to Sherry and her compatriot Heather Hoover, I had the good fortune to work with Jeremy Baskin (Bertozzi Lab, UC Berkeley), Charles Stewart and Mike Austin (Noel Lab, Salk Institute), and Samira Danesh (Nizet Lab, UCSD). Your infinite patience in answering my many ridiculous questions was essential to my growth as a scientist. For that you have my thanks, and my sincere hope that we remain friends and colleagues in the future. Also deserving thanks are the technicians and support staff at UCSD, including Anthony Mrse (NMR), John Lasky (stockroom), Majid Ghassemian and Justin Torpey (Biomolecular MS Core), and Yongxuan Su (small molecule MS).

Finally, the Burkart lab. “There’s no lab like the Burkart lab.” Those were Heriberto Riviera’s words when he came back for a visit a year after leaving his position as a research technician in our lab for the PhD program at UT-Austin, and I think that says it best. This motley crew of scientists with expertise ranging from synthetic organic chemistry to protein purification to soil bacteria microbiology has been my crazy dysfunctional family these past five years. We have laughed, cried, ate, drank, cheered, fought, swore, plotted and schemed together. When I had surgery to repair my right ankle, it was Alex Mandel who drove me home after. When I knocked

my teeth out in a bike accident, it was Andrew and Elinore Mercer who met me at the hospital, and Matt and Jess Alexander who gave me a ride to work the next day. When our car broke down on the way back from Las Vegas, it was Andrew Worthington, Andrew Mercer, and Gene Hur who I waited with in the Mojave desert for a tow-truck, and who sloggged their luggage and hangovers through the sun-baked streets of Barstow with me on May Day the following morning. In happier times, when Aimee and I got engaged in Paris, it was Alex Mandel who we had lunch with the next day and who was the first of our friends to know. Likewise Reymundo Villa will be the officiant at our wedding. You all are so deeply engrained in my life it is difficult and a bit frightening to imagine I will not be seeing you all on a daily basis sometime soon. Thank you for your help, training, ideas, and cameraderie these past five years. I cannot say enough how much I have treasured our time together and I sincerely hope we can keep in touch.

While I have mentioned some of them already, a few Burkart members deserve special recognition. Probably no scientist had a bigger impact on me in my first two years in graduate school, than Matt Alexander. In addition to teaching me what little I know of synthetic organic chemistry, I can think of no finer example for a young graduate student than Matt with regards to the work ethic and intellectual rigor required to be a truly excellent scientist. Something must also be said for his rancourous wit and profane mouth (always set to “11”), which, while sometimes making it feel as if you were doing chemistry on a construction site, contributed greatly to lab morale.

About a third of the work contained in this thesis was performed by Andrew Mercer, and about a third of the work in his thesis was performed by me, so we'll call it even. If Matt Alexander taught me how to work, it was Andrew "Cowboy" Mercer who taught me how to formulate ideas about what to work on. Over the course of many, many Grove coffees we hashed out the ideas and themes that this thesis centers around. Andrew has the rare gift of being among the smartest people you've ever met, but with one of the smallest egos (I attribute this to his wife), which makes him a pleasure to work with. Mercer is one of my best friends and will always remain, for a certain generation of lab members, the true "heart and soul" of the Burkart lab.

It is fitting that Reymundo Villa joined the lab just as Mercer was leaving, because he has stepped solidly into the role of my main lab confidante and smack talker. Thank you for your amazing upbeat optimism, patiently listening to all my proteomic talk, and most importantly for at last bringing a knowledge of real hip-hop to the Burkart lab. You will always be family to me and Aimee. Soon we will sit at Dodger Stadium with the ese's holmes! I couldn't have picked a better lab-mate to sit next to for the last five years than Gene "Pencil" Hur, who keeps an even keel and always gives a ready laugh, even when I'm stealing his glassware. To Chops and Tim, the other two members of our graduate class, I say, we're almost there!! I also enjoyed spending time with Andrew Worthington, another lab member who taught the lab boxing, poker, and is also a helluva scientist, despite his Jersey pedigree. Although Alex Mandel and I did not have much scientific overlap (pantetheine!) he is another genius, and one of my most dependable, best friends. Tiffany Yano and Bill Walkowicz were two other Burkart lab members who, though their tenure was short,

helped make the experience an enjoyable one. Andy Smith is the best undergraduate you could ever ask for, and has a bright future, whatever it is he decides he wants to do with himself. Bob Haushalter works hard, plays hard, and tells the best stories you can ever imagine (all true). With him and Justin Hammons, I can leave knowing the lab is in good hands.

Finally I must say thank you to Mike Burkart. You have been so generous with projects, funding, space, support and most importantly your time that I cannot begin to say enough. The Burkart lab is a fun environment to work on unique, creative scientific problems and I think that spirit comes ultimately from the top. You have always had an open door to me and been willing to hear my conflicts and concerns, in addition to my wingnut ideas. You will be a person I look to for guidance throughout my scientific career, and in the future I expect you will remain not only a mentor to me, but also a colleague and good friend. Thank you again.

Now enough with the fun stuff.....

The text of chapter 2, in full, is a reprint of material as it appears in: Meier, J.L., and Burkart, M.D. (2009). The chemical biology of modular biosynthetic enzymes. Chemical Society Reviews, DOI: 10.1039/b805115c. I was the primary author of this work.

The text of chapter 3, in full, is a reprint of material as it appears in: Meier, J.L., Mercer, A.C., Rivera, H., and Burkart, M.D. (2006). Synthesis and Evaluation of Bioorthogonal Pantetheine Analogues for In Vivo Protein Modification. *J. Am. Chem. Soc. 128*, *12174-12184*. I was the primary author of this work.

The text of chapter 4, in full, is a reprint of material as it appears in: Mercer, A.C., Meier, J.L., Hur, G.H, Smith, A.R., and Burkart M.D. (2008). Antibiotic evaluation and in vivo analysis of alkynyl coenzyme A antimetabolites in Escherichia coli. *Bioorg. Med. Chem. Lett. 18*, *5991-5994*. I was the co-primary author of this work.

The text of chapter 5, in full, is a reprint of material as it appears in: Mercer, A.C., Meier, J.L., Torpey, J.W., and Burkart M.D. (2009). In vivo modification of native carrier protein domains. *Chembiochem. 10*, *1091-1100*. I was the second author of this work and contributed substantially to the results as well as the preparation of the manuscript.

The text of chapter 6, in full, is a reprint of material as it appears in: Meier, J.L., Mercer, A.C., and Burkart, M.D. (2008). Fluorescent profiling of modular biosynthetic enzymes by complementary metabolic and activity based probes. *J. Am. Chem. Soc. 130*, *5443-5445*. I was the primary author of this work.

The text of chapter 7, in full, is a reprint of material as it appears in: Meier, J.L., Barrows-Yano, T., Foley, T.L., Wike, C. L., and Burkart M.D. (2008). The unusual macrocycle forming thioesterase of mycolactone. *Mol. Biosyst.4*, *663-671*. I was the co-primary author of this work.

The text of chapter 8, in full, has been submitted for publication. I am the primary author of the manuscript. Robert Haushalter was responsible for the expression and purification of recombinant CoA biosynthetic enzymes. I designed and performed the experiments. All research was performed under the supervision of Professor Mike Burkart.

The text of chapter 9, in part, is in preparation for publication. I am the primary author of the manuscript. Sherry Niessen and Heather Hoover performed MudPIT analyses and aided in aided in protein enrichment and data analyses. Professor Benjamin Cravatt supervised mass spectrometry experiments and provided fluorophosphonate-biotin. Tim Foley provided biotin-CoA. I designed and performed the experiments. All research was performed under the supervision of Professor Mike Burkart.

VITA

Education

Doctor of Philosophy, Chemistry 2009
University of California at San Diego, San Diego CA

Bachelors of Science, Chemistry 2004
Creighton University, Omaha NE

Experience

- **UCSD** 2004-2009
  Graduate Researcher
  Teaching Assistant, UCSD Chemistry Department

Publications

- **Meier J.L.**, Hoover H.S., Foley T.F., Niessen S., Burkart M.D. Orthogonal active site probes for shotgun proteomics identification of PKS and NRPS enzymes. *In preparation*.
- **Meier J.L.**, Haushalter R.W., Burkart M.D. A mechanism based protein crosslinker for acyl carrier protein dehydratases. *Submitted*.
- **Meier J.L.**, Burkart M.D. The chemical biology of modular biosynthetic enzymes. *Chem. Soc. Rev.* **2009** DOI: 10.1039/b805115c.
- **Meier J.L.**, Burkart M.D. Synthetic probes for polyketide and nonribosomal peptide biosynthetic enzymes. *Methods Enzymol.* **2009** 458, 219-254.
- Hur G.H., **Meier J.L.**, Baskin J.M., Codelli J.A., Bertozzi C.R., Marahiel M.A. Burkart M.D. Crosslinking studies of protein-protein interaction in nonribosomal peptide biosynthesis. *Chem. Biol.* **2009** 16, 372-391.
- Mercer, A.C., **Meier J.L.**, Torpey, J.W., Burkart M.D. In vivo modification of native carrier protein domains. *Chembiochem* **2009** 10, 1091-1100.
- **Meier J.L.**, Mercer A.C., Hur G.H., Smith A.R., Burkart M.D. Antibiotic evaluation and in vivo analysis of alkynyl coenzyme A antimetabolites in Escherichia coli. *Bioorg. Med. Chem. Lett.* **2008** 18, 5991-5994.
- Worthington A.S., Hur G.H., **Meier J.L.**, Chen Q.C., Moore B.S., Burkart M.D. Probing the Compatibility of Type II Ketosynthase – Carrier Protein Partners. *Chembiochem* **2008**, 9, 2096-2103.
- **Meier J.L.**, Yano T.B., Foley T.L., Wike C.L., Burkart M.D. The unusual macrocycle forming thioesterase of mycolactone. *Mol. Biosyst.* **2008**, 4, 663-671.

- **Meier J.L.,** Mercer A.C**.**, Burkart M.D., Fluorescent profiling of modular Biosynthetic enzymes by complementary metabolic and activity based probes. *J. Am. Chem. Soc.* **2008**, 130, 5443-5445.
- **Meier J.L**., Mercer A.C., Rivera H., Burkart, M.D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J. Am. Chem. Soc.* **2006**, 128, 12174-12184.
- L'Amoreaux W.J., **Meier J.L.**, Chauhan B.S. Creation of silver nanoscaffolds using type I collagen as a template. *Microsc. Microanal.* **2004**, 10, 1436-1437.

ABSTRACT OF THE DISSERTATION

Synthetic Active Site Probes for PKS and NRPS Biosynthetic Enzymes

by

Jordan Leslie Meier

Doctor of Philosophy in Chemistry

University of California, San Diego 2009

Professor Michael Burkart, Chair

Polyketides and nonribosomal peptides constitute two classes of natural products which are well known for their application as antibiotic, antiparasitic, antifungal, and anticancer agents. One of the most remarkable discoveries yielded by studies into the biosynthesis of these compounds has been the finding that these seemingly disparate chemical structures are produced by a similar biosynthetic logic, in which multimodular enzymes activate, condense, and tailor a series of monomer units to produce the final natural product. The finding that the sequence and identity of active sites occuring in polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) enzymes dictates the structure and biological activity of the natural product has lead to studies towards reengineering these synthases for the production of novel small molecules, an approach known as combinatorial biosynthesis. However, the many difficulties encountered in such studies emphasize the need for a greater knowledge of these biosynthetic systems at the protein level before the full potential of

such approaches can be achieved. The goal of the work in this thesis was to develop and apply chemical probes capable of facilitating structural and proteomic studies of PKS and NRPS enzymes, two aspects of these systems which have proven problematic to traditional methods of analysis. For this purpose, a series of fluorescence, affinity, and crosslinking probes were synthesized and evaluated for their ability to label enzyme activities commonly utilized by PKS and NRPS enzymes. These studies lead to the development of a chemoenzymatic crosslinking approach capable of directly reporting on the protein-protein interactions of the acyl carrier protein-dehydratase pair found in PKS systems, as well as methods for the specific labeling of PKS and NRPS carrier protein, ketosynthase, thioesterase, and dehydratase domains with fluorescence and affinity reagents in complex proteomic mixtures. Finally, these chemical proteomic probes were integrated with a high-content, mass spectrometry based protein identification platform for the global proteomic analysis of PKS and NRPS biosynthesis in the model bacterium *Bacillus subtilis*. Through their continued application such methods have the potential to provide unique insights into PKS and NRPS biosynthetic pathways, facilitating the discovery and production of new therapeutic agents.

Chapter One

Introduction

Adapted from:

Meier, J. L.; Burkart, M. D. *Methods Enzymol* **458**, 219-54 (2009).

Introduction

The natural product biosynthetic pathways responsible for the production of fatty acids, polyketides, and nonribosomal peptides represent a fascinating intersection of primary and secondary metabolism. In organisms ranging from microbes to man, these enzymes are used to produce small molecules which span the spectrum of biological activity from vital primary metabolites, to therapeutic antibiotic agents, to harmful bacterial virulence factors.

Due to the significant role many of these compounds play in promoting human health and combating disease, the catalysts responsible for their biosynthesis, known as fatty acid synthase (FAS), polyketide synthase (PKS), and nonribosomal peptide synthetase (NRPS) enzymes, have been the subject of intense biochemical, genetic, and structural characterization in the past 30 years.[1,2] One of the most remarkable discoveries of these studies has been the observation that the seemingly disparate chemical structures of fatty acids, polyketides, and nonribsomal peptides are all produced by a similar process. While the individual details of each pathway can be highly nuanced, lying at the center of each is a modular biosynthetic logic, which utilizes covalent acyl-enzyme intermediates, iterative cycles of condensation, and tailoring of monomer units to produce the structural and functional diversity associated with these classes of small molecules.[2,3] A thorough description of this process is provided in Chapter 2.[4] Because PKS and NRPS product formation proceeds from a linear series of biosynthetic steps, each of which is performed by a discrete enzymatic domain, each of which is encoded at the genetic level, analysis of

PKS and NRPS gene clusters can often allow for prediction of the structure of their cognate small molecule product.[5] The modular nature of these molecular assembly lines has also lead to efforts towards the reengineering of these synthases for the production of PKS and NRPS enzymes capable of producing novel small molecules, an approach known as combinatorial biosynthesis.[6] However, despite some notable exceptions, such strategies have been largely unsuccessful, with genetically engineered PKS and NRPS systems often showing reduced or negligible catalytic activity compared to natural systems.[7]

Despite these initial setbacks, the huge potential of combinatorial biosynthetic methods has lead researchers to continue to seek a more complete understanding of the enzymatic chemistry of these systems for the ultimate goal of their efficient reengineering. In more recent years, a number of chemical biological approaches have been developed for the analysis of FAS, PKS, and NRPS enzymes, allowing the study of previously impenetrable aspects of their enzymology. A comprehensive review of such approaches and their historical development is introduced in Chapter 2.[4] The goal of the work in this thesis was to develop and apply chemical probes capable of facilitating structural and proteomic studies of PKS and NRPS enzymes, two facets of these systems which have proven problematic to traditional methods of analysis. Initially our strategy for both of these goals centered around manipulating a unique posttranslational modification found in FAS, PKS and NRPS carrier protein (CP) domains, the 4'-phosphopantetheine (4'-PPant) prosthetic group.[8] This posttranslational modification is introduced by the action of a 4'-phospho-pantetheinyltransferase (PPTase) enzyme, which transfers the 4'-PPant group from

coenzyme A to a conserved serine residue of the PKS or NRPS CP domain. The terminal thiol of this 4'-PPant group then serves to tether intermediates during FAS, PKS, and NRPS biosynthetic cycles through a covalent thioester linkage. Notably, many of the PPTase enzymes which introduce the 4'-PPant modification have a relaxed substrate specificity which allows them to utilize CoA analogues to modify CP domains.[9,10] While in theory this could be used to selectively introduce fluorescence and/or affinity handles into PKS and NRPS enzymes for their visualization and purification from complex proteomic mixtures through applicaton of a permissive PPTase and CoA analogue, in practice this approach is effectively blocked by the endogenous 4'-PPant modification which is introduced into these enzymes during normal cell growth prior to proteomic isolation.[10]

This lead us to explore methods for the introduction of CoA analogues incorporating fluorescence or affinity handles into the intracellular CoA pool of growing natural product producing organisms. Chapter 3 presents a detailed study of this approach, in which a panel of fluorescent and bioorthogonal CoA precursors (denoted as "pantetheine analogues") were evaluated for their ability to cross the cell membrane of living *E. coli* cells and be transformed into CoA analogues by the organism's endogenous CoA biosynthetic pathway.[11] Through careful synthetic elaboration of the pantetheine scaffold, we identified key structural elements necessary for formation of CoA analogues in vivo, and demonstrated these CoA analogues could be used to label and detect CP domains through the PPTase reaction in a model *E. coli* overexpression system. Chapter 4 explores the biological activitiy of CoA precursors from a different perspective, examining their varied properties as antibacterial agents.

Here synthesis and evaluation of a number of CoA precursor analogues allowed us to build a secondary structure-activity relationship providing an explanation for the differential toxicity of these analogues.[12] In chapter 5 we return to the application of biodetectable CoA precursors as in vivo labeling reagents for proteomic analysis of CP domains, demonstrating their utility in the detection and enrichment of FAS enzymes from a variety of naturally occurring bacterial organisms.[13] Here metabolic labeling of CP domains by an azido-CoA precursor was detected after cell lysis by Cu(I)-catalyzed [3+2] cycloaddition reaction with a conjugate reporter alkyne and SDS-PAGE. Notably this approach together with MALDI-TOF/TOF mass spectral analysis allowed identification and cloning of a bacterial CP domain from the unsequenced organism *Brevibacillus brevis*, a validation of the ability of this technique to enable "proteome mining" approaches to modular biosynthetic enzyme discovery.

Because the utility of metabolic labeling approaches for detection of FAS, PKS, and NRPS CP domains is dependent on the permissive nature of an organism's endogenous CoA and PPTase biosynthetic pathways, chapter 6 explores the utility of electrophilic probes targeting alternate enzyme activities commonly found in modular biosynthetic systems for the proteomic detection of PKS and NRPS enzymes.[14] These studies demonstrated that probes of the ketosynthase (KS) and thioesterase (TE) domains could be used for the functional assessment of PKS and NRPS activity in vitro, and could complement the use of CP probes to allow for the gel-based identification of the NRPS enzyme SrfAC in an unfractionated proteome isolated from the well-characterized gram positive bacterium *Bacillus subtilis*.

Chapter 7 then makes use of the in vitro applications of CP and TE active site probes to complement traditional biochemical studies in the study of MLSA2, a PKS TE domain involved in the biosynthesis of the *Mycobacterium ulcerans* virulence factor mycolactone.[15] Chapter 8 contains another in vitro application of these active site probes, this time using chemoenzymatic modification of a FAS CP domain with a 4'-PPant arm containing the canonical suicide substrate 3-decynoyl-N-acetyl cystamine to covalently crosslink ACP and dehydratase (DH) domains from the *E. coli* fatty acid biosynthetic pathway.[16] By covalently immobilizing the normally transient protein-protin interactions of the ACP-DH enzyme pair, this method may have appliations in the structural study of FAS and PKS biosynthetic enzymes.[17]

Finally, chapter 9 contains a comparative analysis of the utility of applying CP and TE active site probes for whole proteome enrichment in combination with the mass spectrometry platform known as multidimensional protein identification technology (MudPIT) for the global proteomic analysis of PKS and NRPS biosynthesis.[18,19] Working again in the model natural product producer *B. subtilis*, we demonstrate the ability of CP and TE active site probes to identify and quantitate PKS and NRPS enzymes at the protein level in two strains of this organisms, and also analzye their enrichment properties on the peptide level as a preliminary assessment of their utility in the proteomic discovery of PKS and NRPS enzymes. Taken as a whole, these studies provide detailed insights into the relative utility and limitations of these methods and set the stage for the routine application of synthetic active site probes in the structural characterization and proteomic analysis of PKS and NRPS enzymes.[20]

References

1. Smith, S.; Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **24**, 1041-72 (2007).

2. Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106**, 3468-96 (2006).

3. Sieber, S. A.; Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105**, 715-38 (2005).

4. Meier, J. L.; Burkart, M. D. The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev* (2009).

5. McDaniel, R.; Welch, M.; Hutchinson, C. R. Genetic approaches to polyketide antibiotics. 1. *Chem Rev* **105**, 543-58 (2005).

6. Weissman, K. J.; Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**, 925-36 (2005).

7. Kennedy, J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat Prod Rep* **25**, 25-34 (2008).

8. Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **69**, 961-1004 (2000).

9. 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* **37**, 1585-95 (1998).

10. La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. Manipulation of carrier proteins in antibiotic biosynthesis. *Chem Biol* **11**, 195-201 (2004).

11. Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J Am Chem Soc* **128**, 12174-84 (2006).

12. Mercer, A. C.; Meier, J. L.; Hur, G. H.; Smith, A. R.; Burkart, M. D. Antibiotic evaluation and in vivo analysis of alkynyl Coenzyme A antimetabolites in Escherichia coli. *Bioorg Med Chem Lett* **18**, 5991-4 (2008).

13. Mercer, A. C.; Meier, J. L.; Torpey, J. W.; Burkart, M. D. In vivo modification of native carrier protein domains. *Chembiochem* **10**, 1091-100 (2009).

14. Meier, J. L.; Mercer, A. C.; Burkart, M. D. Fluorescent profiling of modular biosynthetic enzymes by complementary metabolic and activity based probes. *J Am Chem Soc* **130**, 5443-5 (2008).

15. Meier, J. L.; Barrows-Yano, T.; Foley, T. L.; Wike, C. L.; Burkart, M. D. The unusual macrocycle forming thioesterase of mycolactone. *Mol Biosyst* **4**, 663-71 (2008).

16. Helmkamp, G. M., Jr.; Brock, D. J.; Bloch, K. Beta-hydroxydecanoly thioester dehydrase. Specificity of substrates and acetylenic inhibitors. *J Biol Chem* **243**, 3229-31 (1968).

17. Schnarr, N. A.; Khosla, C. Trapping transient protein-protein interactions in polyketide biosynthesis. *ACS Chem Biol* **1**, 679-80 (2006).

18. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7 (2001).

19. Jessani, N.; Niessen, S.; Wei, B. Q.; Nicolau, M.; Humphrey, M.; Ji, Y.; Han, W.; Noh, D. Y.; Yates, J. R., 3rd; Jeffrey, S. S.; Cravatt, B. F. A streamlined platform for high-content functional proteomics of primary human specimens. *Nat Methods* **2**, 691-7 (2005).

20. Meier, J. L.; Burkart, M. D. Chapter 9. Synthetic probes for polyketide and nonribosomal peptide biosynthetic enzymes. *Methods Enzymol* **458**, 219-54 (2009).

Chapter Two

The chemical biology of modular biosynthetic enzymes

Originally published as:

Meier, J. L.; Burkart, M. D. *Chem Soc. Rev.* In press (2009).

CRITICAL REVIEW www.rsc.org/csr | Chemical Society Reviews

The chemical biology of modular biosynthetic enzymes†

Jordan L. Meier and Michael D. Burkart*

Received 4th December 2008
First published as an Advance Article on the web
DOI: 10.1039/b805115c

Fatty acid synthase (FAS), polyketide synthase (PKS), and nonribosomal peptide synthetase (NRPS) modular biosynthetic enzymes are responsible for the production of a multitude of structurally diverse and biologically important small molecule natural products. Traditional biochemical and genetic studies of these enzymes have contributed substantially to the understanding of their underlying biosynthetic mechanisms. More recently these investigations have been aided by the skillful application of a combination of chemical and biological techniques to aid in overcoming the unique challenges associated with the enzymology of these large multifunctional enzymes. This *critical review* provides a historical context and details studies (through July 2008) which aim to identify and characterize these enzymes using synthetically and/or chemoenzymatically generated small molecule probes.

Department of Chemistry and Biochemistry, University of California-San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0358, USA. E-mail: mburkart@ucsd.edu

† Abbreviations are used as follows: PKS, polyketide synthase, PK, polyketide, NRPS, nonribosomal peptide synthetase, NRP, nonribosomal peptide, FAS, fatty acid synthase, FA, fatty acid, NAC, *N*-acetylcystamine, CP, carrier protein, ACP, acyl carrier protein, PCP, peptidyl carrier protein, ArCP, aryl carrier protein, AT, acyltransferase domain, KS, ketosynthase domain, KR, ketoreductase domain, DH, dehydratase domain, ER, enoylreductase domain, TE, thioesterase domain, A, adenylation domain, C, condensation domain, AMP, adenosine monophosphate, E, epimerization domain, aa, amino acid (general), CoA, coenzyme A, DEBS, 6-deoxyerythronolide, Pik-TE, pikromycin thioesterase, Tyc-TE, tyrocidine thioesterase, MS, mass spectrometry, ESI-FTMS, electrospray ionization with ion cyclotron Fourier-transform mass spectrometry, FRET, fluorescence resonance energy transfer, DBP, 1,3-dibromopropanone, AMS, adenosine-5′-monosulfamate, 6-MSAS, 6-methylsalicylic acid polyketide synthase, FP, fluorophosphonate. Domain abbreviations are listed in the section 'Modular biosynthesis.' In NRPS nomenclature, PCP-bound aminoacyl or peptidyl chains are denoted as aa-S-NRPS when referring to a specific NRPS or aa-S-PCP in general to indicate the carrier protein–substrate thioester bond; polyketides and fatty acids bound to an acyl carrier protein are likewise denoted as acyl-S-ACP substrates.

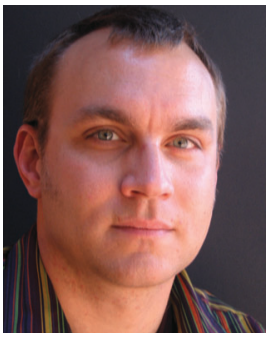
1. Introduction

The natural product biosynthetic pathways responsible for the production of fatty acids, polyketides, and nonribosomal peptides represent a fascinating intersection of primary and secondary metabolism.[1–3] In organisms ranging from microbes to humans these enzymes are used to produce small molecules which span the spectrum of biological activity from the vital, to the therapeutic, to the harmful (Fig. 1).

Remarkably, the unique structures characteristic of these metabolites are all biosynthesized in a similar manner. While the individual details of each pathway can be highly nuanced, lying at the center of each is a modular biosynthetic logic, which utilizes covalent acyl–enzyme intermediates, iterative cycles of condensation, and tailoring of monomer units to produce the structural and functional diversity associated with these classes of small molecules (for a more detailed description of modular biosynthesis see Section 2).

Jordan Meier

Jordan Meier was raised in Iowa Falls, Iowa as the youngest of three brothers. He received his BS in Chemistry from Creighton University, under the tutelage of Dr David Dobberpuhl and Dr Mark Kearley. In 2004 he entered the PhD program at the University of California-San Diego under the guidance of Prof. Michael Burkart. His current research interests include in vivo *carrier-protein labeling, functional proteomics approaches to the identification of PKS and NRPS biosynthetic systems, and the development of new tools for the structural analysis of megasynthetase enzymes.*



Michael Burkart

Michael Burkart grew up in Texas and received a BA in chemistry from Rice University in 1994. He went on to receive his PhD in Organic Chemistry from the Scripps Research Institute in 1999 with Prof. Chi-Huey Wong and pursued postdoctoral studies at Harvard Medical School with Prof. Christopher Walsh. In 2002 he accepted a position as Assistant Professor of Chemistry and Biochemistry at University of California-San Diego, and in 2008 he was promoted to Associate Professor. Prof. Burkart's research interests include natural product synthesis and biosynthesis, enzyme mechanism, and metabolic engineering. Prof. Burkart has been a fellow of the Alfred P. Sloan Foundation and an Ellison Medical Foundation New Scholar in Global Infectious Disease.

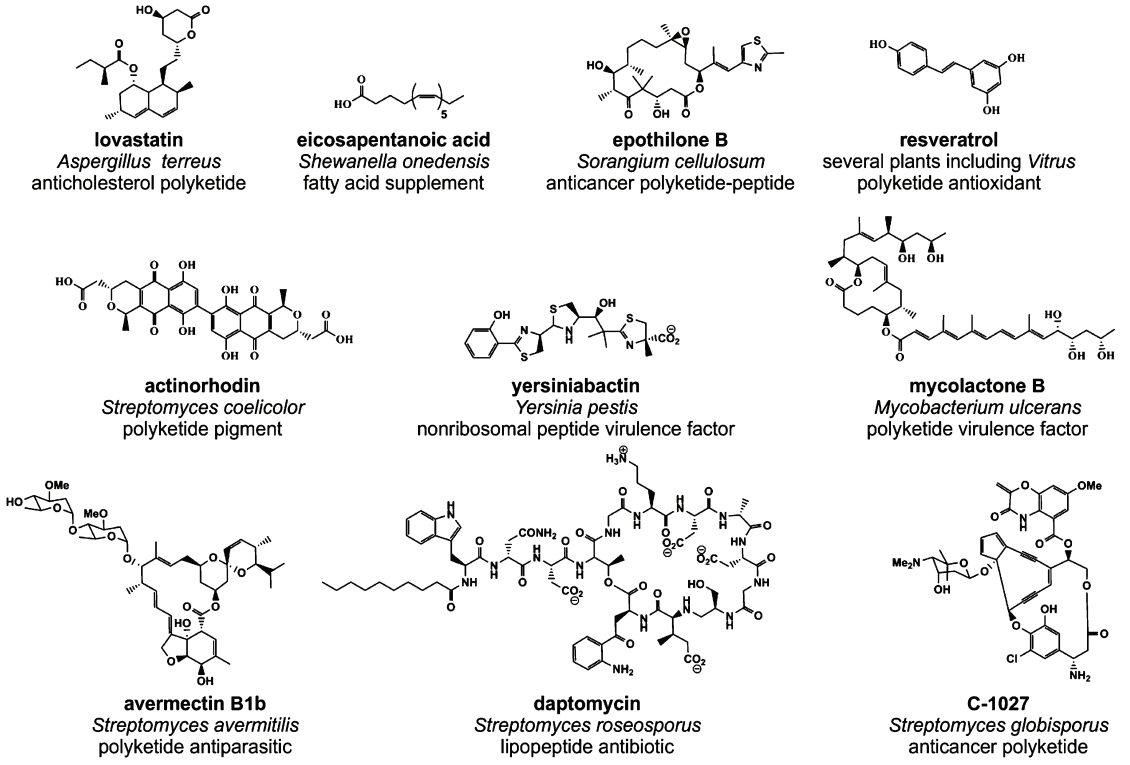


Fig. 1 Structures, producers, and biological activities of some notable fatty acid, polyketide, and nonribosomal peptide natural products.

The past 30 years have seen extraordinary breakthroughs in the biochemical, genetic, and structural understanding of these enzymes (Fig. 2).[4–6] These advances have had many benefits, from the validation of new antimicrobial and antiobesity drug targets, to the discovery of new anticancer agents, to the improvement of batch fermentation processes for industrial antibiotic production.[7,8] The uncovering of the modular nature of these enzymes has also led to hopes that genetic shuffling of the enzymatic domains responsible for PKS and NRPS biosynthesis might allow for the expression of new proteins capable of producing "unnatural" natural products, an approach dubbed combinatorial biosynthesis.[9] However, despite some notable exceptions, such strategies have been largely unsuccessful, with genetically engineered PKS and NRPS systems often showing reduced or negligible catalytic activity compared with natural systems.[10]

In spite of these initial setbacks, the huge potential of combinatorial biosynthetic methods has led researchers to

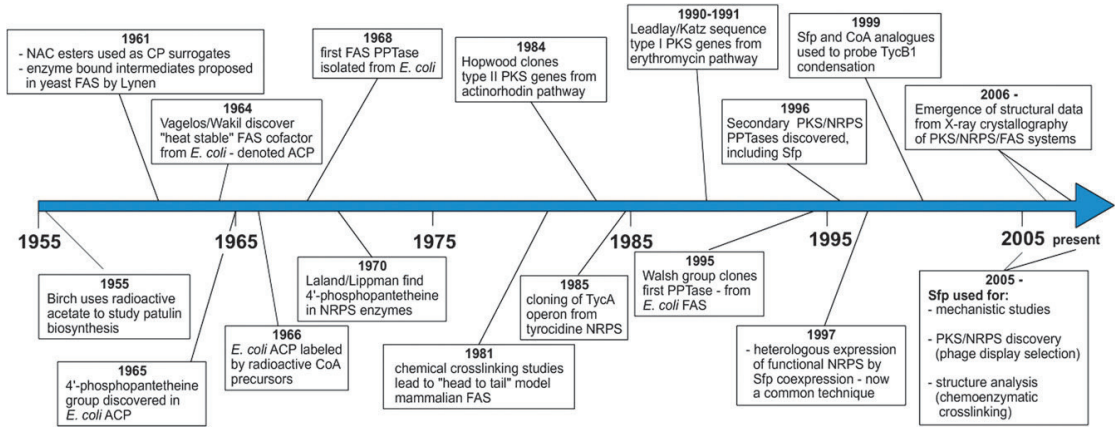


Fig. 2 Timeline of some notable events in the study of carrier-protein-mediated FAS, PKS, and NRPS biosynthetic enzymes.

continue to seek a more complete understanding of the enzymatic chemistry of these systems for the ultimate goal of their efficient reengineering. Most recently, chemical biological approaches have moved to the forefront of methods for the evaluation of FAS, PKS, and NRPS systems, allowing scientists to study these enzymes in new and creative ways. A recently defined (though by no means young) field, chemical biology refers to the application of chemistry to the study of molecular events in biological systems.[11] Chemical biology shares the common goals of traditional biology and biochemistry of seeking to better understand living systems. The element that distinguishes chemical biology from these disciplines is a focus on the design, synthesis, and application of small molecule probes which report on a particular functional aspect of that biological system. It also differs from these fields in that its scope can range broadly from a whole organism, to a whole proteome, to a single enzyme, to a single nucleotide sequence.

Chemical methods played a vital role in the study of natural product biosynthesis long before the term ''chemical biology'' was coined. It was the availability of isotopically labeled acetate which first allowed confirmation of the acetate-derived nature of polyketides by Birch *et al.*, while Stoops and Wakil used a thiol-reactive crosslinking electrophile to probe the domain organization of the analogous fatty acid synthases.[12,13] Similarly, Lynen and then Bloch *et al.* demonstrated the value of *N*-acetyl-cystamine (NAC) thioesters as probes of the individual steps of saturated and unsaturated fatty acid biosynthesis, techniques which have since been successfully adopted by modern researchers to study PKS- and NRPS-mediated processes.[14,15] Any and all of these studies might now be termed chemical biology (Fig. 2).

It is from this perspective that we review efforts in the field which have utilized cutting-edge chemical biological approaches towards four main goals: as (i) probes of mechanism, (ii) selectivity, (iii) structure in FAS, PKS, and NRPS enzyme systems, and (iv) for the identification of new PKS and NRPS systems. We have placed a particular emphasis on methods which use carrier-protein posttranslational modification as a tool for gaining a more complete understanding of these systems. To facilitate this discussion we will provide a brief introduction to the canonical PKS–NRPS–FAS modular biosynthetic paradigm, followed by a historical review of how studies of the carrier protein and its posttranslational modification led to the development of what is now a key tool for the study of these modular biosynthetic enzymes. Finally, after detailing some representative studies in these areas of research, we will conclude with a short perspective on future prospects for this exciting field.

2. Modular biosynthesis

The modular biosynthetic logic utilized by FAS, PKS, and NRPS biosynthetic enzymes is only briefly introduced here for the purpose of facilitating later discussion; for a more detailed description of the underlying biosynthetic mechanisms of these systems the reader is directed to several recent reviews.[1–3]

FAS, PKS, and NRPS enzymes produce their characteristic small molecule metabolites by the loading and condensation of small monomer units. In FAS and PKS systems these monomer units are acyl-CoAs, and structural diversity of PKS compared with FAS natural products results mainly from varying degrees of acyl chain reduction.[16] The discrete enzymes responsible for each of the individual loading, condensation, reducing, and chain-termination steps are referred to as domains. When these catalytic domains are housed on a single massive polypeptide (often >200 kDa in size) the PKS or FAS is termed a type I system, as in the human FAS or 6-deoxyerythronolide (DEBS) PKS systems. These domains can also each be housed on their own discrete polypeptide and function together in *trans* to produce their corresponding natural product in what are termed type II systems, such as those used in the bacterial FAS or actinorhodin polyketide synthase (Fig. 3b).[17]

The loading of acyl monomer units is accomplished in polyketide and fatty acid biosynthesis by the action of acyltransferase (AT) domains. These domains use a nucleophilic serine to selectively attack the thioester of an appropriate acyl-CoA monomer (usually malonyl or methylmalonyl CoA) to form a covalent acyl-O-enzyme intermediate. This intermediate is then transferred to the terminal thiol of a 4′-phosphopantetheine prosthetic group of an acyl carrier protein (ACP) domain.‡ These ACPs are the site of covalent attachment of activated monomers and growing acyl chains throughout the cycles of condensation and reduction during PKS and FAS biosynthesis (Fig. 3c). Condensation is mediated by the ketosynthase (KS) domain, which utilizes a nucleophilic thiol to translocate the growing acyl chain to its active site, before catalyzing decarboxylation of a downstream malonyl-S-ACP which then captures the KS-bound acyl-enzyme through a Claisen condensation. The resulting elongated ketide chain can then be reduced by ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains. In the case of FAS all of these activities work in concert to produce the fully reduced saturated hydrocarbon, while in PKS enzymes some of these reductive activities may be inactive or missing to produce olefin, alcohol, or ketone containing natural products (Fig. 3b and 3c). Finally a thioesterase (TE) domain mediates release of the finished fatty acid or polyketide chain from the ACP domain. Like the AT, thioesterase domains utilize a catalytic serine to form an acyl-O-enzyme intermediate which is then either captured by water to afford the hydrolyzed acyl chain (as in FAS) or an intramolecular nucleophile (such as a downstream hydroxyl group) to produce the macrocyclic structures often observed in PKS natural products (Fig. 3c).

NRPS enzymes introduce functional group diversity in their products by utilizing a much larger monomer pool, including all 20 naturally occurring amino acids as well as a number of unnatural amino and aryl acid substrates.[3] These monomer units are activated by the action of ATP-dependent adenylation (A) domains, which activate these acids as aminoacyl–AMP mixed anhydrides that are in turn captured by the pantetheine thiol of peptidyl or aryl carrier protein (PCP or ArCP) domains. As in PKS and FAS enzymes, the carrier

‡ Carrier protein (CP) domains have been alternately denoted in many studies as thiolation (*T*) domains. In the present manuscript we use the original term as designated by Vagelos and Wakil.

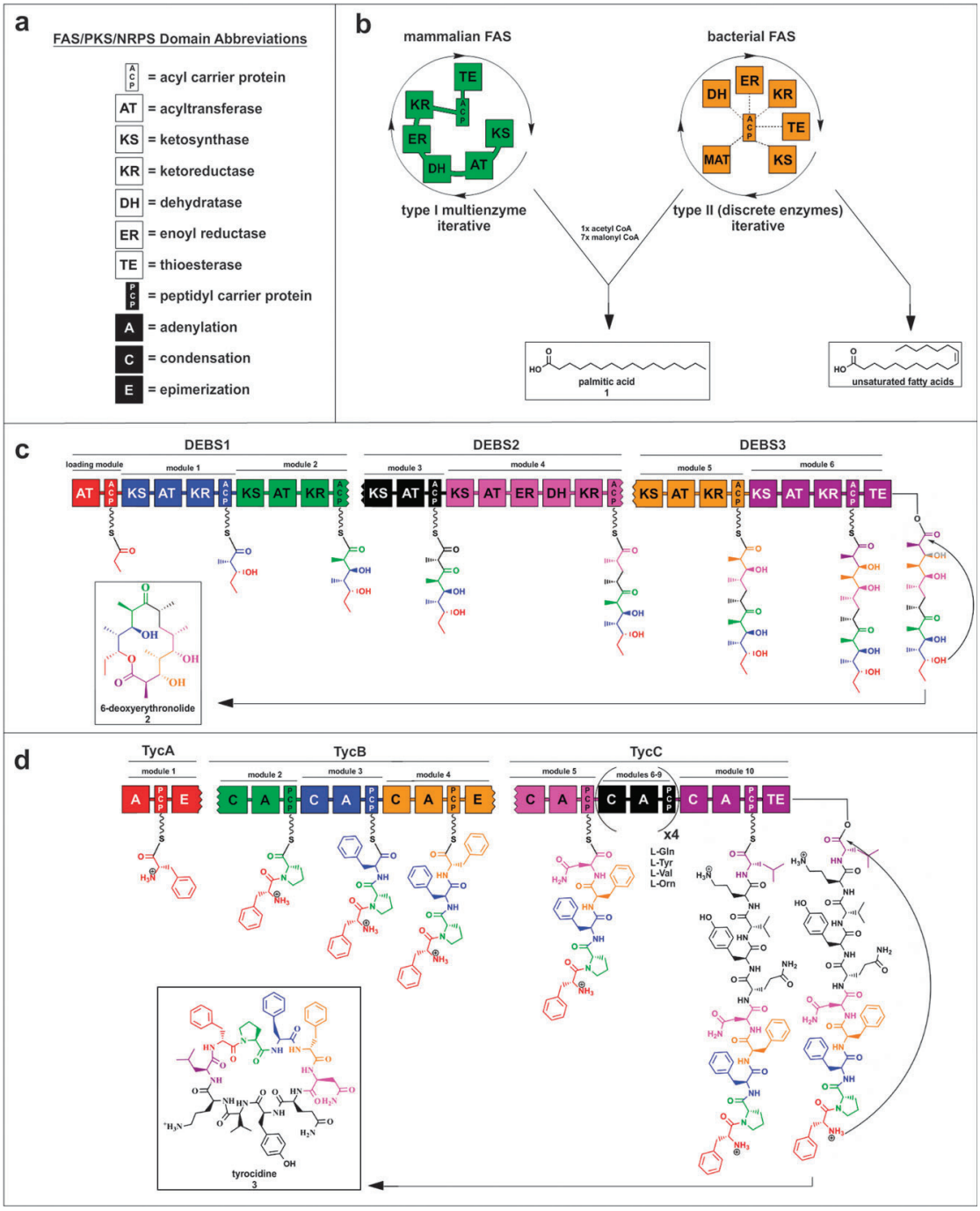


Fig. 3 Domain organization for FAS, PKS, and NRPS enzymes. (a) Domain abbreviations used in this manuscript. (b) Mammalian and bacterial FAS. Both produce the primary metabolite palmitic acid (**1**), but while mammalian FAS houses each necessary enzymatic activity on a single multienzyme, in bacteria they function in *trans*. (c) Type I PKS responsible for 6-deoxyerythronolide (**2**) production. Monomer units are color coded to indicate module of origin. Jagged ACP–KS junctions indicate intermodular communication. (d) Type I NRPS responsible for tyrocidine (**3**) production.

protein (CP) is the point of attachment for these activated aminoacyl monomers and the growing peptidyl chain throughout biosynthesis; however, NRPS multienzymes differ in that PCPS are the only site of covalent attachment to the enzyme during elongation, as neither the activation nor condensation reactions utilize acyl–enzyme intermediates. After PCP loading, dedicated epimerization (E) domains may act on the aminoacyl-S-PCP or peptidyl-S-PCP to introduce D-amino acids into the final nonribosomal peptide product. The condensation reaction itself is performed by the appropriately named condensation (C) domain, which mediates amide bond formation by catalyzing attack of the α-amino group of the downstream aminoacyl-S-PCP on the thioester of the upstream PCP-bound peptide. During cycles of elongation the growing carrier-protein tethered peptidyl chain may also be modified by the action of methylation (M) or reductase (R) domains to introduce additional functionality to the final product. Finally, chain termination is catalyzed by a TE domain in a mechanism similar to that used by PKS enzymes, resulting in release of the straight chain or macrocyclic peptides associated with nonribosomal peptides (Fig. 3d).

3. Carrier-protein-mediated biosynthesis and posttranslational modification

The biosynthetic cycles of polyketide, nonribosomal peptide, and fatty acid biosynthesis are highly dependent on the ability of CP domains to interact with a host of different enzymatic partners in order to efficiently process biosynthetic intermediates. These biosynthetic intermediates are covalently tethered to the CP through a thioester bond formed with the terminal thiol of the 4′-phosphopantetheine prosthetic group. Many of the modern chemical biological approaches detailed in this manuscript utilize manipulation of this prosthetic group to probe FAS, PKS, and NRPS enzymatic machinery in interesting ways. Recalling that there was virtually no knowledge of these classes of enzymes 50 years ago, we must admire the vision of the scientists involved in the initial studies and mechanistic elucidation of carrier-protein-mediated biosynthesis, on whose shoulders the current field stands. In the following section we focus on the unique posttranslational modification utilized by carrier proteins, detailing the history of how it was discovered and the mechanisms by which it is introduced and regulated in modular biosynthetic enzymes, to better set the stage for a discussion of its most recent applications in studying FAS, PKS, and NRPS enzymes. For a more comprehensive look at the historical development of our understanding of fatty acid, polyketide, and nonribosomal peptide biosynthesis, the reader is directed to several superb reviews.[1,18–20]

3.1 Discovery of the carrier protein substrate binding site

In 1961, while attempting to study the enzymes responsible for fatty acid biosynthesis in cell-free yeast extract, Lynen discovered that no low molecular weight intermediates of the biosynthetic process could be detected within the extract.[14] This puzzling observation, followed by a series of elegant experiments which utilized model NAC substrates and thiol-specific affinity agents, led Lynen to propose that the intermediate fatty acyl chains were covalently bound to the FAS through a thioester bond. The site of attachment of these intermediates was tracked to two separate sites on the polypeptide, denoted as the ''central'' and ''peripheral'' thiols of the FAS.[21]

The large multienzymatic nature of the type I yeast FAS precluded further characterization of this system. However, at the same time several groups—most notably those of Vagelos and Bloch—were making headway in the same field by examining crude *Escherichia coli* enzyme fractions capable of catalyzing fatty acid biosynthesis.[22,23] Study of the *E. coli* FAS had many advantages, key among them the freely dissociable type II nature of the system, which allowed the study of each discrete enzymatic activity as a different protein. In the course of fractionating and studying the reaction sequence catalyzed by these enzymes, the groups of Vagelos and Wakil each independently discovered a highly acidic, heat-stable protein which was primarily responsible for the binding of acyl-intermediates through a thioester linkage (Fig. 2). Given the name acyl carrier protein (ACP), this protein was found to function not catalytically but rather as a *cofactor*, acting as an acyl-group acceptor and donor during the iterative cycles of loading, condensation, and reduction involved in fatty acid biosynthesis.[24,25]

The identity of the ACP thiol involved in the formation of these acyl–thioester linkages was a topic of intense research—after first being misidentified as cysteine, a number of chemical and enzymatic degradation studies followed by amino acid determination were performed on ACP.[26] The results definitively demonstrated the substrate tethering group to be the terminal thiol of 4′-phosphopantetheine, bound to ACP through a phosphodiester linkage to a serine residue.[27] This finding was replicated during studies of the type I mammalian FAS.[28] The finding that ACP utilized an identical prosthetic group to that of coenzyme A (CoA) was curious, as previous studies had shown CoA-thioester compounds were poor substrates for ACP-partner proteins. However, the identity of the prosthetic group raised a provocative biochemical hypothesis as to its origin and incorporation into ACP, one Vagelos would investigate shortly.

Around the same time as this revolution in understanding the enzymology underlying fatty acid biosynthesis was coming about, there was increasing evidence that PKS natural products might share a similar province. In 1955, Birch *et al.* demonstrated that the fungal natural product 6-methylsalicylic acid arose from head to tail condensation of acetate units through radiolabeling experiments.[12] However studies of the actual enzymes responsible for polyketide biosynthesis lagged far behind the corresponding studies of fatty acid biosynthesis, mostly due to a lack of reliable techniques for the expression and recovery of these enzymes from crude cell lysate. Nevertheless, Lynen and colleagues isolated the 6-methylsalicylic acid synthase (6-MSAS) from *Penicillum patulum* and used affinity labeling and small-molecule product analysis to postulate a biosynthetic mechanism similar to that used in fatty acid biosynthesis, including the covalent linkage of substrates and intermediates.[29,30] Further characterization of PKS reaction mechanisms would remain elusive until the advent of modern genetics and molecular biology in the 1980s.[31]

Surprisingly, the discovery of covalent channeling of intermediates in fatty acid biosynthesis also had an impact on the biosynthetic hypotheses for a family of cyclic peptides seemingly structurally unrelated to the acetate-based polyketides and fatty acids.[20] Christened "nonribosomal peptides" due to the insensitivity of their production to known inhibitors of ribosomal protein synthesis,[32] very little was understood about the biosynthetic origins of these compounds before 1968, when a series of seminal studies were performed in the laboratory of Lipmann at the Rockefeller University. Using activity-based fractionation, Lipmann and co-workers isolated crude preparations of the enzymes responsible for the production of the cyclic peptide antibiotic gramicidin S.[33] By administration of radiolabeled amino acids and chemical cleavage studies it was quickly discovered that these enzymes utilized covalent aminoacyl–thioester intermediates as intermediates in the formation of nonribosomal peptides.[34,35] As this finding came only a few years after the discovery of ACP and its use of the 4′-phosphopantetheine prosthetic group to form high-energy thioester intermediates, it was logical to explore whether NRPS enzymes utilized a similar mechanism. Indeed Laland *et al*., who had made findings analogous to those of Lipmann *et al*. concerning the enzyme-bound nature of gramicidin biosynthetic intermediates, would observe pantothenic acid from gramicidin S synthetase preparations in 1970.[36,37] Lipmann *et al*. would confirm this finding only months later, this time using a cell-free extract capable of tyrocidine biosynthesis.[38,39] These findings led to the proposal of a thiotemplate model of nonribosomal peptide biosynthesis, in which activated aminoacyl–adenylates were transthioesterified to a single 4′-phosphopantetheine modified carrier protein and underwent multiple cycles of loading, transthioesterification, and condensation to produce the final product.[40] This initial understanding would provide the foundation for our current understanding of NRPS biosynthesis (the so-called "multiple carrier" model).[41] However, the exact mechanism of 4′-phosphopantetheine incorporation into NRPS producing enzymes would remain a mystery for the next 30 years (Fig. 2).

3.2 The PPTase reaction: discovery of primary and secondary PPTases

As mentioned in brief, the discovery of the use of 4′-phosphopantetheine by ACP as the acyl-group carrier in *E. coli* fatty acid biosynthesis immediately suggested the role of CoA as a biosynthetic precursor to ACP. This was established by Alberts and Vagelos, who demonstrated that mutant strains of *E. coli* which incorporated radiolabeled CoA precursors into their intracellular CoA pool similarly transferred this radiolabel to the ACP prosthetic group.[42] This method of metabolic radiolabeling quickly became a powerful tool for establishing 4′-phosphopantetheine content in type I FAS and NRPS enzymes as well.[39,43]

Vagelos further characterized the process by which non-4′-phosphopantetheine containing *apo*-ACP (so called because of its inability to serve as an acyl-group carrier) was converted to *holo*-ACP by purifying a crude preparation of the enzyme responsible for this transfer.[44] This phosphopantetheinyl-transferase (PPTase) from *E. coli* was found to catalyze

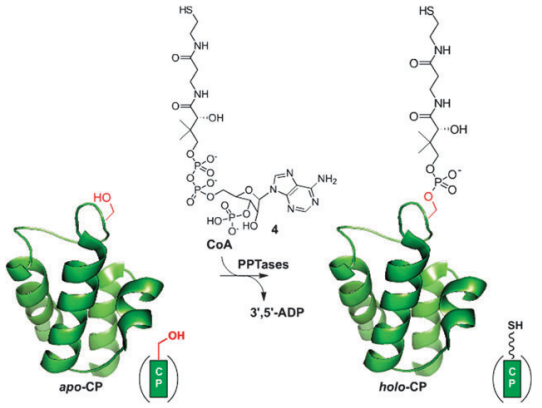


Fig. 4 Carrier protein posttranslational modification by coenzyme A (**4**). Carrier protein is pictured as a helical bundle, with cartoon representation in parentheses. A PPtase mediates transfer of the 4′-phosphopantetheine prosthetic group to a conserved serine residue of the carrier protein (pictured in red), converting it from *apo* to *holo*-form. The terminal thiol can then be used to tether intermediates throughout FAS, PKS, and NRPS biosynthesis (see Fig. 3).

transfer of the 4′-phosphopantetheine arm to *apo*-ACP in a CoA and Mg^{2+}-dependent manner (Fig. 4). Studies showed that dephospho-CoA and oxidized CoA were not substrates for the crude *E. coli* PPTase, and activity was lost upon boiling.

After this initial study, interest in PPTases waned through the 1980s, even as there was a growing awareness of the role played by *holo*-carrier proteins in PKS and NRPS biosynthesis. It was easy to take 4′-phosphopantetheinylation for granted at this time since the limited number of systems studied (type I and II FAS, gramicidin and tyrocidine, 6-MSAS) were obtained as crude cell-free extracts purified directly from the producer organisms, whose endogenous PPTases ensured their efficient posttranslational modification. This all changed as the use of gene cloning technology became routine and molecular biology began to overtake classical biochemical fractionation approaches to PKS and NRPS study. Researchers quickly found that recombinantly expressed PKS and NRPS enzymes were often, or in the case of NRPS enzymes always, produced as *apo*-carrier protein containing synthases.[45] This greatly hindered the study of the reaction mechanisms of these enzymes, and made finding a way to produce *holo*-synthase imperative to the field. This task was undertaken by the laboratory of Walsh at Harvard. Using an isolation strategy based on the known strong interaction between the *E. coli* FAS PPTase and *apo*-ACP, they purified the PPTase enzyme to homogeneity and subjected it to N-terminal sequencing. This allowed identification, cloning, and overexpression of the gene responsible for its production, *acps* (*holo*-acyl carrier protein synthase), encoding the first characterized PPTase.[46] The enzyme was found to efficiently phosphopantetheinylate *E. coli* FAS ACP (as previously shown) as well as a number of *apo*-ACPs from reduced polyketide biosynthetic pathways when incubated with CoA *in vitro*.[47,48] This PPTase was not found to be active toward

NRPS carrier proteins, explaining their poor posttranslational modification during heterologous expression.[49]

Armed with knowledge of the *E. coli* PPTase sequence, researchers involved in the NRPS field began to search for homologues of AcpS which might be responsible for production of *holo*-PCP in these systems. This was accomplished through a collaborative effort when researchers from the groups of Walsh, Khosla, and Marahiel identified a series of proteins in the NRPS producers *E. coli*, *Bacillus subtilis*, and *Bacillus brevis* which showed marginal homology to AcpS.[50] Previous studies had shown these genes, from the enterobactin, gramicidin, and surfactin biosynthetic pathways, were necessary for nonribosomal peptide production in these organisms, although their function was unknown. Recombinant expression and purification demonstrated unambiguously their ability to transfer 4′-phosphopantetheine to *apo*-PCPs *in vitro*.[50] Soon after this it was shown that coexpression of the gramicidin S NRPS module, GrsA, with its corresponding PPTase allowed production of active *holo*-GrsA in large amounts.[51] Additional research on the PPTase from the surfactin synthetase gene cluster, Sfp, would universalize this strategy for production of recombinant *holo*-carrier proteins, by demonstrating this enzyme was capable of modifying carrier protein domains from FAS as well as NRPS enzymes *in vitro*.[52] Heterologous coexpression of Sfp with FAS, NRPS, and PKS carrier proteins results in efficient production of *holo*-carrier proteins, a strategy which is now commonly used for enzymological studies of carrier protein utilizing systems.

3.3 Carrier proteins and PPTases: the current state of the art

Carrier proteins are now known to be active in the tethering of intermediates during biosynthesis of the aforementioned fatty acids, polyketides, and nonribosomal peptides, as well as a number of other secondary metabolites involved in processes such as intercellular signaling and bacterial cell wall biosynthesis.[53] A survey of carrier proteins identified to date shows they are characterized by their small size ($\sim$75 residues) and high number of negatively charged surface residues. NMR solution structures as well as a limited number of crystallographic studies have shown that carrier proteins adopt a helical bundle fold which composes three major alpha helices as well as a fourth minor helix (Fig. 5).[54–57] The serine which acts as the site of 4′-phosphopantetheinylation is found in a loop near the beginning of helix II, most often within a conserved DSX motif (X being lysine or arginine). This three-dimensional motif in combination with electrostatic interactions mediates protein–protein interactions between carrier proteins and the multiple enzymatic partners they must interact with during fatty acid, polyketide, and nonribosomal peptide biosynthesis.[53,58]

Although all carrier proteins share these general characteristics, significant differences are evident between biosynthetic systems. Fatty acid ACPs are extremely abundant relative to PKS and NRPS CPs and are estimated to constitute $\sim$0.25% of total cellular protein in *E. coli* (the prototypical bacterial system).[59] Structurally it has been shown that type II fatty acid and polyketide carrier proteins, which interact with their enzymatic partners in *trans*, usually carry a strongly negative charge, particularly in the long loop between helix I and II, helping to mediate interaction with partner proteins *via* electrostatic interactions. NMR studies and more recently obtained crystal structures of type II ACPs with bound fatty acyl substrates suggest that these carrier proteins also share a common mode of protecting these intermediates through sequestration within a hydrophobic core (Fig. 5a and b).[60] In contrast, NMR analyses of type I PKS and NRPS carrier proteins reveal a more neutral charge on their protein surface, supporting a reduced necessity for guidance to their enzymatic reaction partners by large electrostatic interactions due to their in *cis* positioning in multimodular synthases.[56,61,62] In the crystal structures of type I FAS, PKS, and NRPS enzymes published to date, regions corresponding to the carrier protein have been resolved only in the fungal FAS.[63–67] This study revealed a structural basis for a similar substrate sequestration mechanism as is used by type II FAS carrier proteins, and postulated the ACP transitions between substrate sequestration and substrate processing by movement of the 4′-phosphopantetheine arm *via* a switchblade-like mechanism (Fig. 5c and d).[68]

Since their initial discovery, a large number of PPTases involved in fatty acid, polyketide, and nonribosomal peptide biosynthesis have been identified from a variety of organisms.[69–71] Currently PPTases are classified into three main classes: type I PPTases resemble *E. coli* AcpS and exhibit a

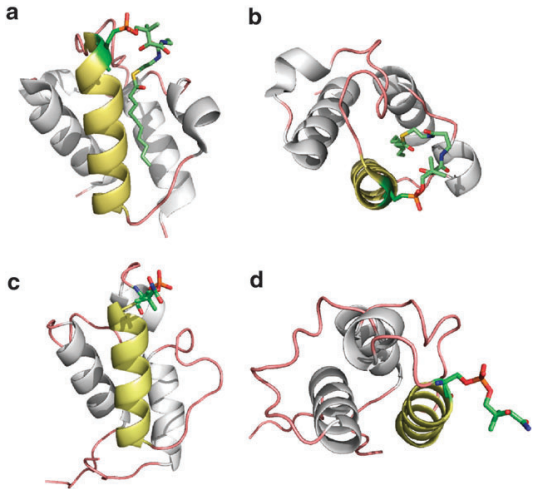


Fig. 5 Crystal structures of fatty acid ACPs from bacterial and fungal systems. (a) Crystal structure of *E. coli* ACP with decanoate bound through thioester linkage to the 4′-phosphopantetheine prosthetic group (PDB:2FAE). The canonical three major helices are observed, with helix II and the conserved Ser 36 colored gold and green respectively. (b) View of "a" rotated 90° toward reader. The bound decanoate moiety is sequestered in a hydrophobic core. (c) Fungal ACP, excised from the crystal structure of the stalled yeast FAS (PDB:2UV8). (d) The pantoic acid portion of the 4′-phosphopantetheine group is observed extended for interaction with the KS domain, in contrast to "b", leading the authors to propose a switchblade mechanism of prosthetic group action in fatty acid biosynthesis.

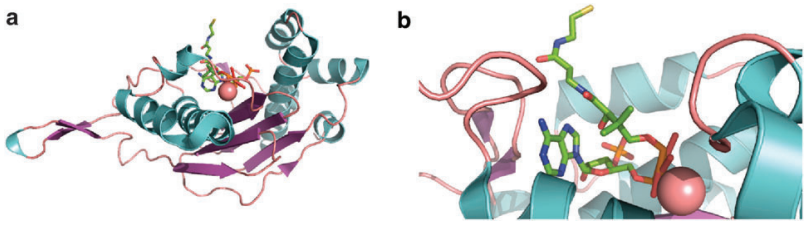
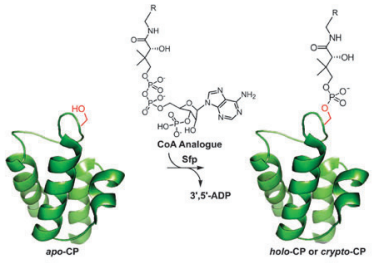


Fig. 6 Crystal structure of Sfp, the promiscuous PPTase from the *B. subtilis* surfactin biosynthetic pathway (PDB:1QR0). (a) Structure of Sfp with CoA substrate and Mg^{2+} cofactor bound in active site. (b) Close-up of CoA bound to Sfp. The terminal β-alanine and cystamine portions of CoA extend into the solvent and make no clear interactions with the enzyme, helping explain the permissivity of Sfp for CoA analogues modified at the thiol-terminus.

strong preference towards modification of carrier proteins from primary metabolic pathways such as fatty acid biosynthesis, possessing little reactivity with PKS and NRPS carrier proteins.[46] Type II PPTases are exemplified by the *B. subtilis* PPTase Sfp, and are often referred to as secondary PPTases due to their *in vivo* role in the modification of CPs involved in polyketide and nonribosomal peptide biosynthesis. In addition to modifying PKS and NRPS carrier proteins, these PPTases possess a relaxed substrate specificity which permits post-translational modification of FAS carrier proteins, both when applied *in vitro* as well as during heterologous coexpression *in vivo*.[72] A subclass of type II PPTase is represented by PcpS found in *Pseudomonas aeruginosa*. This organism does not harbor genes for a type I PPTase and instead utilizes PcpS for the modification of all carrier-protein-dependent biosynthetic pathways.[73] Finally type III PPTases are found integrated within the multimodular fungal and yeast FAS. These enzymes have not been the subject of extensive characterization as they have been found to lose their activity when excised from the FAS complex.[74]

By far the most well-known and useful PPTase from a biotechnology standpoint is the aforementioned Sfp.[50,52] Due to its relaxed specificity in terms of carrier-protein identity, Sfp is now commonly coexpressed with recombinant PKS and NRPS enzymes, co-opting the endogenous CoA of an overexpression host such as *E. coli* for the production of active 4′-phosphopantetheinylated enzyme. More recently researchers have focused on utilizing another characteristic of Sfp, namely its remarkable permissivity for the identity of the CoA analogue whose 4′-phosphopantetheine arm is transferred to the carrier protein. Hinted at in initial studies of Sfp which showed that acyl, aromatic, and non-thioester containing CoA analogues were efficiently transferred to *apo*-carrier proteins,[48] it has now been shown that Sfp is capable of transferring virtually any CoA analogue to any *apo*-carrier protein *in vitro*. Some notable examples of diverse CoA analogues transferred to carrier proteins using Sfp are pictured in Fig. 6.[75–78] The structural basis for this substrate promiscuity is revealed by the crystal structure of Sfp in complex with CoA.[79] While the phosphate groups of CoA make specific contacts with Sfp and the Mg^{2+} cofactor, explaining the absolute requirement for 3′-phosphorylation of the ribose sugar, no defined interactions were observed for the terminal thiol-containing portion of the CoA molecule, which extends out of the CoA binding pocket into the solvent (Fig. 7). To date there has not been a CoA analogue reported in the literature which Sfp is incapable of accommodating, and it is interesting to speculate on the potential limits of this amazingly permissive enzyme. What is known is that Sfp represents one of the trump cards in the toolbox assembled to study fatty acid, polyketide, and nonribosomal peptide biosynthesis, one chemical biologists have been extremely creative in playing.



CoA Analogue (R =)	Mol. weight	Ref.
5	727	46
6	727	48
7	623	75
8	1948	76
9	1425	77
10	1810	78
11	1338	75

Fig. 7 Notable examples of CoA analogues transferred to carrier proteins using Sfp.

4. Probes of PKS and NRPS mechanism

Prior to the molecular biology revolution, the researcher's options for probing the different activities of multidomain FAS, PKS, and NRPS enzymes were few. To test the activity of an acyl carrier-protein-loaded substrate involved either chemical acylation of the carrier protein (which runs a high risk of off-site reactivity and enzyme inactivation) or use of small-molecule carrier-protein surrogates such as NAC, pantetheine, or CoA thioesters.[80–82] Even though such methods continue to play an important role in the mechanistic analysis of modular biosynthetic enzymes, there has been an awareness of the shortcomings of such strategies since the beginning. It has been noted elsewhere that one of the fathers of the FAS field, Feodor Lynen, had a very dim view as to the ability of NAC thioesters to serve as replacements for more complex structures in both the oxidative and reductive fatty acid biosynthetic pathways. Their surprising success led him to say, "Be naïve and try the experiment even if its chances of success *are very small*."[83]

In the field of PKS and NRPS biosynthesis, the newfound availability of recombinant proteins and the ability to load essentially any *apo*-carrier protein with a mechanistically relevant coenzyme A analogue using Sfp has perhaps increased the researchers chances of success, although the discovery of more complex and detailed biosynthetic pathways reminds us that we remain as naïve as ever in some respects. Nevertheless, the power of carrier-protein modification studies in probing several different mechanistic aspects of PKS and NRPS biosynthesis should not be understated. The promiscuous PPTase Sfp is now almost universally applied for a variety of purposes, from the relatively simple posttranslational modification of recombinant carrier proteins to their *holo*-form, to the loading of putative biosynthetic intermediates onto their cognate carrier protein, to the examination of the substrate selectivity of PKS and NRPS enzymes for their potential use in combinatorial biosynthesis.[84–86] Here we focus on a few representative examples which utilize carrier protein 4′-phosphopantetheinylation to probe the mechanisms used by PKS and NRPS condensation and thioesterase domains, as well as examining its more recent use in studies which screen the loading specificity of PKS and NRPS modules using high power mass spectrometry techniques.

4.1 Probes of NRPS condensation mechanism

The ability to load carrier proteins with analogues modified at the thiol-terminus of CoA was first applied to probe the substrate specificity of the condensation domain of the NRPS responsible for tyrocidine biosynthesis in the gram-positive bacterium *Bacillus brevis*.[38,87] TycA (A/PCP/E) and TycB1 (C/A/PCP) are known to catalyze loading, epimerization, and condensation of L-Phe and L-Pro by a canonical NRPS reaction mechanism to form the carrier protein bound dipeptide D-Phe-L-Pro-S-TycB1, an intermediate which undergoes further cycles of elongation and cyclization to form tyrocidine.

Taking advantage of the fact that the donor and acceptor PCPs for this condensation reaction are housed on different peptides, and the ability of the gramicidin initiation module GrsA to substitute for TycA *in vitro*,[88] Belshaw and co-workers used Sfp in combination with aminoacyl-CoA analogues to bypass the adenylation domain specificity of these enzymes and load *apo*-GrsA and *apo*-TycB1 with a number of different aminoacyl 4′-phosphopantetheine arms (Fig. 8). The ability of the TycB1 C-domain was then tested for its ability to catalyze their condensation. They found that when the donor site, GrsA, was loaded with five different aminoacyl CoAs, TycB1 loaded with the natural substrate L-Pro capably processed them all to TycB1-bound dipeptides. Further analysis of this reaction by chiral HPLC showed in the case of L-Phe-S-CoA and L-Ala-S-CoA loading of GrsA, epimerization of the carrier-protein-loaded amino acid preceded condensation, resulting in formation of the DPhe-L-Pro dipeptide over the L-Phe-L-Pro product in a 9 : 1 ratio. In sharp contrast to the broad range of GrsA donor site substrates accepted by the TycB1 C-domain, when TycA was loaded with the natural condensation substrate DPhe and the amino acid loaded at the TycB1 acceptor site was varied from the natural L-Pro,

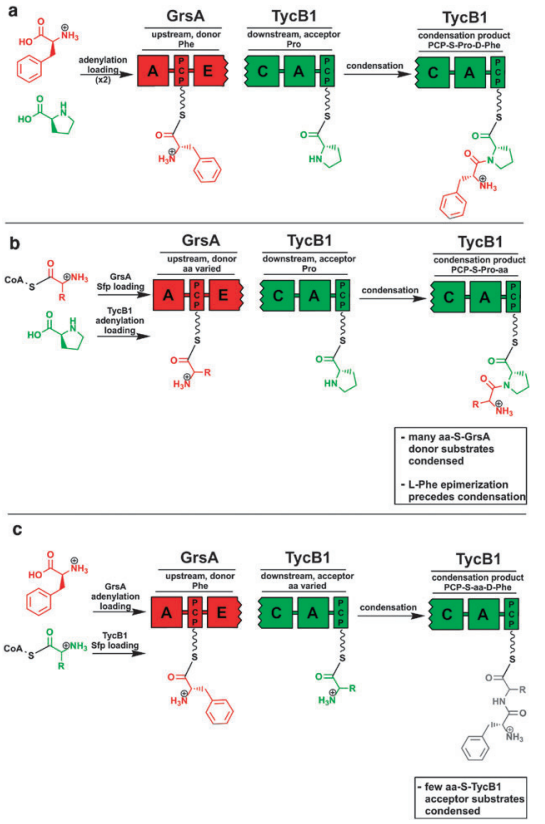


Fig. 8 Aminoacyl CoAs as probes of NRPS condensation mechanism. (a) L-Phe and L-Pro, the natural substrates for GrsA and TycB1, are loaded by A domains, epimerized, and condensed to form a PCP-L-Pro-DPhe dipeptide. (b) GrsA is loaded with a variety of aminoacyl-CoA analogues using Sfp, several of which condense with L-Pro-S-TycB1, implying selectivity is low for the aminoacyl-S-PCP donor. (c) TycB1 is loaded with a variety of aminoacyl CoA analogues using Sfp, few of which are condensed, implying a high selectivity for the aminoacyl-S-PCP acceptor in NRPS biosynthesis.

condensation was observed only with L-Ala (out of five amino acids tested). The stringent substrate specificity and C-domain binding of downstream, single amino acid loaded acceptor domains of an NRPS has been rationalized in terms of the larger context of nonribosomal peptide biosynthesis, during which C-domains must accommodate upstream carrier proteins loaded with aminoacyl polymers of varied size and steric hindrance. By providing evidence that in initiation modules, epimerization precedes condensation, and that the condensation domains confer fidelity to the growing natural product by being highly selective for the correctly loaded aminoacyl-S-PCP acceptor, this study provided some of the first insights into the timing and coordination of nonribosomal peptide biosynthesis.[87] Traditional studies have since confirmed both of these findings are broadly applicable to the canonical enzymes of the NRPS class as a whole, a testament to the information-rich nature of carrier-protein-based analyses.[89,90]

To extend this methodology, and gain insight into the timing of epimerization and condensation in further downstream NRPS elongation modules, the groups of Walsh and Marahiel later used Sfp to load an internal modulae from the tyrocidine NRPS, TycB3 (A/PCP/E) with tetrapeptidyl CoA analogues.[91] The downstream TycC1 (C/A/PCP) module was then auto-aminoacylated with its natural substrate L-Asn and tested for its ability to accept TycB3 loaded tetrapeptides in the condensation reaction in a single turnover fashion. It was found that L-Asp-S-TycC1 promoted condensation when TycB3 was loaded with tetrapeptides containing either the epimerized DPhe adjacent to the C-terminal thioester position, or the analogous L-Phe containing tetrapeptide substrate. However, only the DPhe containing pentapeptide was recovered in both cases, suggesting that the epimerization reaction at the peptidyl-S-TycB3 stage proceeded before TycC1-mediated condensation (Fig. 9). This was consistent with the proposed timing of the epimerization reaction in internal elongation modules, which evidence had suggested to occur at the peptidyl-S-PCP stage in internal elongation modules of NRPS.[89,90] To fully characterize this process a TycB3 mutant was produced with an active site mutation, rendering the epimerization domain inactive, and loaded with the DPhe and L-Phe tetrapeptides. In this case TycC1 produced the pentapeptide only from the TycB3 loaded DPhe tetrapeptide, providing definitive evidence that epimerization at the peptidyl-S-PCP stage precedes condensation in NRPS elongation modules (Fig. 9).[91] According to the nomenclature developed for NRPS C-domains, the TycC1 condensation domain is a $^{D}C_{L}$ catalyst, referring to its stereochemical preference for condensation of an epimerized upstream peptidyl-S-PCP donor with a downstream L-aminoacyl-S-PCP acceptor.

More recently, a novel NRPS condensation domain was characterized from the gene cluster responsible for production of the cyclic lipopeptide arthrofactin in *Pseudomonas* sp. MIS38.[92] In addition to the expected condensation activity, C-domains from the arthrofactin synthetase were found to also catalyze epimerization, consistent with the lack of E-domains and racemases observed in the gene cluster. Sfp-mediated posttranslational modification was again essential in characterizing these enzymes, allowing for the production of

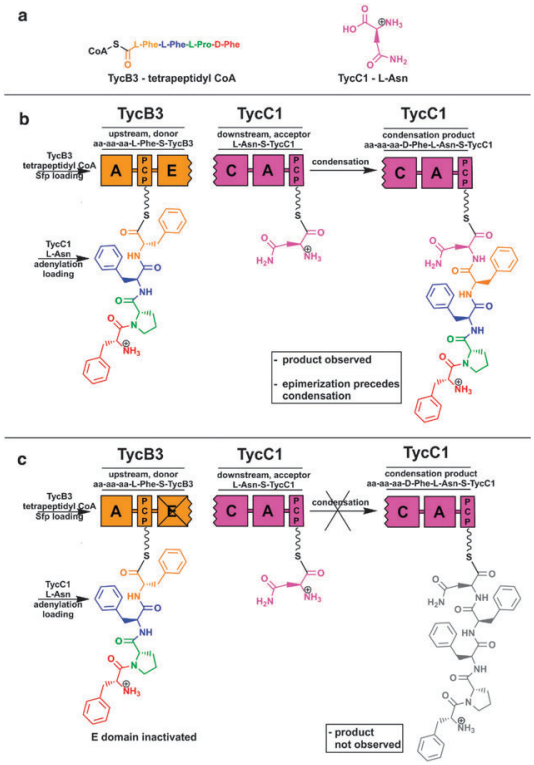


Fig. 9 Peptidyl CoAs as probes of NRPS condensation mechanism. (a) Tetrapeptidyl CoA analogues of the natural TycB4 chain elongation intermediate and L-Asn, the natural substrate for the A domain of TycC1. (b) Upon loading of TycB4 with a L-Phe tetrapeptidyl CoA, and TycC1 with Asn by auto-adenylation, only the epimerized pentapeptide product is observed, implying epimerization precedes condensation in NRPS biosynthesis. (c) A similar experiment using a TycB4 E domain mutant shows no formation of the pentapeptide, implying the non-epimerized peptide is not a substrate for the TycC1 C domain.

a lipopeptidyl-S-PCP which was shown to condense with the downstream aminoacyl-S-PCP only after epimerization. This confirmed that the timing of condensation and epimerization in dual activity C–E domains mirrored that of more traditional $^{D}C_{L}$ C-domains, consistent with phylogenetic analyses which show them to be closely related.[93] In the future, similar studies may be useful in probing the activity of other biologically relevant nonribosomal peptides which are postulated to utilize dual-function C–E domains, including the actinomycete antibiotics ramoplanin and enduracidin.[94,95]

Very recently a third type of condensation domain, catalyzing peptide bond formation between the amine group of a carrier protein bound β-phenylalanine and a octatrienoyl-S-CP intermediate, was discovered during study of the biosynthesis of the antibiotic andrimid.[96] This reaction is catalyzed by AdmF, a stand-alone transglutaminase analogue, which contains a Cys-His-Asp catalytic triad. Once again Sfp proved essential in biochemical characterization of this enzyme, allowing preparation of the putative substrate (octatrienoyl-S-CP) as well as a radiolabeled substrate analogue (^{14}C-butyryl-S-CP) both

of which are transferred *via* amide bond linkage to the amine-group of β-Phe-S-CP upon administration of AdmF. The preparation of ^{14}C-butyryl-S-CP proved especially helpful, demonstrating the broad substrate range of this enzyme as well as providing evidence for the formation of an acyl–enzyme intermediate through the use of its catalytic cysteine. The further characterization of this and transglutaminase-condensation homologues from other gene clusters could prove useful in the development of truly universal PKS–NRPS condensation catalysts.

4.2 Probes of PKS condensation mechanism

Although polyketide substrates are notoriously challenging to prepare, a similar strategy of acyl-CoA substrate loading has been used to examine the kinetic benefit of 4′-phosphopantetheine tethering in the KS-catalyzed condensation reactions of PKS biosynthesis. In 2002 Wu *et al.* compared diffuse, intramolecular, and intermolecular channeling of ketide substrates in the DEBS biosynthetic pathway.[97] By measuring rates of triketide lactone formation upon incubation of DEBS module 2/5/6 + TE with (i) NAC-loaded diketides or (ii) carrier-protein-loaded diketides, the researchers were able to compare the impact of protein–protein interactions and enzyme–substrate interactions in this biosynthetic system. Analysis of the kinetic data showed that similar mechanisms, most likely KS-domain selectivity, were involved in determination of substrate specificity for both NAC and carrier-protein bound diketides. However, comparison of the relative rates of triketide lactone formation showed that ACP-tethered diketides were processed 10–100 times faster than the corresponding NAC-diketides (Fig. 10a and b). Furthermore ACP-mediated substrate channeling resulted in the processing of extremely poor substrates whose turnover could hardly be detected in the case of diffusely presented substrates, displaying the importance of carrier-protein interactions compared to substrate–enzyme interactions in PKS biosynthesis.[97]

Studies of type I PKS systems, as in Fig. 10, are facilitated greatly by the ability to engineer TE domains onto specific modules and truncate the biosynthetic cycle through release of triketide lactones. Application of this approach in combination with synthetic acyl-NAC substrates has allowed for the direct observation or clear inference of nearly every biosynthetic intermediate in the DEBS biosynthetic pathway. However, this approach cannot be utilized with type II PKS enzymes, which exist as discrete domains and function essentially as single-turnover catalysts unless all components of the pathway are added. Additionally, there is often no chemoenzymatic route for release of nascent chains from type II PKS carrier proteins, as these systems often do not utilize TE domains.[17] Chemical release, on the other hand, requires application of strongly basic conditions to unstable polyketone intermediates, whose subsequent side-reactions and degradation can make mechanistic inferences based on their structure unreliable.

Considering these problems, an innovative approach to study type II and III PKS condensation reactions was developed by Spencer and co-workers in 2005.[98] Recalling the

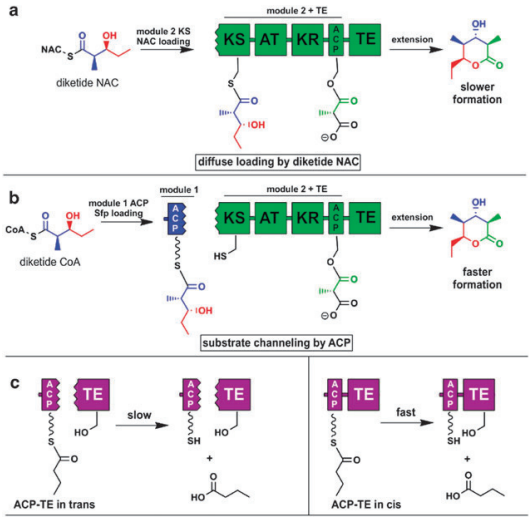


Fig. 10 Acyl CoA probes of PKS condensation and chain termination. (a) Incubation of DEBS module 2 + TE with a diketide-NAC results in diffuse loading of the KS cysteine, followed by condensation and chain cleavage to yield a triketide lactone. (b) Incubation of DEBS module 2 + TE with an Sfp loaded diketide-S-ACP substrate allows ACP-mediated substrate channeling of the diketide intermediate, resulting in faster formation of the triketide lactone and the ability to process noncognate substrates. (c) Kinetic studies of TE-mediated cleavage of an Sfp loaded butyryl-S-ACP in the DEBS system show significantly slower cleavage when the substrate is presented in *trans* (left) *versus* in *cis* on the same polypeptide (right).

normal PKS condensation reaction, in which a downstream malonyl-S-ACP undergoes decarboxylation and the resulting enolate attacks the upstream polyketide-S-ACP, they developed a non-hydrolyzable malonyl CoA analogue (**13**, Fig. 11). The replacement of the thioester bond with a thioether linkage ensured that, upon decarboxylation and capture of the polyketide intermediate, the PKS would be incapable of further condensation, effectively trapping the polyketide intermediate at that step.

This approach was first applied to study stilbene synthase (STS), a type III PKS which catalyzes formation of stilbenes such as resveratrol by condensation of a *p*-coumaroyl-CoA and three malonyl CoA (**12**) units followed by cyclization. Since type III PKS enzymes utilize acyl-CoA intermediates rather than acyl-ACPs, the intermediates captured by the malonyl CoA analogue could be easily detected by LC-MS.[99,100] Applying non-hydrolyzable CoA **13** analogue in combination with a starter-unit aromatic acyl-CoA and varying concentrations of authentic malonyl CoA **12** resulted in observation of diketide and triketide-CoA intermediates by LC-MS (Fig. 11). Interestingly, the tetraketide-CoA intermediate was not observed regardless of the aromatic acyl-CoA starter unit used. This may be due to limitations of the STS active site to accommodate the non-hydrolyzable malonyl CoA analogue, which is one methylene unit longer than the natural substrate.[98] While this method, if applied to type II PKS enzymes, would still be subject to single-turnover conditions specified above, new methods for the direct observation of

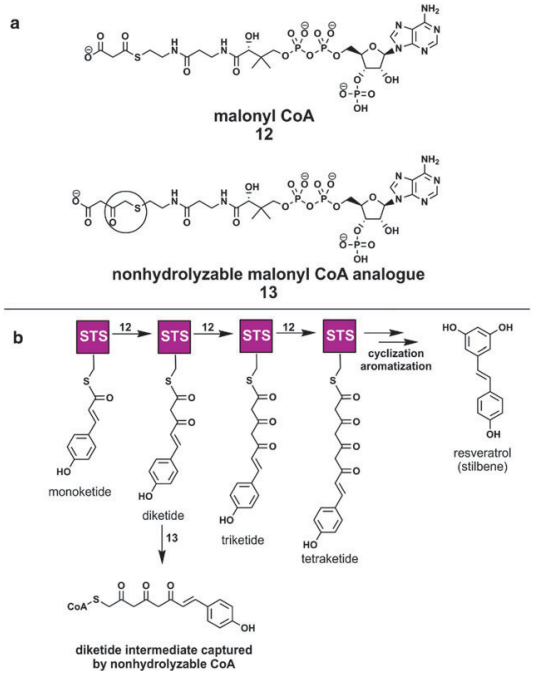


Fig. 11 Malonyl CoA analogues as probes of PKS condensation. (a) Structures of malonyl CoA (top) and Spencer's non-hydrolyzable malonyl CoA analogue (bottom). The thioether linkage is highlighted in gray. (b) Stilbene synthase (STS), a type III PKS, normally performs condensation and cyclization of a coumaroyl CoA starter and 3 malonyl CoAs to form resveratrol. Performing the same enzymatic reaction with varying amounts of the non-hydrolyzable malonyl CoA analogue, allowed the diketide (bottom) and triketide (not pictured) CoA intermediates to be captured.

carrier protein bound intermediates by mass spectrometry (see Section 4.4) in combination with improved routes to non-hydrolyzable CoA analogues[101,102] could make this a powerful tool for mechanistic analysis of carrier protein utilizing PKS systems in the future.

4.3 Probes of PKS and NRPS thioesterase mechanism

Macrocyclization is a strategy used in both polyketide and nonribosomal peptide biosynthesis to produce small molecules constrained in a biologically relevant conformation, as well as protected from hydrolytic or proteolytic degradation.[3] This cyclization is traditionally accomplished through the activity of a dedicated TE domain, located at the C-terminus of the PKS or NRPS enzyme. These domains utilize a Ser-His-Asp catalytic triad to cleave the thioester bond of the full-length peptidyl or acyl-S-CP and produce a peptidyl or acyl-O-TE intermediate, which is in turn hydrolyzed or captured by an intramolecular nucleophile to release the full-length natural product from the enzyme. An excellent perspective on recent studies into both canonical and non-canonical NRPS and PKS macrocyclization strategies has been provided recently by Kopp and Marahiel.[103]

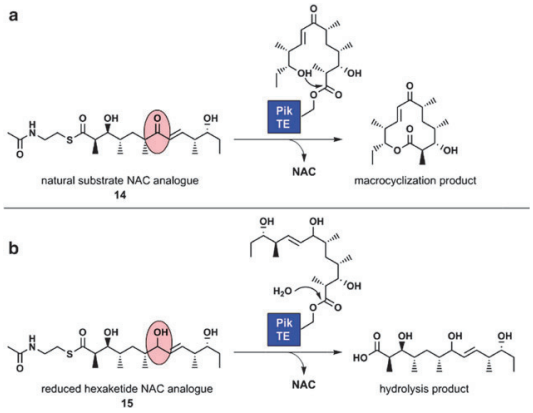


Fig. 12 Acyl-NACs as probes of PKS thioesterase mechanism. (a) The excised Pik-TE efficiently cyclizes a linear hexaketide-NAC to the 10-deoxymethynolide product. (b) Reduction of the C7 carbonyl results in abolition of cyclization activity, implying an essential role for this functionality in TE substrate presentation.

In studies of PKS TE mechanism, acyl-NACs have most often been used as carrier protein surrogates.[104] This is both due to the difficulty involved in the synthesis of sufficient quantities of the linear polyketide-CoA substrates to facilitate carrier-protein loading, as well as the viability of this method in producing scientifically relevant results. For example, acyl-NACs **14** and **15**, analogues of the linear pikromycin precursor *seco*-10-deoxymethynolide, have helped to identify the essential nature of an unreduced ketone at the 7-position to macrolide formation by the pikromycin thioesterase domain (Pik-TE) (Fig. 12).[105,106] This method has also been used to probe the mechanism and the timing of chain release in the biosynthesis of the polyether ionophores nanchangmycin and monensin, providing evidence that chain release occurs subsequent to glycosylation in nanchangmycin biosynthesis.[107,108] In addition, acyl-NACs have been used to probe the cyclization activity of the excised epothilone TE domain, as well as to examine the utility of using excised TE domains from the DEBS, Pik, cryptophycin PKS systems for the cyclization of substrate analogues.[106,109–111]

In very recent work, Tran *et al*. used carrier-protein loading to probe the PKS TE mechanism for the first time, using a simple butyryl-CoA as a polyketide mimic in the study of the DEBS ACP–TE didomain.[112] They observed that TE-catalyzed hydrolysis of the DEBS butyryl-S-ACP was much faster when covalent connection of the two enzymes was maintained than when the same butyryl-S-ACP and TE were studied as individually excised domains which interact in *trans* (Fig. 10c). While detailed kinetic analysis of the in *cis* reaction was not possible due to the fast rate of substrate hydrolysis, binding assays indicated the excised DEBS-TE showed an increased binding affinity for acylated ACP > *holo*-ACP ≫ *apo*-ACP. In the absence of structural data it is not yet understood whether this is due to a significant conformational change in ACP upon its conversion to the *holo*-state (as is observed in the NMR stucture of the tyrocidine PCP),[62] or if this is indicative of a key role played by

recognition of the substrate by the TE in type I PKS chain termination.

Carrier-protein loading has also proven useful in the study of NRPS thioesterase mechanism. For example, the TE of the fengycin NRPS shows no cyclization or hydrolysis activity towards linear peptidyl-NAC or CoA thioester substrates. This led Sieber *et al*., in one of the greatest examples of Sfp substrate promiscuity to date, to test whether loading of the full-length decapeptidyl-CoA onto the excised PCP–TE didomain would reconstitute the predicted cyclization activity of this enzyme *in vitro*.[76] Using solid-phase peptide synthesis followed by solution phase synthesis of the peptidyl-CoA thioester, they produced a panel of polypetidyl-CoAs which could readily be loaded onto the fengycin ACP–TE didomain using Sfp. They found that, after PCP–TE loading of the natural substrate precursor and incubation, the expected cyclized fengycin product was observed. In addition, examination of a number of linear fengycin analogues in which the position of the intramolecular nucleophile (in this case a tyrosine) had been altered demonstrated differential cyclization activity and provided insight into the regioselectivity of the fengycin TE. While it has since been shown that peptidyl-thiophenols can also act as fengycin TE substrates,[113] similar carrier-protein-based approaches could prove useful in the future characterization of PKS and NRPS thioesterase domains which prove recalcitrant to the use of NAC surrogates.

4.4 Probing PKS and NRPS mechanism and selectivity using ESI-FTMS

The widespread use of high-powered mass spectrometry (MS) technology in the study of modular biosynthetic enzymes is a relatively recent development but has rapidly accelerated and broadened in scope. MS first proved a valuable tool in the probing of NRPS mechanism, validating the multiple carrier model of nonribosomal peptide biosynthesis by facilitating identification of radiolabeled PCP active sites following proteolytic digest of the gramicidin NRPS.[41] Since this initial effort, the development of MS technology has transformed mass spectrometry into a leading tool for the study of mechanism and selectivity in polyketide and nonribosomal peptide biosynthesis. Particularly noteworthy has been the development of ESI-FTMS (electrospray ionization with ion cyclotron Fourier-transform mass spectrometry) and tandem MS fragmentation methods that can be readily applied to multienzyme synthases. Here we provide a few representative examples in which chemical biological techniques have been used in combination with MS to probe PKS and NRPS enzymes; for the definitive review on the topic, including a detailed discussion of methods and strategy, the reader is directed to a recent article by Dorrestein and Kelleher.[114]

As early as 1999 it was shown that ESI-FTMS could be used to observe intermediates in modular biosynthetic processes.[115] In a study of the NRPS responsible for production of enterobactin, it was found that proteolytic digest of the reconstituted NRPS followed by calculation of the mass shift of the PCP-active site containing peptide allowed the molecular weights of intermediates to be determined with high accuracy, and from this their structures inferred. Such an approach is absolutely dependent on the generation of carrier-protein containing peptide samples which (i) are sufficiently pure to allow observation of the ionized carrier protein active site peptide and (ii) do not cause thioester cleavage of PKS–NRPS reaction intermediates.[114] Fulfilling these requirements can necessitate time-consuming optimization of digestion and separation protocols which allow CP active site identification by either tandem MS or chromatographic subtraction of Sfp–CoA treated and nontreated samples.

In order to streamline interrogation of polyketide and nonribosomal peptide biosynthesis by ESI-FTMS, the groups of Walsh and Kelleher developed a method for facile detection of carrier protein active site peptides.[116] Burkart and co-workers observed in 2004 that the substrate tolerance of Sfp allowed transfer of fluorescent CoA analogues onto a variety of carrier proteins.[117] Using this approach, McLoughlin *et al*. loaded modules of the recombinant pyochelin NRPS with fluorescent or UV-detectable CoA analogues. Cyanogen bromide digest followed by chromatographic fractionation allowed the identification of PCP-containing fractions based on their incorporated fluorescence or unique UV absorbance. This allowed observation of missed cleavages in the cyanogen bromide digestion procedure and determination of elution times of PCP active site peptides which could not otherwise have been predicted (Fig. 13a). Applying this knowledge, the researchers then loaded the pyochelin NRPS with natural substrates and found they were able to utilize similar digestion and fractionation methods to observe PCP-bound intermediates from the reconstituted pyochelin biosynthetic pathway *in vitro*. This approach, as well as recent advances in tandem MS fragmentation methods which allow observation of characteristic carrier protein degradation fragments (Fig. 13b),[118,119] has greatly facilitated detection of 4′-phosphopantetheinylated PKS–NRPS proteins and peptides. The development of optimized protocols for observation of recombinant carrier protein active sites has a variety of applications in the mechanistic analysis of modular biosynthetic enzymes. One notable example has been the use of ESI-FTMS for the rapid screening of PKS and NRPS loading domain specificity. In NRPS biosynthesis, A domains are responsible for activation of a specific amino acid as an aminoacyl–AMP mixed anhydride. This aminoacyl–AMP is then captured by the 4′-phosphopantetheine thiol of the adjacent PCP to complete the loading process.[3] The traditional biochemical assay for this process uses the reverse reaction, in which radiolabeled pyrophosphate is incorporated into ATP, as a measure of A domain substrate specificity for an amino acid. While such studies have been extremely useful, in some cases amino acid activation as measured by the pyrophosphate exchange does not correlate to PCP loading of a substrate.[120] In addition, such assays must be performed individually with each amino acid, limiting the throughput of this approach.

These considerations led Dorrestein *et al*. to develop an ESI-FTMS based assay capable of determining the selectivity of A domains in substrate mixtures containing multiple amino acids (Fig. 13c).[121] *Holo*-PCP–A domain pairs (produced through posttranslational modification using Sfp and CoA) were incubated with a mixture of potential acid substrates in the presence of ATP. Inspection of the corresponding mass

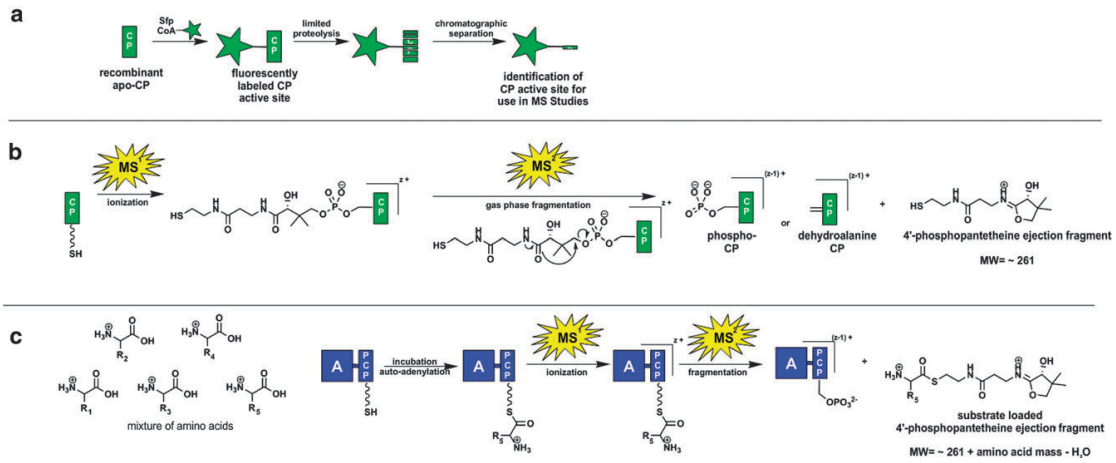


Fig. 13 ESI-FTMS as a tool for probing mechanism and selectivity in PKS and NRPS biosynthesis. (a) Strategy for detection of carrier protein active site peptides using fluorescent labeling by CoA analogues. Identification of the CP-peptide can greatly streamline interrogation of modular biosynthetic enzymes when using limited proteolysis-based approaches. (b) Tandem MS (MS^2) fragmentation of the 4′-phosphopantetheine prosthetic group in the gas phase. Using ESI-FTMS this property can be observed to arise from intact CP-containing enzymes >100 kDa in weight. This top-down method greatly simplifies observation of the modified and unmodified phosphopantetheine group, allowing the weights of substrates and biosynthetic intermediates to be deduced and from them mechanism inferred. (c) Using ESI-FTMS as a substrate loading assay for A domain activity. By allowing multiple substrates to be analyzed at once, FTMS allows insight into both mechanism and selectivity in a high-throughput manner. This approach was first used with limited proteolysis, but a tandem MS fragmentation is depicted here for simplicity.

shift allowed rapid determination of the loading specificity of a variety of modular biosynthetic enzymes. One illustration of the power of this method was its use in combination with unbiased, commercially available metabolite pools, to determine A domain specificity of two A–PCP didomains from the then uncharacterized PksX gene cluster. This allowed the authors to disqualify the small molecule difficidin as a potential product of this orphan gene cluster due to its lack of incorporation of the observed amino acids. This proved prescient, as the cluster was later shown to encode the hybrid PKS–NRPS product bacillaene.[122] In addition, by omission of the preferred substrate, loading of alternate substrates could be observed using ESI-FTMS and used as a measure of the permissivity of different A domains for noncognate substrates. Such insights are not always evident from bioinformatic analyses and may be useful in future biocombinatorial efforts.

While these initial studies utilized cyanogen bromide digestion and chromatographic separation, this method has since been shown to be compatible with tandem MS methods developed for analysis of PKS and NRPS proteins. Such methods reproducibly induce phosphopantetheinyl-elimination of posttranslationally modified carrier proteins in the gas phase, greatly simplifying the carrier-protein loading analyses of high molecular weight PKS and NRPSs whose digestion would result in a large number of peptide fragments (Fig. 13b).[118] For example, tandem MS has allowed observation of L-Phe loading of *holo*-GrsA at isotopic resolution without digestion of the 127 kDa construct. While the substrate loading has been most readily screened to date in NRPS systems, the synthesis and application of CoA analogues in a similar fashion should allow probing of PKS loading specificity.[98,123] In the future these methods promise to be a key tool in the screening of mechanism and substrate selectivity in both PKS and NRPS biosynthesis.

5. Probes of PKS and NRPS selectivity

One goal of the staggering amount of mechanistic work which has been performed on FAS, PKS, and NRPS enzymes is to provide a foundation for their rational reengineering, from the bottom up. While these efforts may hold the key to the ultimate success of combinatorial biosynthetic efforts, just as interesting have been studies which seek to establish and push the limits of selectivity and substrate tolerance of currently known PKS and NRPS catalysts. As will be seen, there is considerable overlap in studies of mechanism and selectivity of these enzymes, and the two often go hand in hand. The following section details the ways in which substrate tolerance has been used to probe and classify PPTase enzymes, as well as the utilization of ''top-down'' approaches to determination and application of the selectivity of NRPS epimerization and thioesterase domains.

5.1 Probes of PPTase selectivity

While this review illustrates (at some length) uses of the natural promiscuity of the PPTase Sfp, substrate selectivity of this enzyme has been well-demonstrated. For example, in their study of thioesterase activity, Sieber *et al.* noted that decapeptidyl-CoAs exhibited Sfp-catalyzed PCP loading at rates approximately 100 times slower than the natural CoA substrate. Mercer *et al.* explored PPTase selectivity using fluorescent CoA analogues possessing orthogonal red, green, and blue excitation characteristics.[124] Incubation of an equimolar mixture of these CoAs with a carrier protein and Sfp resulted in production of a highly defined red–green–blue ratio for each carrier protein tested. Carrier proteins from FAS, PKS, and NRPS biosynthetic pathways showed distinct labeling patterns, both alone and in complex mixtures, most

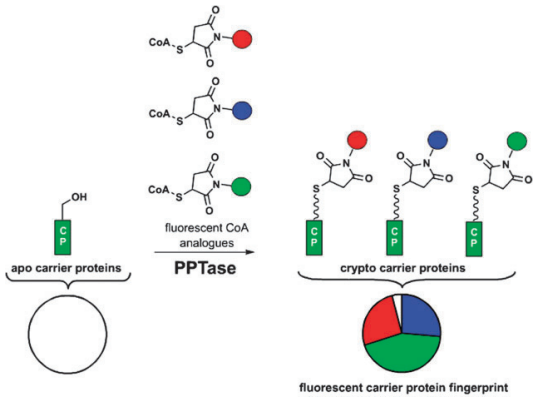


Fig. 14 A multiplex approach to analyzing PPTase loading specificity. The combined incubation of orthogonal fluorescent CoA analogues with an *apo*-CP results in reproducible loading of well-defined ratios of each fluorescent CoA. This approach can be used to generate specificity fingerprints for both CPs and PPTases.

likely due to the differential labeling rates of each structurally distinct fluorescent CoA analogue (Fig. 14). In the future such multiplex approaches could be used for the development of substrate specificity-based "fingerprinting" for classification of both carrier proteins and PPTases.

PPTase selectivity was again probed using fluorescent CoA analogues by the groups of Klok and Johnsson in 2005. This study used an O^6-alkylguanine-DNA alkyltransferase fusion tag to covalently immobilize the wild type as well as a number of helix II mutants of the *E. coli* fatty acid ACP onto a glass slide.[125] The effect of these mutants on ACP–PPTase interaction was then explored by monitoring the ability of the PPTases AcpS (from *E. coli*) and Sfp to catalyze transfer of a fluorescent CoA analogue to the immobilized ACPs (Fig. 15). The excised PCP from the tyrocidine NRPS was also examined. The results demonstrated that Sfp was broadly tolerant of all the single mutant ACPs tested, efficiently transferring the fluorescent CoA to the immobilized carrier protein. AcpS proved less promiscuous, particularly of ACPs mutated at Asp35 (the residue adjacent to the posttranslationally modified Ser in ACP). This led the authors to posit a more significant role for this acidic residue in the ACP–AcpS interaction, and propose a hypothesis for carrier protein recognition by Sfp.[125] While this study was greatly facilitated

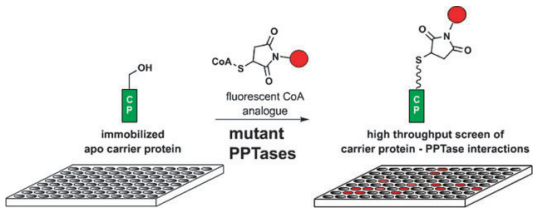


Fig. 15 Site-specific immobilization of *apo*-ACP (depicted in a 96-well plate) combined with use of fluorescent CoA analogues allows high-throughput screening of mutant PPTases based on ACP fluorescence. This can be used to determine essential ACP–PPTase interactions.

by the fluorescent nature of the assay used, which provided a simple and qualitative measure of the PPTase reaction, similar protein immobilization techniques may prove useful in future studies of protein–protein interactions of ACP.

5.2 Probes of NRPS epimerization selectivity

The study of epimerase activity has always been closely associated with the study of NRPS condensation reactions, as researchers have sought to understand the natural interplay of these two domains in nonribosomal peptide biosynthesis. Early work to probe the selectivity of E-domains utilized aminoacyl-NAC and pantetheine derivatives; however, the intrinsically low binding of these substrates for their cognate E-domains was problematic from the standpoint of gaining insight into the native substrate tolerance of these enzymes.[126]

To address this issue, Marahiel and co-workers launched an exhaustive effort aimed at characterization of the epimerase activity of PCP–E didomains towards carrier-protein loaded di-, tri-, and tetrapeptidyl thioester substrates.[127] This involved the recombinant expression and purification of four different *apo*-PCP–E didomains, the initiation module PCP–E didomains from tyrocidine (TycA) and gramicidin (GrsA) and the elongation module PCP–E didomains of the tyrocidine (TycB2) and fengycin (FenD2) synthases. Each of these didomains were then individually loaded with members of a 16-member peptidyl-CoA library consisting of peptidyl CoAs varying in length, sequence, and *N*-methylation using Sfp (Fig. 16a). Incubation of the loaded didomain followed by substrate hydrolysis allowed determination of the *in vitro* activity of each E-domain against each peptidyl-S-PCP substrate. Thus Sfp allowed not only the loading of unnatural substrates onto each excised NRPS-enzyme, but also circumvented the previously observed low affinity of E-domains for in *trans* substrates.[128]

The results of this study showed that initiation E-domains, such as TycA and GrsA, are capable of epimerization of longer peptidyl-S-PCP substrates, although at slower rates than their natural aminoacyl-S-PCP substrates. Significantly, these epimerized peptidyl-S-PCP species were substrates for the downstream TycB1 condensation domain, indicating this may be a viable approach for reengineering of nonribosomal peptide length. In many of these substrates a single stereochemical product was formed, implying that the C-domain retained its specificity as a $^{D}C_{L}$ catalyst in the context of unnatural substrates, although some ambiguity was seen.[127] The two elongation E-domains, TycB2 and FenD2, demonstrated a broad substrate tolerance which was remarkably similar to one another, an unexpected finding given that they epimerize non-homologous substrates in the larger context of their NRPS megasynthase. All four PCP–E domains showed efficient processing of peptidyl-S-PCP substrates *N*-methylated adjacent to the thioester bond, possibly due to the stabilizing influence of *N*-alkylation on the enolate-S-PCP intermediate formed during epimerization (Fig. 16b and c). Finally, in comparing the observed rates of epimerization of both initiation and elongation E-domains with aminoacyl *versus* peptidyl substrates, it was observed that aminoacyl initiation E-domains are seemingly capable of activating either aminoacyl-S-PCP or

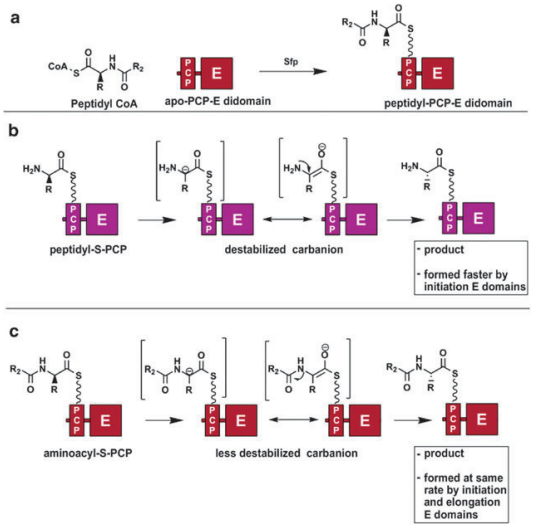


Fig. 16 Peptidyl and aminoacyl CoAs as probes of NRPS epimerization selectivity. (a) Strategy for loading PCP–E didomains with peptidyl CoAs using Sfp. (b) Aminoacyl-S-PCPs are epimerized substantially faster by E domains from initiation modules than elongation modules. This is postulated to be due to their ability to better accommodate the destabilized carbanion intermediate which results from deprotonation of their natural substrate than elongation E domains. (c) Peptidyl-S-PCPs are epimerized at similar rates by both enzymes, indicating they both can accommodate the less destabilized carbanion intermediate formed during epimerization of this substrate.

peptidyl-S-PCP substrates with equal efficiency, while peptidyl elongation E-domains show a strong preference for peptidyl-S-PCP substrates. This correlates with the hypothesis, previously proposed, that the aminoacyl-S-PCP epimerization is more electronically challenging due to the destabilizing influence of the free amine group, and thus initiation E-domains are preferentially optimized by nature to accomplish this transformation.[89,127]

5.3 Probes of NRPS thioesterase selectivity

Due to the critical nature of macrocyclization in promoting biological activity as well as the technical challenge it often presents in organic synthesis, efforts to probe the substrate selectivity of TE domains have been an area of intense focus since their discovery. While some studies exist of substrate selectivity using PKS TE domains,[104,105,111] the majority of work to date has focused on NRPS TE selectivity, in large part due to the ease with which their linear peptidyl–thioester substrate analogues are prepared through solid-phase peptide synthesis.

The most well-characterized NRPS TE domain to date is the terminal domain of the tyrocidine synthetase (Tyc-TE) (Fig. 17). In 2000 Trauger and colleagues demonstrated that when recombinant Tyc-TE was excised from its ~700 kDa megasynthetase complex, it retained cyclization activity for decapeptidyl-NAC thioesters analogues of the tyrocidine *seco*-acid.[129] In addition, Tyc-TE was capable of dimerization and cyclization of a pentapeptidyl-NAC to reconstitute

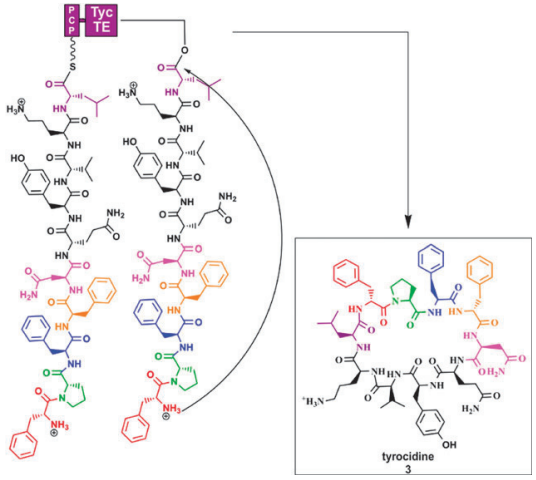


Fig. 17 Normal decapeptide-S-PCP substrate and mechanism of tyrocidine thioesterase (Tyc-TE).

formation of the gramicidin macrocycle *in vitro*. Perhaps most surprisingly, the peptide backbone of the linear decapeptide proved resilient to single amino-acid substitutions, demonstrating an absolute requirement for only two residues (D-Phe_1 and L-Orn_9) of the decapeptide for promotion of cyclization activity.[129] Ensuing studies probed the effect replacement of multiple decapeptide-NAC amino acids with short ethylene glycol linkers or hydroxy acids had on cyclization. These studies identified backbone peptide bonds whose substitution completely abrogated cyclization activity, suggesting a role for them in preorganization of the linear peptide in a product-like conformation inside the Tyc-TE active site.[130] In an accompanying paper, the authors also demonstrated that Tyc-TE macrocyclization activity was tolerant of changes in both overall chain length and internal amino acid identity of the peptidyl-NAC.[131]

Together, these studies provided a preliminary structure–activity relationship for tyrocidine macrocyclization, which suggested Tyc-TE might accept a wide range of peptidyl-NAC substrates possessing the requisite length, nucleophile position, and intramolecular hydrogen bonding interactions. Indeed this has proven to be the case, as Tyc-TE has since been shown to cyclize peptidyl-NACs substituted with a number of unnatural amino acids. This remarkable substrate promiscuity was first applied by Kohli and co-workers, who inserted into the linear decapeptidyl-NAC many alterations including an integrin binding motif.[132] Upon cyclization by Tyc-TE, this resulted in generation of a cyclic peptide with high affinity for this cell-surface receptor. Addressing a different target, Kohli *et al.* sought to increase the throughput of this method by testing the ability of Tyc-TE to cyclize peptides directly from the surface of a solid-phase resin.[133] Using this method, over 300 resin-bound linear decapeptidyl substrates were tested for cyclization by Tyc-TE and the corresponding cyclic peptides screened for antibiotic activity. This resulted in the discovery and rapid optimization of a tyrocidine analogue with increased selectivity for bacterial *versus* human membranes.

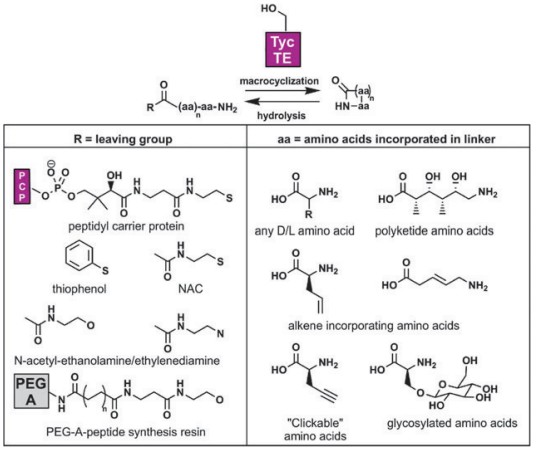


Fig. 18 Carrier protein surrogates (left) and notable unnatural amino acids (aa's—right) used as substrates for macrocyclization by Tyc-TE.

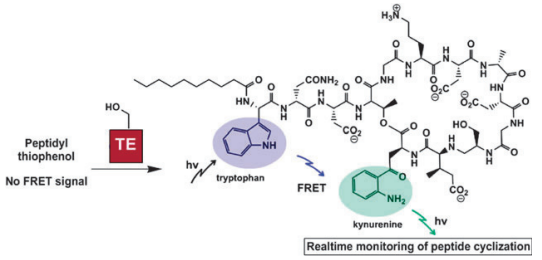


Fig. 19 Monitoring cyclization of daptomycin analogues in real time using FRET. Macrocyclization brings the FRET donor tryptophan and FRET acceptor kynurenine into spatial proximity, resulting in a detectable signal.

Since these initial studies, Tyc-TE has been used to cyclize linear peptides incorporating diverse functionalities including allyl and propargyl substituents (for later elaboration by olefin metathesis and Huisigen [3 + 2] cycloaddition, respectively), glycosylated residues, and embedded polyketide chains (Fig. 18).[134–137] In addition, a number of NRPS TE domains in addition to Tyc-TE have been explored as macrocyclization catalysts.[113,131,138,139] The study of one such system, the daptomycin TE, fostered the development of an inventive method to probe peptide cyclization using Förster resonance energy transfer (FRET).[140] In its cyclic peptide structure, daptomycin incorporates tryptophan as well as the unnatural amino acid kynurenine. These amino acids possess a natural spectral overlap of their excitation and emission spectra (Trp $\lambda_{Em} \sim 330$ nm; Kyn $\lambda_{Ex} \sim 350$ nm) which results in fluorescent excitation of Kyn ($\lambda_{Em} \sim 450$ nm) when they are brought into proximity by cyclization. After showing that the observed FRET signal was highly distance dependent and not observed in the *seco*-acid of the linear peptide, Grunewald *et al.* used this interaction to follow peptide cyclization in real time by monitoring fluorescent emission at 450 nm after incubation of linear peptidyl thioester analogues of daptomycin with daptomycin TE (Fig. 19). It was shown that this process was highly sensitive and could be applied to surface immobilized daptomycin TE domains, foreshadowing a possible use in combination with directed evolution for high-throughput screening of TE domains with altered substrate selectivities.[140,141]

These studies all emphasize the essential nature of NRPS TE domains for rapid production and diversification of chemically distinct cyclic peptide natural product libraries. Also highlighted in studies of TE specificity are questions as to the most relevant qualities of carrier protein surrogates. NAC esters have been long thought to be active in FAS, PKS, and NRPS systems partly due to natural binding interactions resulting from structural similarities to the carrier protein bound 4′-phosphopantetheine arm (Fig 18). Tyc-TE and Srf-TE have been shown to catalyze acyl transfer and cyclization using both NAC substrates as well as less activated esters, such as the Fmoc synthesis compatible *N*-(β-alanyl)-*N*-β-amino-ethanol esters.[133,138] The fengycin TE, in contrast, shows no cyclization or hydrolysis activity towards traditional CP surrogates such as NAC thioesters, only cyclizing PCP or thiophenol-bound peptides.[113] Understanding the basis for leaving group stabilization of 4′-phosphopantetheine and its surrogates during the transacylation reaction in each of these systems, as well as the unique interplay between molecular recognition elements and inherent thermodynamic activation of a substrate, will likely be important to future combinatorial biosynthetic endeavors.

6. Probes of megasynthetase structure

When Lynen reported the apparent multienzymatic nature of the yeast FAS in the mid-60s he was greeted with some skepticism, as the presence of multiple enzymatic activities on a single polypeptide was still a fairly revolutionary idea. Such qualms led Otto Warburg, a fellow researcher in the yeast field, to memorably suggest that Lynen, "Go back to the laboratory and crystallize the fatty acid synthetase," in order to definitively answer the questions concerning the tertiary structure of the FAS.[83] Neither Warburg nor Lynen could have predicted at the time just how evasive this pursuit would prove. Not only are multienzyme megasynthases extremely large and architecturally complex, properties not well-suited for collection of high quality crystallographic data, but they are also inherently flexible, owing to the fact each must deliver covalently tethered intermediates to several geometrically distinct catalytic domains. Despite the potential utility of structural insights to facilitate combinatorial biosynthetic endeavors, only recently have such data began to emerge for type I FAS, PKS, and NRPS systems, over 40 years after Warburg's comment! This has required not only technological innovation in genetics, crystallography,[142] and data analysis, but also in the creative design and application of chemical probes. Here we highlight the roles such probes have played, both in classical studies of FAS domain organization, as well as in more recent efforts which use inhibitor design and chemoenzymatic methods to investigate FAS, PKS, and NRPS systems.

6.1 Chemical crosslinking of FAS and PKS by bifunctional electrophiles

By the early 1980s the general reaction sequence of fatty acid biosynthesis was fairly well understood, in no small part due to studies of the type II *E. coli* FAS which could be broken down into its constitutive activities.[143] However, very little was known about the domain organization of type I systems, such as those used by animals or fungi. A particularly salient question was raised by the observation that when the animal FAS, normally a homodimer, was disassociated into its monomeric subunits it showed no FAS activity.[144] Assays traced this inactivation to a lack of KS-catalyzed condensation activity. What was the molecular basis for the lack of KS activity in the monomer?

Stoops and Wakil attempted to answer this question by applying a chemical crosslinking probe, 1,3-dibromopropanone (DBP, **16**), to the chicken FAS.[13] As early as Lynen's first study of fatty acid biosynthesis in 1961, it was understood that components of both type I and type II FAS enzymes were susceptible to inactivation by thiol-specific electrophiles such as *N*-ethylmaleimide and iodoacetamide.[14,145] Furthermore, by this time both the identity of the ACP 4′-prosthetic group as well as the KS catalytic cysteine had been uncovered.[143] It was therefore anticipated that a bifunctional electrophile such as DBP, with alkylating groups joined by an ~5 Å linker, could provide insight into the spatial proximity and mechanistic cooperation of intra- and intersubunit ACP and KS domains. Indeed, while SDS-PAGE analysis of the animal FAS normally results in observation of the denatured FAS monomer at ~250 kDa, a similar analysis following DBP treatment showed migration of three new higher molecular weight bands at ~400–500 kDa, indicative of monomer–monomer crosslinking (Fig. 20).[13] Although the number of crosslinked oligomers formed was not explained, the site of crosslinking was tracked to the ACP and KS thiols of opposing monomers based on substrate protection by acetyl and malonyl CoA, as well as competitive inhibition of DBP crosslinking by the known KS-specific reagent iodoacetamide.[146] The finding that the ACP and KS domains lie in such close (~5 Å) proximity led to the proposal of the head to tail model of mammalian fatty acid biosynthesis, in which the ACP of each FAS monomer lies juxtaposed with the KS of the other monomer, creating two distinct and non-complementary reaction chambers. Additional evidence for this model was provided by a study which used chloroacetyl-CoA and iodoacetamide to specifically inactivate the ACP and KS active sites of the FAS. While each individual modification led to a loss of FAS activity, upon disassociation and mixture of the two inactivated enzymes it was found that a percentage of FAS activity was regained.[147] This gain in activity was presumably due to heterodimer formation of ACP and KS deficient monomers, reaffirming the intersubunit mechanism of ACP–KS interaction in fatty acid biosynthesis. Through such studies, chemical biology aided in the creation of the initial model of a type I megasynthetase.

In some senses these studies represent both a pioneering success as well as a cautionary tale, as 18 years later technological advances would allow an alternate interpretation of

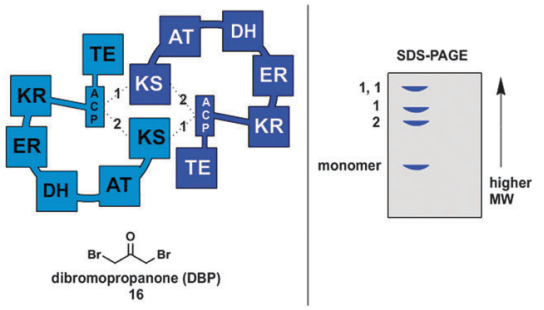


Fig. 20 Chemical crosslinking of mammalian FAS using dibromopropanone (DBP, **16**). (a) Mammalian FAS is pictured as a head-to-head dimer. DBP can crosslink the dimer between ACP and KS domains in an intersubunit (route 1) or intrasubunit (route 2) fashion. (b) Cartoon representation of SDS-PAGE analysis of crosslinking reaction. Smith successfully identified the crosslinked species as the doubly intersubunit crosslinked (1, 1), the singly intersubunit crosslinked (1), and the singly intrasubunit crosslinked FAS (2) by repeating the original DBP crosslinking experiments with epitope tagged FAS heterodimers, a key piece of evidence in support of the head-to-head model of mammalian FAS.

both the chemical crosslinking results and the heterodimerization study, ultimately leading to an entirely different model of domain organization in the type I FAS.[1,148,149] A modern version of the heterodimerization study was first performed by Smith and colleagues which utilized the newfound ability to recombinantly express active site mutants of the animal FAS, a more specific method of active site inactivation than the previously used chemical reagents. Among other notable findings, it was observed that heterodimers of animal FAS in which one monomer contained both ACP and KS inactivating mutations were still capable of palmitate biosynthesis.[148] This was unexpected as it necessitated cooperation of ACP and KS domains within the same monomer, a physical impossibility when considering the head to tail model. This led to a reexamination of the chemical crosslinking studies of Stoops and Wakil, this time with an emphasis on unambiguous identification of the high molecular weight species formed.[149] Through the use of epitope tags along with purified homo- and heterodimer FAS enzymes incorporating mutations in one, both, or neither of the ACP and KS domains, the identity of the three species was determined. In addition to double and single intersubunit crosslinking of ACP and KS domains producing FAS dimers (explained by the head to tail model), the third species migrating at a molecular weight of ~100 kDa greater than the FAS monomer was identified as the intrasubunit ACP–KS crosslinked monomer (Fig. 20). Although the reason for the greatly reduced electrophoretic mobility of the crosslinked FAS dimer is not well understood, this study overturned a key piece of evidence supporting the head to tail model, therefore allowing its revision to the current model in which ACP and KS domains of opposing subunits are aligned head-to-head.[1] Further evidence for this head-to-head arrangement has since been provided by site-specific chemical crosslinking of cysteines engineered near the N-termini of animal FAS.[150]

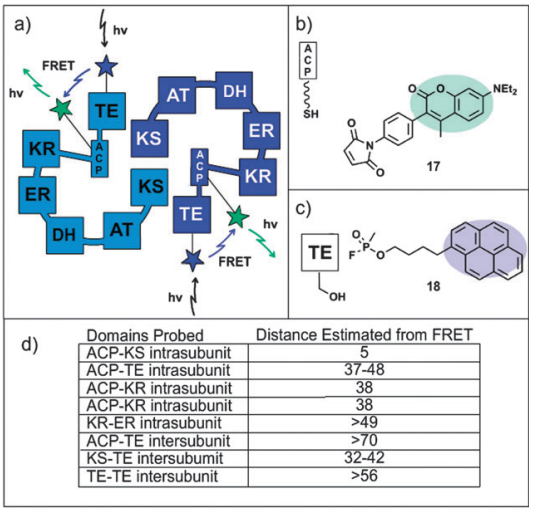
While this comparatively dramatic story was unfolding in the study of the animal fatty acid biosynthesis, DBP

crosslinking was also applied to study the domain organization of other type I enzymes. Both the fungal FAS as well as the fungal PKS 6-MSAS were shown to migrate as dimers on SDS-PAGE following DBP treatment, suggesting ACP–KS interaction occurred at the dimer interface in these enzymes as well.[151,152] As genetic information became available for non-iterative type I PKS enzymes, most notably the DEBS synthase, the DBP crosslinking of what were implied to be ACP and KS domains was also cited as evidence for a double helical, head-to-head model of PKS quaternary structure.[153] Since this study, crosslinking by DBP and other bifunctional electrophiles has been used to verify the correctly folded, dimeric nature of recombinant PKS modules by intersubunit crosslinking, although it has been observed that when modules of DEBS are exposed to bifunctional electrophiles with longer linkers (~15 Å) than that of DBP (~5 Å) they undergo non-specific crosslinking at sites other than the ACP and KS thiols.[154,155] Notably, similar applications of bifunctional electrophilic crosslinkers targeting thiol and amino groups to NRPS enzymes have failed to identify any dimerized species, an observation cited as evidence that these enzymes do not readily form dimers in solution and may exist in multiple oligomeric states.[156] Since these studies, both the head-to-head models of mammalian FAS as well as the multimerization states of PKS and NRPS enzymes have been verified by recently acquired structural data.[63,66,67,157]

6.2 FRET studies of interdomain distance in animal FAS

Another classic chemical method which has proven capable of providing insight into the domain arrangement of multi-domain biosynthetic enzymes is the use of fluorescent small molecule affinity tags to estimate interdomain distances. This method to date has only been applied to the animal FAS, in a number of inquiries which culminated in the mid-1980s. In one such study, Kolattukudy and co-workers utilized the well-chronicled susceptibility of serine hydrolases to fluorophosphonate (FP) inhibitors to specifically label the TE domain of the FAS with a pyrene-fluorophosphonate (**17**, λ_{Ex} ~ 346 nm, λ_{Em} ~ 376 nm) capable of acting as a FRET donor.[158] They then utilized a coumarin-maleimide (**18**, λ_{Ex} ~ 398 nm, λ_{Em} ~ 459 nm) to label the 4′-phosphopantetheine thiol of the ACP domain with an acceptor fluorophore (Fig. 21). As with DBP crosslinking, specificity of modification was monitored largely by activity assays and substrate protection studies, with SDS-PAGE and HPLC analysis of fluorescent fragments following proteolytic digest providing a new method of analysis. Although FRET measurements were hindered by partial off-site alkylation of a non-ACP thiol by the reactive maleimide, labeling of ACP and TE active sites in this manner was found to result in an observable enhancement in the coumarin emission spectrum upon excitation of the pyrene moiety. The well-known methods for disassociation of the FAS dimer allowed the researchers to track this interaction to the intrasubunit interaction of the ACP and TE domains from the same monomer, and calculations estimated the distance between the two at ~37 Å.[158]

In the same time period Cardon and Hammes was developing a method which utilized nitrobenzaoxazole-CoA



Domains Probed	Distance Estimated from FRET
ACP-KS intrasubunit	5
ACP-TE intrasubunit	37-48
ACP-KR intrasubunit	38
ACP-KR intrasubunit	38
KR-ER intrasubunit	>49
ACP-TE intersubunit	>70
KS-TE intersubumit	32-42
TE-TE intersubunit	>56

Fig. 21 Estimation of interdomain distances in mammalian FAS using FRET. (a) Labeling of the TE with a pyrene donor, and the ACP with a coumarin-acceptor moiety, results in production of a FRET signal on excitation of the donor fluorophore. The intensity of this signal can be used to estimate interdomain distances. (b) Structure of coumarin-maleimide used to label ACP. (c) Structure of pyrene-fluorophosphonate used to label TE. (d) Accumulated inter- and intrasubunit interdomain distances estimated by FRET study of the mammalian FAS by the groups of Kolattukudy[158] and Hammes.[159]

(λ_{Ex} ~ 488 nm, λ_{Em} ~ 539 nm) to specifically acylate ACP.[159] The choice of this fluorophore allowed it to act as a FRET acceptor to both pyrene-fluorophosphonate as well as the NADPH cofactors (λ_{Ex} ~ 365 nm, λ_{Em} ~ 460 nm) which were well-known to bind to the KR and ER sites, allowing many intra- and intersubunit distances for the animal FAS to be calculated.[160] The independent analysis of ACP–KR, ACP–ER, ACP–TE, TE–TE, and KR–ER pairs led to the observation that the interdomain distances were considerably larger than expected (sometimes >59 Å, Fig. 21), and would likely not be accessible to a statically located, 20 Å long 4′-phosphopantetheine arm.[160,161] This led the authors to note that "relatively large conformational changes may be part of the catalytic cycle [of animal FAS]," another finding which has been echoed in recent commentary on the newly available crystal structures of this enzyme. While FRET based approaches have been largely superceded by other methods of structural analysis, in the future they may be useful in providing quick and reliable assays for the structural integrity of engineered PKS and NRPS enzymes by allowing researchers to verify retention of interdomain distances and quaternary architecture similar to the native enzyme. Such efforts may be facilitated by the increasing number of active site directed reagents for FAS, PKS, and NRPS reported in recent years.[7,162,163]

6.3 Chemoenzymatic crosslinking to capture transient protein–protein interactions in FAS, PKS, and NRPS

The past two years have seen multiple breakthroughs in the structural analysis of type I modular biosynthetic enzymes.

These achievements have included multiple refinements of the fungal FAS crystal structures, generation of high resolution structural data for multidomain PKS fragments, and most recently the first ever crystallization of an entire NRPS module.[4,65,67,164] A full commentary on many of these findings has been provided by three recent reviews.[1,4,165] While these studies have answered many questions concerning the quaternary architecture of these enzymes, they have raised many more, a large number of which are related to the protein–protein interactions of ACP. Namely, what is the physiochemical nature of the large conformational changes which must occur in FAS, PKS, and NRPS enzymes in order for the carrier protein to interact with its partner proteins in these systems? How is this accomplished when the conformational freedom of the carrier protein is constrained, as it is when it lies in the center of several hundred kDa PKS and NRPS enzymes? What is the molecular basis for the transient protein–protein and enzyme–substrate interactions which drive these processes?

One approach which has been employed for years to facilitate crystallization of transient enzyme conformations has been the use of tight-binding inhibitors and cofactor analogues. Recalling that *E. coli* ACP was originally understood not as an enzyme, but as a cofactor,[24,25] one can take this analogy to the extreme and imagine the synthesis of a carrier-protein analogue which would act as a tight-binding inhibitor of its FAS, PKS, and NRPS accessory enzymes, providing insight into the stepwise protein–protein interactions which drive modular biosynthetic processes.

Indeed, while *E. coli* ACP has been chemically synthesized in its apo form by solid-phase peptide synthesis,[166] the most successful routes to ACP analogues have used molecular biology in combination with organic and chemoenzymatic synthesis. The first study pursued with the explicit goal of capturing the protein–protein interactions of carrier-protein mediated biosynthesis through an inhibition-based approach was undertaken by Worthington *et al.* in 2006.[167] Employing the well-characterized ACP–KS enzyme pairs used during type II fatty acid biosynthesis in *E. coli*, the authors investigated the ability of ACP analogues modified in the 4′-phosphopantetheine moiety to specifically bind and inhibit their cognate KS domains. These ACP analogues (denoted *crypto*-ACPs due to the "hidden" nature of the natural thiol of the phosphopantetheine moiety) were produced through a one-pot, chemoenzymatic transformation of pantetheine analogues into CoA analogues, which were then transferred onto the *apo*-ACP by Sfp (Fig. 22).[168] As the in *trans* interactions of ACP and KS domains are necessarily weak, a non-hydrolyzable amide bond to an electrophilic inhibitor moiety was synthetically incorporated into the 4′-phosphopantetheine arm of the ACP analogues in order to irreversibly capture any association of the ACP and KS active sites. The inhibitor "warheads" chosen for this task were an electrophilic epoxide, modeled after the known KS inhibitor cerulenin, and a chloroacrylamide Michael acceptor (**19**), another known affinity label for enzymes which utilize active site cysteines.

Upon incubation of the electrophilic *crypto*-ACPs with the *E. coli* KS enzymes, it was found that KASI and KASII site-specifically crosslinked to the *crypto*-ACP (as analyzed

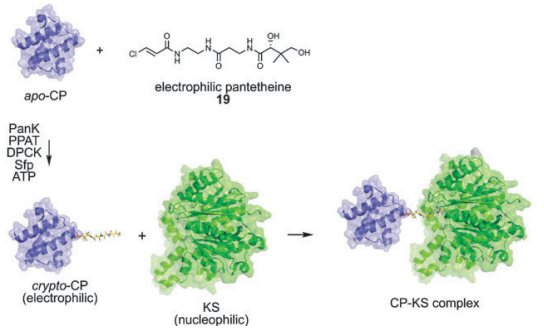


Fig. 22 General strategy for chemoenzymatic crosslinking of ACP and KS domains. A one-pot chemoenzymatic method is used to transform an electrophilic pantetheine analogue into an electrophilic CoA analogue and load it onto an *apo*-CP. This electrophilic ACP is then incubated with the KS, whose nucleophilic cysteine attacks the electrophile, trapping the interaction of the two proteins.

by gel-shift and MS-analysis), while a third *E. coli* ketosynthase, KASIII, did not. This is consistent with the known mechanism and substrate selectivity of these KS domains, as KASI and KASII normally facilitate condensation by translocation of the growing chain from acyl-S-ACP to their own nucleophilic cysteine, while KASIII initiates fatty acid biosynthesis by acting on an acyl-CoA substrate.[59] This method also proved capable of shedding light on the suitability of protein–protein interactions between unnatural ACP–KS pairs, as electrophile incorporating carrier proteins from PKS and NRPS pathways showed differential crosslinking efficiency with the *E. coli* FAS ketosynthase KASI. This latter finding has potential utility in the directed evolution of compatible ACP–KS pairs for use in combinatorial biosynthesis.

Since the initial report of this technique, it has been applied with success to a number of modular biosynthetic systems, allowing site-specific crosslinking of the discrete ACP and KS-CLF of the type II enterocin PKS,[169] a discrete ACP and AT–DH–KS tridomain of type I animal FAS,[170] and two discrete ACPs and AT–KS didomains from modules 3 and 5 of the type I DEBS PKS (Fig. 23a).[171] This last study found that ACPs and KS–AT pairs from the same module of the DEBS synthase showed preferential reactivity, reaffirming the ability of ACP–KS crosslinking to identify complementary enzymatic partners. Another notable aspect of these studies was the finding that the crosslinking efficiency of *crypto*-ACPs incorporating fatty-acyl or aromatic groups into the electophilic 4′-phosphopantetheine arm showed a good correspondence with substrate specificity of the KS in the study of ACP–KS crosslinking of type II FAS and PKS systems.[169] This indicates that through skillful inhibitor design it may be possible to use ACP–KS crosslinking to gauge the relative contributions of the substrate specificity conferred by the group located on the 4′-phosphopantetheine chain and compatibility of protein–protein interactions for ACP–KS association.

The ideal use of this method would be its application in concert with X-ray crystallography to yield insight into the protein–protein interactions which lie at the heart of the

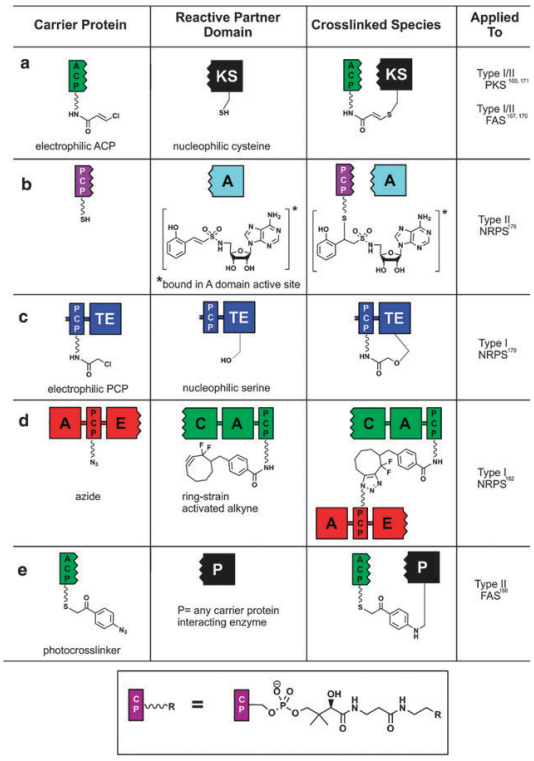


Fig. 23 Chemoenzymatic crosslinkers applied to date for structural analysis of modular biosynthetic enzymes. (a) ACP–KS crosslinking. (b) PCP–A crosslinking. (c) PCP–TE crosslinking. The EntF construct is abbreviated for simplification. (d) COM-mediated crosslinking. (e) ACP–partner enzyme photocrosslinking. (Bottom) Structural shorthand used to denote unnatural 4′-phosphopantetheine arms used in a, c, d, and e.

ACP–KS condensation reaction. For successful crystallization of these species, purity of the crosslinked ACP–KS pair will be crucial. This creates a potential barrier, as the one-pot chemoenzymatic method used to create *crypto*-ACPs utilizes four enzymes (pantothenate kinase, PanK; 4′-phosphopantetheine adenyltransferase, PPAT; dephospho-CoA kinase, DPCK; and Sfp) along with the ACP and KS of interest.[167,168] One way to decrease the complexity of the reaction mixture would be to isolate the electrophilic CoA analogue prior to ACP loading; however, it has been observed that some of the most useful crosslinkers incorporate reactive tags which are prone to hydrolysis under the strongly acidic conditions typically utilized during HPLC purification of CoA analogues.[172] These challenges led Haushalter *et al.* to develop a strategy for the purification of crosslinked ACP–KS domains.[173] This involved first optimizing the stoichiometry of the ACP–KS crosslinking reaction in order to drive it to consumption of the 6xHis-tagged KS domain. Then, through use of orthogonally tagged CoA biosynthetic enzymes along with native Sfp, affinity chromatography was used to isolate purified preparations of the ACP–KS complex in good yield. This approach has since allowed isolation of multimilligram quantities of crosslinked ACP–KS enzymes from the *E. coli* FAS pathway. In addition to the disassociable *E. coli* FAS enzymes, another particularly promising system for crystallographic analysis is the aforementioned DEBS synthase. The KS–AT didomain from this PKS has already been crystallized at high resolution, and the demonstration that excised ACP domains are capable of reconstituting PKS activity with these fragments indicates the ACP and KS–AT pair are able to take on natural protein conformations even when interacting in *trans*.[174] Both of these observations bode well for the ultimate success of ACP–KS crosslinking applications to the DEBS PKS system.

In addition to PKS and FAS condensation, carrier protein crosslinking has also been used to probe the interaction of PCP and partner domains in nonribosomal peptide biosynthesis. The group of Aldrich first applied this approach to an NRPS system for the site-specific crosslinking of PCP and A domains.[175] The A domains of NRPS enzymes are known to be extremely susceptible to inhibition by non-hydrolyzable aminoacyl–AMP analogues, in particular aminoacyl–AMS (“adenosine-5′-monosulfamate”) mimics which exhibit K_i values in the single nanomolar range (Fig. 24b, **20**).[162] These analogues have found their main application to date as antibiotics which target the production of NRPS and PKS virulence factors in pathogenic bacteria, and an overview of such efforts has been provided very recently by Cisar and Tan.[176] In order to apply these same tools to structural analysis and PCP–A crosslinking, Qiao *et al.* engineered a vinyl sulfonamide, normally a Michael acceptor of low reactivity, into a non-hydrolyzable aryl-adenylate which was a tight-binding inhibitor of the A domain from the mycobactin NRPS.[177] Incubation of the *holo*-PCP from the mycobactin system with this analogue did not result in modification of the terminal thiol, indicating its non-specific reactivity was low. However, upon addition of the discrete mycobactin A domain, modification of the terminal thiol of the PCP-4′-phosphopantetheine arm was observed by MS analysis, indicative of the ability of the unsaturated adenylate to mediate formation of a PCP–A complex (Fig. 23b). That this phenomenon was driven by specific protein–protein interactions of the mycobactin PCP–A domain pair was demonstrated by the lack of PCP-modification observed when

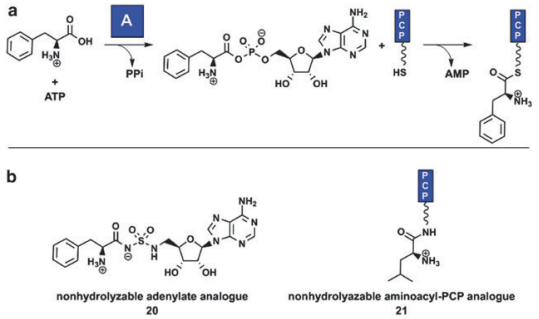


Fig. 24 A domain activity. (a) Reactions catalyzed by the A domain. (b) Inhibitors of A domain adenylation (**20**) and aminoacyl-S-PCP (**21**) forming half reactions.

noncognate A domains from the yersinabactin and enterobactin NRPS pathways were applied in a similar manner.[175]

Almost concurrently with this study, Liu and Bruner reported their own use of PCP-based inhibitors of the A and TE domains within the context of the terminal module of the enterobactin NRPS.[178] The 140 kDa multienzyme EntF contains C, A, PCP, and TE activities within a single polypeptide. They first sought to inhibit the second half reaction catalyzed by the A domain, in which the bound aminoacyl-adenylate is transferred to the free thiol of the *holo*-PCP (Fig. 24a). By loading the *apo*-PCP with a non-hydrolyzable, amide-linked version of the ordinary product of this reaction, Ser-S-PCP, they observed decreased A domain activity as measured by pyrophosphate exchange assay. The decreased ability of the A domain to catalyze the first half reaction of adenylation is possibly indicative of its preferential locking in a second half reaction-like conformation (Fig. 24b, **21**). Also shown was the ability of the EntF A domain to catalyze transfer of serine to the amino-CoA loaded PCP for formation of its own non-hydrolyzable, aminoacyl-PCP inhibitors.[178]

Turning their attention to the interaction of the PCP and TE domains, Liu and Bruner next tested the viability of a 4′-phosphopantetheine incorporating a chloroacetamide moiety for covalent crosslinking of the PCP and the active site serine of the TE domain (Fig. 23c). Although haloacetamides have traditionally been found to target activated cysteine residues,[179,180] in this case it was hoped that proximity effects resulting from covalent tethering of the inhibitor would overcome this natural preference. Indeed, upon PPTase-catalyzed loading of the EntF PCP domain with the non-hydrolyzable, electrophilic chloroacetamide CoA, it was found that the V_{max} of TE hydrolysis was reduced by approximately fivefold.[178] This is indicative of covalent modification of the NRPS, although evidence for the site-selectivity and nature of this modification proved difficult to obtain due to the large size of the EntF enzyme.

Finally, switching focus to intermodular communication in nonribosomal peptide biosynthesis, a chemoenzymatic crosslinking approach was recently developed for the study of NRPS interactions facilitated by communication-mediating (COM) domains.[181] Found at adjacent termini of NRPS modules that perform condensation reactions in which the peptidyl-PCP donor and aminoacyl-PCP acceptor are housed on separate polypeptides, COM domains are short, ~25 residue protein segments which mediate molecular recognition between NRPS megasynthetases.[182] Due to the crucial role they play in promoting intermodular condensation, COM domains represent a key target for engineering in combinatorial biosynthetic approaches.[183]

Hur *et al*. sought to study this interaction by using the [3 + 2] Huisigen cycloaddition of azides and alkynes. In particular, it was hypothesized that a strong, COM domain-mediated interaction of NRPS modules loaded with complementary azides and alkynes on their PCPs would facilitate [3 + 2] cycloaddition and concurrent crosslinking. To examine this approach, a panel of azide and alkyne pantetheine analogues was synthesized and chemoenzymatically loaded onto the well-studied NRPS multienzymes TycA (A/PCP/E) and TycB1 (C/A/PCP). Loading of the PCPs of TycA with an azide and TycB1 with a difluorinated cyclooctyne184 allowed the transient interaction of TycA and TycB1 to be captured through strain-promoted, [3 + 2] cycloaddition, as observed by gel-shift and MS identification (Fig. 23d). Evidence that this crosslinking event was promoted by the endogenous protein–protein interactions of the TycA–TycB1 pair was provided by the finding that crosslinking was abrogated by pre-denaturation of the enzyme and the failure to observe a similar gel-shift when noncognate PCP domains from the enterobactin and vibriobactin biosynthetic pathways were similarly loaded with azides and incubated with cyclooctyne TycB1. Finally, the role of the COM domain in negotiating this process was specifically investigated through the use of TycAΔ23, a TycA mutant in which the 23 residues corresponding to the COM domain were deleted. It had previously been shown that this mutation leads to a loss of condensation activity with the downstream, L-Pro loaded TycB1, presumably due to a lack of interaction of the two megasynthetase modules.[182] As expected, upon incubation of azide-loaded TycAΔ23 with cyclooctyne–TycB1 crosslinking was not observed, providing ultimate confirmation of the ability of this method to discriminate between native and non-native intermodular protein–protein interactions in nonribosomal peptide biosynthesis.[181] The use of [3 + 2] cycloaddition to assay for complementary COM domain interactions could be useful in the engineering of complementary enzyme pairs for combinatorial biosynthetic approaches, as well as for conformational locking of NRPS systems to simplify structural analysis.

Carrier-protein based chemoenzymatic crosslinking methods have now been applied to almost every modular biosynthetic architecture besides iterative fungal FAS and PKS systems (Fig. 23), and the future will no doubt see the development of more probes to study the interaction of ACP with DH, KR, ER, and TE domains in polyketide biosynthesis, and PCP domains with E and C domains in nonribosomal peptide biosynthesis. As many of these domains do not contain hydrolytic active sites or well-known affinity agents, one approach which may prove useful is the use of photoaffinity reagents. This approach has previously been applied to the *E. coli* FAS ACP, which was chemically modified at the phosphopantetheine thiol with an aromatic azide and photocrosslinked to an inner membrane protein found in membrane vesicles (Fig. 23e).[185] It remains to be seen whether the noncovalent interactions of ACP and PCPs with their partner enzymes in fatty acid, polyketide, and nonribosomal peptide biosynthesis are sufficiently strong to yield site-specifically crosslinked complexes through similar application of photocrosslinking methods. Regardless, chemoenzymatic crosslinking of modular biosynthetic enzymes promises to remain a fertile area of research for natural product biosynthesis.

6.4 Affinity labels of thioesterase domains in FAS, PKS, and NRPS

In addition to using carrier protein analogues as inhibitors for the structural study of modular biosynthetic enzymes, more traditional small molecule affinity labels have also been employed. Most notable among such efforts have been inhibitors designed to probe TE-catalyzed cyclization in PKS and NRPS

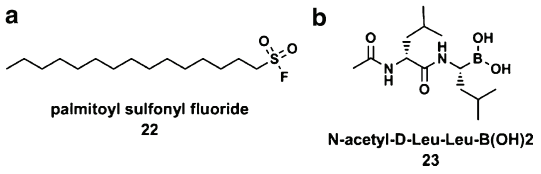


Fig. 25 Electrophilic substrate analogues which have been used as structural probes of (a) FAS and (b) NRPS TE domains.

systems. This focus on TE domains stems both from the importance of this process in promoting biological activity among nonribosomal peptides and polyketides, as well as from the fact that these enzymes belong to the α–β hydrolase family, for which a large body of literature on active site directed inhibition already exists.[186]

The first affinity label to be used in combination with structural analysis of a megasynthetase TE was actually applied to the study of fatty acid biosynthesis, when palmitoyl sulfonyl fluoride was used to label the excised TE from the animal FAS in 2000 (Fig. 25a).[187,188] Around the same time, a similar approach was applied independently to probe TE-catalyzed peptide cyclization in the surfactin biosynthetic pathway. Surfactin TE normally cyclizes a linear lipoheptapeptide-S-PCP substrate through transacylation to the active site serine of the TE, followed by intramolecular attack of the N-terminal hydroxy-fatty acid on the lipoheptapeptide–O–TE bond. Tseng *et al.* attempted to mimic this TE-bound intermediate by synthesizing linear surfactin analogues in which a boronic acid moiety, a known irreversible inhibitor of serine hydrolases, was inserted at the C-terminal end in place of the electrophilic thioester of the natural PCP-bound substrate.[138] The active site was then visualized by soaking these peptidyl–boronic acids into crystals of the excised surfactin TE followed by structural analysis. Interestingly, clear electron density was observed only for the two C-terminal amino acids located closest to the boronic acid-bound active site serine of surfactin TE (Fig. 25b). This phenomenon was found to be consistent regardless of the identity of the other residues housed on the boronic acid inhibitor. The two C-terminal residues that were visualized, D-Leu_6 and L-Leu_7, had previously been shown to be essential to TE-catalyzed hydrolysis or cyclization. Analysis of the electron density generated by the bound inhibitor showed they were also the only two residues which were bound in the hydrophobic pocket of the enzyme, with the rest of the peptidyl moiety extending into the solvent. These results along with modeling studies of the full-length substrate were consistent with the previously proposed mechanism of nonribosomal peptide cyclization in which the nascent peptidyl chain assumes a product-like conformation prior to macrocycle formation.[138,189]

Pikromycin (Pik) TE was one of the first PKS TE domains to be crystallized, and is of interest from a combinatorial biosynthetic standpoint due to its unique ability to catalyze macrocyclization of both 12- and 14-membered rings.[190] However, investigating PKS macrolactonization with substrate analogues has proven a challenge relative to the above studies of FAS and NRPS chain termination, primarily due to the complex nature of the chemical syntheses necessary to

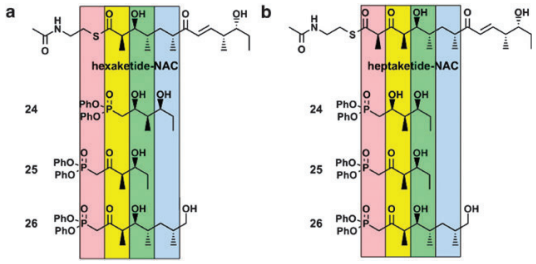


Fig. 26 Structural probes of pikromycin PKS TE domain. Shown are comparisons of tri- (**24**/**25**) and pentaketide (**26**) diphenylphosphonates to NAC analogues of Pik-TE natural (a) hexaketide and (b) heptaketide-S-ACP substrates.

produce fully extended polyketide substrate analogues. In order to circumvent this difficulty and still gain insight into the structural basis for PKS macrolactonization, the groups of Fecik, Sherman, and Smith engaged in a collaborative effort to synthesize and evaluate truncated tri- and pentaketide affinity labels as structural probes of the excised Pik-TE domain (Fig. 26, **24–26**).[191,192] In these studies a reactive diphenylphosphonate was chosen as the TE-directed affinity label, due to its compatibility with the potentially nucleophilic hydroxyl groups contained in the tri- and pentaketide analogues. Treatment of Pik-TE with these diphenylphosphonate polyketides resulted in moderate to complete inhibition of TE activity. Interestingly **24**, which bears a hydroxyl at the C3 position (adjacent to the diphenylphosphonate), was found to be a significantly better inhibitor than both linear tri- and pentaketide affinity labels **25** and **26** which incorporate the ketone found in the natural substrate at the same C3 position (Fig. 26). It is unclear whether this is simply a result of differential electronic activation of the reduced triketide **24**, which is capable of forming an intramolecular hydrogen bond with the adjacent diphenylphosphonate, or whether this is more deeply indicative of a preference of the Pik-TE for a hydroxyl group at this position, as is found in the linear hexaketide chain of *seco*-10-deoxymethynolide (see comparison in Fig. 26). Structural analysis showed no differences in the binding modes of the differentially reduced analogues.[191]

Structures of Pik-TE with each of the linear polyketide diphenylphosphonates bound were obtained through crystal soaking. Surprisingly, while these bound substrate analogues allowed clear visualization of the active site and oxyanion hole, they showed no specific hydrogen bonding contacts to Pik-TE and the enzyme did not demonstrate any significant conformational shifts relative to its previously observed uninhibited structure.[191,192] This was used to argue against an "induced fit" model of substrate binding by the thioesterase. Perhaps the most notable result of this study was the use of electron density obtained from binding of pentaketide diphenylphosphonate **26** to build off and model the full-length linear heptaketide bound in the Pik-TE active site. This led to a model of pikromycin macrocyclization in which conformational restrictions resulting from keto–enol tautomerization of the sole unreduced ketone of the natural substrate were essential in constraining the linear polyketide into a product-like

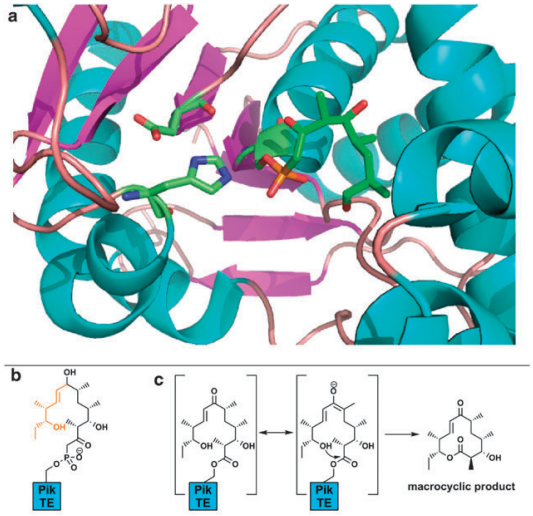


Fig. 27 Structural basis for pikromycin macrocyclization. (a) Crystal structure (PDB:2HFJ) of pentaketide affinity label bound to Pik-TE active site. The catalytic triad (S148, Asp176, His268) and bound affinity label are depicted in ball and stick form with carbons in green. (b) Close-up of active site and catalytic triad with bound **26**. (c) Cartoon depiction of enzyme-bound inhibitor pictured in Pik-TE crystal structure. The portion in orange was modeled using insights from the Pik-TE bound product. (d) Structural basis for TE-catalyzed macrocyclization based on molecular modeling of full-length hexaketide-O-TE. The C7 carbonyl is key to maintaining the planar configuration necessary for TE-catalyzed polyketide cyclization.

conformation for formation of the intramolecular ester bond (Fig. 27).[192] This is consistent with the previously noted inability of Pik-TE to cyclize linear substrate analogues reduced at this position (Section 4.3).[105,106] It will be interesting to see if the similar application of this approach to less permissive PKS macrocyclization catalysts will confirm these as general characteristics of polyketide macrolactonization.

While we constrain ourselves here to a discussion of inhibitors specifically designed for the purpose of probing PKS and NRPS TE domains, also notable are recent studies, chiefly from the group of Bruner, which have applied mechanism-based inhibitors to excellent effect in the structural analysis of the PKS–NRPS accessory enzymes responsible for the production of unnatural amino acids used in biosynthesis of the antibiotic vancomycin and the antitumor enediyne C-1027.[193–195] The ability of such studies to yield fundamental advances in our understanding of the molecular basis for natural product biosynthesis will likely be key to future efforts at rational reprogramming of these enzymes.

7. Reverse genetics approaches to identification of FAS, PKS, and NRPS biosynthetic enzymes

In addition to their application in probing the mechanism and structure of modular biosynthetic enzymes, techniques lying at the interface of chemistry and biology have also been used in "reverse genetics" approaches to the discovery of FAS, PKS, and NRPS multienzymes. In this context, reverse genetics refers to the purification and partial sequencing of a protein of interest in order to facilitate identification and cloning of the gene responsible for its production (for a figure illustrating this method see Section 7.2). While such approaches have largely been superceded by genetic and bioinformatic based approaches they constitute the original method for FAS, PKS, and NRPS gene discovery and have a rich, though often forgotten, history. For example, the first FAS PPTase gene was discovered through the use of an immobilized *apo*-ACP, which allowed the enzyme to be isolated and N-terminally sequenced, providing insight into the sequence of an enzyme class for which, at the time, no homologues existed.[46] Similarly, in 1990 Beck *et al*. used traditional protein-purification techniques to isolate 6-MSAS, a fungal PKS. This allowed production of a polyclonal antibody against 6-MSAS which was used to screen an expression library and ultimately facilitate identification of the PKS coding sequence.[196]

In addition to discovery, the dependence of these methods on protein isolation means they may be used to study natural product enzymes directly from fractionated cell lysate of a producer organism. The utility of this method is illustrated by its use in the study of the cyclosporin NRPS. This enzyme is encoded by a 45.8 kb segment of DNA, one of the largest known genomic open reading frames, making cloning and manipulation of this gene for recombinant expression a challenge even by the standards of modern technologies.[197] In contrast, Lawen and Zocher used protein-purification methods to isolate and study multiple activities of the entire 1600 kDa cyclosporin NRPS directly from the producer organism *Tolypocladum niveum* as early as 1990.[198]

While the revolutionary insights gained through genetic analyses of polyketide and nonribosomal peptide producers are well-chronicled[6,199] and have attracted the majority of resources in this field, recently there has been a modest renewal in interest in protein-based approaches to the study of modular biosynthetic enzymes. This is in no small part due to the development of new methods for their specific labeling, many of which utilize the newfound understanding of carrier protein posttranslational modification and PPTase promiscuity (Section 3.3).[52] Protein-based approaches to isolate and study complete PKS and NRPS enzymes directly from their natural producer organisms, as well as to discover and classify mega-synthases from unsequenced organisms on a proteomic time scale, can be seen as complementary and extremely well-paired with genetic techniques and represent a unique opportunity to understand natural product biosynthesis from the gene, to the protein, to the small molecule. As this topic has not been recently reviewed, we provide here a thorough discussion of both general strategies and specific methods for labeling of modular biosynthetic proteins in proteomic settings, as well as a commentary on the advantages and future applications each of these techniques may have.

7.1 *In vitro* labeling of carrier protein domains

The majority of techniques developed for specific labeling of FAS, PKS, and NRPS multienzymes utilize manipulation of the 4′-phosphopantetheine prosthetic group in some form.

This PPTase-catalyzed posttranslational modification is common to modular biosynthetic enzymes but rare among cellular proteins as a whole, and thus represents an avenue for their selective labeling, identification, and isolation. This is a particularly salient approach given the known promiscuity many PPTases exhibit towards CoA analogues modified at the terminal thiol. While the ability of Sfp to posttranslationally modify carrier proteins with CoA analogues was first applied for the mechanistic analysis of recombinantly expressed PKS and NRPS enzymes (Sections 5 and 6), employing a similar approach using CoA analogues incorporating fluorescence or affinity tags could allow for the isolation and visualization of PKS and NRPS carrier proteins directly from their producer organisms. Such an approach was first pursued by La Clair *et al*. who described a simple synthesis of fluorescent and affinity-labeled CoA analogues from reporter-labeled maleimides. When added to cell lysate in the presence of Sfp, these analogues demonstrated modification of *apo*-carrier proteins, a modification easily visualized by gel fluorescence scanning or Western blot analysis (Fig. 28).[117] This provided an effective way of monitoring posttranslational modification of recombinant carrier proteins during coexpression with Sfp. Since this initial report, the labeling of *apo*-carrier proteins with reporter-labeled CoA analogues has been used to facilitate detection of carrier protein active sites for FTMS based approaches,[116] as well as in a variety of site-specific protein labeling applications,[77,200,201] the subject of a recent review.[202]

Perhaps more interestingly in the context of the current discussion, this method showed modest labeling of the native DEBS PKS from proteomic preparations of *Saccharopolyspora erythraea*, although this result was noted to be dependent on growth conditions.[117] The low labeling efficiency of DEBS was attributed to the fact that the majority of ACP domains (>95%) from native producer organisms exist in their *holo*-form, posttranslationally modified by their endogenous PPTase and acetyl CoA. Upon cell lysis of a natural product producer of interest, this modification blocks chemoenzymatic labeling by reporter-labeled CoAs and Sfp, as no substrate (*apo*-ACP) is present! As this greatly reduces the utility of PPTase-based chemoenzymatic labeling protocols in reverse genetics approaches to the analysis of PKS and NRPS systems, several approaches have been pursued to reverse the PPTase reaction

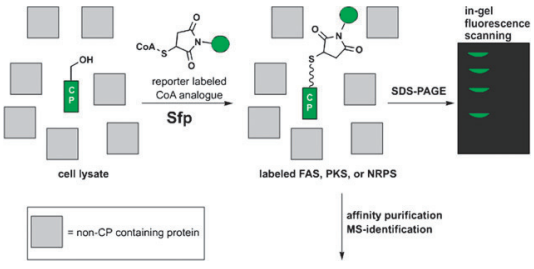


Fig. 28 General strategy for *in vitro* labeling of CP domains in proteomic preparations with reporter-labeled CoA analogues. Methods for the specific labeling of CP-containing enzymes would facilitate reverse genetics approach to the identification of FAS, PKS, and NRPS proteins.

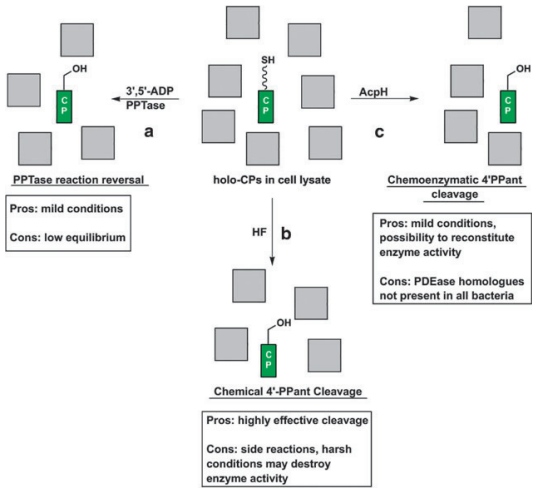


Fig. 29 Strategies for production of *apo*-CPs in proteomic preparations to enable detection of FAS, PKS, and NRPS enzymes by *in vitro* labeling approaches. (a) Reversability of PPTase reaction. (b) Chemical cleavage of 4′-phosphopantetheine prosthetic group. (c) Use of AcpH to produce *apo*-CP. The production of *apo*-CPs from crude cell lysates would greatly increase the utility of *in vitro* CP labeling approaches.

and convert proteomic preparations of *holo*-carrier proteins to *apo*-carrier proteins *in vitro* (Fig. 29).

Many of these approaches are based on observations made in the initial studies of *E. coli* fatty acid biosynthesis in the 1960s regarding the reversibility of the PPTase reaction, chemical methods for 4′-phosphopantetheine cleavage, and the existence of an enzyme (acyl carrier protein hydrolase, AcpH) whose catalytic function is to cleave the 4′-phosphopantetheine group for production of *apo*-ACP. The reversibility of the PPTase reaction was initially investigated by Elovson and Vagelos, who observed that when *holo*-ACP radiolabeled in the 4′-phosphopantetheine arm was treated with a 40-fold excess of 3′,5′-ADP and the PPTase, a small percentage (~3.6%) of the radiolabel became acid soluble, presumably due to formation of radiolabeled CoA and, concurrently, *apo*-ACP.[44] However, more recent studies have not found this method viable for the production of *apo*-ACPs capable of being labeled by Sfp and reporter-labeled CoAs at detectable levels.[203] Also reported in this same era was chemical cleavage of the ACP prosthetic group through the use of HF.[204] Although this specific approach is too harsh for application to proteomic preparations of natural product producers, chemical approaches to cleavage of the phosphodiester bond remain a potentially promising approach for production of *apo*-carrier proteins *in vitro*.

A more useful enzymatic approach to *apo*-carrier proteins may be provided by the recent cloning of an acyl carrier protein hydrolase (also known as ACP phosphodiesterase), AcpH, from *E. coli*.[205] Although its physiological role is still unclear, this enzyme was shown to remove radiolabeled 4′-phosphopantetheine groups from the *E. coli* ACP, as well as from a number of related fatty acid ACPs *in vitro* (Fig. 29).

One potential limitation of this method was the finding that the true AcpH gene product has homologues only in a subset of Gram negative bacterial species. This is puzzling, as AcpH activity has been reported in a number of other organisms, including rat and pigeon.[206,207] Also, the best substrates for AcpH were found to be ACPs which are also substrates for the primary PPTase AcpS,[205] making it unlikely this particular enzyme would be useful in the study of NRPS enzymes. Still, the application of AcpH for the production of *apo*-ACPs from FAS and PKS systems represents an intriguing avenue for future use in combination with proteomic CP labeling methods.

7.2 *In vivo* labeling of carrier proteins

An alternative approach to the labeling of FAS, PKS, and NRPS enzymes in natural product producer organisms through the PPTase reaction would be to apply reporter-labeled CoA analogues *in vivo*. Incorporation of these CoA analogues into the intracellular CoA pool would allow them to access and modify *apo*-carrier proteins directly as they came off the ribosomal assembly line, for later visualization after cell lysis and SDS-PAGE. Unfortunately this approach is hindered by the strongly charged nature of CoA, which prevents it from crossing the cell membrane. Thus, any attempts to manipulate the intracellular CoA pool *in vivo* must utilize CoA precursors.

In an earlier section, we described the *in vivo* delivery of a radiolabeled CoA precursor, 1-^{14}C-pantothenate, to the 4′-phosphopantetheine prosthetic group of *E. coli* ACP by Alberts and Vagelos.[42] This implies metabolic conversion of radiolabeled pantothenate to CoA *via* the five-step enzymatic pathway pictured in Fig. 30,[208] followed by PPTase-catalyzed loading of the ACP. While radiolabeled CoA precursors have emerged as valuable tools for the study of CoA and fatty acid biosynthetic processes, these studies mainly utilize bacterial auxotrophs deficient in pantothenate biosynthesis.[209,210] Also, at least one series of functionalized CoA precursors (the fatty alkyl pantothenate analogues known as pantothenamides) have demonstrated antibiotic activity, partially due to labeling of the bacterial FAS ACP and interference with fatty acid biosynthesis.[211,212] In contrast, an ideal CoA precursor for use in studying carrier-protein mediated biosynthesis would be non-toxic, applicable to genetically unmodified natural product producers, and contain functionalities compatible with fluorescence and immunoprecipitation techniques for facile visualization and isolation of labeled proteins.

The first indication such an approach may be feasible was provided by Clarke and co-workers in 2005.[213] Using the observation that CoA analogues could be produced *in vitro* by a truncated CoA biosynthetic pathway consisting of PanK, DPCK, and PPAT, a fluorescent CoA precursor (**28**) was designed and synthesized for evaluation as an *in vivo* carrier protein label. While this compound did not show labeling of carrier proteins in native bacterial organisms, its application to genetically modified *E. coli* which were actively overexpressing both a carrier protein and PPTase resulted in fluorescent labeling of the overexpressed carrier protein. This result indicated fluorescent **28** was compatible with both uptake and processing by the native CoA biosynthetic pathway of

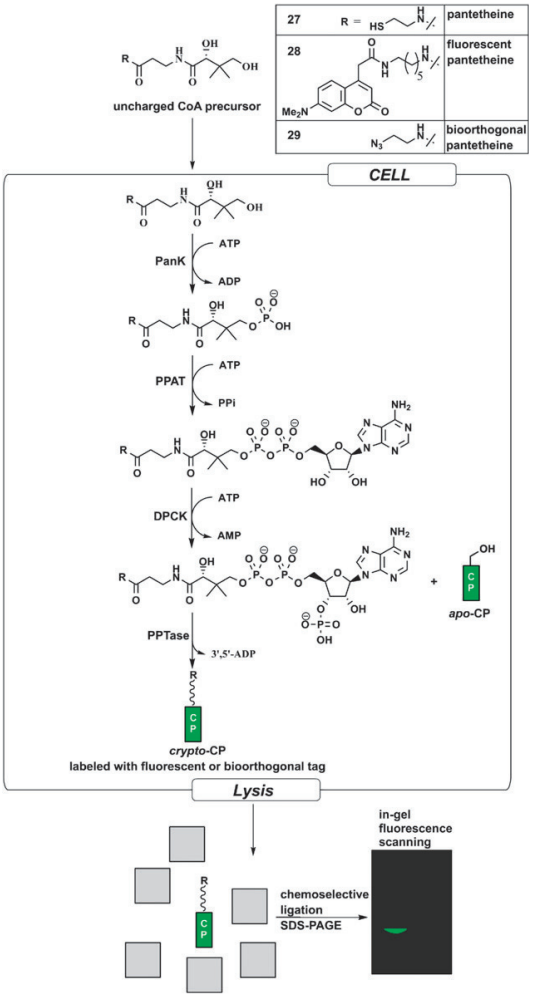


Fig. 30 *In vivo* metabolic labeling of CP domains by uptake and biosynthetic processing of cell permeable CoA precursors. Both pantetheine (**27**) and biodetectable pantetheine analogues (**28**, **29**) can act as substrates for the truncated CoA biosynthetic pathway *in vivo* and form CoA analogues which are appended to *apo*-CPs by a PPTase. After cell lysis the labeled CP can be visualized by SDS-PAGE either directly (**28**) or following chemoselective ligation to a bioorthogonal reporter molecule (**29**).

E. coli, although the fluorescent CoA was apparently *not* produced at high enough concentrations *in vivo* to compete with endogenous CoA for PPTase labeling of non-overexpressed carrier proteins.[213] Importantly, this was the first demonstration of intracellular biosynthesis of a biodetectable CoA analogue for use in protein labeling methodologies.

One reason for the low intracellular concentrations of the fluorescent CoA analogue produced in this study may have been due to the low uptake of the CoA precursor, estimated at only ~4% in *E. coli* cultures grown with 1 mM of the fluorescent pantothenate. In order to improve uptake and metabolic processing of CoA precursors, the group of Burkart

next investigated the utilization of bioorthogonal chemical reporters in their *in vivo* carrier protein labeling approach.[75] Bioorthogonal chemical reporters are functional groups such as azides, alkynes, and ketones which are not normally found in cellular contexts and are relatively unreactive with commonly occurring biological functionalities.[214] However, these functionalities contain an intrinsic reactivity which is triggered upon exposure to a cognate chemical partner, leading to chemoselective ligation of the two groups. A classic example of a bioorthogonal chemical reporter is the azide functionality, which is biologically unreactive and can be metabolically incorporated into biological polymers for their later visualization by reaction with fluorophore-linked triphenylphosphines or alkynes, neither of which are commonly found in biological systems.[215,216] The advantage of this two-step labeling strategy lies in the small size of the bioorthogonal functionality compared to traditional affinity and fluorescence agents, which can result in better uptake and more efficient processing by endogenous biosynthetic enzymes in metabolic labeling strategies.

This led to synthesis of a small panel of bioorthogonal CoA precursors incorporating azide, alkyne, and ketone functionalties, which were then examined for metabolic labeling in the *E. coli* overexpression system (Fig. 29, compound **29**). Upon cell lysis and reaction with their cognate biotin-linked chemical reporter, several of these analogues showed carrier protein labeling as detected by Western blot. Use of the Cu-catalyzed azide–alkyne [3 + 2] cycloaddition proved an especially efficient method for detection of labeled carrier proteins.[217] In addition, kinetic analyses were performed on a variety of biodetectable CoA precursors with PanK, the rate-limiting enzyme for CoA biosynthesis *in vivo*. This allowed identification of structural characteristics of CoA precursors crucial to biosynthetic processing and *in vivo* labeling efficiency. Notably, many of the bioorthogonal pantothenate analogues were processed by CoA with improved catalytic efficiency relative to the bulkier fluorescent analogues, validating this two-step labeling strategy.[75]

Most recently, this same bioorthogonal labeling approach was applied to the labeling of endogenous ACPs in native, non-genetically modified natural product producers. By growth in media supplemented with an azide incorporating CoA precursor, the endogenous fatty acid ACPs of a number of bacterial organisms, including *E. coli*, *Schewanella onedensis*, and *B. subtilis*, could be visualized post-lysis by chemoselective ligation to a fluorophore-alkyne.[218] Alternatively, ligation to a biotin-alkyne could be used for affinity purification of ACPs. Despite labeling the fatty acid ACP of primary metabolism, these analogues were found to be non-toxic, a result which has led to a reevaluation of the antibiotic activity of the pantothenamides.[219] Perhaps most importantly, this method was used to facilitate a reverse genetics approach to cloning of an ACP from an unsequenced organism (Fig. 31). Metabolic labeling of the unsequenced bacterium *B. brevis* 26A1 with a bioorthogonal pantothenate analogue allowed visualization of an ~9 kDa protein following chemoselective ligation to a fluorophore-alkyne and SDS-PAGE. The fluorescent band was excised, analyzed by tandem MS analysis, and the resulting peptides screened against a database of known bacterial fatty acid ACPs. The identification of a single peptide facilitated the design of degenerate primers for use in arbitrary PCR, ultimately leading to cloning of the ACP.[218] While only FAS carrier proteins have so far been reported by this method, its future application in combination

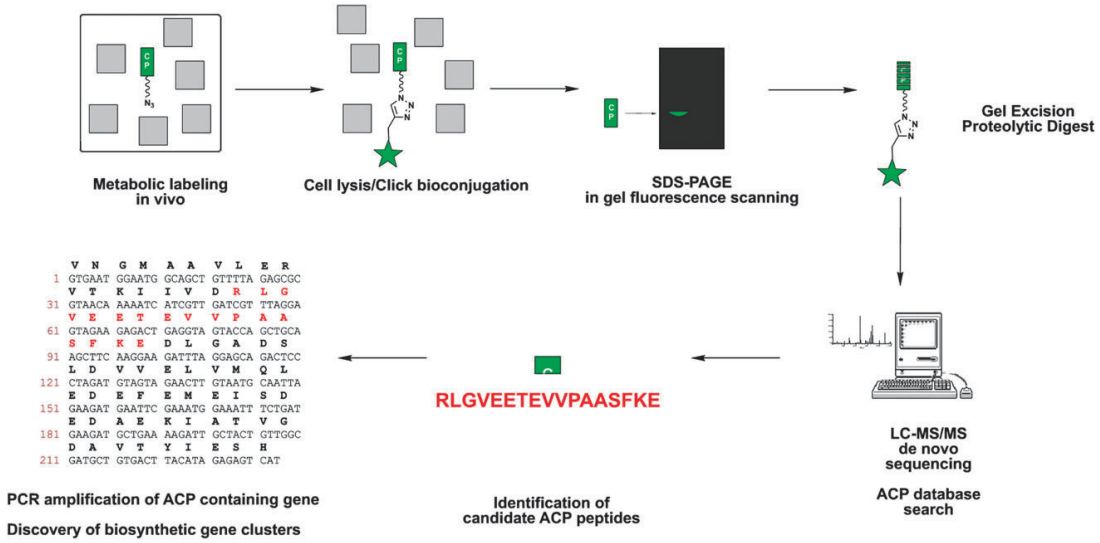


Fig. 31 A reverse genetics approach to identification and cloning of a fatty acid ACP. Following metabolic labeling by azide **29**, the natural product producer organism can be lysed and azide-labeled CP domains detected by chemoselective ligation to a bioorthogonal reporter followed by SDS-PAGE. Excision of the labeled band from the gel followed by tryptic digest allows *de novo* sequencing of the carrier protein containing peptide by tandem MS. These peptides are searched against a database of known ACPs, and used to generate primers which facilitate the cloning of the carrier protein gene by PCR.

with metabolomic analysis of natural product production may be a promising approach to the discovery and visualization of active PKS and NRPS enzymes.

7.3 Active site profiling of modular biosynthetic enzymes

In addition to their use of carrier proteins, another unique characteristic of FAS, PKS, and NRPS multienzymes is the presence of multiple active sites on a single polypeptide. Theoretically this attribute could also be used to distinguish modular synthases from the proteomic mileau and facilitate their identification and analysis (Fig. 32). This approach was examined in a preliminary study by Meier *et al.*, in which fluorescent analogues of known KS and TE affinity labels were used in combination with carrier protein labeling methods (described above) to investigate a number of FAS, PKS, and NRPS systems.[220] First examining a panel of fluorescent KS and TE affinity agents *in vitro*, it was found that reporter-labeled fluorophosphonates showed excellent labeling of TE but not AT-type serine hydrolases in PKS systems. This is in agreement with the earlier observation of Foster *et al.* (Section 6.2) concerning the specificity of these reagents against type I mammalian FAS.[158] Haloacetamides were found to show the best, although relatively modest, labeling of PKS and

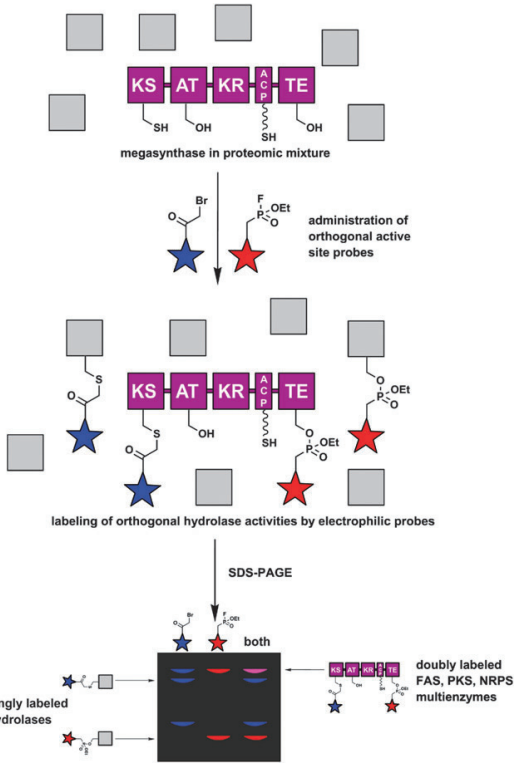


Fig. 32 Strategy for labeling type I FAS, PKS, and NRPS enzymes based on their multienzymatic nature using orthogonal active site probes. While such active site probes will label multiple proteomic enzymes, the use of well-designed probes can ensure only FAS, PKS, and NRPS enzymes are multiply labeled, facilitating their identification.

FAS KS domains among the reagents surveyed. Perhaps more notably, the application of these reagents to proteomic extracts was shown to facilitate identificiaton of type I FAS and NRPS enzymes. For example, application of a fluorescent fluorophosphonate to lysate from *B. subtilis* 168 showed labeling of a multitude of proteins, presumably serine hydrolases, consistent with the known utility of this activity-based protein profiling reagent.[221,222] However, upon treatment of the same lysate with a reporter-labeled CoA analogue and the PPTase Sfp, only a single band corresponding to a serine hydrolase was labeled, suggesting this protein was a carrier protein–TE containing multienzyme. Inspection of the sequenced *B. subtilis* genome indicated the most likely PKS–NRPS candidate in the observed molecular weight range was SrfAC, the terminal module of the surfactin NRPS, a conclusion confirmed by tandem MS identification of the excised band.

While this initial study was facilitated both by the ease of carrier protein labeling in this particular organism as well the availability of genomic information, this method is promising in its ability to simplify complex protein labeling patterns generated by orthogonal protein labels into a subset of candidate PKS and NRPS proteins. The development of probes targeting alternate active sites of modular biosynthetic enzymes will be necessary both to uncouple this strategy from carrier protein labeling methods (and their current shortcomings) as well as for its application to non-terminal modules of PKS and NRPS megasynthases. The recent development of reporter-labeled inhibitors targeting CoA acyltransferases,[223,224] NRPS A domains,[162] and bacterial FAS KS-domains[225] bodes well for the future application of active site probes to natural product proteomics (Fig. 33).

7.4 Mass spectrometry based approaches

A limited number of whole proteome, mass spectrometry (MS) studies of sequenced natural product producers have been performed, providing insight into many cellular processes including regulation of secondary metabolism.[226–228] In contrast, the specific study of PKS and NRPS biosynthetic enzymes isolated from whole proteomes by MS represents a field which is largely undeveloped, but extremely promising. Most encouraging on this front are recently described tandem MS techniques which have the capacity to specifically observe the 4′-phosphopantetheine modification of carrier proteins by specific fragmentation of this posttranslational modification in the gas phase (Fig. 14b).[114,118] While this "phosphopantetheinylation ejection assay" was initially reported on a high-end, custom-built FTMS instrument, a recent report by Melluzi *et al.* extended the use of this method to commercially available linear quadripole ion trap (LTQ) utilizing mass spectrometers.[119] LTQ-MS instruments are readily available and commonly utilized for proteomic analyses, but can lack the high resolution and signal to noise of their FTMS counterparts that is crucial for observation of the 4′-phosphopantetheine posttranslational modification. In order to increase the utility of this method on lower resolution MS instruments, it was investigated whether subjection of the putative phosphopantetheine peak to a third round of fragmentation

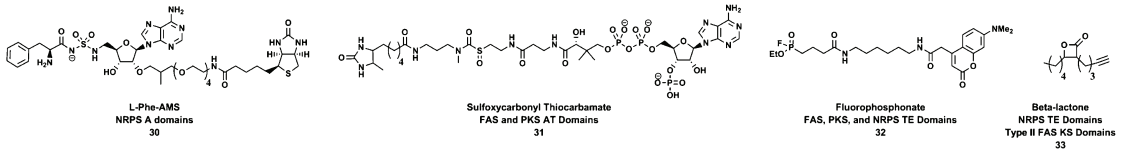


Fig. 33 Examples of active site probes which have been either directly applied (**30**, **32**, **33**) or are good candidates (**31**) for the study of activities commonly found in FAS, PKS, and NRPS enzymes.

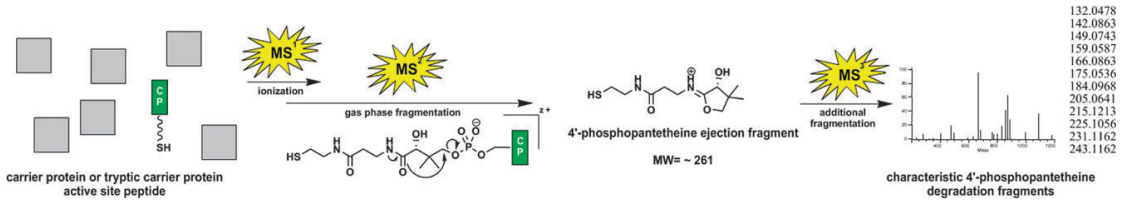


Fig. 34 MS^3 identification of CP-active site peptides. The carrier protein or a tryptic fragment containing the 4′-phosphopantetheine moiety is fragmented in the gas phase to yield a 4′-phosphopantetheine ejection fragment, which can be further fragmented through use of an ion trap to yield a characteristic degradation spectrum. While this approach has to date only been applied *in vitro*, its use in cell-free extracts or in combination with *in vivo* labeling strategies could prove a boon to reverse genetics approaches to the discovery and identification of PKS and NRPS enzymes.

(MS/MS/MS or MS^3) could be used to unambigously identify this modification. Remarkably, fragmentation of the 4′-phosphopantetheine moiety led to the observation of 12 associated ions. Based on molecular formula analysis nine of these ions were attributed to fragments which contained the phosphopantetheine thiol and three were characteristic of the pantothenate terminus (Fig. 34).[119] These three "pantothenate" MS^3 fragments are particularly notable, as they provide a characteristic MS^3 signature for the 4′-phosphopantetheine modification. While this method has to date only been utilized for the study of recombinant ACP and PCP domains *in vitro*, its future optimization for use in proteomic preparations, particularly when applied in combination with the aforementioned carrier protein labeling techniques, may prove an extremely robust method for identification of PKS and NRPS proteins in unsequenced organisms.

7.5 Genome mining by phage display

Though not a proteomic method in the strict sense, the chemoenzymatic labeling of CP domains has also been applied to the identification of PKS and NRPS gene clusters through a phage display method.[229] This approach involves random fragmentation of an organism's genomic DNA and cloning of these fragments into phagemid vectors, allowing a phage library to be constructed of appropriate size for display of all possible proteins of an organism. FAS, PKS, and NRPS carrier protein utilizing enzymes are then selected for by treatment of the phage library with biotin-CoA and Sfp, resulting in biotin modification of *apo*-carrier proteins which are displayed on the phage particle surface. Repeated cycles of selection allow preferential enrichment of carrier protein encoding phages, which can then be used to infect a bacterial host and the phagemid sequenced to determine the identity of the genomically encoded carrier protein and potentially facilitate cloning of PKS or NRPS gene clusters (Fig. 35).

To explore the feasibility of this method it was first applied to genomic DNA from the prototypical bacteria *B. subtilis*, a sequenced organism and well-characterized producer of PKS and NRPS derived metabolites.[230] Several rounds of selection of phages encoding *B. subtilis* genome fragments through Sfp–biotin-CoA modification allowed substantial enrichment of phage particles. Sequencing of phagemids from infected *E. coli* clones showed 85% to contain DNA coding for *apo*-carrier proteins from *B. subtilis* PKS or NRPS systems, providing sequence data for an overall 17 out of 39 PKS and NRPS carrier proteins. Interestingly, only PKS and NRPS carrier proteins were observed, presumably due to endogenous posttranslational modification of the phage protein encoding the *B. subtilis* FAS ACP by the primary PPTase of *E. coli*. This approach was also applied to the recently sequenced myxobacterium *Myxococcus xanthus*, allowing enrichment of a multitude of phages encoding DNA fragments annotated in the *M. xanthus* genome as PKS and NRPS carrier proteins. A caveat to this approach was the finding that from both organisms some phages came to dominate the library through repeated cycles of selection, presumably due to preferential reactivity with Sfp in the PPTase reaction. It was proposed that this bias could be guarded against by the use of multiple PPTases, as they are known to show distinct substrate preferences, and careful monitoring of the selection process itself.[229]

Another notable finding of this study was the enrichment of phages from both *B. subtilis* and *M. xanthus* shotgun libraries using Sfp–biotin-CoA selection which showed no homology to known PKS and NRPS carrier proteins. Two of these proteins from *B. subtilis* were expressed as full-length constructs and demonstrated posttranslational modification upon administration of CoA and Sfp. While the identification of such enzymes may provide unwanted background in the search for PKS and NRPS catalysts, this demonstrates the utility of phage display methods are for identification of unexpected

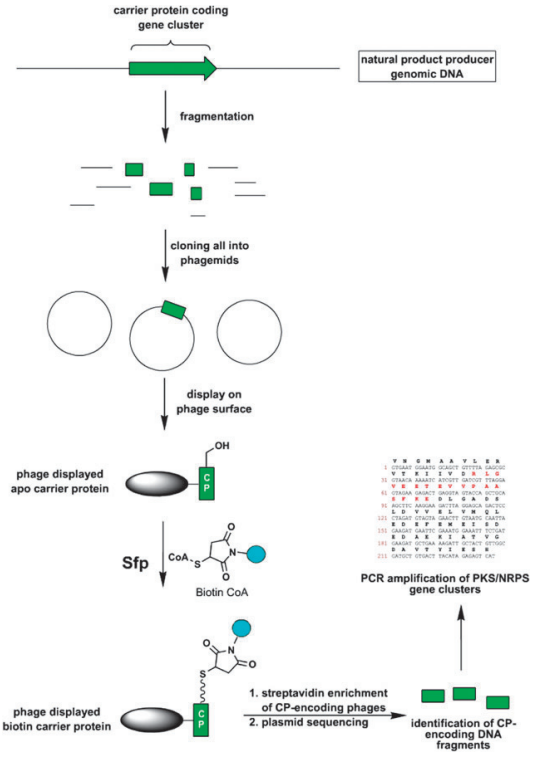


Fig. 35 Discovery of PKS and NRPS gene clusters by phage display. A shotgun library of fragmented genomic DNA from a natural product producing organism is cloned into phagemids and displayed on a phage capsid. Fragments coding for carrier proteins can be selectively enriched using Biotin-CoA and Sfp, facilitating their identification and cloning by PCR.

posttranslational modifications and carrier proteins of low sequence homology. These unexpected targets of PPTase modification identified by phage display have also led to the unexpected discovery of minimal peptide sequences for use in fusion protein labeling applications.[201,231]

Comparatively, the use of PPTase enrichment with phage display and proteomic identification of PKS and NRPS proteins are complementary in many ways. Phage display may be more useful in profiling the overall biosynthetic potential of an organism, as it studies a library of all the possible proteins encoded by an organism, providing it with an advantage in the discovery and analysis of normally ''silent'' gene clusters. Also, since it does not rely on culture of a natural product producer this approach may prove useful in the study of metagenomic preparations from environmental samples.[232] In contrast, proteomic methods may prove more useful in identification of specific modular biosynthetic enzymes which encode a natural product of interest, provided an organism is culturable, as they can be applied in combination with monitoring of natural product production. Nevertheless, phage display is a prototypical example of an emerging area where chemical and biological techniques are being used in concert to explore previously cryptic topics in natural product biosynthesis.

8. Conclusions

When looking back at the difficulties encountered during the initial elucidation of the structure and metabolic role of CoA, Fritz Lipmann, one of the progenitors of the NRPS field, remarked that, ''All this sums up to teaching how difficult it is to see the new because it is new, and how badly one may be handicapped by preconceived notions.''[233] A similar thought might be expressed concerning how easy it is to overlook the past simply because it is the past. Although this review has covered a wide variety of topics, one unifying theme lies in the ability of classical chemical biological methods to be updated and recombined with more recent innovations in seemingly unrelated fields to yield fresh insights into modular biosynthesis. In such a manner did knowledge of PPTase promiscuity together with advances in solid-phase peptide synthesis allow peptidyl CoA analogues to be used as probes of NRPS condensation mechanism. Similarly, chemical crosslinking of ACP and KS domains by bifunctional electrophiles was revised in the context of the newfound ability to site-specifically modify ACPs to provide chemoenzymatic crosslinking approaches for use in structure elucidation, and basic research into CoA biosynthesis provided insights which were used with recently developed metabolic labeling and chemoselective ligation strategies to allow *in vivo* labeling and identification of fatty acid ACPs from unsequenced organisms. By seeing the new in not only the present, but also the past, the future application of chemical biological techniques to study natural product biosynthesis appears bright indeed.

Acknowledgements

The authors wish to thank Andrew Worthington, Gene Hur, Jillian Blatti, Joseph Lipson, and Michael Rothmann of the Burkart lab for careful reading of the manuscript and several helpful suggestions. This work was supported by NIH grant RO1GM075797.

Notes and references

1 S. Smith and S. C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.
2 M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.
3 S. A. Sieber and M. A. Marahiel, *Chem. Rev.*, 2005, **105**, 715–738.
4 K. J. Weissman and R. Muller, *ChemBioChem*, 2008, **9**, 826–848.
5 A. M. Hill, *Nat. Prod. Rep.*, 2006, **23**, 256–320.
6 R. McDaniel, M. Welch and C. R. Hutchinson, *Chem. Rev.*, 2005, **105**, 543–558.
7 Y. M. Zhang, S. W. White and C. O. Rock, *J. Biol. Chem.*, 2006, **281**, 17541–17544.
8 A. Cropp, S. Chen, H. Liu, W. Zhang and K. A. Reynolds, *J. Ind. Microbiol. Biotechnol.*, 2001, **27**, 368–377.
9 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936.
10 J. Kennedy, *Nat. Prod. Rep.*, 2008, **25**, 25–34.
11 T. P. Begley, *Nat. Chem. Biol.*, 2005, **1**, 236–238.
12 A. J. Birch, R. A. Massy-Westropp and C. J. Moye, *Aust. J. Chem.*, 1955, **8**, 539–544.
13 J. K. Stoops and S. J. Wakil, *J. Biol. Chem.*, 1981, **256**, 5128–5133.
14 F. Lynen, *Federation Proc.*, 1961, **20**, 941–951.
15 A. T. Norris, S. Matsumura and K. Bloch, *J. Biol. Chem.*, 1964, **239**, 3653–3662.
16 B. S. Moore and C. Hertweck, *Nat. Prod. Rep.*, 2002, **19**, 70–99.

17 C. Hertweck, A. Luzhetskyy, Y. Rebets and A. Bechthold, *Nat. Prod. Rep.*, 2007, **24**, 162–190.
18 J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.
19 R. Bentley and J. W. Bennett, *Annu. Rev. Microbiol.*, 1999, **53**, 411–446.
20 E. A. Felnagle, E. E. Jackson, Y. A. Chan, A. M. Podevels, A. D. Berti, M. D. McMahon and M. G. Thomas, *Mol. Pharmacol.*, 2008, **5**, 191–211.
21 G. B. Kresze, D. Oesterhelt, F. Lynen, H. Castorph and E. Schweizer, *Biochem. Biophys. Res. Commun.*, 1976, **69**, 893–899.
22 P. Goldman and P. R. Vagelos, *Biochem. Biophys. Res. Commun.*, 1962, **7**, 414–418.
23 W. J. Lennarz, R. J. Light and K. Bloch, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 840–846.
24 P. W. Majerus, A. W. Alberts and P. R. Vagelos, *Proc. Natl. Acad. Sci. U. S. A.*, 1964, **51**, 1231–1238.
25 S. J. Wakil, E. L. Pugh and F. Sauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1964, **52**, 106–114.
26 F. Sauer, E. L. Pugh, S. J. Wakil, R. Delaney and R. L. Hill, *Proc. Natl. Acad. Sci. U. S. A.*, 1964, **52**, 1360–1366.
27 P. W. Majerus, A. W. Alberts and P. R. Vagelos, *Proc. Natl. Acad. Sci. U. S. A.*, 1965, **53**, 410–417.
28 A. R. Larrabee, E. G. McDaniel, H. A. Bakerman and P. R. Vagelos, *Proc. Natl. Acad. Sci. U. S. A.*, 1965, **54**, 267–273.
29 P. Dimroth, H. Walter and F. Lynen, *Eur. J. Biochem.*, 1970, **13**, 98–110.
30 P. Dimroth, G. Greull, R. Seyffert and F. Lynen, *Hoppe–Seyler Z. Physiol. Chem.*, 1972, **353**, 126.
31 D. A. Hopwood, *Chem. Rev.*, 1997, **97**, 2465–2498.
32 B. Mach, E. Reich and E. L. Tatum, *Proc. Natl. Acad. Sci. U. S. A.*, 1963, **50**, 175–181.
33 T. L. Berg, L. O. Froholm and S. G. Laland, *Biochem. J.*, 1965, **96**, 43–52.
34 H. Kleinkauf, W. Gevers and F. Lipmann, *Proc. Natl. Acad. Sci. U. S. A.*, 1969, **62**, 226–233.
35 W. Gevers, H. Kleinkauf and F. Lipmann, *Proc. Natl. Acad. Sci. U. S. A.*, 1969, **63**, 1335–1342.
36 O. Froshov, T. L. Zimmer and S. G. Laland, *FEBS Lett.*, 1970, **7**, 68–71.
37 C. C. Gilhuus-Moe, T. Kristensen, J. E. Bredesen, T. Zimmer and S. G. Laland, *FEBS Lett.*, 1970, **7**, 287–290.
38 R. Roskoski, Jr, W. Gevers, H. Kleinkauf and F. Lipmann, *Biochemistry*, 1970, **9**, 4839–4845.
39 H. Kleinkauf, W. Gevers, R. Roskoski, Jr and F. Lipmann, *Biochem. Biophys. Res. Commun.*, 1970, **41**, 1218–1222.
40 S. G. Laland and T. L. Zimmer, *Essays Biochem.*, 1973, **9**, 31–57.
41 T. Stein, J. Vater, V. Kruft, A. Otto, B. Wittmann-Liebold, P. Franke, M. Panico, R. McDowell and H. R. Morris, *J. Biol. Chem.*, 1996, **271**, 15428–15435.
42 A. W. Alberts and P. R. Vagelos, *J. Biol. Chem.*, 1966, **241**, 5201–5204.
43 D. A. Roncari, R. A. Bradshaw and P. R. Vagelos, *J. Biol. Chem.*, 1972, **247**, 6234–6242.
44 J. Elovson and P. R. Vagelos, *J. Biol. Chem.*, 1968, **243**, 3603–3611.
45 E. Pfeifer, M. Pavela-Vrancic, H. von Dohren and H. Kleinkauf, *Biochemistry*, 1995, **34**, 7450–7459.
46 R. H. Lambalot and C. T. Walsh, *J. Biol. Chem.*, 1995, **270**, 24658–24661.
47 C. W. Carreras, A. M. Gehring, C. T. Walsh and C. Khosla, *Biochemistry*, 1997, **36**, 11757–11761.
48 A. M. Gehring, R. H. Lambalot, K. W. Vogel, D. G. Drueckhammer and C. T. Walsh, *Chem. Biol.*, 1997, **4**, 17–24.
49 T. Stachelhaus, A. Huser and M. A. Marahiel, *Chem. Biol.*, 1996, **3**, 913–921.
50 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923–936.
51 J. Ku, R. G. Mirmira, L. Liu and D. V. Santi, *Chem. Biol.*, 1997, **4**, 203–207.
52 L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585–1595.
53 A. C. Mercer and M. D. Burkart, *Nat. Prod. Rep.*, 2007, **24**, 750–773.
54 Y. Kim and J. H. Prestegard, *Proteins: Struct., Funct., Genet.*, 1990, **8**, 377–385.
55 M. P. Crump, J. Crosby, C. E. Dempsey, J. A. Parkinson, M. Murray, D. A. Hopwood and T. J. Simpson, *Biochemistry*, 1997, **36**, 6000–6008.
56 T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel and T. A. Holak, *Structure (London)*, 2000, **8**, 407–418.
57 A. Roujeinikova, C. Baldock, W. J. Simon, J. Gilroy, P. J. Baker, A. R. Stuitje, D. W. Rice, A. R. Slabas and J. B. Rafferty, *Structure (London)*, 2002, **10**, 825–835.
58 J. R. Lai, A. Koglin and C. T. Walsh, *Biochemistry*, 2006, **45**, 14869–14879.
59 K. Magnuson, S. Jackowski, C. O. Rock and J. E. Cronan, Jr, *Microbiol. Rev.*, 1993, **57**, 522–542.
60 A. Roujeinikova, W. J. Simon, J. Gilroy, D. W. Rice, J. B. Rafferty and A. R. Slabas, *J. Mol. Biol.*, 2007, **365**, 135–145.
61 V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi and C. Khosla, *Protein Sci.*, 2007, **16**, 2093–2107.
62 A. Koglin, M. R. Mofid, F. Lohr, B. Schafer, V. V. Rogov, M. M. Blum, T. Mittag, M. A. Marahiel, F. Bernhard and V. Dotsch, *Science*, 2006, **312**, 273–276.
63 T. Maier, S. Jenni and N. Ban, *Science*, 2006, **311**, 1258–1262.
64 I. B. Lomakin, Y. Xiong and T. A. Steitz, *Cell (Cambridge, Mass.)*, 2007, **129**, 319–332.
65 S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolasek and N. Ban, *Science*, 2007, **316**, 254–261.
66 Y. Tang, A. Y. Chen, C. Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 931–943.
67 A. Tanovic, S. A. Samel, L. O. Essen and M. A. Marahiel, *Science*, 2008.
68 M. Leibundgut, S. Jenni, C. Frick and N. Ban, *Science*, 2007, **316**, 288–290.
69 K. J. Weissman, H. Hong, M. Oliynyk, A. P. Siskos and P. F. Leadlay, *ChemBioChem*, 2004, **5**, 116–125.
70 C. Chalut, L. Botella, C. de Sousa-D'Auria, C. Houssin and C. Guilhot, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 8511–8516.
71 C. Sanchez, L. Du, D. J. Edwards, M. D. Toney and B. Shen, *Chem. Biol.*, 2001, **8**, 725–738.
72 R. C. Smith, K. N. Wills, D. Antelman, H. Perlman, L. N. Truong, K. Krasinski and K. Walsh, *Circulation*, 1997, **96**, 1899–1905.
73 R. Finking, J. Solsbacher, D. Konz, M. Schobert, A. Schafer, D. Jahn and M. A. Marahiel, *J. Biol. Chem.*, 2002, **277**, 50293–50302.
74 F. Fichtlscherer, C. Wellein, M. Mittag and E. Schweizer, *Eur. J. Biochem.*, 2000, **267**, 2666–2671.
75 J. L. Meier, A. C. Mercer, H. Rivera, Jr and M. D. Burkart, *J. Am. Chem. Soc.*, 2006, **128**, 12174–12184.
76 S. A. Sieber, C. T. Walsh and M. A. Marahiel, *J. Am. Chem. Soc.*, 2003, **125**, 10862–10866.
77 N. George, H. Pick, H. Vogel, N. Johnsson and K. Johnsson, *J. Am. Chem. Soc.*, 2004, **126**, 8896–8897.
78 F. Vitali, K. Zerbe and J. A. Robinson, *Chem. Commun.*, 2003, 2718–2719.
79 K. Reuter, M. R. Mofid, M. A. Marahiel and R. Ficner, *EMBO J.*, 1999, **18**, 6823–6831.
80 J. E. Cronan, Jr and A. L. Klages, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 5440–5444.
81 J. Crosby, K. J. Byrom, T. S. Hitchman, R. J. Cox, M. P. Crump, I. S. Findlow, M. J. Bibb and T. J. Simpson, *FEBS Lett.*, 1998, **433**, 132–138.
82 C. Khosla, R. S. Gokhale, J. R. Jacobsen and D. E. Cane, *Annu. Rev. Biochem.*, 1999, **68**, 219–253.
83 H. Krebs and K. Decker, *Biogr. Mem. Fellows R. Soc.*, 1982, **28**, 261.
84 K. Woithe, N. Geib, K. Zerbe, D. B. Li, M. Heck, S. Fournier-Rousset, O. Meyer, F. Vitali, N. Matoba, K. Abou-Hadeed and J. A. Robinson, *J. Am. Chem. Soc.*, 2007, **129**, 6887–6895.
85 Y. Ding, W. H. Seufert, Z. Q. Beck and D. H. Sherman, *J. Am. Chem. Soc.*, 2008, **130**, 5492–5498.
86 M. Fridman, C. J. Balibar, T. Lupoli, D. Kahne, C. T. Walsh and S. Garneau-Tsodikova, *Biochemistry*, 2007, **46**, 8462–8471.

87 P. J. Belshaw, C. T. Walsh and T. Stachelhaus, *Science*, 1999, **284**, 486–489.
88 T. Stachelhaus, H. D. Mootz, V. Bergendahl and M. A. Marahiel, *J. Biol. Chem.*, 1998, **273**, 22773–22781.
89 U. Linne and M. A. Marahiel, *Biochemistry*, 2000, **39**, 10439–10447.
90 L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2002, **41**, 9184–9196.
91 S. L. Clugston, S. A. Sieber, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2003, **42**, 12095–12104.
92 C. J. Balibar, F. H. Vaillancourt and C. T. Walsh, *Chem. Biol.*, 2005, **12**, 1189–1200.
93 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evol. Biol.*, 2007, **7**, 78.
94 D. G. McCafferty, P. Cudic, B. A. Frankel, S. Barkallah, R. G. Kruger and W. Li, *Biopolymers*, 2002, **66**, 261–284.
95 X. Yin and T. M. Zabriskie, *Microbiology (Reading, U. K.)*, 2006, **152**, 2969–2983.
96 P. D. Fortin, C. T. Walsh and N. A. Magarvey, *Nature*, 2007, **448**, 824–827.
97 N. Wu, D. E. Cane and C. Khosla, *Biochemistry*, 2002, **41**, 5056–5066.
98 D. Spiteller, C. L. Waterman and J. B. Spencer, *Angew. Chem., Int. Ed.*, 2005, **44**, 7079–7082.
99 M. B. Austin and J. P. Noel, *Nat. Prod. Rep.*, 2003, **20**, 79–110.
100 K. Watanabe, A. P. Praseuth and C. C. Wang, *Curr. Opin. Chem. Biol.*, 2007, **11**, 279–286.
101 A. L. Mandel, J. J. La Clair and M. D. Burkart, *Org. Lett.*, 2004, **6**, 4801–4803.
102 M. van Wyk and E. Strauss, *Chem. Commun.*, 2007, 398–400.
103 F. Kopp and M. A. Marahiel, *Nat. Prod. Rep.*, 2007, **24**, 735–749.
104 C. N. Boddy, T. L. Schneider, K. Hotta, C. T. Walsh and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 3428–3429.
105 W. He, J. Wu, C. Khosla and D. E. Cane, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 391–394.
106 C. C. Aldrich, L. Venkatraman, D. H. Sherman and R. A. Fecik, *J. Am. Chem. Soc.*, 2005, **127**, 8910–8911.
107 B. M. Harvey, H. Hong, M. A. Jones, Z. A. Hughes-Thomas, R. M. Goss, M. L. Heathcote, V. M. Bolanos-Garcia, W. Kroutil, J. Staunton, P. F. Leadlay and J. B. Spencer, *ChemBioChem*, 2006, **7**, 1435–1442.
108 T. Liu, X. Lin, X. Zhou, Z. Deng and D. E. Cane, *Chem. Biol.*, 2008, **15**, 449–458.
109 R. Aggarwal, P. Caffrey, P. F. Leadlay, C. J. Smith and J. Staunton, *J. Chem. Soc., Chem. Commun.*, 1995, **15**, 1519.
110 Y. Yin, H. Lu, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2003, **125**, 5671–5676.
111 Z. Q. Beck, C. C. Aldrich, N. A. Magarvey, G. I. Georg and D. H. Sherman, *Biochemistry*, 2005, **44**, 13457–13466.
112 L. Tran, M. Tosin, J. B. Spencer, P. F. Leadlay and K. J. Weissman, *ChemBioChem*, 2008, **9**, 905–915.
113 S. A. Sieber, J. Tao, C. T. Walsh and M. A. Marahiel, *Angew. Chem., Int. Ed.*, 2004, **43**, 493–498.
114 P. C. Dorrestein and N. L. Kelleher, *Nat. Prod. Rep.*, 2006, **23**, 893–918.
115 C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg and C. T. Walsh, *Chem. Biol.*, 1999, **6**, 385–400.
116 S. M. McLoughlin, M. T. Mazur, L. M. Miller, J. Yin, F. Liu, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2005, **44**, 14159–14169.
117 J. J. La Clair, T. L. Foley, T. R. Schegg, C. M. Regan and M. D. Burkart, *Chem. Biol.*, 2004, **11**, 195–201.
118 P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 12756–12766.
119 D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet and P. C. Dorrestein, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3107–3111.
120 G. J. Gatto, Jr, S. M. McLoughlin, N. L. Kelleher and C. T. Walsh, *Biochemistry*, 2005, **44**, 5993–6002.
121 P. C. Dorrestein, J. Blackhall, P. D. Straight, M. A. Fischbach, S. Garneau-Tsodikova, D. J. Edwards, S. McLaughlin, M. Lin, W. H. Gerwick, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 1537–1546.
122 R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1506–1509.
123 L. M. Hicks, M. T. Mazur, L. M. Miller, P. C. Dorrestein, N. A. Schnarr, C. Khosla and N. L. Kelleher, *ChemBioChem*, 2006, **7**, 904–907.
124 A. C. Mercer, J. J. La Clair and M. D. Burkart, *ChemBioChem*, 2005, **6**, 1335–1337.
125 I. Sielaff, A. Arnold, G. Godin, S. Tugulu, H. A. Klok and K. Johnsson, *ChemBioChem*, 2006, **7**, 194–202.
126 L. Luo, M. D. Burkart, T. Stachelhaus and C. T. Walsh, *J. Am. Chem. Soc.*, 2001, **123**, 11208–11218.
127 D. B. Stein, U. Linne and M. A. Marahiel, *FEBS J.*, 2005, **272**, 4506–4520.
128 U. Linne, S. Doekel and M. A. Marahiel, *Biochemistry*, 2001, **40**, 15824–15834.
129 J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel and C. T. Walsh, *Nature*, 2000, **407**, 215–218.
130 J. W. Trauger, R. M. Kohli and C. T. Walsh, *Biochemistry*, 2001, **40**, 7092–7098.
131 R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2001, **40**, 7099–7108.
132 R. M. Kohli, J. Takagi and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 1247–1252.
133 R. M. Kohli, C. T. Walsh and M. D. Burkart, *Nature*, 2002, **418**, 658–661.
134 D. Garbe, S. A. Sieber, N. G. Bandur, U. Koert and M. A. Marahiel, *ChemBioChem*, 2004, **5**, 1000–1003.
135 H. Lin and C. T. Walsh, *J. Am. Chem. Soc.*, 2004, **126**, 13998–14003.
136 H. Lin, D. A. Thayer, C. H. Wong and C. T. Walsh, *Chem. Biol.*, 2004, **11**, 1635–1642.
137 R. M. Kohli, M. D. Burke, J. Tao and C. T. Walsh, *J. Am. Chem. Soc.*, 2003, **125**, 7160–7161.
138 C. C. Tseng, S. D. Bruner, R. M. Kohli, M. A. Marahiel, C. T. Walsh and S. A. Sieber, *Biochemistry*, 2002, **41**, 13350–13359.
139 C. Mahlert, S. A. Sieber, J. Grunewald and M. A. Marahiel, *J. Am. Chem. Soc.*, 2005, **127**, 9571–9580.
140 J. Grunewald, F. Kopp, C. Mahlert, U. Linne, S. A. Sieber and M. A. Marahiel, *Chem. Biol.*, 2005, **12**, 873–881.
141 Z. Zhou, J. R. Lai and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 11621–11626.
142 M. Mueller, S. Jenni and N. Ban, *Curr. Opin. Struct. Biol.*, 2007, **17**, 572–579.
143 J. J. Volpe and P. R. Vagelos, *Physiol. Rev.*, 1976, **56**, 339–417.
144 J. K. Stoops, P. Ross, M. J. Arslanian, K. C. Aune, S. J. Wakil and R. M. Oliver, *J. Biol. Chem.*, 1979, **254**, 7418–7426.
145 S. Kumar, K. R. Srinivasan and N. Asato, *Biochim. Biophys. Acta*, 1977, **489**, 32–47.
146 J. K. Stoops and S. J. Wakil, *J. Biol. Chem.*, 1982, **257**, 3230–3235.
147 Y. S. Wang, W. X. Tian and R. Y. Hsu, *J. Biol. Chem.*, 1984, **259**, 13644–13647.
148 A. K. Joshi, A. Witkowski and S. Smith, *Biochemistry*, 1998, **37**, 2515–2523.
149 A. Witkowski, A. K. Joshi, V. S. Rangan, A. M. Falick, H. E. Witkowska and S. Smith, *J. Biol. Chem.*, 1999, **274**, 11557–11563.
150 A. Witkowski, A. Ghosal, A. K. Joshi, H. E. Witkowska, F. J. Asturias and S. Smith, *Chem. Biol.*, 2004, **11**, 1667–1676.
151 J. K. Stoops and S. J. Wakil, *J. Biol. Chem.*, 1981, **256**, 8364–8370.
152 J. B. Spencer and P. M. Jordan, *Biochem. J.*, 1992, **288**, 839–846.
153 J. Staunton, P. Caffrey, J. F. Aparicio, G. A. Roberts, S. S. Bethell and P. F. Leadlay, *Nat. Struct. Biol.*, 1996, **3**, 188–192.
154 R. Pieper, R. S. Gokhale, G. Luo, D. E. Cane and C. Khosla, *Biochemistry*, 1997, **36**, 1846–1851.
155 P. Kumar, C. Khosla and Y. Tang, *Methods Enzymol.*, 2004, **388**, 269–293.
156 S. A. Sieber, U. Linne, N. J. Hillson, E. Roche, C. T. Walsh and M. A. Marahiel, *Chem. Biol.*, 2002, **9**, 997–1008.

157 F. J. Asturias, J. Z. Chadick, I. K. Cheung, H. Stark, A. Witkowski, A. K. Joshi and S. Smith, *Nat. Struct. Mol. Biol.*, 2005, **12**, 225–232.
158 R. J. Foster, A. J. Poulose, R. F. Bonsall and P. E. Kolattukudy, *J. Biol. Chem.*, 1985, **260**, 2826–2831.
159 J. W. Cardon and G. G. Hammes, *Biochemistry*, 1982, **21**, 2863–2870.
160 Z. Y. Yuan and G. G. Hammes, *J. Biol. Chem.*, 1986, **261**, 13643–13651.
161 S. I. Chang and G. G. Hammes, *Biochemistry*, 1988, **27**, 4753–4760.
162 R. Finking, A. Neumuller, J. Solsbacher, D. Konz, G. Kretzschmar, M. Schweitzer, T. Krumm and M. A. Marahiel, *ChemBioChem*, 2003, **4**, 903–906.
163 J. A. Ferreras, J. S. Ryu, F. Di Lello, D. S. Tan and L. E. Quadri, *Nat. Chem. Biol.*, 2005, **1**, 29–32.
164 S. Jenni, M. Leibundgut, T. Maier and N. Ban, *Science*, 2006, **311**, 1263–1267.
165 C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr and D. E. Cane, *Annu. Rev. Biochem.*, 2007, **76**, 195–221.
166 W. S. Hancock, D. J. Prescott, G. R. Marshall and P. R. Vagelos, *J. Biol. Chem.*, 1972, **247**, 6224–6233.
167 A. S. Worthington, H. Rivera, J. W. Torpey, M. D. Alexander and M. D. Burkart, *ACS Chem. Biol.*, 2006, **1**, 687–691.
168 A. S. Worthington and M. D. Burkart, *Org. Biomol. Chem.*, 2006, **4**, 44–46.
169 A. S. Worthington, G. H. Hur, J. L. Meier, Q. Cheng, B. S. Moore and M. D. Burkart, *ChemBioChem*, 2008, **9**, 2096–2103.
170 A. S. Worthington, M. D. Burkart and S. Smith, unpublished results.
171 S. Kapur, A. Worthington, Y. Tang, D. E. Cane, M. D. Burkart and C. Khosla, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3034–3038.
172 P. K. Mishra and D. G. Drueckhammer, *Chem. Rev.*, 2000, **100**, 3283–3310.
173 R. W. Haushalter, A. S. Worthington, G. H. Hur and M. D. Burkart, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3039–3042.
174 A. Y. Chen, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 784–792.
175 C. Qiao, D. J. Wilson, E. M. Bennett and C. C. Aldrich, *J. Am. Chem. Soc.*, 2007, **129**, 6350–6351.
176 J. S. Cisar and D. S. Tan, *Chem. Soc. Rev.*, 2008, **37**, 1320–1329.
177 C. Qiao, A. Gupte, H. I. Boshoff, D. J. Wilson, E. M. Bennett, R. V. Somu, C. E. Barry, 3rd and C. C. Aldrich, *J. Med. Chem.*, 2007, **50**, 6080–6094.
178 Y. Liu and S. D. Bruner, *ChemBioChem*, 2007, **8**, 617–621.
179 K. T. Barglow and B. F. Cravatt, *Angew. Chem., Int. Ed.*, 2006, **45**, 7408–7411.
180 E. Weerapana, G. M. Simon and B. F. Cravatt, *Nat. Chem. Biol.*, 2008, **4**, 405–407.
181 G. H. Hur, J. L. Meier and M. D. Burkart, 2008, manuscript in preparation.
182 M. Hahn and T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15585–15590.
183 C. Chiocchini, U. Linne and T. Stachelhaus, *Chem. Biol.*, 2006, **13**, 899–908.
184 J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16793–16797.
185 N. Bayan and H. Therisod, *Biochim. Biophys. Acta*, 1992, **1123**, 191–197.
186 J. C. Powers, J. L. Asgian, O. D. Ekici and K. E. James, *Chem. Rev.*, 2002, **102**, 4639–4750.
187 A. K. Das, J. J. Bellizzi, 3rd, S. Tandel, E. Biehl, J. Clardy and S. L. Hofmann, *J. Biol. Chem.*, 2000, **275**, 23847–23851.
188 B. Chakravarty, Z. Gu, S. S. Chirala, S. J. Wakil and F. A. Quiocho, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15567–15572.
189 S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh and M. T. Stubbs, *Structure (London)*, 2002, **10**, 301–310.
190 S. C. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, 2002, **41**, 12598–12606.
191 J. W. Giraldes, D. L. Akey, J. D. Kittendorf, D. H. Sherman, J. L. Smith and R. A. Fecik, *Nat. Chem. Biol.*, 2006, **2**, 531–536.
192 D. L. Akey, J. D. Kittendorf, J. W. Giraldes, R. A. Fecik, D. H. Sherman and J. L. Smith, *Nat. Chem. Biol.*, 2006, **2**, 537–542.
193 C. V. Christianson, T. J. Montavon, G. M. Festin, H. A. Cooke, B. Shen and S. D. Bruner, *J. Am. Chem. Soc.*, 2007, **129**, 15744–15745.
194 P. F. Widboom, E. N. Fielding, Y. Liu and S. D. Bruner, *Nature*, 2007, **447**, 342–345.
195 T. J. Montavon, C. V. Christianson, G. M. Festin, B. Shen and S. D. Bruner, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3099–3102.
196 J. Beck, S. Ripka, A. Siegner, E. Schiltz and E. Schweizer, *Eur. J. Biochem.*, 1990, **192**, 487–498.
197 G. Weber, K. Schorgendorfer, E. Schneider-Scherzer and E. Leitner, *Curr. Genet.*, 1994, **26**, 120–125.
198 A. Lawen and R. Zocher, *J. Biol. Chem.*, 1990, **265**, 11355–11360.
199 H. B. Bode and R. Muller, *Angew. Chem., Int. Ed.*, 2005, **44**, 6828–6846.
200 J. Yin, F. Liu, X. Li and C. T. Walsh, *J. Am. Chem. Soc.*, 2004, **126**, 7754–7755.
201 Z. Zhou, P. Cironi, A. J. Lin, Y. Xu, S. Hrvatin, D. E. Golan, P. A. Silver, C. T. Walsh and J. Yin, *ACS Chem. Biol.*, 2007, **2**, 337–346.
202 T. L. Foley and M. D. Burkart, *Curr. Opin. Chem. Biol.*, 2007, **11**, 12–19.
203 T. L. Foley and M. D. Burkart, unpublished results.
204 D. J. Prescott, J. Elovson and P. R. Vagelos, *J. Biol. Chem.*, 1969, **244**, 4517–4521.
205 J. Thomas and J. E. Cronan, *J. Biol. Chem.*, 2005, **280**, 34675–34683.
206 C. Sobhy, *J. Biol. Chem.*, 1979, **254**, 8561–8566.
207 M. Kim, A. A. Qureshi, R. A. Jenik, F. A. Lornitzo and J. W. Porter, *Arch. Biochem. Biophys.*, 1977, **181**, 580–590.
208 R. Leonardi, Y. M. Zhang, C. O. Rock and S. Jackowski, *Prog. Lipid Res.*, 2005, **44**, 125–153.
209 S. Jackowski and C. O. Rock, *J. Bacteriol.*, 1981, **148**, 926–932.
210 S. Jackowski and C. O. Rock, *J. Bacteriol.*, 1984, **158**, 115–120.
211 E. Strauss and T. P. Begley, *J. Biol. Chem.*, 2002, **277**, 48205–48209.
212 Y. M. Zhang, M. W. Frank, K. G. Virga, R. E. Lee, C. O. Rock and S. Jackowski, *J. Biol. Chem.*, 2004, **279**, 50969–50975.
213 K. M. Clarke, A. C. Mercer, J. J. La Clair and M. D. Burkart, *J. Am. Chem. Soc.*, 2005, **127**, 11234–11235.
214 J. A. Prescher and C. R. Bertozzi, *Nat. Chem. Biol.*, 2005, **1**, 13–21.
215 E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007–2010.
216 N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem. Biol.*, 2006, **1**, 644–648.
217 A. E. Speers and B. F. Cravatt, *Chem. Biol.*, 2004, **11**, 535–546.
218 A. C. Mercer, J. L. Meier, J. W. Torpey and M. D. Burkart, 2008, submitted.
219 A. C. Mercer, J. L. Meier, G. H. Hur, A. R. Smith and M. D. Burkart, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 5991–5994.
220 J. L. Meier, A. C. Mercer and M. D. Burkart, *J. Am. Chem. Soc.*, 2008, **130**, 5443–5445.
221 Y. Liu, M. P. Patricelli and B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14694–14699.
222 M. J. Evans and B. F. Cravatt, *Chem. Rev.*, 2006, **106**, 3279–3301.
223 Y. Hwang, P. R. Thompson, L. Wang, L. Jiang, N. L. Kelleher and P. A. Cole, *Angew. Chem., Int. Ed.*, 2007, **46**, 7621–7624.
224 M. Yu, L. P. de Carvalho, G. Sun and J. S. Blanchard, *J. Am. Chem. Soc.*, 2006, **128**, 15356–15357.
225 T. Bottcher and S. A. Sieber, *Angew. Chem., Int. Ed.*, 2008, **47**, 4600–4603.
226 P. R. Jungblut, U. E. Schaible, H. J. Mollenkopf, U. Zimny-Arndt, B. Raupach, J. Mattow, P. Halada, S. Lamer, K. Hagens and S. H. Kaufmann, *Mol. Microbiol.*, 1999, **33**, 1103–1117.
227 K. P. Jayapal, R. J. Philp, Y. J. Kok, M. G. Yap, D. H. Sherman, T. J. Griffin and W. S. Hu, *PLoS One*, 2008, **3**, e2097.
228 P. Tafelmeyer, C. Laurent, P. Lenormand, J. C. Rousselle, L. Marsollier, G. Reysset, R. Zhang, A. Sickmann,

T. P. Stinear, A. Namane and S. T. Cole, *Proteomics*, 2008, **8**, 3124–3138.
229 J. Yin, P. D. Straight, S. Hrvatin, P. C. Dorrestein, S. B. Bumpus, C. Jao, N. L. Kelleher, R. Kolter and C. T. Walsh, *Chem. Biol.*, 2007, **14**, 303–312.
230 T. Stein, *Mol. Microbiol.*, 2005, **56**, 845–857.
231 J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15815–15820.
232 J. Handelsman, M. R. Rondon, S. F. Brady, J. Clardy and R. M. Goodman, *Chem. Biol.*, 1998, **5**, R245–R249.
233 F. Lipmann, *Wanderings of a Biochemist*, Wiley, 1971.

Acknowledgments

The text of chapter 2, in full, is a reprint of material as it appears in: Meier, J.L., and Burkart, M.D. (2009). The chemical biology of modular biosynthetic enzymes. Chemical Society Reviews, DOI: 10.1039/b805115c. I was the primary author of this work.

Chapter Three

Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification

Originally published as:

Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. *J Am Chem Soc* **128**, 12174-84 (2006).

J|A|C|S
ARTICLES
Published on Web 08/26/2006

Synthesis and Evaluation of Bioorthogonal Pantetheine Analogues for in Vivo Protein Modification

Jordan L. Meier, Andrew C. Mercer, Heriberto Rivera, Jr., and Michael D. Burkart*

Contribution from the Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358

Received May 8, 2006; E-mail: mburkart@ucsd.edu

Abstract: In vivo carrier protein tagging has recently become an attractive target for the site-specific modification of fusion systems and new approaches to natural product proteomics. A detailed study of pantetheine analogues was performed in order to identify suitable partners for covalent protein labeling inside living cells. A rapid synthesis of pantothenamide analogues was developed and used to produce a panel which was evaluated for in vitro and in vivo protein labeling. Kinetic comparisons allowed the construction of a structure−activity relationship to pinpoint the linker, dye, and bioorthogonal reporter of choice for carrier protein labeling. Finally bioorthogonal pantetheine analogues were shown to target carrier proteins with high specificity in vivo and undergo chemoselective ligation to reporters in crude cell lysate. The methods demonstrated here allow carrier proteins to be visualized and isolated for the first time without the need for antibody techniques and set the stage for the future use of carrier protein fusions in chemical biology.

Introduction

Recent years have seen intense research effort focused toward the development of new methods for the study and manipulation of covalently modified proteins, with particular attention given to in vivo methodologies.[1] Fluorescent protein fusions[2] and antibody conjugates[3] provide powerful tools for protein imaging and manipulation. However drawbacks of these methods, such as structural perturbations sometimes induced by large fusions and general membrane impermeability of antibodies, have lead researchers to devise methods for the site-specific modification of proteins by small-molecule probes. Ideally these probes should be low molecular weight, covalent in nature, and possessed of fluorescence or affinity properties allowing for facile imaging and manipulation. We recently introduced one such technique, demonstrating cellular uptake and covalent modification of carrier protein fusions by pantetheine analogues.[4] These coenzyme A (CoA) precursors were shown to penetrate the cell membrane and be transformed into fully formed CoA derivatives via the endogenous CoA metabolic pathway, whereupon they were transferred to a carrier protein by the promiscuous phosphopantetheinyltransferase (PPTase; E.C. 2.7.7.7) Sfp (Figure 2). This advance allows carrier protein labeling, a technique first developed from cell lysates[5] and since demonstrated on the cell surface,[6] to be performed within the cell, opening the door for more sophisticated labeling systems. Recent developments have seen the trimming of the carrier protein domain down to just 11 amino acids,[7] offering a fusion tag of the size and flexibility to be competitive with contemporary tagging systems and further highlighting the importance of techniques for the labeling of intracellular carrier proteins.

Several strategies for site-specific labeling of proteins in vivo have been previously demonstrated. Examples include Bertozzi's manipulation of the sialic acid biosynthetic pathway for the introduction of keto and azido functionalized cell-surface glycoproteins,[8] Cravatt's introduction of azido/alkyne functionalities by covalent irreversible inhibition of protein active sites,[9] and Hsieh−Wilson's chemoenzymatic introduction of a keto-functionality for capture of *O*-GlcNAc-modified proteins.[10] In each of these examples the protein is not directly labeled with a fluorescence or affinity tag, but rather a unique and biologically inert chemical functionality is introduced. This functional-

(1) Bertozzi, C. R.; Prescher, J. A. *Nat. Chem. Biol.* **2005**, *1*, 13-21.
(2) (a) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−544. (b) Lippincott-Schwartz, J.; Patterson, G. H. *Science* **2003**, *300*, 87−91.
(3) (a) Fritze, C. E.; Anderson, T. R. *Methods Enzymol.* **2000**, *327*, 3−16. (b) Massoud, T. F.; Gambhir, S. S. *Genes Dev.* **2003**, *17*, 545−580.
(4) Clarke, K. M.; Mercer, A. C.; La Clair, J. J.; Burkart, M. D. *J. Am. Chem. Soc.* **2005**, *127*, 11234−11235.
(5) (a) La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. *Chem. Biol.* **2004**, *11*, 195−201. (b) Yin, J.; Liu, F.; Li, X.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, *126*, 7754−7755. (c) Yin, J.; Liu, F.; Schinke, M.; Daly, C.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, *126*, 13570−13571.
(6) (a) George, N.; Pick, H.; Vogel, H.; Johnsson, N.; Johnsson, K. *J. Am. Chem. Soc.* **2004**, *126*, 8896−8897. (b) Yin, J.; Lin, A. J.; Buckett, P. D.; Wessling-Resnick, M.; Golan, D. E.; Walsh, C. T. *Chem. Biol.* **2005**, *12*, 999−1006. (c) Vivero-Pol L.; George, N.; Krumm, H.; Johnsson, K.; Johnsson, N. *J. Am. Chem. Soc.* **2005**, *127*, 12770−12771.
(7) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15815−15820.
(8) (a) Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. *Science* **1997**, *276*, 1125−1128. (b) Yarema, K. J.; Mahal, L. K.; Bruehl, R.; Rodriguez, E. C.; Bertozzi, C. R. *J. Biol. Chem.* **1998**, *273*, 31168−31179. (c) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007−2010.
(9) (a) Speers, A. E.; Cravatt, B. F. *Chem. Biol.* **2004**, *11*, 535−546. (b) Speers, A. E.; Adam, G. C.; Cravatt, B. F. *J. Am. Chem. Soc.* **2003**, *125*, 4686−4687. (c) Alexander, J. P.; Cravatt, B. F. *Chem Biol.* **2005**, *12*, 1179−1187.
(10) (a) Hwan-Ching, T.; Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. *J. Am. Chem. Soc.* **2004**, *126*, 10500−10501. (b) Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *36*, 13132−13137.

10.1021/ja063217n CCC: $33.50

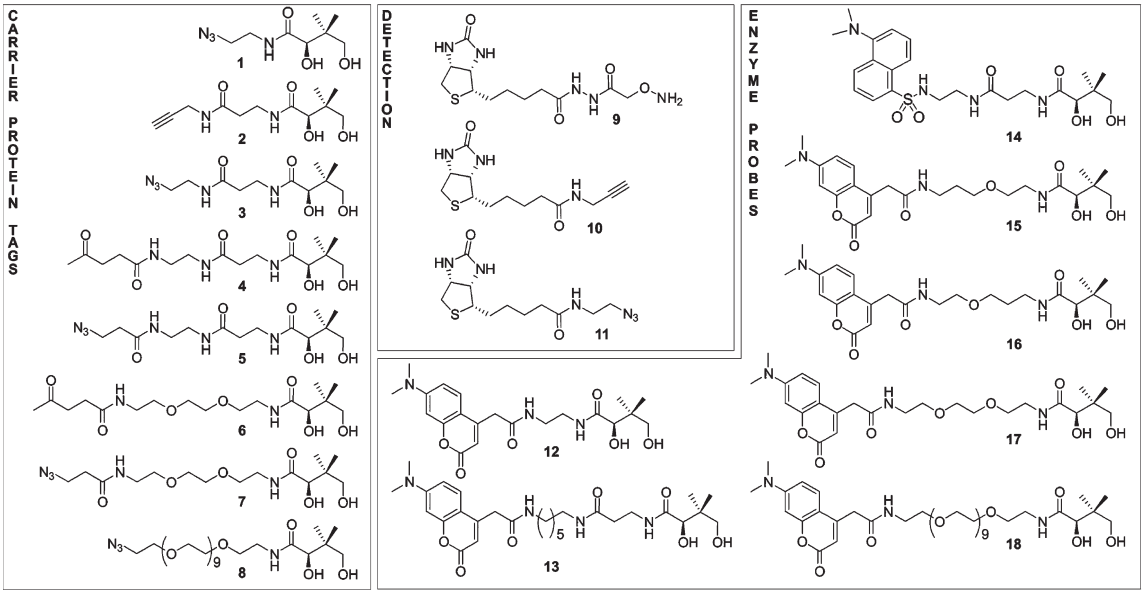


Figure 1. Structures of pantetheine analogues and biotin detection agents used in this study.

ity can then undergo reaction with exogenously delivered reporters to label the protein of interest for detection and/or isolation, depending on the nature of the reporter. Advantages of this two-step labeling process include (i) better uptake of smaller probes due to increased membrane permeability, (ii) increased incorporation of probes into native biosynthetic pathways due to a greater similarity to natural substrate, and (iii) the ability to conjugate a protein to virtually any reporter possessing reactivity with the bioorthogonal functionality.[1,9] Here we present a full study of simplified pantetheine analogues that harness the power of such bioorthogonal ligation reactions. First we optimize the synthesis of simplified pantetheine analogues via a one-step reaction with pantolactone. Next, the specificity of the CoA biosynthetic pathway is probed by a small panel of these simplified substrates. Finally, we validate the utility of this strategy by demonstrating and comparing the delivery of bioorthogonal chemical functionalities to carrier proteins in vitro and in vivo and using the newly tagged carrier proteins for two widely applied chemoselective ligations: the reaction of ketones and hydroxylamines to form oximes and the Cu(I)-catalyzed azide−alkyne [3 + 2] cycloaddition reaction ("click" chemistry). This new ability to manipulate a bioorthogonally tagged carrier protein in vivo promises to be a valuable tool for both new approaches to natural product proteomics as well as the study of novel intracellular carrier protein fusion systems.

Results and Discussion

Analogue Synthesis: Pantolactone-Ring Opening. In our efforts to address CoA biosynthesis with novel analogues, we had initially investigated the synthesis of pantetheine and phosphopantetheine derivatives that could be assembled in a manner analogous to peptide library synthesis.[11] This necessitated addressing the synthetic challenges associated with pantolactone, namely the lability of the α-proton following protection of the pantolactone secondary alcohol. At this point, the presumption was made that the identities of cystamine and β-alanine were necessary for turnover by the CoA metabolic pathway. However our own in vitro studies and recent work by Lee[12] questioned the degree of selectivity gained through specific interactions between the β-alanine moiety of pantothenate and PANK (used in this manuscript to denote all enzymes with pantothenate kinase activity; E.C. 2.7.1.33), the first enzyme in the CoA biosynthetic pathway and gatekeeper for downstream metabolism.[13] Further, we found that *E. coli* PANK (CoaA) catalyzed the phosphorylation of pantetheine and analogues with variations at the cystamine moiety almost as well as pantothenate itself.[14] Given this newly revealed permissiveness in the CoA biosynthetic pathway, we reasoned that simplified analogues of pantetheine could be used for both in vitro and in vivo applications. Elimination of the amide bond between cystamine and β-alanine significantly simplifies synthetic access to reporter-modified pantetheine analogues by reducing overall molecule polarity and solubility issues, eliminating time-consuming protection/deprotection steps of the 1,3-diol, and replacing the multiple peptide coupling and purification steps of previous syntheses[4] with a simple one-step nucleophilic ring-opening of pantolactone. With this in mind we chose to mimic the aletheine moiety (*N*-(β-alanyl)-β-aminoethanethiol) with more synthetically flexible poly(ethylene glycol) (PEG) linkers. In addition to the synthetic utility of this substitution, PEG spacers have the advantages of increasing the aqueous solubility of small molecule probes and distancing reporter labels from a labeled protein with the effect of both enhancing secondary detection properties and reducing any negative effect of reporter/protein interactions.[15] We sought to incorporate these

(11) Mandel, A. L.; La Clair, J. J.; Burkart, M. D. *Org. Lett.* **2004**, *6*, 4801−4803.
(12) Virga, K. G.; Zhang, Y. M.; Leonardi, R.; Ivey, R. A.; Hevener, K.; Park, H. W.; Jackowski, S.; Rock, C. O.; Lee, R. E. *Bioorg. Med. Chem.* **2006**, *14*, 1007−1020.
(13) Jackowski, S.; Rock, C. O. *J. Bacteriol.* **1981**, *148*, 926−932.
(14) Worthington, A. S.; Burkart, M. D. *Org. Biomol. Chem.* **2005**, *4*, 44−46.
(15) Kumar, V.; Aldrich, J. V. *Org. Lett.* **2003**, *5*, 613−616.

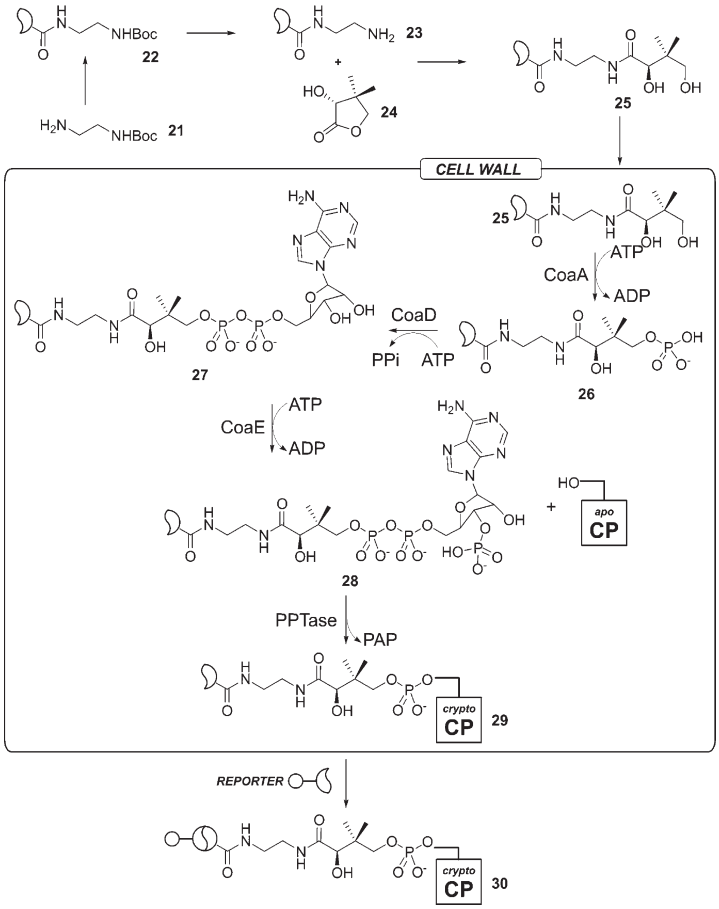


Figure 2. General strategy for in vivo labeling of carrier protein by pantetheine analogues. Virtually any monoprotected amine (**21**) can be transformed into a pantetheine analogue (**25**) by the three-step coupling/deprotection/ring-opening sequence. Cellular uptake and biosynthetic processing by CoaA, CoaD, and CoaE yield the CoA analogue **28**, which is then transferred to the carrier protein by a PPTase to yield bioorthogonally labeled carrier protein **28**. After cell lysis this carrier protein can now be conjugated to the reporter of choice via an appropriate chemoselective ligation reaction.

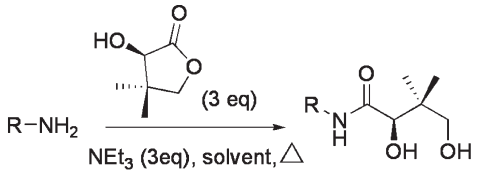
advantages into the design of an ideal, synthetically straightforward, biodetectible pantetheine analogue.

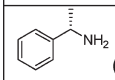
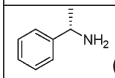
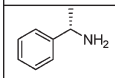
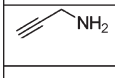
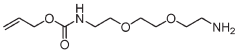
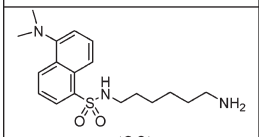
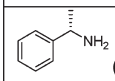
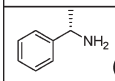
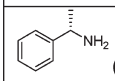
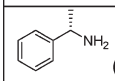
To quickly access a large selection of analogues it was deemed appropriate to first revise the current methodology for pantothenamide synthesis. Previous protocols calling for the base-promoted nucleophilic ring opening of pantolactone by an amine could be subject to racemization or hampered by long reaction times (>24 h).[12,16,17] To address these problems we turned to microwave-assisted organic synthesis. By using (*S*)-(−)-α-methylbenzylamine we were able to test a variety of conditions for their ability to open pantolactone with a fairly hindered chiral nucleophile and analyze enantiopurity by ^{1}H NMR (Table 1, see Supporting Information for ^{1}H NMR data).

The study showed pantolactone to be surprisingly robust to a variety of conditions, and reaction times could be reduced nearly 50-fold compared to previous preparations with a retention of optical purity. As expected from the hypothesized transition state of this reaction, protic solvents proved ideal for nucleophilic ring-opening, with ethanol providing the best balance of energy-absorbance and solubilization. Moving from our model system to usefully functionalized amines, it was shown that alkyne (**41**), PEG (**44**), and fluorophore (**20**) containing pantetheine analogues could be synthesized in good to moderate yields within 30 min using microwave assistance. Interestingly a very recent report also presented ethanol as the solvent of choice for this transformation under thermal conditions; however without the addition of any base these large-scale syntheses suffered from very long reaction times (72−120 h).[18] Accordingly our ideal microwave reaction conditions were also tested under simple reflux. Triethylamine proved to be a sufficient base, as replacement with Hunig's base showed no significant effect on the reaction outcome. Stronger bases were avoided. Again it was found that reflux of (*S*)-(−)-α-methylbenzylamine with excess pantolactone and triethylamine provided pantetheine analogues with no apparent racemization in excellent yields in 7−12 h. This alternative synthesis provides

(16) Dueno, E. E.; Chu, F.; Kim, S. I.; Jung, K. W. *Tetrahedron Lett.* **1999**, *40*, 1843−1846.
(17) (a) Michelson, A. M. *Biochim. Biophys. Acta* **1964**, *93*, 71−77. (b) Moffatt, J. G.; Khorana, H. G. *J. Am.* Chem. Soc. **1961**, *83*, 663−675.
(18) Krause, B. R. et al. *Synth. Commun.* **2006**, *36*, 365−391.

Table 1. Data Table for One-Step Synthesis of Pantetheine Analogues via Nucleophilic Ring-Opening of Pantolactone



Amine	Solvent	Temp (°C)	Time (hr)	Yield
 (**19**)	EtOH	160 (a)	0.5	91%
 (**19**)	DMF	165 (a)	0.5	63%
 (**19**)	THF	110 (a)	0.5	44%
 (**41**)	EtOH	160 (a)	0.5	42%
 (**44**)	EtOH	160 (a)	0.5	82%
 (**20**)	EtOH	160 (a)	0.5	75%
 (**19**)	EtOH	Reflux (b)	7	97%
 (**19**)	MeOH	Reflux (b)	7	84%
 (**19**)	CH_3CN	Reflux (b)	22	83%
 (**19**)	DME	Reflux (b)	12	30%

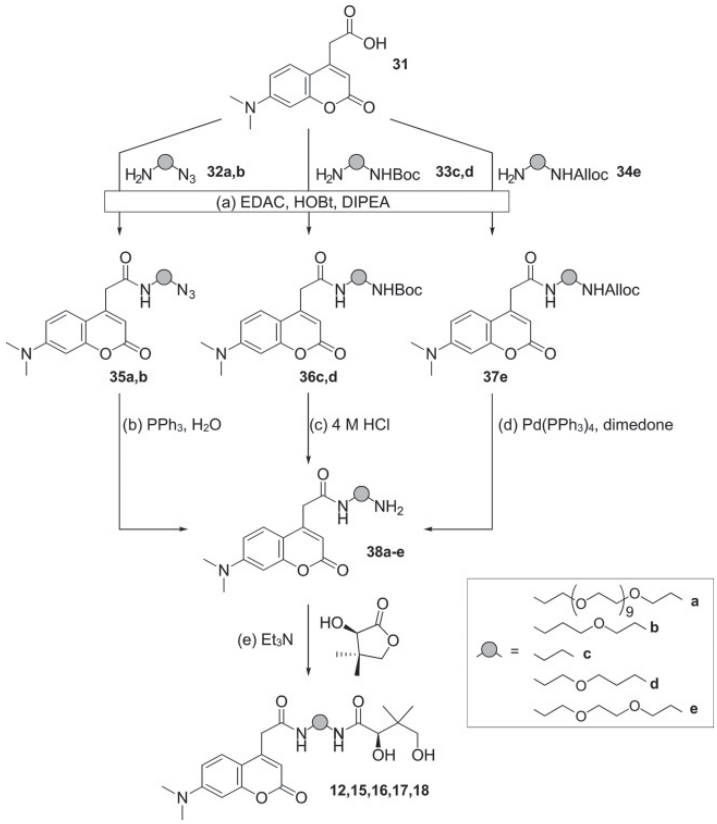
[a] Microwave-assisted. [b] Thermal condition.

another avenue for analogue preparation in cases where the reporter or linker is sensitive to decomposition under microwave conditions.[19]

Synthesis of Bioorthogonal and Fluorescent Pantetheine Analogues. The general strategy for chemoenzymatic synthesis of CoA analogues is depicted in Figure 2. First a monoprotected amine (depicted in this example by *N*-Boc ethylenediamine **21**) is conjugated to the biodetectible tag of choice by standard peptide coupling conditions. After deprotection, this amine can be conjugated to either pantolactone **24** through nucleophilic ring opening or pantothenic acid via EDAC mediated coupling. The newly formed pantetheine analogue is then processed via

(19) Microwave-assisted conjugation of 7-dimethylaminocoumarin-4-acetic acid containing amines to pantolactone resulted in the formation of unidentified decomposition products. For these couplings the classical condition (MeOH, NEt_3, reflux) was used.

Scheme 1. Synthesis of CoaA Probes[a]



[a] (a) EDAC (2 equiv), DIPEA (2 equiv), DMF, rt 12 h; (b) PPh_3 (1.2 equiv), 10:1 THF/H_2O, rt 12−24 h; (c) 4 M HCl/dioxanes, rt 24 h; (d) $Pd(PPh_3)_4$ (cat.), PPh_3 (2 equiv), dimedone (7 equiv), rt 12 h; (e) pantolactone (3 equiv), NEt_3 (3 equiv), MeOH, reflux 24−72 h.

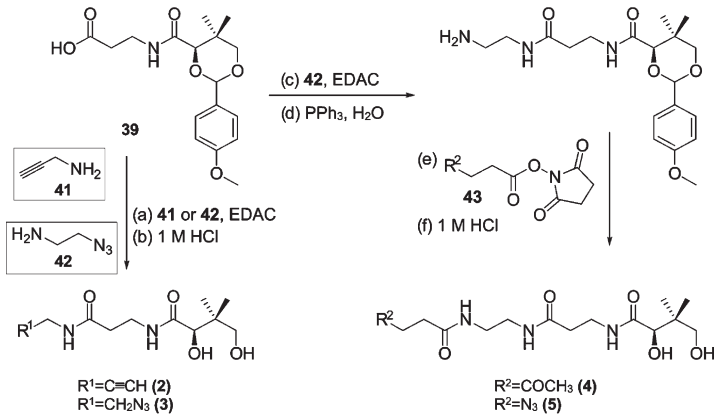
stepwise conversion by CoaA, a phosphopantetheine adenylytranserase (CoaD; E.C. 2.7.7.3), and a dephospho-CoA kinase (CoaE; E.C. 2.7.1.24) to form CoA analogue **28**, which is subsequently transferred to a conserved residue of the carrier protein by a PPTase to produce reporter-modified *crypto*-carrier protein **30**.

Analogues **12**−**18** were synthesized by this route in order to test the permissibility of CoA biosynthesis toward unnatural pantetheine analogues, particularly the effect of changes in the β-alanine/cystamine region (Scheme 1). We chose three parallel amino-protecting group strategies (azide, Boc, Alloc) based on the commercial availability (**32a**), simple synthesis from literature preparations (**33c**, **34e**), and orthogonal protecting group traits (**32b**, **33d**) of the specified diamines (Scheme 1). This strategy allowed compound **12** to be synthesized in two steps from *N*-Boc ethylenediamine conjugated 7-dimethylaminocoumarin-4-acetic acid (DMACA) **36c**.[20] Acid-catalyzed deprotection afforded the free amine, which performed nucleophilic ring opening of pantolactone to afford the final product in 85% yield. Compound **13** was synthesized as previously described,[4] while dansylated pantetheine **14** was synthesized by an analogous route. Compounds **15** and **16** were chosen to probe the effect of replacement of the strong H-bond accepting carbonyl and H-bond donating nitrogen of the natural substrate amide with weak H-bond accepting ether oxygens. Their synthesis made use of a common orthogonally protected diamine **48** (see Supporting Information), which underwent differential deprotection to give monoprotected diamines **32b** and **33d**. Subsequent EDAC mediated conjugation to dye **31**, azido/Boc deprotection by standard conditions, and nucleophilic ring opening of pantolactone afforded enzyme probes **15** and **16**. Our interest in incorporating the advantageous properties of PEG spacers in our library of pantetheine analogues lead us to synthesize PEG-linked-pantoic acid conjugate **17** in a 41% overall yield through an analogous dye conjugation/deprotection/nucleophilic ring-opening sequence starting from previously described *N*-Alloc diaminoethylene glycol **34e**. Compound **18** was chosen to test the limits of linker length in CoA biosynthesis and was easily attainable from the commercially available monoazido/monoamino terminal nonaethylene glycol **32a** through a similar series of reactions.

Compounds **1**−**8** sought to create a wide spectrum of pantetheine analogues incorporating some of the most commonly used bioorthogonal tags such as ketones, azides, and alkynes. Particular attention was paid to the azido moiety, where analogues replacing the β-Ala (**1**), cystamine (**2**), and thiol (**5**) regions of natural pantetheine with terminal azide moieties were

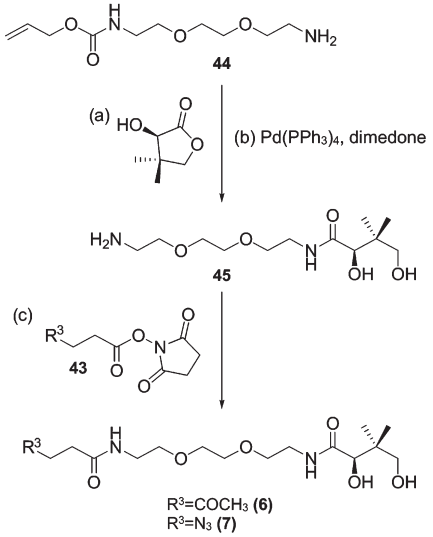
(20) La Clair, J. J.; et al. *ChemBioChem* **2006**, *7*, 409−416.

Scheme 2. Synthesis of β-Ala-Linked Bioorgthogonal Pantetheine Analogues[a]



[a] (a) **41** or **42**, EDAC (2 equiv), HOBt (2.5 equiv), DIPEA (2 equiv), DMF, rt 12 h; (b) 1 M HCl, 1:1 THF/H_2O, rt 1 h; (c) **42**, EDAC (2 equiv), HOBt (2.5 equiv), DIPEA (2 equiv), DMF, rt 12 h; (d) PPh_3 (2 equiv), 10:1 THF/H_2O, rt 12 h; (e) NHS ester **43a** or **43b**, NEt_3 (2 equiv), rt 8 h; (f) 1 M HCl, 1:1 THF/H_2O rt 1 h.

Scheme 3. Synthesis of PEG-Linked Bioorthogonal Pantetheine Analogues[a]



[a] (a) Pantolactone (3 equiv), NEt_3 (3 equiv), EtOH, 160 °C 0.5 h; (b) $Pd(PPh_3)_4$ (0.05 equiv), PPh_3, dimedone; (c) NHS ester **43a** or **43b**, NEt_3 (2 equiv), rt 4 h.

synthesized, in addition to experimentation with short (**7**) and long (**8**) PEG-linked azido-pantetheine analogues. Compound **1** was synthesized in one step from the nucleophilic ring-opening of pantolactone by 2-azidoethanamine (**42**). Compounds **2** and **3** (Scheme 2) were synthesized by standard peptide coupling of *p*-methoxybenzylideneacetal-protected (PMB) pantothenate **39** and the corresponding alkynyl and azido-amines, followed by acidic deprotection of the 1,3-diol. Compounds **4** and **5** again utilized **39** as a starting material, which was coupled to 2-azidoethanamine, deprotected to the corresponding amine, and coupled to the corresponding *N*-hydroxysuccinimidyl keto/azido (**43a,b**) ester. PEG linked pantetheine conjugates **6** and **7** (Scheme 3) were constructed in a similar fashion by nucleophilic ring-opening of pantolactone under microwave-conditions by *N*-Alloc protected diaminoethylene glycol **44** to give a common intermediate, followed by $Pd(PPh_3)_4$ mediated deprotection and coupling of the purified free amine to an keto/azido acid activated as the succinimidyl ester. Finally, analogue **8** was synthesized in one step through microwave-assisted nucleophilic ring-opening of pantolactone by monoazido/monoamino terminated nonaethylene glycol **32a**.

In Vitro Pathway Incorporation: Kinetics with CoaA. As mentioned earlier, phosphorylation of the primary hydroxyl group of pantothenate by the first protein in the CoA biosynthesis pathway, CoaA, is believed to be the rate-limiting step in vivo.[13] Due to its role as the gatekeeper of CoA biosynthesis, we assayed each of our new analogues for kinetic activity with CoaA (Table 2).

The assay was performed as previously described using the prototypical bacterial pantothenate kinase, CoaA from *E. coli*.[15] The reaction of CoaA with the natural substrate pantothenate gave values that conformed to those previously reported in the literature.[14,21] Pantothenate mimics **2** and **3** show k_{cat} and K_m values closely approaching those of the natural substrate. As seen in our previous studies pantetheine is also a substrate for CoaA, and pantetheine-resembling substrates **4**, **5**, **13**, and **14** show turnovers near equal to (**4**, **13**, **14**) or greater than (**5**) that of pantetheine. Compounds **5**, **13**, and **14** are processed particularly efficiently by CoaA. Conversely, compounds with PEG linkers between the pantoic acid moiety and the bioorthogonal terminus were poor substrates for CoaA. The length of the linker region was a strong factor in determining substrate suitability, with the longest compounds **8** and **18** proving such poor substrates that kinetic data could not be generated. Shorter PEG linked compounds (**7** and **17**) were viable substrates in the assay but turned over at a rate 2−5-fold less than that for derivatives containing β-Ala/diamine linkers. PEG linked pantoic acid conjugate **6** differs only from **7** by the exchange of an azide for an acyl substitutent but shows markedly decreased kinetic activity.

To isolate and investigate the effect of H-bond accepting heteroatoms in the linker region, pantetheine analogues **15** and

(21) Strauss, E.; Begley, T. P. *J. Biol. Chem.* **2002**, *277*, 48205−48209.

Table 2. Kinetic Parameters of *E. coli* CoaA with Natural Substrates and Pantetheine Analogues and Summary of in Vivo/in Vitro Results[a]

compound	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m ($s^{-1} M^{-1}$) ($\times 10^3$)	in vitro	in vivo
pantothenate	31.27 ± 0.58	28.56 ± 1.77	18.25 ± 0.68	na	na
pantetheine[14]	19.2 ± 0.1	91 ± 10	3.53 ± 0.44	na	na
1	22.33 ± 1.54	692.3 ± 89.13	0.54 ± 0.07	++	−
2	31.45 ± 0.75	32.85 ± 2.54	15.96 ± 0.76	++	++
3	28.02 ± 0.42	43.53 ± 1.99	10.73 ± 0.32	++	++
4	20.64 ± 0.45	62.65 ± 3.89	5.49 ± 0.24	++	+
5	40.91 ± 1.12	71.89 ± 6.98	9.48 ± 0.52	++	++
6	1.60 ± 0.06	0.196 ± 0.09	136.4 ± 10.63	++	−
7	6.16 ± 0.44	53.76 ± 11.16	1.91 ± 0.27	++	−
8	na	na	na	+	−
12	11.56 ± 0.57	37.24 ± 5.82	5.17 ± 0.51	++	+
13	19.16 ± 1.41	28.40 ± 6.92	11.25 ± 1.65	++	++
14	17.26 ± 0.62	27.33 ± 3.28	10.53 ± 0.76	++	−
15	1.36 ± 0.05	0.76 ± 0.32	0.03 ± 0.04	++	−
16	1.00 ± 0.03	0.14 ± 0.18	0.12 ± 0.01	++	−
17	10.06 ± 0.43	51.18 ± 6.42	3.28 ± 0.28	++	−
18	na	na	na	+	−

[a] Kinetic data with CoaA and summary of in vivo/in vitro labeling efficiency of carrier proteins using reporter modified pantetheine analogues. Kinetic values for the natural substrate pantothenate and pantetheine are given for comparison. In vivo/in vitro labeling is based on data from gel shift, fluorescent gel, Western blot, and MALDI-MS data (see Figures 3,4,5, Supporting Information) and represented semiquantitatively as follows: ++ for excellent labeling of carrier protein, + for low but detectable labeling, and − for no detectable labeling of carrier protein.

16 were synthesized. These compounds were designed to replace the amide bond between β-Ala and cystamine of pantetheine with a single ether oxygen, allowing investigation of subtle substrate−enzyme interactions in the CoaA active site.[12] To our surprise these compounds were extremely poor substrates. While compounds with the aforementioned short PEG linkers (**7**, **17**) showed K_m values 2-fold higher than that for natural substrate pantothenate, and compounds with traditional β-Ala/diamine linkers showed values that were either the same (**13**, **14**) or 2-fold higher (**4**, **5**), the K_m values for **15** and **16** were 100-fold *lower*. Turnover for these compounds was proportionately low, indicative of tight binding. To test if these compounds were acting as inhibitors of CoaA, a competitive kinetic assay was set up using pantotenate as the substrate. The results (see Supporting Information) show that **15** acts as a noncompetitive inhibitor, suggesting that it may bind the allosteric regulation site of CoaA.[23] Investigation of these compounds as potential inhibitor scaffolds is ongoing. Azide **1**, which omitted entirely the β-Ala/carbon diamine or PEG linkages of other analogues, showed turnover within the range of the natural substrate; however the K_m was 20-fold higher than that for pantothenate suggesting deletion of an interaction involved in active site binding. Another interesting pantoic acid analogue **12**, which shortened the pantoic acid-reporter linker length to four carbons and reversed the carbonyl and amide β-Ala/cystamine linkage of natural pantetheine, showed lower turnover and catalytic efficiency than analogues containing the natural β-Ala/pantoic acid linkage (**5**, **13**, **14**).

In Vitro Pathway Incorporation: Gel Shift. The conversion of *apo*-ACP to *holo*-ACP or reporter-modified *crypto*-ACP causes a change in the mobility of the protein on a nondenaturing polyacrylamide gel.[12] To assay each compound for activity throughout the entire co-opted CoA pathway, we ran covalently modified carrier protein on a native PAGE gel and compared the mobility to *apo*-ACP (Figure 3a). Pantetheine analogues were reacted as previously reported[4] with the enzymes CoaA-E to create CoA analogues, followed by the addition of the PPTase Sfp and *apo*-ACP. As seen in Figure 3a, all of the compounds tested demonstrated some change in mobility with relation to *apo*-ACP. For most compounds full conversion to *crypto*-ACP is obtained; however analogues with long PEG linkers (**8**, **18**) give multiple bands on the gel that suggest *apo*-protein remains.

To confirm the results from this assay, we subjected the reaction mixtures to matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). The MALDI-MS data (Figure 4 and Supporting Information) confirm that all pantetheine analogues in our panel are indeed converted into CoA derivatives and transferred onto a carrier protein in vitro. *Apo*-ACP (Figure 4a) shows a characteristic peak with a mass of 8505 Da. Compound **13** was reacted with the CoaA biosynthesis

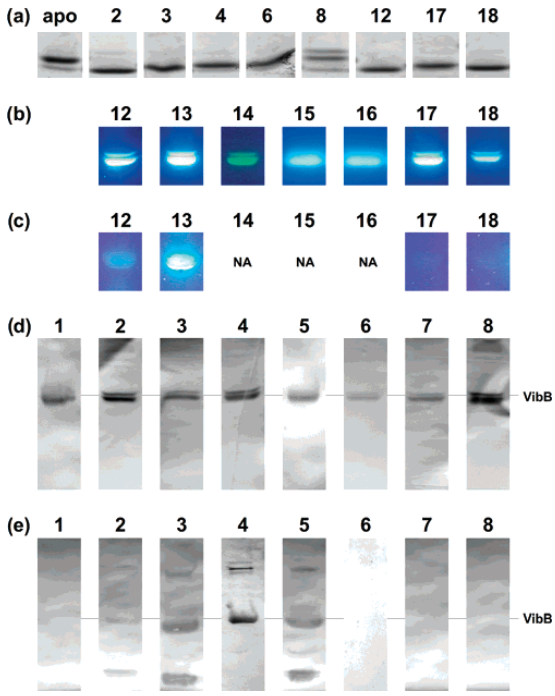


Figure 3. In vivo and in vitro activity of pantetheine analogue panel. (a) Analogues were assayed for gel shift after reaction with CoA biosynthesis enzymes (CoaA/D/E), PPTase (Sfp), and carrier protein (*E. coli* ACP). Conversion of *apo*-ACP to reporter modified *crypto*-ACP causes a change in the mobility of the protein on native PAGE. (b) In vitro modification of the carrier protein VibB by reaction with fluorescent pantetheine analogues, CoA biosynthesis enzymes (CoaA/D/E), and Sfp. (c) In vivo modification of carrier protein by incubation of fluorescent pantetheine analogues with *E. coli* overexpressing VibB and the PPTase Sfp. (d) In vitro modification of VibB by reaction with bioorthogonal pantetheine analogues, CoA biosynthesis enzymes (CoaA/D/E), and Sfp. Labeled carrier protein is visualized by chemoselective ligation to the appropriate biotin reporter (**9**−**11**) followed by SDS-PAGE, blotting onto nitrocellulose, and incubation with streptavidin-linked alkaline phosphatase. (e) In vivo modification of carrier protein by incubation of bioorthogonal pantetheine analogues with *E. coli* overexpressing VibB and the PPTase Sfp. Visualization as in (d).

(22) Mercer, A. C.; La Clair, J. J.; Burkart, M. D. *ChemBioChem* **2005**, *8*, 1335−1337.
(23) Ivey, R. A.; Zhang, Y.; Virga, K. G.; Hevener, K.; Lee, R. E.; Rock, C. O.; Jackowski, S.; Park, H. *J. Biol. Chem.* **2004**, *279*, 35622−35629.

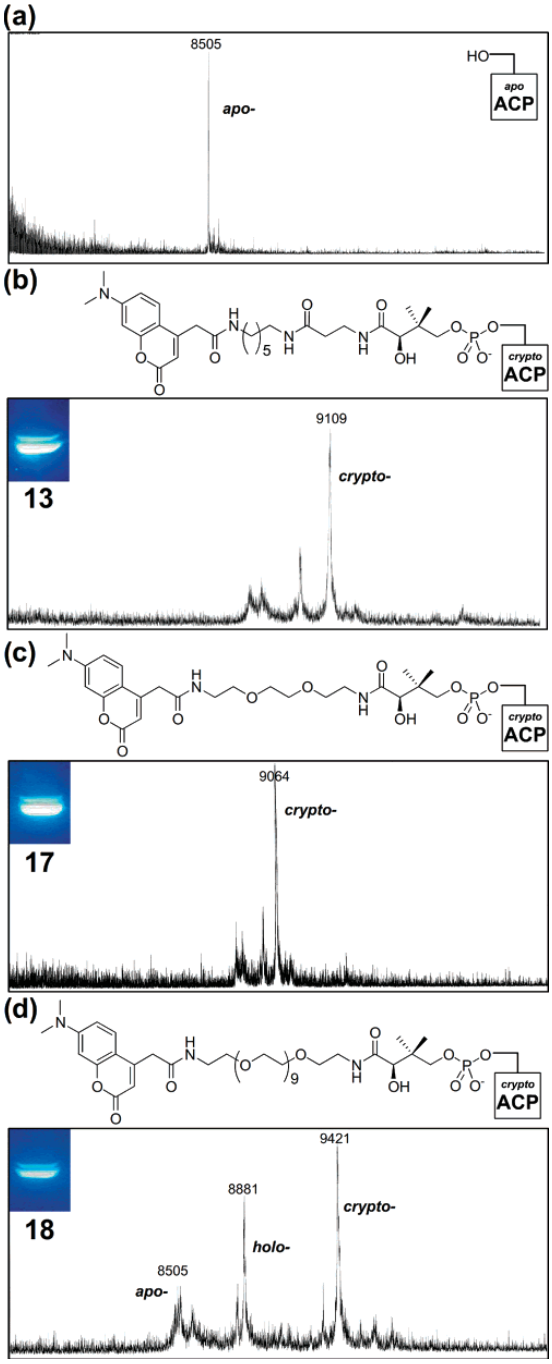


Figure 4. In vitro labeling of ACP. *Apo*-ACP (a) is reacted with pantetheine analogues, CoA biosynthetic enzymes (CoaA/D/E), and the PPTase Sfp. Fluorescent compounds **13** (b), **17** (c), and **18** (d) are all shown to be converted to CoA analogues and modify ACP by gel and mass spectral analysis.

enzymes, ACP, and Sfp as described above, and the reaction mix was analyzed by MALDI without further purification. As seen in Figure 4b, reaction with the CoA analogue of **13** causes the expected mass change of 604 mass units. PEG analogues **17** and **18** (Figure 4c,d) also show the expected mass shifts of 559 units and 916 units, respectively, corresponding to the formation of a *crypto*-carrier protein. Only in the case of the PEG-linked derivatives **8** and **18** does the conversion appear significantly incomplete, supporting the results of the gel shift assay.

In Vitro Pathway Incorporation: Biodetectability. Having confirmed that the pantetheine analogues were suitable substrates for the CoA biosynthesis enzymes and PPTase/carrier protein reaction, we next investigated the biodetectability of each analogue. The fluorescent analogues (**12**−**18**) were detected by UV visualization on PAGE gels as previously described.[4] Bioorthogonally tagged pantetheine analogues **1**−**8** were detected by chemoselective ligation to the appropriate alkoxyamine/azide/alkyne functionalized biotin followed by PAGE and visualization by western blotting and incubation with streptavidin-conjugated alkaline phosphatase. Keto-pantetheine compounds **4** and **6** were reacted overnight with biotin hydroxylamine (**9**) at room temperature, while pantetheine analogues with azide and alkyne functionalities were reacted with the corresponding alkynyl/azido (**10**/**11**) biotin following the procedure of Alexander.[9c] Inspection of the fluorescent gels (Figure 3b) and western blots (Figure 3d) confirmed the results of the gel shift assay and mass spectral data, indicating biodetectible covalent modification of carrier proteins in vitro for compounds **1**−**8** and **12**−**18**.

In Vivo Pathway Incorporation. Next we tested our library of pantetheine analogues for integration into the *E. coli* CoA pathway using our in vivo assay.[4] *E. coli* overexpressing the carrier protein VibB and the PPTase Sfp were incubated with 1 mM of each compound in 1 mL of culture. After 4 h of growth cells were pelleted, washed, and lysed. Lysate from these cultures was run on SDS-PAGE gels, and detection was carried out as described for the in vitro studies. Compounds **2**, **3**, **5**, **12**, and **13** demonstrated detectible modification of VibB in vivo (Figure 3c,e). Keto-pantetheine analogue **4** was also detectable but showed much weaker labeling than the similarly linked azido-analogue **5** (Figure 3e). The compounds most active in vivo show a strong correlation with the CoaA kinetic profile. To verify the results of the gels and blots, samples of crude lysate were assayed by MALDI-MS (Figure 5). For MALDI-MS analysis doubly transformed *E. coli* containing plasmids for the carrier protein Fren and the PPTase Sfp were used. As seen in Figure 5a compound **13** was taken up by the cell, processed into a CoA analogue and attached onto Fren. The *apo* peak can be seen at 8570 mass units. *Holo*-carrier protein is also visible at 8910 Da. This peak arises from the fact that natural CoA is available in the cell and readily ligated by PPTases to the overexpressed *apo*-Fren. Carrier protein modified with **13** can be seen at 9179 Da giving the expected 609 mass unit change. Similarly mass spectral analysis of cell lysate after incubation of bioorthogonal pantetheine analogue **2** with Fren/Sfp overexpressing *E. coli* shows an observable 315 Da shift of the known *apo* peak indicating formation of an alkyne-modified *crypto*-carrier protein. Subjection of the same crude cell lysate to click reaction conditions with biotin reporter **11** resulted in another 317 Da mass shift, indicating successful formation of biotinylated Fren via a Cu(I)-catalyzed [3 + 2] cycloaddition process.

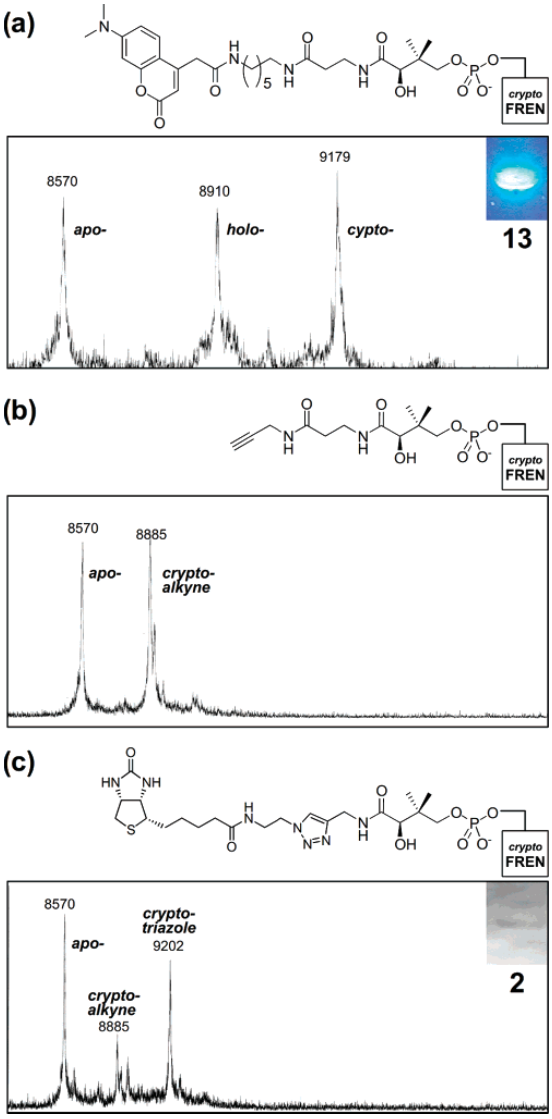


Figure 5. In vivo carrier protein labeling. The carrier protein Fren is labeled in vivo by incubation of *E. coli* overexpressing Fren and Sfp with (a) a fluorescent pantothenate analogue **13** and (b) a bioorthogonally tagged analogue **2**. Click reaction of alkyne-modified *crypto*-carrier protein with biotin reporter **11** affords triazole-linked biotinylated carrier protein (c), resulting in the expected shift in mass and allowing protein visualization by western blot.

Insights into CoaA Substrate Specificity from Kinetic, in Vivo, and in Vitro Analyses. Comparisons of the kinetic, in vitro, and in vivo assay results for different members of the panel yield important insights into the structure−activity relationships between *E. coli* PANK and pantetheine analogues. Despite the poor performance of compounds **1**, **6**−**8**, **12**, and **15**−**18** in the CoaA kinetic assay, all were shown to be converted to CoA analogues and loaded onto the carrier protein ACP by Sfp in vitro. This is both a testament to the incredible efficiency of Sfp in transferring unnatural CoA derivatives to *apo*-carrier proteins and a further demonstration of the utility of the chemoenzymatic approach to synthesis of unnatural CoA analogues. As mentioned above this finding greatly simplifies the synthetic task of constructing CoA derivatives for in vitro applications which utilize carrier protein tagging.[6] This allows access to virtually any reporter labeled-CoA analogue from a monoprotected amine in three steps, one of which can be expedited using microwave technology. Additionally the subtle substrate preferences Sfp has been shown to exhibit in systems incorporating multiple carrier proteins can be used for selective coding based on differential phosphopantetheinylation by unnatural CoA analogues,[22] adding another layer of complexity to in vitro and cell-surface carrier protein fusion systems.

Detailed analyses of the kinetic data demonstrate several important relationships. In general compounds containing a similar β-Ala linker region to that of natural pantothenate show the best kinetic profiles. In all compounds of the panel the pantoic acid region (C1−N5, numbering from the terminal primary hydroxyl of pantetheine) was conserved. Lee et al. indicated low binding of PANK inhibitors which lacked a proton-donating amide NH at the N8 position, pointing to a model of the PANK−ADP−pantothenate ternary complex in which the C-7 acid of pantothenate acts as an H-bond donor toward two key residues.[12,23] Our results verify the importance of this interaction. Compounds **1** and **3** contain the same 2-azidoethanamine-derived terminal azide, differing only in the β-Ala linkage of **3**. This change results in a 10-fold increase in K_m and near 20-fold increase in catalytic efficiency, indicating much better binding of the substrate with the H-bond donating β-Ala linker. Pantetheine analogues containing ethylene glycol based linkers incapable of acting as efficient H-bond donors (**6**−**8**, **17**−**18**) showed similarly poor kinetics compared with those containing β-Ala linkers. Compounds **15** and **16** provide perhaps the strongest evidence of the importance of this interaction, losing almost all substrate activity when reducing the strength of the electron pair donor and removing the H-bond donating NH completely from the pantetheine analogue. Interestingly these compounds show markedly different kinetics than PEG-linked pantetheine **17**, which differs by only two atoms, demonstrating the limitations of this model in predicting effects caused by alternate variables such as the reduced rotation around an sp^2 hybridized carbon at C8 and addition of an extra H-bond accepting heteroatom further down the linker. The kinetics of analogue **12**, in which the connectivity of the amide at C8 and N9 of the natural substrate pantetheine is reversed, indicate that there is some flexibility in the pocket around this position. Compound **12** shows good binding but slow turnover, with a catalytic efficiency poorer than that of any of the compounds with the H-bond donator in its natural position (**2**−**5**, **13**−**14**) but better than that of every compound in which the H-bond donating NH is absent (**1**, **6**−**8**, **15**−**18**). While the general trend toward an H-bond donor effect is large, other interactions resulting from the proximity of the aromatic coumarin-reporter molecule to the active site in this analogue must also be considered.

Several other trends which may be important for future design of carrier protein tags are of interest. In general, analogues terminating in alkynes show better kinetic parameters than those of azides, with azides showing better kinetic parameters than those of ketones. For PEG-based pantetheine derivatives, chain length proved an important factor, with longer chains showing

negligible activity with CoaA and poor in vivo protein tagging. Appending a different dye to the end of the pantetheine had no statistically significant impact on CoaA kinetics; however substitution of DMACA with dansyl lead to a complete loss of in vivo activity, suggesting a possible lack of a viable membrane-transport mechanism for dansyl-pantetheine analogues or an intracellular degradation process. It has been shown previously that DMACA is an excellent dye for in vivo applications.[20] The kinetics of β-Ala containing bioorthogonal pantetheines (**2**−**5**) compared with β-Ala containing pantetheines in which a fluorescent reporter that was directly appended (**13**−**14**) showed slightly better turnover and similar catalytic efficiency. On the whole the binding site of CoaA beyond the β-Ala moiety appears to be quite promiscuous, with little kinetic effect observed on substitution of the hexanediamine linker of **13** with a ethylenediamine linker or substitution of the ethylenediamine linker of **14** with a short (8-atom) PEG linker.[24]

Perhaps the most important conclusion that can be drawn from the assay results is that the in vivo activity of a pantetheine analogue has a direct correlation with kinetic activity with CoaA. Analogues **2**, **3**, **4**, **5**, **12**, and **13** were shown to be biodetectible in *E. coli* by Western blot analysis and fluorescence visualization. These show a k_{cat} between 11.5 and 40.9 min^{-1} and k_{cat}/K_m of 5.2×10^3−16.0×10^3 compared with the values of 19.2 min^{-1} (k_{cat}) and 3.5×10^3 s^{-1} M^{-1} (k_{cat}/K_m) for unmodified pantetheine (Table 2). In order for a pantetheine analogue to be processed into a CoA analogue in vivo, not only must an organism contain a promiscuous PANK with pantetheine kinase activity but also the pantetheine analogue must have comparable or better kinetics with PANK than pantetheine and (ideally) natural substrate pantothenate. Given that enterococci produce far more pantothenate than they require for primary metabolism,[25] any modified pantetheine analogue must make extremely efficient use of CoaA in order for CoA conversion and subsequent protein labeling to occur at detectible levels. Although the structural requirements for the import of pantetheine analogues have not been studied in detail, *E. coli* has been shown to have a pantothenate import system, the pantothenate permease (panF) symporter,[26] which also may exert some selectivity in the import of analogues and thus influence the ability of pantetheine analogues to be integrated into the CoA pathway in vivo. However previous studies showed that while overexpression of the panF gene resulted in elevated pantothenate uptake, a concurrent increase in CoA production was not observed, indicating PANK activity as the principal regulator of CoA biosynthesis. With that in mind these kinetic parameters should prove useful for the future design and assay of in vivo carrier protein tags.

Conclusions

The molecules described here have varied applications and provide insight valuable in the expanding field of proteomics. For in vitro applications and in vivo cell-surface labeling of carrier protein fusions, all pantetheine analogues studied in this manuscript are efficiently converted into CoA analogues and tethered to the protein via the four-step enzymatic pathway. This alternative methodology negates the purification steps necessary in the production of maleimide−CoA analogues from commercial sources, allows almost any variance of the chemical identity of the linker region, and provides an economical substitute for producing large amounts of CoA analogues in cases where large quantities of the desired CoA−maleimide−reporter conjugate may be prohibitively expensive. In addition the expansion of analogues with bioorthogonal reporters allows for increased detection and sensitivity. However despite these subtle advances, it is in the prospect of in vivo labeling that these tools become particularly important. The covalent modifications described herein have practical value in the study of in vivo activity of proteins and a place among the ever increasing myriad of proteomic techniques used to study them.

Of particular importance are the chemoselective ligation reactions demonstrated by the ketone, azide, and alkyne protein labels. These tools allow carrier proteins to be visualized and isolated for the first time without the expense and complication of antibody techniques. While our survey of bioorthogonal coupling partners was not exhaustive, the functionalities introduced should be applicable to other published methods. For instance, one can easily envisage analogues **3** and **5** being modified by Bertozzi's covalent Staudinger ligation with a reporter-conjugated triarylphosphine analogue for in vivo applications in which more stringent Cu(I)-catalyzed click chemistry conditions are not ideal.[27] It was to our disappointment that the PEG-incorporating pantetheine analogues proved nonamenable to in vivo protein labeling, most likely due to deletion of a crucial H-bonding interaction. Yet while PEG analogues are often useful for distancing reporter labels from the protein of interest for downstream modification, most likely they are not a necessity in this instance by virtue of the 4′-phosphopantetheine moiety. Indeed the 4′-phosphopantetheine is commonly believed to be appended to carrier proteins as a means to distance substrates and products from the protein core, a concept reinforced by recent structural studies of the fatty acid synthase.[28]

The system used in this study was limited to the *E. coli* CoA pathway and relied on the overexpression of a carrier protein PPTase pair to facilitate detection. However it has already been demonstrated that AcpS, a widely expressed PPTase, has the ability to transfer CoA analogues chemoenzymatically generated from panthothenamide type substrates to apo-ACP.[12] Immediate work will aim at lowering the detection limit to native protein levels in *E. coli*, a reasonable goal given the demonstrated ability of bioorthogonally labeled carrier protein to undergo ligation to affinity agents readily amenable to enrichment strategies. Also of interest to the growth of this methodology is the promiscuity of CoA biosynthetic enzymes in species other than *E. coli*. Little is known about the interspecies substrate specificity of CoaD and CoaE, although studies of the mechanism of inhibition of N-alkylpantothenamides have shown these enzymes capable of catalyzing the formation of unnatural CoA analogues which modify the fatty acid ACP in both *E. coli* and *S. aureus*,[21,29] despite a disparity in sequence homology between the PANK

(24) This comparison refers to kinetic data compiled for alternatively linked dansyl and DMACA−pantetheine analogues in ref 14.
(25) Rock, C. O.; Calder, R. B.; Karim, M. A.; Jackowski, S. *J. Biol. Chem.* **2000**, *275*, 1377−1383.
(26) Jackowski, S.; Alix, J. H. *J. Bacteriol.* **1990**, *172*, 3842−3848.
(27) Kohn M, Breinbauer R. *Angew. Chem., Int. Ed.* **2004**, *43*, 3106−3116.
(28) (a) Simon Jenni, S.; Leibundgut, M.; Maier, T.; Ban, N. *Science* **2006**, *311*, 1258−1262. (b) Maier, T.; Jenni, S.; Ban, N. *Science* **2006**, *311*, 1263−1267.
(29) Leonardi, R.; Chohnan, S.; Zhang, Y.; Virga, K. G.; Lee, R. E.; Rock, C. O.; Jackowski, S. *J. Biol. Chem.* **2005**, *280*, 3314−3322.

enzymes of the different species.[30] Studies in our own lab have shown the mammalian form of CoaD and CoaE, which exists as a bifunctional fusion, efficiently processes 4′-phosphopantetheine analogues into CoA analogues in vitro.[14] One system where this approach may be limited is in a subset of pathogenic bacteria which contain only the recently discovered CoaX type pantothenate kinase. This PANK isoform has been shown to be resistant to inhibition by N-alkylpantothenamides, possibly indicating an inability of pantetheine analogues to act as competitive substrates.[31] To date most studies of pantetheine analogues have focused on their properties of inhibitors; the new amenability of carrier protein systems to fusion tagging strategies and our demonstration of protein modification in living systems provide strong incentive for the reexamination of the substrate properties of this class of small molecules.

Here we have performed a detailed investigation of pantetheine analogues to identify suitable partners for covalent protein labeling inside living cells. A rapid synthesis of pantothenamide analogues was developed for this purpose and used to produce a panel which was evaluated for in vitro and in vivo protein labeling. Kinetic comparisons allowed the construction of structure−activity relationships to pinpoint the linker, dye, and bioorthogonal reporter of choice for protein labeling. Finally bioorthogonal pantetheine analogues were shown to target carrier proteins with high specificity in vivo and undergo chemoselective ligation to reporters in crude cell lysate. The paucity of site-specific protein labeling tools represents a major obstacle to the routine application of such small-molecule probes in vivo.[32] Our increasing understanding of the kinetic parameters and structural limitations of pantetheine analogues sets the stage for the future use of 4′-phosphopantetheine analogue labeling in chemical biology.

Acknowledgment. This manuscript is dedicated to Bill Fenical on the occasion of his 65th birthday. Funding was provided by the University of California, San Diego, Department of Chemistry and Biochemistry, NSF CAREER Award 0347681 and NIH RO1GM075797. J.M. was supported by NIH Training Grant T32DK007233. H.R. was supported as an NIH NIGMS PREP scholar. We thank Joseph Noel and Thomas Baiga of the Salk Institute for Biological Studies for the use of a microwave reactor and Jessica Alexander of the Scripps Research Institute for helpful discussions.

Supporting Information Available: Abbreviations used are given in ref 33. General procedures for microwave assisted ring-opening of pantolactone; experimental details for the synthesis of compounds **1**−**18**; kinetic, in vitro, and in vivo assay details; full author listings for refs 18 and 20; and ^{1}H and ^{13}C NMR spectra of all final compounds and synthetic intermediates. This material is available free of charge via the Internet at http://pubs.acs.org.

JA063217N

(30) Choudhry, A. E.; Mandichak, T. L.; Broskey, J. P.; Egolf, R. W.; Kinsland, C.; Begley, T. P.; Seefeld, M. A.; Ku, T. W.; Brown, J. R.; Zalacain, M.; Ratnam, K. *Antimicrob. Agents Chemother.* **2003**, *47*, 2051−2055.
(31) Brand, L. A.; Strauss, E. *J. Biol. Chem.* **2005**, *280*, 20185−20188.
(32) Chen, I.; Ting, A. Y. *Curr. Opin. Biotechnol.* **2005**, *16*, 35−40.
(33) Abbreviations: CoA, Coenzyme A; PPTase, phosphopantetheinyltransferase; *O*-GlcNAc, *O*-linked *β*-*N*-acetylglucosamine; *β*-Ala, *β*-alanine; PANK, pantothenate kinase; CoaA, ecoli PANK; CoaD, *E. coli* phosphopantetheine adenylytransferase; CoaE, *E. coli* dephospho-CoA kinase PEG, poly(ethylene glycol); Pantolactone, (D)-(−)-pantolactone; Boc, *tert*-butyl carbamate; Alloc, allyl-carbamate; DMACA, 7-dimethylaminocoumarin-4-acetic acid; CP, carrier protein; ACP, fatty acid synthase acyl carrier protein from *E. coli*; VibB, vibriobactin synthase carrier protein from *Vibrio cholerae*; Fren, frenolicin synthase carrier protein from *Streptomyces roseofulvus*; panF, pantothenate permease.

Synthesis and Evaluation of Bioorthogonal Pantetheine Analogs for in Vivo Protein Modification

*Jordan L. Meier, Andrew C. Mercer, Heriberto Rivera Jr., and Michael D. Burkart**
Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358
mburkart@ucsd.edu

Supporting Information I

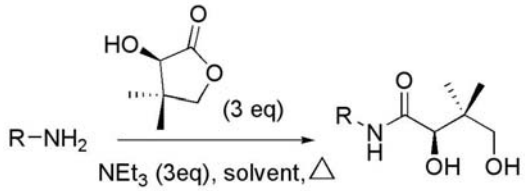
Contents

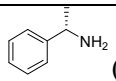
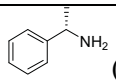
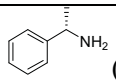
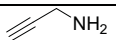
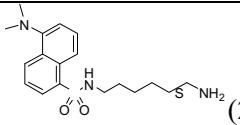
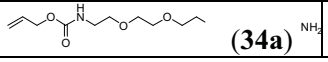
General synthetic procedures

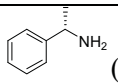
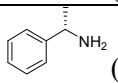
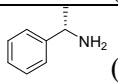
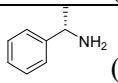
All commercial reagents (Sigma-Aldrich, Spectrum, MP Biomedicals, Alfa Aesar, TCI America, Acros) were used as provided unless otherwise indicated. Aldehyde-Reactive Probe **9** (*N*-(aminooxyacetyl)-*N*'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt) was purchased from Molecular Probes. Allyl 2-(2-(2-aminoethoxy)ethoxy)ethylcarbamate[1] (**34a**), levulinic acid succinimidyl ester[2] **(43a)**, 3-azidopropionic acid succinimidyl ester[3] **(43b)**, 2-azidoethanamine[4] **(42),** N-(7-dimethylaminocoumarin-4-acetate)-ethylenediamine[5] **(38b)**, 7-dimethylaminocoumarin-4-acetic acid[5] (**31**), para-methoxy-phenyl (PMP) protected pantothenate[6] **(39),** fluorescent enzyme probe **13**[6], 3-(2-hydroxyethoxy)propanenitrile[7] (**46**), and mono-dansylated ethylenediamine[8] (**54**) were prepared according to published literature procedures. All reactions were carried out under argon atmosphere in dry solvents with oven-dried glassware and constant magnetic stirring unless otherwise noted. Triethylamine and ethyl-*N,N*-diisopropylamine (DIPEA) were dried over sodium and freshly distilled. ^{1}H-NMR spectra were taken at 300, 400, or 500 MHz and ^{13}C-NMR spectra were taken at 100.6 or 75.5 MHz on Varian NMR spectrometers and standardized to the NMR solvent signal as reported by Gottlieb.[9] Multiplicities are given as s=singlet, d=doublet, t=triplet, q=quartet, p=pentet, dd=doublet of doublets, bs=broad singlet, bt=broad triplet, m=multiplet using integration and coupling constant in Hertz. TLC analysis was performed using Silica Gel 60 F254 plates (EM Scientific) and visualization was accomplished with UV light (λ=254 nm) and/or the appropriate stain (iodine, 2,4-dinitrophenylhydrazine, cerium molybdate, ninhydrin). Silica gel chromatography was carried out with Silicycle 60 Angstrom 230-400 mesh according to the method of Still[10]. TLC prep plate purification was performed with EMD Silica Gel 60 F_{254} pre-coated plates. Electrospray (ESI) and fast atom bombardment (FAB) mass spectra were obtained at the UCSD Mass Spectrometry Facility using a Finnigan LCQDECA mass spectrometer and a ThermoFinnigan MAT- 900XL mass spectrometer, respectively.

General Procedures and Spectroscopic Data for Microwave Assisted Nucleophilic Ring Opening of D-(-)-Pantolactone

General synthetic procedures for microwave chemistry All commercial reagents (Spectrum, MP Biomedicals, Lancaster, TCI America) were used as provided unless otherwise indicated. In both general procedures detailed below the reaction vessel was *not* purged or put under an inert atmosphere. Mono-dansylated hexane diamine (**20**) was synthesized by a procedure analogous to **54**.[8] Allyl 2-(2-(2-aminoethoxy)ethoxy)ethylcarbamate (**34a**) was prepared as previously described.[1] Microwave heated reactions were performed in a SmithSynthesizer from Biotage AB. Reaction temperatures were determined using the built-in, on-line IR-sensor. Racemization was checked for by reaction of the chiral amine (*S*)-(-)-α-methylbenzylamine (**19**) with (D)-(-)-pantolactone. Alternately in the case of the ring-opening of D-(-)-pantolactone by allyl 2-(2-(2-aminoethoxy)ethoxy)ethylcarbamate (**34a**) the di-ester with R-(-)-2-phenylbutyric acid was synthesized (see SI II for spectra). Examination of the ^{1}H-NMR spectra in multiple solvents showed a single diastereomer. The product of table entries 1 and 7 (see below data table for table entry #s) were isolated as described in the general procedures below. Comparison of the yield compared to that predicted from the ^{1}H-NMR spectra of the homogeneously stirred crude reaction mixture showed good correlation, so yield of table entries 2-3 and 8-10 were estimated from evaluation of the crude ^{1}H-NMR spectra without isolation.



Entry	Amine	Solvent	Temp (^{o}C)	Time (hr)	Yield
1	 (**19**)	EtOH	160	0.5	91%
2	 (**19**)	DMF	165	0.5	63%
3	 (**19**)	THF	110	0.5	44%
4	 (**41**)	EtOH	160	0.5	42%
5	 (**20**)	EtOH	160	0.5	82%
6	 (**34a**)	MeOH	135	0.5	70%

7	 (**19**)	EtOH	Reflux	7	97%
8	 (**19**)	MeOH	Reflux	7	84%
9	 (**19**)	CH_3CN	Reflux	22	83%
10	 (**19**)	DME	Reflux	12	30%

Model procedure for microwave assisted nucleophilic ring-opening of pantolactone (D)-(-)-pantolactone (390 mg, 3 mmol), NEt_3 (410 μL, 3 mmol), and (*S*)-(-)-α-methylbenzylamine (130 μL, 1 mmol) were dissolved in EtOH (2 mL). The reaction vessel was sealed and subjected to microwave reaction conditions with mechanical stirring. Specific temperature data for nucleophilic ring opening of D-(-)-pantolactone in EtOH was: 120 s ramp time (rt-160 ^{o}C) and 1800 s hold time at 160 ^{o}C. After cooling to room temperature the reaction mixture was vented and the solvent removed under reduced pressure, before being purified by flash chromatography (1:4 EtOAc/hexanes to 2:1 EtOAc/hexanes) to yield a thin white film (228 mg, 91%). NMR data was collected in both deuterated MeOH and DMSO to determine whether a single diastereomer had been formed. ^{1}H-NMR (400 MHz, $(CD_3)_2SO$) δ 8.00 (d, *J*= 8.0 Hz, 2H), 7.33-7.17 (m, 5H), 5.33 (d, *J*=6.0 Hz, 1H), 4.94 (p, *J*=8.0 Hz, 1H), 4.47 (t, *J*=5.6 Hz, 1H), 3.74 (d, *J*=6.0 Hz, 1H), 3.28-3.24 (dd, *J*=10, 6.0 Hz, 1H), 3.15-3.11 (dd, *J*=10.4, 5.6 Hz, 1H), 1.36 (d, *J*=7.2 Hz, 3H), 0.76 (s, 3H), 0.75 (s, 3H). ^{1}H-NMR (400 MHz, CD_3OD) δ 8.06 (d, *J*=7.6 Hz, 1H), 7.36-7.20 (m, 5H), 5.05 (p, *J*=7.2 Hz, 1H), 3.94 (s, 1H), 3.45 (d, *J*=10.8 Hz, 1H), 3.35 (d, *J*=11.2 Hz, 1H), 0.884 (s, 6H). ^{13}C-NMR (100.6 MHz, CD_3OD) δ 173.8, 143.9, 128.4, 127.0, 126.1, 76.2, 69.3, 48.7, 39.4, 21.2, 20.3, 19.8. MS (ESI) $[M+H]^+$ 252.01.

Model procedure for thermal (reflux) nucleophilic ring-opening of pantolactone (D)-(-)-pantolactone (390 mg, 3 mmol), NEt_3 (410 μL, 3 mmol), and (*S*)-(-)-α-methylbenzylamine (130 μL, 1 mmol) were dissolved in EtOH (3.5 mL). A reflux condenser was attached and the vessel was warmed to a gentle reflux with mechanical stirring for 7 hours. Reaction work-up as described above yielded the product pantoic acid conjugate as a thin white film with spectral properties identical to those listed above.

Spectral Data for Table Entry 4 (propargyl amine pantolactone conjugate) Table entry 4 was subjected to the microwave reaction conditions detailed above, using propargyl amine **41** as the limiting reagent (1 mmol). Reaction mixture was purified by flash chromatography (EtOAc to 5% MeOH/EtOAc) to yield a clear

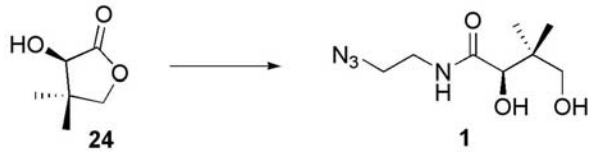
oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.19 (bt), 4.05 (m, 3H), 3.50 (s, 2H), 2.24 (t, *J*=2.8 Hz, 1H), 0.99 (s, 3H), 0.92 (s, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 173.4, 79.4, 77.7, 71.8, 71.4, 39.6, 28.9, 21.2, 20.6. MS (ESI) $[M+H]^+$ 185.99.

Spectral Data for Table Entry 5 (mono-dansyl hexanediamine pantolactone conjugate) Table entry 5 was subjected to microwave reaction conditions in EtOH as detailed above, using mono-dansyl hexanediamine **20** as the limiting reagent (0.5 mmol). Reaction mixture was purified by flash chromatography (EtOAc to 5 % MeOH/EtOAc) to yield a green fluorescent oil. ^{1}H-NMR (400 MHz, $(CD_3)_2SO$) δ 8.43 (d, *J*=8.4 Hz, 1H), 8.28 (d, *J*=8.4Hz, 1H), 8.07 (dd, *J*=1.6, 7.2 Hz, 1H), 7.85 (t, *J*=5.6 Hz), 7.61-7.54 (m, 2H), 7.22 (d, *J*=7.6 Hz, 1H), 5.27 (d, *J*=5.2 Hz, 1H), 4.47 (t, *J*=5.6 Hz, 1H), 3.65 (d, *J*=5.6 Hz, 1H), 3.26 (dd, *J*=6.0, 10.4 Hz), 3.14 (dd, *J*=4.8, 10.0 Hz), 3.00-2.86 (m, 1H), 2.80 (s, 6H), 2.73 (q, *J*=6.8 Hz, 1H), 1.25-1.13 (m, 4H), 1.05-0.97 (m, 4H), 0.76 (s, 3H), 0.75 (s, 3H). ^{13}C-NMR (100.6 MHz, $(CD_3)_2SO$) δ 173.4, 152.0, 136.9, 129.9, 129.8, 129.7, 128.9, 128.4, 124.2, 119.9, 115.8, 75.9, 68.8, 45.8, 42.9, 39.6, 38.6, 29.7, 29.6, 26.5, 26.3, 21.6, 21.1. MS (ESI) $[M+H]^+$ 480.14.

Spectral Data for R-(-)-2-phenylbutyric acid Di-Ester of Table Entry 6 (Allyl 2-(2-(2-aminoethoxy)ethoxy)ethylcarbamate pantolactone conjugate) The di-ester of the reaction product of table-entry 6 was prepared according to the procedure of Mosher.[11] ^{1}H-NMR (400 MHz, ($CDCl_3$) δ 7.32-7.27 (m, 5H), 5.91 (m, 1H), 5.44 (s, 1H), 5.29 (d, *J*=17 Hz, 2H) 5.27 (bs, 1H), 5.19 (d, *J*=10.5 Hz, 1H), 4.94 (**s**, 1H), 4.55 (d, *J*=5.5 Hz, 2H), 3.92 (d, *J*=11 Hz, 1H), 3.79 (d, *J*=11 Hz, 1H), 3.52-3.23 (m, 10H), 2.97 (t, *J*=8.5 Hz, 2H), 2.12 (m, 4H), 1.83 (m, 2H), 0.91-0.83 (m, 9H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 173.9, 172.2, 168.0, 156.5, 139.1, 138.6, 133.1, 129.17, 128.8, 128.3, 128.2, 127.9, 127.5, 117.9, 76.6, 70.5, 70.3, 69.5, 69.4, 65.7, 53.7, 53.4, 41.0, 38.8, 37.6, 26.3, 25.7, 21.4, 20.7, 12.4, 12.2. (*m/z*) $[M]^+$ calcd for $C_{36}H_{50}N_2O_9$ 654.3511, found 654.3513.

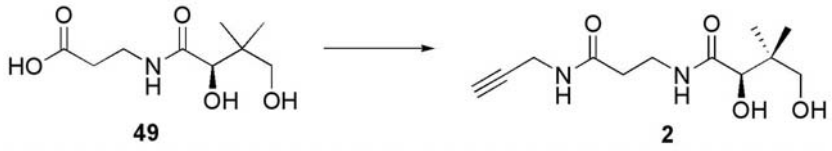
Synthetic Procedures and Spectroscopic Data for Pantetheine Analogs and Detection Agents 1-18

(a) Synthesis of C7-azido-pantothenate (1)



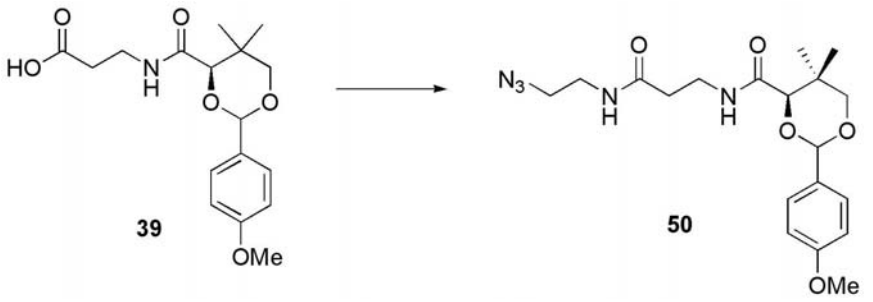
2-azidoethanamine **42** (42 mg, 0.49 mmol), (D)-(-)-pantolactone (191 mg, 1.47 mmol), and triethylamine (0.134 mL, 0.98 mmol) were dissolved in MeOH (2 mL) and heated to a gentle reflux for four hours. The solvent was removed under reduced pressure to yield a clear oil, which was purified by flash chromatography (EtOAc to 5% MeOH: EtOAc) to yield **1** (35.5 mg, 33.8%) as an oil. ^{1}H-NMR (500 MHz, CD_3OD) δ 7.22 (s, 1H), 4.04 (s, 1H), 3.49-3.44 (m, 6H), 0.99 (s, 3H), 0.92 (s, 3H); ^{13}C-NMR (75.5 MHz, CD_3OD) δ 174.0, 77.9, 71.5, 51.0, 39.5, 38.3, 21.3, 21.3, 20.5. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_8H_{16}N_4O_3$ 216.1217, found 216.1214.

(b) Synthesis of C10-Propargyl Pantetheine (2)

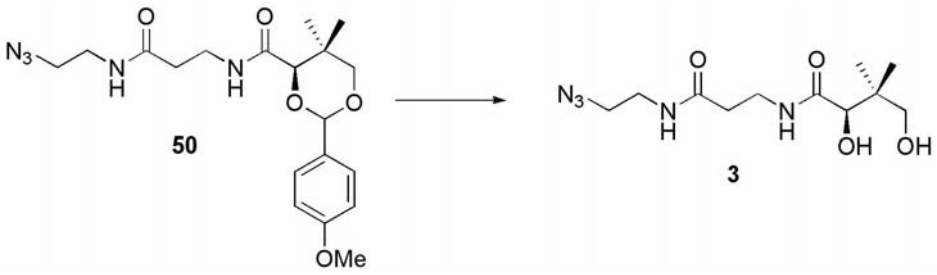


Propargylamine **41** (220mg, 4.00 mmol), pantothenate **49** (876 mg, 4.00 mmol), EDAC (1533 mg, 8.00 mmol), and HOBt monohydrate (1530 mg, 10.00 mmol) were combined and dissolved in dry DMF (10 mL). DIPEA (1.39 mL, 8.00 mmol) was added, and the reaction was allowed to stir overnight. The solvent was removed under reduced pressure, and the resultant oil was purified by column chromatography (EtOAc to 5% MeOH/EtOAc) to yield **2** (387 mg, 38%) as a thick yellow oil. ^{1}H-NMR (300 MHz, $(CD_3)_2CO$) δ 7.95 (s, 1H), 7.76 (s, 1H), 5.17 (bs, 1H), 4.55 (bs, 1H), 3.96 (m, 3H), 3.46 (m, 5H), 2.48 (t, *J*= 6.6 Hz, 2H), 0.91 (s, 3H), 0.87 (s, 3H). ^{13}C-NMR (75.5 MHz, $(CD_3)_2CO$) δ 174.1, 171.5, 80.5, 76.6, 71.8, 69.9, 39.5, 35.4, 35.3, 28.5, 20.9, 20.4. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{12}H_{20}O_4N_2$, 256.1418, found 256.1414.

(c) Synthesis of C11-Azido-Pantetheine (3)

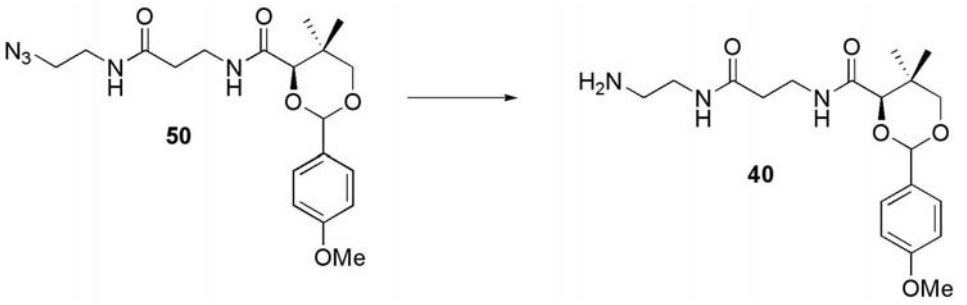


PMP-protected pantothenate **39** (3.11g, 9.23 mmol), 2-azidoethanamine **42** (1.58 g, 9.23 mmol), EDAC (3.49 g, 18.22 mmol), and HOBt (3.52 g, 22.99 mmol) were combined and dissolved in dry DMF (10 mL). DIPEA (4.7 mL, 27.03 mmol) was added, and the reaction was allowed to stir overnight. The solvent was removed under reduced pressure, and the resultant oil was resuspended in EtOAc and washed with saturated $NaHCO_3$ (3 x 45 mL) and brine (1 x 45 mL). Elution through a plug of Si-gel with EtOAc gave the pure product **50** (3.03 g, 82%) as a yellow oil. ^{1}H-NMR (500 MHz, $(CD_3)_2SO$) δ 8.13 (1H, bt), 7.46 (1H, bt), 7.42 (d, 2H, *J*= 9.0 Hz), 6.91 (d, 2H, *J*= 8.5Hz), 5.50 (s, 1H), 4.06 (s, 1H), 3.744 (s, 2H), 3.61 (q, 2H, *J*= 8.0 Hz), 3.35-3.19 (m, 6H), 2.27 (t, 2H, *J*= 7.0 Hz), 0.98 (s, 3H), 0.93 (s, 3H). ^{13}C-NMR (75.5 MHz, $(CD_3)_2CO$) δ 171.7, 169.7, 160.4, 130.3, 127.8, 113.9, 101.5, 84.0, 78.5, 50.7, 39.1, 35.8, 35.1, 33.2, 22.0, 19.3. MS (ESI) $[M+Na]^+$ 428.08. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{19}H_{27}O_5N_5$, 405.2007, found 405.2005.

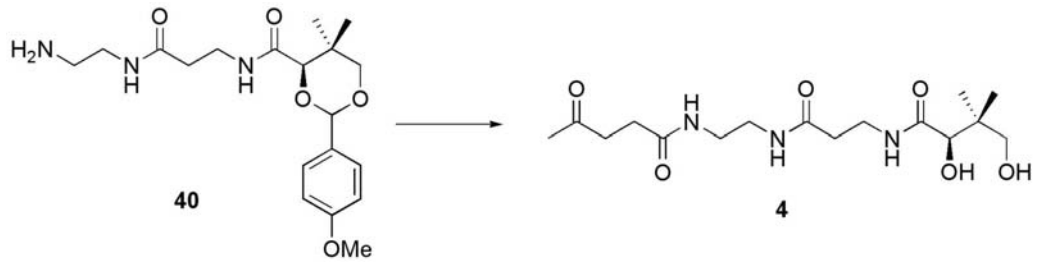


Compound **50** (136 mg, 0.34 mmol) was dissolved in THF (1 mL) and 1 M HCl (1 mL) was added. After 1.5 h at rt, AG-1-X8 Strong Basic anionic exchange resin was added to the reaction mixture until the solution was neutral. The crude product was filtered and solvent removed under reduced pressure and re-dissolved in methanol (40 mL) and washed with hexanes (3 x 20 mL). The crude product was concentrated and purified by flash chromatography (CH_2Cl_2 to 15% MeOH: CH_2Cl_2) to yield **3** (76 mg, 79%) as a clear oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.55 (s, 1H), 6.89 (s, 1H), 3.99 (s, 1H), 3.55 (q, *J*=6.0 Hz, 2H), 3.40 (m, 6H), 2.47(t, *J*= 6.0 Hz, 2H), 0.97 (s, 3H), 0.91(s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 174.2, 172.2, 77.6, 70.9, 50.8, 39.5, 39.1, 35.9, 35.4, 21.5, 20.7. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{11}H_{21}O_4N_5$, 287.1592, found 287.1588.

(d) Synthesis of N12-Pantetheine-Levulinate (4)



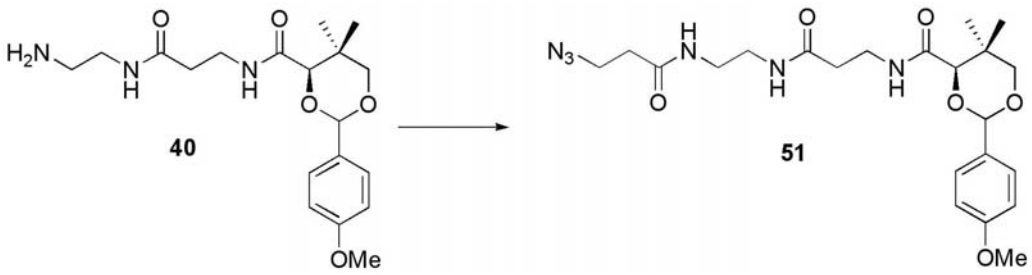
To a solution of the azide **50** (475 mg, 1.17 mmol) in THF (2.5 mL), PPh_3 (338 mg, 1.29 mmol), and water (250 μL) were added and stirred at room temperature for 24 h. Evaporation of the solvent under reduced pressure gave oil from which amine **40** was isolated by flash chromatography (CH_2Cl_2 to 15% MeOH: CH_2Cl_2 : 1.5% NEt_3) as a brown-yellow oil (350 mg, 78%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.41 (d, *J*= 8.5 Hz, 2H), 7.06 (bt, 1H), 6.92 (d, *J*= 8.5 Hz, 2H), 6.37 (bt, 1H), 5.45 (s, 1H), 4.08 (s, 1H), 3.82 (s, 3H), 3.68 (q, , *J*= 7.3 Hz, 2H), 3.55 (m, 2H), 3.25 (m, 2H), 2.76 (bs), 2.66 (q, *J*= 8.5 Hz, 2H), 2.43 (t, *J*= 7.3 Hz, 2H), 1.10 (s, 3H), 1.09 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 171.7, 169.8, 160.4, 130.3, 127.8, 113.9, 101.5, 84.0, 78.6, 55.6, 41.2, 36.2, 33.4, 33.3, 22.1, 19.4. MS (ESI) $[M+H]^+$ 380.11. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{19}H_{29}O_5N_3$, 379.2102, found 379.2109.



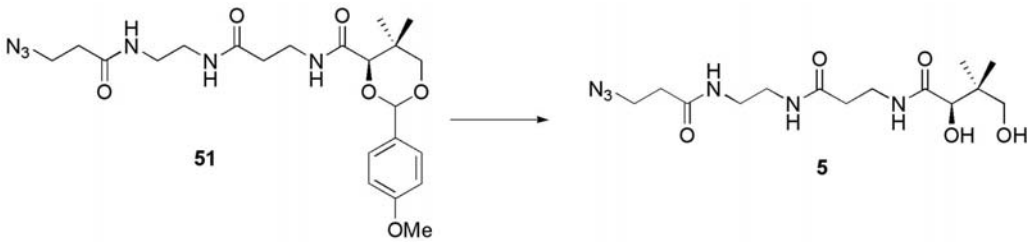
To a stirring solution of NHS ester **43a** (126 mg, 0.59 mmol) in DCM (2 mL) was added triethylamine (122 μL, 0.89 mmol) and amine **40** (112 mg, 0.30 mmol). The reaction was monitored by TLC and stirred at room temperature for 8h. Evaporation of the solvent gave an oil which was ran through a plug of silica (10% MeOH/DCM eluent) to give the crude PMP-protected ketone product, as evidenced by TLC and ^{1}H-NMR analysis of the crude. This material was dissolved in THF (1 mL) and 1M HCl (1 mL) was added without further purification. After 1.5 h at rt, AG-1-X8 Strong Basic anionic exchange resin was added to the reaction mixture until the solution was neutral. The crude product was filtered and solvent removed under reduced pressure and re-dissolved in water (30 mL) and washed with CH_2Cl_2 (4 x 25 mL). The water layer was frozen and lyophilized to afford the product **4** (45 mg, 42%) as the hydrate of the ketone.

^{1}H-NMR (500 MHz, CD_3OD) δ 3.89 (s, 1H), 3.24-3.48 (m, 6H), 2.66 (bs, 2H), 2.26-2.63 (m, 2H), 2.03-2.18 (m, 2H), 1.91 (s, 1H), 1.48 (s, 3H), 0.890 (s, 3H), 0.882 (s, 3H). ^{13}C-NMR (100.6 MHz, D_2O, internal standard 0.01% MeOH) δ 178.7, 175.6, 174.6, 91.7, 76.4, 69.0, 39.2, 38.9, 38.6, 35.9, 35.9, 34.4, 29.3, 25.6, 21.1, 19.7. MS (ESI) $[M+Na]^+$ 382.16. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{16}H_{29}O_6N_3$, 359.2056 found 359.2056.

(e) Synthesis of C15-Azido-Pantetheine (5)



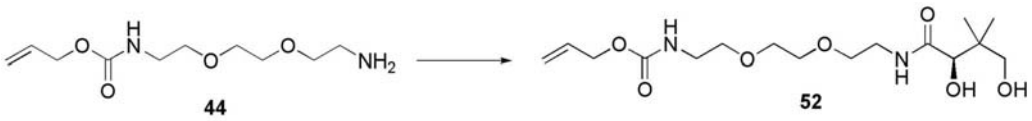
To a stirring solution of NHS ester **43b** (134 mg, 0.63 mmol) in dry DCM (2 mL) was added triethylamine (150 µL, 1.05 mmol) and amine **40** (80 mg, 0.21 mmol). The reaction was monitored by TLC and stirred at room temperature for 8 h. Evaporation of the solvent gave a crude oil from which product **51** was isolated by flash chromatography ($CHCl_3$ to 5% MeOH: $CHCl_3$) as a yellow oil (95 mg, 95%) contaminated with N-hydroxysuccinimide. This product was carried onto the next step without further purification. ^{1}H-NMR (300MHz, $CDCl_3$) δ 7.4 (d, *J*= 9 Hz, 1H), 7.1-7.3 (m, 3H), 6.88 (d, *J*= 8.7 Hz, 1H), 5.43 (s, 1H), 4.06 (s, 1H), 3.78 (s, 3H), 3.60-3.69 (m, 3H), 3.40-3.55 (m, 3H), 3.31 (s, 2H), 2.39 (m, 4H), 1.06 (s, 3H), 1.02 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 178.7, 175.6, 174.6, 91.7, 76.4, 69.0, 39.2, 38.9, 38.6, 35.9, 35.9, 34.4, 29.3, 25.6, 21.1, 19.7. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{22}H_{32}N_6O_6$, 476.2378 found 476.2381.



Compound **51** (80 mg, 0.17 mmol) was dissolved in THF (1 mL) and 1 M HCl (1 mL). After 2 h at rt, AG-1-X8 Strong Basic anionic exchange resin was added to the reaction mixture until the solution was neutral. The crude product was filtered and solvent removed under reduced pressure and re-dissolved in methanol (30 mL) and washed with hexanes (3 x 20 mL). The crude product was concentrated and purified by flash chromatography (1% MeOH: CH_2Cl_2 to 15% MeOH: CH_2Cl_2) to yield pure **5** (45 mg, 75%) as a colorless oil. ^{1}H-NMR (500 MHz,

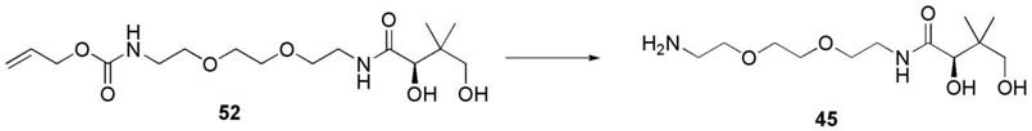
$(CD_3)_2CO$) δ 7.65 (bs, 1H), 7.54 (bs, 1H), 7.48 (bs, 1H), 5.00 (d, *J*=5 Hz, 1H), 4.26 (t, *J*=3.5 Hz, 1H), 3.94 (d, *J*=5 Hz, 1H), 3.57 (t, *J*=6.25 Hz, 2H), 3.50 (q, *J*=6.0 Hz, 4H), 3.49 (t, *J*=6.25 Hz, 2H), 3.27 (m, 4H), 2.47 (t, *J*=6.5 Hz, 2H), 2.41 (t, *J*=6.5 Hz, 2H), 0.95 (s, 3H), 0.87 (s, 3H). ^{13}C-NMR (75.5 MHz, $(CD_3)_2CO$) δ 173.7, 171.7, 170.5, 76.8, 70.0, 47.4, 39.5, 39.2, 35.8, 35.3, 35.2, 21.2, 20.0. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{14}H_{26}N_6O_5$ 358.1959, found 358.1961

(f) Synthesis of PEG2-Levulinate-Pantoic Acid Conjugate (6)



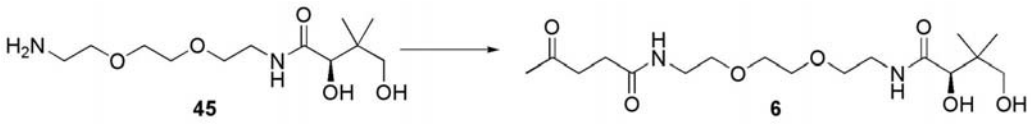
44 52

Mono-allyl carbamate **44** (228 mg, 0.98 mmol), (D)-(-)-pantolactone (408 mg, 3.14 mmol), and triethylamine (430 µL, 3.14 mmol) were dissolved in EtOH (5 mL). The reaction vessel was sealed and subjected to microwave reaction conditions (160°C, 30 min, see above) with mechanical stirring. After cooling to room temperature the reaction mixture was vented and the solvent removed under reduced pressure, before being purified by flash chromatography (EtOAc to 7.5% MeOH: EtOAc) to yield **52** (288 mg, 82%) as a yellow oil. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.55 (bs, 1H), 5.91 (m, 1H), 5.5 (bs, 1H), 5.30 (dd, *J*= 1.5, 17 Hz, 1H), 5.21 (d, *J*= 10.5 Hz, 1H), 4.55 (d, *J*= 5 Hz, 2H), 4.02 (s, 1H), 3.63-3.36 (m, 14H), 1.05 (s, 3H), 0.90 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 173.7, 156.8, 133.0, 118.0, 77.6, 71.0, 69.8, 65.8, 41.0, 39.5, 38.9, 21.9, 20.3. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{16}H_{30}N_2O_7$, 362.2048, found 362.2052.



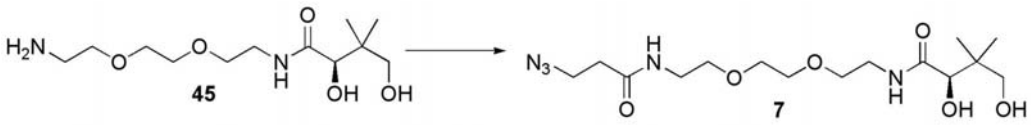
52 45

To a solution of allyl-carbamate **52** (628 mg, 1.74 mmol) in dried THF (7 mL) was added dimedone (1.70 g, 12.2 mmol), PPh_3 (91 mg, 0.35 mmol) and $Pd(Ph_3P)_4$ (100 mg, 0.087 mmol). The flask was allowed to stir at room temperature overnight. The solvent was removed under reduced pressure to give a crude oil, which was taken up in CH_2Cl_2 and purified by flash chromatography (CH_2Cl_2 to 15% MeOH: CH_2Cl_2: 1.5% NEt_3) to yield the pure product **45** as a yellow oil (310 mg, 64%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.43 (bs, 1h), 3.94 (s, 1H), 3.59-3.49 (m, 11H), 3.31 (m, 1H), 2.83 (m, 2H), 0.98 (s, 1H), 0.87 (s, 1H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 174.4, 77.1, 71.8, 71.0, 70.7, 70.2, 69.8, 41.2, 39.4, 38.9, 22.2, 20.5. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{12}H_{26}N_2O_3$, 278.1836, found 278.1835.



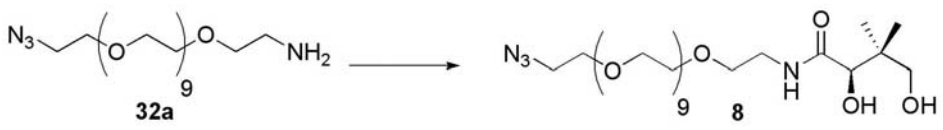
To a stirring solution of NHS ester **43a** (114 mg, 0.54 mmol) in dry DCM (1 mL) was added triethylamine (222 µL, 1.62 mmol) and amine **45** (75 mg, 0.27 mmol). The reaction was monitored by TLC and stirred at room temperature for 4 h. Evaporation of the solvent gave a crude oil from which product **6** was isolated by flash chromatography ($CHCl_3$ to 5% MeOH: $CHCl_3$). Pure fractions were taken to give the product as a colorless oil (34 mg, 34 %). N-hydroxysuccinimide contaminated product was also recovered. δ ^{1}H-NMR (500 MHz, $CDCl_3$) 7.34 (bs, 1H), 6.50 (bs, 1H), 4.45 (d, *J*=5 Hz, 1H), 4.01 (d, *J*= 4.5 Hz, 2H), 3.96 (bs, 1H), 3.57-3.32 (m, 14H), 2.80 (t, *J*= 6.5 Hz, 2H), 2.41 (t, *J*= 6.5 Hz, 2H), 2.17 (s, 3H), 1.02 (s, 3H), 0.90 (s, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 20.19, 173.3, 172.2, 77.2, 70.9, 70.4, 70.3, 70.1, 69.6, 40.9, 39.4, 38.7, 38.5, 29.9, 29.8, 22.1, 20.0. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{17}H_{32}N_2O_7$, 276.2204, found 376.2216.

(g) Synthesis of PEG2-(β-Ala-N_3)-Pantoic Acid Conjugate (7)



To a stirring solution of NHS ester **43b** (141 mg, 0.67 mmol) in dry DCM (1 mL) was added triethylamine (91 µL, 0.72 mmol) and amine **45** (62 mg, 0.22 mmol). The reaction was monitored by TLC and stirred at room temperature for 4 h. Evaporation of the solvent gave a crude oil which was taken up in CH_2Cl_2 and isolated by flash chromatography (EtOAc to 10% MeOH: EtOAc). Pure fractions were taken to give **7** as a colorless oil (25 mg, 30 %). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.29 (s, 1H), 6.57 (s, 1H), 4.36 (d, *J*= 5.5 Hz, 1H), 4.00 (d, *J*= 5.5 Hz, 2H), 3.93 (bs, 1H), 3.60-3.40 (m, 16H), 2.43 (t, *J*= 6.0 Hz, 2H), 1.03 (s, 3H), 0.90 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 173.6, 170.6, 77.5, 71.0, 70.5, 70.5, 70.3, 69.9, 47.6, 39.7, 39.6, 38.9, 36.0, 22.2, 20.3. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{15}H_{29}N_5O_6$, 375.2112, found 375.2121.

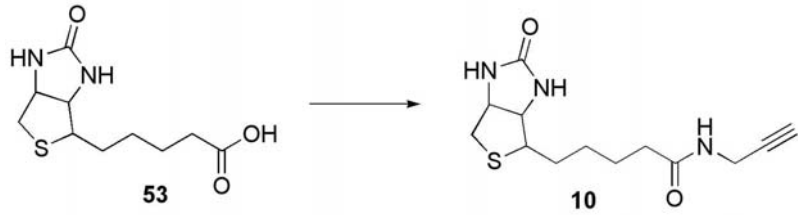
(h) Synthesis of PEG10-N_3-Pantoic Acid Conjugate (8)



32a (50 mg, 0.095 mmol), (D)-(-)-pantolactone (37 mg, 0.285 mmol), and triethylamine (39 µL, 0.0285 mmol) were dissolved in EtOH (2.25 mL). The reaction vessel was sealed and subjected to microwave reaction conditions

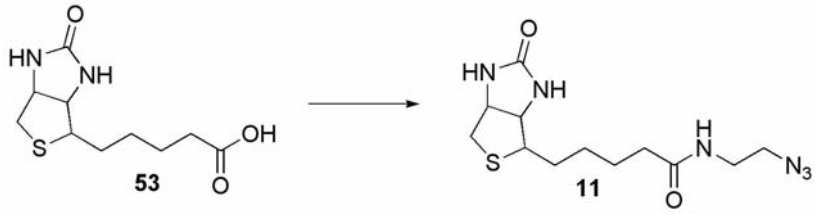
(160°C, 30 min, see above) with mechanical stirring. After cooling to room temperature the reaction mixture was vented and the solvent removed under reduced pressure, before being purified by flash chromatography (EtOAc to 15% MeOH: EtOAc) to yield **8** (44 mg, 71%) as a yellow oil. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 4.45 (bs, 1H), 4.12 (d, *J*=6.0 Hz, 1H), 3.99 (d, *J*= 6.0 Hz, 1H), 3.81 (bs, 1H), 3.66-3.35 (m, 46H), 1.03 (s, 3H), 0.86 (s, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 173.6, 77.2, 71.0, 70.9, 70.9, 70.8, 70.8, 70.5, 70.3, 69.9, 50.9, 39.0, 22.5, 20.2. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{28}H_{56}N_4O_{13}$ 656.3838, found 656.3833.

(i) Synthesis of Biotin-Alkyne (10)



Biotin **53** (40 mg, 0.16 mmol), propargylamine **41** (15 µL, 0.21 mmol), HATU (68 mg, 0.18 mmol), and HOBt monohydrate (27 mg, 0.18 mmol) were combined and dissolved in dry DMF (3 mL). DIPEA (85 µL, 0.50 mmol) was added, and the reaction was allowed to stir for 4 hours. The solvent was removed under reduced pressure to yield a crude oil which and purified by column chromatography (CH_2Cl_2 to 12.5% MeOH/CH_2Cl_2) to yield **10** (34 mg, 74%)**,** a white precipitate with limited organic solubility. ^{1}H-NMR (400 MHz,$(CD_3)_2SO$) δ 8.21 (bs, 1H), 6.42 (s, 1H), 6.35 (s, 1H), 4.28 (t, *J*= 6.4 Hz, 1H), 4.10 (q, *J*= 3.2 Hz, 1H) 3.81 (d, *J* = 2.4 Hz, 2H), 3.06 (m, 1H), 2.80 (dd, *J*= 4.8, 12.4 Hz, 1H), 2.55 (d, *J*=12.4 Hz, 1H), 2.06 (t, *J*= 7.2 Hz, 2H), 1.59-1.24 (m, 6H). ^{13}C-NMR (100.6 MHz, $(CD)_2SO$) δ 172.5, 163.4, 82.0, 73.5, 61.7, 59.9, 56.1, 40.5, 28.8, 28.7, 28.3. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{13}H_{19}N_3O_2S$ 281.1192, found 281.1191.

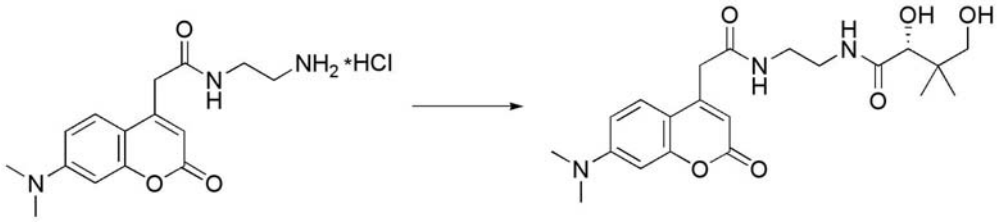
(j) Synthesis of Biotin-Azide (11)



Biotin **53** (100 mg, 0.41 mmol), 2-azidoethanamine **42** (70 mg, 0.82 mmol), HATU (171 mg, 0.45 mmol), and HOBt monohydrate (69 mg, 0.45 mmol) were combined and dissolved in dry DMF (2 mL). DIPEA (350 µL, 2.05 mmol) was

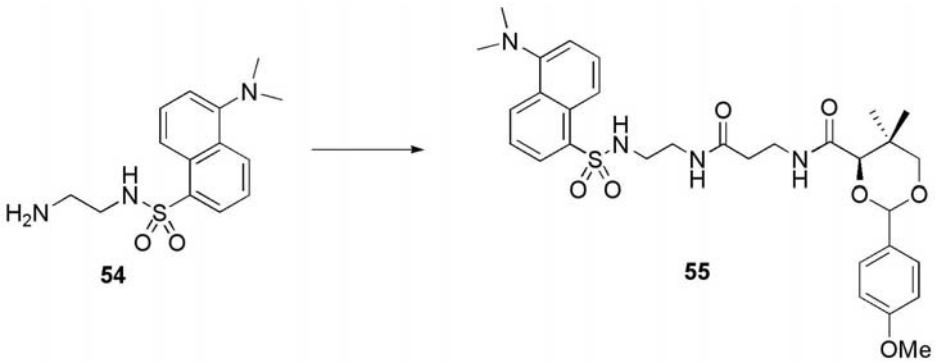
added, and the reaction was allowed to stir for 4 hours. The solvent was removed under reduced pressure to yield a crude oil which was taken up in CH_2Cl_2 and purified by column chromatography (CH_2Cl_2 to 10% MeOH/CH_2Cl_2) to yield **11** contaminated with HOBt. Tituration with acetone yielded the pure product **11** (87 mg, 68%)**,** again as a white precipitate with limited organic solubility. ^{1}H-NMR (500 MHz,$(CD_3)_2SO$) δ 8.04 (s, 1H), 6.42 (s, 1H), 6.35 (s, 1H), 4.28 (t, *J*= 7.0 Hz, 1H), 4.11 (q, *J*= 5 Hz, 1H), 3.22 (m, 2H), 3.08 (m, 1H), 2.80 (dd, *J*= 5.5, 12.5 Hz, 1H), 2.56 (d, *J*= 12 Hz, 1H), 2.06 (t, *J*= 7.0 Hz, 2H), 1.6-1.2 (m, 6H). ^{13}C-NMR (75.5 MHz, $(CD_3)_2SO$) δ 174.4, 164.0, 61.9, 60.0, 56.1, 50.5, 40.4, 35.8, 28.8, 28.5, 25.8. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{12}H_{20}N_6O_2S$, 312.1363, found 312.1366.

(k) Synthesis of DMACA-ethylenediamine-Pantoic Acid Conjugate (12)

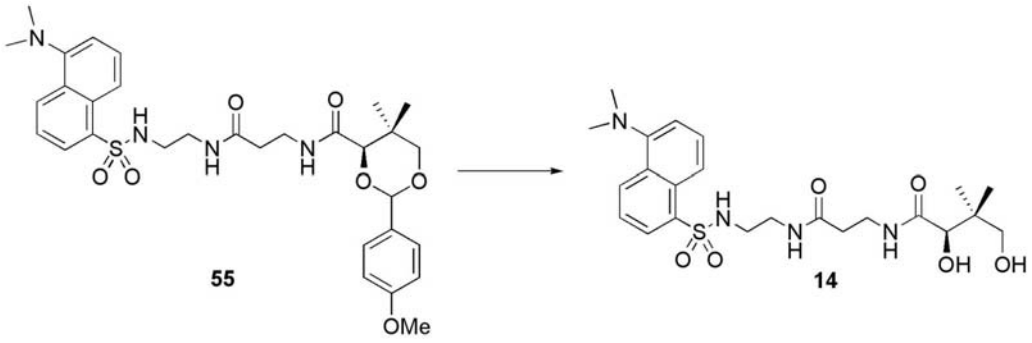


38c (20 mg, 0.069 mmol), (D)-(-)-pantolactone (26 mg, 0.204 mmol), and triethylamine (92 µL, 0.69 mmol) were dissolved in MeOH (2 mL), heated to a gentle reflux, and monitored by TLC. After 18 hrs the solvent was removed under reduced pressure to yield an oil, which was purified by flash chromatography (EtOAc to 15% MeOH: EtOAc) to yield **12** (25 mg, 85%) as a fluorescent oil. ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.55 (d, *J*= 9.0 Hz, 1H), 6.77 (dd, *J*= 2.7, 9.0 Hz, 1H), 6.56 (d, *J*= 2.7 Hz, 1H), 6.05 (s, 1H), 3.86 (s, 1H), 3.68 (d, *J*= 0.6 Hz, 2H), 3.46-3.34 (m, 6H), 3.07 (s, 6H), 0.904 (s, 3H), 0.902 (s, 3H). ^{13}C-NMR (75.5 MHz, CD_3OD) δ 175.3, 170.3, 163.2, 155.9, 153.6, 151.5, 125.7, 109.4, 109.2, 108.6, 97.5, 76.2, 68.9, 39.2, 39.2, 39.1, 38.9, 38.3, 20.3, 20.0. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{29}N_3O_6$, 419.2051, found 419.2048.

(l) Synthesis of Dansyl-ethylenediamine-Pantothenate Conjugate (14)



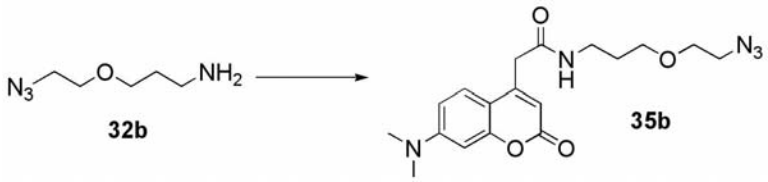
Fluorescent amine **54** (50 mg, 0.15 mmol) was coupled to PMP-protected pantothenate **39** (50 mg, 0.15 mmol) by the method of Clarke[7] to yield **55** (65 mg, 71%) as a green fluorescent oil. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 8.52 (d, *J* = 8.5 Hz, 1H), 8.28 (d, *J*= 8.5 Hz), 8.18 (d, *J*= 7.5 Hz, 1H), 7.54-7.47 (m, 2H), 7.40 (d, *J*= 8.5 Hz, 1H), 7.16 (d, *J*= 7.5 Hz, 1H), 7.05 (bt, *J*= 6.5 Hz, 1H), 6.90 (d, *J*=8.5 Hz, 1H), 6.45 (bt, *J*= 6 Hz, 1H), 6.17 (bt, *J*= 5.5 Hz, 1H), 5.43 (s, 1H), 4.10 (s, 1H), 3.77 (s, 3H), 3.65 (q, *J*= 6.5 Hz, 2H), 3.49-3.4 (m, 2H), 3.28-3.16 (m, 2H), 2.99-2.95 (m, 2H), 2.87 (s, 6H), 2.26 (q, *J*=6 Hz, 2H), 1.06 (d, *J*= 10 Hz, 6H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 172.0, 170.1, 160.5, 152.2, 135.0, 130.7, 130.4, 130.1, 129.7, 129.7, 128.6, 127.8, 123.5, 119.1, 115.5, 114.0, 84.0, 78.6, 77.6, 77.3, 76.9, 55.5, 43.1, 39.6, 36.3, 35.3, 33.3, 22.1, 19.4. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{31}H_{40}N_4O_7S$ 612.2612, found 612.2618.



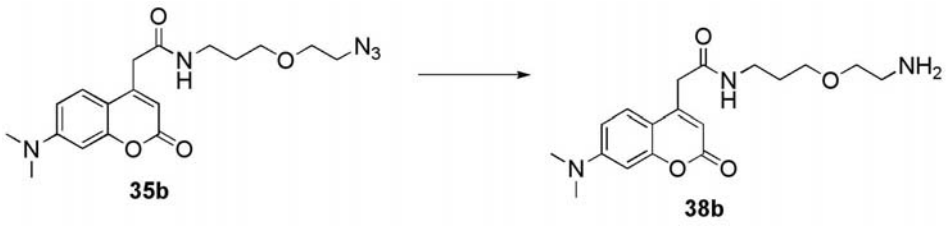
Compound **55** (40 mg, 0.065 mmol) was dissolved in THF (1 mL) and 1 M HCl (0.5 mL). After 1.5 h at rt, AG-1-X8 Strong Basic anionic exchange resin was added to the reaction mixture until the solution was neutral. The reaction mixture was filtered and solvent removed under reduced pressure to yield a crude oil which was purified by flash chromatography (CH_2Cl_2 to 15% MeOH: CH_2Cl_2) to yield **14** as a fluorescent green oil (25 mg, 79%). ^{1}H-NMR (500 MHz, $(CD_3)_2SO$) δ 8.44 (d, *J*= 8.5 Hz, 1H), 8.25 (d, *J*= 8 Hz, 1H), 8.08 (dd, *J*= 1, 7.5 Hz, 1H), 7.97 (bs, 1H), 7.83 (t, *J*= 6 Hz, 1H), 7.65-7.56 (m, 3H), 7.23 (d, *J*= 7.5 Hz, 1H), 5.32 (d, *J*= 5.5 Hz, 1H), 4.44 (t, *J*= 5.5 Hz, 1H), 3.66 (d, *J*= 5.5 Hz, 1H), 3.27-3.22 (m, 2H), 3.18-3.11 (m, 2H), 3.01 (q, *J*= 7.5 Hz, 2H), 2.83-2.75 (m, 8H), 2.14 (m, 2H), 0.72 (d, *J*= 12.5 Hz, 6H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 174.7, 172.9, 152.1, 134.7, 130.7, 130.0, 129.7, 129.5, 128.7, 123.5, 119.0, 115.5, 77.2, 70.5, 45.6, 42.8, 39.8, 39.4, 36.1, 35.8, 21.3, 20.9. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{20}H_{34}N_4O_6S$ 494.2194, found 494.2199.

(m)Synthesis of DMACA-β-Ala Probe–Pantoic Acid Conjugate I (15)

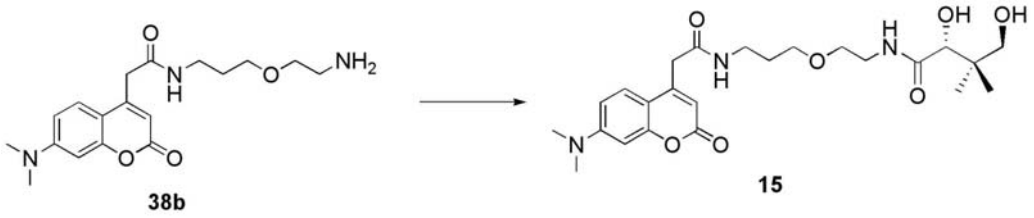
Compound **48** (300 mg, 1.23 mmol) was dissolved in dioxane (4 ml) and cooled to 0°C. 4M HCl in dioxane (2 mL 61.4 mmol) was added and the resulting reaction mixture was allowed to warm to room temperature and stirred overnight. The solvent was removed under reduced pressure, giving **32b** (214mg, 97%) as a light brown oil. ^{1}H-NMR (400MHz, CD_3OD) δ 3.64 (m, 4H), 3.40 (t, *J*= 4.8 Hz, 2H), 3.07 (t, *J*= 6.8 Hz, 2H), 1.943 (q, *J*= 6 Hz, 2H). ^{13}C-NMR (100.6 MHz, CD3OD) δ 69.9, 68.2, 50.6, 37.9, 27.4. HRMS (EI) calcd for $C_5H_{12}N_2O$ $[M]^+$: *m/z* 144.1006, found 144.1004.



Compound **32b** (200 mg, 1.11mmol), 7-dimethylaminocoumarin-4-acetic acid **31** (273 mg, 1.11 mmol), and EDAC (530.6 mg, 2.77 mmol) were combined and dissolved in dry DMF (8 mL). DIPEA (0.96 ml, 5.54 mmol) was added, and the reaction was allowed to stir 48 hours. The solvent was removed under reduced pressure, and the resultant oil and purified by flash chromatography (100% EtOAc) to yield the product **35b** (94 mg, 23%) as a yellow solid. ^{1}H-NMR (500 MHz, $CDCl_3$); δ 7.49 (d, *J*= 9, 1H), 6.63 (dd, *J*= 2.5, 8.5, 1H), 6.53(d, *J*= 2.5, 1H), 6.18 (b s, 1H), 6.053 (s, 1H), 3.61 (s, 2H), 3.53 (t, *J*=4.5, 2H), 3.49 (t, *J*= 5, 2H), 3.37 (q, *J*= 6, 2H), 3.33 (q, *J*= 5, 2H), 3.06 (s, 6H), 1.74 (q, *J*= 5.5, 2H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 168.2, 162.1, 156.2, 153.3, 150.3, 125.9, 110.4, 109.4, 108.6, 98.3, 70.2, 69.9, 50.9, 40.9, 40.4, 38.4, 29.1. HRMS (EI) calcd for $C_{19}H_{22}O_4N_3$ $[M]^+$: *m/z* 373.1745, found 373.1749.



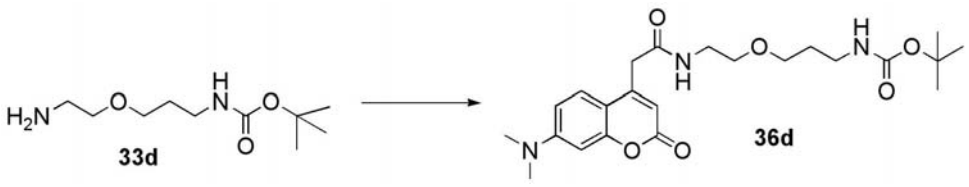
To a solution of the azide **35b** (90 mg, 0.20 mmol) in THF (5 mL), PPh_3 (105 mg, 0.402 mmol) and water (500 μL) were added and stirred at room temperature overnight. Evaporation of the solvent under reduced pressure gave a crude oil from which amine **38b** was isolated by flash chromatography (5% MeOH/CH_2Cl_2 to 15% MeOH: CH_2Cl_2 : 1.5% NEt_3) as a light oil (44 mg, 63%). ^{1}H-NMR (400 MHz, CD_3OD) δ 7.55 (d, *J*= 9.2 Hz, 1H), 6.74 (dd, *J*= 2.8, 9.2 Hz, 1H), 6.53 (d, *J*= 2.8 Hz, 1H), 6.03 (s, 1H), 3.64 (s, 2H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 168.5, 162.4, 156.1, 153.2, 150.8, 126.1, 110.1, 109.4, 108.7, 98.2, 72.2, 69.3, 50.6, 40.63, 40.3, 37.9, 30.0, 29.0. HRMS (EI) calcd for $C_{18}H_{25}O_4N_3$ $[M]^+$: *m/z* 347.1840, found 347.1839.



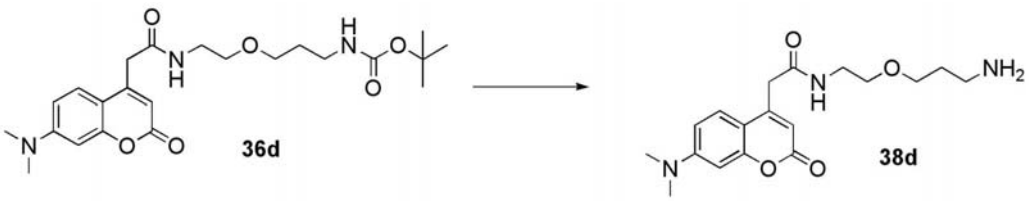
Compound **38b** (44 mg, 0.127 mmol), (D)-(-)-pantolactone (32.9 mg, 0.253 mmol), and triethylamine (70 µL, 0.47 mmol) were dissolved in MeOH (5 mL), heated to a gentle reflux, and monitored by TLC. After 48 hrs the solvent was removed under reduced pressure to yield a crude brown oil, which was purified by flash chromatography (EtOAc to 20% MeOH: EtOAc) and preparative thin-layer chromatography (15% MeOH:EtOAc) sequentially to yield **15** (16 mg, 26%) as a light fluorescent oil. ^{1}H-NMR (300 MHz, CDCl3) δ 7.50 (d, *J*= 9 Hz, 1H), 7.30 (bs, 1H), 6.62 (dd, *J*= 2.7, 6.3 Hz, 1H), 6.54 (bs, 1H) 6.46 (d, *J*= 1.8 Hz, 1H), 6.03 (s, 1H), 4.59 (bs, 1H), 4.05 (s, 1H), 3.62 (s, 1H), 3.41 (m, 10H), 3.05 (s, 6H), δ 1.84 (s, 1H), 1.68 (p, *J*=5.7 Hz, 2H), 1.03 (s, 3H), 0.89 (s, 1H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 173.9, 168.7, 162.6, 156.2, 153.4, 150.7, 126.1, 110.0, 109.6, 108.6, 98.3, 77.8, 71.3, 69.8, 68.6, 40.7, 40.3, 39.5, 38.9, 37.7, 29.3, 21.9, 20.4. HRMS (EI) calcd for $C_2H_{35}N_3O_7$ $[M]^+$: *m/z* 477.2470, found 477.2471.

(n) Synthesis of DMACA-β-Ala Probe–Pantoic Acid Conjugate II (16)

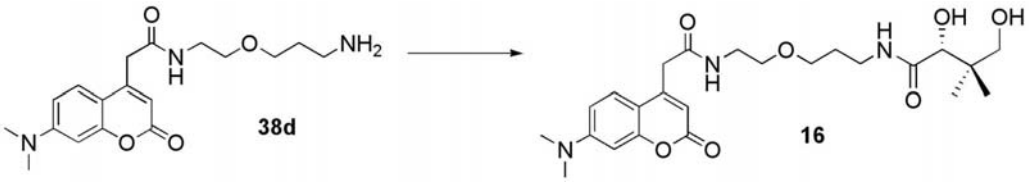
To a solution of azide **48** (111 mg, 0.45 mmol) in THF (5 mL), PPh_3 (131 mg, 0.50 mmol) and water (500 µL) were added and stirred at room temperature overnight. Evaporation of the solvent under reduced pressure gave a crude oil from which amine **33d** was isolated by flash chromatography (5% MeOH/CH_2Cl_2 to 7.5% MeOH: CH_2Cl_2: 1.5% NEt_3) as a light brown oil (54 mg, 54%). ^{1}H-NMR (300 MHz. CD3OD) 3.49 (t, *J*= 5.7 Hz, 2H), 3.48 (t, *J*= 5.4 Hz), 3.13 (t, *J*= 6.9 Hz, 2H), 2.81(t, *J*= 4.8 Hz, 2H), 1.72 (p, *J*= 6.3 Hz, 2H), 1.43 (s, 9H). ^{13}C-NMR (100 MHz, CD_3OD) δ 160.2, 80.1, 68.3, 40.7, 37.3, 29.8, 27.6. HRMS (EI) calcd for $C_{10}H_{22}N_2O_3$ $[M]^+$: *m/z* 218.1625, found 218.1626.



Compound **33d** (54 mg, 0.25 mmol), 7-dimethylaminocoumarin-4-acetic acid **31** (67 mg, 0.27 mmol), and EDAC (118 mg, 0.61 mmol) were combined and dissolved in dry DMF (5 mL). DIPEA (170 µl, 0.98 mmol) was added, and the reaction was allowed to stir 48 hours. The solvent was removed under reduced pressure, and the resultant oil purified by flash chromatography (5% MeOH: $CHCl_3$) to yield the product **36d** (55 mg, 50%) as a yellow solid. ^{1}H-NMR (400 MHz, $(CD_3)_2CO$) δ 7.62 (d, *J*= 9.2 Hz, 1H), 6.72 (dd, *J*= 2.8,6.0 Hz, 1H), 6.5 (d, *J*=2.4 Hz, 1H), 6.05 (s, 1H), 3.7 (s, 2H), 3.46 (t, *J*= 5.6 Hz, 2H), 3.43 (t, *J*= 6.0 Hz, 2H), 3.16 (q, *J*= 6.4 Hz, 2H), 3.08 (s, 6H), 1.67 (p, *J*= 6.4 Hz, 2H), 1.4 (s, 9H). ^{13}C-NMR (100.6 MHz, $(CD_3)CO$) δ 168.4, 162.0, 156.6, 156.1, 153.0, 150.3, 126.1, 110.7, 109.4, 109.2, 98.7, 77.4, 69.1, 67.3, 40.4, 40.3, 39.9, 37.0, 29.9, 28.6. HRMS (EI) calcd for $C_{23}H_{33}N_3O_6$ $[M]^+$: *m/z* 447.2364, found 447.2362.



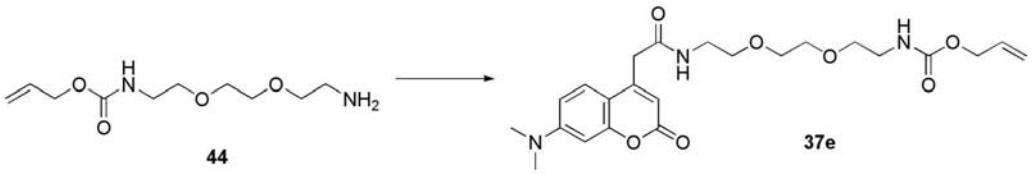
Compound **36d** (45 mg, 0.10 mmol) was dissolved in dioxane (4 ml) and cooled to 0°C. 4M HCl in dioxane (0.3 mL, 61.4 mmol) was added and the resulting reaction mixture was allowed to warm to room temperature and stirred overnight. The solvent was removed under reduced pressure, yielding **38d** (36 mg, 94%) as a dark yellow oil. ^{1}H-NMR (300 MHz, CD_3OD) δ 7.68 (d, *J*= 9 Hz, 1H), 6.96 (dd, *J*= 2.4, 9 Hz, 1H), 6.81 (d, *J* = 2.7 Hz, 1H), 6.17 (s, 1H), 3.76 (s, 2H), 3.63-3.43 (m, 6H), 3.14 (s, 6H), 3.04 (q, *J*= 6.9 Hz, 2H), 1.91 (p, *J*= 6.0 Hz, 2H). ^{13}C-NMR (75.5 MHz, CD_3OD) δ 169.7, 163.2, 155.8, 153.6, 151.8, 126.0, 109.4, 109.1, 108.7, 97.4, 69.3, 68.5,41.1, 40.7, 40.1, 39.6, 29.8. HRMS (EI) calcd for $C_{18}H_{25}N_3O_4$ $[M]^+$: *m/z* 347.1845, found 347.1847.



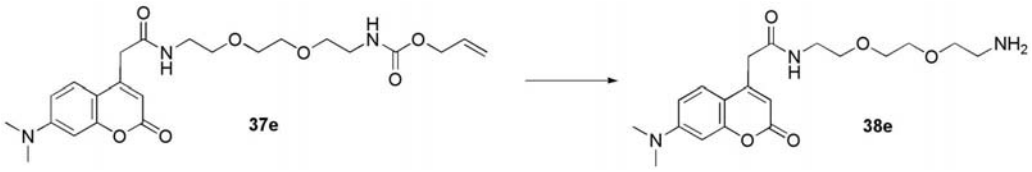
Compound **38d** (36 mg, 0.09 mmol), (D)-(-)-pantolactone (18 mg, 0.14 mmol), and triethylamine (70 µL, 0.47 mmol) were dissolved in MeOH (5 mL), heated to a gentle reflux, and monitored by TLC. After 72 hrs the solvent was removed under reduced pressure to yield a crude brown oil, which was purified by flash chromatography (EtOAc to 20% MeOH: EtOAc) and preparative thin-layer chromatography (15% MeOH:EtOAc) sequentially to yield **16** (12 mg, 27%) as a light fluorescent oil. ^{1}H-NMR (400 MHz), $CDCl_3$) δ 7.56 (d, *J*= 8.8 Hz, 1H), 7.33 (bs, 1H), 7.17 (bs, 1H), 6.62 (dd, *J*= 2.4, 8.8 Hz, 1H), 6.46 (d, *J*= 2.4 Hz, 1H), 6.08 (s, 1H), 4.05 (s, 2H), 3.44 (m, 10H), 3.05 (s, 6H), 1.71 (p, *J*= 8 Hz), 1.01(s,

3H), 0.92 (s, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 173.5, 168.7, 162.9, 156.1, 153.4, 151.1, 126.2, 109.9, 109.6, 108.8, 98.3, 78.2, 71.6, 69.6, 69.3, 40.5, 40.3, 40.1, 39.9, 39.3, 37, 29.6, 21.4, 20.7. HRMS (FAB) calcd for $C_{24}H_{36}N_3O_7$ [M+H]: *m/z* 478.2548, found 478.2554.

(o) PEG2 Linked-DMACA-Pantoic Acid Conjugate (17)

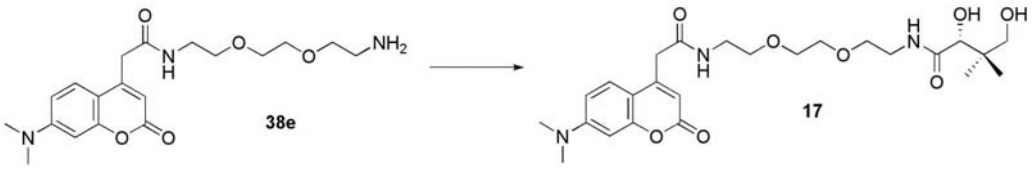


Mono-allyl-carbamate **44** (469 mg, 2.02 mmol), 7-dimethylaminocoumarin-4-acetic acid **31** (500 mg, 2.02 mmol), and EDAC (860 mg, 4.48 mmol) were combined and dissolved in dry DMF (6 mL). DIPEA (778 µL, 4.48 mmol) was added, and the reaction was allowed to stir overnight. The solvent was removed under reduced pressure, and the resultant oil was resuspended in EtOAc and washed with saturated $NaHCO_3$ (3 x 45 mL) and brine (1 x 45 mL). Flash chromatography (EtOAc to 10% MeOH:EtOAc) gave the pure product **37e** (700 mg, 75%) as a fluorescent pink oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.40 (d, *J*= 8.8 Hz, 1H), 6.80 (bs, 1H), 6.51 (d, *J*= 8.8 Hz, 1H), 6.35 (s, 1H), 5.98 (s, 1H), 5.85-5.77 (m, 1H), 5.49 (bs, 1H), 5.20 (dd, *J*= 1.2, 17.2 Hz, 1H), 5.10 (d, *J*= 10.4 Hz, 1H), 4.45 (d, *J*=5.2 Hz, 2H), 3.55 (s, 2H), 3.46 (s, 8H), 3.37 (t, *J*= 6.1 Hz, 2H), 3.27 (q, *J*= 5.2 Hz, 2H), 2.95 (s, 6H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.3, 161.9, 155.8, 152.9, 150.3, 133.0, 125.8, 117.7, 110.1, 109.2, 108.6, 98.1, 79.3, 07.7, 70.7, 70.6, 70.6, 70.6, 70.4, 70.3, 70.2, 69.8, 69.8, 69.7, 65.7, 41.6, 41.0, 40.4, 40.4, 39.9. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{20}H_{31}N_3O7$, 461.2157, found 461.2160.



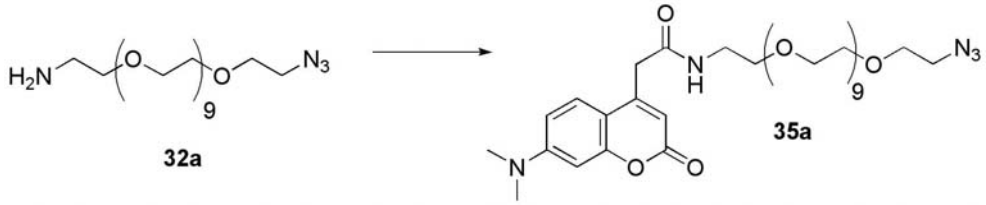
To a solution of allyl-carbamate **37e** (440 mg, 0.95 mmol) in dried THF (10 mL) was added dimedone (930 mg, 6.7 mmol), PPh_3 (50 mg, 0.19 mmol) and $Pd(Ph_3P)_4$ (54 mg, 0.048 mmol). The flask was allowed to stir at room temperature overnight. The solvent was removed under reduced pressure to give a crude oil, which was taken up in CH_2Cl_2 and purified by flash chromatography (CH_2Cl_2 to 15% MeOH: CH_2Cl_2: 1.5% NEt_3) to yield the pure product **38e** as a thin fluorescent film (215 mg, 60%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.42 (d, *J* = 8.8 Hz, 1H), 7.15 (bs, 1H), 6.52 (dd, *J*= 2, 8.8 Hz, 1H), 6.38 (d, *J*= 2.4, 1H), 5.99 (s, 1H), 3.55 (s, 2H), 3.48-3.36 (m, 12H), 2.96 (s, 6H) 2.79 (bs, 1H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.2, 161.9, 155.9, 152.9, 150.2, 125.8, 110.1, 109.2,

108.3, 98.2, 73.2, 70.5, 70.2, 69.8, 41.8, 40.5, 40.4, 39.9. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{19}H_{27}N_3O_5$, 377.1945, found 377.1948.

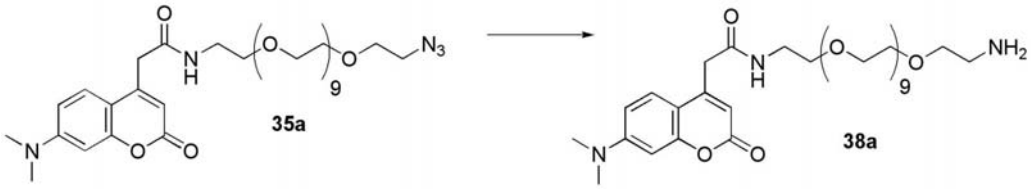


Compound **38e** (176 mg, 0.46 mmol), (D)-(-)-pantolactone (181 mg, 1.4 mmol), and triethylamine (192 µL, 1.4 mmol) were dissolved in MeOH (5 mL), heated to a gentle reflux, and monitored by TLC. After 48 hrs the solvent was removed under reduced pressure to yield a crude oil, which was purified by flash chromatography (EtOAc to 15% MeOH: EtOAc) to yield **17** (214 mg, 92%) as a pink fluorescent oil. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.53 (d, *J*= 9 Hz, 1H), 6.63 (dd, *J*= 2.5, 9 Hz, 1H), 6.51(d, *J*=3 Hz, 1H), 6.06 (s, 1H), 4.04 (s, 1H), 3.63 (s, 2H), 3.54-3.4 (m, 14H), 3.06 (s, 6H), 1.04 (s, 3H), 0.91 (s, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 171.3, 166.2, 159.9, 153.4, 150.6, 148.1, 123.3, 107.3, 106.8, 106.1, 95.6, 74.9, 68.4, 67.8, 67.8, 67.2, 37.6, 37.6, 37.3, 36.9, 36.3, 19.2, 17.8. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{25}H_{37}N_3O_8$, 507.2575, found 507.2581.

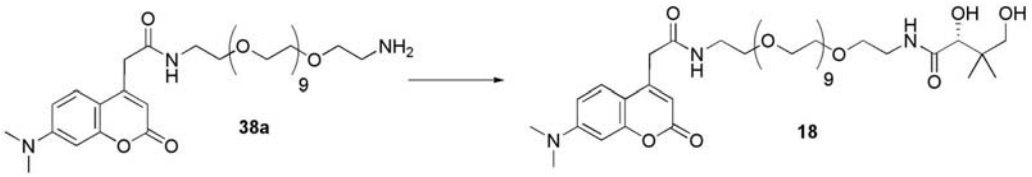
(p) PEG10 Linked-DMACA-Pantoic Acid Conjugate (18)



Mono-azido PEG **32a** (121 mg, 0.23 mmol), 7-dimethylaminocoumarin-4-acetic acid **31** (80 mg, 0.32 mmol), and EDAC (100 mg, 0.52 mmol) were combined and dissolved in dry DMF (3 mL). DIPEA (79 µL, 0.46 mmol) was added, and the reaction was allowed to stir overnight. The solvent was removed under reduced pressure, and the resultant oil was resuspended in $CHCl_3$ and purified by flash chromatography (EtOAc to 15% MeOH:EtOAc) gave the pure product **35a** (125 mg, 72%) as a fluorescent pink oil. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.49 (d, *J*= 9 Hz, 1H), 6.79 (bs, 1H), 6.59 (dd, *J*=2.6, 9 Hz, 1H), 6.46 (d, *J*= 1.5 Hz, 1H), 6.05 (s, 1H), 3.65-3.37 (m , 46H), 3.02 (s, 6H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 168.4, 161.9, 156.1, 153.0, 150.7, 126.0, 110.4, 109.3, 108.6, 98.3, 70.8, 70.8, 70.8, 70.7, 70.7, 70.6, 70.4, 70.2, 69.8, 50.8, 40.5, 40.3, 39.8. MS (ESI) $[M+H]^+$ 756.2.



To a solution of azide **35** (115 mg, 0.15 mmol) in THF (2 mL), PPh_3 (44 mg, 0.17 mmol) and water (500 μL) were added and stirred at room temperature for 24 hours. Evaporation of the solvent under reduced pressure gave a crude oil which was resuspended $CHCl_3$ and extracted into water (3x25 mL). The aqueous layer was washed with CH_2Cl_2 (3x15mL) and lyophilized to yield **38a** (81 mg, 74%). as a light fluorescent oil. ^{1}H-NMR (300 MHz, $CDCl_3$) δ 7.42 (d, *J*=9 Hz, 1H), 6.81 (bs, 1H), 6.52 (dd, *J*= 2.7, 9 Hz, 1H), 6.39 (d, *J*= 2.7 Hz, 1H), 5.98 (s, 1H), 3.6-3.32 (m, 46H), 2.96 (s, 6H), 1.66 (bs, 1H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 168.3, 161.9, 156.0, 153.1, 150.3, 125.9, 110.3, 109.2, 108.7, 98.2, 73.5, 70.6, 70.4, 70.3, 69.7, 41.9, 40.4, 40.2, 39.8. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{35}H_{59}N_3O_{13}$, 729.4042, found 729.4051.

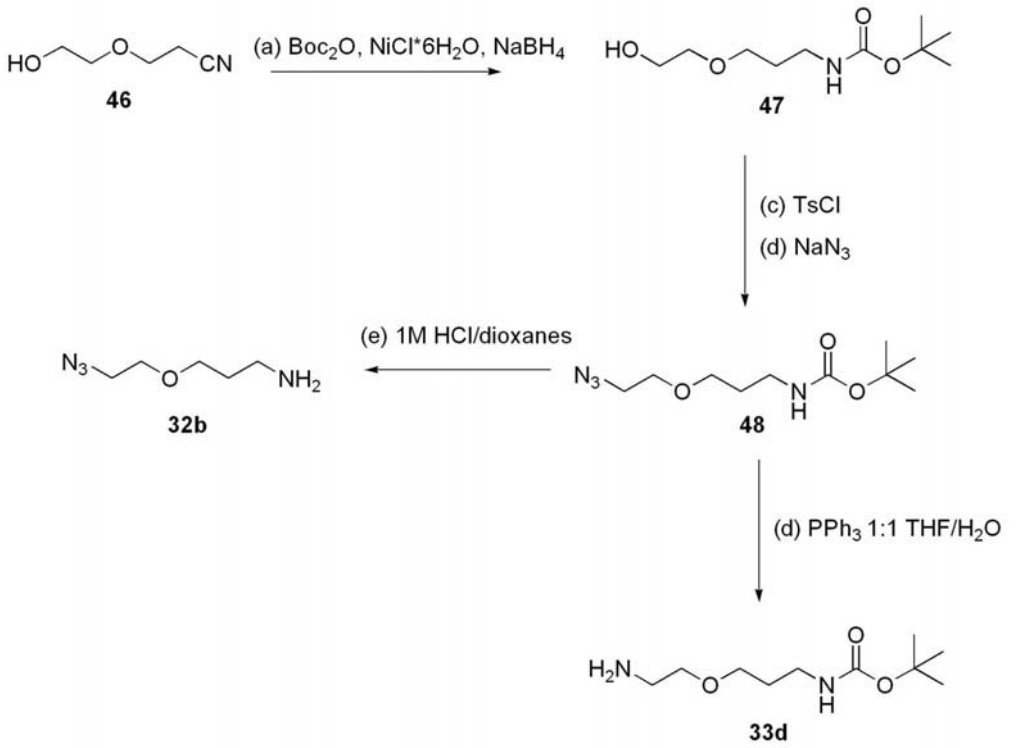


Compound **38a** (42 mg, 0.058 mmol), (D)-(-)-pantolactone (37 mg, 0.28 mmol), and triethylamine (39 μL, 0.288 mmol) were dissolved in MeOH (5 mL), heated to a gentle reflux, and monitored by TLC. After 48 hrs the solvent was removed under reduced pressure to yield a crude oil, which was purified by flash chromatography (EtOAc to 12.5% MeOH: EtOAc) to yield **18** (27 mg, 55%) as a pink fluorescent oil. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.52 (d, *J*= 9 Hz, 1H), 6.84 (bs, 1H), 6.61 (dd, *J*= 2.5, 9 Hz, 1H), 6.49 (d, *J*=3 Hz, 1H), 6.07 (s, 1H), 4.01 (s, 1H), 3.67-3.51 (m, 48H), 3.04 (s, 6H), 1.03 (s, 3H), 0.88 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 173.7, 168.5, 162.1, 156.1, 153.2, 150.4, 126.1, 110.4, 109.3, 108.7, 98.4, 77.2, 70.9, 70.7, 70.7, 70.6, 70.4, 70.3, 69.8, 40., 40.4, 39.8, 39.5, 38.9, 22.5, 20.2. HRMS (FAB) (*m/z*) $[M+H]^+$ calcd for $C_{41}H_{69}N_3O_{16}$, 860.4751, found 860.4742

Route, Synthetic Procedures, and Spectroscopic Data for Diprotected Linker 48

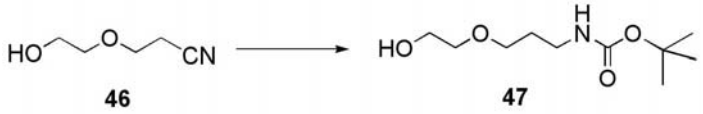
General Synthetic Procedures General materials and methods as described above (pg. S2).

Synthetic Route to Diprotected Linker 48



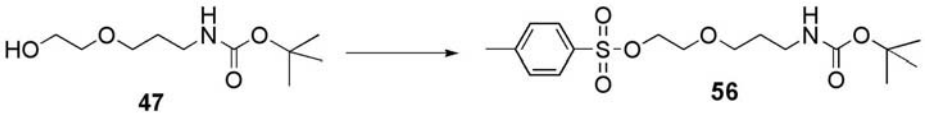
(a) Boc_2O (2eq), $NiCl*6H_2O$(0.1eq), $NaBH_4$ (7eq), MeOH, O° --> RT (b) TsCl, NEt_3, DCM (c) NaN_3, DMF (d) PPh_3 1:1 THF/H_2O (e) 1M HCl/dioxanes

Synthetic Procedures and Spectroscopic Data for Diprotected Linker 48



A solution of **46** (2 g, 17.4 mmol) in dry methanol was prepared and cooled to 0°C. To the reaction mixture di-*tert*-butyl dicarbonate (7.5 g, 34.7 mmol) and nickel (II) chloride hexahydrate (413 mg, 1.7 mmol) were added sequentially. This was followed by the addition of $NaBH_4$ (4.6 g, 121.6 mmol) in small aliquots. The reaction was then let warm to room temperature and stirred over night. Upon completion ethylene diamine (1.2 ml, 17.4 mmol) was added to the reaction mixture and let stir for an additional 30mins. The resulting mixture was

concentrated and dissolved in EtOAc (300 ml). The resulting solution was then washed with saturated $NaHCO_3$ (2x300 ml) and dried over Na_2SO_4. The solvent was removed under reduced pressure to afford a clear oil (2.7 g, 71%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 4.77 (bs, 1H), 3.72 (m, 2H), 3.53 (m, 4H), 3.3 (d, *J*= 4.4, 2H), 1.75 (p, J= 6, 2H); ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 156.4, 79.5, 72.4, 68.8, 61.9, 38.1, 28.6, 28.5; HRMS (EI) calcd for $C_{10}H_{21}NO_4$ $[M]^+$: *m/z* 219.1465, found 219.1463.



Tert-butyl carbamate protected alcohol **47** (1.60g, 7.3 mmol) was dissolved in CH_2Cl_2 (15 ml) and cooled to -10°C. After the solution had cooled triethylamine (1.22 ml, 8.8 mmol), dimethyl-aminopyridine (223 mg , 1.8 mmol), and tosyl chloride (1.53 g, 8 mmol) were added sequentially. The reaction was warmed to room temperature and allowed to stir overnight. The reaction mixture was concentrated under reduced pressure and taken up in CH_2Cl_2 (250 ml). The organic layer was washed with 0.5M HCl (250 ml), saturated aqueous $NaHCO_3$ (250 ml), H_2O (250 ml) and then dried over Na_2SO_4 and the solvent removed to yield **56** (2.29 g, 85%) as a dark yellow oil. ^{1}H-NMR (300MHz, CD_3OD) δ 7.81-7.77 (m, 1H), 7.46-7.43 (m, 1H), 4.15-4.12 (m, 2H), 3.59-3.56 (m, 2H), 3.40 (t, *J*= 6.0 Hz, 2H), 3.04 (t, *J*= 6.9 Hz, 2H), 1.63 (p, *J*= 6.6 Hz, 2H), 1.43 (s, 9H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 156.2, 145.1, 133.2 130.1 128.2 69.6, 69.4, 68.5, 38.4, 29.9, 28.7, 28.5, 21.9. HRMS (EI) calcd for $C_{17}H_{27}NO_6S$ $[M]^+$: *m/z* 373.1554, found 373.1559.

To a solution of compound **56** (1.60 g, 4.3 mmol) dissolved in DMF, NaN_3 (1.25 g, 19.3 mmol) was added and stirred at 100°C overnight. The reaction mixture was concentrated under reduced pressure, taken up in EtOAc (350 ml), and washed with H_2O (3x 250 ml), and saturated aqueous NaCl (250 ml). The organic layer was dried over Na_2SO_4 and solvent removed to afford di-protected linker **48** (996 mg, 95%) as a light brown oil. ^{1}H-NMR (500 MHz, CDCl3) δ 4.83 (bs, 1H), 3.62 (t, *J*=5 Hz, 2H), 3.55 (t, *J*= 6.5 Hz, 2H), 3.39 (t, *J*= 5 Hz, 2H) 3.25 (q, *J*= 5.5 Hz, 2H) 1.79 (p, *J*= 6.5 Hz, 2H) 1.45 (s, 9H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 156.3, 79.3, 69.9, 69.7, 50.9, 38.5, 29.8, 28.6. HRMS (EI) calcd for $C_{10}H_{20}O_3N_4$ $[M]^+$: *m/z* 244.1530, found 244.1523.

Kinetic, In Vitro, and In Vivo Assay Data

Kinetic Assays

Kinetics with CoAA for the panel of compounds were performed as previously described in the literature.[12] To test compound **16** as an inhibitor a competition assay was designed. The assay was set up as in the other kinetic runs with pantothenate as the substrate (500μM-3μM). A dilution series of compound **16** was added, keeping the total volume and the concentration of all other reagents the same. At 125μM of compound **16**, the K_{cat} for CoAA with pantothenate was reduced between 30 and 40 percent, without any change in K_m.

In Vitro Assays

The gel shift assay was performed by incubating compounds with the CoA biosynthetic enzymes, the carrier protein ACP (*E. coli*), and the PPTase Sfp, as previously described.[13] Samples were then run on a native 15% SDS PAGE gel.

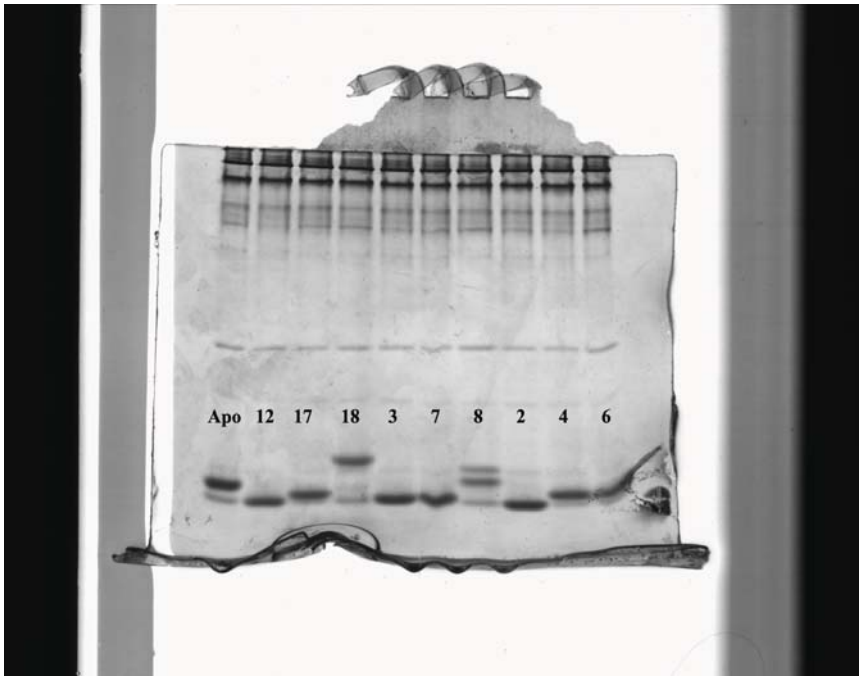


Figure S1. Gel Shift Assay. Apo lane corresponds to the *apo* form of the fatty acid synthase acyl carrier protein from *E. coli*, ACP. Gel-shift indicates formation of *crypto*-ACP by reaction of ACP, Sfp, and the appropriate pantetheine analog.

In vitro assays were done as previously described,[6] with 0.4mM analogue per reaction. Reactions with fluorescent compounds were run on 12% PAGE and visualized with UV. Bioorthogonally tagged pantetheine analogs **1**-**8** were incubated with the enzyme mixture and then the appropriate reactive biotin reporter (**9**, **10**, **11**) was added. In the case of azide and alkyne derivatives (**1-3**, **5**,

7-8) the procedure of Alexander was followed.[13] For keto-derivatives **4** and **6** reactions were incubated with biotin hydroxylamine (**9**, Molecular Probes) as follows. 5µL of 100mM stock (DMSO) was added and the reaction was incubated at room temperature overnight. Samples were then run on 12% PAGE gels, blotted onto nitrocellulose and developed with streptavidin-alkaline phosphotase using standard procedures.

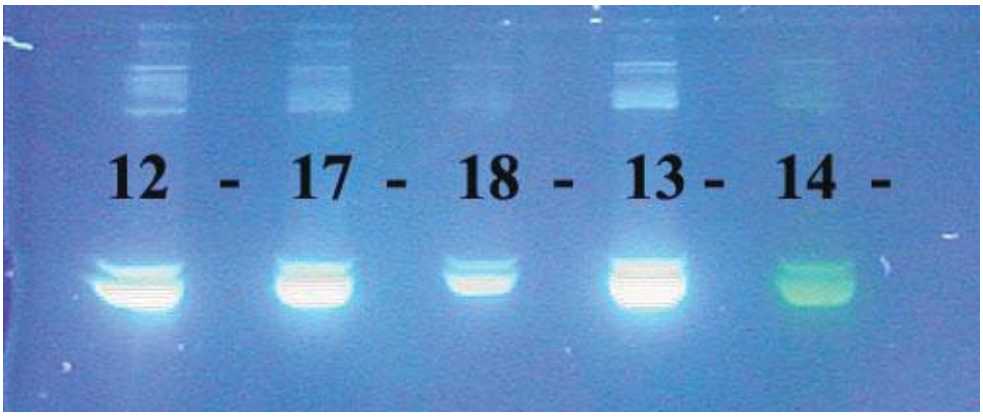


Figure S2. In vitro reactions of fluorescent pantetheine derivatives **12-14**, **17-18** with negative controls. Bands are fluorescently tagged VibB, the vibriobactin synthase carrier protein from *Vibrio cholerae*. Negative controls lack Sfp.

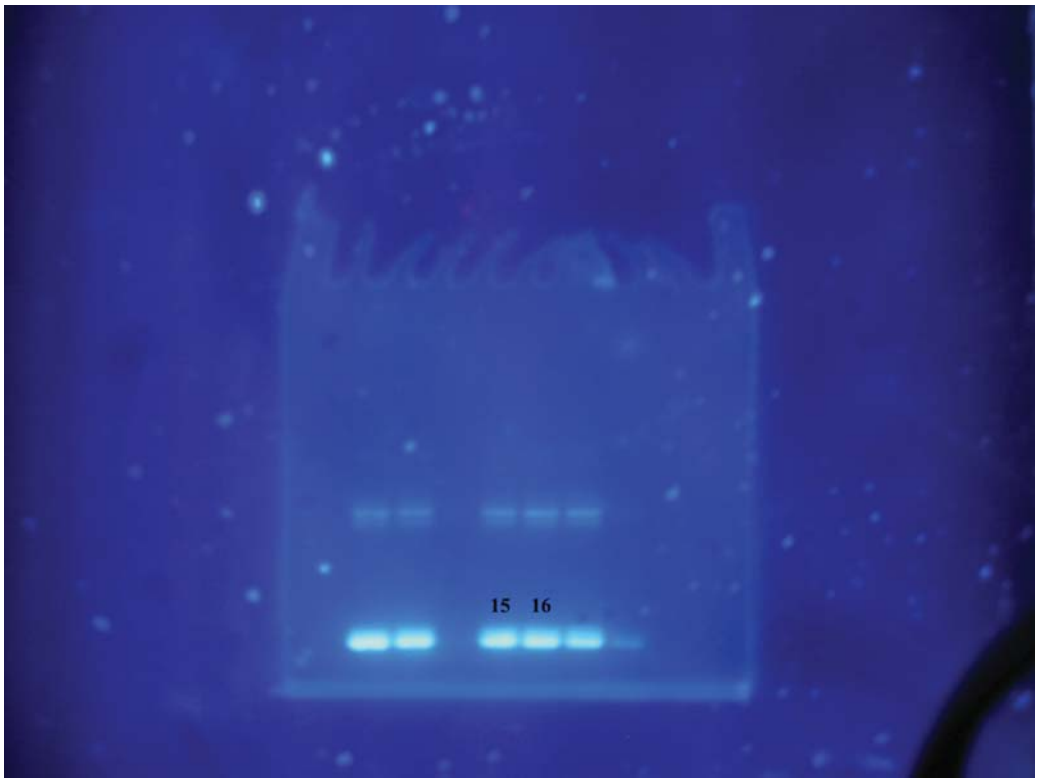


Figure S3. In vitro reactions of fluorescent pantetheine derivatives **15** and **16.**

Figure S4. Western blot showing in vitro reaction of bioorthogonally tagged pantetheine analogs **2-4, 6-8,** with negative controls. Bands are biotin-ligated VibB. Negative controls lack Sfp.

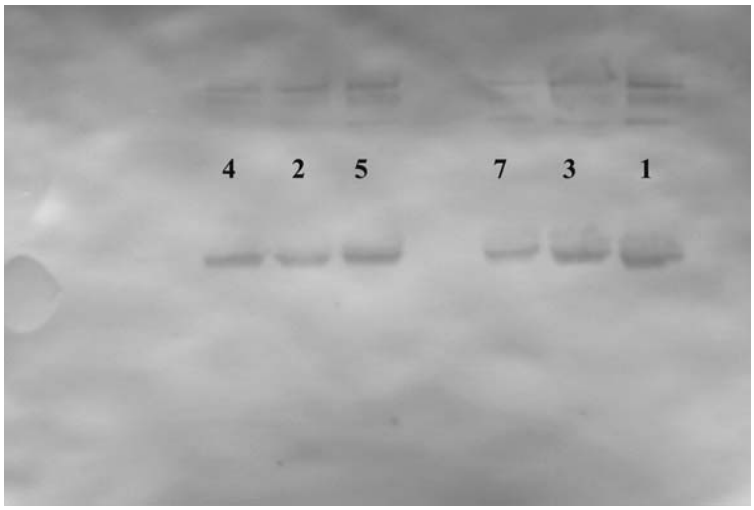


Figure S5. Western blot showing in vitro reaction of bioorthogonally tagged pantetheine analogs **1-5, 7,** with negative controls. Bands are biotin-ligated VibB after reaction with strepdavidin-linked alkaline phosphatase. Negative controls lack Sfp.

In Vivo Assays

In vivo assays were performed as previously described, but samples were grown in a volume of 1mL LB.[6] Samples were visualized as in the in vitro assays. In the case of azide and alkyne derivatives (**1-3**, **5**, **7-8**) 43µL of lysate was used in the click reaction following the procedure of Alexander.[14] For keto-derivatives **4** and **6** 35µL of lysate was added to 25µL of 0.5 Tris buffer pH 6.0 and 5µL of biotin hydroxylamine (**9**, Molecular Probes). The reaction proceeded overnight at room temperature.

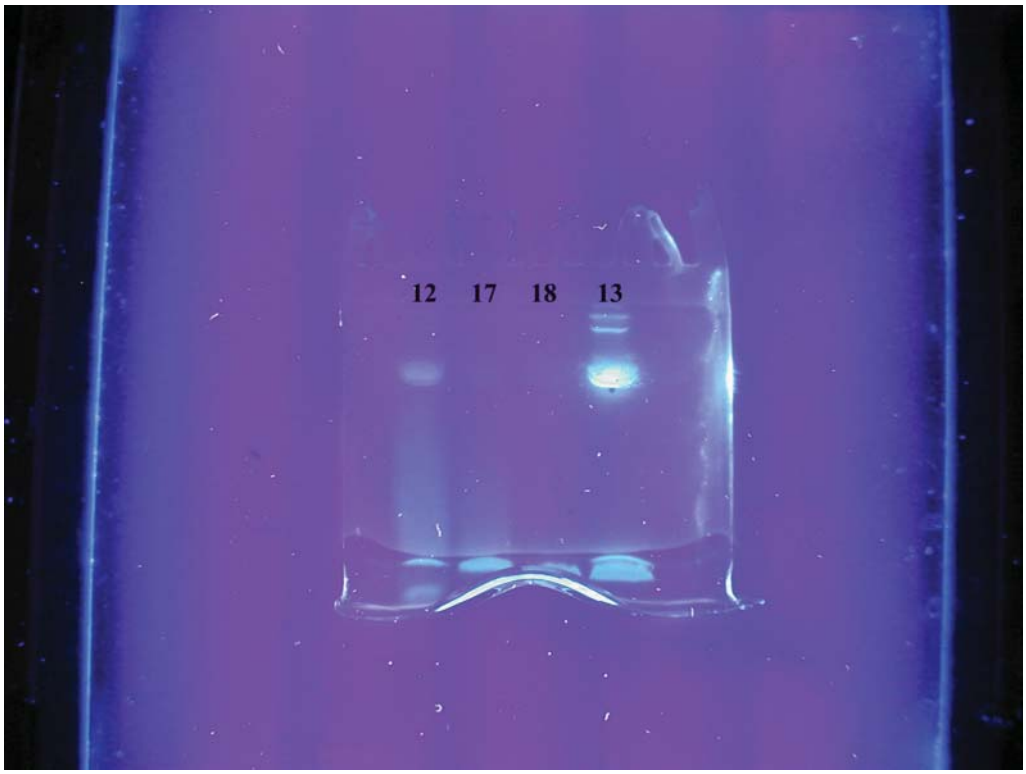


Figure S6. In vivo assay of fluorescent pantetheine derivatives **12-13, 17-18**. Bands are fluorescently tagged VibB.

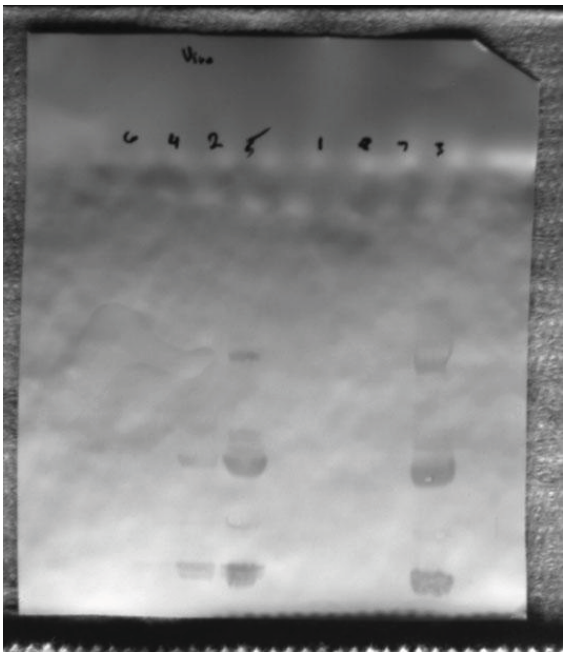


Figure S7. Western blot showing in vivo assay of bioorthogonally tagged pantetheine analogs **1-8**. Bands are biotin-ligated VibB after reaction with strepdavidin-linked alkaline phosphatase.

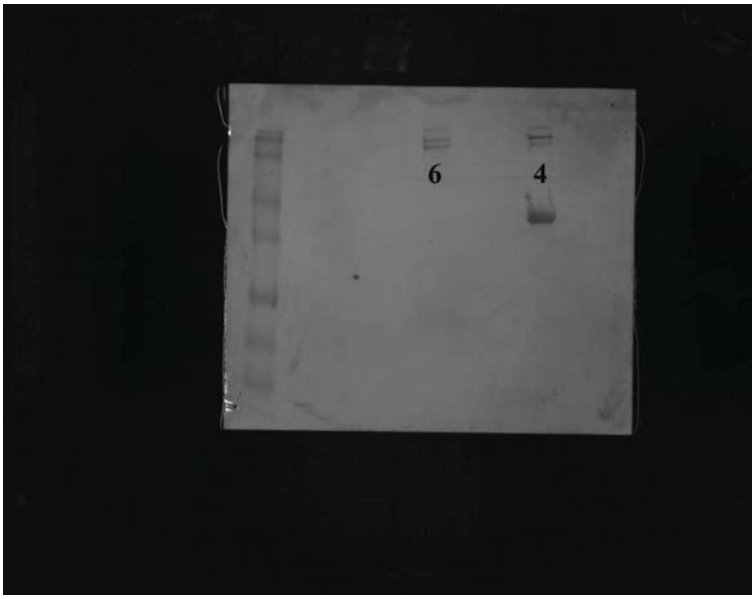


Figure S8. Western blot showing in vivo assay of bioorthogonally tagged pantetheine analogs **6** and **4**. Bands are biotin-ligated VibB after reaction with strepdavidin-linked alkaline phosphatase.

Full Author Listing for References 18 and 20

18. Daniela C. Oniciu, D.C.; Bell R.P.L.; McCosar, B.H.; Bisgaier, C.L.; Dasseux, J.H.; Verdijk, D.; Relou, M.; Smith, D.; Regeling, H.; Leemhuis, F.M.C.; Ebbers, E.J.; Mueller, R.; Zhang, L.; Pop, E.; Cramer, C.T.; Goetz, B.; McKee, A.; Pape, M.E.; Krause, B.R. *Syn. Comm.* **2006**, *36*, 365-391.

20. Alexander, M. D.; Burkart, M. D; Leonard, M. S.; Portonovo, P.; Liang, B; Ding, X.; Joullie, M. M.; Gulledge, B. M.; Aggen, J. B.; Chamberlin, A. R.; Sandler, J.; Fenical, W.; Cui, J.; Gharpure, S. J.; Polosukhin, A.; Zhang, H.-R.; Evans, P. A.; Richardson, A. D.; Harper, M. K.; Ireland, C. M.; Vong, B. H.; Brady, T. P; Theodorakis, E. A.; La Clair, J. J. *ChemBioChem* **2006**, 7, 409-416.

References:

1. Kumar, V.; Aldrich, J. V., A Solid-Phase Synthetic Strategy for Labeled Peptides: Synthesis of a Biotinylated Derivative of the δ Opioid Receptor Antagonist TIPP (Tyr-Tic-Phe-Phe-OH). *Org. Lett.* **2003,** 5, 613-616.

2. Lee, J.H.; Baker, T.J.; Mahal, L.K.; Zabner, J.; Bertozzi, C.R.; Wiemer, D.F.; Welsh, M.J., Engineering Novel Cell Surface Receptors for Virus-mediated Gene Transfer. *J. Biol. Chem.* **1999**, 274, 21878-21884.

3. Grandjean, C.; Boutonnier, A.; Guerreiro, C.; Fournier, JM.; Mulard, L.A., On the Preparation of Carbohydrate-Protein Conjugates Using the Traceless Staudinger Ligation. *J. Org. Chem.* **2005,** 70, 7123–7132.

4. Newman, S.H.; Forster, M.O., The Triazo-group. Part XVIII. β-Triazoethylamine. *J. Chem. Soc*., **1911**, 99, 1277-1282.

5. Alexander, M. D.; Burkart, M. D; Leonard, M. S.; Portonovo, P.; Liang, B; Ding, X.; Joullie, M. M.; Gulledge, B. M.; Aggen, J. B.; Chamberlin, A. R.; Sandler, J.; Fenical, W.; Cui, J.; Gharpure, S. J.; Polosukhin, A.; Zhang, H.-R.; Evans, P. A.; Richardson, A. D.; Harper, M. K.; Ireland, C. M.; Vong, B. H.; Brady, T. P; Theodorakis, E. A.; La Clair, J. J., A Central Strategy for Converting Natural Products into Fluorescent Probes. *ChemBioChem* **2006**, 7, 409-416.

6. Clarke, K.C.; Mercer, A.C.; LaClair, J.J.; Burkart, M.D., In Vivo Reporter Labeling of Proteins via Metabolic Delivery of Coenzyme A Analogues. *J. Am. Chem. Soc*. **2005**, *127*, 11234 – 11235.

7. Mathisen, T.; Albertson, A.C. Polymerization of 1,5-dioxepan-2-one. I. Synthesis and characterization of the monomer 1,5-dioxepan-2-one and its cyclic dimer 1,5,8,12-tetraoxacyclotetradecane-2,9-dione. *Macromolecules* **1989**, *22*, 3838-3842.

8. Misra, A.; Mishra, S.; Mishra, K. Synthesis and Fluorescence Studies of Multiple Labeled Oligonucleotides Containing Dansyl Fluorophore Covalently Attached at 2' Terminus of a Cytidine via a Carbamate Linkage. *Bioconj. Chem.* **2004**, *15*, 638-646.

9. Gottlieb, H.E.; Kotlyar V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents As Trace Impurities. *J. Org. Chem.* **1997**, *62*, 7512-7515.

10. Still, W.C.; Kahn, A.; Mitra, A. Rapid Chromatographic Technique For Preparative Separations With Moderate Resolution. *J. Org. Chem.* **1978**, *43*, 2923-2925.

11. Dale, J.A.; Mosher, H.S**.** Nuclear magnetic resonance nonequivalence of diastereomeric esters of alpha.-substituted phenylacetic acids for the determination of stereochemical purity. *J. Am. Chem. Soc.* **1968**, *90*, 3732.

12. Worthington, A.S.; Burkart, M.D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem.* **2005**, *4*, 44-46.

13. Virga K.G.; Zhang Y.M.; Leonardi R., Ivey RA, Hevener K, Park HW, Jackowski S, Rock CO, Lee RE. Structure–activity relationships and enzyme inhibition of pantothenamide-type pantothenate kinase inhibitors. *Bioorg. Med. Chem*. **2006**, *14*, 1007-1020.

14. Alexander, J.P.; Cravatt, B.F. Mechanism of carbamate inactivation of FAAH: implications for the design of covalent inhibitors and in vivo functional probes for enzymes. *Chem Biol.* **2005**, *12*, 1179-1187.

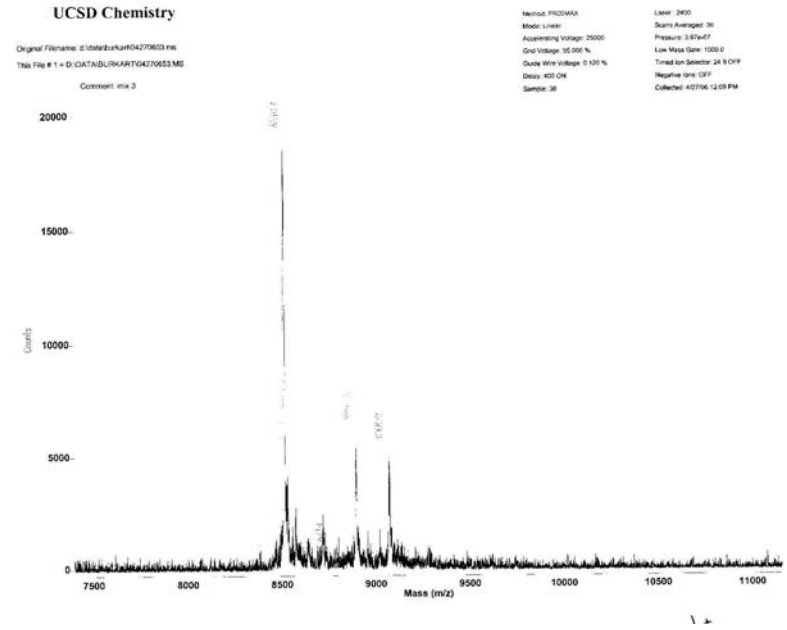
Synthesis and Evaluation of Bioorthogonal Pantetheine Analogs for In Vivo Protein Modification

*Jordan L. Meier, Andrew C. Mercer, Heriberto Rivera Jr., and Michael D. Burkart**
Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358
mburkart@ucsd.edu

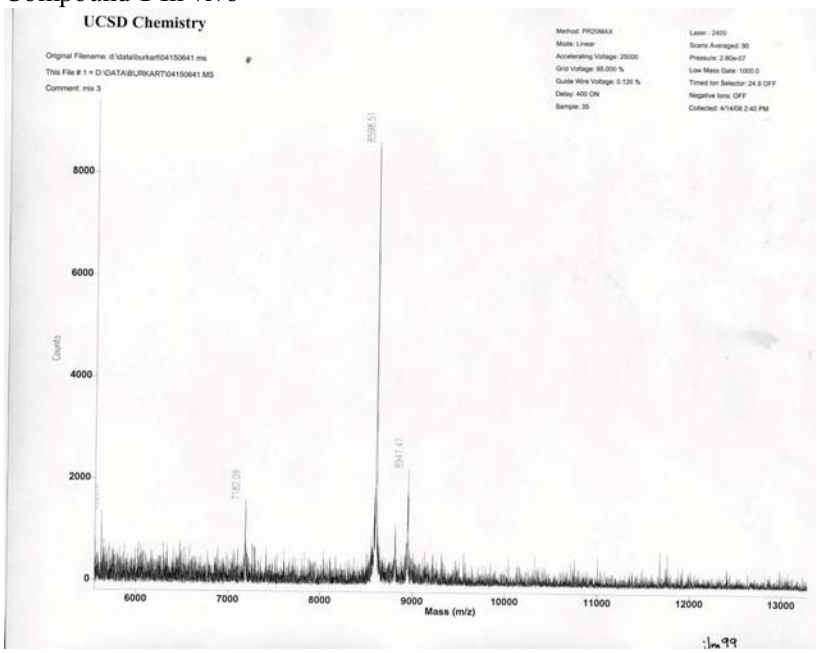
Supporting Information II

Contents

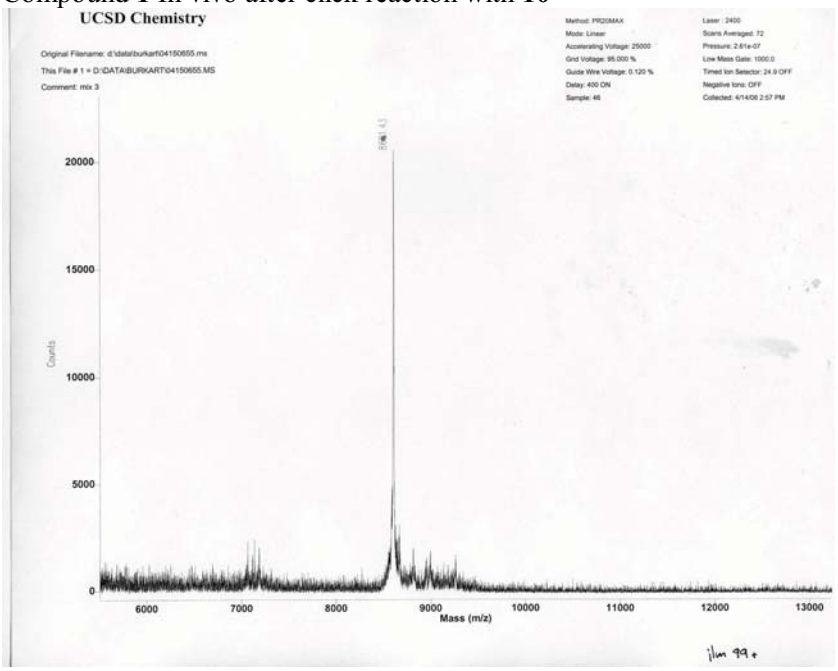
Compound **1** In vitro after click reaction with **10**



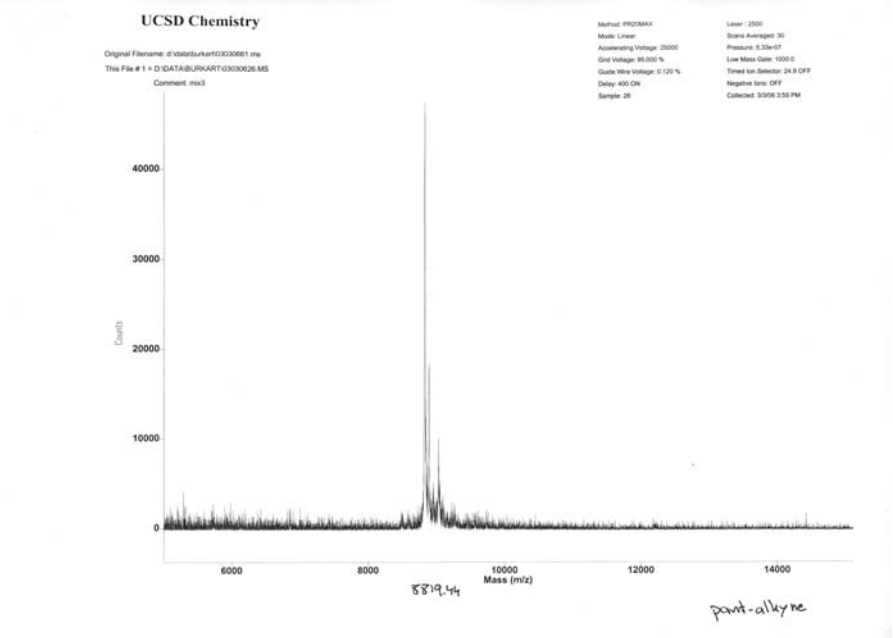
Compound **1** In vivo



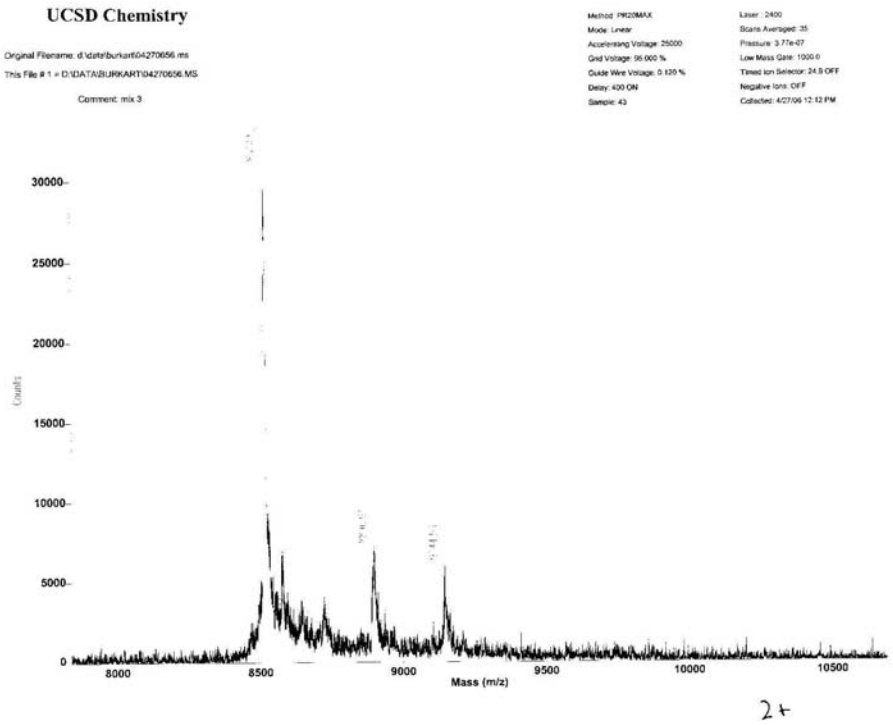
Compound **1** In vivo after click reaction with **10**



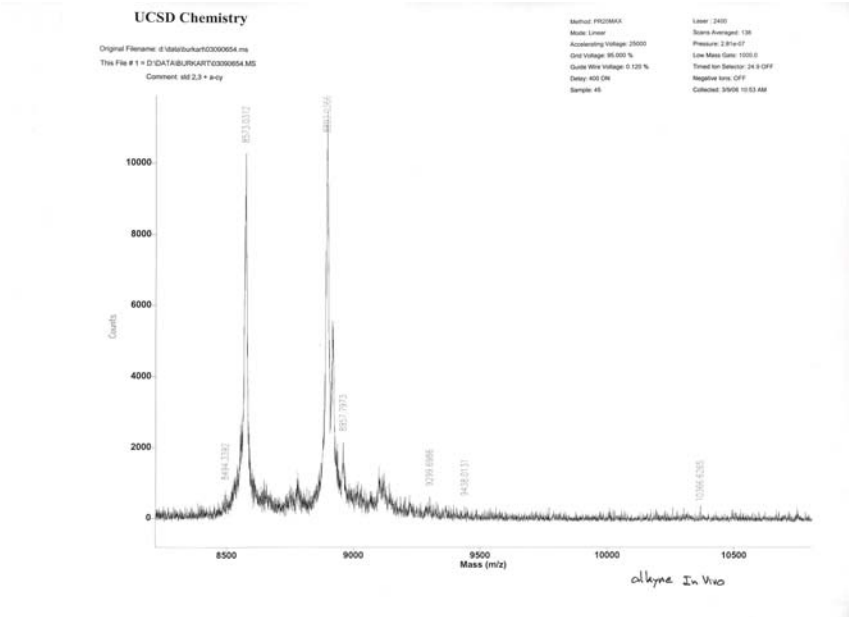
Compound **2** In vitro



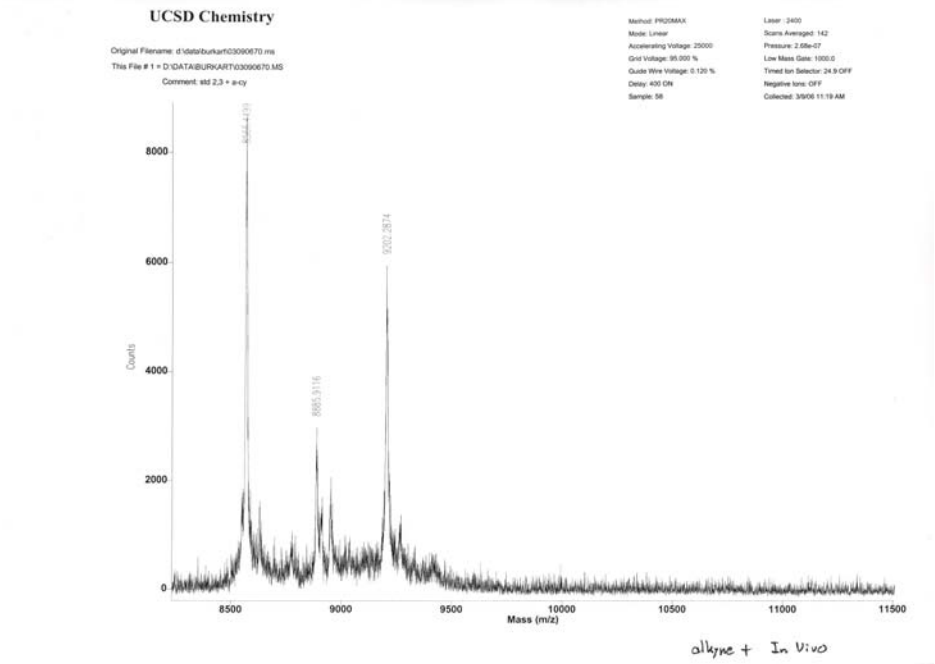
Compound **2** In vitro after click reaction with **11**



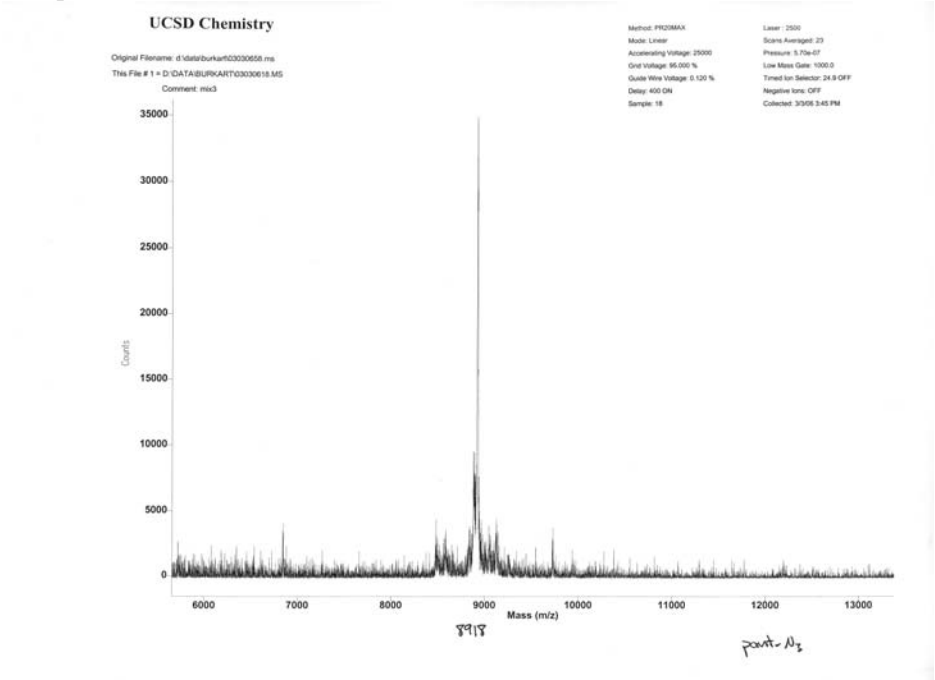
Compound **2** In vivo



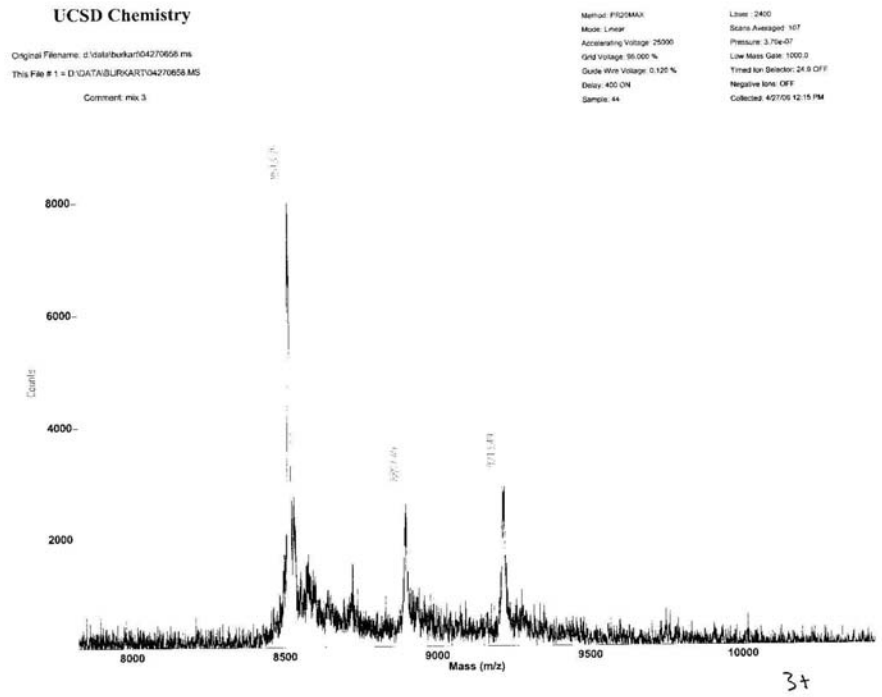
Compound **2** In vivo after click reaction with **11**



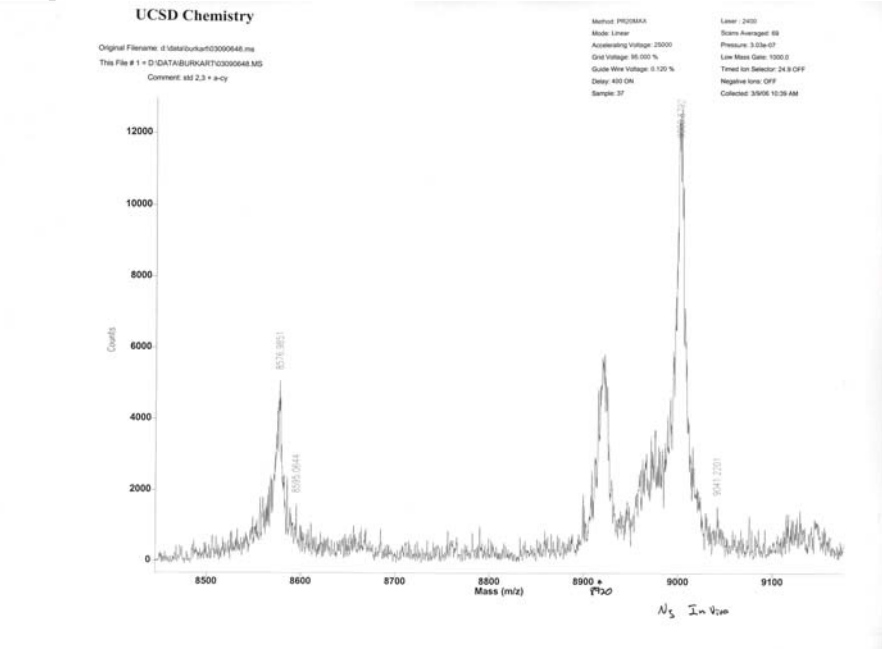
Compound **3** In vitro



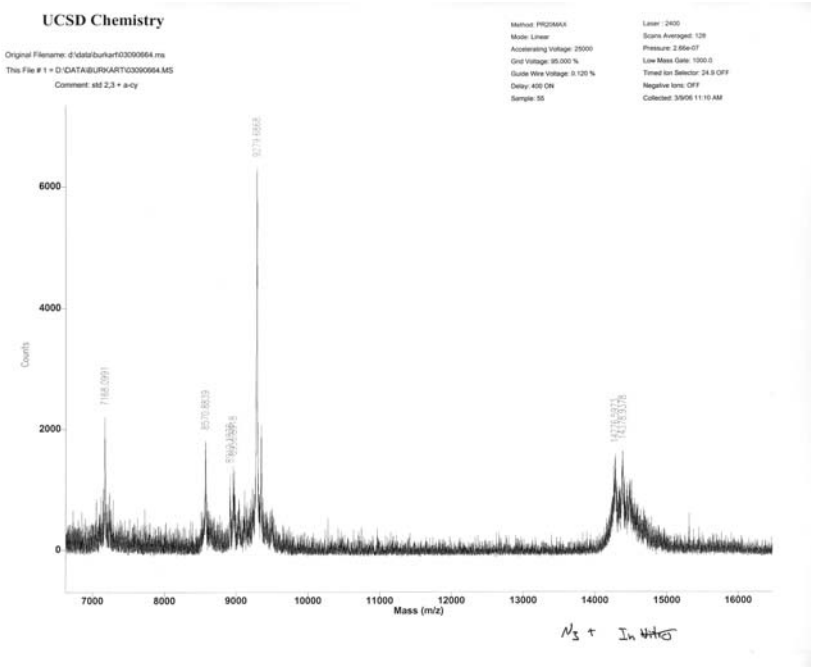
Compound **3** In vitro after click reaction with **10**



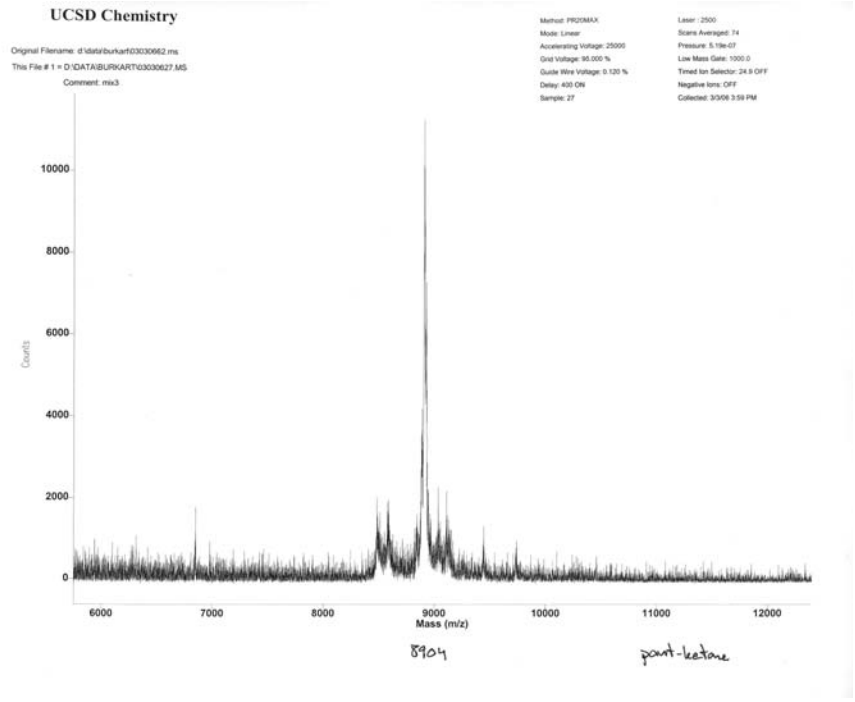
Compound **3** In vivo



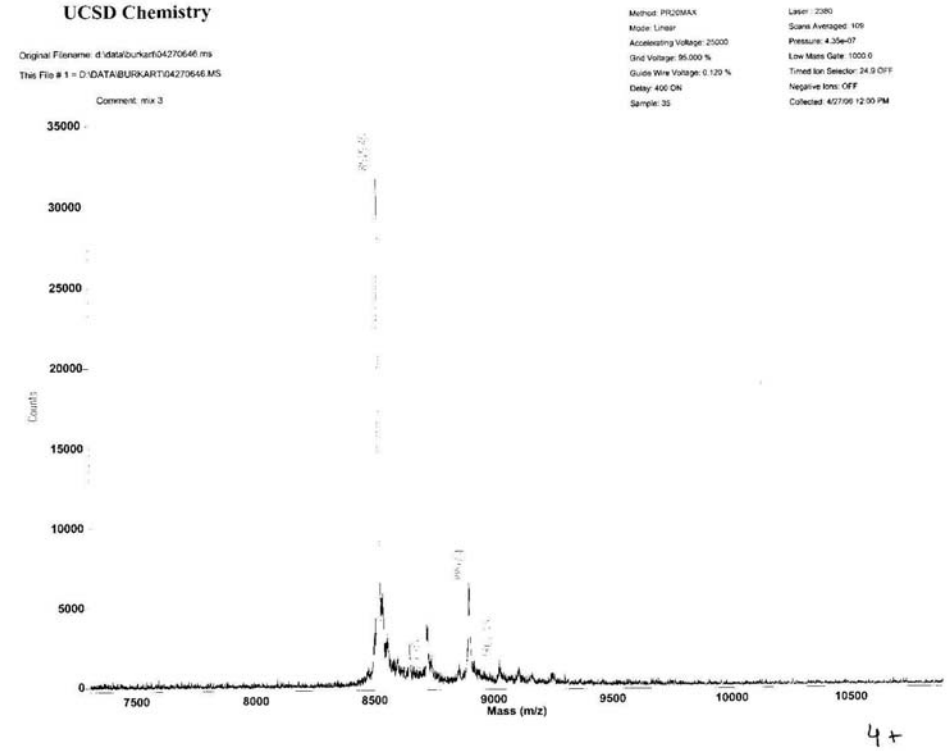
Compound **3** In vivo after click reaction with **10**



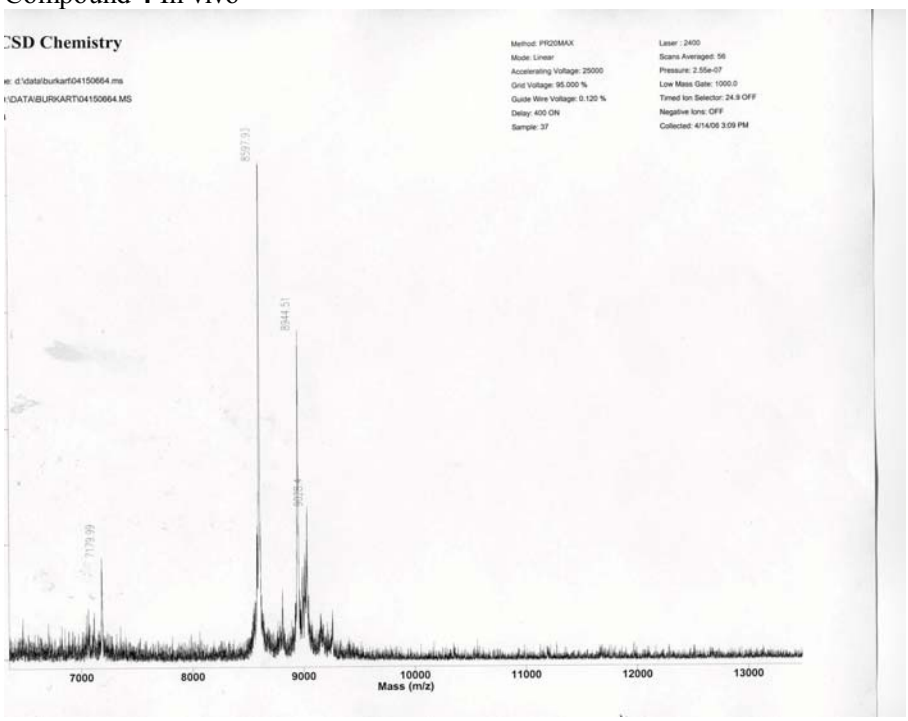
Compound **4** In vitro



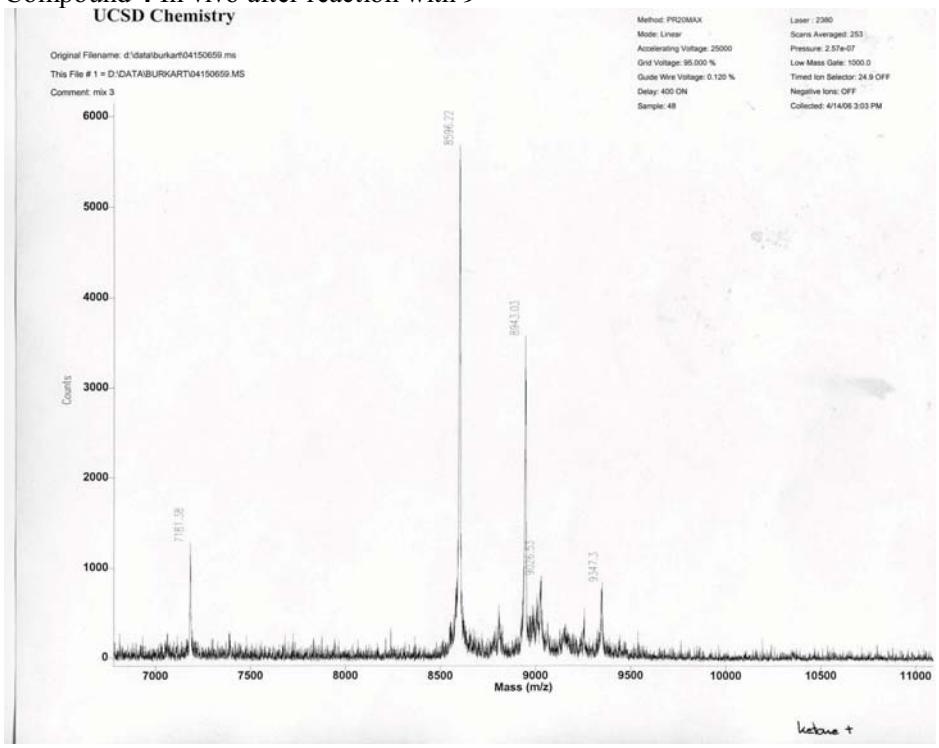
Compound **4** In vitro after reaction with **9**



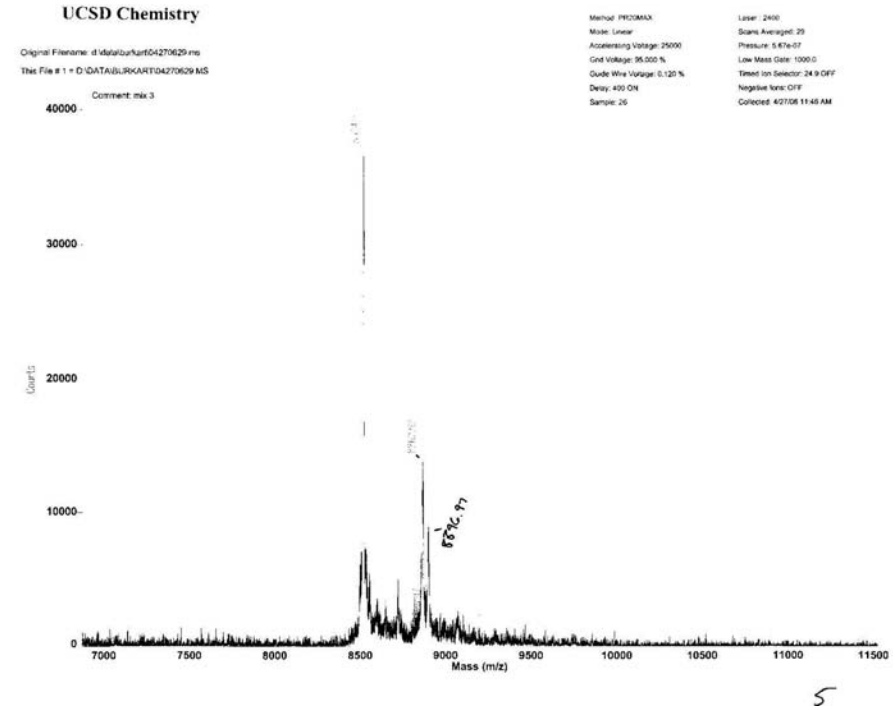
Compound **4** In vivo



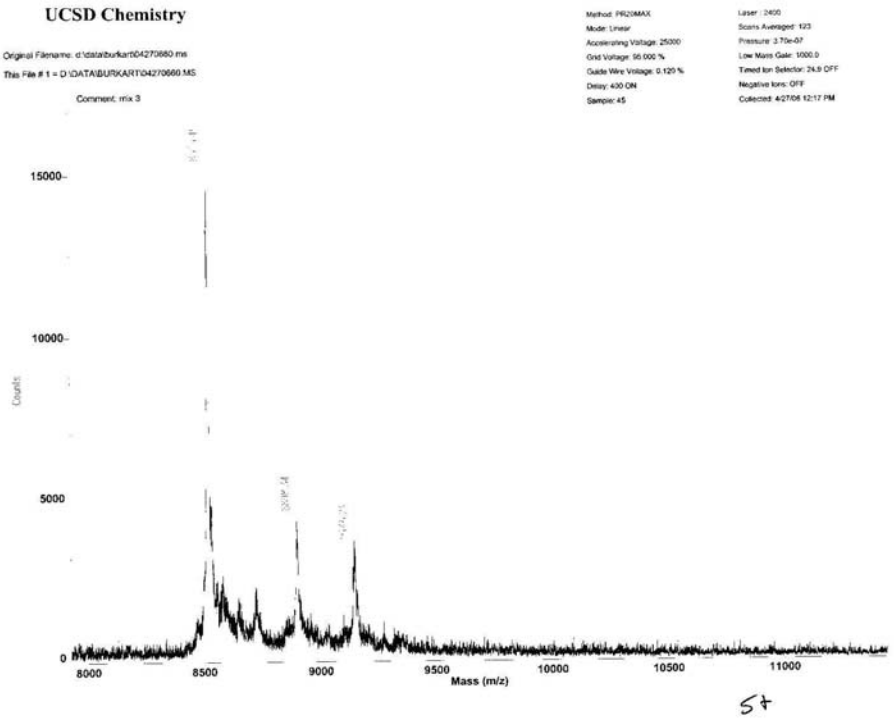
Compound **4** In vivo after reaction with **9**



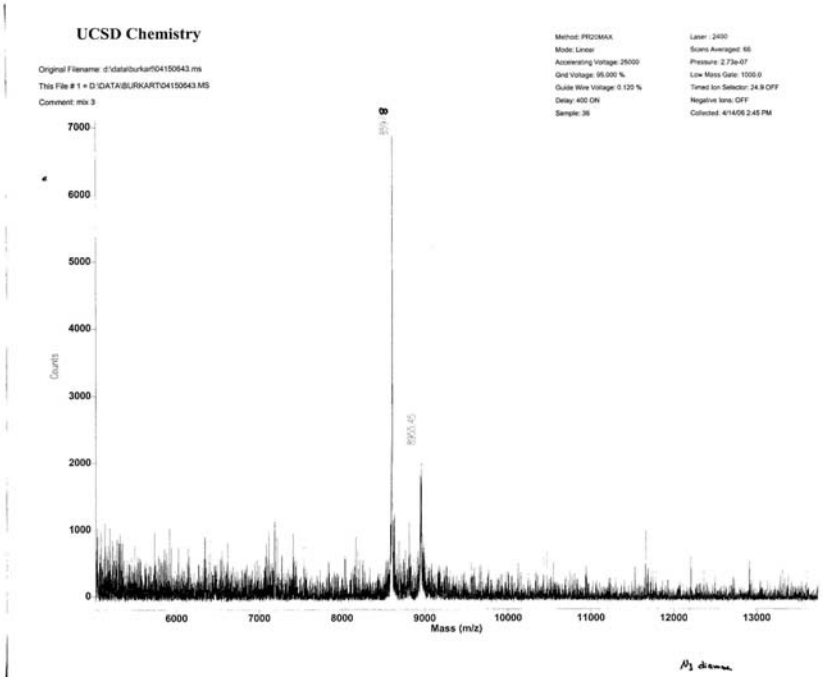
Compound **5** In vitro



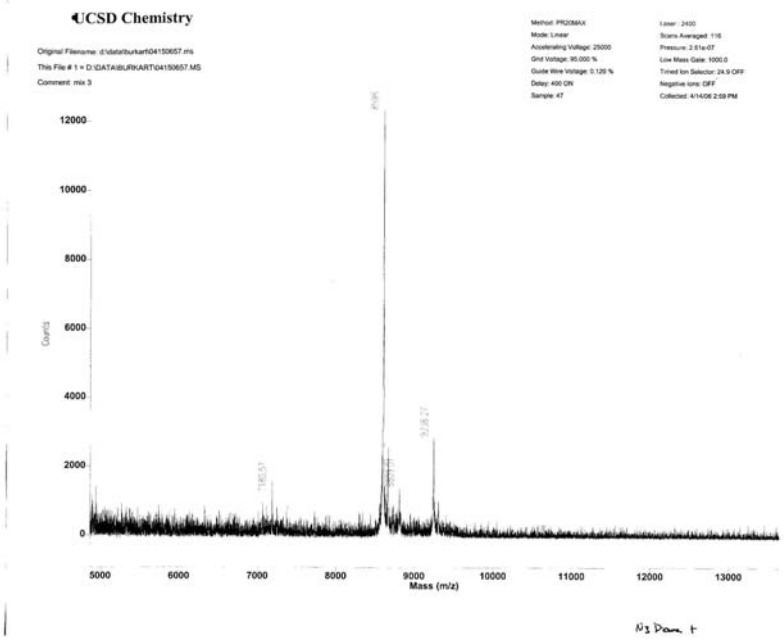
Compound **5** In vitro after click reaction with **10**



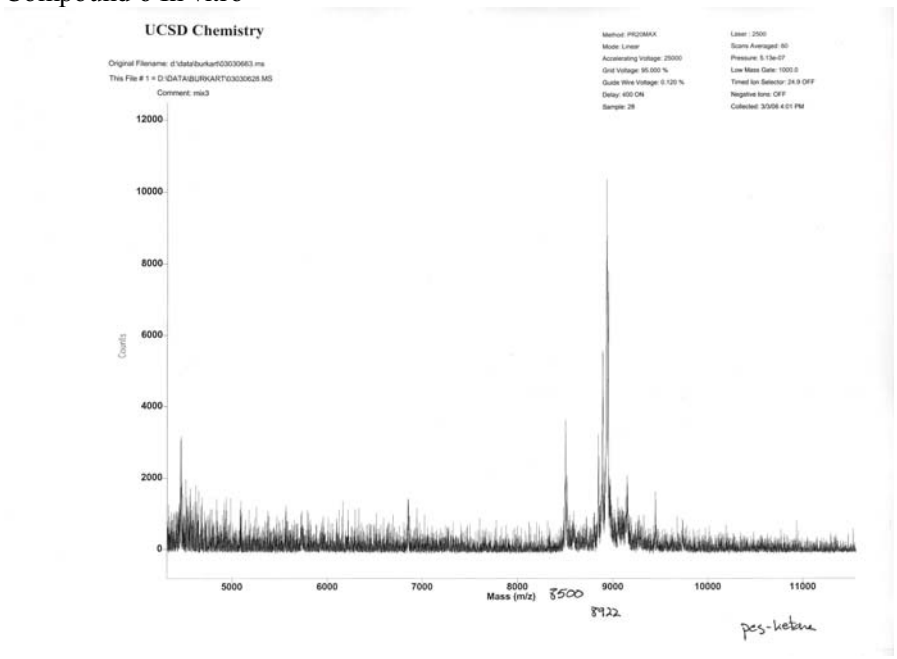
Compound **5** In vivo



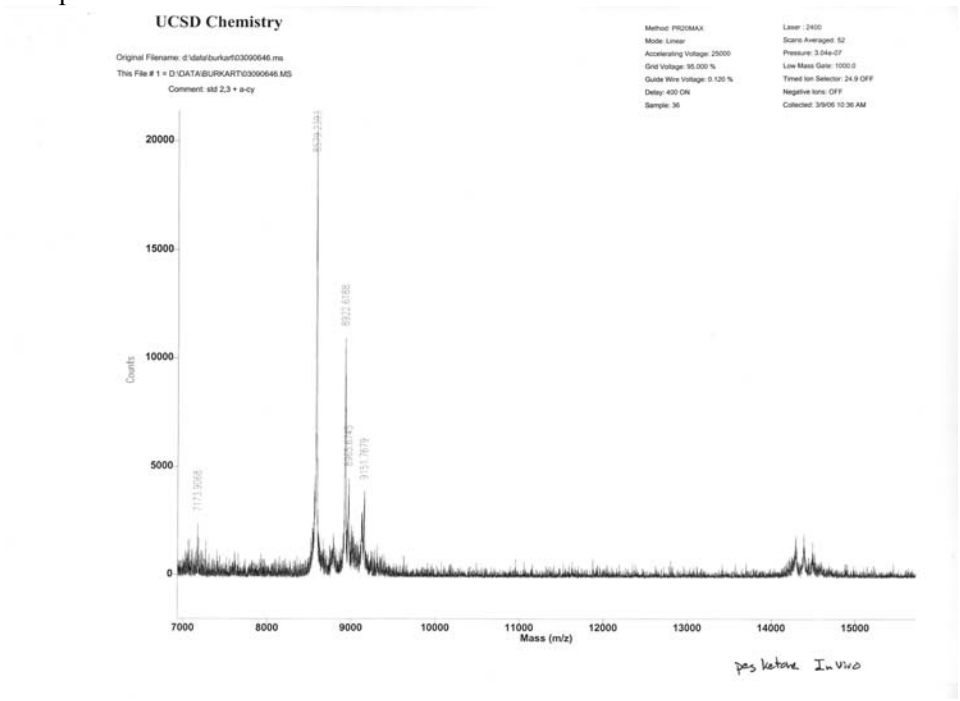
Compound **5** In vivo after click reaction with **10**



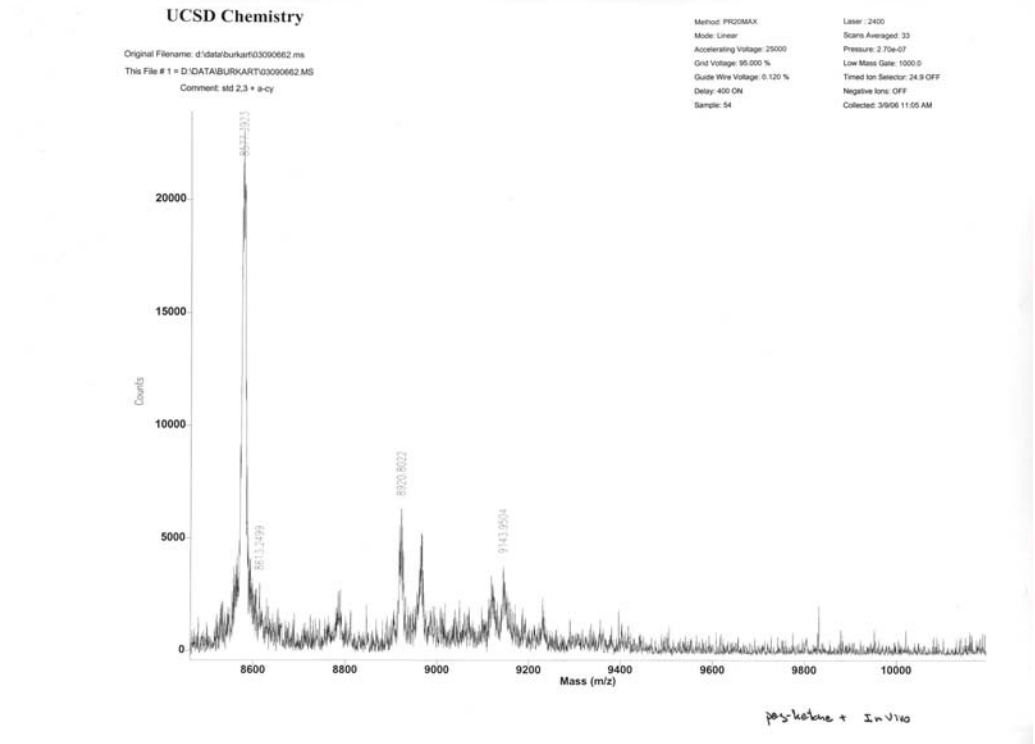
Compound **6** In vitro



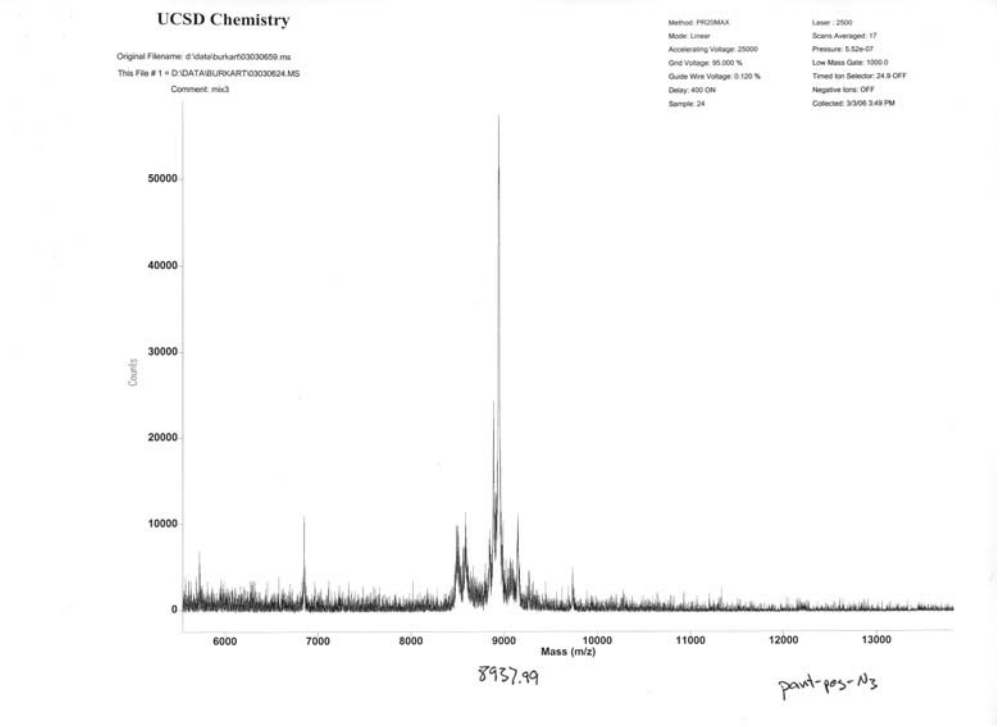
Compound **6** In vivo



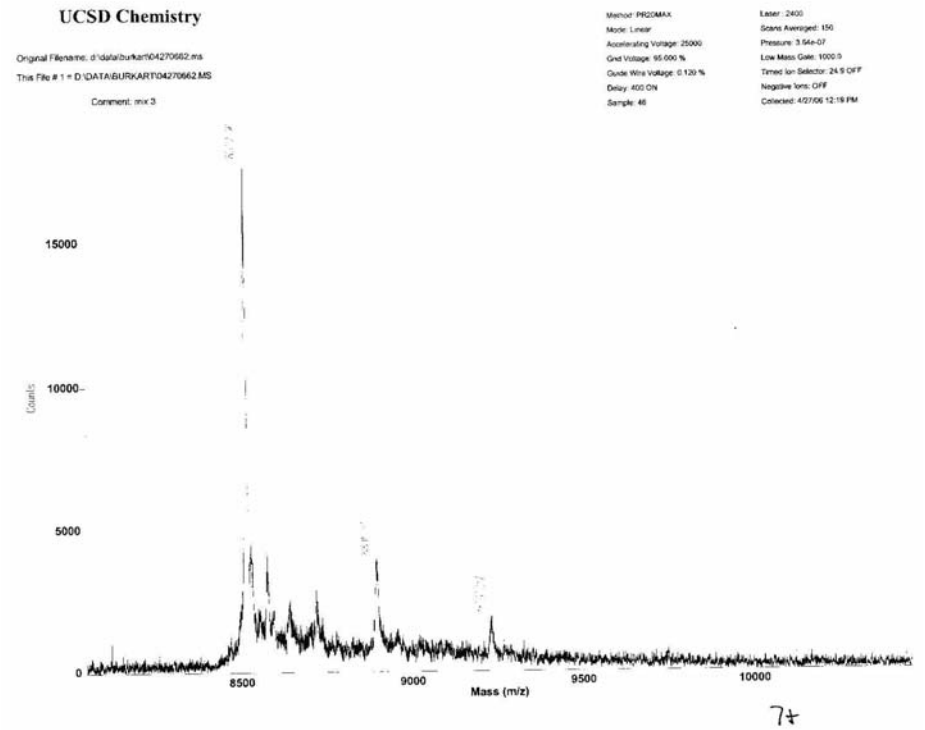
Compound **6** In vivo after reaction with **9**



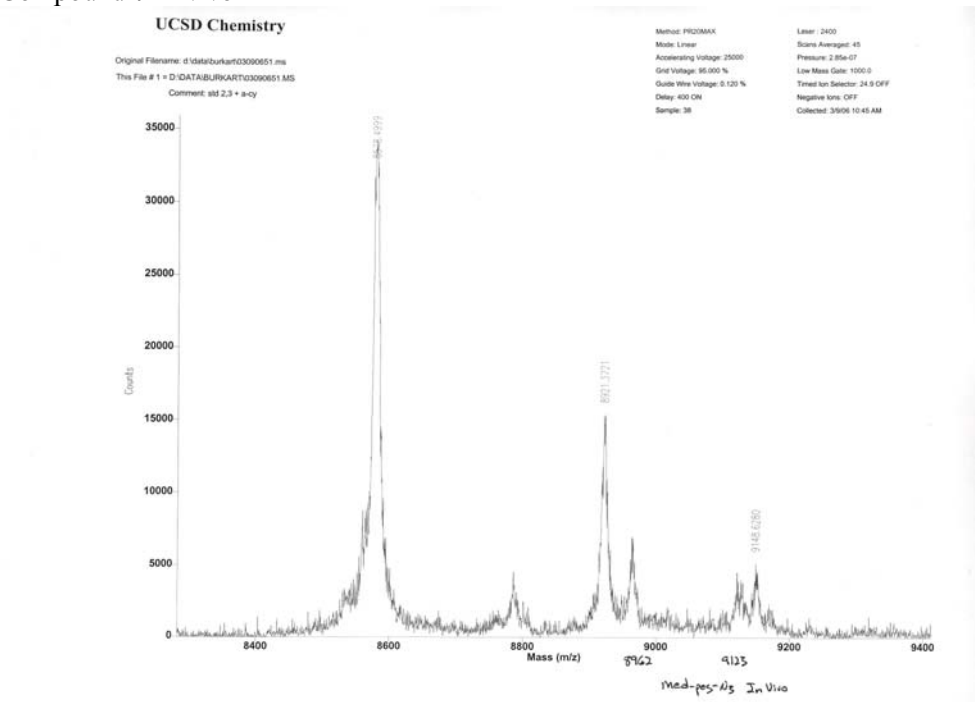
Compound **7** In vitro



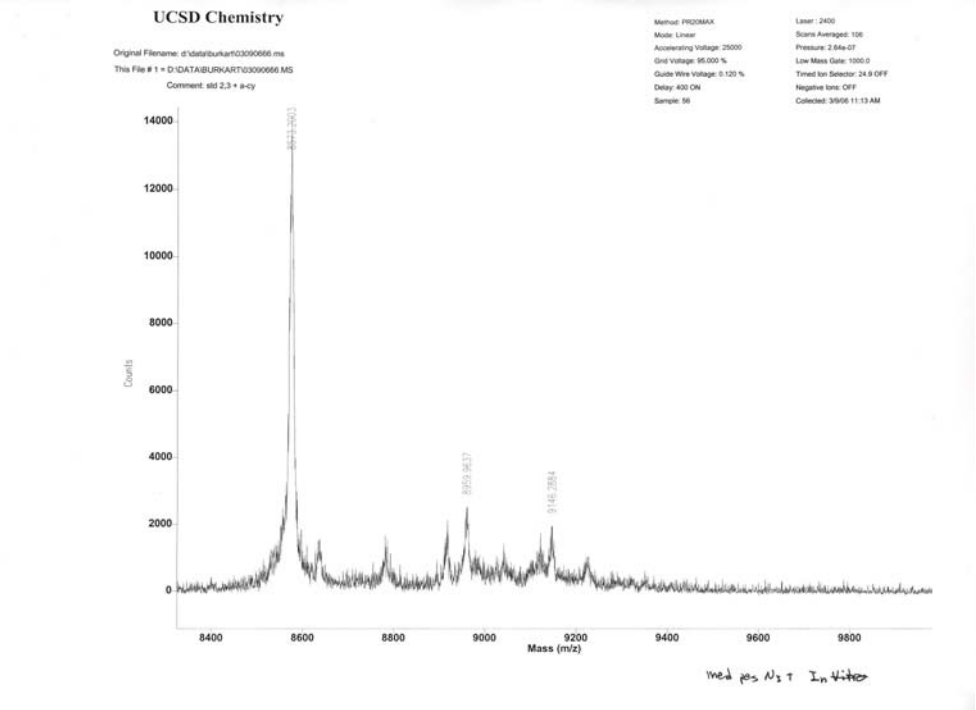
Compound **7** In vitro after click reaction with **10**



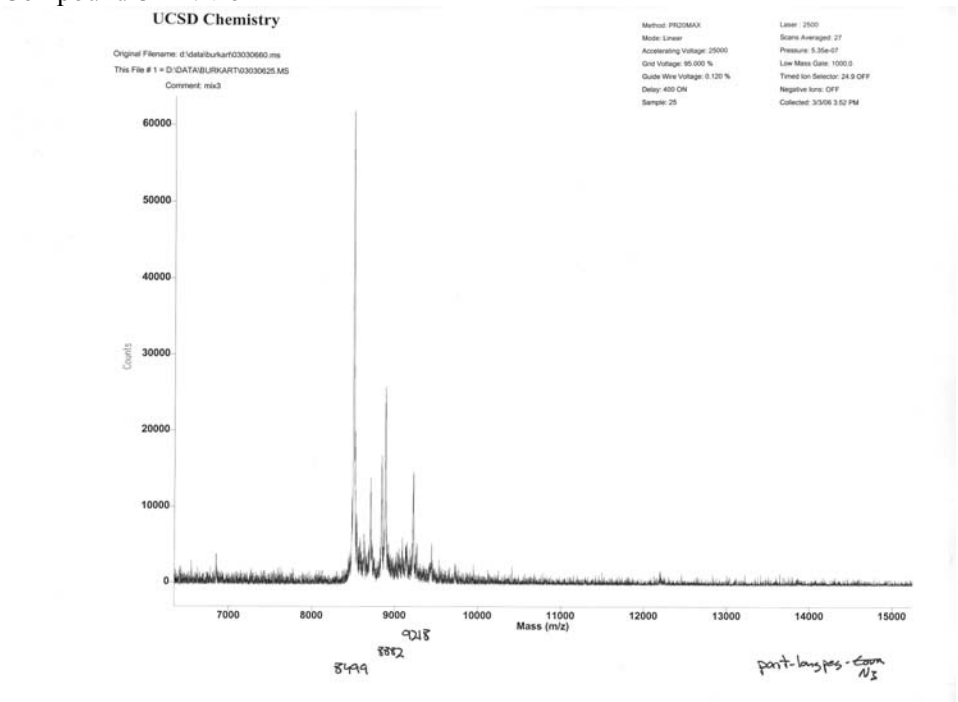
Compound **7** In vivo



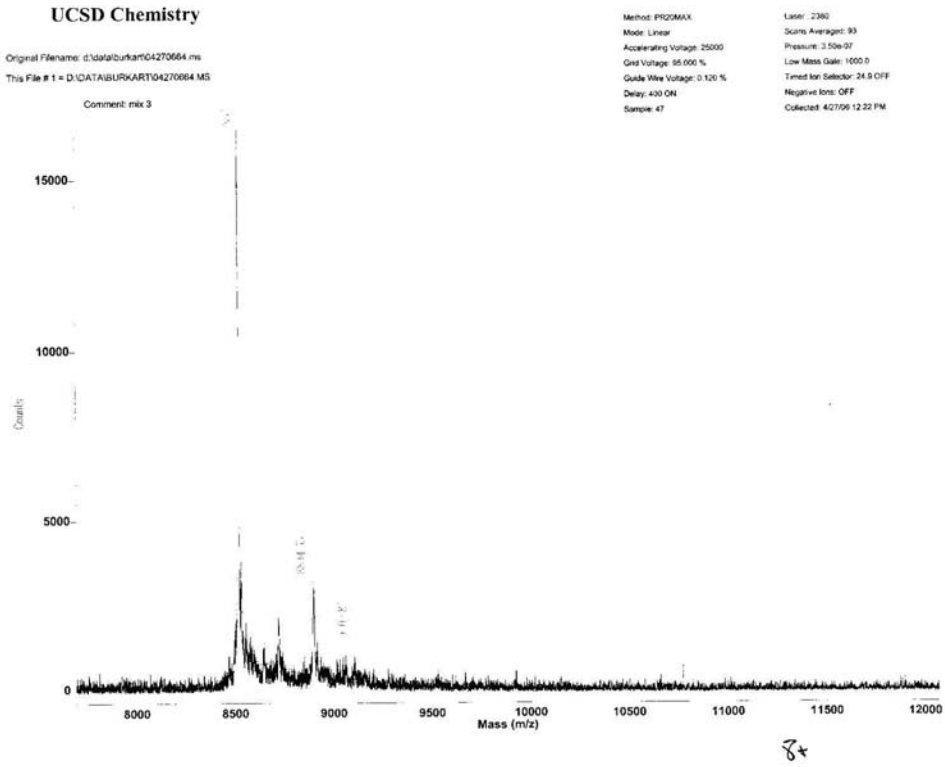
Compound **7** In vivo after click reaction with **10**



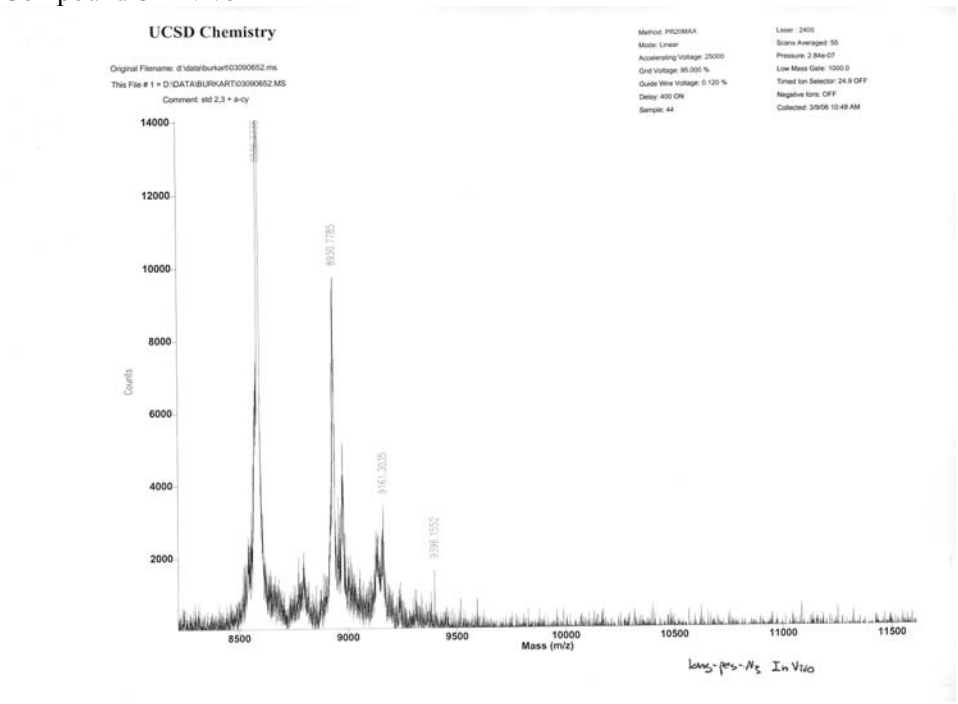
Compound **8** In vitro



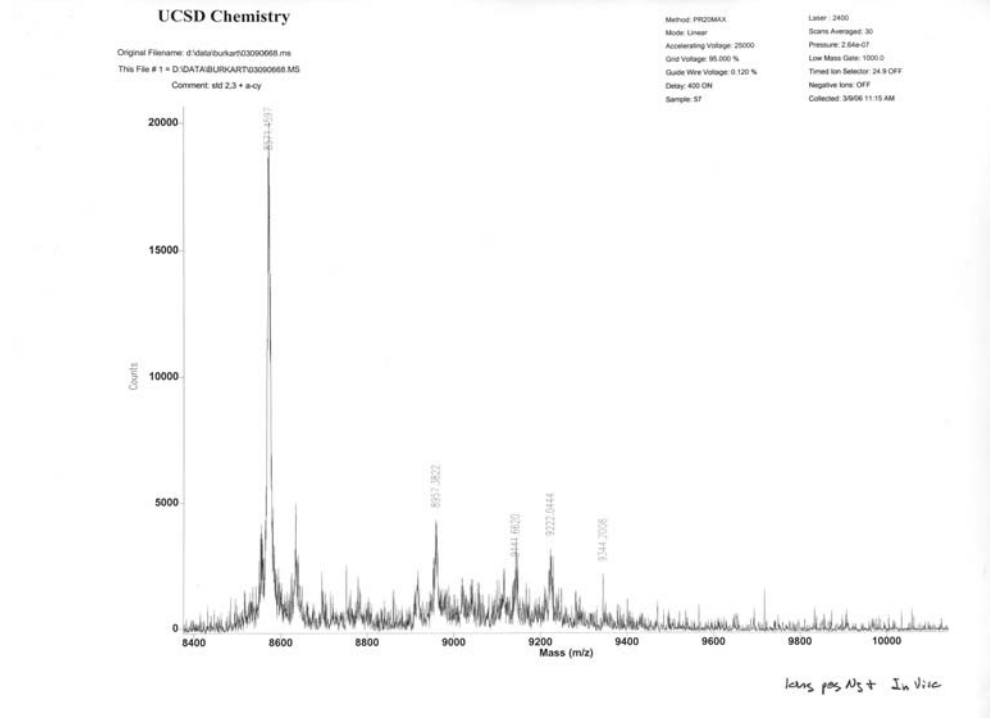
Compound **8** In vitro after click reaction with **10**



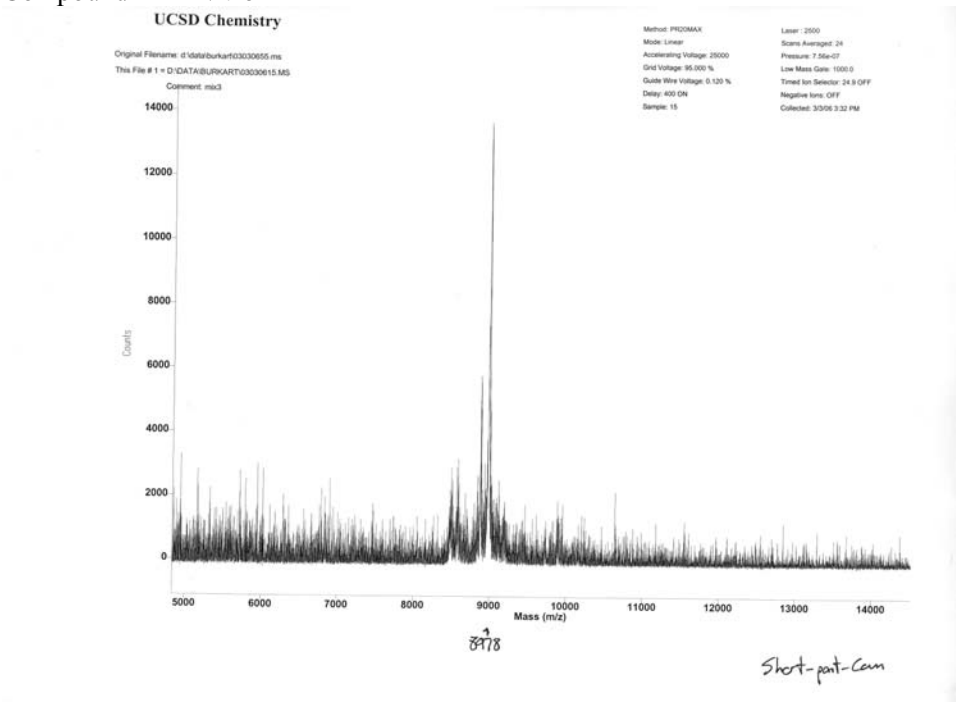
Compound **8** In vivo



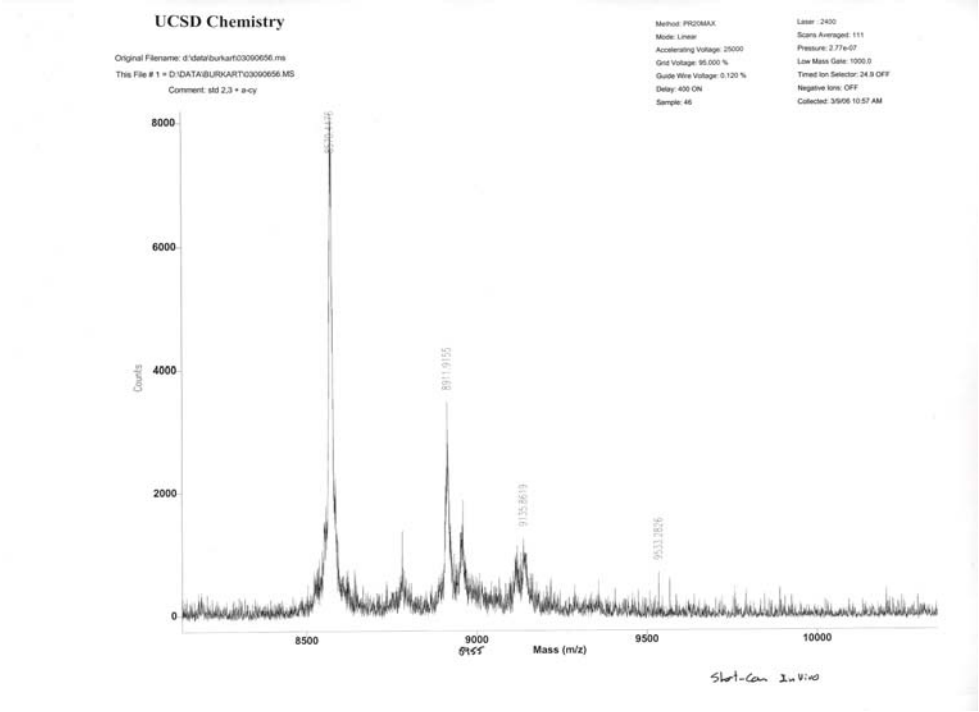
Compound **8** In vivo after click reaction with **10**



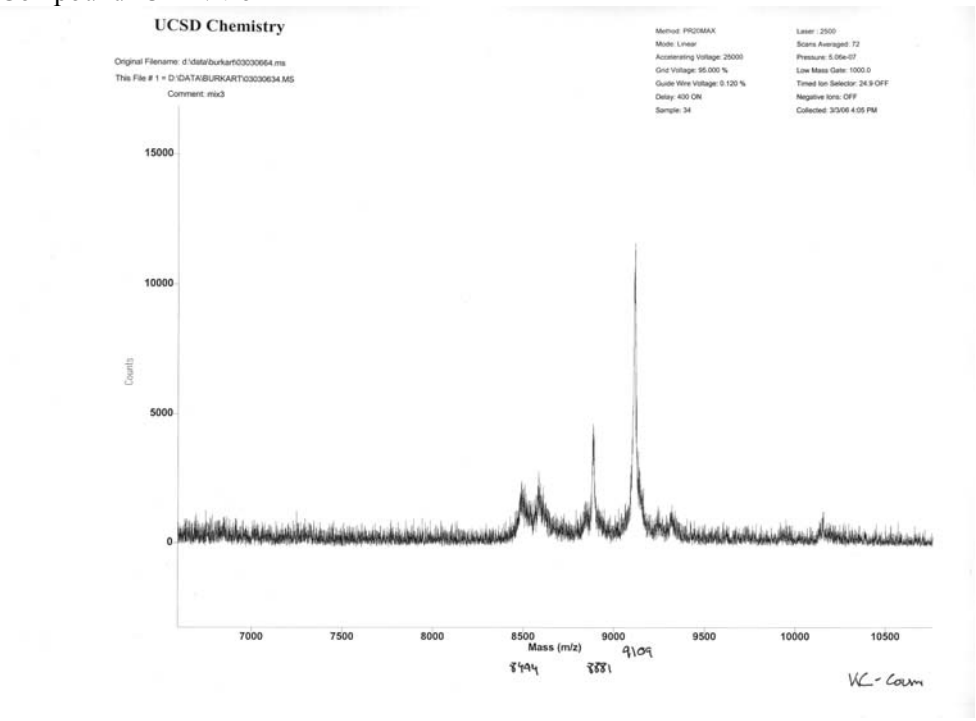
Compound **12** In vitro



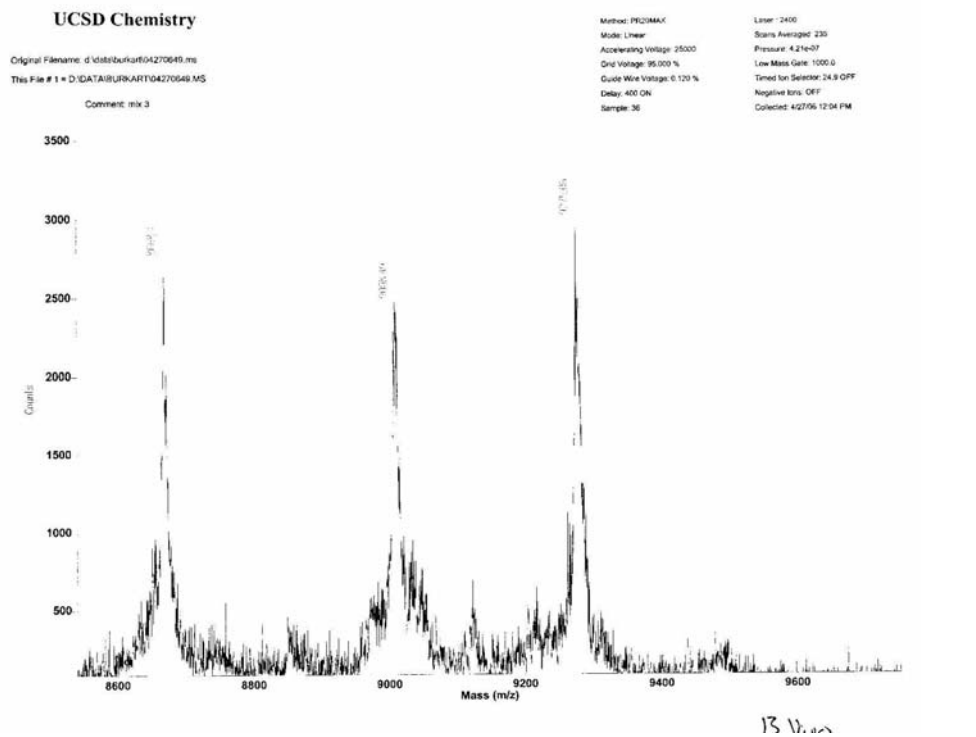
Compound **12** In vivo



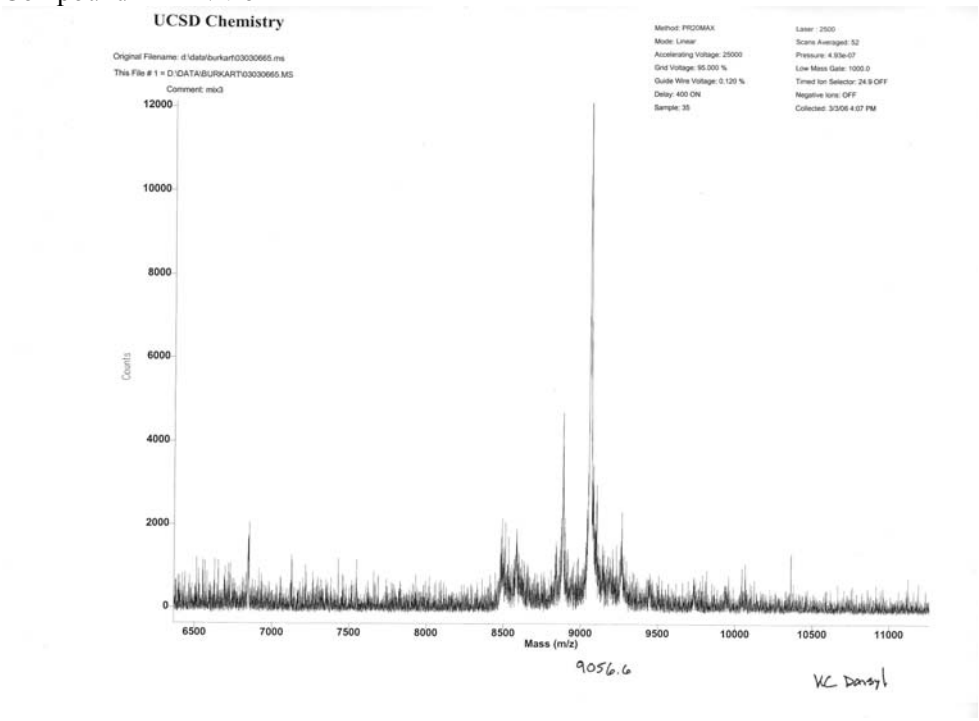
Compound **13** In vitro



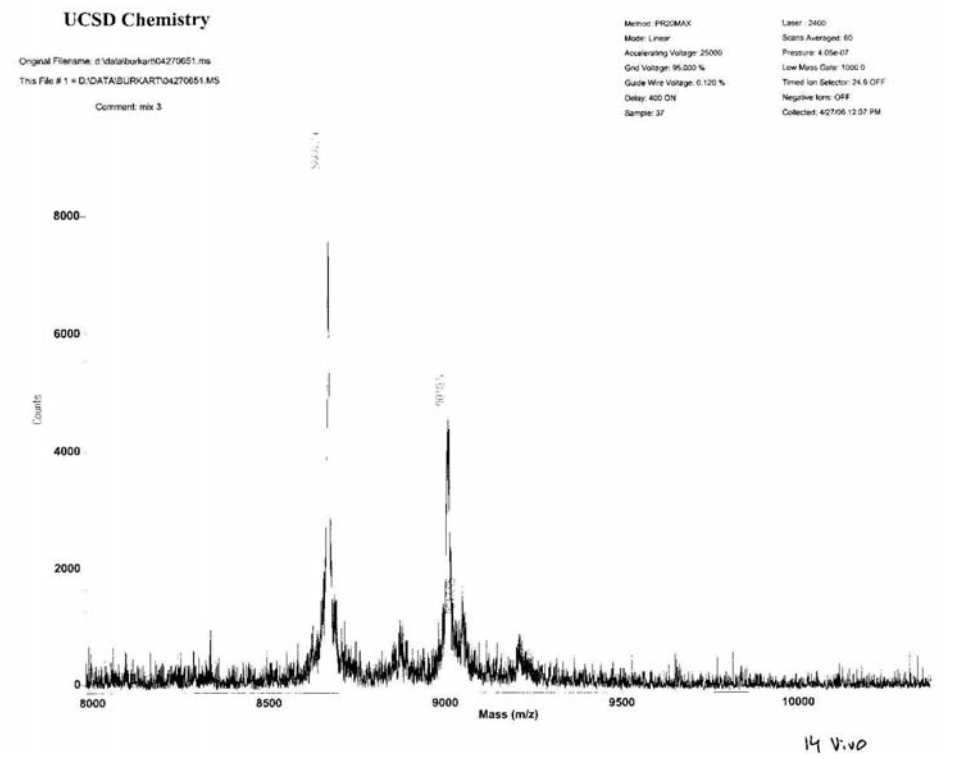
Compound **13** In vivo



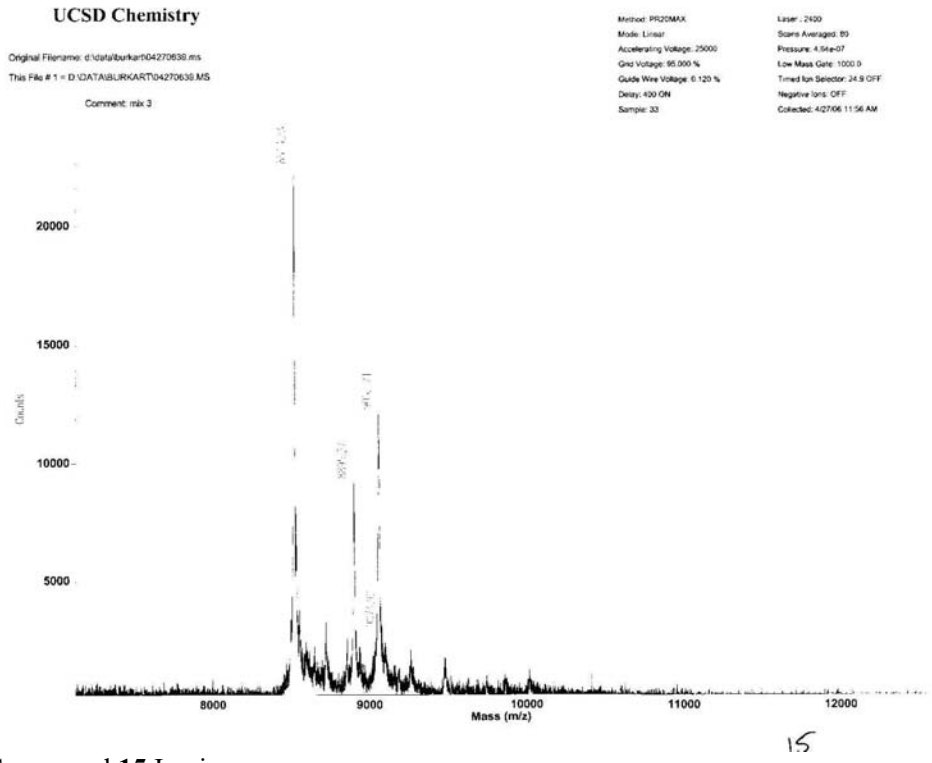
Compound **14** In vitro



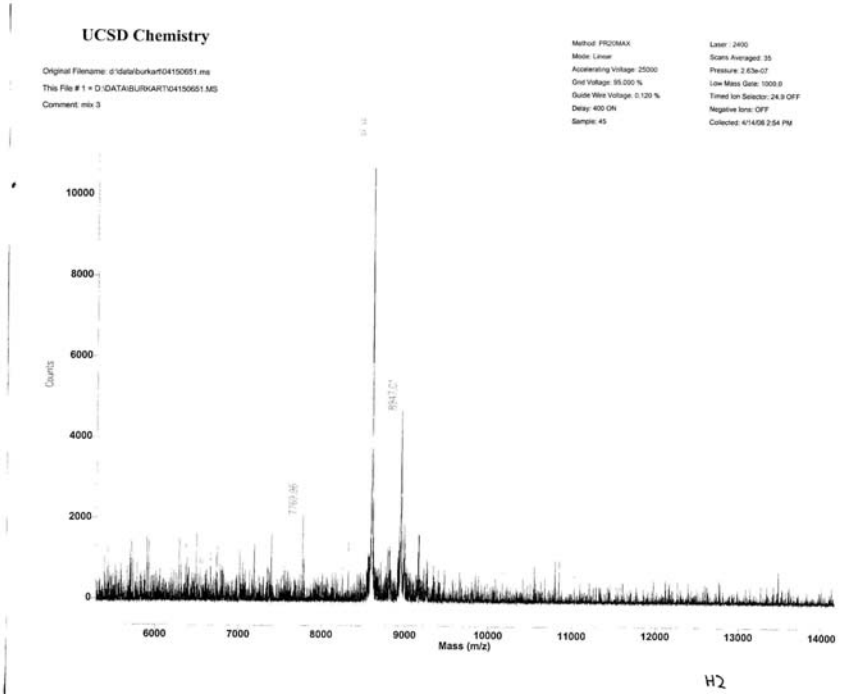
Compound **14** In vivo



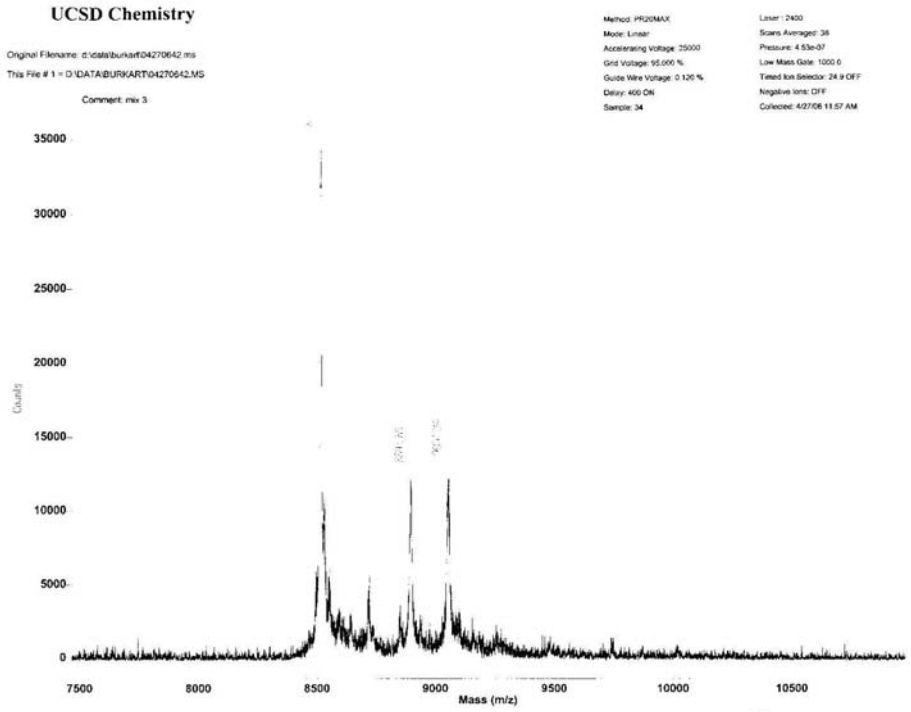
Compound **15** In vitro



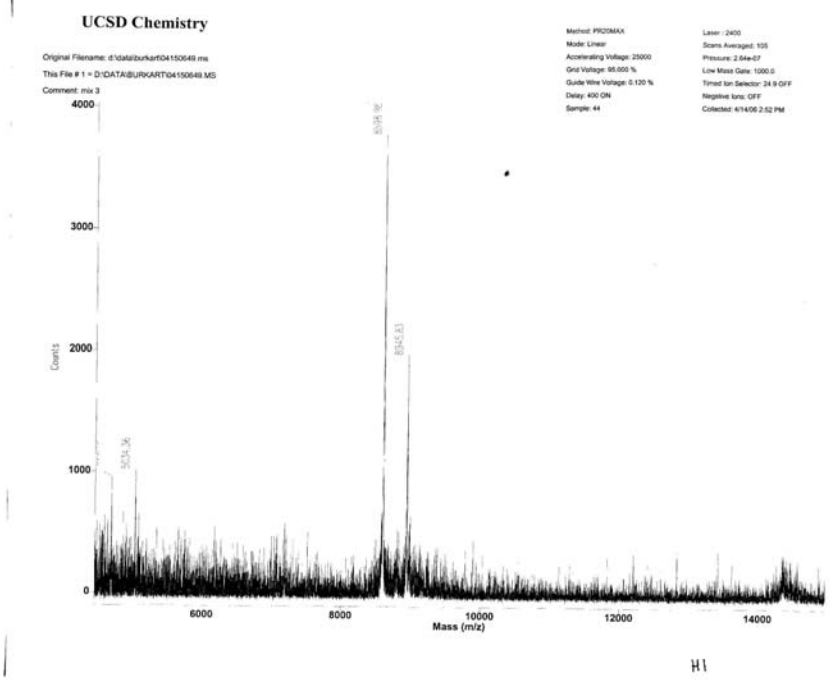
Compound **15** In vivo



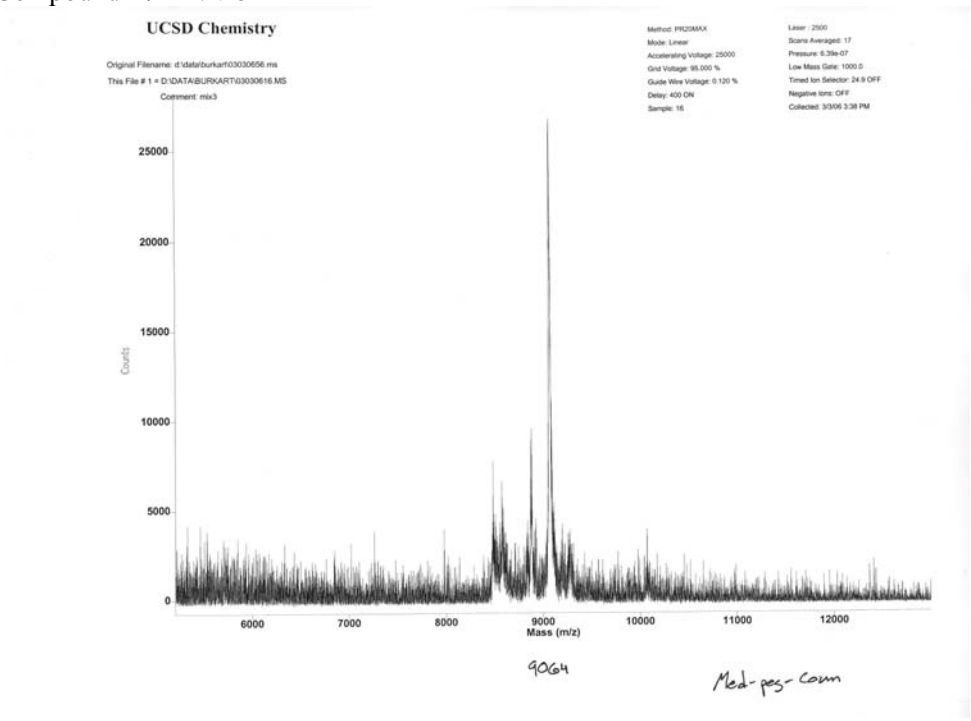
Compound **16** In vitro



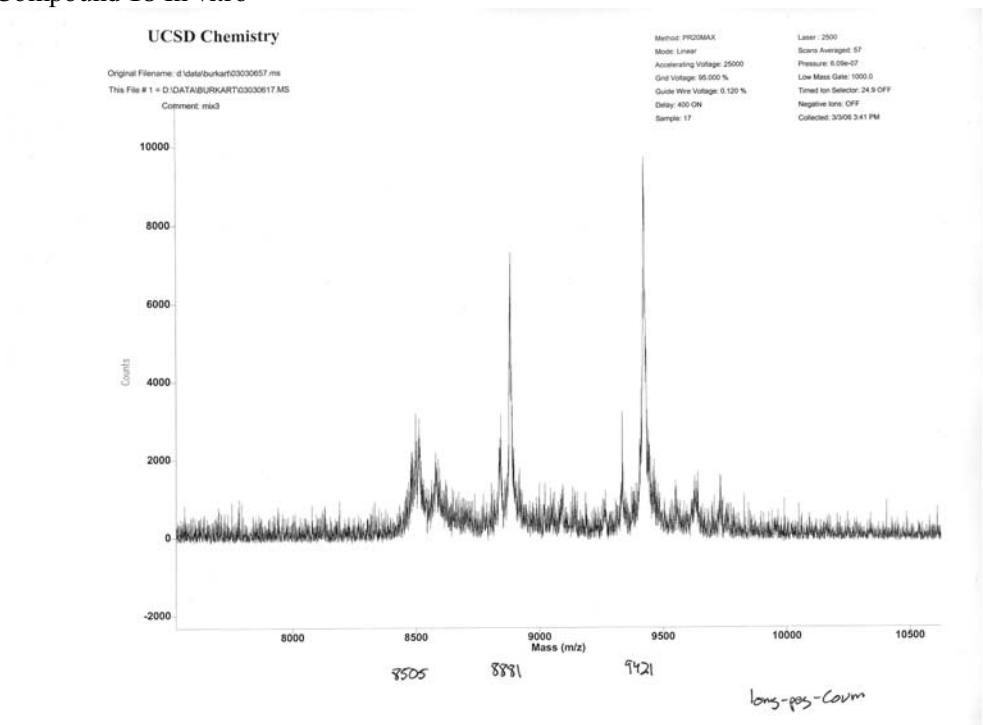
Compound **16** In vivo



Compound **17** In vitro



Compound **18** In vitro



Acknowledgments

The text of chapter 3, in full, is a reprint of material as it appears in: Meier, J.L., Mercer, A.C., Rivera, H., and Burkart, M.D. (2006). Synthesis and Evaluation of Bioorthogonal Pantetheine Analogues for In Vivo Protein Modification. *J. Am. Chem. Soc. 128*, *12174-12184*. Permission was obtained from all co-authors. I was the primary author of this work. Andrew Mercer performed biochemical assays and Heriberto Rivera synthesized compounds **15** and **16**. I synthesized all other compounds and wrote the manuscript. This research was performed under the supervision of Prof. Michael Burkart.

Chapter Four

Antibiotic evaluation and in vivo analysis of alkynyl coenzyme A antimetabolites in Escherichia coli

Originally published as:

Mercer, A. C.; Meier, J. L.; Hur, G. H.; Smith, A. R.; Burkart, M. D. *Bioorg Med Chem Lett* **18**, 5991-4 (2008).

Bioorganic & Medicinal Chemistry Letters 18 (2008) 5991–5994

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Antibiotic evaluation and in vivo analysis of alkynyl Coenzyme A antimetabolites in *Escherichia coli*

Andrew C. Mercer [†], Jordan L. Meier [†], Gene H. Hur, Andrew R. Smith, Michael D. Burkart [*]

Department of Chemistry and Biochemistry, University of California-San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0358, USA

ARTICLE INFO

Article history:
Received 8 July 2008
Accepted 18 July 2008
Available online 24 July 2008

Keywords:
Fatty acid
Coenzyme A
Antibiotic
Pantothenamides
Structure–activity relationships

ABSTRACT

Pantothenamides have been the subject of much study as potential inhibitors of CoA and carrier protein dependent biosynthetic pathways. Based on an initial observation of growth inhibition in *Escherichia coli* by **3**, we have synthesized a small panel of pantetheine analogues and re-examined the inhibitory properties of this class of antibiotics with an emphasis on understanding the ability of these compounds to act as substrates of native CoA and carrier protein utilizing biosynthetic pathways. Our findings suggest that a secondary structure–activity relationship is an important factor in the antibiotic activity of these compounds.

Coenzyme A (CoA) is an essential cofactor in living systems, functioning as the major acyl group carrier in numerous metabolic pathways.[1] The bacterial CoA biosynthetic pathway was first elucidated in *Escherichia coli*, where it was shown that five enzymes (CoaA, CoaB, CoaC, CoaD, and CoaE) are responsible for the conversion of pantothenate (vitamin B_5) to CoA (Fig. 1A).[2] While this pathway also exists in humans, comparison of the individual enzymes of the prokaryotic and eukaryotic CoA biosynthetic pathways has shown that the two demonstrate very little sequence homology.[3] With reports of bacterial resistance to traditional antibiotics becoming increasingly common, the design of selective inhibitors of bacterial CoA biosynthesis presents a potentially novel antibiotic target.[4]

In addition to functioning directly as an acyl group carrier, CoA is also the source of the 4′-phosphopantetheine arm used for acyl group transfer in fatty acid biosynthesis (Fig. 1A).[5] In this pathway, a phosphopantetheinyltransferase (PPTase) enzyme functions to transfer the 4′-phosphopantetheine arm of CoA to a conserved serine residue of an *apo*-acyl carrier protein (ACP). The free thiol of this posttranslational modification is then used as the site of acyl-intermediate tethering during the loading, condensation, and reduction reactions necessary for production of fatty acids. ACPs and peptidyl carrier proteins (PCPs) are used similarly in polyketide and nonribosomal peptide biosynthesis.[6]

In the past several years, it has been shown that many PPTases, most notably Sfp from the surfactin synthetase pathway in *B. subtilis*, have a relaxed substrate specificity which allows for the modification of ACPs with CoA analogues in vitro.[7] When derivatized with fluorescence or affinity tags, this property can be used for the selective visualization and isolation of carrier protein utilizing biosynthetic enzymes.[8] Our group has had a long-standing interest in using PPTase promiscuity as a strategy for the investigation of primary and secondary biosynthetic pathways in bacterial organisms, which has lead us to investigate methods for the manipulation of intracellular CoA pool as a means of labeling carrier proteins. As CoA analogues cannot cross the cell membrane due to their strong negative charge, we have examined the utility of CoA precursors as in vivo carrier protein labels.[9]

Perhaps the most thoroughly investigated CoA precursors to date have been the antibacterial pantothenamides.[10] This class of antibiotics, typified by N5-Pan (**1**), has been shown to inhibit *E. coli* and *Staphylococcus aureus* growth.[11] Originally postulated as inhibitors of the pantothenate kinase (CoaA) enzyme, it has since been shown that these compounds act as competitive substrates of CoaA, and are believed to exert their antibacterial activity through interference with fatty acid biosynthesis by labeling of the *E. coli* fatty acid ACP (Fig. 1B).[12,13]

During our own studies of CoA precursors we encountered an interesting phenomenon relevant to the continued development and study of this antibiotic class. Through the synthesis and evaluation of a large number of CoA precursors (henceforth referred to as pantetheine analogues), we identified a single compound (**2**) capable of modifying the *E. coli* fatty acid ACP in the native organism (Fig. 1B).[14] Further, these studies showed that compound **2** was non-toxic to *E. coli* grown in minimal media at concentrations >1 mM, while a structurally similar alkynyl pantetheine analogue (**3**) possessed cytotoxic activity near equivalent to that of **1**

* Corresponding author. Tel.: +1 858 534 5673; fax: +1 858 822 2182.
E-mail address: mburkart@ucsd.edu (M.D. Burkart).
† These authors contributed equally to this work.

0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.07.078

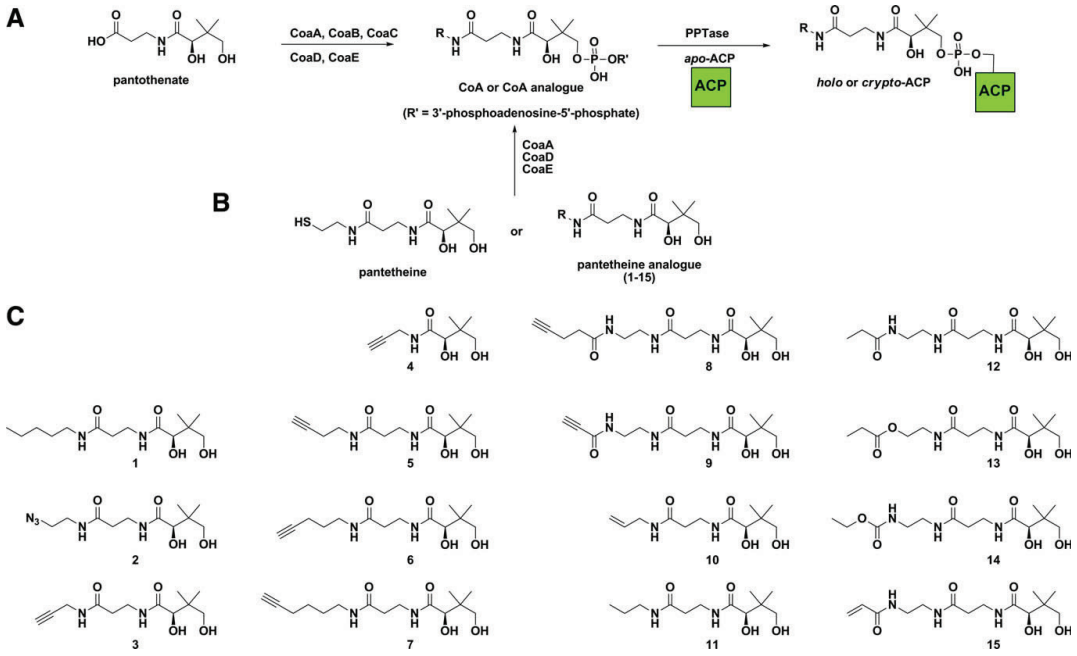


Figure 1. (A) Pathway for CoA biosynthesis and posttranslational modification of carrier proteins. (B) Pantetheine analogues can act as alternate substrates for these pathways to form CoA antimetabolites. These CoA analogues can be processed by PPTases to form *crypto*-carrier proteins. (C) Structures of pantothenate and pantetheine analogues (**1**–**15**) examined in this study.

(Table 3). These observations ran contrary to expectations, as it was anticipated that the most toxic pantetheine analogues would be those which labeled the *E. coli* fatty acid ACP most efficiently, according to the proposed mode of action of these compounds.[13]

These findings lead us to synthesize a small panel of pantetheine analogues in order to examine in detail the influence of chain length (**4–9**), oxidation state (**10–11**), and functionality (**12–15**) on the ability to serve as CoA antimetabolites and inhibit growth in *E. coli* (Fig. 1C).[15] We first tested these compounds as substrates of CoaA, the first enzyme of CoA biosynthetic pathway and the rate-limiting step for the accumulation of CoA metabolites.[16] In past studies, we have correlated good turnover by CoaA with in vivo CoA production and carrier protein modification.[17] As can be seen in Table 1, many of the analogues exhibit good turnover with some (**1–3**, **5**, **10–11**, **13**) approaching the kinetic efficiency of pantothenate. When examining the effect of chain length on turnover of alkynyl pantetheines, **4** shows very poor catalytic efficiency due mainly to its high K_m. This is indicative of weak binding, a conclusion consistent with the crystal structure of *E. coli* CoaA which shows the terminal acid of pantothenate having an important role in binding. This can be replaced by interaction with an amide which is missing in compound **4**.[4] In contrast, **3** and **5** show catalytic efficiencies in the range of the native substrate, while **6–9** show a decrease in efficiency that correlates to the length of the

Table 1
Kinetic data for CoaA turnover of natural substrate (pantothenate) as well as pantetheine analogues **1**–**15**[a]

Compound #	K_m (μM)	V_{max} (μM s^{-1})	k_{cat} (min^{-1})	k_{cat}/K_m (mM^{-1} min^{-1})
Pantothenate	35.81 (±6.00)	0.31 (±0.01)	0.76 (±0.03)	21.13 (±1.93)
1	33.98 (±7.11)	0.34 (±0.02)	0.84 (±0.05)	24.59 (±2.88)
2	43.51 (±3.4)	0.28 (±0.002)	0.69 (±0.01)	16.1 (±0.53)
3	36.04 (±6.05)	0.31 (±0.02)	0.77 (±0.04)	21.38 (±2.11)
4	396.50 (±123.7)	0.26 (±0.05)	0.64 (±0.12)	1.62 (±0.61)
5	34.04 (±10.72)	0.33 (±0.03)	0.81 (±0.07)	23.78 (±4.23)
6	40.34 (±12.95)	0.29 (±0.03)	0.72 (±0.07)	17.95 (±3.41)
7	40.00 (±14.26)	0.29 (±0.03)	0.72 (±0.07)	17.97 (±3.75)
8	47.80 (±14.62)	0.23 (±0.02)	0.56 (±0.05)	11.70 (±2.22)
9	48.49 (±10.99)	0.22 (±0.02)	0.54 (±0.04)	11.11 (±1.53)
10	41.16 (±5.94)	0.38 (±0.02)	0.94 (±0.04)	22.85 (±1.96)
11	36.91 (±7.52)	0.37 (±0.02)	0.91 (±0.05)	24.65 (±2.88)
12	34.57 (±6.07)	0.22 (±0.01)	0.53 (±0.03)	15.43 (±1.52)
13	27.78 (±5.21)	0.26 (±0.01)	0.65 (±0.03)	23.49 (±2.33)
14	33.70 (±3.98)	0.25 (±0.01)	0.62 (±0.02)	18.29 (±1.20)
15	57.97 (±12.01)	0.26 (±0.02)	0.64 (±0.04)	10.96 (±1.49)

[a] Values are means of three experiments, standard deviation is given in parentheses.

analogue. This decrease in efficiency is from both binding and turnover, as K_m increases and V_{max} decreases with the growing length of the alkynyl pantetheine substrates. Compounds **8** and **9** highlight another interesting trend seen in this panel, which is the detrimental effect of an amide bond at the natural thiol position of pantetheine on CoaA turnover. This effect is most readily observed by comparison of **12**, **13**, and **14**, which demonstrates a steady decrease in catalytic efficiency with decreasing polarization of the carbonyl (amide < carbamate < ester) at this position (see Table 1).

In comparing what effect degree of unsaturation has on turnover of pantetheine analogues by CoaA, the overall trend appears to indicate marginally improved turnover for fully saturated alkyl-pantetheine analogues. While this effect is slight among those pantetheine analogues terminating in propyl-derived chains (**3**, **10**, **11**), it can be clearly observed upon comparison of **1** to **6**. A similar effect is observed on comparison of amide bond-extended analogues **12** and **15.**

While kinetic values of pantetheine analogues with CoaA are a good predictor of in vivo activity, many other factors, including cell permeability and susceptibility to efflux pumps, impact the performance of antimetabolites when administered to living cells. To analyze the ability of these compounds to be processed by the CoA biosynthetic pathway in vivo and to interact with carrier proteins as CoA analogues, we utilized an in vivo assay.[17] This assay provides a qualitative measure of the ability of pantetheine analogues to be processed by the endogenous CoA biosynthetic enzymes of *E. coli* by coupling CoA analogue production to the modification of a carrier protein. To facilitate detection and isolate CoA biosynthesis from variables such as carrier protein expression and PPTase promiscuity, *E. coli* are first transformed with expression plasmids for a carrier protein, in this case the Fren-ACP from the frenocylin polyketide synthase, as well as the PPTase Sfp, which is known to have a very broad substrate specificity. After growth to mid-log phase, the pantetheine analogue (**1–15**) is added at the same time as IPTG, which induces expression of the reporter system.[15] Compounds that exhibit uptake and processing by the native *E. coli* CoA biosynthetic pathway produce modified ACPs, which demonstrate a mass shift characteristic of posttranslational modification by each analogue, and can be observed by MALDI-TOF (Table 2).

Having confirmed that the majority of these compounds are capable of formation of CoA analogues in vivo, we sought to correlate our findings with their antibacterial activity in native *E. coli*. To investigate the effects that additives present in the media might have on antibiotic activity, we determined the MIC values for **1–15** using *E. coli* K12 grown in both minimal (M9) media, as well as in a richer, 1% tryptone broth which had been used to determine MIC values in an earlier study of pantothenamides.[11,13]

Inspecting the results, all of the analogues tested showed greater growth inhibition in minimal media compared to rich media (Table 3). These results show a direct correlation between toxicity and CoaA kinetic profile for these compounds. This is to be expected, as it has previously been shown that CoaA is the rate-limiting step for CoA biosynthesis in vivo, and the antibacterial activity of these compounds is believed to be dependent on their in vivo transformation to CoA analogues. The major outliers in this respect are **2** and **13**, which possess good kinetics but do not show inhibition of *E. coli* at concentrations up to 500 μM even in minimal media. Additional evidence that these compounds act as CoA antimetabolites was provided by the observation that the inhibitory effects of the most toxic members of this panel (**1**, **3**, **5–6**, **10–11**) were greatly decreased by addition of the CoA precursors pantothenate and β-alanine to the growth medium (Table 3). Among the alkynyl analogues which initially inspired this study (**3, 5–9**), an increasing MIC value is observed with growing chain length, mirroring the decline in catalytic efficiency observed among this group. Interestingly, among alkyl pantetheine analogues of the same chain-length (**3** and **11, 1** and **6**), changing the oxidation state from an alkyne to a saturated alkyl chain lowers the MIC by a factor of two to four. However, while **11** is sixfold more active than **1** in minimal media, administration of these same compounds to *E. coli* grown on rich media shows **11** to be at least 10× less toxic under these conditions.

The kinetic profiles, in vivo analysis, and inhibitory data generated here all support the previously held hypothesis that the antibiotic activity of pantetheine analogues is due to the production of CoA analogues in vivo.[12] However, the finding that saturated and unsaturated pantetheine analogues demonstrate rates of CoaA turnover within error of one another (i.e., **3** and **11**), yet show drastically different MICs suggests that CoA analogue production alone is not sufficient for antibacterial activity. Based on our results it appears that of the pantetheine analogues processed efficiently by CoaA, those terminating in fully saturated alkyl groups are ideal for activity, while substitution by unsaturated alkynyl chains and polar terminal groups on the pantetheine chain (i.e., **2** and **13**) results in decreased or no growth inhibition. This suggests a secondary structure–activity relationship for pantetheine analogue inhibition, in which one set of structural characteristics is necessary for biosynthetic processing and formation of CoA or ACP-analogues in vivo, while the identity

Table 2
MALDI-TOF data for in vivo modification of Fren-ACP by compounds **1–15**

Compound #	*apo*	Modified	Difference	Expected
None (control)	8663	9000[a]	337[a]	342[a]
1	8660	9011	351	351
2	8663	9018	355	358
3	8663.7	8982	318.3	319
4	8661.1	—	—	248
5	8664.9	8996	331.1	333
6	8662.2	9008	345.8	347
7	8665.8	9025.9	360.1	361
8	8663.4	9065	401.6	404
9	8666	9042 & 9350[b]	376[b]	376
10	8658	8979	321	321
11	8660	8982	322	323
12	8660	9042	382	380
13	8660	9046	386	381
14	8660	9058	398	396
15	8666	9039	373	376

[a] Modification by native CoA of *E. coli* results in the expected mass shift.
[b] Compound **9** gives a large peak at 9350 in addition to the expected peak at 9042, possibly due to matrix interactions with the unsaturated activated alkyne.

Table 3
Minimum inhibitory concentrations of pantetheine analogues **1–15** to *E. coli* grown in different media and effect of additives

Media compound #	M9 MIC (μM)	Tryptone MIC (μM)	M9 + Pan[a] MIC (μM)	M9 + β-Ala[b] MIC (μM)
1	6	50	500	250
2	>500	>500	nd[c]	nd
3	4	>500	>500	>500
4	500	>500	nd	nd
5	7.5	>500	>500	>500
6	15	>500	>500	>500
7	500	>500	nd	nd
8	>500	>500	nd	nd
9	>500	>500	nd	nd
10	3	>500	250	125
11	1	500	500	250
12	>500	>500	nd	nd
13	500	>500	nd	nd
14	>500	>500	nd	nd
15	500	>500	nd	nd

[a] Pan, addition of 1 mM pantothenate to growth medium.
[b] β-Ala, addition of 1 mM β-alanine to growth medium.
[c] nd, not determined.

of the terminal group facilitates interaction of the CoA or ACP-analogue with a biologically relevant target.

In *E. coli*, fatty acid ACP is estimated to be present in the cytosol at concentrations approaching 1 mM. This abundance may explain the observation that **2** is capable of ACP modification in native *E. coli* without toxicity. Why then do alkyl pantetheine analogues **1** and **10**–**11** show increased inhibitory properties? Based on their kinetics with CoaA, these analogues seem unlikely to modify ACP at substantially higher levels than **2** in vivo. More likely, and in agreement with the secondary structural features of antimicrobial pantetheine analogues observed here, is the hypothesis that the activity of alkyl pantetheine analogues is due to the differential activity of alkyl-ACP-prodrugs to bind and disrupt the associated lower abundance enzymes of fatty acid biosynthesis. Further elucidation of this process may have important implications for design of new members of this antibiotic class.

Acknowledgment

This work was funded by NIH RO1GM075797; JM was supported by T32DK007233.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2008.07.078.

References and notes

1. Leonardi, R.; Zhang, Y. M.; Rock, C. O.; Jackowski, S. *Prog. Lipid Res.* **2005**, *44*, 125.
2. Strauss, E.; Kinsland, C.; Ge, Y.; McLafferty, F. W.; Begley, T. P. *J. Biol. Chem.* **2001**, *276*, 13513.
3. Gerdes, S. Y.; Scholle, M. D.; D'Souza, M.; Bernal, A.; Baev, M. V.; Farrell, M.; Kurnasov, O. V.; Daugherty, M. D.; Mseeh, F.; Polanuyer, B. M.; Cmapbell, J. W.; Anantha, S.; Shatalin, K. Y.; Chowdhury, S. A.; Fonstein, M. Y.; Osterman, A. L. *J. Bacteriol.* **2002**, *184*, 4555.
4. Ivey, R. A.; Zhang, Y. M.; Virga, K. G.; Hevener, K.; Lee, R. E.; Rock, C. O.; Jackowski, S.; Park, H. W. *J. Biol. Chem.* **2004**, *279*, 35622.
5. Mercer, A. C.; Burkart, M. D. *Nat. Prod. Rep.* **2007**, *24*, 750.
6. Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468.
7. Quadri, L. E.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585.
8. La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. *Chem. Biol.* **2004**, *11*, 195.
9. Clarke, K. M.; Mercer, A. C.; La Clair, J. J.; Burkart, M. D. *J. Am. Chem. Soc.* **2005**, *127*, 11234.
10. Spry, C.; Kirck, K.; Saliba, K. J. *FEMS Microbiol. Rev.* **2008**, *32*, 56.
11. Virga, K. G.; Zhang, Y. M.; Leonardi, R.; Ivey, R. A.; Hevener, K.; Park, H. W.; Jackowski, S.; Rock, C. O.; Lee, R. E. *Bioorg. Med. Chem.* **2006**, *14*, 1007.
12. Strauss, E.; Begley, T. P. *J. Biol. Chem.* **2002**, *277*, 48205.
13. Zhang, Y. M.; Frank, M. W.; Virga, K. G.; Lee, R. E.; Rock, C. O.; Jackowski, S. *J. Biol. Chem.* **2004**, *279*, 50969.
14. Mercer, A. C.; Meier, J. L.; Torpey, J. W.; Burkart, M. D. Submitted for publication.
15. See Supporting information for synthetic procedures and data as well as assay details.
16. Jackowski, S.; Rock, C. O. *J. Bacteriol.* **1981**, *148*, 926.
17. Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. *J. Am. Chem. Soc.* **2006**, *128*, 12174.

Antibiotic Evaluation and In Vivo Analysis of Alkynyl Coenzyme A Antimetabolites in Escherichia Coli

*Andrew C. Mercer, Jordan L. Meier, Gene H. Hur, Andrew R. Smith, and Michael D. Burkart**
Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358
mburkart@ucsd.edu

Online Supporting Information

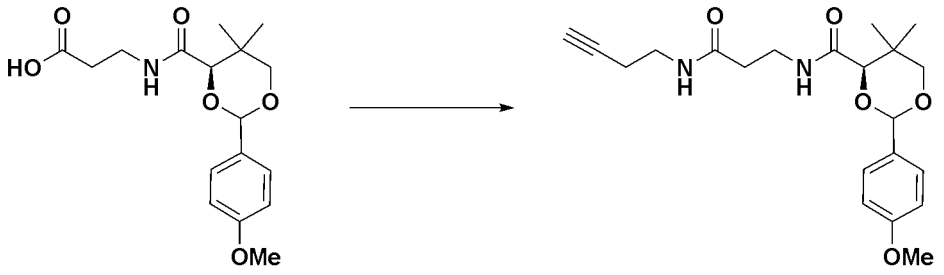
Contents

General synthetic procedures and materials

All commercial reagents (Sigma-Aldrich, Alfa Aesar, TCI America, Acros) were used as provided unless otherwise indicated. Analogues **1-4**, **11**, and **15** were prepared according to published literature procedures.[1-3] All reactions were carried out under argon atmosphere in dry solvents with oven-dried glassware and constant magnetic stirring unless otherwise noted. Triethylamine (TEA), and ethyl-*N,N*-diisopropylamine (DIPEA) were dried over sodium and freshly distilled. ^{1}H-NMR spectra were taken at 300, 400, or 500 MHz and ^{13}C-NMR spectra were taken at 100.6 or 75.5 MHz on Varian NMR spectrometers and standardized to the NMR solvent signal as reported by Gottlieb.[4] Multiplicities are given as s=singlet, d=doublet, t=triplet, q=quartet, p=pentet, dd=doublet of doublets, bs=broad singlet, bt=broad triplet, m=multiplet using integration and coupling constant in Hertz. TLC analysis was performed using Silica Gel 60 F254 plates (EM Scientific) and visualization was accomplished with UV light (λ=254 nm) and/or the appropriate stain (iodine, 2,4-dinitrophenylhydrazine, cerium molybdate, ninhydrin). Silica gel chromatography was carried out with Silicycle 60 Angstrom 230-400 mesh according to the method of Still.[5] TLC prep plate purification was performed with EMD Silica Gel 60 F_{254} pre-coated plates. Electrospray (ESI) and fast atom bombardment (FAB) mass spectra were obtained at the UCSD Mass Spectrometry Facility by Dr. Yongxuan Su using a Finnigan LCQDECA mass spectrometer and a ThermoFinnigan MAT- 900XL mass spectrometer, respectively.

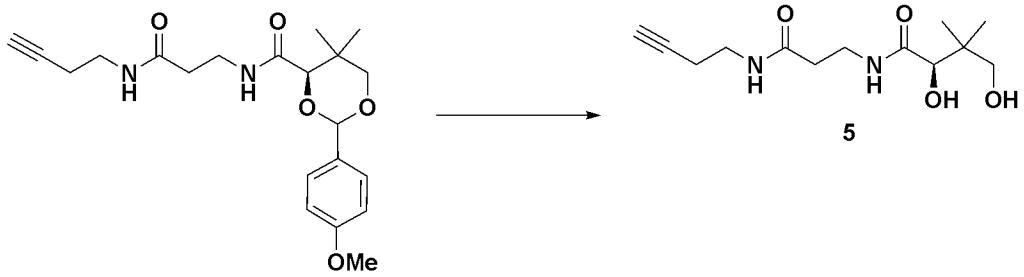
Representative Procedures and Spectroscopic Data for Compounds 1-15

A representative procedure is provided for preparation of compound **5**; all other pantetheine analogues were prepared analogously from the corresponding acids or acyl chlorides. Spectroscopic data for previously published analogues (**1-4**, **11**, and **15**) were in accordance with the published literature values.



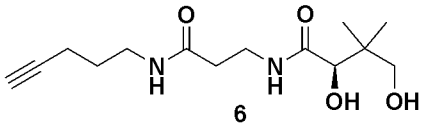
PMB-protected pantothenate[2] (434 mg, 1.28 mmol), PyBOP (732 mg, 1.41 mmol), and 4-amino-1-butyne[6] (88mg, 1.28 mmol) were combined and dissolved in dry CH_2Cl_2 (50 mL). DIPEA (677 µl, 3.84 mmol) was added at 0 ^{o}C, and the reaction was allowed to stir overnight at room temperature. The solvent was removed under reduced pressure, and the resultant oil was purified by column

chromatography (Hex:EtOAc, 2:1 to EtOAc) to yield compound PMB-protected **5** as an oil (405 mg, 80%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.42 (d, *J*= 8.5 Hz, 2H), 7.00 (bt, 1H), 6.92 (d, *J*= 9.0 Hz, 2H), 6.03 (bt, 1H), 5.45 (s, 1H), 4.08 (s, 1H), 3.82 (s, 3H), 3.68 (dd, *J*= 23.0 Hz, *J*= 12.0 Hz, 2H), 3.59-3.49 (m, 2H), 3.43-3.32 (m, 2H), 2.44 (t, *J*= 6.5 Hz, 2H), 2.34 (td, *J*= 7.1 Hz, *J*= 2.0 Hz, 2H), 1.98 (t, *J*= 2.5 Hz, 1H), 1.10 (s, 3H), 1.09 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 171.1, 169.7, 160.4, 130.3, 127.7, 113.9, 101.5, 84.0, 81.6, 78.7, 70.2, 55.5, 38.2, 36.2, 35.1, 33.3, 22.0, 19.6, 19.3. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{21}H_{28}O_5N_2$, 388.1993, found 388.1987.



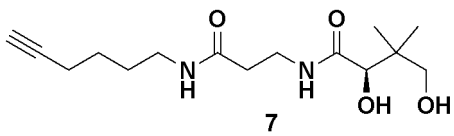
PMB-protected **5** (87 mg, 0.22 mmol) was dissolved in 1N HCl:THF, 1:1 (10 ml) and stirred for 1 hr or until the starting material was gone by TLC, at which point 1N NaOH or AG-1-X8 Strong Basic anionic exchange resin was used to neutralize the solution. The solvent was removed under reduced pressure, and the crude product was purified by column chromatography (EtOAc to 15% MeOH-EtOAc) to afford compound **5** as an oil (11 mg, 18%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.43 (bt, 1H), 6.39 (bt, 1H), 4.23 (d, *J*= 6.5 Hz, 1H), 3.99 (d, *J*= 5.2 Hz, 1H), 3.72 (bt, 1H), 3.56 (p, *J*= 5.3 Hz, 2H), 3.47 (d, *J*= 5.3Hz, 2H), 3.38 (qd, *J*= 6.2 Hz, *J*= 1.2 Hz, 2H), 2.46 (t, *J*= 6.0 Hz, 2H), 2.39 (td, *J*= 6.4 Hz, *J*= 2.4 Hz, 2H), 2.03 (t, *J*= 2.7 Hz, 1H), 0.99 (s, 3H), 0.91 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 173.8, 171.8, 81.5, 77.8, 71.1, 70.4, 39.5, 38.3, 35.9, 35.4, 21.7, 20.6, 19.5. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{13}H_{23}O_5N_2$, 271.1652, found 271.1657.

Pantetheine Analogue 6



^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.46 (bt, 1H), 6.43 (bt, 1H), 4.38 (d, *J*= 4.3 Hz, 1H), 3.98 (d, *J*= 5.2 Hz, 1H), 3.83 (bt, 1H), 3.59-3.50 (m, 2H), 3.46 (d, *J*= 4.3Hz, 2H), 3.33 (q, *J*= 6.2 Hz, 2H), 2.43 (t, *J*= 5.6 Hz, 2H), 2.23 (td, *J*= 6.9 Hz, *J*= 2.5 Hz, 2H), 1.99 (t, *J*= 2.5 Hz, 1H), 1.71 (p, *J*= 6.7 Hz, 2H), 0.98 (s, 3H), 0.90 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 174.0, 171.8, 83.5, 77.7, 71.0, 69.5, 39.5, 38.9, 35.9, 35.4, 28.1, 21.6, 20.6, 16.3. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{14}H_{25}O_4N_2$, 285.1809, found 285.1805.

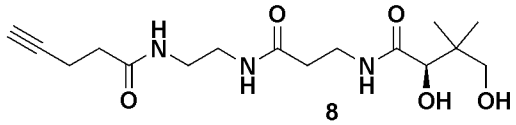
Pantetheine Analogue 7



7

^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.51 (bt, 1H), 6.53 (bt, 1H), 4.68 (d, *J*= 5.1 Hz, 1H), 4.10 (bt, 1H), 3.97 (d, *J*= 4.7 Hz, 1H), 3.58-3.49 (m, 2H), 3.49-3.40 (m, 2H), 3.21 (q, *J*= 6.4 Hz, 2H), 2.42 (t, *J*= 6.2 Hz, 2H), 2.19 (td, *J*= 6.5 Hz, *J*= 2.5 Hz, 2H), 1.96 (t, *J*= 2.3 Hz, 1H), 1.63-1.48 (m, 4H), 0.96 (s, 3H), 0.89 (s, 3H). ^{13}C-NMR (75.5 MHz, $CDCl_3$) δ 174.2, 171.7, 84.2, 77.5, 71.0, 69.1, 39.5, 39.3, 36.0, 35.5, 28.6, 25.8, 21.5, 20.6, 18.2. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{15}H_{26}O_4N_2$, 298.1887, found 298.1883.

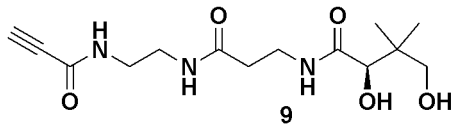
Pantetheine Analogue 8



8

^{1}H-NMR (400 MHz, CD_3OD) δ 3.88 (s, 1H), 3.52-3.41 (m, 2H), 3.41 (dd, *J*= 31.8 Hz, *J*= 10.6 Hz, 2H), 3.29-3.26 (m, 2H), 2.48-2.36 (m, 8H), 2.27 (t, *J*= 2.3 Hz, 1H), 0.91 (s, 6H). ^{13}C-NMR (75.5 MHz, CD_3OD) δ 174.9, 173.1, 172.9, 82.3, 76.1, 69.1, 39.1, 38.8, 38.7, 35.4, 35.2, 34.8, 20.1, 19.6, 14.4. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{16}H_{27}O_5N_3$, 341.1945, found 341.1943

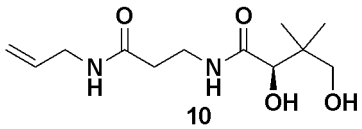
Pantetheine Analogue 9



9

^{1}H-NMR (400 MHz, (CD_3OD) δ 3.88 (s, 1H), 3.54-3.43 (m, 2H), 3.41 (dd, *J*= 31.3 Hz, *J*=10.6 Hz, 2H), 3.36-3.25 (m, 2H), 2.41 (t, *J*= 6.5 Hz, 2H), 0.91 (s, 6H). ^{13}C-NMR (75.5 MHz, (CD_3OD) δ 174.8, 173.0, 153.8, 77.0, 76.1, 74.9, 69.1, 39.1, 38.9, 38.5, 35.3, 35.2, 20.2, 19.7. HRMS (EI) (*m/z*): $[M+Na]^+$ calcd for $C_{14}H_{23}O_5N_3Na$, 336.1530, found 336.1529.

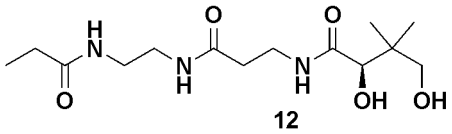
Pantetheine Analogue 10



10

H-NMR (400 MHz, $CDCl_3$) 7.39, (bs, 1H), 6.00, (bs, 1H), 5.78-5.46 (m, 1H), 5.18 (d, *J*=16 Hz, 1H), 5.13 (d, *J*=10.8 Hz, 1H), 4.00 (s, 1H), 3.87 (t, *J*=5.6Hz, 2H), 3.65-3.45 (m, 4H), 2.47 (t, *J*=6.5 Hz, 2H), 1.01 (s, 6H) C-NMR (100.6 MHz, $CDCl_3$) δ 174.5, 171.0, 134.1, 116.7, 77.3, 70.7, 42.2, 39.5, 35.9, 35.6, 21.2, 20.8.

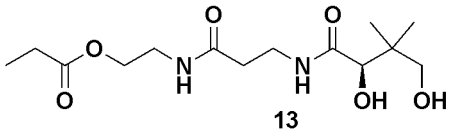
Pantetheine Analogue 12



12

^{1}H-NMR (400 MHz, CD_3OD) δ 3.89 (s, 1H), 3.47-3.27 (m, 8H), 2.41 (t, *J*=6.4 Hz, 2H), 2.20 (q, *J*=7.2 Hz, 2H), 1.12 (t, *J*=7.6 Hz, 3H), 0.91 (s, 6H). ^{13}C-NMR (100.6 MHz, CD_3OD) δ 176.2, 174.9, 173.0, 76.1, 69.1, 39.2, 38.9, 38.8, 35.4, 35.2, 29.0, 20.2, 19.7, 9.3. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{14}H_{27}N_3O_5$, 317.1945, found 317.1948.

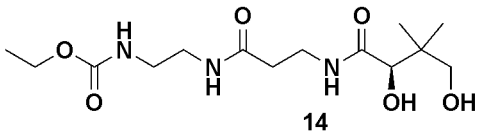
Pantetheine Analogue 13



13

^{1}H-NMR (400 MHz, CD_3OD) δ 4.12 (t, *J*=5.2 Hz, 2H), 3.88 (s, 1H), 3.50-3.31 (m, 6H), 2.43 (t, *J*=6.8 Hz, 2H), 2.35 (q, *J*=8.0 Hz, 2H), 1.11 (t, *J*=7.6 Hz, 3H), 0.91 (s, 6H). ^{13}C-NMR (100.6 MHz, CD_3OD) δ 174.9, 174.8, 172.8, 76.0, 69.2, 62.8, 39.2, 38.3, 35.2, 35.1, 27.0, 20.2, 19.7, 8.2. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{14}H_{26}N_2O_6$, 318.1785, found 318.1780.

Pantetheine Analogue 14



14

^{1}H-NMR (400 MHz, CD_3OD) δ 4.05 (q, *J*=7.2 Hz, 2H), 3.88 (s, 1H), 3.49-3.19 (m, 8H), 2.41 (t, *J*=6.4 Hz, 2H), 1.22 (t, *J*=7.2 Hz, 3H), 0.91 (s, 6H). ^{13}C-NMR

(100.6 MHz, CD_3OD) δ 174.9, 172.9, 158.4, 76.1, 69.1, 60.7, 39.9, 39.3, 39.2, 35.4, 35.25, 20.2, 19.7, 13.8.
HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{14}H_{27}N_3O_6$, 333.1894, found 333.1888.

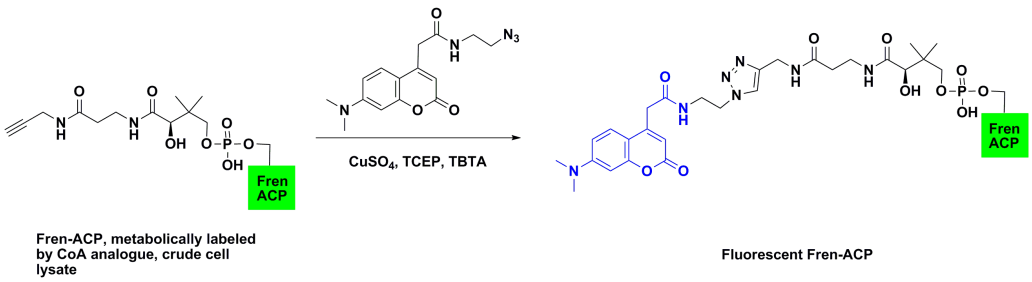
Assay Procedures

CoaA Kinetics: Kinetics with CoaA were collected using a continuous spectrophotometric assay as previously described.[1] Briefly, pantetheine analogues **1**-**15** were serially diluted into 96-well UV transparent plates (Grenier Bio) containing ATP (1.5 mM), $MgCl_2$ (10 mM), and KCl (20 mM) in 50 mM Tris-HCl buffer (pH 7.6). Phosphoenolpyruvate (0.5 mM), NADH (0.3 mM), pyruvate kinase (5 units), lactate dehydrogenase (5 units) were used in the reaction mixture to couple consumption of ATP to NADH oxidation. Reactions were started by addition of CoaA, and monitored at 340 nM for one hour. Initial velocities were calculated for each dilution point and plotted using KaleidaGraph software to generate Km and Vmax values. All measurements were obtained in triplicate.

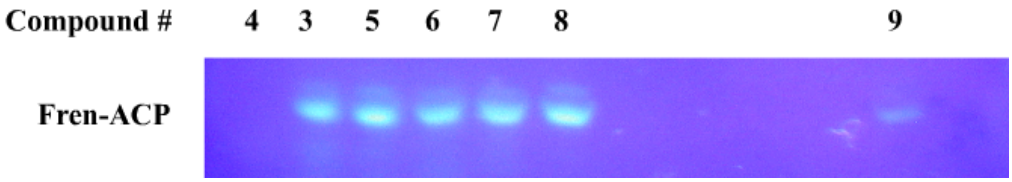
In Vivo CoA Biosynthetic Assay. The assay for conversion of pantetheine analogs to CoA analogs in vivo was preformed as described previously.[2] Briefly, *E. coli* BL21 cells were transformed with expression plasmids for the PPTase Sfp as well as the carrier protein Fren-ACP and grown in LB-Amp/Kan media. Cells were growth with shaking at 37 ^{o}C to an OD_{600} of 0.6 and then induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). At the same time 1 mM of analogue **1**-**15** (200 mM stocks in DMSO) was added to bring the total culture volume to ~ 2 mL. After incubation in a rotary shaker at 37 ^{o}C for five hours cells were pelleted, resuspended in 5 mL of lysis buffer (50 mM sodium phosphate pH 7.4, 300 mM NaCl), and washed three times in an equal volume of lysis buffer. Following this cells were resuspended in 0.3 mL lysis buffer and lysed by addition of 3 mg/mL lysozyme (Worthington) for 1 hr followed by sonication, treatment with DNAase I, and centrifugation. Lysate from cells was collected and analyzed by MALDI-TOF. For terminal alkynes **3**-**9**, in addition to MALDI-TOF analysis lysates were reacted with a fluorescent azide and analyzed by SDS-PAGE according to the procedure of Cravatt.[7]

Determination of Minimum Inhibitory Concentration (MIC). Pantetheine analogs were assayed for MIC following the procedure of Virga et al.[10] Briefly, *E. coli* K12 were grown overnight in either M9 or 1% tryptone media. Cultures were then diluted 1:10,000 in fresh media and added to a 96 well sterile tissue culture pate containing serial dilutions of the analogs in the same media. After incubation for 20 hours at 37 ^{o}C, plates were read at 590 nM to check for growth. *E. coli* grown with vehicle DMSO in place of pantetheine analogue were used as positive controls, and wells without *E. coli* were used as controls against contamination.

In Vivo Analysis of Analogues 3-9 by Fluorescent Bioconjugation

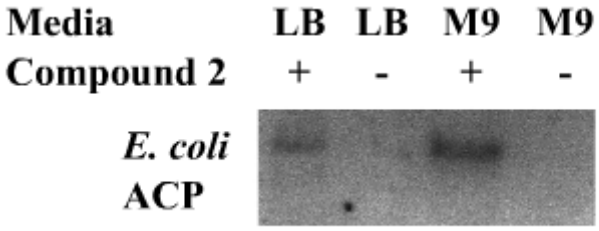


General scheme depicting analysis of in vivo labeling by terminal alkynes **3-9** using bioconjugation to fluorescent azide and SDS-PAGE.



Click reaction with fluorescent azide of cell lysate generated during in vivo analysis. Terminal alkynes **3-9** show labeling of overexpressed Fren-ACP, indicating they are being processed by the endogenous CoA biosynthetic pathway of *E. coli*. Compound 9 shows weaker labeling than the rest of the panel, consistent with the poor kinetics of this compound and low intensity signal observed for **9** by MALDI-TOF analysis.

In Vivo Labeling of Native *E. coli* ACP by Non-toxic Pantetheine Analogue 2



Administration of **2** (1 mM) to *E. coli* K12 grown in either nutrient rich (LB) or minimal (M9) media followed by cell lysis and Cu-catalyzed bioconjugation to a fluorescent alkyne results in fluorescent labeling of the fatty acid ACP (fluorescent gel pictured above). The identity of the labeled band was confirmed by gel-excision and tandem MS analysis. A complete analysis of these results will be the subject of a separate publication (reference 13 in text).

References:

1. Strauss E, Begley TP (2002) The antibiotic activity of N-pentylpantothenamide results from its conversion to ethyldethia-coenzyme a, a coenzyme a antimetabolite. J. Biol. Chem. 277: 48205-48209.

2. Meier JL, Mercer AC, Rivera H Jr, Burkart MD (2006) Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. J. Am. Chem. Soc. 128: 12174-12184.

3. Virga KG, Zhang YM, Leonardi R, Ivey RA, Hevener K, Park HW, Jackowski S, Rock CO, Lee RE (2006) Structure-activity relationships and enzyme inhibition of pantothenamide-type pantothenate kinase inhibitors. Bioorg. Med. Chem. 14:1007-1020.

4. Gottlieb HE, Kotlyar V, Nudelman A (1997) NMR Chemical Shifts of Common Laboratory Solvents As Trace Impurities. J. Org. Chem. 62: 7512-7515.

5. Still WC, Kahn A, Mitra A (1978) Rapid Chromatographic Technique For Preparative Separations With Moderate Resolution. J. Org. Chem. 43: 2923-2925.

6. Müller TE, Lercher JA, Nhu NV (2003) Hydroamination on homogeneous and heterogeneous catalysts: kinetic study. *AIChE*. 49: 214-224.

7. Speers AE, Cravatt BF (2004) Profiling enzyme activities in vivo using click chemistry methods. Chem. Biol. 11: 535-546.

Acknowledgments

The text of chapter 4, in full, is a reprint of material as it appears in: Mercer, A.C., Meier, J.L., Hur, G.H, Smith, A.R., and Burkart M.D. (2008). Antibiotic evaluation and in vivo analysis of alkynyl coenzyme A antimetabolites in Escherichia coli. *Bioorg. Med. Chem. Lett. 18*, *5991-5994*. Permission was obtained from all co-authors. I was the co-primary author of this work. Andrew Mercer performed biochemical assays and Gene Hur and Andrew Smith synthesized compounds **6-11**. I synthesized all other compounds, assisted in biological assays, and wrote the manuscript. This research was performed under the supervision of Prof. Michael Burkart.

Chapter Five

In vivo modification of native carrier protein domains

Originally published as:

Mercer, A. C.; Meier, J. L.; Torpey, J. W.; Burkart, M. D. *Chembiochem* **10**, 1091-100 (2009).

DOI: 10.1002/cbic.200800838

In Vivo Modification of Native Carrier Protein Domains

Andrew C. Mercer, Jordan L. Meier, Justin W. Torpey, and Michael D. Burkart*[a]

Carrier proteins are central to the biosynthesis of primary and secondary metabolites in all organisms. Here we describe metabolic labeling and manipulation of native acyl carrier proteins in both type I and II fatty acid synthases. By utilizing natural promiscuity in the CoA biosynthetic pathway in combination with synthetic pantetheine analogues, we demonstrate metabolic labeling of endogenous carrier proteins with reporter tags in Gram-positive and Gram-negative bacteria and in a human carcinoma cell line. The highly specific nature of the post-translational modification that was utilized for tagging allows for simple visualization of labeled carrier proteins, either by direct fluorescence imaging or after chemical conjugation to a fluorescent reporter. In addition, we demonstrate the utility of this approach for the isolation and enrichment of carrier proteins by affinity purification. Finally, we use these techniques to identify a carrier protein from an unsequenced organism, a finding that validates this proteomic approach to natural product biosynthetic enzyme discovery.

Introduction

The carrier protein (CP) is central to the biosynthetic pathways of many important primary and secondary metabolites.[1] These small proteins, found either as discrete polypeptides or as small domains within larger synthases, facilitate biosynthesis by acting as a scaffold for growing natural products. Due to their key role in the biosynthesis of fatty acids,[2] polyketides,[3] and nonribosomal peptides,[4] methods for the specific labeling of CPs in vivo would be useful for visualizing the localization and dynamics of these biosynthetic lynchpins, as well as to facilitate the proteomic identification of CPs from unsequenced natural-product-producing organisms.

In CP-mediated biosynthesis, the growing acyl chain is linked to the enzyme through a thioester bond to the terminal thiol of a coenzyme A (CoA)-derived 4′-phosphopantetheine (4′-PPant) prosthetic group. This post-translational modification is introduced by a phosphopantetheinyltransferase (PPTase, E.C. 2.7.7.7), which transfers the moiety from CoA to a conserved serine of the *apo*-CP.[5] The finding that many PPTases are capable of accepting functionalized CoA analogues as substrates offers an opportunity for selective labeling of CP domains with fluorescence and affinity reporters.[6,7] However, in order for this approach to succeed, reporter-labeled CoA analogues must be available to the PPTase, or else the PPTase will utilize endogenous CoA to produce *holo*-CPs, which are not easily visualized or detected. Due to the highly charged nature of CoA and its analogues, which precludes their use in intracellular-labeling approaches,[8] our previous work has focused on studying the ability of reporter-labeled CoA precursors to cross the cell membrane and be transformed into fully formed CoA analogues in vivo via the endogenous CoA biosynthetic pathway (Scheme 1).[9,10] Past studies have applied similar metabolic delivery strategies to the labeling of post-translationally lipidated and glycosylated proteins.[11–13] However, while our earlier studies were essential to the identification of potentially useful CoA precursors, this work was only explored in the context of a single Gram-negative bacterial organism, *Escherichia coli*, and only proved capable of labeling CPs when they were heterologously co-overexpressed with a promiscuous PPTase.

Here we report a significant advance in CP-labeling methodology by describing for the first time the labeling of endogenous CP domains in native bacterial systems by azide-labeled CoA precursor **1**. We demonstrate its utility in both Gram-positive and Gram-negative bacterium, as well as in a number of knockout strains, which probe its requirements for uptake and PPTase type. We also explore the activity of this analogue in a human carcinoma cell line, and perform an in-depth analysis of labeling of the human FAS in this cell line by **2**, a previously reported fluorescent CoA precursor. Finally, to evaluate the utility of this method for proteomic identification of CP-containing enzymes from unsequenced organisms, we demonstrate its application towards affinity purification of CPs, as well as for the de novo sequencing and genetic identification of a fatty acid biosynthetic CP from an unsequenced bacterium.

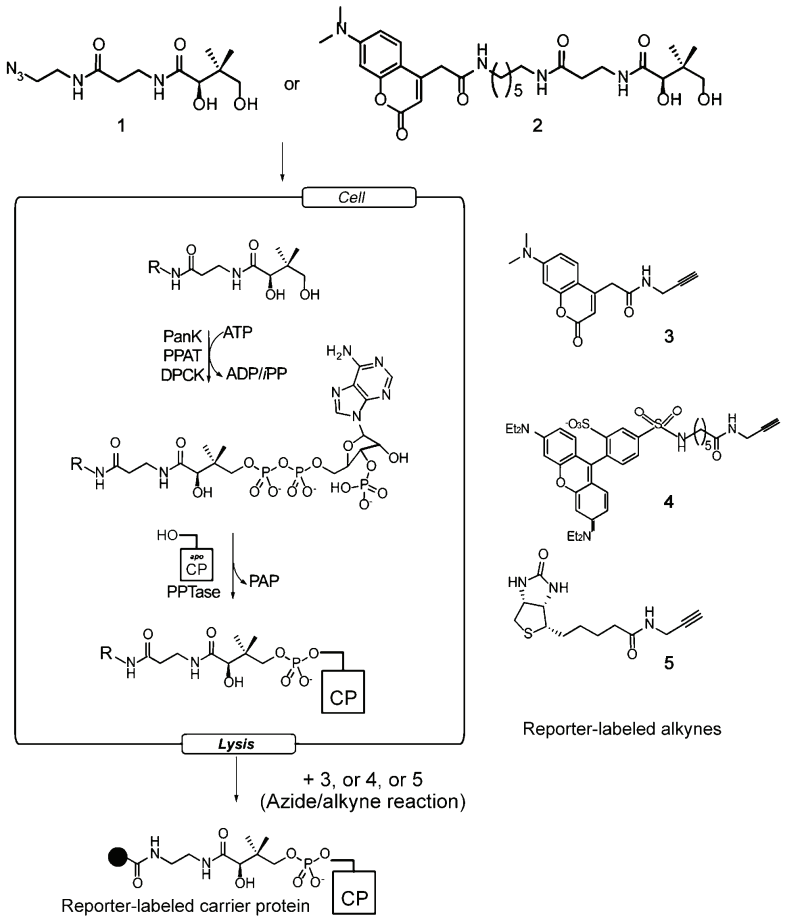
Results

In vivo labeling of native ACP

In our previous studies, pantetheine analogues **1** and **2** demonstrated cellular uptake and processing by the endogenous *E. coli* CoA biosynthetic enzymes pantothenate kinase (PanK, E.C. 2.7.1.33), phosphopantothenoyl adenyltransferase (PPAT, E.C. 2.7.7.3), and dephospho-CoA kinase (DPCK, E.C. 2.7.1.24) to form reporter-labeled CoA in vivo.[9,10] The formation of these

[a] *Dr. A. C. Mercer, J. L. Meier, Dr. J. W. Torpey, Prof. M. D. Burkart*
Department of Chemistry and Biochemistry
University of California, San Diego
9500 Gilman Drive, La Jolla, California (US)
Fax: (+1)
E-mail: mburkart@ucsd.edu

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.200800838.



Scheme 1. In vivo labeling strategy. Cells are grown in the presence of azido-pantetheine **1** or fluorescent pantetheine **2**. After uptake, the native CoA biosynthetic enzymes convert the pantetheine analogues to CoA analogues, which then are appended to endogenous CPs by the native PPTase enzyme. After cell lysis azido-modified CPs can be detected by reaction with alkyne probes **3**–**5**.

CoA analogues was demonstrated by their ability to undergo reaction with a heterologously expressed CP and PPTase, which resulted in labeling of the overexpressed CP with a fluorescence or affinity tag. While valuable for method development, the heterologous PPTase and CP were present within the cells at levels that far exceeded those of endogenous proteins. To explore the utility of this method for the metabolic labeling of endogenous CPs in a native organism, we directed our efforts at the fatty acid ACP of *E. coli*. Fatty acid and CoA metabolism are well understood in this organism, and our previous reports indicated that the CoA biosynthetic pathway in BL-21 strains of *E. coli* is permissive enough to allow for the in vivo conversion of pantetheine analogues **1** and **2** to CoA analogues.

By using the wild-type K12 strain of *E. coli* as a model system, we set out to demonstrate the in vivo modification of the native fatty acid CP AcpP. Overnight growth in a 1 mM solution of compound **1** in Luria-Bertani (LB) media followed by cell lysis and a copper-catalyzed cycloaddition reaction with fluorescent alkyne (**3** or **4**) allowed visualization of the modified CP with SDS-PAGE (Figure 1). Fluorescent pantetheine **2** showed no detectable labeling of native *E. coli* CPs under identical growth conditions. In addition, control cultures grown with vehicle (DMSO) showed no similar fluorescence after incubation with **3** or **4** under cycloaddition reaction conditions. The protein modified by **1** was found to run at the same apparent molecular weight as recombinantly purified and labeled AcpP on both denaturing and native PAGE gels, a compelling finding given that AcpP is readily separated from most other proteins on high-percentage native gels due to its small size and charge.[14] To definitively identify the labeled protein, the fluorescent band was excised, proteolytically digested, and analyzed by LC–MS/MS. The results confirmed that the labeled band was indeed the AcpP protein from *E. coli* (Figure S1 in the Supporting Information).

Because a traditional obstacle to the use of in vivo labeling techniques has been the ability of the small molecule to penetrate the cell, we also examined in detail the method of uptake of **1** by *E. coli*. While *E. coli* are capable of synthesizing pantothenate de novo, they also express a pantothenate transport system. The *panF* gene encodes pantothenate permease, a sodium-dependant pantothenate symporter.[15] If implicated in the uptake **1**, PanF ac-

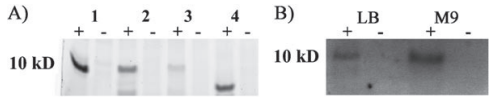


Figure 1. Detection of in vivo modified carrier proteins. A) Cultures were grown with or without compound **1** and reacted with RRX-alkyne **4** after lysis. Distinct CP labeling is seen in *E. coli* (1), *B subtilis* 168 (2), *B. subtilis* 6051 (3), and *S. oneidensis* (4) as compared to negative controls B) Labeling of *E. coli* ACP by **1** with visualization by DMAC-alkyne **3** was effective in both LB (lane 1) and M9 minimal media (lane 2) as compared to negative controls (lanes 2, 4).

tivity could represent a caveat to the general applicability of this labeling method to organisms that lack PanF homologues. To decouple this dedicated pantothenate-uptake system from the metabolic-labeling protocol, we performed metabolic labeling in *E. coli* strains that contain an in-frame deletion in the *panF* gene to knock-out activity.[16] The result of these labeling experiments were qualitatively indistinguishable from those performed with wild-type *E. coli* (Figure S6); this suggests that pantetheine analogues such as **1** enter *E. coli* by passive diffusion or an alternate uptake mechanism. Additionally, strains of *E. coli* with the pantothenate biosynthetic genes panD and panC knocked out showed no difference in CP labeling by **1** relative to wild-type *E. coli*; this indicates the presence or absence of active pantothenate biosynthesis has no effect on labeling efficiency.

Finally we examined the toxicity of metabolic labeling by **1** to growing *E. coli*. An ideal metabolic label would allow for the study of CPs in live organisms with minimal perturbation to their natural growth and metabolism. However, past studies have associated the in vivo modification of the fatty acid acyl carrier protein (ACP) by CoA analogues with growth inhibition in a number of bacteria, including *E. coli*.[17,18] To examine the toxicity of the metabolic-labeling procedure we determined minimum inhibitory concentrations for **1**, **2**, and N5-Pan, a prototypical bacteriostatic pantetheine analogue (see Figure S15 for structure), towards *E. coli* grown in nutrient-rich and minimal media. As opposed to N5-Pan, which inhibits *E. coli* growth in minimal media at 30 μM, the compounds used in this study (**1**, **2**) showed no growth inhibition of *E. coli* at concentrations up to 500 μM. This lack of growth inhibition was not due to a lack of ACP modification under these conditions as growth of *E. coli* supplemented with **1** in minimal media showed modification of ACP at similar levels to those observed in LB (Figure 1B). To reconcile the nontoxic ACP labeling of **1** with the proposed mode of action of bacteriostatic pantetheine analogues, we have since performed a full study of the effects that secondary structural characteristics have on the toxicity of CoA precursors.[19] These findings indicated that the more polar reporter labels incorporated in **1** and **2** might result in decreased toxicity relative to N5-Pan-modified ACP due to an inability of ACPs modified by **1** and **2** to interact with partner enzymes of the fatty acid biosynthetic pathway; this is an optimal property for non-toxic, in vivo labeling of CPs.

In vivo modification of ACPs from other bacteria

Having demonstrated in vivo CP modification in wild-type *E. coli*, we sought to next test the generality of this methodology for labeling of CP domains in other bacterial species whose CoA and fatty acid biosynthetic pathways have been less well studied (Figure 1). Accordingly *Bacillus subtilis* 6051 and *Shewanella oneidensis* MR-1, and *Bacillus brevis* 8246 were grown with 1 mM pantetheine azide **1** overnight. Lysis followed by chemical conjugation with fluorescent alkyne **4** allowed low-molecular-weight proteins corresponding to the fatty acid ACPs in each of these strains to be visualized by SDS-PAGE (Figures 1A and 4). Once again, this tentative identification was verified by band excision and LC–MS/MS analysis. Database searching allowed identification of the fatty acid ACPs from *B. subtilis* and *S. oneidensis*. This demonstrates that **1** possesses properties compatible with cellular uptake and processing by the CoA biosynthetic pathways in both Gram-positive and Gram-negative bacterial species, which together constitute a large number of known natural-product-producing organisms.

In addition to demonstrating metabolic labeling of the fatty acid ACPs from these organisms, we explored in further detail the biosynthetic capabilities necessary for transfer of the 4′-phosphopantetheine arm from CoA to the conserved serine residue of ACP by a PPTase (Scheme 1). There are two classes of bacterial PPTase enzymes: the AcpS type, which is associated with primary metabolism and necessary for modification of the fatty acid ACP, and the Sfp type, which is associated with secondary metabolism and is involved in the modification of CPs in NRPS and PKS systems.[6,20,21] While some cross-reactivity exists, AcpS-type PPTases are less permissive of variation in CP or CoA analogue identity.[22] Conversely, the Sfp type shows relaxed specificity in terms of CP or CoA identity.[6] We had previously observed CP labeling in *B. subtilis* 6051, which carries a functional copy of Sfp. To ensure that CP labeling by **1** would not be hindered at the last step in organisms lacking a Sfp-type PPTase we also tested this method in *B. subtilis* strain 168, which contains an in-frame deletion in its Sfp gene and therefore contains only the less-permissive AcpS-type PPTase. Metabolic labeling by **1** followed by chemoselective ligation to fluorescent alkyne **4** demonstrated labeling at similar levels to those observed by using strain 6051 (Figure 1A). This indicates the azido-CoA analogue formed from **1** is compatible with less-promiscuous AcpS-type PPTases; this is a desirable feature for the labeling of CPs from disparate biosynthetic systems.

Labeling of human type I FAS in living cells

Fatty acid, polyketide, and nonribosomal peptide synthases can be classified as type I or II depending on whether the active domains required for product biosynthesis are located on discrete polypeptides (type II) or are housed on multimodular megasynthases (type I).[1] Bacterial fatty acid synthases are type II and as such have discrete active domains, including the ACP. In animals, most fatty acids are biosynthesized by type I synthases. Our success in labeling ACPs of type II fatty acid biosynthesis in bacterial organisms with **1** lead us to test whether this approach would also be feasible in eukaryotic cells, particularly in the labeling of type I fatty acid ACP domains.

We chose as a model system the SKBR3 cell line, which is derived from human breast cancer cells.[23] SKBR3 cells were plated at 25% confluency and allowed to grow for 24 h at 37 °C before introduction of the pantetheine analogue. Incubation of the cells with compound **1** for up to 60 h did not result in detectable labeling of the FAS after cell lysis and reaction with fluorescent alkyne **4**. This was surprising, because we had expected labeling in eukaryotic cells to be enhanced compared

to bacteria because mammalian cells cannot produce endogenous pantothenate.[24] The hypothesis that this difference could be due to differential pantothenate import mechanisms in eukaryotic and prokaryotic cells led us to test another pantetheine analogue, fluorescent pantetheine **2**, which had shown good compatibility with the CoA biosynthetic pathway in other studies.[25,26] Interestingly, this compound, which had proven ineffective in modifying fatty acid ACPs from endogenous bacterial systems effectively labeled the type I FAS in human cells (Figure 2A). Growth of SKBR3 cells with **2** for between 40 and 60 h resulted in visibly bright staining of the cells, and an increased amount of observable cell death compared to incubation with **1** or DMSO. To quantify the level of uptake, cell cultures grown with **2** or vehicle DMSO were subjected to analysis by flow cytometry (FACS, Figure 2B). Of cells grown with **2**, 100% of the cells could be identified with FACS as containing significant fluorescence. Lysis of the cells and visualization of the fluorescent labeled FAS by using SDS-PAGE gel confirmed labeling of the FAS protein (Figure 2A). However, this labeling was relatively modest compared to expectations based on the qualitative observation of strong uptake. To reconcile these results, lysate from the cells was examined by HPLC along with chemoenzymatically prepared standards of the four possible intermediates in the processing of pantetheine **2** by the CoA biosynthetic pathway. The major fluorescent peak found in SKBR3 lysate was observed to co-elute with the product of **2** and PanK, leading us to identify it as the singly processed phosphopantetheine analogue. (Figure 2C) This is consistent with previous reports on the processing of pantetheine prodrugs by human cells that indicated the bifunctional PPAT/DPCK as the major point of blockage for formation of CoA analogues in vivo.[27] While we have previously shown that recombinant human PPAT/DPCK can be used to convert fluorescent pantetheine analogues into CoA analogues for CP labeling in vitro,[26] in living cells these enzymes are most likely present at far lower levels and only process the most efficient 4′-phosphopantetheine substrates. Additionally the human FAS is cytosolic, but CoA is known to be seques-

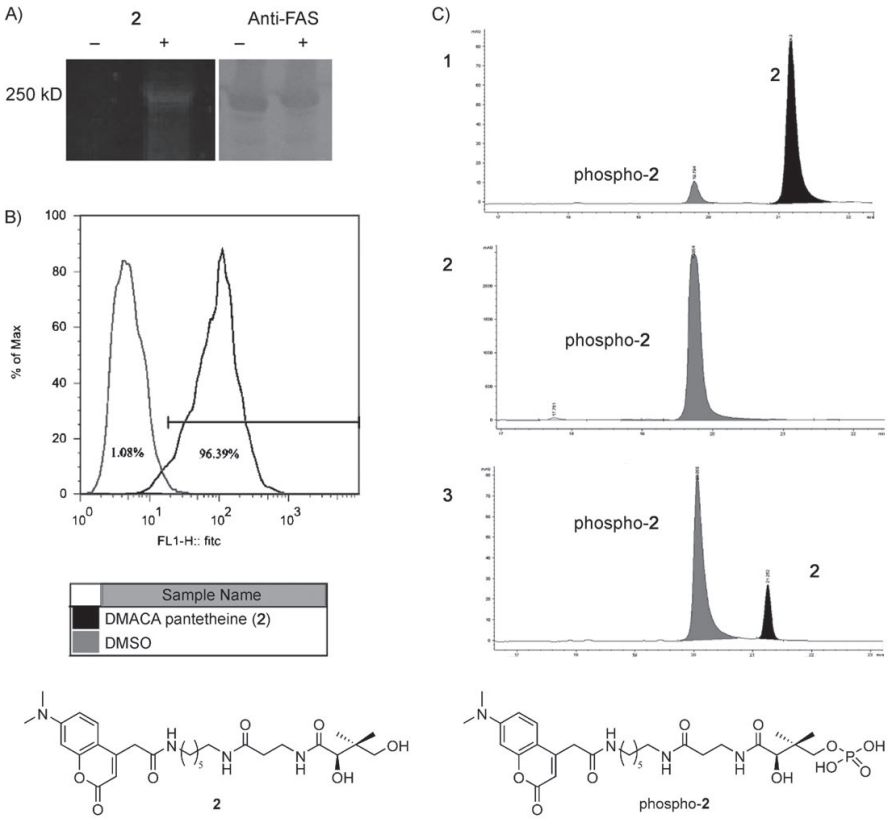


Figure 2. Analysis of type I FAS ACP labeling in SKBR3 cells A) In vivo modification of the human type I FAS. Lysate from cultures grown with compound **2** (+) show fluorescent modification of the FAS megasynthase as compared to negative controls (left). Blot with anti-FAS antibody confirms the presence and location of FAS on the gel (right). B) FACS analysis of SKBR3 cells grown with compound **2** shows 99% of the SKBR3 cells grown with **2** took up the compound. C) (panel 1) Compound **2** (black) incubated with PanK+ATP for 15 min results in addition of a phosphate on the 4′-hydroxyl and a shift to lower retention time (grey). (panel 2) After 120 min the reaction is complete, with all of **2** converted to the phospho analogue. (panel 3) Lysate from SKBR3 cells grown with **2** contains mostly phospho analogue. No further conversion to the CoA analogue in the SKBR3 cells could be detected (Supporting Information).

tered at its highest intracellular levels in the mitochondria; this makes localization of the CoA analogue of **2** another complicating factor.[28] Efforts are currently ongoing to develop an analogue that shows uptake similar to **2** and demonstrates improved compatibility with the human CoA biosynthetic enzymes.

Affinity purification of a metabolically labeled bacterial carrier protein

One of the major goals of this research has been to isolate and identify new modular biosynthetic enzymes with these techniques. The studies described above had shown that metabolic labeling in genetically unmodified cultures by **1** or **2** allowed for the visualization and identification of fatty acid ACPs from human, *E. coli*, *B. subtilis*, *S. oneidensis*, and *B. brevis* lysates. Here, labeled CPs from sequenced organisms were readily identified by proteolytic digest and MS analysis of excised gel bands, followed by comparison of the observed peptides to genomic databases. However, these analyses also commonly identified background proteins that migrated at the same molecular weight on SDS-PAGE. In unsequenced or unannotated organisms, this phenomenon could complicate de novo sequencing efforts and hinder identification of novel synthases by diluting the pool of labeled CP peptides or identifying false positives based on sequence homology. To remove contaminating proteins and also to verify the specificity of CP modification by metabolic labeling, we tested the utility of a biotin-streptavidin affinity purification method to isolate modified fatty acid ACPs.

This technique was demonstrated by using *S. oneidensis* MR-1, a bacterial strain noted for its production of polyunsaturated fatty acid natural products.[29] The finding that metabolic labeling by **1** is not toxic to growing bacteria suggested that it modifies only a fraction of the total cellular ACP content because it allows for at least a basal level of vital cellular processes such as fatty acid biosynthesis. Therefore, in order to facilitate the recovery of labeled CPs we performed the metabolic labeling on a large scale by using one-liter cultures. *S. oneidensis* MR-1 was grown overnight in the presence of 1 mM pantetheine analogue **1**, centrifuged to a pellet, washed, and lysed to yield azide-labeled CPs within the crude cell lysate. The lysate was then subjected to a copper-catalyzed [3+2] cycloaddition reaction with the biotin alkyne **5**. After removal of excess **5** by a desalting column, followed by incubation with streptavidin agarose beads, the affinity-purified proteins were recovered by the addition of SDS-PAGE running buffer and boiling. Figure 3 shows the results of enrichment of lysate from *S. oneidensis* that was metabolically labeled with **1** (lanes 1–6) in contrast to a DMSO-treated control (7–9). Comparison by SDS-PAGE shows enrichment of an approximately 10 kDa protein from the lysate of MR-1 grown with **1** and treated with biotin **5** (lane 3). This protein is not enriched when alkyne **5** is not added (lane 6), and is also not present in non-metabolically labeled, DMSO-treated cultures of *S. oneidensis* (lane 9). This indicates that the enriched band does not result from endogenous biotinylation or non-specific reaction

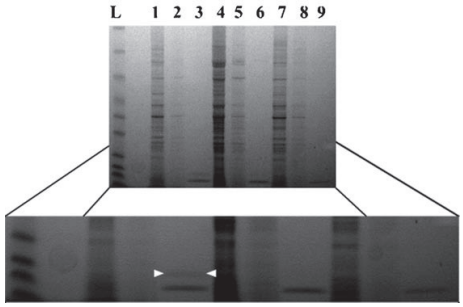


Figure 3. Affinity purification of ACP. Lysate from cultures of *S. oneidensis* MR1 that were grown with compound **1** and then treated with biotin-alkyne **5** were incubated with streptavidin agarose, which allowed for the purification of modified CP (crude lysate lane 1, desalting column flow lane 2, recovered ACP band—between white arrows—10 kDa lane 3). Cultures grown with compound **1** but not treated with biotin-alkyne **5** (lanes 4–6) or grown without compound **1** (lanes 7–9) showed no protein enrichment after incubation with streptavidin agarose. The strong band below ACP is a contaminant from the resin and was seen under all elution condtions.

of biotin-probe **5** in cell lysates. To our satisfaction, excision, tryptic digest, and MS analysis of this band resulted in the identification of the fatty acid ACP. This experiment shows that in vivo CP-labeling techniques can be used to specifically recover labeled CPs from their native systems.

Identification of a carrier protein from an unsequenced organism

B. brevis is a Gram-positive bacterium responsible for the production of the natural products gramicidin and tyrocidine.[30] We chose the nonproducing strain (ATCC 8246) of this unsequenced organism to test whether this chemical proteomic approach was capable of identifying CPs without the aid of genomic sequence data. As shown above with *S. oneidensis*, native CPs labeled by **1** and **2** can be reliably identified through database searches of known genomes after tryptic digest and MS. The ability to extend this methodology to the large number of unsequenced natural-product-producing organisms would further validate this approach for studying CP-based natural product biosynthetic enzymes.

Accordingly, *B. brevis* 8246 was grown overnight in the presence or absence of compound **1**. The cultures were lysed and prepared as in the other in vivo labeling experiments, followed by reaction with cognate fluorescent alkyne **4**. As can be seen in Figure 4, *B. brevis* 8246 cultures grown in the presence of **1** show distinct labeling of a protein with an apparent M_W of 15 kDa as compared to control cultures. Although this organism is unsequenced many other Bacillus species have been well studied, and these lead us to the expectation that this labeled protein was the type II fatty acid ACP, as we had seen in previous work with *B. subtilis*. To verify this hypothesis, we again excised the fluorescent band and submitted it to MS/MS analysis (Figure 4B). Because no database was available for this unsequenced organism, we constructed a data set comprising all known bacterial fatty acid ACPs against which to search.

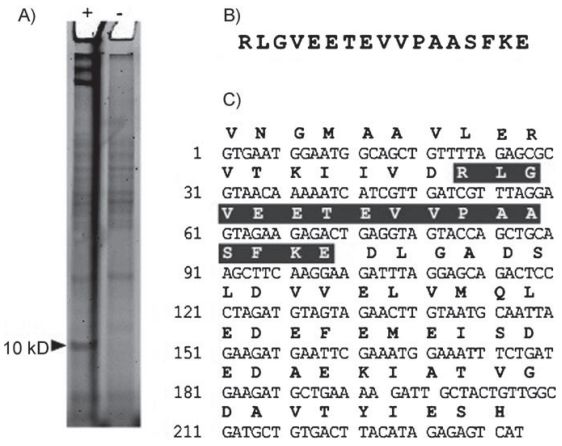


Figure 4. Identification of a carrier protein from an unsequenced organism. A) Metabolic labeling of *B. brevis* 8246 by **1** gives a band around 10 kDa that is not present in the negative control. B) Tryptic MS analysis of this band by using a database of 200 known bacterial ACP sequences identified a peptide fragment matching an ACP sequence. C) Degenerate and arbitrary PCR allowed for sequencing of the gene that encodes this peptide and its identification as the fatty acid ACP.

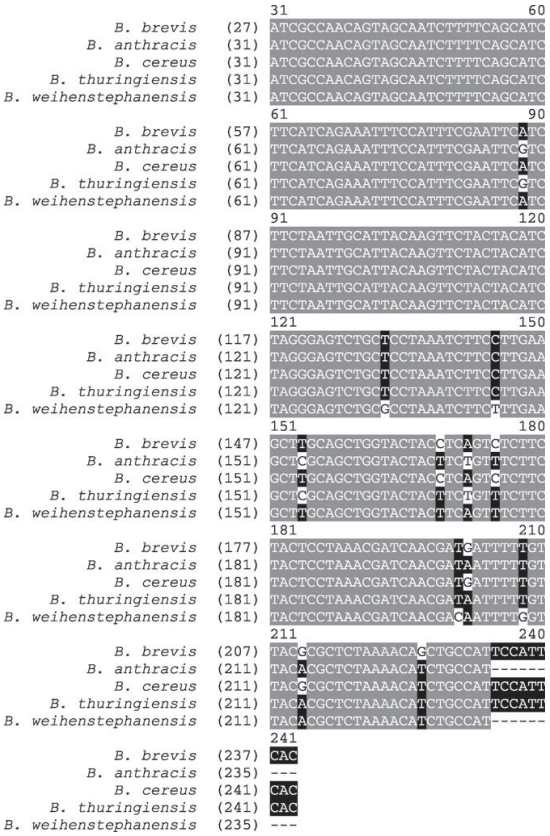


Figure 5. Alignment of *Bacillus* ACP sequences. Alignment of the new ACP sequence from *B. brevis* shows high homology with known *Bacillus* ACP sequences.

Searching the observed peptide fragments against this database returned a single hit that matched a tryptic fragment from the ACP in *B. cereus*. Because of the small size of type II fatty acid ACPs, only a small number of tryptic peptides are generated on proteolytic digest, and a single matching peptide represents a significant percentage of sequence coverage.

To complete the identification of the CP from *B. brevis* 8246 we constructed a set of degenerate primers from the identified peptide. By using degenerate and arbitrary PCR techniques, we were able to generate PCR fragments from genomic DNA.[31]

Sequencing confirmed that we had amplified the ACP gene from 8246. As seen in Figure 5, the sequence is nearly identical to the ACP gene from *B. cereus* and has high homology to the ACP genes from *B. thuringenes* and *B. anthrasis*. It differs considerably from the sequence for *B. subtilis* ACP. The taxonomy of *B. brevis* has been reclassified a number of times, but this result concurs with recent reports indicating *B. brevis* and *B. cereus* are closely related.[32] While this analysis relied on the relatively good sequence homology among small CP domains, the application of this method in combination with affinity tags targeting alternate active sites on PKS and NRPS multienzymes[33] should facilitate de novo sequencing of a variety of CP-containing systems from unsequenced organisms.

Discussion

Previous attempts to label CPs in vivo with functionally useful labels have been hindered by the inability of charged CoA analogues to permeate the cell membrane. In our prior studies we addressed this issue by developing uncharged pantetheine analogues and studying the ability of the native CoA biosynthetic pathway to convert them into reporter-labeled CoA analogues in living cells.[25,34] Although these investigations provided insight into the structural features necessary for in vivo production of CoA analogues from synthetic pantetheine analogues, all of these experiments were conducted with heterologously expressed CPs and PPTases in a single Gram-negative bacteria, *E. coli*. Here we have extended this technique to metabolically deliver CoA analogues to *apo*-CPs at native levels. By using this strategy, we have modified CPs from both Gram-positive and Gram-negative bacteria. The modified type II fatty acid ACPs in these organisms are detectable after cell lysis by bioconjugation of fluorescently tagged alkynes with azido-pantetheine **1**. After visualization by PAGE, these labeled proteins can be readily identified from LC–MS/MS analysis of excised gel fragments. Alternatively, cell lysate from organisms that were grown with **1** can be subjected to copper-catalyzed reaction with biotin-alkyne **5**, and their CPs can be isolated. The metabolic incorporation of **1** shows minimal toxicity and is compatible with CoA and PPTase pathways from a variety of different natural-product-producing organisms. This new ability to selectively label and affinity purify endogenous CPs from

native organisms should allow for the discovery of novel CPs as well as their associated biosynthetic systems.

We have also demonstrated the metabolic labeling of an ACP from a type I FAS. This indicates that our methodology is also applicable to type I modular biosynthetic systems and could provide a means for isolation and identification of the analogous type I PKS and NRPS megasynthases, which are responsible for the production of diverse natural products. Interestingly, whereas **1** was effective in modifying type II fatty acid ACPs in prokaryotes, it did not appear to be efficiently processed by the same pathway in the human SKBR3 cell line. Our previous work has demonstrated that turnover of an analogue in both the CoA biosynthetic pathway and the phosphopantetheinylation reaction can be greatly influenced by small changes in the structure of the particular analogue, and it has been observed previously that the human PanK isoform shows a very low sequence homology to those that are used by many prokaryotes.[35] One explanation for the extremely robust uptake of fluorescent pantetheine **2**, as was observed by FACS and HPLC analysis, could be that **2** possesses greater activity with the human PanK enzyme than **1**. Efficient processing of **2** by the PanK enzyme would lead to production of a phosphorylated, cell-impermeable intermediate. Because there are no currently annotated import or export mechanisms for pantetheine analogues in human cells, continued passive diffusion of the compound into the human cells could result in increasing accumulation of this cell-impermeable intermediate, a finding consistent with the increased cell death observed in cells grown with **2** for longer periods of time. With such a concentration mechanism in place, it seems credible that despite its low turnover by the PPAT/DPCK enzyme, a small amount of **2** is converted to CoA and shuttled to the ACP by the human PPTase; this leads to the observed modest labeling of human FAS. This preliminary observation of the differential activity of pantetheine analogues **1** and **2** in bacteria and humans could eventually prove useful for the specific tuning of the CP-labeling method to different organisms, or for the in vivo labeling of specific CPs in complex mixtures of organisms. With further refinement this technique may prove practical for the study of symbiotic natural-product-producing communities by allowing for time-dependent CP-labeling in an organism-specific fashion.[36]

In addition to possessing compatibility with several different biosynthetic enzymes, a useful metabolic label must also compete with the natural substrate without inducing cell death. In this study pantetheine analogues **1** and **2** have been shown to be compatible with three CoA biosynthetic enzymes and the CP post-translational modification process in a full spectrum of prokaryotic and eukaryotic organisms, including Gram-positive and Gram-negative prokaryotes and eukaryotes. These analogues show model properties in terms of their ability to successfully compete with the natural substrates for the CoA biosynthetic enzymes and PPTases without any genetically or media-based advantages. One disadvantage of this technique from the perspective of its use in proteomic identification of CPs is that this competition likely results in only intermediate to low-level labeling of CPs by our unnatural CoA precursors. However, metabolic labeling of CPs offers an advantage compared to genetic methods (RT-PCR) or post-lysis protein identification techniques in its ability to detect proteins throughout all stages of growth, rather than representing a single temporal snapshot of the dynamic proteome.[33] This characteristic should prove valuable to the study of the synthesis and degradation of biosynthetic enzymes.

Finally, the use of native metabolic pathways as a means of in vivo modification of CPs has enabled the identification of a previously unknown CP. Especially in systems with large or unsequenced genomes, a protein-based approach allows for a directed search for CP-dependent biosynthetic pathways without the use of laborious techniques such as genomic library construction and screening. No genetic information is required, and complications arising from silent biosynthetic gene clusters are removed. Furthermore, because this method depends only on the activity of a protein to be phosphopantetheinylated, it should hold no obvious bias against CP-containing systems that are not readily identifiable by homology alone, including those belonging to orphan gene clusters.[37] We have initially developed this methodology by using type I and II fatty acid biosynthetic CPs that are constitutively expressed in growing cultures. In these studies no effort was made to correlate polyketide or nonribosomal peptide production with CP labeling by **1** or **2**. However, the demonstrated compatibility of this method with uptake, CoA biosynthesis, and PPTase activity in a wide range of organisms should allow for its similar application to PKS and NRPS-producing organisms when correlated with small-molecule production or antibiotic activity. In the future, we believe that in vivo labeling of native CPs should prove highly complementary to existing methods and add greatly to the knowledge of natural product producing synthases.

Conclusions

A significant number of these biologically relevant natural products are produced in systems that tether the growing product to a 4′-phosphopantetheine modified CP. Selective labeling of these enzymes could allow for a better understanding of these proteins in their native environments and provide a platform for discovery of natural product synthases from unsequenced organisms. Here we have demonstrated a new set of tools for investigation and discovery of natural-product-producing systems. Using native metabolic pathways as a means of in vivo modification of endogenous CPs has allowed for the identification of a previously unknown CP. Because CPs and CP domains are found in fatty acid, polyketide, nonribosomal peptide, and other metabolite biosynthetic pathways, the ability to specifically modify these peptides in vivo allows one to probe numerous biochemical pathways. We have demonstrated for the first time the modification of native CPs by reporter-labeled CoA precursors in both prokaryotic and eukaryotic organisms; this gives access to the full range of CP-dependent systems.

Experimental Section

Experimental procedures

See Table 1.

Table 1. Strains and cell lines used in this study.

Organism	Medium	Ref.
E. coli K12	LB/M9	[38]
B. subtilis 168	LB	[39]
B. subtilis 6051	LB	ATCC 6051
B. brevis 26A1	RM	ATCC 8246
S. oneidensis MR-1	RM	[40]
human SKBR3	McCoy's 5A	[41]

Primers used in this study:

ARB1	5′-GGCCACGCGTCGACTAGTACNNNNNNNNNN-GATAT-3′
ARB2	5′-GGCCACGCGTCGACTAGTAC-3′
ARB6	5′-GGCCACGCGTCGACTAGTACNNNNNNNNN-NACGCC-3′
Internal frag	5′-CGCCTGGGCGTAGAAGAAACGGAA-3′
External frag	5′-CGCCTGGGCGTAGAAGAAACGGAA-3′
B. brevis acpp F	5′-ATGGAATGGCAGCTGTTTTAGAGCGCGT-3′
B. brevis acpp R	5′-TAGATGACTCTCTATGTAAGTCACAGCATCGC-3′

In primers ARB1 and ARB6 N is an equimolar mix of A/T/G/C bases.

Media: Luria Broth (LB) was purchased from Fisher Scientific (Waltham, MA). M9 was made according to the standard procedure.[42] Rich media (RM) contained 1% bactopeptone, 1% beef extract, 0.25% NaCl. McCoy's 5A media (Invitrogen) was modified to contain a final concentration of 10% fetal bovine serum. *E. coli* knockout strains were grown in LB.

In vivo carrier protein labeling: Bacteria were grown in the above-noted media with or without **1** (1 mM) for 12–17 h in an orbital shaker at 37°C. Cells were harvested by centrifugation at 20000 rpm for 5 min and resuspended in lysis buffer (100 mM NaCl, 25 mM potassium phosphate pH 7.0). For small cultures (<10 mL), cells were lysed by addition of lysozyme (3 mg mL^{-1}, Worthington Biochemical Corporation, Lakewood, NJ) incubated at 25°C for 1 h and sonicated (3×30 s with a microtip at low power). For large cultures (>10 mL) cells were resuspended in lysis buffer (10 mL per liter of culture) with lysozyme (0.1 mg mL^{-1}). After 1 h incubation with shaking on ice, cells were lysed by two passes through a French pressure cell. The lysate was then subjected to reaction with fluorescent alkyne **3** or **4** as previously reported.[34 43]

Determination of kinetic parameters of pantetheine analogues with PanK: Analogues were assayed for turnover with *E. coli* PanK in a coupled enzyme assay as previously described.[34]

Identification of *B. brevis* ACP: Cultures of *B. brevis* were grown in the presence of **1** as described above. MS identification of a labeled protein gave a single peptide that matched the *B. cereus* AcpP protein. By using this sequence, primers were designed (ARB1,2,6, internal frag, external frag), and arbitrary PCR was carried out by following the procedure of Caetano-Anolles.[31] Amplified bands were ligated into *Taq*-amplified (TA) vectors (Invitrogen) and sent for sequencing. The resulting sequence allowed for the design of specific primers and the cloning of the full acpp gene (*B. brevis* F and R). (Table 1).

Tissue culture: SKBR3 human breast cancer cells were grown in McCoy's 5A medium (Invitrogen) supplemented with FBS (10%, HyClone, Logan, UT, USA) and penicillin–streptomycin-glutamine (2%, Invitrogen). Cells were plated at ~25% confluency on day 0. On day one, the media was changed to a media that contained either fluorescent pantetheine **2** (1 mM), or DMSO (0.5%) as a control. On day three cells were trypsinized and washed twice with cold PBS. Cells were lysed by incubation in lysis buffer and Triton-X 100 (0.1%) on ice. Protein was recovered by centrifugation.

FACS analysis: 10^6 cells were trypsinized and washed once in cold PBS containing 1% FBS (HyClone). Cells were resuspended in PBS containing 1% FBS (1 mL) and analyzed by flow cytometry. Live cells, as determined by forward and side scatter profile, were analyzed by using a FACScalibur (Becton Dickinson) for uptake of compound **2**.

Large-scale in vivo carrier protein labeling and affinity purification: Lysate from MR-1 culture (1 L) grown with or without compound **1** (1 mM) was subjected to the bioconjugation reaction as previously reported with alkyne 5 as the conjugate probe.[34 43] Reactions (1 mL) were diluted to 3 mL in RIPA buffer (10 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5% deoxycholate, 0.1% SDS, 1% TritonX 100) and run over two desalting columns (Econo-Pac 10 DG, Bio-Rad) to remove unreacted **5**. The reaction was eluted in lysis buffer and incubated with streptavidin agarose resin (Thermo Sci, Waltham, MA, USA) for 2 h at room temperature. The resin was collected by centrifugation and washed in RIPA buffer (5×). The bound protein was then recovered from the resin by the addition of SDS-PAGE loading buffer and incubation at 95°C for 10 min. After brief centrifugation to remove the melted agarose, the supernatant was loaded onto a SDS-PAGE gel (12%) for analysis.

Mass spectroscopic analysis, 1D *SDS-PAGE*, and in-gel digestion: The proteins were separated by 1D SDS-PAGE by using Bis-Tris NuPAGE gels (10%, Invitrogen). The gel was fixed overnight in MeOH (50%)/H_2O (43%)/AcOH (7%), washed twice with H_2O (200 mL) for 10 min, and stained overnight with Gel Code Blue (Pierce, Rockford, IL, USA). Subsequently, the gel was washed with H_2O (200 mL) for 1 h. Excised gel bands were washed with 50% acetonitrile (MeCN) and 50% DTT (5 mM)/NH_4HCO_3 (25 mM) (2× 200 μL) with vortexing for 10 min, and finally washed with MeCN (200 μL). The dehydrated gel piece was rehydrated by addition of ice-cold trypsin (20 μL, 10 ng μL^{-1}, Promega) in DTT (5 mM)/ NH_4HCO_3 (25 mM), and incubated on ice for 30 min, and the remaining trypsin solution was removed and replaced with fresh DTT (5 mM)/ NH_4HCO_3 (25 mM). The digestion was allowed to continue at 37°C overnight. The peptide mixture was then acidified with trifluoroacetic acid (2%, 2 μL; TFA; Sigma), vortexed for 30 min, and the supernate was extracted. Finally, MeCN (20%)/TFA (0.1%) (20 μL) was added, and the mixture was vortexed to extract the remaining peptides and combined with the previous fraction. The combined extractions are analyzed directly by nanobore LC–MS/MS.

Mass spectrometry: The samples were analyzed by electrospray ionization by using a QSTAR-Elite hybrid mass spectrometer (Applied Biosystems/MDS Sciex) interfaced to a Tempo nanoscale reversed-phase HPLC (Eksigent/Applied Biosystems) by using a 75 μm×15 cm column (Grace Davison, Columbia, MD, USA) that was packed with Vydac MS C18 (300 A, 5 μm) packing material. The buffer compositions were as follows: buffer A: H_2O (98%), MeCN (2%), acetic acid (0.1%, Fluka), heptafluorobutyric acid (0.005%, Fluka); buffer B: MeCN (98%), H_2O (2%), acetic acid (0.1%), heptafluorobutyric acid (0.005%). Samples (10 μL) were injected by the Tempo autosampler onto a C18 PepMap pre-column

(5 mm × 300 μm, LC Packings, Sunnyvale, CA, USA) by using the channel 1 loading pump at a flow rate of 15 μL min^{-1} buffer A. After washing for 5 min, the peptides were transferred onto the analytical column and eluted directly into the mass spectrometer with a 25 min linear gradient from 5 to 40% buffer B at a flow rate of 300 nL min^{-1} by using channel 2. LC–MS/MS data were acquired in a data-dependent fashion by selecting the most intense peak with charge state 2–4 that exceeds 40 counts with exclusion of former target ions set to "always" and the mass tolerance for exclusion set to 100 ppm. TOF MS were acquired at *m/z* 500–1800 Da for 0.5 s with 20 time bins to sum. MS/MS are acquired from *m/z* 65–2000 Da by using "enhance all" and 20 time bins to sum, Dynamic Background Subtract, Automatic Collision Energy, and Automatic MS/MS Accumulation with the Fragment Intensity Multiplier set to 12 and Maximum Accumulation set to 3 s before returning to the survey scan.

Database search: The MS/MS were analyzed by Analyst 2.0 (Applied Biosystems) and subjected to database search by using Mascot 2.2.1 (Matrix Science, Boston, MA, USA) with Mascot Daemon 2.2 (Matrix Science) data import filter parameters set as follows: default precursor charge state 2–4; precursor and MS/MS data centroiding by using 50% height and 0.05 amu merge distances. MS/MS peaks with intensity less than 1% of the base peak were discarded, as were MS/MS spectra with fewer than 22 peaks remaining. Data were searched against the Swissprot database obtained at ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/ containing 237 168 sequences. The search identified tryptic peptides with up to two missed cleavages and used mass tolerances of 100 ppm (MS) and 0.10 Da (MS/MS), with variable modifications as follows: deamidation (NQ), oxidation (M), pyro-Glu (N-term Q). The search results indicated that individual ion scores >42 indicate identity or extensive homology ($P<0.05$).

Growth-inhibition assay: Assays were modified from reported procedure.[35] A single colony of K12 was picked into M9 media for overnight growth and the resulting culture was diluted 1:10 000. This diluted culture was plated into 96-well tissue culture plates that contained dilutions of pantetheine analogues in M9 medium. The plate was incubated overnight at 37 °C, and the wells were assessed for growth.

HPLC analysis of SKBR3 lysate: Cytosolic extract from the growth of SKBR3 cells with **2** was precipitated by trichloroacetic acid, centrifuged, and the supernatant was analyzed by reversed-phase HPLC by using a Burdick and Jackson OD5 column (4.6 mm by 25 cm, Morristown, NJ, USA) with monitoring at 254 and 360 nm. The solvents that were used were TFA (0.05%) in H_2O (solvent A) and TFA (0.05%) in MeCN (solvent B). Compounds were eluted at a flow rate of 1 mL min^{-1}. The method used an isocratic step from 0 to 5 min with 100% A, followed by a linear gradient to 45% B over 15 min, followed by an increasing gradient with solution B until at 25 min the solvent composition was 100% solution B. Under these conditions, **2** and its CoA analogue intermediates eluted between 17 and 21 min. For comparison of the fluorescent compounds observed in SKBR3 lysate to authentic CoA analogue intermediates of **2**, **2** was incubated in a stepwise manner with the recombinant *E. coli* biosynthetic enzymes PanK, PPAT, and DPCK as previously described.[25] The product of the reaction of **2** and PanK was combined with a small amount of SKBR3 lysate to verify co-elution of the two peaks.

General synthetic procedures and materials: All commercial reagents (Sigma–Aldrich, Alfa Aesar, TCI America) were used as provided unless otherwise indicated. The 5-isomer of Rhodamine-Red X NHS Ester was purchased from Molecular Probes (Carlsbad, CA, USA). Azido-pantetheine **1**, fluorescent pantetheine **2**, biotin alkyne **5**, and 7-dimethylaminocoumarin-4-acetic acid were prepared as previously described.[25 34] All reactions were carried out under an argon atmosphere in dry solvents with oven-dried glassware and constant magnetic stirring unless otherwise noted. Triethylamine (TEA), and ethyl-*N*,*N*-diisopropylamine (DIPEA) were dried over Na and freshly distilled. ^{1}H NMR spectra were taken at 300, 400, or 500 MHz and ^{13}C NMR spectra were taken at 100.6 or 75.5 MHz on Varian NMR spectrometers and standardized to the solvent signal as reported by Gottlieb.[44] TLC analysis was performed by using silica gel 60 F254 plates (EM Scientific) and visualization was accomplished with UV light ($\lambda=254$ nm) and/or the appropriate stain (iodine, 2,4-dinitrophenylhydrazine, cerium molybdate, ninhydrin). Silica gel chromatography was carried out with Silicycle 60 Å 230–400 mesh according to the method of Still.[45] TLC prep plate purification was performed with EMD silica gel 60 F_{254} precoated plates. Electrospray (ESI) and fast atom bombardment (FAB) mass spectra were obtained at the UCSD Mass Spectrometry Facility by Dr. Yongxuan Su by using a Finnigan LCQDECA mass spectrometer and a ThermoFinnigan MAT- 900XL mass spectrometer, respectively.

Synthesis of fluorescent 2-(7-(dimethylamino)-2-oxo-2*H*-chromen-4-yl)-*N*-(prop-2-ynyl)acetamide (3): 7-dimethylaminocoumarin-4-acetic acid (250 mg, 1.01 mmol), DIPEA (440 μL, 2.53 mmol), propargyl amine (70 μL, 1.01 mmol), and 1-hydroxybenzotriazole (484 mg, 3.16 mmol) were dissolved in DMF (25 mL) with stirring and cooled to 0 °C. *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide (485 mg, 2.53 mmol) was added in one portion, and the reaction was allowed to slowly warm to RT and followed by TLC. After 20 h the solvent was removed under reduced pressure and the sparingly soluble crude reaction mixture was taken up in EtOAc with heating. After filtration the soluble portion was purified by column chromatography (1:1 EtOAc/hexanes to EtOAc to 5% MeOH/CH_2Cl_2). This procedure was repeated several times on the EtOAc-resuspended filter cake to afford propargyl-DMC **3** (87 mg, 30%) as an orange solid. ^{1}H NMR (500 MHz, $CDCl_3$): $\delta=7.43$ (d, $J=$ 9.0 Hz, 1H), 6.61 (dd, $J=2$, 9.0 Hz, 1H), 6.51 (d, $J=2$ Hz, 1H), 6.05 (s, 1H), 5.92 (br s, 1H), 4.02 (dd, $J=2.0$, 4.5 Hz, 2H), 3.64 (s, 2H), 3.06 (m, 7H); ^{13}C NMR (75.5 MHz, [D$_6$]DMSO): $\delta=168.4$, 161.4, 156.1, 153.5, 151.7, 126.7, 110.1, 109.7, 108.8, 98.2, 81.5, 74.0, 41.0, 39.1, 28.9; HRMS (EI): *m/z* calcd for $C_{16}H_{16}N_2O_3$: 284.1155 $[M]^+$; found: 284.1154.

Synthesis of fluorescent alkyne 2-(6-(diethylamino)-3-(diethyliminio)-3*H*-xanthen-9-yl)-5-(*N*-(6-oxo-6-(prop-2-ynylamino)hexyl)sulfamoyl)benzenesulfonate (4): Rhodamine Red-X NHS Ester (10 mg, 0.013 mmol), DIPEA (10 μL, 0.06 mmol), and propargyl amine (5 μL, 0.07 mmol), were dissolved in DMF (0.5 mL), covered with aluminum foil, and allowed to stir for 1 h. After removal of the solvent under reduced pressure, the bright-red residue was redissolved in MeCN (1.4 mL) and purified by RP-HPLC. The solvents used were TFA (0.05%) in H_2O (Solvent A) and TFA (0.05%) in MeCN (solvent B). Compounds were eluted at a flow rate of 1 mL min^{-1} with monitoring at 560 nm. The method used an isocratic step from 0 to 5 min with 40% B, followed by a linear gradient to 58% B over 15 min, followed by an increasing gradient with solution B until at 16 min the solvent composition was 100% solution B. Under these conditions **4** eluted around 11.3 min. Pooling and lyophilization of multiple HPLC runs yielded **4** (6.5 mg, 72%) as a red solid. ^{1}H NMR (500 MHz, [D$_6$]DMSO): $\delta=8.39$ (s, 1H), 8.20 (br t, $J=5.0$ Hz, 1H), 7.91, (d, $J=8.0$ Hz, 1H), 7.90 (br s, 1H), 7.45 (d, $J=8.0$ Hz, 1H), 7.03 (d, $J=9.5$ Hz, 2H), 6.94 (d, $J=9.5$ Hz, 2H), 6.91

(s, 2H), 3.79 (d, J=2.5 Hz, 2H), 3.62 (q, J=7.5 Hz, 8H), 3.01 (s, 1H), 2.83 (q, J=6.5 Hz, 2H), 2.04 (t, J=7.5 Hz, 2H), 1.42 (p, J=8.0 Hz, 2H), 1.37 (p, J=7.5 Hz, 2H), 1.19 (t, J=7.0 Hz, 12H), 1.15 (t, J=7.0 Hz, 2H). MS (ESI): m/z calcd for $C_{36}H_{44}N_4O_7S_2$: 708.26 $[M]^+$; found: 731.25 $[M+Na]^+$.

Acknowledgements

The authors thank Elinore Mercer for assistance with FACS analysis. This work was funded by NIH R01M075797 and NSF CAREER MCB-0347681.

Keywords: carrier protein · fatty acids · natural products · protein labeling · proteomic

[1] A. C. Mercer, M. D. Burkart, *Nat. Prod. Rep.* **2007**, *24*, 750.
[2] S. W. White, J. Zheng, Y. M. Zhang, C. O. Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791.
[3] A. M. Hill, *Nat. Prod. Rep.* **2006**, *23*, 256.
[4] D. Schwarzer, R. Finking, M. A. Marahiel, *Nat. Prod. Rep.* **2003**, *20*, 275.
[5] R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem. Biol.* **1996**, *3*, 923.
[6] L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585.
[7] J. J. La Clair, T. L. Foley, T. R. Schegg, C. M. Regan, M. D. Burkart, *Chem. Biol.* **2004**, *11*, 195.
[8] B. H. Meyer, J. M. Segura, K. L. Martinez, R. Hovius, N. George, K. Johnsson, H. Vogel, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2138.
[9] K. M. Clarke, A. C. Mercer, J. J. LaClair, M. D. Burkart, *J. Am. Chem. Soc.* **2005**, *127*, 11234.
[10] J. L. Meier, A. C. Mercer, H. Rivera, M. D. Burkart, *J. Am. Chem. Soc.* **2006**, *128*, 12174.
[11] S. T. Laughlin, N. J. Agard, J. M. Baskin, I. S. Carrico, P. V. Chang, A. S. Ganguli, M. J. Hangauer, A. Lo, J. A. Prescher, C. R. Bertozzi, *Methods Enzymol.* **2006**, *415*, 230.
[12] H. C. Hang, E. J. Geutjes, G. Grotenbreg, A. M. Pollington, M. J. Bijlmakers, H. L. Ploegh, *J. Am. Chem. Soc.* **2007**, *129*, 2744.
[13] T. L. Foley, M. D. Burkart, *Curr. Opin. Chem. Biol.* **2007**, *11*, 12.
[14] S. Jackowski, C. O. Rock, *J. Biol. Chem.* **1984**, *259*, 1891.
[15] S. Jackowski, J. H. Alix, *J. Bacteriol.* **1990**, *172*, 3842.
[16] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, *Mol. Syst. Biol.* **2006**, *2*.
[17] Y.-M. Zhang, M. W. Frank, K. G. Virga, R. E. Lee, C. O. Rock, S. Jackowski, *J. Biol. Chem.* **2004**, *279*, 50969.
[18] E. Strauss, T. P. Begley, *J. Biol. Chem.* **2002**, *277*, 48205.
[19] A. C. Mercer, J. L. Meier, G. H. Hur, A. R. Smith, M. D. Burkart, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5991.
[20] R. H. Lambalot, C. T. Walsh, *J. Biol. Chem.* **1995**, *270*, 24658.
[21] C. G. Marshall, M. D. Burkart, R. K. Meray, C. T. Walsh, *Biochemistry* **2002**, *41*, 8429.
[22] A. C. Mercer, J. J. La Clair, M. D. Burkart, *ChemBioChem* **2005**, *6*, 1335.
[23] B. J. Thompson, A. Stern, S. Smith, *Biochim. Biophys. Acta Enzymol.* **1981**, *662*, 125.
[24] Y. Hwang, S. Ganguly, A. K. Ho, D. C. Klein, P. A. Cole, *Bioorg. Med. Chem.* **2007**, *15*, 2147.
[25] K. M. Clarke, A. C. Mercer, J. J. LaClair, M. D. Burkart, *J. Am. Chem. Soc.* **2005**, *127*, 11234.
[26] A. S. Worthington, M. D. Burkart, *Org. Biomol. Chem.* **2006**, *4*, 44.
[27] M. Cebrat, C. M. Kim, P. R. Thompson, M. Daugherty, P. A. Cole, *Bioorg. Med. Chem.* **2003**, *11*, 3307.
[28] R. Leonardi, Y. M. Zhang, C. O. Rock, S. Jackowski, *Prog. Lipid Res.* **2005**, *44*, 125.
[29] D. R. Lovley, D. E. Holmes, K. P. Nevin in *Advances in Microbial Physiology, Vol. 49*, **2004**, p. 219.
[30] K. Okuda, G. C. Edwards, T. Winnick, *J. Bacteriol.* **1963**, *85*, 329.
[31] G. Caetanoanolles, *Pcr-Methods and Applications* **1993**, *3*, 85.
[32] D. Xu, J.-C. Cote, *Int. J. Syst. Evol. Microbiol.* **2003**, 53, 695.
[33] J. L. Meier, A. C. Mercer, M. D. Burkart, *J. Am. Chem. Soc.* **2008**, *130*, 5443.
[34] J. L. Meier, A. C. Mercer, H. Rivera, M. D. Burkart, *J. Am. Chem. Soc.* **2006**, *128*, 12174.
[35] K. G. Virga, Y.-M. Zhang, R. Leonardi, R. A. Ivey, K. Hevener, H.-W. Park, S. Jackowski, C. O. Rock, R. E. Lee, *Bioorg. Med. Chem.* **2006**, *14*, 1007.
[36] S. Angell, B. J. Bench, H. Williams, C. M. H. Watanabe, *Chem. Biol.* **2006**, *13*, 1349.
[37] H. Gross, *Appl. Microbiol. Biotechnol.* **2007**, *75*, 267.
[38] F. R. Blattner, G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, Y. Shao, *Science* **1997**, *277*, 1453.
[39] H. D. Mootz, R. Finking, M. A. Marahiel, *J. Biol. Chem.* **2001**, *276*, 37289.
[40] O. Bretschger, A. Obraztsova, C. A. Sturm, I. S. Chang, Y. A. Gorby, S. B. Reed, D. E. Culley, C. L. Reardon, S. Barua, M. F. Romine, J. Zhou, A. S. Beliaev, R. Bouhenni, D. Saffarini, F. Mansfeld, B.-H. Kim, J. K. Fredrickson, K. H. Nealson, *Appl. Environ. Microbiol.* **2007**, *73*, 7003.
[41] B. J. Thompson, S. Smith, *Biochim. Biophys. Acta Lipids Lipid Metab.* **1982**, *712*, 217.
[42] O. Paliy, T. S. Gunasekera, *Appl. Microbiol. Biotechnol.* **2007**, *73*, 1169.
[43] E. Weerapana, A. E. Speers, B. F. Cravatt, *Nat. Protocols* **2007**, *2*, 1414.
[44] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512.
[45] W. C. Still, M. Kahn, A. Mitra, *J. Org. Chem.* **1978**, *43*, 2923.

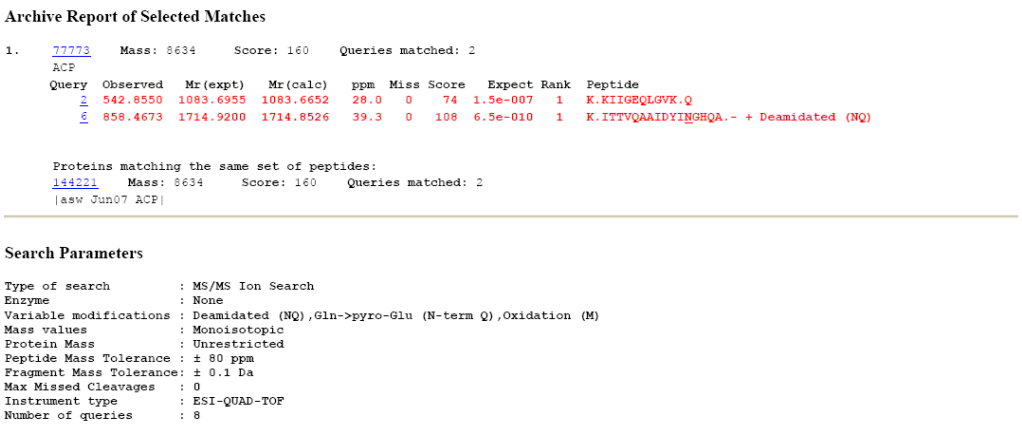
Received: December 20, 2008
Published online on March 23, 2009

CHEMBIOCHEM

Supporting Information

Mascot Reports from MS/MS Analysis

Figure S1. Mascot search report of labeled band from *E. coli* K12 lysate.



Archive Report of Selected Matches

1. 77773 Mass: 8634 Score: 160 Queries matched: 2
ACP

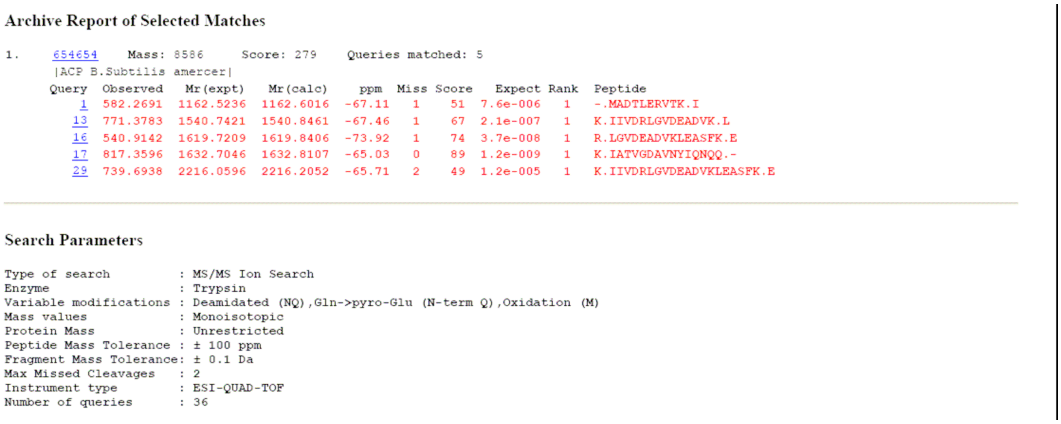
Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
2	542.8550	1083.6955	1083.6652	28.0	0	74	1.5e-007	1	K.KIIGEQLGVK.Q
6	858.4673	1714.9200	1714.8526	39.3	0	108	6.5e-010	1	K.ITTVQAAIDYINGHQA.- + Deamidated (NQ)

Proteins matching the same set of peptides:
144221 Mass: 8634 Score: 160 Queries matched: 2
|asw Jun07 ACP|

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : None
Variable modifications : Deamidated (NQ),Gln->pyro-Glu (N-term Q),Oxidation (M)
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 80 ppm
Fragment Mass Tolerance: ± 0.1 Da
Max Missed Cleavages   : 0
Instrument type        : ESI-QUAD-TOF
Number of queries      : 8
```

Figure S2 Mascot search results for labeled band from *B. subtilis* 6051 lysate



Archive Report of Selected Matches

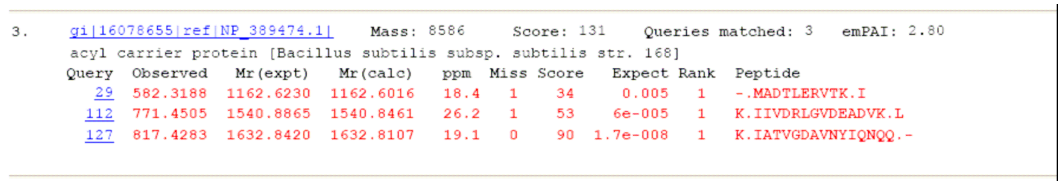
1. 654654 Mass: 8586 Score: 279 Queries matched: 5
|ACP B.Subtilis amercer|

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
1	582.2691	1162.5236	1162.6016	-67.11	1	51	7.6e-006	1	-.MADTLERVTK.I
13	771.3783	1540.7421	1540.8461	-67.46	1	67	2.1e-007	1	K.IIVDRLGVDEADVK.L
16	540.9142	1619.7209	1619.8406	-73.92	1	74	3.7e-008	1	R.LGVDEADVKLEASFK.E
17	817.3596	1632.7046	1632.8107	-65.03	0	89	1.2e-009	1	K.IATVGDAVNYIQNQQ.-
29	739.6938	2216.0596	2216.2052	-65.71	2	49	1.2e-005	1	K.IIVDRLGVDEADVKLEASFK.E

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : Trypsin
Variable modifications : Deamidated (NQ),Gln->pyro-Glu (N-term Q),Oxidation (M)
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance: ± 0.1 Da
Max Missed Cleavages   : 2
Instrument type        : ESI-QUAD-TOF
Number of queries      : 36
```

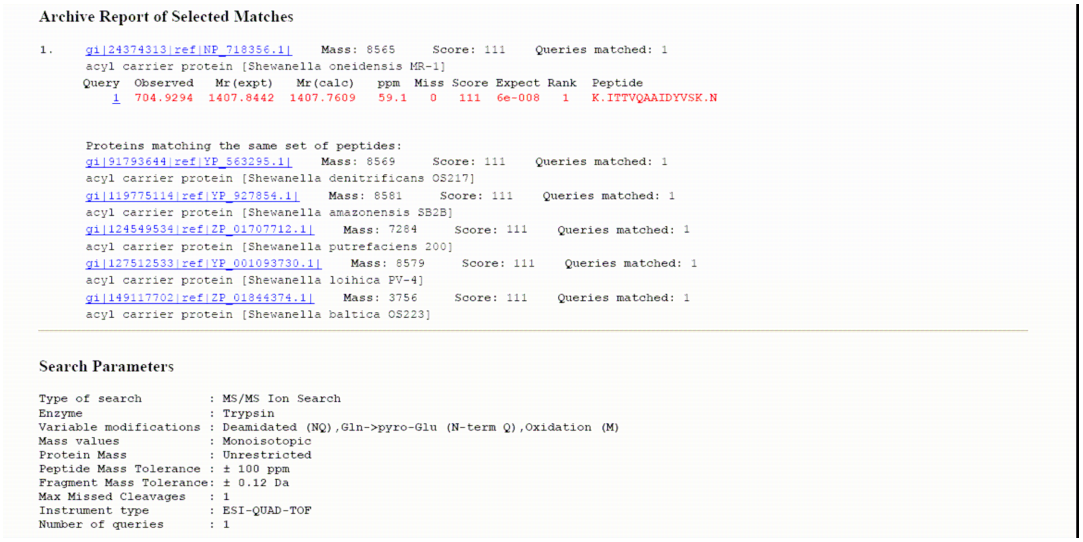
Figure S3 Mascot search results for labeled band from *B. subtilis* 168 lysate



3. gi|16078655|ref|NP_389474.1| Mass: 8586 Score: 131 Queries matched: 3 emPAI: 2.80
acyl carrier protein [Bacillus subtilis subsp. subtilis str. 168]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
29	582.3188	1162.6230	1162.6016	18.4	1	34	0.005	1	-.MADTLERVTK.I
112	771.4505	1540.8865	1540.8461	26.2	1	53	6e-005	1	K.IIVDRLGVDEADVK.L
127	817.4283	1632.8420	1632.8107	19.1	0	90	1.7e-008	1	K.IATVGDAVNYIQNQQ.-

Figure S3 Mascot search results for labeled band from *S. oneidensis* MR-1



Archive Report of Selected Matches

1. gi|24374313|ref|NP_718356.1| Mass: 8565 Score: 111 Queries matched: 1
acyl carrier protein [Shewanella oneidensis MR-1]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
1	704.9294	1407.8442	1407.7609	59.1	0	111	6e-008	1	K.ITTVQAAIDYVSK.N

Proteins matching the same set of peptides:
gi|91793644|ref|YP_563295.1| Mass: 8569 Score: 111 Queries matched: 1
acyl carrier protein [Shewanella denitrificans OS217]
gi|119775114|ref|YP_927854.1| Mass: 8581 Score: 111 Queries matched: 1
acyl carrier protein [Shewanella amazonensis SB2B]
gi|124549534|ref|ZP_01707712.1| Mass: 7284 Score: 111 Queries matched: 1
acyl carrier protein [Shewanella putrefaciens 200]
gi|127512533|ref|YP_001093730.1| Mass: 8579 Score: 111 Queries matched: 1
acyl carrier protein [Shewanella loihica PV-4]
gi|149117702|ref|ZP_01844374.1| Mass: 3756 Score: 111 Queries matched: 1
acyl carrier protein [Shewanella baltica OS223]

Search Parameters

Type of search : MS/MS Ion Search
Enzyme : Trypsin
Variable modifications : Deamidated (NQ),Gln->pyro-Glu (N-term Q),Oxidation (M)
Mass values : Monoisotopic
Protein Mass : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance: ± 0.12 Da
Max Missed Cleavages : 1
Instrument type : ESI-QUAD-TOF
Number of queries : 1

Figure S4 Mascot search result for labeled band in *B. brevis* 26A1 lysate.

Archive Report of Selected Matches

2. gi|30021938|ref|NP_833569| Mass: 8809 Score: 83 Queries matched: 1
acyl carrier protein

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
18	788.4492	1574.8838	1574.8192	41.0	0	83	2.1e-007	1	R.LGVEETEVVPAASFK.E

Proteins matching the same set of peptides:
gi|30263853|ref|NP_846230| Mass: 8507 Score: 83 Queries matched: 1
acyl carrier protein
gi|42782942|ref|NP_980189| Mass: 8507 Score: 83 Queries matched: 1
acyl carrier protein
gi|47529279|ref|YP_020628| Mass: 8507 Score: 83 Queries matched: 1
acyl carrier protein
gi|49186700|ref|YP_029952| Mass: 8507 Score: 83 Queries matched: 1
acyl carrier protein
gi|49479144|ref|YP_037911| Mass: 8507 Score: 83 Queries matched: 1
acyl carrier protein
gi|52141635|ref|YP_085191| Mass: 8507 Score: 83 Queries matched: 1
acyl carrier protein
gi|118479072|ref|YP_896223| Mass: 8809 Score: 83 Queries matched: 1
acyl carrier protein

Search Parameters

Type of search : MS/MS Ion Search
Enzyme : None
Variable modifications : Deamidated (NQ),Gln->pyro-Glu (N-term Q),p-pantetheine (S)
Mass values : Monoisotopic
Protein Mass : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance: ± 0.1 Da
Max Missed Cleavages : 0
Instrument type : ESI-QUAD-TOF
Number of queries : 26

Figure S5 Mascot search results for labeled band in SKBR3 lysate.

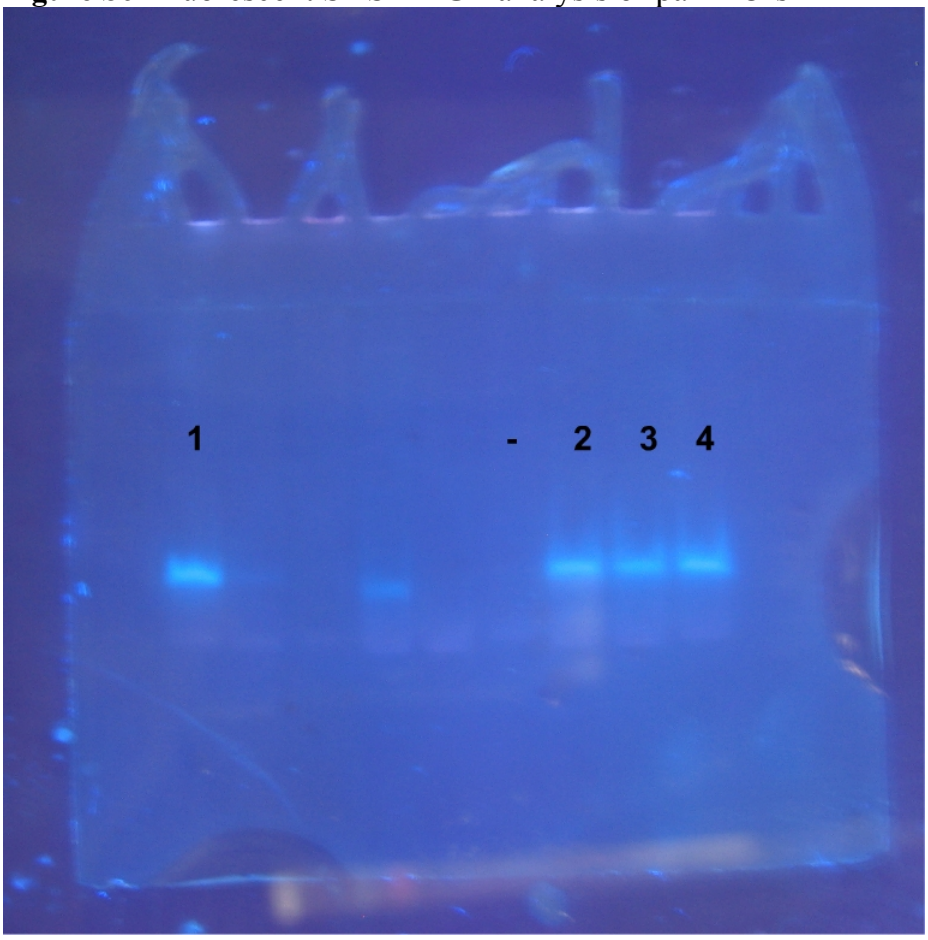
Archive Report of Selected Matches

1. gi|67476453|sp|P49327|FAS_HUMAN Mass: 273227 Score: 3083 Queries matched: 74 emPAI: 2.05

Fatty acid synthase [Includes: [Acyl-carrier-protein] S-acetyltransferase ; [Acyl-carrier-protein] S-malonyltransferase ; 3-oxo

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
1	506.8353	1011.6561	1011.6328	23.0	0	75	7.6e-006	1	R.VAAAVDLIIK.S
4	519.7886	1037.5626	1037.5393	22.4	0	56	0.002	1	R.VFTTVGSAEK.R
5	519.8468	1037.6791	1037.6485	29.5	0	49	0.0015	1	R.GTPLISPLIK.W
7	521.8283	1041.6421	1041.6182	22.9	0	54	0.0022	1	R.ELNLVLSVR.E
8	522.3081	1042.6016	1042.5771	23.5	0	60	0.00083	1	K.AQVADVVVSR.W
11	535.2791	1068.5436	1068.5200	22.1	0	74	3.2e-005	1	R.DPSQQELPR.L
13	537.2875	1072.5605	1072.5302	28.3	0	53	0.0062	1	R.QVQPEGPYR.V
15	543.3201	1084.6256	1084.6128	11.8	0	66	0.00019	1	R.QEPLLIGSTK.S
17	558.3566	1114.6987	1114.6598	34.9	0	63	0.00024	1	R.SEGVVAVLLTK.K
23	590.3109	1178.6073	1178.5502	48.4	0	45	0.032	1	K.QAHTMDPQLR.L + Gln->pyro-Glu (N-term Q)
26	595.8389	1189.6632	1189.6343	24.3	0	57	0.0021	1	R.SDEAVKPFGLK.V
28	598.2979	1194.5812	1194.5452	30.1	0	(41)	0.081	1	K.QAHTMDPQLR.L + Gln->pyro-Glu (N-term Q); Oxidation (M)
29	598.8069	1195.5993	1195.5768	18.8	0	(42)	0.06	1	K.QAHTMDPQLR.L
45	621.3683	1240.7220	1240.6928	23.6	0	79	7.8e-006	1	R.EVWALVQAGIR.D
47	623.3479	1244.6813	1244.6434	30.4	0	44	0.043	1	R.AALQEELQLCK.G
48	624.3424	1246.6703	1246.6418	22.9	0	57	0.0024	1	K.VSVHVIEGDHR.T
53	632.3867	1262.7589	1262.7347	19.2	0	72	2.8e-005	1	R.LQVVDQPLPVR.G
61	645.8050	1289.5954	1289.5677	21.5	0	71	8.7e-005	1	K.VYQWDDPDPR.L
62	646.8413	1291.6681	1291.6152	41.0	0	109	1.4e-008	1	-.MEEVVIAGMSGK.L + Acetyl (Protein N-term)
63	649.8613	1297.7080	1297.6626	35.0	0	107	1.7e-008	1	K.VGDPQELNGITR.A
64	650.3605	1298.7064	1298.6466	46.0	0	(87)	1.8e-006	1	K.VGDPQELNGITR.A + Deamidated (NQ)
69	661.8680	1321.7214	1321.6878	25.4	0	55	0.0032	1	R.QQEQQVPILEK.F + Gln->pyro-Glu (N-term Q)
71	662.8294	1323.6443	1323.6050	29.7	0	(82)	6.9e-006	1	-.MEEVVIAGMSGK.L + Acetyl (Protein N-term); 2 Oxidation
72	665.8840	1329.7535	1329.7153	28.7	0	69	0.00011	1	R.VTAIHIDPATHR.Q
76	670.3823	1338.7501	1338.7143	26.7	0	(51)	0.0055	1	R.QQEQQVPILEK.F
81	676.8826	1351.7506	1351.7282	16.5	0	68	0.00012	1	K.MVVPGLDGAQIPR.D
86	693.9030	1385.7915	1385.7402	37.0	0	105	2.5e-008	1	K.GVDLVLNSLAEEK.L
88	701.8559	1401.6973	1401.6446	37.6	0	81	9.7e-006	1	K.ADEASELACPTPK.E + Propionamide (C)
89	703.4254	1404.8362	1404.8089	19.4	0	98	8.5e-008	1	R.DLVEAVAHILGIR.D
90	703.8970	1405.7795	1405.7388	29.0	0	101	7.4e-008	1	K.VLQGDLVMNVYR.D
95	711.9018	1421.7890	1421.7337	38.9	0	(82)	5.9e-006	1	K.VLQGDLVMNVYR.D + Oxidation (M)
97	713.8993	1425.7841	1425.7650	13.4	0	77	1.7e-005	1	R.SLLVNPEGPTLMR.L
102	721.9174	1441.8202	1441.7599	41.8	0	(57)	0.0018	1	R.SLLVNPEGPTLMR.L + Oxidation (M)
105	735.3692	1468.7239	1468.6947	19.9	0	70	0.00012	1	R.FPQLDSTSFANSR.D
108	748.4380	1494.8615	1494.8154	30.8	1	52	0.0046	1	R.RQQEQQVPILEK.F
113	768.9401	1535.8656	1535.8209	29.1	0	83	4.5e-006	1	K.VVEVLAGHGHLYSR.I
119	775.9112	1549.8078	1549.7624	29.3	0	90	1.3e-006	1	R.AFEVSENGNLVVSGK.V + Deamidated (NQ)
124	797.4619	1592.9092	1592.8967	7.85	0	53	0.0042	1	R.VLFPATGYLSIVWK.T
127	807.4366	1612.8586	1612.8169	25.9	0	101	1e-007	1	K.EDGLAQQQTQLNLR.S
128	811.9984	1621.9822	1621.9291	32.8	0	115	1e-009	1	K.VVVQVLAEEPEAVLK.G
132	825.9376	1649.8606	1649.8195	24.9	0	(55)	0.0042	1	K.SNMGHPEPASGLAALAK.V
135	833.9291	1665.8437	1665.8144	17.6	0	72	8.3e-005	1	K.SNMGHPEPASGLAALAK.V + Oxidation (M)

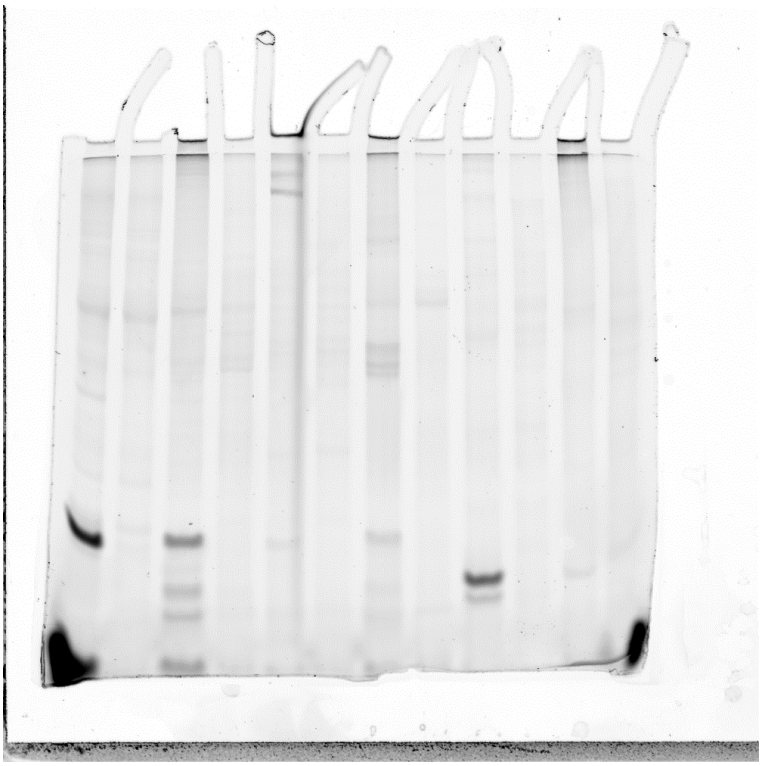
Figure S6 Fluorescent SDS-PAGE analysis of pan KO's



E.coli lacking the genes panF (2), panC (3), and panD (4), were incubated overnight with compound **1**. After lysis, the lysate was subjected to the conjugation reaction with fluorescent alkyne **3**. In each of the knockout strains, fatty acid ACP was labeled similarly as ACP in the native organism (1).

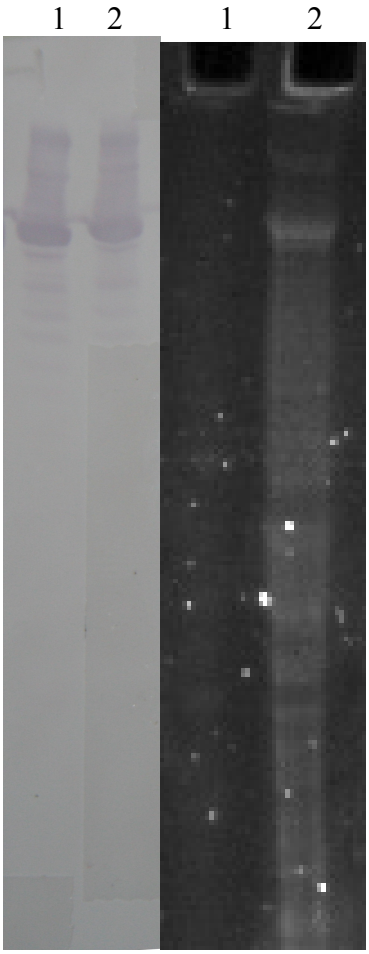
Complete gel images

Figure S7 In vivo ACP labeling in bacteria.



E. coli lysate from treated and untreated cultures (lanes 1&2). *B. subtilis* 6051 lysate from treated and untreated cultures (3&4). *B. brevis* 26A1 lysate from treated and untreated cultures (5&6). *B. subtilis* 168 lysate from treated and untreated cultures (7&8). *S. oneidensis* MR-1 lysate from treated and untreated cultures (9&10).

Figure S8 SKBR3 lysate.



Human FAS detected by western blotting with Anti-FAS (Left). Lysate from treated (2) and untreated cultures (1)

HPLC traces for all compound 2 CoA pathway intermediates

Figure S9 HPLC trace for compound **2**

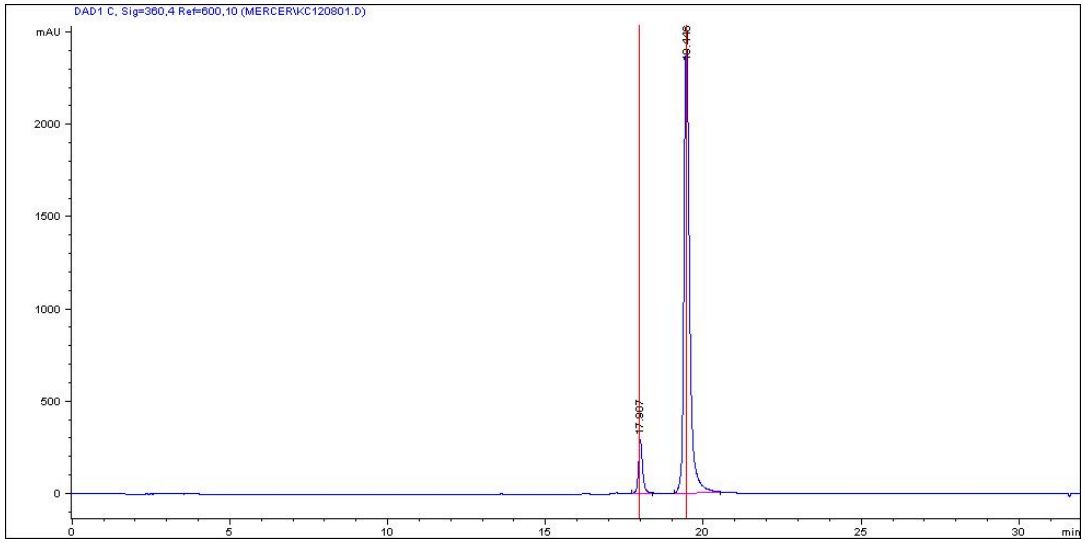


Figure S10 HPLC trace for compound **2** + PanK

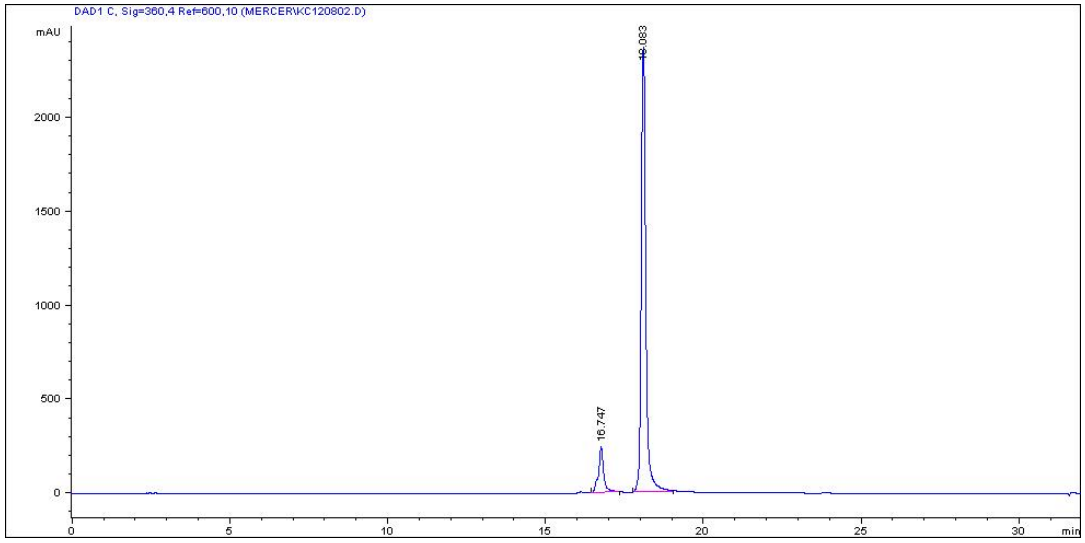


Figure S11 HPLC trace for compound **2** + PanK + PPAT

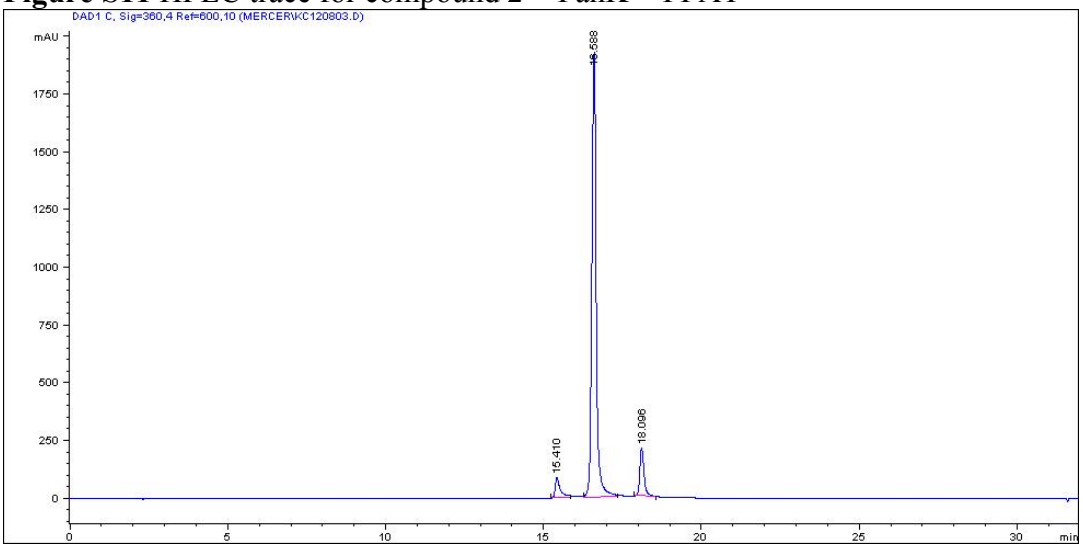


Figure S12 HPLC trace for compound **2** + PanK + PPAT + DPCK

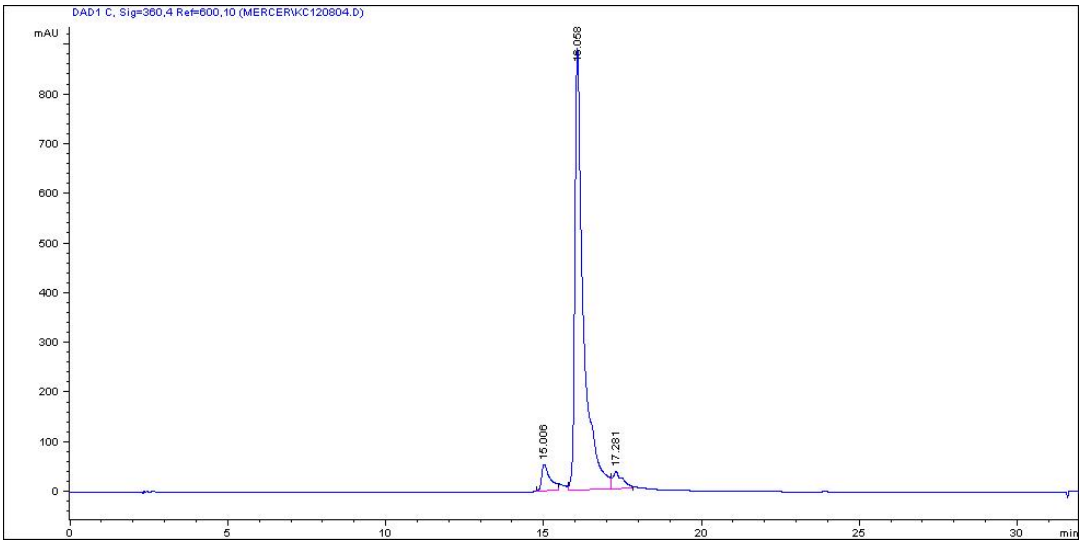


Figure S13 HPLC trace of SKBR3 lysate.

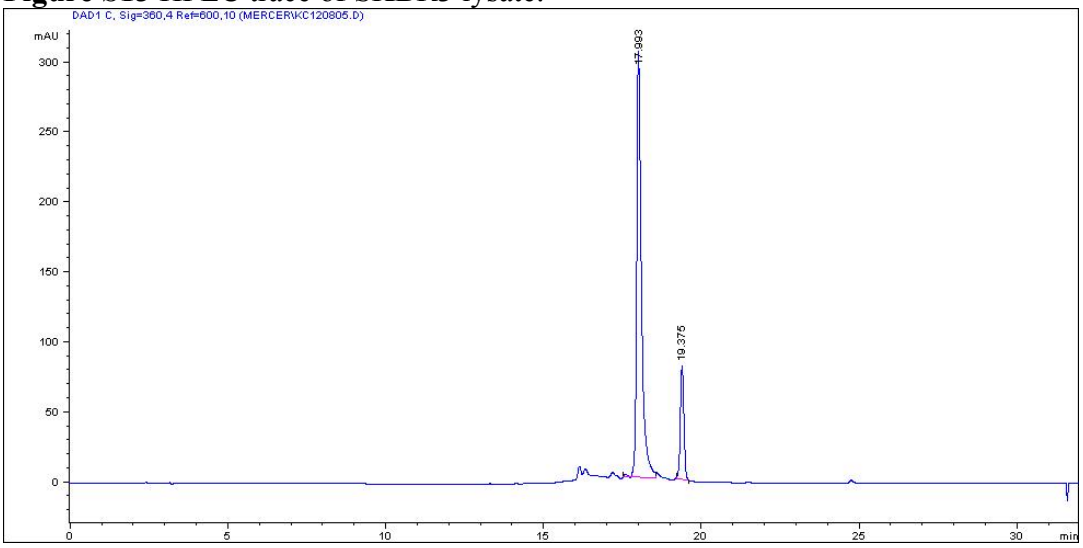
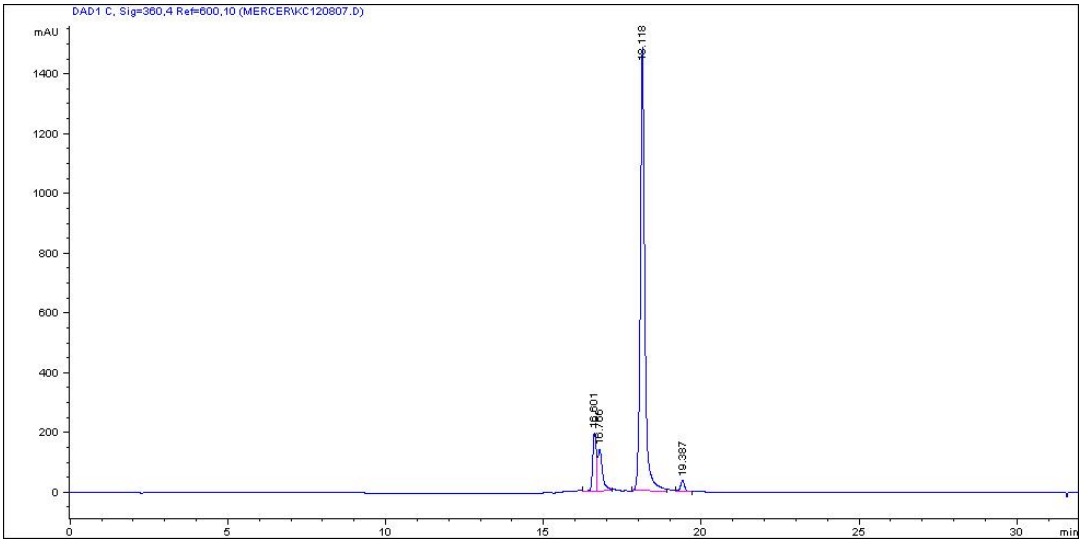


Figure S14 HPLC trace showing co-elution of phospho-**2** and SKBR3 lysate



Kinetics for pantetheine analogs

Table S1 Kinetic values of pantetheine analogues with *E. coli* PanK

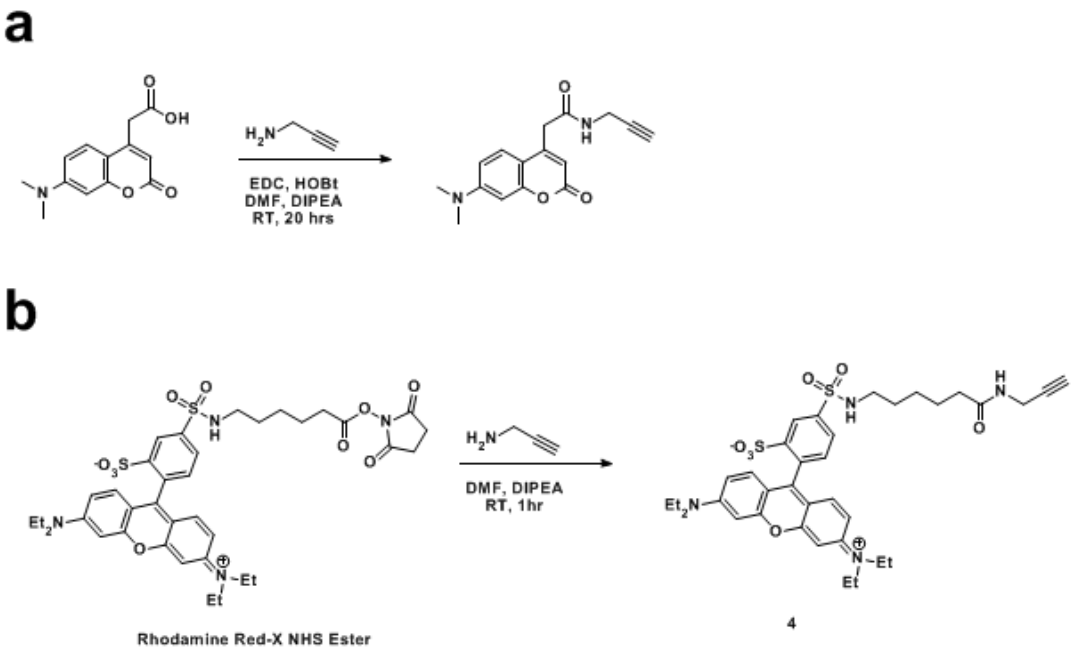
Compound #	Km	Vmax	Kcat	Kcat/Km
1	36.04±6.05	0.31±0.02	0.77±0.04	21.38±2.1
2	28.40±6.92	0.12±0.01	0.32±0.01	11.25±1.4
N5-Pan	33.98±7.11	0.34±0.02	0.84±0.05	24.59±2.9
Pantothenic acid	28.56±1.76	0.21±0.00	0.52±0.02	18.25±1.9

.

Figure S15 Structure of N5-Pan

Scheme for synthesis of fluorescent alkynes 3 and 4

Synthesis detailed in main text.



Acknowledgments

The text of chapter 5, in full, is a reprint of material as it appears in: Mercer, A.C., Meier, J.L., Torpey, J.W., and Burkart M.D. (2009). In vivo modification of native carrier protein domains. *Chembiochem. 10*, *1091-1100*. Permission was obtained from all co-authors. I was the second author of this work and contributed substantially to the results as well as the preparation of the manuscript. Andrew Mercer performed biochemical assays and Justin Torpey performed mass spectral analysis. I assisted with biological assays, synthesized all compounds, and wrote the manuscript. This research was performed under the supervision of Prof. Michael Burkart.

Chapter Six

Fluorescent profiling of modular biosynthetic enzymes by complementary metabolic and activity based probes.

Originally published as:

Meier, J. L.; Mercer, A. C.; Burkart, M. D. *J Am Chem Soc* **130**, 5443-5445 (2008).

J|A|C|S COMMUNICATIONS

Published on Web 04/01/2008

Fluorescent Profiling of Modular Biosynthetic Enzymes by Complementary Metabolic and Activity Based Probes

Jordan L. Meier, Andrew C. Mercer, and Michael D. Burkart*

Department of Chemistry and Biochemistry, University of California—San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358

Received December 19, 2007; E-mail: mburkart@ucsd.edu

Natural products compose a wide-ranging milieu of antibiotic and anticancer drug leads, virulence factors, and signaling molecules. Many of these small molecules are produced by highly versatile and modular polyketide synthases (PKS), nonribosomal peptide synthetases (NRPS), or PKS−NRPS hybrids[1] that are structurally and functionally analogous to fatty acid synthase (FAS) systems.[2] The past 25 years have seen significant progress in the isolation and structure elucidation of PKS and NRPS small molecule metabolites and the genetics and enzymology of the modular synthases which produce them.[3] Proteomic studies of natural product producers have lagged behind, in part because of increasingly streamlined genomic approaches which allow access to natural product gene clusters and recombinant enzymes. However, some of these enzymes, particularly multidomain PKS and NRPS systems, are resistant to study as recombinant enzymes both due to their large size and the intractability of their producer organisms to genetic manipulation and heterologous expression.[4] Direct profiling of microbial proteomes could prove highly complementary to genetic approaches by allowing us to understand the activity, transcriptional control, and post-translational modification of these enzymes in their native and dynamic proteomic environments.

A common feature of PKS, NRPS, and FAS systems is the use of carrier protein (CP) domains as a scaffold for the tethering and elongation of biosynthetic intermediates.[5] The site of this enzymatic tethering is the thioester of a post-translationally appended, coenzyme A (CoA) derived 4′-phosphopantetheine group. We have described two methods which utilize this unique post-translational modification for fluorescence/affinity labeling of CP domains in proteomic environments, first through an in vitro chemoenzymatic approach in which CoA analogues such as **1** along with the promiscuous phosphopantetheinyltransferase (PPTase) Sfp[6] are used to label and enrich CPs in crude cell lysate (Figure 1A), and more recently through the cellular uptake and in vivo metabolic pathway incorporation of fluorescent (**2**) or bioorthogonally tagged CoA precursors (Figure 1B).[7,8] While the specificity of these methods for the labeling of CP domains is ideal in many respects, it also restricts the information CP labeling methodologies can offer in terms of the type (PKS, NRPS, or FAS), identity, or alternate activities present in the labeled modular synthase. CP labeling methods are also reliant on a lack of either endogenous CP-phosphopantetheinylation or substrate promiscuity in an organism's CoA biosynthetic pathway, both of which represent potential limitations.

The focus of the present study is the supplementation of these CP based protein profiling approaches with activity based protein profiling (ABPP).[9] ABPP is a proteomic method which utilizes irreversible enzyme inhibitors, specific for a given enzyme class, labeled with fluorescent or affinity reporters to divide and classify a proteome based on enzyme activity. While such probes are not specific for PKS or NRPS enzymes, type I modular synthases of this type are uniquely susceptible to such a method of interrogation due to the presence of multiple active sites on a single polypeptide, each of which can be potentially targeted by an activity based probe (Figure 1C). The combination of CP-specific labeling methods with the proteome wide reactivity of activity based probes offers a powerful method for the identification, domain characterization, and inhibitor discovery of these biosynthetic enzymes.

To test the feasibility of this approach, we first generated a panel of fluorescently labeled activity based probes (Figure 2, **3**–**8**) and tested them for in vitro reactivity against a number of purified PKS, NRPS, and FAS hydrolytic enzymes. The fluoro-

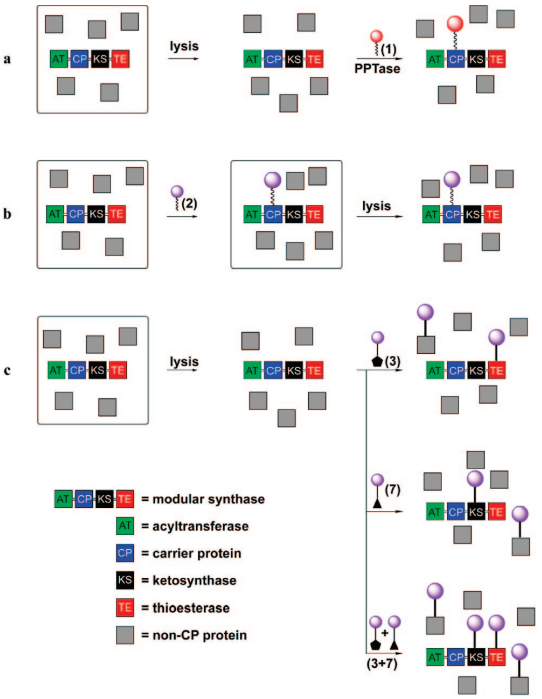


Figure 1. Methods for the proteomic analysis of FAS, PKS, and NRPS enzymes (PKS pictured). For probe structures, see Figure 2. (A) Chemoenzymatic labeling of apo-CP domains by CoA analogue **1** and Sfp in crude cell lysate. (B) Metabolic labeling of CP domains by uptake, biosynthetic processing, and in vivo labeling by fluorescent pantetheine analogue **2**. (C) Activity based protein profiling of KS and TE domains by probes **3** and/or **7**.

10.1021/ja711263w CCC: $40.75

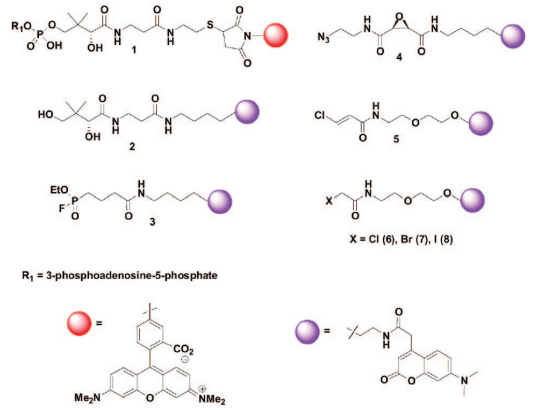


Figure 2. Chemoenzymatic tag (**1**), metabolic label (**2**), and activity based probes (**3**−**8**) utilized in this study.

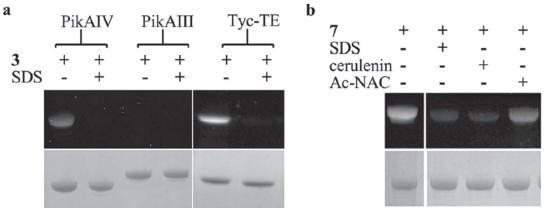


Figure 3. (A) In vitro labeling of purified recombinant PKS and NRPS acyltransferase (AT) and thioesterase (TE) domains by fluorophosphonate **3** (25 uM). (B) In vitro labeling of purified recombinant PikAIV PKS by bromoacetamide **7** (12.5 uM). PikAIII = module 3 of pikromycin PKS containing AT, CP, KR, and KS domains. PikAIV = module 4 containing AT, CP, KS, and TE domains. Tyc-TE = excised tyrocidine NRPS TE domain.

phosphonate warhead of probe **3** is a specific inhibitor of serine hydrolases which has been extensively applied toward a multitude of ABPP applications in eukaryotic systems.[10] Multimodular PKS enzymes utilize two distinct serine hydrolases during biosynthesis, an acyltransferase (AT) domain for substrate loading, and a thioesterase (TE) domain for chain termination. To test the specificity of **3**, we incubated it with recombinant enzymes from the pikromycin PKS and found it showed strong labeling of only TE containing domains (PikAIV, Figure 3A).[11] Probe **3** also showed strong labeling of the excised NRPS TE domain from the tyrocidine synthase. In both cases, this signal was lost upon active-site disruption by addition of detergent (SDS). This indicates the reactivity of **3** with modular biosynthetic enzymes should be limited to the labeling of terminal TE containing modules (Figure 3A).

For the labeling of PKS ketosynthase (KS) cysteine esterases, we synthesized **4**, a duel bioorthogonal/fluorescently labeled analogue of the well-known KS inhibitor cerulenin.[12] Unfortunately, this probe showed low levels of labeling of KS enzymes at concentrations up to 300 uM, as did a similarly labeled chloroacrylamide[13] **5**, making them technically unfeasible for use as activity based reagents (Supporting Information Figure S1). While chloroacetamide **6** showed slightly higher levels of labeling, it was haloacetamides **7** and **8** which showed the greatest SDS and cerulenin-sensitive labeling of KS enzymes at low (<25 μM) probe concentrations.[14] Although these reagents are known as nonspecific cysteine alkylating reagents at high concentrations, we found that performing our labeling

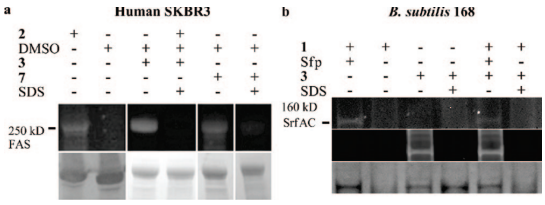


Figure 4. Metabolic labeling, chemoenzymatic labeling, and ABPP of eukaryotic and prokaryotic type I modular synthases. (A) FAS labeling of SKBR3 by metabolic tag **2**, probe **3**, or probe **7**. Top: fluorescence. Bottom left: anti-FAS blot. Bottom middle/right: Coomassie. (B) Chemoenzymatic and activity based labeling of *Bacillus subtilis* strain 168 by **1**/Sfp and **3**. Top: 532 nM excitation of probe **1**. Middle: UV excitation with 437 nM emission filter for visualization of probe **3**. Bottom: Coomassie. Note the SrfAC signal observed on coadministration of Sfp, **1**, and **3**.

reactions with 12.5 μM **7** or **8** in the presence of high (10 mM) concentrations of the scavenging nucleophile DTT led to active-site-directed labeling of the KS domain of PikAIV, as inferred from SDS-sensitive labeling and the decrease in labeling intensity observed upon preincubation of PikAIV with KS-reactive agents cerulenin or diacetyl cystamine (Ac-NAC) (Figure 3B).[15]

With these ABPP tools in hand, we first tested the compatibility of our duel labeling strategies in a proteomic context by applying them to a model system for PKS labeling, the eukaryotic FAS. The human breast cancer cell line SKBR3 is known to produce FAS at high levels, a characteristic phenotype associated with aggressive tumor growth.[16] SKBR3 cells were grown under standard tissue culture conditions in the presence of either 1 mM metabolic label **2** or vehicle DMSO for 48 h. Workup and analysis of the cytosolic fraction showed that, in contrast to the DMSO treated control, cells grown in the presence of fluorescent pantetheine analogue **2** show specific labeling of a high molecular weight (HMW) band corresponding to the approximate size of the human FAS (Figure 4A). Fluorophosphonate **3** and haloacetamide **7** showed SDS-sensitive labeling of the same band, confirmed as FAS by LC MS/MS analysis and treatment with an anti-FAS antibody, although **7** suffered from higher levels of nonspecific background than **2** and **3** (Figure 4A). Notably, FAS was the only protein which appeared to possess labeling with **2**, **3**, and **7**, demonstrating the complementarity of the probe set. We also performed competitive ABPP[10b] by preincubating SKBR3 lysate with known FAS inhibitors orlistat and cerulenin prior to administration of **3** or **7**. The results showed that orlistat decreased FAS labeling by **3** but not **7**, while cerulenin decreased FAS labeling by **7** but not **3** (Figure S2).[17] This is indicative of the ability of these probes to delineate the domain specificity of inhibitors of modular synthases in complex proteomes.

Finally, we sought to apply these probes directly to the analysis of natural product producer proteomes. *Bacillus subtilis* was the first sequenced Gram-positive bacteria and represents a model organism which is known to produce small molecule natural products characteristic of multidomain PKS and NRPS synthases.[18] Strain 168 of this organism contains a mutation in the allele coding for its secondary metabolism PPTase, lowering the amount of endogenous CP-phosphopantetheinylation and making it susceptible to our chemoenzymatic CP labeling technique (Figure 1A). Strain 6051 contains the wild-type allele, whose endogenous PPTase activity is necessary for metabolic

labeling by CoA precursor **2** (Figure 1B). Both strains in turn should be accessible by our activity based labeling strategy (Figure 1C).

Accordingly, *B. subtilis* 168 and 6051 were grown up to late log phase, lysed, and labeled chemoenzymatically by **1** and Sfp. The Sfp-dependent labeling of several HMW bands was observed in strain 168 but not strain 6051 (Figure 4B and Figure S3). Since labeling by our chemoenzymatic method can be blocked by endogenous phosphopantetheinylation in wild-type strain 6051, we also grew 6051 in the presence of metabolic label **2** and observed no fluorescent labeling of the HMW bands characteristic of multidomain modular synthases (Figure S4).

Having thus profiled the carrier proteome of *B. subtilis*, we sought to use ABPP to further characterize the CP labeling in strain 168. Significantly, none of the HMW bands showed active-site-dependent labeling by KS probe **7**, leading us to tentatively assign them the role of NRPS proteins. Administration of **3** to strain 168 showed strong denaturation-sensitive labeling of a ~150 kD protein corresponding to a band also labeled by **1** (Figure 4B). Inspection of the *B. subtilis* genome[19] led us to identify this protein as SrfAC, the terminal module of the surfactin synthase, on the basis of size, presence of CP and TE domains, and apparent absence of a KS domain. Band excision and MS/MS analysis confirmed this hypothesis, demonstrating 22% sequence coverage of the 144 kD polypeptide. MS analysis also provided evidence as to the site-specific nature of these probes as S1003, which represents the unmodified catalytic serine of the CP domain, and was found in MS spectra of samples treated with **3** but not **1**. Applying **3** to wild-type 6051 lysate, we did not observe similar labeling of this ~150 kD band, a finding which combined with the lack of CP labeling and differences in the Coomassie stain (Figure S5) led us to conclude this strain did not produce Srf proteins under the two growth conditions examined.

In summary, we have shown that metabolic and chemoenzymatic methods of CP labeling can be used in combination with ABPP to probe inhibitor specificity, assign domain structure, and assign identity of natural product producing modular synthases in vitro and in vivo. Of particular significance is the highly complementary nature of the methods used, which together provide a level of information not available to either of these methods of analysis alone. The results of our protein profiling of *B. subtilis* indicate this technology should be immediately useful for the analysis of natural product gene expression at the protein level. Using ABPP metabolic labeling as a link between genomic analysis and natural product isolation would help obviate one of the common difficulties of natural product genome mining by providing a simple and general assay for the expression of PKS and NRPS gene clusters.[20] While this is commonly followed by RT-PCR, since ABPP probes are not specific for PKS and NRPS enzymes, they may be useful in identifying other enzymes upregulated during natural product expression and delineating signaling pathways involved in secondary metabolite production. Such associations could prove highly valuable in terms of engineering biosynthetic pathways and providing new targets for the inhibition of production of PKS and NRPS virulence factors.[21] A current limitation of our activity based probe set is the small number of domains (KS and TE) targeted, which are absent in many modular biosynthetic enzymes; the design and synthesis of ABPP reagents targeting additional domains should be aided by growing knowledge of structure and inhibition of FAS and NRPS proteins.[22,23] The methods presented here should also lend themselves readily to enrichment and coupling to online tandem MS strategies which have demonstrated significant advantages in the detection and resolution of low-abundance proteins in proteomic samples.[9,24] Such technology has already been shown to be a powerful and highly compatible method for the analyses of purified modular synthases.[25] We are currently applying these complementary methods toward studies of secondary metabolism in a variety of natural product producer organisms.

Acknowledgment. This work was supported by the American Cancer Society RSG-06-011-01-CDD and Biochemistry and NIH Grant RO1GM075797. J.M. was supported by NIH Training Grant T32DK007233. We thank Justin Torpey of the UCSD Mass Spectrometry Core for MS/MS analysis, Timothy Foley and Bill Walkowicz for materials, Elinore Mercer and Andrew Worthington for enzymatic preparations, and Dr. Cleo Salisbury of The Scripps Research Institute for helpful discussions.

Supporting Information Available: Supplementary figures, procedures for the synthesis of probes **1**–**8**, complete gel images, and full experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468.
(2) Smith, S.; Tsai, S. C. *Nat. Prod. Rep.* **2007**, *24*, 1041.
(3) (a) Blunt, J. W.; Copp, B. R.; Hu, W.; Munro, M. H.; Northcote, P. T.; Prinsep, M. R. *Nat. Prod. Rep.* **2007**, *24*, 31. (b) Danadio, S.; Monciardini, P.; Sosio, M. *Nat. Prod. Rep.* **2007**, *24*, 1073. (c) Staunton, J.; Weissman, K. *Nat. Prod. Rep.* **2001**, *18*, 380.
(4) Pfeifer, B. A.; Khosla, C. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 106.
(5) (a) Mercer, A. C.; Burkart, M. D. *Nat. Prod. Rep.* **2007**, *24*, 750. (b) Lai, J. R.; Koglin, A.; Walsh, C. T. *Biochemistry* **2006**, *45*, 14869.
(6) Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. *Biochemistry* **1998**, *37*, 1585.
(7) LaClair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. *Chem. Biol.* **2004**, *11*, 195.
(8) (a) Clarke, K. M.; Mercer, A. C.; LaClair, J. J.; Burkart, M. D. *J. Am. Chem. Soc.* **2005**, *127*, 11234. (b) Meier, J. L.; Mercer, A. C.; Rivera, H.; Burkart, M. D. *J. Am. Chem. Soc.* **2006**, *128*, 12174.
(9) Evans, M. J.; Cravatt, B. F. *Chem. Rev.* **2006**, *106*, 3279.
(10) (a) Kidd, D.; Liu, Y.; Cravatt, B. F. *Biochemistry* **2001**, *40*, 4005. (b) Leung, D.; Hardouin, C.; Boger, D. L.; Cravatt, B. F. *Nat. Biotechnol.* **2003**, *21*, 687.
(11) Giraldes, J. W.; Akey, D. L.; Kittendorf, J. D.; Sherman, D. H.; Smith, J. L.; Fecik, R. A. *Nat. Chem. Biol.* **2006**, *2*, 531.
(12) Omura, S. *Methods Enzymol.* **1981**, *72*, 520.
(13) Worthington, A. S.; Rivera, H.; Torpey, J. W.; Alexander, M. D.; Burkart, M. D. *ACS Chem. Biol.* **2006**, *1*, 687.
(14) (a) Child, C. J.; Shoolingin-Jordan, P. M. *Biochem. J.* **1998**, *330*, 933. (b) Jez, J. M.; Noel, J. P. *J. Biol. Chem.* **2000**, *275*, 39640.
(15) Rendina, A. R.; Cheng, D. *Biochem. J.* **2005**, *388*, 895.
(16) Oskovian, B. *Cancer Lett.* **2000**, *149*, 43.
(17) Kridel, S. J.; Axelrod, F.; Rozenkrantz, N.; Smith, J. W. *Cancer Res.* **2004**, *64*, 2070.
(18) Stein, T. *Mol. Microbiol.* **2005**, *56*, 845.
(19) The Subtilist World Wide Web Server, http://genolist.pasteur.fr/SubtiList; accessed October 28th, 2007.
(20) Bode, H. B.; Muller, R. *Angew. Chem., Int. Ed.* **2005**, *44*, 6828.
(21) (a) Bergmann, S.; Schumann, J.; Scherlach, K.; Lange, C.; Brakhage, A. A.; Hertweck, C. *Nat. Chem. Biol.* **2007**, *3*, 213. (b) Ferreras, J. A.; Stirrett, K. L.; Lu, X.; Ryu, J.; Soll, C. E.; Tan, D. S.; Quadri, L. E. N. *Chem. Biol.* **2008**, *25*, 51.
(22) (a) Maier, T.; Jenni, S.; Ban, N. *Science* **2006**, *311*, 1258. (b) Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. *Science* **2007**, *316*, 288.
(23) (a) Zhang, Y-M.; White, S. W.; Rock, C. O. *J. Biol. Chem.* **2006**, *281*, 17541. (b) Qiao, C.; Wilson, D. J.; Bennett, E. M.; Aldrich, C. C. *J. Am. Chem. Soc.* **2007**, *129*, 6350.
(24) Jessani, N.; Niessen, S.; Wei, B. Q.; Nicolau, M.; Humphrey, M.; Ji, Y.; Han, W.; Noh, D. Y.; Yates, J. R.; Jeffrey, S. S.; Cravatt, B. F. *Nat. Methods* **2005**, *2*, 691.
(25) Dorrestein, P. C.; Kelleher, N. L. *Nat. Prod. Rep.* **2006**, *23*, 893.

JA711263W

Fluorescent Profiling of Modular Biosynthetic Enzymes by Complementary Metabolic and Activity Based Probes

*Jordan L. Meier, Andrew C. Mercer, and Michael D. Burkart**
Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358
mburkart@ucsd.edu

Supporting Information

Contents

Supplementary Figures

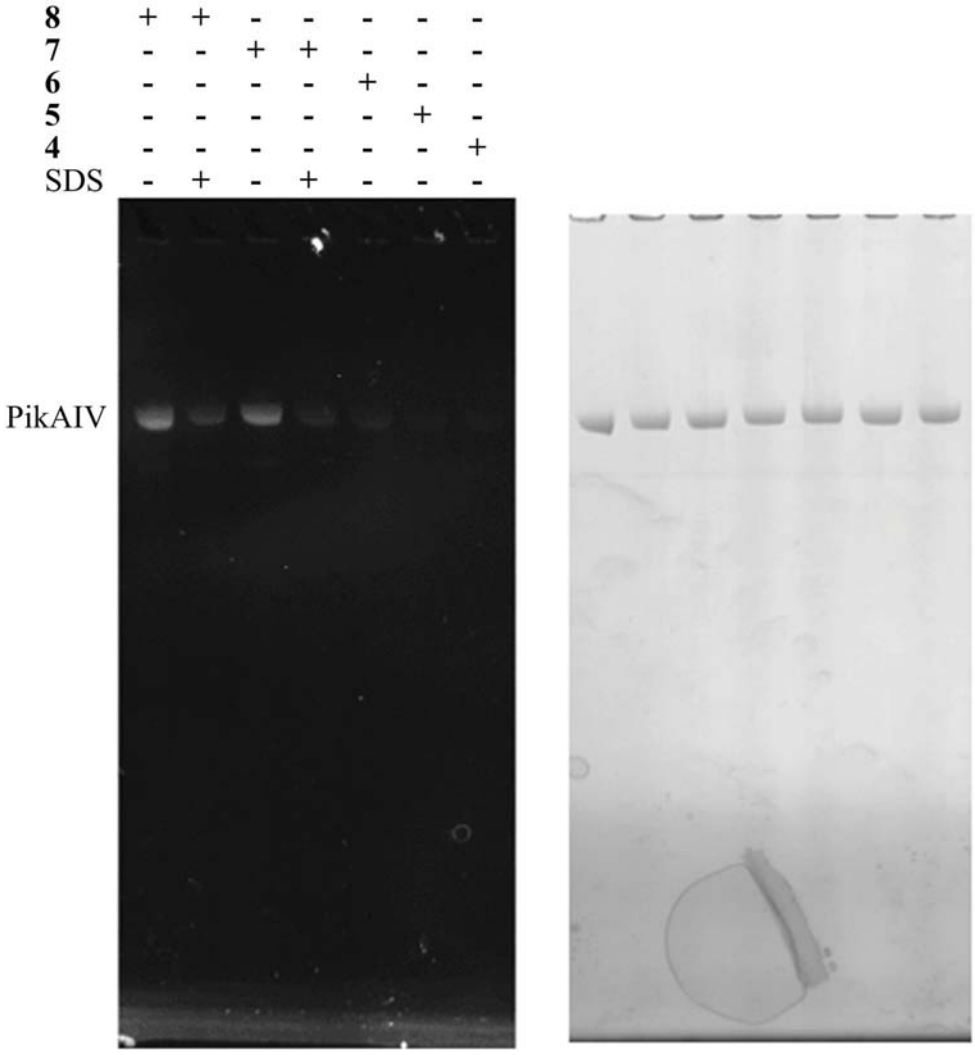


Figure S1: In vitro evaluation of labeling of purified recombinant PikAIV by KS probes **8** (12.5uM), **7** (12.5uM), **6** (300uM), **5** (300uM), or **4** (300 uM) in the presence or absence of 0.5% SDS. Only **8** and **7** provide high level labeling of PikAIV. Left: fluorescence. Right: Coomassie. PikAIV = module 4 of pikromycin PKS, containing AT, CP, KS, and TE domains.

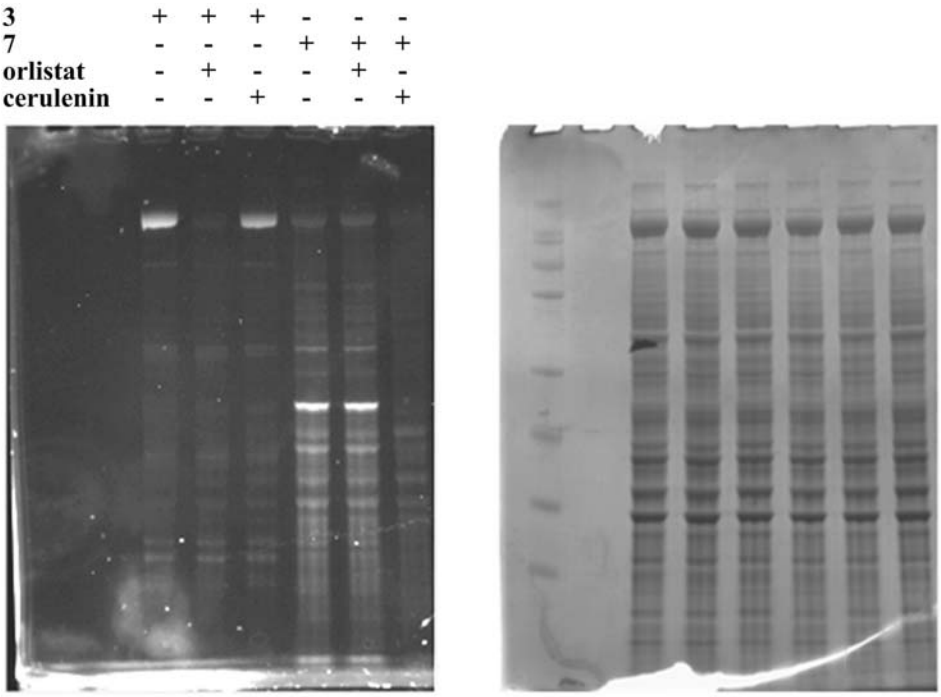


Figure S2: Competitive ABPP of human-FAS from SKBR3 lysate. Lanes 1`-3: Effect of preincubation of SKBR3 lysate with orlistat (500 uM) on labeling by **3** and **7**. Lanes 4-6: Effect of princubation of SKBR3 lysate with cerulenin (1 mM) on labeling by **3** and **7**.

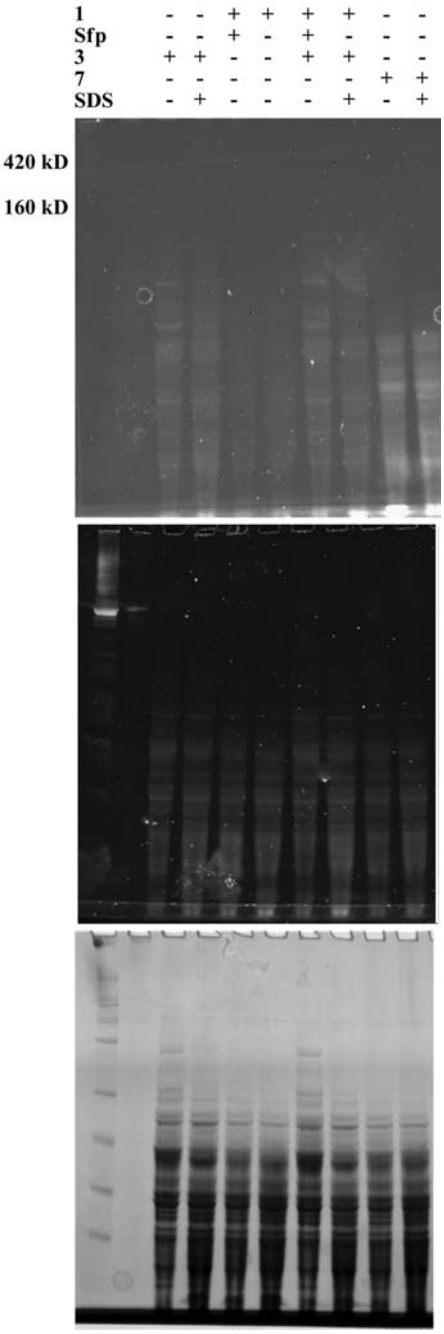


Figure S3: SDS-PAGE analysis of chemoenzymatic and activity-based labeling of strain 6051 grown in LB media.

Top: fluorescence (UV, 460 nM emission filter, coumarin).

Middle: fluorescence (Typhoon, 532 nM excitation, TAMRA).

Bottom: Coomassie

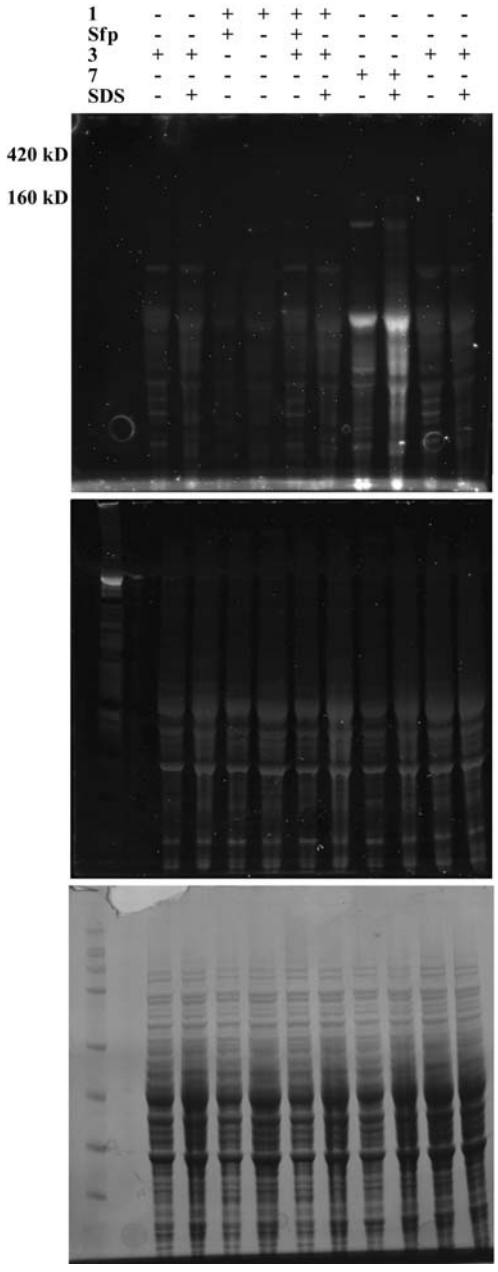


Figure S3 (continued): SDS-PAGE analysis of chemoenzymatic and activity-based labeling of strain 6051 grown in YEME media. Metabolic labeling by **2** (not shown) showed no apparent labeling of HMW proteins. Fluorophosphonate **3** (labeled lanes 1+2, 9+10) showed a similar lack of labeling, leading us to the conclusion this strain did not produce surfactin under these conditions. Lanes 9+10 were loaded with 25% more protein in an attempt to observe surfactin synthase (unsuccessful).

Top: fluorescence (UV, 460 nM emission filter, coumarin).

Middle: fluorescence (Typhoon, 532 nM excitation, TAMRA).

Bottom: Coomassie

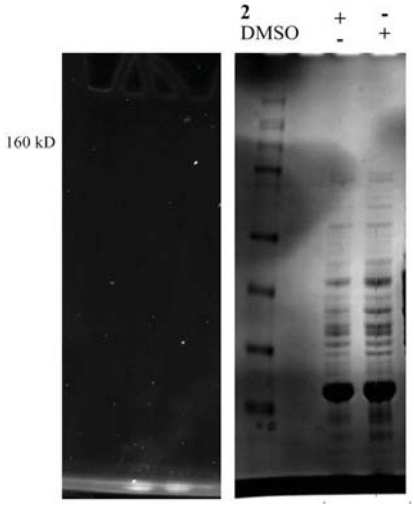


Figure S4*:* Labeling of *B. subtilis* 6051 grown in LB with metabolic probe **2**. Left: fluorescence (UV, 460 nM emission filter, coumarin). Right: Coomassie. No labeling of HMW proteins is observed. No >160 kD proteins are observed by Coomassie.

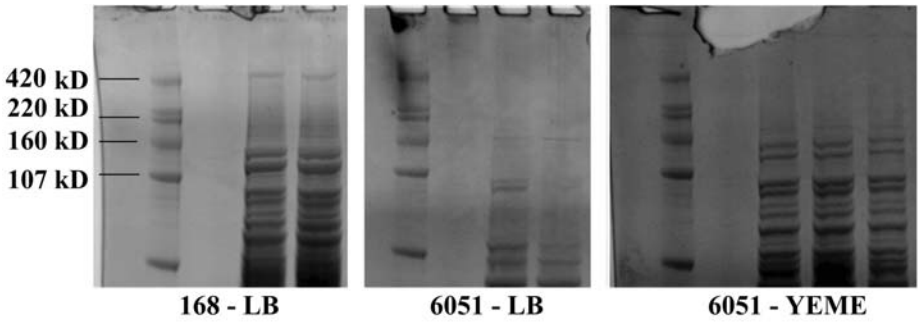


Figure S5: Comparison of *B. subtilis* 168 and 6051 protein expression as observed by Coomassie stain.

Left: Strain 168 grown in LB media.
Middle: Strain 6051 grown in LB media.
Right: Strain 6051 grown in YEME media.

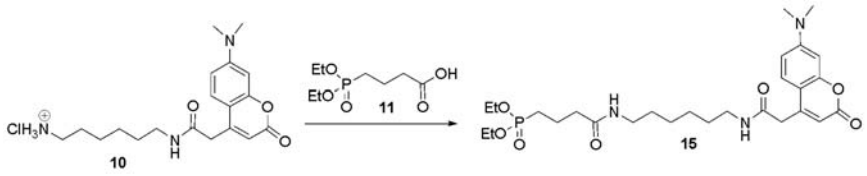
Note the absence under both 6051 growth conditions of the protein observed at MW~420 kD in 168. The size and labeling of this protein by **1** and Sfp (Figure 4B, Complete gel on page S24) suggests this is a component of the surfactin biosynthetic system which is presumably not being expressed in strain 6051 or not at levels high enough for detection by **1**, **2**, **3**, or Coomassie stain.

General synthetic procedures and materials

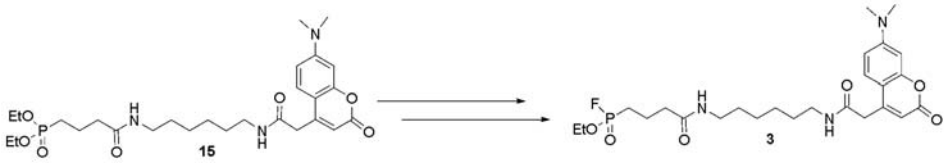
All commercial reagents (Sigma-Aldrich, Spectrum, MP Biomedicals, Alfa Aesar, TCI America, Acros) were used as provided unless otherwise indicated. Cerulenin was purchased from VWR. Orlistat was purchased from Aldrich. Cis-2,3-oxiranedicarboxylic acid monoethyl ester **12** was prepared from cis-2,3-oxiranedicarboxylic acid by the method of Bogyo.[1] TAMRA-CoA[2] **1**, coumarin-pantetheine[3] **2**, 7-dimethylaminocoumarin-4-acetic acid[4] **9**, 4-(diethoxyphosphinyl)-butanoic acid[5] **10**, aminocoumarin[3] **11**, 2-azidoethylamine hydrochloride[6] **13**, and PEG-coumarin[7] **14** were prepared according to published literature procedures. All reactions were carried out under argon atmosphere in dry solvents with oven-dried glassware and constant magnetic stirring unless otherwise noted. Triethylamine (TEA), N-methyl morpholine (NMM), and ethyl-*N,N*-diisopropylamine (DIPEA) were dried over sodium and freshly distilled. ^{1}H-NMR spectra were taken at 300, 400, or 500 MHz and ^{13}C-NMR spectra were taken at 100.6 or 75.5 MHz on Varian NMR spectrometers and standardized to the NMR solvent signal as reported by Gottlieb.[8] Multiplicities are given as s=singlet, d=doublet, t=triplet, q=quartet, p=pentet, dd=doublet of doublets, bs=broad singlet, bt=broad triplet, m=multiplet using integration and coupling constant in Hertz. TLC analysis was performed using Silica Gel 60 F254 plates (EM Scientific) and visualization was accomplished with UV light (λ=254 nm) and/or the appropriate stain (iodine, 2,4-dinitrophenylhydrazine, cerium molybdate, ninhydrin). Silica gel chromatography was carried out with Silicycle 60 Angstrom 230-400 mesh according to the method of Still.[9] TLC prep plate purification was performed with EMD Silica Gel 60 F_{254} pre-coated plates. Electrospray (ESI) and fast atom bombardment (FAB) mass spectra were obtained at the UCSD Mass Spectrometry Facility by Dr. Yongxuan Su using a Finnigan LCQDECA mass spectrometer and a ThermoFinnigan MAT- 900XL mass spectrometer, respectively.

Synthetic Procedures and Spectroscopic Data for Activity Based Probes 3-8

(a) Synthesis of FP-DMC (3)



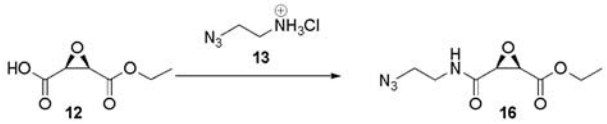
4-(diethoxyphosphinyl)-butanoic acid **10** (220 mg, 0.98 mmol), DIPEA (677 uL, 3.88 mmol), aminocoumarin-HCl **11** (368 mg, 0.970 mmol), and HOBt (558 mg, 3.71 mmol) were dissolved in DMF (20 mL) with stirring and cooled to 0°C. EDC (558 mg, 2.91 mmol) was added in one portion and the reaction was allowed to slowly warm to RT and stirred overnight. The solvent was removed under reduced pressure to yield a crude oil which was taken up in EtOAc (150 mL) and washed with water (1x100 mL), saturated $NaHCO_3$ (2x125 mL), and brine (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure to yield **15** as a fluorescent oil (405 mg, 76%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.48 (d, *J*=8.5 Hz, 1H), 6.95 (bt, *J*=6.0 Hz, 1H), 6.55 (dd, *J*=2.5 Hz, 9.0 Hz, 1H), 6.48 (bt, *J*=6.0 Hz, 1H), 6.39 (d, *J*=2.5 Hz, 1H), 6.03 (s, 1H), 4.01 (q, *J*=7.5 Hz, 4H), 3.57 (s, 2H), 3.13 (m, 4H), 2.99 (s, 6H), 2.26 (t, *J*=7.5 Hz, 2H), 1.86 (p, *J*=7.5 Hz, 2H), 1.73 (m, 2H), 1.37 (m, 4H), 1.26 (t, *J*=7.5 Hz, 6H), 1.21 (m, 4H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 172.4, 168.5, 162.3, 156.1, 153.2, 151.0, 126.1, 110.0, 109.3, 108.8, 98.2, 61.8 (d, *J*=6.8 Hz), 40.5, 40.3, 39.6, 39.1, 36.6 (d, *J*=13.8 Hz), 29.6, 29.3, 26.3, 25.5, 24.1, 19.2 (d, *J*=4.6 Hz), 16.6 (d, *J*=6.1 Hz). HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{48}N_3O_7P_1$, 551.2755 found 551.2757.



Bromotrimethylsilane (0.418 mL, 3.24 mmol) was added to a solution of **15** (350 mg, 0.635 mmol) in DCM (6.5 mL) at 0°C. The reaction was allowed to slowly attain RT and stir for 2.5 hrs. The reaction was quenched with H_2O and the solvent was removed under reduced pressure via rotovap with an inline base (NaOH) trap. The crude monophosphinic acid was carried forward as a low R_f oil as observed by TLC. This monophosphinic acid residue, was then dissolved in dry DMF (5 mL), transferred to a plastic reaction vial, and cooled to –78°C with stirring. DAST (50 uL, 0.39 mmol) was added dropwise via microsyringe. After 10 minutes the reaction was allowed to slowly attain RT and diluted with EtOAc (250 mL). This solution was washed with H_2O (1x100 mL) and brine (1x100 mL), and the H_2O layer saved and lyophilized to recover the crude

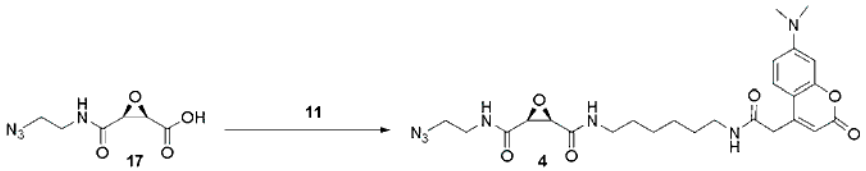
monophosphinic acid (282 mg, 85%). The organic layer was then dried over Na_2SO_4, filtered, and evaporated under reduced pressure to yield the crude fluorophosphonate. This residue was redissolved in $CHCl_3$ and purified by flash chromatography ($CHCl_3$ to 1:3 $CHCl_3$:acetone to 1:3:0.1 $CHCl_3$:acetone:MeOH) to give **3** as a fluorescent oil. This procedure was repeated on the recovered monophosphinic acid an additional time to provide a total of 22.5 mg (6.5 % yield from **15**) of fluorophosphonate **3** (estimated 90% purity by ^{1}H-NMR). Note: Although the above procedure provided material sufficient for this study, we have since found that the method of Cravatt[10] (since modified by Zhang[11]), which utilizes a less polar N-hydroxysuccinimidyl carbonate functionalized fluorophosphonate intermediate, provides a higher yielding and more versatile route to functionalized fluorophosphonates and is highly recommended. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.47 (m, 1H), 6.60 (m, 2H), 6.48 (m, 1H), 6.23 (bs, 1H), 6.04 (s, 1H), 3.60 (m, 2H), 3.18 (m, 4H), 3.03 (s, 6H), 2.31 (t, *J*=7.0 Hz, 2H), 1.99 (m, 4H), 1.40 (m, 4H), 1.25 (m, 7H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 171.9, 168.1, 160.7, 155.1, 153.3, 153.2, 150.9, 126.1, 113.8, 110.2, 109.5, 98.3, 53.4 (d, *J*=6.9 Hz), 43.5, 41.7, 40.4, 39.5, 36.0 (d, *J*=15.7 Hz), 29.5, 29.2, 26.1, 24.0, 22.6, 18.5 (d, *J*=4.6 Hz), 14.5. ^{19}F-NMR (282 Hz, $CDCl_3$) δ -66.9 (d, J_p=1071 Hz, F-P). ^{31}P-NMR (162 Hz, $CDCl_3$) δ 33.3 (d, J_F=1071 Hz, P-F). HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{25}H_{37}N_3O_6F_1P_1$, 525.2399, found 525.2398.

(b) Synthesis of Succinyl-Epoxyamide DMC (4)



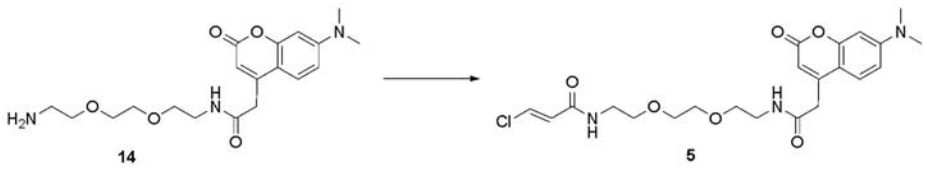
Cis-2,3-oxiranedicarboxylic acid monoethyl ester **12** (784 mg, 4.90 mmol) was dissolved in THF (35 mL) and N-methyl morpholine (520 uL, 4.90 mmol) was added. Upon dropwise addition of isobutylchloroformate (635 uL, 4.90 mmol) a white precipitate formed and the reaction was allowed to stir at room temperature for 10 minutes. Over this period the reaction mixture slowly turned a pale yellow. At this point 2-azidoethylamine hydrochloride **13** (480 mg, 5.0 mmol) and N-methyl morpholine (572 uL, 5.0 mmol) were dissolved in DMF (10 mL) and added to the stirring reaction mixture. The reaction was monitored by TLC and stirred at room temperature for 1 hour. Evaporation of the solvent gave a crude oil that was redissolved in EtOAc (150 mL) and extracted with saturated $NaHCO_3$ (2x100 mL), 5% citric acid (1x50 mL), and brine (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure to yield **16** (911 mg, 81%) as a brown oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 6.77 (bs, 1H), 4.24 (m, 2H), 3.70 (s, 2H), 3.42 (m, 4H), 1.29 (t, *J*=7.2 Hz, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 166.3, 165.4, 62.3, 54.6, 53.4, 50.5, 38.4, 14.1. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_8H_{12}N_4O_4$, 228.0853, found 228.0855.

Azido-epoxide **16** (911 mg, 4.0 mmol) was dissolved in absolute ethanol (20 mL) and cooled to 0°C. A solution of KOH (595 mg, 10.6 mmol) in absolute ethanol (11 mL) was added dropwise over 10 minutes to the stirring epoxide. The reaction mixture was allowed to slowly attain room temperature and stirred for 5 hours. The solvent was removed under reduced pressure and the crude reaction redissolved in H_2O (25 mL). The aqueous layer was washed with DCM (2x50 mL) and acidified by dropwise addition of concentrated HCl (0.85 mL). After extraction with EtOAc (2x200 mL) the organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure to yield **17** (614 mg, 77%) as a clear oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.96 (bs, 1H), 6.98 (bs, 1H), 3.77 (s, 2H), 3.45 (m, 4H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.9, 166.5, 54.6, 53.4, 50.5, 38.7. HRMS (EI) (*m/z*) $[M+H]^+$ calcd for $C_6H_9N_4O_4$, 201.0618, found 201.0617.



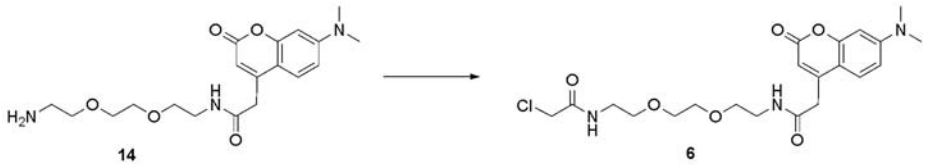
Carboxylic acid **17** (50 mg, 0.25 mmol) was dissolved in THF (7 mL) and N-methyl morpholine (51 uL, 0.48 mmol) was added. Upon dropwise addition of isobutylchloroformate (35 uL, 0.27 mmol) a white precipitate formed and the reaction was allowed to stir at room temperature for 10 minutes. At this point aminocoumarin-HCl **11** (90 mg, 0.24 mmol) and N-methyl morpholine (51 uL, 0.48 mmol) were dissolved in DMF (5 mL) and added to the stirring reaction mixture. The reaction was monitored by TLC and stirred at room temperature for 1 hour. Evaporation of the solvent gave a crude oil that was redissolved in EtOAc (125 mL) and extracted with saturated $NaHCO_3$ (2x50 mL), 5% citric acid (1x25 mL), and brine (1x50 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The resultant oil was purified by flash chromatography (1:1 EtOAc:hexanes to EtOAc to 9:1 EtOAc/MeOH) to yield **4** (109 mg, 87%) as a brown-yellow solid. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.53 (d, *J*=9.2 Hz, 1H), 7.47 (bs, 1H), 7.11 (bs, 1H), 6.75 (bs, 1H), 6.60 (dd, *J*=2.8 Hz, 9.2 Hz, 1H), 6.44 (d, *J*=2.8 Hz, 1H), 6.08 (s, 1H), 3.70 (d, *J*=4.8 Hz, 1H), 3.65 (d, *J*=4.8 Hz, 1H), 3.63 (s, 2H), 3.39 (m, 4H), 3.18 (m, 4H), 3.03 (s, 6H), 1.42 (m, 4H), 1.22 (m, 4H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.7, 166.2, 165.5, 162.8, 156.1, 153.3, 151.3, 126.1, 109.7, 109.5, 108.7, 98.1, 55.6, 55.2, 46.6, 40.4, 40.3, 39.5, 39.0, 38.7, 29.3, 29.1, 26.1, 26.0. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{25}H_{33}N_7O_6$, 527.2487, found 527.2481.

(c) Synthesis of Chloroacrylamide DMC (5)



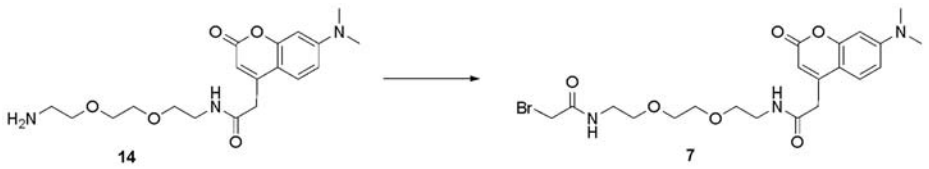
Trans-chloroacrylic acid (42 mg, 0.4 mmol), PEG-coumarin **14** (75 mg, 0.2 mmol), and DIPEA (104 uL, 0.6 mmol) were dissolved in DCM (20 mL) with stirring and PyBOP (208 mg, 0.4 mmol) was added in a single portion. The reaction was allowed to stir at room temperature and monitored by TLC. After 1 hour the reaction was complete and the solvent was removed under reduced pressure. The resultant oil was purified by column chromatography (DCM to 3:2 DCM:acetone) to yield **5** (66mg, 71 %) as a yellow solid. ^{1}H-NMR (400 MHz, $(CD_3)_2SO$) δ 8.30 (bs, 1H), 8.25 (bs, 1H), 7.52 (d, *J*=8.8 Hz, 1H), 7.19 (d, *J*=13.2 Hz, 1H), 6.68 (d, *J*=9.2 Hz, 1H), 6.53 (s, 1H), 6.44 (d, *J*=13.2 Hz, 1H), 5.98 (s, 1H), 3.59 (s, 2H), 3.39 (m, 8H), 3.24 (m, 4H), 2.99 (s, 6H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.7, 164.0, 162.5, 156.0, 153.2, 150.9, 133.3, 127.5, 126.0, 109.9, 109.4, 108.6, 98.1, 70.4, 70.4, 69.8, 69.7, 41.1, 40.3, 39.8, 39.5. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{30}N_3O_6Cl$ 455.1818, found 455.1820.

(d) Synthesis of Chloroacetamide DMC (6)



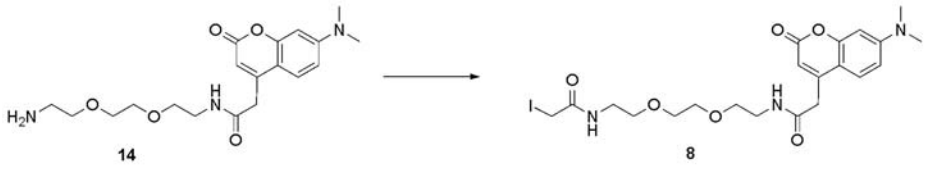
PEG-coumarin **14** (49 mg, 0.13 mmol), and DIPEA (67uL, 0.39mol) were dissolved in THF (10 mL) with stirring and cooled to 0°C. Chloroacetyl chloride (20 uL, 0.26 mmol) was added dropwise over 5 minutes. After 30 minutes the reaction was quenched by the addition of MeOH (5 mL) and the solvent removed under reduced pressure. The crude reaction mixture was redissolved in DCM and purified by flash chromatography (DCM to 19:1 DCM: MeOH) to yield **6** (56 mg, 94%) as a fluorescent yellow-green oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.45 (d, *J*=8.8 Hz, 1H), 7.14 (bs, 1H), 6.70 (bs, 1H), 6.56 (d, *J*=8.0 Hz, 1H), 6.41 (s, 1H), 6.02 (s, 1H), 4.00 (s, 2H), 3.59 (s, 2H), 3.47 (m, 8H), 3.39 (m, 4H), 3.00 (s, 6H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.5, 166.6, 162.2, 156.1, 153.3, 150.5, 125.9, 110.2, 109.4, 108.6, 98.2, 70.5, 70.4, 69.8, 69.6, 42.9, 40.5, 40.3, 39.8, 39.7. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{28}N_3O_6Cl$ 453.1661, found 453.1657.

(e) Synthesis of Bromoacetamide DMC (7)



PEG-coumarin **14** (49 mg, 0.13 mmol), and DIPEA (67uL, 0.39mol) were dissolved in THF (10 mL) with stirring and cooled to 0°C. Bromoacetyl bromide (22 uL, 0.26 mmol) was added dropwise over 5 minutes and the reaction mixture turned a dark brown. After 30 minutes the reaction was quenched by the addition of MeOH (5 mL) and the solvent removed under reduced pressure. The crude reaction mixture was redissolved in DCM (75 mL) and washed with saturated $NaHCO_3$ (2x75 mL), 5% citric acid (2x50 mL) and brine (1x75 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The resultant oil was purified by flash chromatography (DCM to 32:1 DCM: MeOH) to yield **7** (40 mg, 62.5%) as a fluorescent yellow-green oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.48 (d, *J*=8.8 Hz, 1H), 7.14 (bs, 1H), 6.59 (m, 2H), 6.45 (d, *J*=2.4 Hz, 1H), 6.05 (s, 1H), 3.86 (s, 2H), 3.62 (s, 2H), 3.50 (m, 12H), 3.03 (s, 6H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 168.4, 166.3, 162.3, 156.2, 153.3, 150.4, 125.9, 110.2, 109.4, 108.6, 98.3, 70.5, 69.8, 69.6, 40.7, 40.4, 40.1, 39.8, 29.4. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{28}N_3O_6Br$, 497.1156, found 497.1160.

(f) Synthesis of Iodoacetamide DMC (8)



Iodoacetic acid (93 mg, 0.50 mmol) was dissolved in THF (10 mL) and N-methyl morpholine (53 uL, 0.50 mmol) was added. Upon dropwise addition of isobutylchloroformate (65 uL, 0.50 mmol) a white precipitate formed and the reaction was allowed to stir at room temperature for 10 minutes, during which it slowly turned a deep yellow. At this point PEG-coumarin **14** (113 mg, 0.30 mmol) and N-methyl morpholine (53 uL, 0.50 mmol) were dissolved in THF (3 mL) and added to the stirring reaction mixture. The reaction was monitored by TLC and stirred at room temperature for 25 minutes. Evaporation of the solvent gave a crude oil that was redissolved in DCM (125 mL) and extracted with 1:1 brine/saturated $NaHCO_3$ (2x50 mL), 5% citric acid (1x50 mL), and brine (1x50 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The resultant oil was purified by flash chromatography (DCM to 19:1 DCM/MeOH) to yield **8** (118 mg, 72%) as a brown-yellow solid. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.44 (d, *J*=9.2 Hz, 1H), 7.25 (bs, 1H), 6.94 (bt, *J*= 5.2 Hz, 1H), 6.55 (dd, *J*=2.4, 8.8 Hz, 1H), 6.37 (d, *J*=2.4 Hz, 1H), 6.01 (s, 1H), 3.66 (s, 2H), 3.59 (s, 2H), 3.47 (m, 8H), 3.37 (m, 4H), 2.97 (s, 6H). ^{13}C-NMR

(100.6 MHz, $CDCl_3$) δ 168.6, 168.3, 162.4, 156.1, 153.2, 150.7, 126.0, 110.0, 109.4, 108.7, 98.2, 70.5, 70.4, 69.7, 69.6, 40.4, 40.3, 40.2, 39.8, -0.10. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{28}N_3O_6I$, 545.1017, found 545.1026.

Strains, Culture, Media, and Materials

Human breast cancer cell line SKBR3 was a kind gift of Dr. Jeff Smith. *B. subtilis* strain 168 was a kind gift of Dr. Kit Pogliano. FAS antibody was a kind gift of Dr. Stuart Smith. Pikromycin and MCAT expression plasmids were kind gifts of Dr. David Sherman and Dr. Andrei Osterman respectively. Wild type *B. subtilis* (ATCC 6051) was obtained from ATCC. Culture medium components were purchased from VWR. Lysozyme and DNAase were obtained from Worthington Biochemical. For proteome labeling experiments *B. subtilis* (168 and 6051) cells were grown in Luria Bertani medium for 12-16 hours before being pelleted by centrifugation (8000g),frozen, and resuspended in lysis buffer (100 mM NaCl, 25mM K_3PO_4 pH 7.0). Cells were lysed by addition of lysozyme (0.1 mg/mL) and passage through a French Pressure Cell and shook with DNAase for 1 hour at 0^{o}C. Pelleting of cellular debris afforded lysate (20 mg/mL) which was used for activity-based experiments or stored as a 20% glycerol mixture at –80^{o}C for later use. The same procedure was used to work up *B. subtilis* cells grown in YEME media according to the method of Straight.[12] Human SKBR3 cells were grown and worked up to afford cytosolic lysate according to the procedure of Smith.[13] SDS-PAGE was ran on 4-12% Bis-Tris NuPAGE gels (Invitrogen) unless otherwise noted. MALDI mass spectrometry was conducted at the UCSD Chemistry and Biochemistry Mass Spectrometry Core Facility by Justin Torpey.

Procedures for Chemoenzymatic Labeling by TAMRA-CoA/Sfp

Lysates were prepared as specified above. To a 30 uL reaction mixture containing *B. subtilis* lysate (10 mg/mL), DTT (10 mM), and Tris-HCl pH 6.8 (100 mM) was added sequentially TAMRA-CoA (10uM), $MgCl_2$ (40 mM), and Sfp. After incubation at room temperature for 1 hr the reaction was quenched with 5x SDS- loading buffer (strong reducing). Sfp minus controls were performed to control for non-specific background labeling by TAMRA-CoA. For visualization of SrfAC in *B. subtilis* 168, fluorophosphonate **5** (25 uM) was added and the reaction allowed to proceed under identical conditions. To better separate HMW proteins SDS-PAGE of strain 168 was performed on 3-8% Bis-Tris NuPAGE gels (Invitrogen) as well as 4-12% gels. Fluorescent visualization was performed using a Typhoon Laser Flatbed Scanner (GE Healthcare) with excitation at 532 nm. Color photographs were obtained after excitation at 360 nm using a standard CCD camera.

Procedures for Metabolic Labeling

Human SKBR3 Cells

Human SKBR3 cells were grown for 50 hrs in the presence or absence of **2** (1 mM) followed by workup to afford cytosolic lysate by the conditions specified above and SDS-PAGE.[13] Fluorescent gel visualization was performed using a Fluor-S Multiimager equipped with a 460 nm emission filter.

B. subtilis strain 6051

B. subtilis strain 6051 was grown in the presence or absence of **2** (1 mM) according to the conditions specified above followed by lysis and SDS-PAGE according to the procedure of Clarke.[3] Fluorescent gel visualization was performed using a BioRad Fluor-S Gel Doc equipped with a 460 nm emission filter.

Procedures for Activity Based Labeling

In Vitro Activity Based Labeling of Recombinant PKS/NRPS/FAS proteins

TycC-TE, PikAIV, MCAT, DEBS ACP-TE were prepared according to previously published procedures. Fluorophosphonate **5** (25 uM) was applied using conditions analogous to those optimized by Cravatt (100 mM Tris-HCl pH 8.0, 10 mM DTT, 1 hour RT).[10] For KS activity based labeling experiments, a 30uL reaction containing enzyme (1 ug), Tris-HCl pH 6.8 (100 mM), and DTT (10 mM) was cooled to 0^{o}C and activity based probe **4** (300 uM), **5** (300 uM), **6** (300 uM), **7** (12.5 uM), or **8** (12.5 uM) was added before or after addition of 0.5% SDS. For competitive activity-based protein profiling experiments enzyme was preincubated with inhibitor for 1 hour at the specified concentration before addition of activity based probe. In all experiments total DMSO concentration was kept at or below 3.3%. After reaction for 30 minutes at 0^{o}C labeling reactions were quenched with 5x SDS-loading buffer (strong reducing) and subjected to SDS-PAGE. Performing labeling reactions for up to two hours at room temperature did not result in a greatly enhanced signal for probes **4-6** although these conditions did increase background labeling of probes **7-8**. Probes **7** and **8** also showed high off-site reactivity when heated prior to SDS-PAGE, so boiling of these samples was avoided. DTT was not required to scavenge off-site reactivity of **7** and **8** in in vitro experiments, but appeared to be beneficial for proteomic treatments. Fluorescent gel visualization was performed using a BioRad Fluor-S Gel Doc equipped with a 460 nm emission filter.

Activity-Based Labeling of Human SKBR3, B. subtilis 0168, and B. subtilis 6051 Lysates

Lysates were prepared as specified above. For SKBR3 lysate labelings were performed at 1 mg/mL; for *B. subtilis* lysate labelings were performed at 7mg/mL to optimize signal. Fluorophosphonate **3** (25 uM) was applied using conditions analogous to those optimized by Cravatt (100 mM Tris-HCl pH 8.0, 10 mM DTT, 1 hour RT).[10] Bromoacetamide **7** (12.5 uM) was applied using conditions as specified above (100 mM Tris-HCl pH 6.8, 10 mM DTT, 0.5 hour 0^{o} C) followed by addition of 5x SDS-loading buffer and SDS-PAGE. 0.5% SDS was added as a protein denaturant for ABPP control reactions. For competitive activity-based protein profiling of SKBR3 lysate, protein was preincubated with inhibitor for 1 hour at the specified concentration before addition of activity based probe. Multiprobe experiments were performed at pH 6.8. In all experiments total DMSO concentration was kept at or below 3.3%. Fluorescent gel visualization was performed using a BioRad Fluor-S Gel Doc equipped with a 460 nm emission filter. Color photographs were obtained after excitation at 360 nm using a standard CCD camera.

MS Analysis of Fluorescently Labeled Bands

1-D SDS-PAGE and in-gel digestion.

After treatment of lysate with probes **1**, **2**, **3**, or **7** as specified above, the proteins were separated by 1-D SDS-PAGE using 4-12% Bis-Tris NuPAGE gels (Invitrogen). Excess probe fluorescence was removed by destaining in a solution of 50% MeOH / 40% H_2O / 10% AcOH overnight. For identification of SrfAC from *B. subtilis* 168, lysate was treated with chemoenzymatic probe **1**/Sfp or activity based probe **3**, and subjected to SDS-PAGE. Excitation at 365 nM and comparison with a HMW Benchmark Protein Ladder (Invitrogen) allowed excision of a single approximately 150 kD protein band from each lane. These two bands were then washed twice with 300 uL of water with vortexing for 10 min, and washed with 300 uL of MeOH for 10 minutes and the solvent removed. Subsequently, gel bands were washed twice with 200 uL of 50% acetonitrile (ACN) and 50% 5 mM DTT / 25 mM NH_4HCO_3 with vortexing for 10 min, and finally washed with 200 uL ACN. The dehydrated gel piece was rehydrated by addition of 20 uL of ice-cold 10 ng/uL trypsin (Promega) in 5 mM DTT / 25 mM NH_4HCO_3, and incubated on ice for 30 min and the remaining trypsin solution was removed and replaced with fresh 5 mM DTT / 25 mM NH_4HCO_3. The digestion was allowed to continue at 37°C overnight. The peptide mixture was then acidified with 2uL of 2% trifluoroacetic acid (TFA) (Sigma), vortexed for 30 min, and the supernate extracted. Finally, 20uL of 20% acetonitrile/ 0.1% TFA was added followed by vortexing to extract the remaining peptides and combined with the previous fraction. The combined extractions are analyzed directly by nanobore LCMSMS

.

Mass spectrometry

The samples were analyzed by electrospray ionization using a QSTAR-Elite hybrid mass spectrometer (Applied Biosystems/MDS Sciex) interfaced to a Tempo nanoscale reverse-phase HPLC (Eksigent/Applied Biosystems) using a 75 um x 15 cm column (Grace Davison) packed with Vydac MS C18 (300 A, 5 um) packing material by Justin Torpey at the UCSD Chemistry and Biochemistry Mass Spectrometry Core Facility. The buffer compositions were as follows: buffer A: 98% H_2O, 2% ACN, 0.1% acetic acid (Fluka), 0.005% heptafluorobutyric acid (Fluka); buffer B: 98% ACN, 2% H_2O, 0.1% acetic acid, 0.005% heptafluorobutyric acid. Samples of 10 uL were injected by the Tempo autosampler onto a C18 PepMap pre-column (5 mm x 300 um, LC Packings) using the Channel 1 loading pump at a flow rate of 15 uL/min buffer A. After washing for 5 min, the peptides were transferred onto the analytical column and eluted directly into the mass spectrometer with a 25-min linear gradient from 5 to 40% buffer B at a flow rate of 300 nL/min using Channel 2. LCMSMS data were acquired in a data-dependent fashion by selecting the most intense peak with charge state 2-4 that exceeds 40 counts with exclusion of former target ions set to always and the mass tolerance for exclusion set to 100 ppm. TOF MS were acquired at *m/z* 500-1800 Da for 0.5 sec with 20 time bins to sum. MSMS are acquired from *m/z* 65 – 2000 Da using "enhance all" and 20 time bins to sum, Dynamic Background Subtract, Automatic Collision Energy, and Automatic MS/MS Accumulation with the Fragment Intensity Multiplier set to 12 and Maximum Accumulation set to 3 sec before returning to the survey scan.

Database search.

The MSMS were analyzed by Analyst 2.0 (Applied Biosystems) and subjected to database search using Mascot 2.2.1 (Matrix Science) with Mascot Daemon 2.2 (Matrix Science) data import filter parameters set as follows: default precursor charge state 2-4; precursor and MSMS data centroiding using 50 % height and 0.05 amu merge distances. MSMS peaks with intensity less than 1% of the base peak were discarded, as were MSMS spectra with less than 22 peaks remaining. Data were searched against the *Bacillus subtilis* ssp. *subtilis* 168 database obtained from SubtiList (genolist.pasteur.fr/SubtiList/) containing 4037 sequences. The search identified tryptic peptides with up to 2 missed cleavages and used mass tolerances of 100 ppm (MS) and 0.10 Da (MSMS), with variable modifications as follows: deamidation (NQ), oxidation (M), pyro-Glu (N-term Q). The search results indicated that individual ion scores > 42 indicate identity or extensive homology ($P < 0.05$).

Compiled MS/MS Data of Gel Excised Fluorescent SrfAC bands

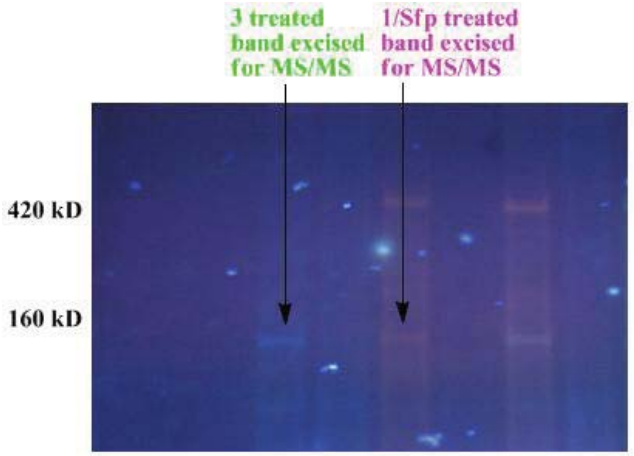
Srf AC

1 msqfskdqvq dmyylspmqe gmlfhpilnp gqtfyleqit mkvkgslnik cleesmnvim
61 drydvfrtvf ihekvkrpvq vvlkkrqfhi eeidlthltg seqtakiney keqdkirgfd
121 ltrdipmraa ifkkaeesfe wvwsyhhiil dgwcfgivvq dlfkvynalr eqkpyslppv
181 kpykdyikwl ekqdkqaslr ywreylegfe gqttfaeqrk kqkdgyepke llfslseaet
241 kaftelaksq httlstalqa vwsvlisryq qsgdlafgtv vsgrpaeikg vehmvglfin
301 vvprrvklse gitfnglllkr lqeqslqsep hqyvplydiq sqadqpklid hiivfenypl
361 qdakneesse ngfdmvdvhv feksnydlnl maspgdemli klaynenvfd eafilrlksq
421 lltaiqqliq npdqpvstin lvddrerefl ltglnppaqa hetkpltywf keavnanpda
481 paltysgqtl syreldeean riarrlqkhg agkgsvvaly tkrslelvig ilgvlkagaa
541 ylpvdpklpe drisymlads aaacllthqe mkeqaaelpy tgttlfiddq trfeeqasdp
601 ataidpndpa yimytsgttg kpkgnittha niqglvkhvd ymafsdqdtf lsvsnyafda
661 ftfdfyasml naarliiade htlldterlt dlilqenvnv mfattalfnl ltdagedwmk
721 glrcilfgge rasvphvrka lrimgpgkli ncygptegtv fatahvvhdl pdsisslpig
781 kpisnasvyi lneqsqlqpf gavgelcisg mgvskgyvnr adltkekfie npfkpgetly
841 rtgdlarwlp dgtieyagri ddqvkirghr ieleeiekql qeypgvkdav vvadrhesgd
901 asinaylvnr tqlsaedvka hlkkqlpaym vpqtftflde lplttngkvn krllpkpdqd

SrfAC CP-TE region (active sites in bold)

961 qlaeewigprnemeetiaqiwsevlgrkqigihddffalg**ghslk**amtavphqqelgidl
1021 pvkllfeaptiagisaylknggsdglqdvtimnqdqeqiifafppvlgyglmyqnlssrl
1081 psyklcafdfieeedrldryadliqklqpegpltlf**gysag**cslafeaakkleeqgrivq
1141 riimvdsykkqgvsdldgrtvesdvealmnvnrdnealnseavkhglkqkthafysyyvn
1201 listgqvkadidlltsgadfdmpewlasweeattgvyrvkrgfgthaemlqgetldrnae
1261 illeflntqtvtvs

Purple = peptide sequence coverage received for proteome treated with **1**/Sfp and band corresponding to FP labeling excised and subjected to tryptic digest
Green = peptide sequence coverage received for proteome treated with **3** and excised
Red = peptide sequence coverage received from both MS/MS spectra



Full Gel Data for Figures 3 and 4

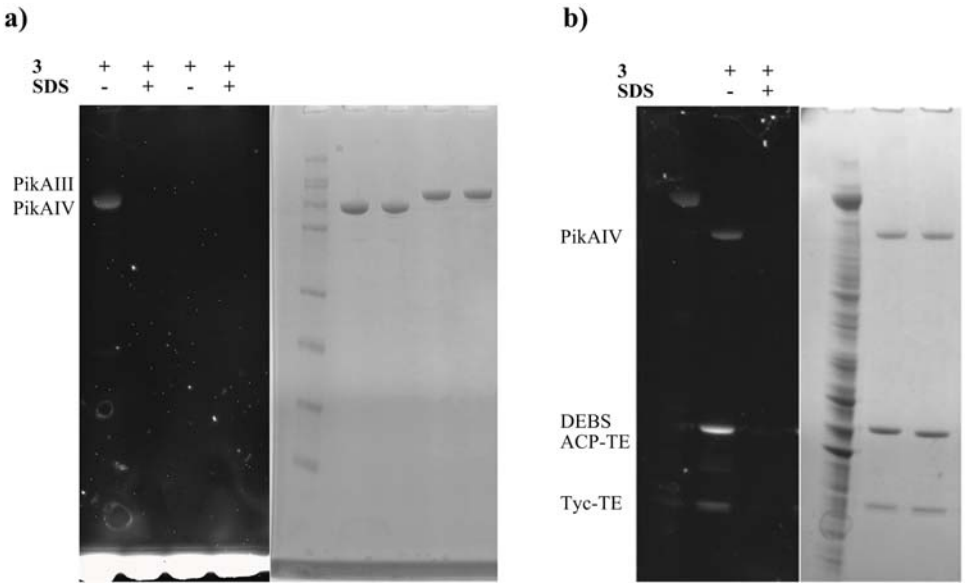


Figure 3A: In vitro labeling of purified recombinant PKS and NRPS acyltransferase (AT) and thioesterase (TE) domains by fluorophosphonate **3** (25uM). Left: fluorescence. Right: Coomassie. (A) Active site dependent labeling PikAIV and PikAIII. The lack of labeling of PikAIII lanes suggests AT-type serine hydrolases are not susceptible to this probe. (B) active site dependent labeling of a mixture of PKS and NRPS TE including Tyc-TE. PikAIV = module 4 of pikromycin PKS, containing AT, CP, KS, and TE domains. DEBS ACP-TE = excised ACP-TE didomain from 6-deoxyerythronolide synthase. Tyc-TE = excised tyrocidine NRPS TE domain from TycC.

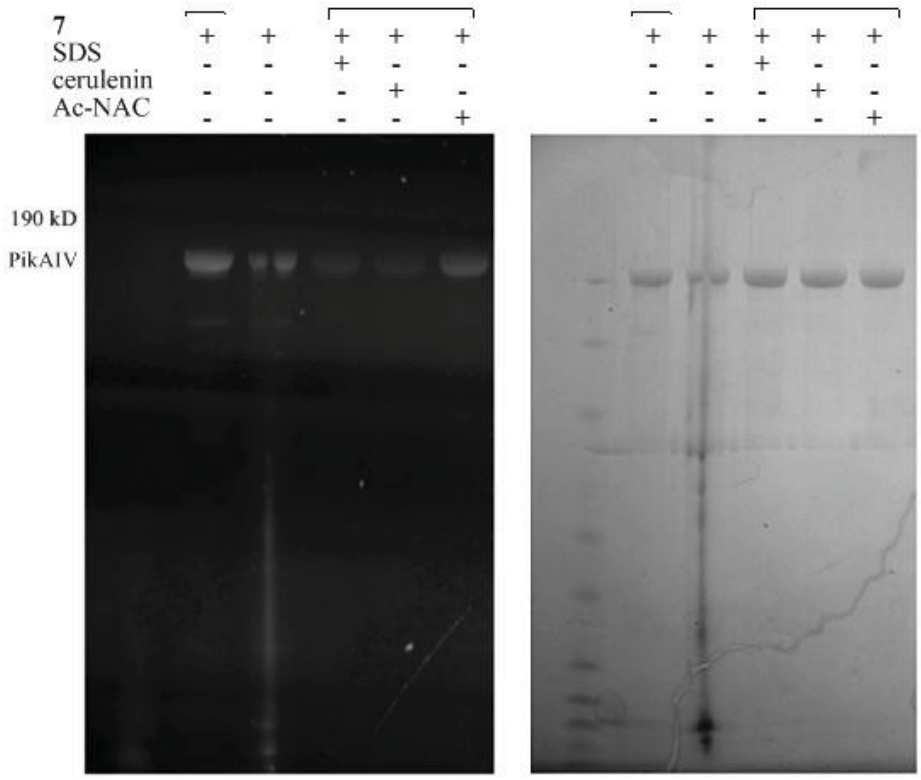


Figure 3B: Evidence for the site specificity of KS-probe **7**. Labeling of purified recombinant PikAIV in the presence or absence of SDS (0.5%) cerulenin (1mM), or Ac-NAC (7.5mM). Lane 2 was misloaded, as shown in the Coomassie stain. Lanes used in figure 3B are bracketed. Left: fluorescence. Right: Coomassie.

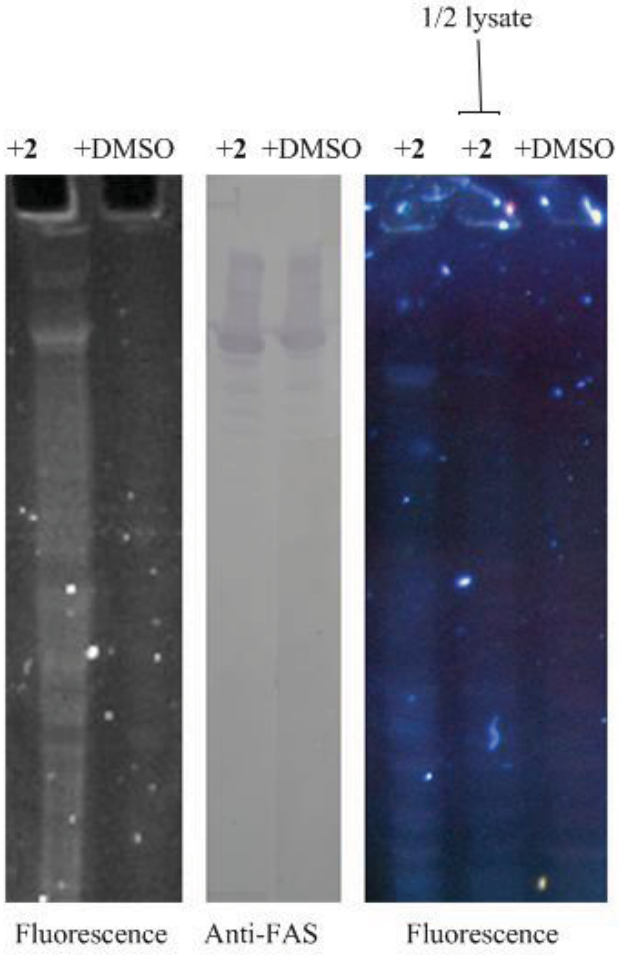


Figure 4A: In vivo metabolic labeling of FAS in SKBR3 cells by probe **2**. Left: fluorescence. Middle: Anti-FAS blot. Right: effect of gel loaded protein concentration on fluorescence.

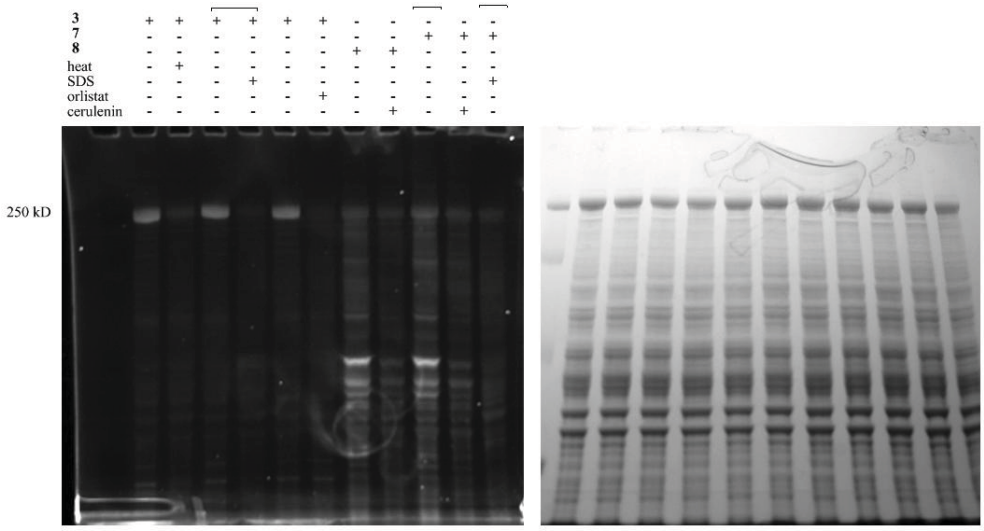


Figure 4A: Labeling of SKBR3 lysate by probes **3**, **7**, and **8**. The effect of heat, SDS, and preincubation with known FAS inhibitors (orlistat – TE, cerulenin – KS) were studied. The lack of labeling by fluorophosphonate **3** in the presence of known TE-directed agent orlistat provides further evidence for the lack of reactivity of this probe with AT domains. Lanes used in Figure 4A are bracketed. Left: fluorescence. Right: Coomassie.

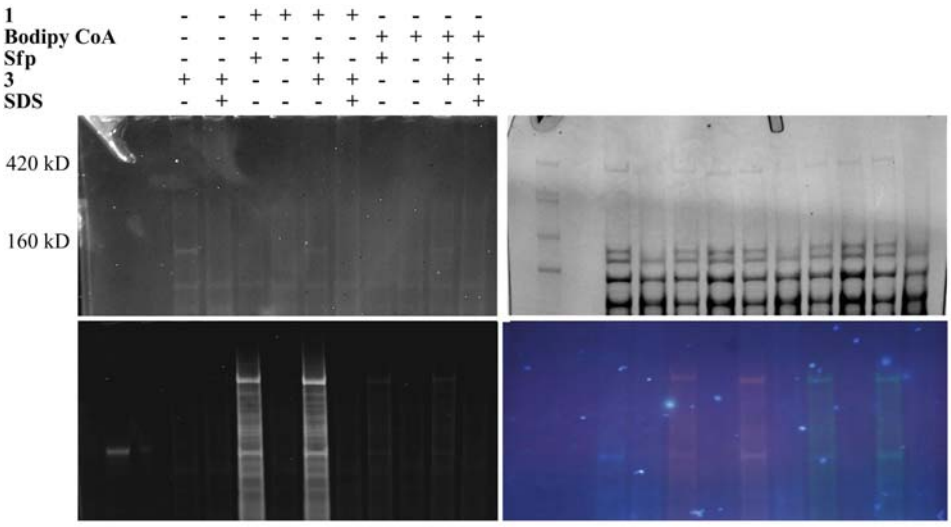


Figure 4B: 3-8% SDS-PAGE analysis of chemoenzymatic and activity-based labeling of HMW proteins from *B. subtilis* strain 168 grown in LB Media. Fluorophosphonate **3** (labeled lanes 1+2) shows an SDS-sensitive band just below 160 kD corresponding to SrfAC. This band is also visible upon administration of **1**/Sfp (lanes 3+4). Note the distinctive signal observed in the color gel upon labeling of SrfAC with both **1** and **3** in labeled lane 5. Addition of SDS caused some pertubation of the banding on the 3-8% gel as evidenced by Coomassie. This effect was not observed in the 4-12% SDS-PAGE of these proteins (see below) or any other **3** treated samples throughout this study.

Top left: fluorescence (UV, 460 nM emission filter, coumarin).

Top right: Coomassie

Bottom left: fluorescence (Typhoon, 532 nM excitation, TAMRA).

Bottom right:color photograph of fluorescence observed upon 360 nM irradiation

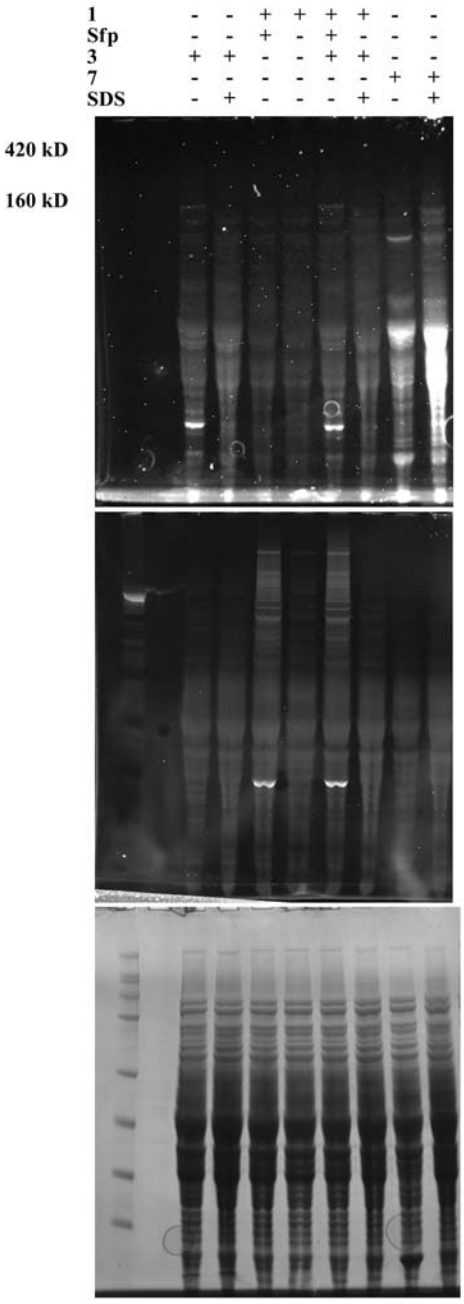


Figure 4B (continued): 4-12% SDS-PAGE analysis of chemoenzymatic and activity-based labeling of strain 168 grown in LB Media. Fluorophosphonate **3** (labeled lanes 1+2) shows an SDS-sensitive band just below 160 kD corresponding to SrfAC. This band is also visible upon administration of **1**/Sfp (lanes 3+4). Equal loading is observed by Coomassie.

Top: fluorescence (UV, 460 nM emission filter, coumarin).

Middle: fluorescence (Typhoon, 532 nM excitation, TAMRA).

Bottom: Coomassie

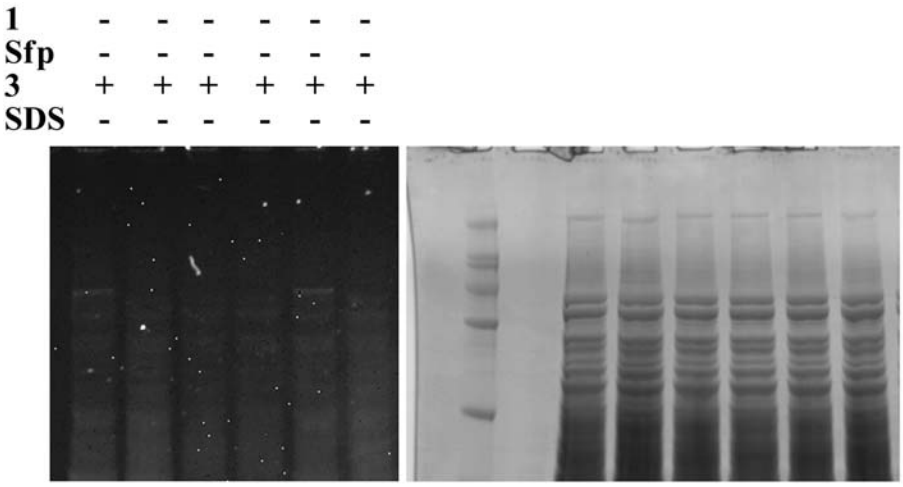


Figure 4B (zoom): 4-12% SDS-PAGE analysis of chemoenzymatic and activity-based labeling of strain 168 grown in LB Media. Fluorophosphonate **3** (labeled lanes 1+2) shows an SDS-sensitive band just below 160 kD corresponding to SrfAC. This band is also visible upon administration of **1**/Sfp (lanes 3+4). Equal loading is observed by Coomassie.

Left: fluorescence (UV, 460 nM emission filter, coumarin).

Right: Coomassie

MASCOT Report for MS/MS analysis of excised gel bands from 168 labelings.

gi|16077420|ref|NP_388239.1| Mass: 143727 Score: 371 Queries matched: 8
surfactin synthetase [Bacillus subtilis subsp. subtilis str. 168]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
23	683.9018	1365.7890	1365.7027	63.2	0	87	2.4e-008	1	K.ELLFSLSEAETK.A
31	733.4478	1464.8810	1464.8049	52.0	0	85	2.5e-008	1	K.GNITTEANIQGLVK.H
51	619.0339	1854.0800	1853.9424	74.2	0	63	5.4e-006	1	K.QIGIHDDFFALGGHSLK.A
59	702.4154	2104.2244	2104.0953	61.4	0	53	4.7e-005	1	R.LLPKPDQDQLAEEWIGPR.N
64	741.7627	2222.2662	2222.1331	59.9	0	57	2.7e-005	1	R.YQQSGDLAPGTVVSGRPAEIK.G
67	761.0776	2280.2110	2280.0910	52.6	0	32	0.0094	1	R.QFHIEEIDLTHLTGSEQTAK.I + Deamidated (NQ); Gln->pyro-Glu
69	780.1124	2337.3153	2337.1237	82.0	0	96	3.4e-009	1	K.EAVNANPDAPALTYSGQTLSYR.E
76	916.8257	2747.4553	2747.2919	59.5	1	46	0.00037	1	R.TVESDVEALMNVNRDNEALNSEAVK.H + Deamidated (NQ)

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : Trypsin
Variable modifications : Deamidated (NQ),Gln->pyro-Glu (N-term Q),Oxidation (M)
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance: ± 0.12 Da
Max Missed Cleavages   : 2
Instrument type        : ESI-QUAD-TOF
Number of queries      : 78
```

gi|16077420|ref|NP_388239.1| Mass: 143727 Score: 423 Queries matched: 9
surfactin synthetase [Bacillus subtilis subsp. subtilis str. 168]

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Peptide
11	548.8404	1095.6663	1095.5998	60.7	1	47	0.00011	1	R.IINVDSYKK.Q
12	551.3315	1100.6485	1100.5866	56.3	0	58	2e-005	1	K.AGAAYLPVDPK.L
30	689.3864	1376.7583	1376.6725	62.4	0	51	0.0001	1	R.WLPDGTIEYAGR.I
38	718.4180	1434.8215	1434.7466	52.2	2	75	4e-007	1	K.INEYKEQDKIR.G
39	733.4555	1464.8964	1464.8049	62.5	0	94	2.7e-009	1	K.GNITTHANIQGLVK.H
49	856.0048	1709.9950	1709.8777	68.6	0	64	4.1e-006	1	K.FIENPFKPGETLYR.T
53	881.4558	1760.8971	1760.8152	46.5	0	64	6.2e-006	1	R.GFGTHAENLQGETLDR.N
61	702.4305	2104.2695	2104.0953	82.8	0	53	3.6e-005	1	R.LLPKPDQDQLAEEWIGPR.N
69	1057.2393	3168.6961	3168.5363	50.4	0	92	7.9e-009	1	R.LQEQSLQSEPEQYVPLYDIQSQADQPK.L

Search Parameters

```
Type of search         : MS/MS Ion Search
Enzyme                 : Trypsin
Variable modifications : Deamidated (NQ),Gln->pyro-Glu (N-term Q),Oxidation (M)
Mass values            : Monoisotopic
Protein Mass           : Unrestricted
Peptide Mass Tolerance : ± 100 ppm
Fragment Mass Tolerance: ± 0.12 Da
Max Missed Cleavages   : 2
Instrument type        : ESI-QUAD-TOF
Number of queries      : 69
```

See MS section for details. Top: peptide sequence coverage received for proteome treated with **3** and excised. Bottom: peptide sequence coverage received for proteome treated with **1**/Sfp and excised.

References:

1. Baruch, A.; Verhelst, S.H.; Bogyo, M., An Improved Preparation of the Activity-Based Probe JPM-Oet and In Situ Applications. *Synthesis* **2005**, *2*, 240-244.

2. La Clair, J.J.; Foley, T.L.; Schegg, T.R.; Regan, C.M.; Burkart, M.D. Manipulation of Carrier Proteins in Antibiotic Biosynthesis. *Chem. Biol.* **2004**, *11*, 195–201.

3. Clarke, K.C.; Mercer, A.C.; LaClair, J.J.; Burkart, M.D., In Vivo Reporter Labeling of Proteins via Metabolic Delivery of Coenzyme A Analogues. *J. Am. Chem. Soc*. **2005**, *127*, 11234 – 11235.

4. Alexander, M. D.; Burkart, M. D; Leonard, M. S.; Portonovo, P.; Liang, B; Ding, X.; Joullie, M. M.; Gulledge, B. M.; Aggen, J. B.; Chamberlin, A. R.; Sandler, J.; Fenical, W.; Cui, J.; Gharpure, S. J.; Polosukhin, A.; Zhang, H.-R.; Evans, P. A.; Richardson, A. D.; Harper, M. K.; Ireland, C. M.; Vong, B. H.; Brady, T. P; Theodorakis, E. A.; La Clair, J. J. A Central Strategy for Converting Natural Products into Fluorescent Probes. *ChemBioChem* **2006**, 7, 409-416.

5. Castilho, M.S.; Pavao, F.; Oliva, G.; Ladame, S.; Willson, M.; Perie, J. Evidence for the Two Phosphate Binding Sites of an Analogue of the Thioacyl Intermediate for the *Tyrpanosoma cruzi* Glyceraldehyde-3-phosphate Dehydrogenase-Catalyzed Reaction, from Its Crystal Structure. *Biochemistry* **2003**, *42*, 7143-7151.

6. Ritter, S.C.; Konig, B. Signal Amplification and transduction by photo-activated catalysis. *Chem. Commun.* **2006**, 4694-4696.

7. Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. Synthesis and Evaluation of Bioorthogonal Pantetheine Analogues for in Vivo Protein Modification. *J. Am. Chem. Soc.* **2006**, *128*, 12174-12184.

8. Gottlieb, H.E.; Kotlyar V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents As Trace Impurities. *J. Org. Chem.* **1997**, *62*, 7512-7515.

9. Still, W.C.; Kahn, A.; Mitra, A. Rapid Chromatographic Technique For Preparative Separations With Moderate Resolution. *J. Org. Chem.* **1978**, *43*, 2923-2925.

10. Kidd, D.; Liu, Y.; Cravatt, BF. Profiling serine hydrolase activities in complex proteomes. *Biochemistry*. **2001**, *40*, 4005-4015.

11. Zhang, D.; Saraf, A.; Kolasa T.; Bhatia P.; Zheng, G.Z.; Patel, M.; Lannoye, G.S.; Richardson, P.; Stewart, A.; Rogers, G.C.; Brioni, J.D.; Surowy, C.S. Fatty acid amide hydrolase inhibitors display broad selectivity and inhibit multiple carboxylesterases as off-targets. *Neuropharmacology*. **2007**, *52*, 1095-1105.

12. Straight, P.D.; Fischbach, M.A.; Walsh, C.T.; Rudner, D.Z.; Kolter, R. A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*. **2007**, *104*, 305-310.

13. Thompson, B.J.; Stern, A.; Smith, S. Purification and properties of fatty acid synthetase from a human breast cell line. *Biochim Biophys Acta.* **1981**, *662*, 125–130.

Acknowledgments

The text of chapter 6, in full, is a reprint of material as it appears in: Meier, J.L., Mercer, A.C., and Burkart, M.D. (2008). Fluorescent profiling of modular biosynthetic enzymes by complementary metabolic and activity based probes. *J. Am. Chem. Soc. 130*, *5443-5445*. Permission was obtained from all co-authors. I was the primary author of this work. Andrew Mercer and Elinore Mercer performed in vivo labeling experiments in SKBR3 cells. I synthesized all compounds, performed all assays, and wrote the manuscript. This research was performed under the supervision of Prof. Michael Burkart.

Chapter Seven

The unusual macrocycle forming thioesterase of mycolactone

Originally published as:

Meier, J. L.; Barrows-Yano, T.; Foley, T. L.; Wike, C. L.; Burkart, M. D. *Mol Biosyst* **4**, 663-671 (2008).

PAPER www.rsc.org/molecularbiosystems | Molecular BioSystems

The unusual macrocycle forming thioesterase of mycolactone†‡

Jordan L. Meier,§ Tiffany Barrows-Yano,§ Timothy L. Foley, Candice L. Wike and Michael D. Burkart*

Received 29th January 2008, Accepted 12th March 2008
First published as an Advance Article on the web 16th April 2008
DOI: 10.1039/b801397g

Mycolactone is a polyketide natural product secreted by *Mycobacterium ulcerans*, the organism responsible for the tropical skin disease Buruli ulcer. The finding that this small molecule virulence factor is sufficient to reconstitute the necrotic pathology associated with Buruli ulcer suggests that a better understanding of mycolactone biosynthesis, particularly the processes which are distinct from those in human metabolism, may provide a unique avenue for the development of selective therapeutics. In the present study we have cloned, expressed, and biochemically characterized the putative macrocycle forming thioesterase for mycolactone, MLSA2 TE. We have evaluated the enzyme both as the truncated thioesterase domain and as a carrier protein-linked didomain construct. The results of these analyses distinguish MLSA2 TE from traditional fatty acid and polyketide synthase TE-domains in terms of its sequence, kinetic parameters, and susceptibility to traditional active-site directed inhibitors. These findings suggest that MLSA2 TE utilizes a unique biochemical mechanism for macrocycle formation.

Introduction

Buruli ulcer is a skin disease caused by *Mycobacterium ulcerans* associated with the necrosis of subcutaneous tissue leading to massive skin ulcerations.[1] Incidences of the disease have been reported in over 30 countries with the highest incidence occurring in tropical, humid areas of Africa. *M. ulcerans* now stands as the third most prevalent mycobacterial infection, behind *M. tuberculosis* and *M. leprae*.[2] In clinical trials the disease has shown some response to contemporary anti-mycobacterial antibiotics such as rifampicin, but the occurrence of the disease in rural communities with limited medical facilities and delayed diagnosis have led to surgical excision of infected tissue being the most common treatment.[3,4]

M. ulcerans is unique among mycobacterial pathogens in that it secretes a polyketide toxin, mycolactone **1**, which has been shown to reproduce the pathology associated with infection in a guinea pig model.[5] This small molecule virulence factor has cytotoxic, immunosuppressive, and analgesic properties, which result in the unusual painless nature of Buruli ulcer lesions. Mycolactone consists of an ester-linked polyketide chain attached to a 12-membered macrolide core and is produced by three giant plasmid-encoded polyketide synthase (PKS) enzymes: MLSA1 and MLSA2, which are responsible for production of the core, and MLSB, which synthesizes the southern chain (Fig. 1).[6]

PKS enzymes are structurally and functionally similar to the type I fatty acid synthase (FAS) which is responsible for fatty acid biosynthesis in animals, but differ in substrate utilization, number of steps, degree of reduction, and mechanism of chain release in important ways which lead to the structural diversity seen in PKS natural products such as mycolactone.[7] These differences represent unique druggable targets, an increased understanding of which would complement the steady advances in the genomic understanding,[8] diagnostic methods,[9] and investigations into mode of transmission[10] of this neglected tropical disease.

The plasmid-encoded polyketide synthase (PKS) responsible for the biosynthesis of mycolactone closely follows the canonical multimodular PKS paradigm but has some unusual features, such as the large size of MLSA1 and MLSB (1.8 and 1.2 MDa, respectively) and an unusually high interdomain sequence identity.[6] The large size of the enzymes responsible for the biosynthesis of mycolactone make them prohibitively difficult to characterize structurally or functionally as recombinant enzymes. To date only the discrete recombinant KR domain has been studied, with results indicating a lack of substrate promiscuity for the discrete enzyme and underlining the importance of substrate tethering and acyl carrier protein (ACP) interactions.[11]

With this in mind we have turned our attention towards the terminal thioesterase (TE) domain of MLSA2. MLSA2 TE is presumed to catalyze the formation of the mycolactone core (**2**) by intramolecular attack of the C-11 hydroxyl group on an appropriately activated acyl–enzyme intermediate (Fig. 1). This process is of interest therapeutically as it represents a mode of chain release distinct from that catalyzed by the mammalian FAS TE, a better understanding of which will aid efforts at constructing inhibitors designed to inactivate the biosynthetic machinery responsible for producing mycolactone. Past structure–activity relationship studies of mycolactone have shown that the

Department of Chemistry and Biochemistry, University of California-San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358. E-mail: mburkart@ucsd.edu
† This article is part of a *Molecular BioSystems* 'Emerging Investigators' issue highlighting the work of outstanding young scientists at the chemical- and systems-biology interfaces.
‡ Electronic supplementary information (ESI) available: Synthesis of fluorophosphonate **3**, kinetic data and gels for MLSA2 TE monodomain, complete gel images for Fig. 4 and 5, and sequence coverage for MS Analysis of MLSA2 ACP-TE active site labeling experiments. See DOI: 10.1039/b801397g
§ These authors contributed equally to this work.

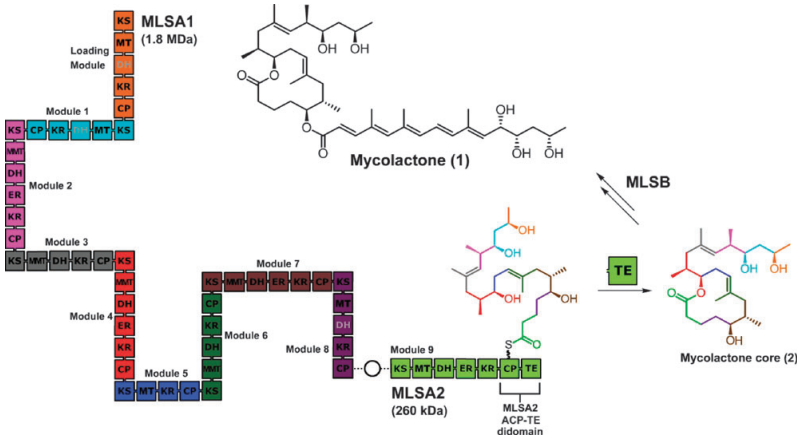


Fig. 1 Domain and module organization of the mycolactone core biosynthetic enzymes MLSA1 and MLSA2. MLSA2 TE is presumed to catalyze the cyclization of the macrolactone core by intramolecular formation of an ester bond. The structure of the mycolactone core has been color coded to specify the module responsible for each cycle of chain extension.

macrolide core produced by MLSA1 and MLSA2 is sufficient for cytopathicity, while the hydrolyzed open chain *seco*-acid is biologically inactive, demonstrating the importance of the macrolide core to the molecule's mode of action.[12] Here we explore the biosynthetic mechanism of mycolactone TE-catalyzed macrocyclization by describing the heterologous expression, purification, kinetic characterization, and substrate specificity of a previously unstudied TE from the mycolactone synthase MLSA2. Our findings differentiate the MLSA2 TE from traditional FAS and PKS TE-domains in terms of its kinetic parameters and vulnerability to traditional active-site directed inhibitors, suggesting that MLSA2 TE utilizes a non-canonical mode of macrocyclization in the biosynthesis of this polyketide virulence factor.

Results

Primary sequence analysis of MLSA2 TE

Initial BLAST analysis of the MLSA2 TE sequence revealed strong homology with other polyketide TE domains which catalyze similar macrolactonization reactions. However sequence comparison with six disparate macrocycle-forming TE domains demonstrated several interesting differences (Fig. 2). In each of these TE domains, two of which have published crystal structures, a serine hydrolase fold is present, and a Ser-His-Asp catalytic triad is observed.[13] Our analysis revealed that MLSA2 lacks the terminal Gly residue of the consensus Gly/Ala-X-Ser-X-Gly. This motif, characteristic of thioesterases, is associated with the nucleophilic elbow of the α/β hydrolase family.[14] Perhaps even more strikingly, while candidate residues S126 and D153 corresponding to the catalytic Ser and Asp of the catalytic triad were readily identified by alignment, a candidate for the catalytic His was not observed within the sequence. Overall, the region of the protein C-terminal to Y192 gave poor alignment. Only two His residues are present in this region of the protein, H208 and H212, and neither of these could be confidently aligned with the catalytic His of the other TE domains due to excessive gap lengths that would disrupt the putative fold.

To improve the alignment of the C-terminus, we turned to secondary structure prediction to identify motifs that may be used to anchor the sequence. This identified an α-helical structure spanning P223 to M242. This region contained a conserved AxxxAxxhxxWh motif that was not apparent from the initial analysis. Subsequently, this motif was anchored with DIALIGN and used to give the final alignment shown in Fig. 2. The finalized alignment suggests a Pro residue has replaced the His of the catalytic triad, a mutation that could occur by a single base mutation at the DNA level. Analysis of the conserved regions within MLSA2 clearly defines putative secondary structures, except for α7, which has recently been reported to be involved in linear ketide interaction and macrocyclization.[13,15] In addition sequence comparison of the MLSA2 and MLSB TE domains themselves (presumed to catalyze macrolactonization and hydrolysis, respectively) showed 100% sequence identity in the first 244 residues of the domain, differing only in the terminal 24 residues of MLSA2 and 10 residues of MLSB (data not shown).

Cloning, expression, and purification of MLSA2 ACP-TE didomain

To explore the questions raised by the sequence data and probe what effect, if any, interdomain communication may play in the mycolactone macrolactonization mechanism, the ACP-TE didomain and discrete TE monodomain were cloned out of *mlsa2* using genomic DNA from *M. ulcerans* with primers based on the previously reported sequence.[6] Gene products were amplified by PCR and digested with NcoI and HindIII before ligation into pET28b expression vectors in frame with a C-terminal 6x histidine tag. To account for the high occurrence of rare codons, constructs were expressed in Rosetta 2 *E. coli* cells (Strategene). While the discrete monodomain MLSA2 TE was attained in good yield using standard growth in LB medium at 37 °C with IPTG induction, similar conditions yielded only small quantities of

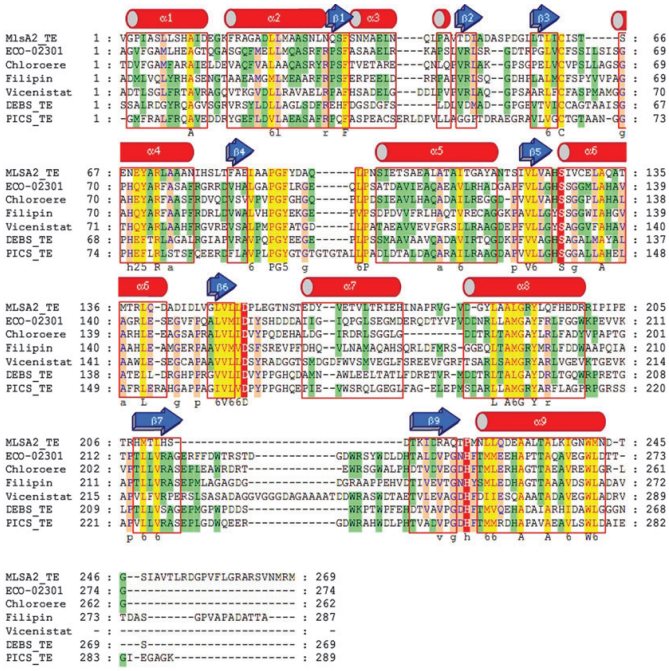


Fig. 2 Partial sequence alignment between MLSA2 TE, MLSB TE, and various macrolactonization catalysts from annotated polyketide synthases. The MLSA2 TE sequence was aligned with seven disparate polyketide-terminating thioesterase domains and colored for homology as follows: catalytic residues red; conserved identity, yellow; conserved similarity at 80% to consensus, pink; and conserved similarity at 60%, green. Secondary structural information was extracted from the DEBS-TE crystal structure (PDB identifier: 1KEZ) and demonstrates that regions of high conservation are contained within secondary structural elements. Notable variances in the MLSA2 TE include the absence of glycines flanking the nucleophilic serine S147 (positions 145 and 149 of the MLSA2 TE sequence), the latter of which constitutes the oxyanion hole. Sequences above are abbreviated, and their PubMed accession numbers follow: MLSA2 TE (YP_025561); ECO-02301(AAX98192); Chloroere: chloroeremomycin (ABM47024); Filipin (BAC68125); Vicenistat: vicenistatin (BAD08360); PICS-TE: picromycin (AAC69332); DEBS-TE: 6-deoxyerythronolide A (X56107).

soluble MLSA2 ACP-TE didomain protein, leading us to pellet and collect the insoluble fraction for purification from inclusion bodies. A brief survey of the literature revealed refolding conditions for an esterase from *M. tuberculosis* with 37% sequence identity to MLSA2 TE.[16] After resuspension, centrifugation, and Ni-NTA purification of the insoluble fraction, it was found that slow dilution of the urea denatured protein into 50 mM Tris-HCl pH 8.0 resulted in pure MLSA2 ACP-TE didomain (Fig. 3). This refolded protein was of equal hydrolytic activity to the small quantities of MLSA2 ACP-TE originally obtained as soluble protein, indicating that our purification protocol did not disrupt activity. For comparison studies the discrete monodomain TE from the 6-deoxyerythronolide (DEBS) PKS system was also overexpressed and purified as previously described.[17,18]

Kinetic characterization and substrate specificity of MLSA2 ACP-TE didomain

TE activity was monitored by the hydrolysis of *p*-nitrophenyl propionate (pNPP) using a variation of the procedure of Leadlay *et al.* for the assay of the DEBS ACP-TE didomain.[19,20] PNP esters are well known substrates of serine esterases which lend themselves to continuous spectrophotometric monitoring and avoid the use of thiol-modifying reagents such as 5,5′-dithio-bis-2-nitrobenzoic acid, which have been shown in the past to rapidly inactivate oxidation sensitive thioesterases.[21]

The observed steady state parameters are summarized in Table 1. MLSA2 ACP-TE didomain displayed saturation kinetics with pNPP, giving a K_m of 1.63 mM and k_{cat}/K_m of 0.20 min^{-1} mM^{-1} with pNPP (Table 1 and Fig. 3). While similar to values previously reported in the literature,[22,23] comparison with DEBS monodomain showed that MLSA2 TE catalyzed substrate hydrolysis nearly 20× slower than DEBS TE. To verify that this effect was not a result of misfolding or interference due to linkage of the ACP to MLSA2 TE, kinetic analyses were also performed using the monodomain TE, which demonstrated similar behavior, although the protein did not display saturable catalysis at the limit of substrate solubility (Supplementary Information‡). We investigated this phenomenon further *via* gel filtration chromatography and found that the monodomain TE exists

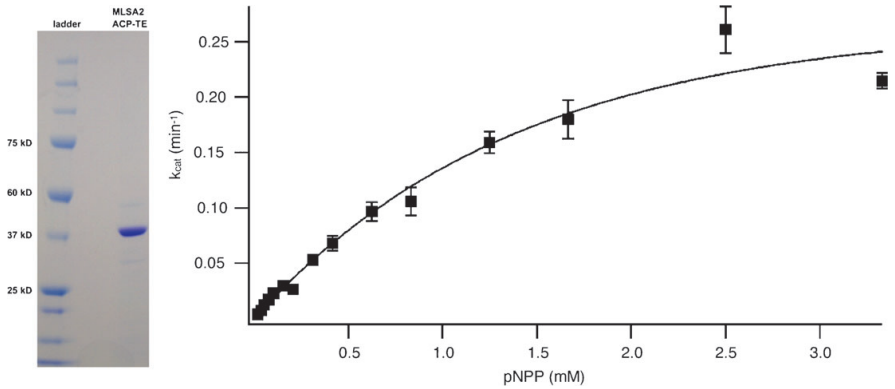


Fig. 3 Purification of the heterologously expressed MLSA2 ACP-TE didomain. (A) Lane 1 molecular weight markers; lane 2 purified MLSA2 ACP-TE didomain after purification from inclusion bodies. (B) Saturation kinetics of recombinant MLSA2 ACP-TE didomain with pNPP.

as a homodimer, consistent with the commonly observed multimerization state for PKS thioesterases. This indicated that the kinetic performance of MLSA2 TE in the pNPP assay did not result from interruption of the dimerization interface by improper choice of an N-terminal domain boundary point.

Further differences between the MLSA2 and DEBS TEs were displayed in the substrate specificity and pH profile of the two enzymes (Tables 2 and 3). While neither enzyme showed any significant hydrolytic activity toward the *p*-nitrophenyl anilide substrate, MLSA2 TE showed a modest selectivity toward shorter chain pNP esters, while DEBS showed a distinct preference for pNP caproate. Both TE domains showed little activity with pNP laurate, although this observation may be skewed by the low solublility of the substrate in this case.

The pH profile of pNPP hydrolysis was examined for both TEs over a limited range (Table 3). As previously reported, DEBS showed increasing activity at alkaline pH, with a four-fold increase in activity at pH 9.0 compared with neutral pH. MLSA2 ACP-TE didomain meanwhile demonstrated an opposite trend, demonstrating a 50% decrease in apparent k_{cat} over the same range. This is contrary to the trend observed in canonical serine proteases, in which acidic pH decreases the nucleophilicity of the catalytic serine.[21,24]

Phosphopantetheinylation of the recombinant carrier protein active site of MLSA2 ACP-TE didomain

An advantage to studying the MLSA2 TE as the didomain construct is the opportunity to analyze the activity and influence of the carrier protein domain on TE activity. A typical assay for carrier protein folding and activity is PPTase modificiation.[25] To investigate whether the carrier protein of MLSA2 was properly folded for native activity, we assayed for post-translation modification of recombinant *apo*-carrier protein (Fig. 4). Upon incubation of purified MLSA2 ACP-TE didomain with a fluorescent BODIPY-CoA analogue and the promiscuous PPTase Sfp (under previously described conditions), we verified the PPTase-dependent active site labeling of MLSA2 ACP.[26] The ability of the carrier protein domain to be phosphopantetheinylated by CoA analogues provides further evidence as to the competent folding of the MLSA2 ACP-TE didomain and suggests that this protein may be amenable to mechanistic analyses by recently proposed ACP-TE crosslinking strategies.[27,28]

Table 1 Kinetic parameters

Protein	k_{cat}/min^{-1}	K_m/mM	k_{cat}/K_m/min^{-1} mM^{-1}
MLSA2 ACP-TE	0.33 ± 0.02	1.63 ± 0.21	0.20
DEBS TE	6.4 ± 0.3	1.46 ± 0.03	4.40

Kinetic parameters for MLSA2 and DEBS TE-domains observed with pNPP at 405 nm.

Affinity labeling studies of the thioesterase active site

In addition to labeling of the ACP active site of the ACP-TE didomain, we also sought to verify the presence of serine esterase activity by affinity labeling of the TE active site. Small molecule fluorophosphonates are well-known class-wide inhibitors of serine hydrolases and have been utilized in the past to irreversibly modify and identify the active site Ser of both type I FAS and PKS TE domains.[17,29] This reactivity has proven itself amenable to gel-based activity assays,[30] as the attachment of a fluorophore to the fluorophosphonate active site

Table 2 Substrate specificity

Protein	pNP propionate	pNP caproate	pNP laurate	pNA (aniline)
MLSA2 ACP-TE	0.025 ± 0.001	0.022 ± 0.005	No activity	No activity
DEBS TE	5.33 ± 0.37	3071 ± 55	0.31 ± 0.04	No activity

Apparent k_{cat} (min^{-1}) of turnover of 280 μM ester substrates observed at 405 nm (aniline observed at 382 nm). Apparent k_{cat} values were generated at non-saturating concentrations of pNP substrates due to the low solubility of caproate and laurate substrates respectively.

Table 3 Limited pH profile

Protein	pH 7.0	pH 7.5	pH 8.0	pH 8.5	pH 9.0
MLSA2 ACP-TE	0.13 ± 0.01	0.112 ± 0.002	0.09 ± 0.01	0.08 ± 0.01	0.07 ± 0.01
DEBS TE	10.3 ± 0.4	11.1 ± 0.2	13.8 ± 0.2	35.3 ± 0.8	46.6 ± 1.3

Apparent k_{cat} (min^{-1}) of turnover of 280 μM substrate observed at the isobestic point of the *p*-nitrophenol/*p*-nitrophenoxide couple (λ = 346 nm, ε = 4800 M^{-1} cm^{-1}).

"warhead" allows a functional readout of serine hydrolase activity based on fluorescent labeling (Fig. 5A). In such assays intact active-sites are fluorescently labeled by the affinity probe while inhibitor-blocked or denatured active-sites are not.[31]

Consequently, we synthesized fluorescent fluorophosphonate **3** and tested the susceptibility of the MLSA2 ACP-TE and DEBS-TE to affinity labeling by this well-known covalent inhibitor. DEBS-TE showed strong labeling by **3**, which was disrupted upon denaturation by pre-heating of the sample at 80 °C for 5 minutes (Fig. 5C, bottom row). In stark contrast, MLSA2 ACP-TE showed only low-level labeling even at high enzyme concentrations, with no heat-dependence. To verify this unanticipated lack of reactivity, MLSA2 was preincubated with 50× fluorophosphonate for 1 hour and submitted to pepsin digest and MALDI-TOF-TOF analysis. No modification of the putative TE active site serine 126 was observed (Supplementary Information‡). Since the fluorescent nature of **3** interferes with our kinetic assay, we also tested the effect of preincubation of MLSA2 ACP-TE with phenylmethylsulfonyl fluoride (PMSF, **4**), another well-known small molecule inhibitor of thioesterases,[20] on pNPP turnover by this enzyme. Consistent with the results of the fluorophosphonate activity assay, no significant inhibition was observed.

Discussion

The secondary biosynthetic capacities of virulent organisms are often distinct from those of eukaryotes and represent an attractive target for therapeutics.[32] Such an approach would greatly benefit from a knowledge of the structure and enzymology of prospective enzyme targets. This strategy, along with our interest in the biosynthesis of PKS natural products, led us to study the thioesterification mechanism of the PKS virulence factor mycolactone. Our studies of the MLSA2 TE commenced with cloning and overexpression of the C-terminal TE monodomain and the ACP-TE didomain from *M. ulcerans*.

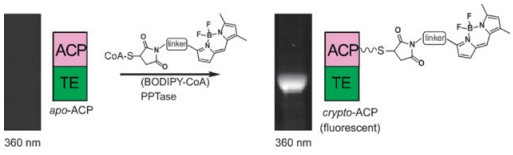


Fig. 4 Fluorescent labeling of MLSA2 ACP-TE ACP active site. Treatment of MLSA2 ACP-TE didomain (20 μM) with Sfp (0.5 μg), $MgCl_2$ (11 mM) and BODIPY-CoA (200 μM), results in fluorescent labeling of the carrier protein domain. Upon omission of Sfp, no labeling is observed.

Our choice to study the ACP-TE didomain was influenced by the previously inferred importance of ACP interactions in mycolactone biosynthesis,[11] as well as by studies of PKS TE domains which showed that covalent connection of the TE domain and ACP could both enhance the rate of polyketide chain release by as much as 100× compared to that observed with the excised TE domain[33] and avoid cleavage of important ACP-TE linker regions which may result in a loss of hydrolytic activity.[21] While MLSA2 TE monodomain proved attainable using standard protein overexpression and purificaiton conditions, we were forced to devise a refolding strategy for the purification of the ACP-TE didomain from inclusion bodies. This agrees with recent studies which have suggested that incorporation of an *apo*-ACP domain can result in insolubility of multidomain PKS proteins, presumably due to loss of the stabilizing effect of the 4′-phosphopantetheine posttranslational modification.[34] The protein purification and refolding methods demonstrated here may be useful for the preparation of *apo*-ACP domains for applications which utilize their site-specific labeling properties as mechanistic and structural probes.[27,34]

Fig. 5 Affinity labeling of MLSA2 ACP-TE TE active site. (A) Active-site directed inhibitors examined in this study, and (B) mode of action of fluorophosphonate labelling of nucleophilic serines. (C) Fluorescent gel image generated on treatment of MLSA2 ACP-TE didomain (20 μM) with fluorophosphonate **3** (50 μM) before or after denaturation. The equivalent labeling observed indicates that **3** does not label MLSA2 TE in an active-site dependent manner. DEBS-TE (2 μM) treated with fluorphosphonate **3** (50 μM) demonstrates the expected pattern for labeling of the catalytic serine of a thioesterase.

The purified ACP-TE didomain was an active, natively folded protein as verified by several criteria. This protein showed saturation kinetics in the pNPP assay and the ability to be phosphopantetheinylated by a fluorescent BODIPY CoA analogue and Sfp. Although interference of the conformationally flexible pendant ACP domain hindered multimerization studies of the didomain, a monodomain TE construct which demonstrated similar kinetic values in the pNPP assay was shown to form a discrete dimer. All of these properties are consistent with those observed in typical PKS thioesterases.[13] However, upon comparison with the monodomain DEBS TE, it was found that MLSA2 ACP-TE catalyzes the release of the *p*-nitrophenoxide anion at a rate nearly 20× lower than DEBS. Similarly, the pH profile and substrate specificity of MLSA2 ACP-TE differ from those usually observed, showing increasing activity at acidic pH and a slight preference for shorter alkyl pNP esters.

These findings were coherent with the overall picture of MLSA2 TE generated from sequence analyses. Sequence alignment revealed that MLSA2 TE lacks several normally conserved features of macrolactonization catalyzing thioesterases, including a Gly-X-Ser-X-Gly consensus sequence around the catalytic serine or a plausible candidate residue for the His of the catalytic triad. The finalized alignment suggests a proline residue has replaced the histidine of the catalytic triad, a mutation that could occur by a single base mutation at the DNA level. However, while past studies have suggested that the extreme level of sequence identity found in the mycolactone gene locus is indicative of a high level of genetic instability,[6,35] a single mutation seems insufficient to explain the gaps in primary and secondary structure observed in the C-terminus of MLSA2 TE.

One effective method for comparing the activity of proteins with divergent sequences is to examine their susceptibility to active-site directed inhibitors.[30] Since the reactivity of such inhibitors is promoted by conserved enzymatic motifs, they provide a direct report of enzyme activity which may not be immediately apparent from sequence alignment or traditional substrate turnover assays. Therefore in an attempt to further probe the mechanism and identify the active site residues that are important in the formation of the mycolactone core, we examined the reactivity of MLSA2 ACP-TE with the fluorescent fluorophosphonate activity based probe **3**. In contrast to the DEBS-TE monodomain, MLSA2 ACP-TE showed no signs of active-site dependent labeling as determined by sensitivity to heat or SDS treatment. Furthermore tandem MS analysis of MLSA2 ACP-TE treated with **3** showed no modification of Ser126, the presumed catalytic serine as identified by sequence alignment.

These results indicate that MLSA2 TE differs in important ways in both its kinetic and inhibitory profile from canonical PKS and FAS thioesterases. This finding is notable in itself, as it challenges the currently held dogma of mycolactone ring formation and highlights a biosynthetic process which is of potential medicinal relevance and currently poorly understood. Recent elucidation of cryptic cyclization methods in PKS biosynthesis has shown the value of crystallographic information, mutational analysis, and the study of hydrolysis of natural substrate NAC esters, and these will be the subject of future inquiry.[36–39] The preparation of linear substrate analogues should be greatly simplified by the recent publication of multiple modular syntheses of the mycolactone core.[40,41] Such studies are the first step toward developing chemotherapeutic strategies targeted directly at the biosynthetic capabilities of the microbe responsible for this fascinating and destructive disease.

Experimental procedures

Strains, culture, media and general materials

M. ulcerans DNA was a generous gift from Dr Pamela Small. Oligonucleotides were purchased from Allele Biotechnology (San Diego, CA). Standard protocols were applied for all DNA manipulations. Rosetta 2 *E. coli* were purchased from Stratagene. The T7-derived expression vector was obtained from Novagen. MALDI mass spectrometry was conducted at the UCSD Chemistry and Biochemistry Mass Spectrometry Core Facility by Justin Torpey.

Primary sequence analysis

A homology tree generated by alignment of the MLSA2, DEBS, and pikromycin/methymycin (PICS) TE sequences was used to guide selection of seven disparate sequences for alignment to avoid biasing from highly similar genes. Sequence analysis was performed using ClustalW.[42] Due to poor homology of the C-terminal portion of MLSA2 TE to other sequences, the MLSA2 TE sequence was submitted to the PsiPred[43] and J-Pred[44] servers to assess the secondary structure of the protein, and these were compared to outputs yielded for the DEBS-TE and PIK TE sequences. Predicted secondary structure for the C-terminal region of MLSA2 TE was used to identify disparately conserved motifs of α9–10. These were anchored and aligned using DIALIGN[45] and homologous regions colored using GeneDoc.[46] Secondary structural data was loaded into GeneDoc from the DEBS TE crystal structure (PDB identifier: 1KEZ),[18] and secondary structure designations follow the conventions from this report.

Construction of expression plasmids

The gene fragment encoding for MLSA2 ACP-TE (1200 bp) and MLSA2 TE monodomain (1000 bp) were amplified by PCR from genomic DNA of *M. ulcerans* MU1615 using Taq polymerase with primers based on the gene sequence reported by Cole *et al.*[6] A 1.2 kb gene product was produced for the MLSA2 ACP-TE didomain using the following primers: 5′-CAACAAGTTGGTACCAAACCATGGGCTCCACACT-GGCC-3′ (5′ primer) and 5′-GTCCAGAAGCTTCCTCCTC ATG TTGAC-3′ (3′ primer). The N-terminal domain boundary for MLSA2 TE mondomain was based on primary sequence analysis as described above. A 1.1 kb gene product was produced for the MLSA2 TE monodomain using the primers: 5′- CCGTGCCGGCCGTGACCTATCCATGG-CCGC -3′ (5′ primer) and 5′-GTCCAGAAGCTTCCTCCTC ATG TTGAC-3′ (3′ primer). The PCR products were cloned into the HindIII/NcoI site of pET28B (Novagen). All inserted fragments were confirmed by sequencing (Eton Biotechnology).

Expression and purification of recombinant enzymes

DEBS-TE was cloned and overexpressed from *E. coli* BL21 cells as previously reported.[18] For expression of MLSA2 TE monodomain *E. coli* BL21 cells were transformed with the desired plasmid and grown in Luria Bertani medium at 37 °C to OD 0.5 (600 nm). Protein expression was induced at 17 °C using 1 mM isopropyl β-D-thiogalactopyranoside and cultured for approximately 16 hr. Cells were pelleted and resuspended in lysis buffer (100 mM potassium phosphate, 300 mM NaCl [pH 8.0]) and incubated with lysozyme (1 mg mL^{-1}) for 1 hr followed by treatment with DNase and sonication. The cellular debris was separated from solublized MLSA2 TE by centrifugation and the supernatant was loaded onto a Ni-NTA column and eluted with a gradient of imidazole (5–300 mM). Similar expression conditions yielded only a small amount of protein when applied to the MLSA2 ACP-TE didomain plasmid. Due to the high occurrence of rare codons, the MLSA2 ACP-TE didomain was expressed in Rosetta 2 *E. coli* cells (Stratagene). Cells were transformed with the desired plasmid and grown in Luria Bertani medium at 37 °C to OD 0.6 (600 nm). Protein expression was induced using 1 mM isopropyl β-D-thiogalactopyranoside and cultivated for approximately 16 hr. Cells were pelleted and resuspended with three volumes of lysis buffer (100 mM potassium phosphate, 300 mM NaCl [pH 8.4]) before being frozen at −80 °C. Thawed cells were lysed by two passes through a french pressure cell, treated with DNaseI for 30 min at 0 °C, and the pellet was collected for purification of the didomain from inclusion bodies. The pellet was resuspended (6 M urea, 50 mM potassium phosphate, 300 mM NaCl [pH 8.4]) and incubated for 30 minutes at room temperature. Cell debris was separated from the solublized inclusion bodies by centrifugation and the supernatant was loaded onto a Ni-NTA column. The unfolded didomain was refolded by slow dilution into 50 mM Tris-HCl, pH 8.0. This protein was then extensively dialyzed against 100 mM Tris-HCl (pH 8.0) to remove excess urea and imidazole. For all enzymes purity was determined by SDS PAGE, and concentration was estimated using the Bradford assay.[47]

Native molecular weight evaluation of MLSA2 TE monodomain

The multerimization state of the standalone MLSA2 TE was determined by gel filtration chromatography performed with an FPLC system on a 1.6 × 60 cm Sephacryl S-100 column at a flow rate of 0.25 mL min^{-1} with a mobile phase consisting of 100 mM Tris-HCl (pH 8.0) and 300 mM NaCl. The column was calibrated with standards (Sigma-Aldrich # MW-GF-200) immediately prior to chromatography. Separation of MLSA2 TE in this fashion yielded a peak with a retention time corresponding to an apparent molecular weight of 62 000 Da, with no leading peak (corresponding to higher order multimerization or aggregate) or peak eluting with apparent molecular weight of 32 000 Da observed.

Determination of the kinetic parameters for TE-catalyzed hydrolysis

Hydrolysis of the *p*-nitrophenyl laurate, caproate, and propionate (pNPP) was monitored by following the rate of formation of the *p*-nitrophenoxide anion (λ_{max} = 400 nm, ε = 8570 M^{-1} cm^{-1}) at 30 °C. Hydrolysis of *p*-nitrophenyl anilide (pNA) substrates were monitored by following the rate of formation of *p*-nitrophenyl aniline (λ_{max} = 382 nm, ε = 13 500 M^{-1} cm^{-1}). In all cases substrate turnover was monitored on a Perkin-Elmer HTS 7000 Bio-Assay Reader. Each assay contained (in a total volume of 150 μL) 19.2 mM Tris-HCl (pH 8.0), 20 μM MLSA2 ACP-TE didomain or DEBS-TE, 1% ethanol, and a variable concentration of substrate in CH_3CN. Stock solutions of all substrates were 250 mM in CH_3CN. In all cases, the total content of CH_3CN in the reaction was adjusted to 1.3%. Reactions were carried out in triplicate and the rate of hydrolysis calculated from the initial linear portion of the curves. In all cases, the rates were corrected for the background rate of chemical hydrolysis in the absence of enzyme and analyzed by least squares analysis of the data. Hydrolysis rates for pNPP were measured over the concentration range 0.05–3.3 mM. Hydrolysis rates for laurate and caproate derivatives and pNA were measured at 0.28 mM using 0.1% Triton-X to aid substrate solubility. The pH dependence of the reaction was monitored at 1.25 mM pNPP in Tris-HCl buffer adjusted to the appropriate pH with NaOH or HCl and measured at the isobestic point of the *p*-nitrophenol/*p*-nitrophenoxide couple (λ = 346 nm, ε = 4800 M^{-1} cm^{-1}).

Determination of carrier protein active site labeling by BODIPY-CoA

BODIPY-CoA was prepared as previously described.[26] MLSA2 ACP-TE didomain (20 μM) was incubated with $MgCl_2$ (11 mM), BODIPY-CoA (200 μM), and Sfp (0.5 μg) in Tris-HCl pH 8.0 (64 mM) at 37 °C for 1 hour before quenching with 5× SDS loading buffer and separation by 1D SDS-PAGE with visualization by fluorescence and Coomasie Brilliant Blue.

Thioesterase active site affinity labeling experiments

The synthesis of fluorophosphonate **3** is detailed in the Supplementary Information.‡ MLSA2 ACP-TE didomain (20 μM) was incubated with **3** (50 μM) in Tris-HCl pH 8.0 (64 mM) at room temperature for 30 minutes. As a control this procedure was also performed on DEBS-TE monodomain (2 μM). Labeling reactions were quenched with 5× SDS loading buffer (strong reducing) and subjected to SDS-PAGE with visualization by fluorescence and Coomassie Brilliant Blue.

MS analysis of MLSA2 ACP-TE active site labeling experiments

MLSA2 was treated with 50× fluorophosphonate **3** (1000 μM) or DMSO (control) in Tris-HCl pH 8.0 at room temperature for 1 hour after which the reaction was quenched and the enzyme subjected to proteolytic digest by pepsin or trypsin or chymotrypsin. LC-MALDI was acquired on a Tempo LC MALDI spotter (MDS Sciex). A Chromolith CapRod monolithic capillary column (150 × 0.1 mm) (EMD, VWR International) packed with endcapped RP-18 was run at 3 μL min^{-1} on channel 1 using 98% buffer A (98% H_2O–2%

acetonitrile (ACN, Fisher Optima) with 0.1% acetic acid (AcOH, Fluka) and 0.005% heptafluorobutyric acid (HFBA, Fluka)) and 2% buffer B (98% ACN–2% H_2O with 0.1% AcOH and 0.005% HFBA). 5 μL were injected onto a 3 μL loop and after holding initial conditions for 3 min, a linear gradient was run with increasing buffer B from 2–30% over 10 min. The eluant was mixed with channel 2 running matrix solution at a flow rate of 3 μL min^{-1}. A matrix was made by mixing 2.0 mL α-cyano-4-hydroxycinnamic acid (Agilent) with 480 μL of H_2O and 60 μL of 5 mg ml^{-1} ammonium hydrogen citrate (Sigma) to give the required volume and concentration.

LC MALDI spots were acquired at a rate of 4 s $spot^{-1}$ for 8–16 minutes. 120 spots were obtained in the run in 3 rows. A plate model and default calibration were performed for reflector positive mode as well as a default calibration for 2 kV positive mode on the 4800 MALDI TOF/TOF Anaylzer (Applied Biosystems). Reflector positive spectra were obtained by acquiring 500 shots from m/z 1000–4000 amu. Tandem mass spectra (MS/MS) were acquired in 2 kV positive mode with CID on at resolution 200 (FWHM) for 3000 shots. Peaks occurring with minimum chromatogram peak width of 2 fractions, a minimum S/N filter of 40, and a fraction-to-fraction precursor mass tolerance of 100 ppm were submitted for MSMS acquired in the order of strongest precursors first with a max. precursors/fraction of 8. Sodium and potassium adducts were excluded with an adduct tolerance of ±0.03 Da, and precursors <200 resolution were excluded.

The MSMS were analyzed by Global Proteome Server 3.0 (Applied Biosystems) and subjected to a database search using Mascot 2.1.1 (Matrix Science). Data were searched against the recombinant MLSA2 ACP-TE sequence of interest. The search employed up to 2 missed cleavages and used mass tolerances of 100 ppm (MS) and 0.20 Da (MSMS), with variable modifications as follows: fluorophosphonate labeling of serine (S), deamidation (NQ), oxidation (M), pyro-Glu (N-term Q). The search results indicated that individual ion scores >42 indicate identity or extensive homology ($P < 0.05$).

Acknowledgements

This work was supported by the Ellison Medical Foundation, NSF CAREER Award 0347681 and NIH RO1GM075797. T.Y. was supported by a Ruth L. Kirschstein National Research Service Award 5T32 CA009523-22. J.M. was supported by NIH Training Grant T32DK007233. We thank Marie Tran, and Matt Alexander for preliminary work in cloning and assay development, Justin Torpey of the UCSD Mass Spectrometry Core for MS Analyses, and Dr Pamela Small of the University of Tennessee for providing *M. ulcerans* genomic DNA.

References

1 M. Wansbrough-Jones and R. Phillips, *Lancet*, 2006, **367**, 1849.
2 P. D. R. Johnson, T. Stinear, P. L. Small, G. Plushke, R. W. Merritt, F. Portaels, K. Huygen, J. A. Hayman and K. Asiedu, *PLoS Med.*, 2005, **2**, e108.
3 Buruli ulcer: *Mycobacterium ulcerans* infection, ed. K. Asiedu, R. Sherpbier and M. Raviglione, Geneva, 2000, WHO_CDS_CPE_GBUI_2000.1.pdf. Available: http://www.who.int/buruli/information/publications/en/index.html. Accessed: July 10th, 2007.
4 A. Chauty, M. F. Aradant, A. Adeye, H. Euverte, A. Guedenon, C. Johnson, J. Aubry, E. Nuermberger and J. Grosset, *Antimicrob. Agents Chemother.*, 2007, **51**, 4029.
5 K. M. George, D. Chatterjee, G. Gunawardana, D. Welty, J. Hayman, R. Lee and P. L. Small, *Science*, 1999, **283**, 854.
6 T. P. Stinear, A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay and S. T. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1345.
7 S. Smith and S. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041.
8 T. P. Stinear, T. Seemann, S. Pidot, W. Frigui, G. Reysset, T. Garnier, G. Meurice, D. Simon, C. Bouchier, L. Ma, M. Tichit, J. L. Porter, J. Ryan, P. D. Johnson, J. K. Davies and G. A. Jenkin *et al.*, *Genome Res.*, 2007, **17**, 192.
9 J. A. Fyfe, C. J. Lavender, P. D. Johnson, M. Globan, A. Sievers, J. Azuolas and T. P. Stinear, *Appl. Environ. Microbiol.*, 2007, **73**, 4733.
10 L. Marsollier, P. Brodin, M. Jackson, J. Kordulakova, P. Tafelmeyer, E. Carbonnelle, J. Aubry, G. Milon, P. Legras, J. P. Andre, C. Leroy, J. Cottin, M. L. Guillou, G. Reysset and S. T. Cole, *PLoS Pathog.*, 2007, **3**, e62.
11 S. Bali and K. J. Weissman, *ChemBioChem*, 2006, **7**, 1935.
12 A. Mve-Obiang, R. E. Lee, F. Portaels and P. L. Small, *Infect. Immun.*, 2003, **71**, 774.
13 F. Kopp and M. Marahiel, *Nat. Prod. Rep.*, 2007, **24**, 735.
14 M. Nardini and B. W. Dijkstra, *Curr. Opin. Struct. Biol.*, 1999, **9**, 732.
15 D. L. Akey, J. D. Kittendorf, J. W. Giraldes, R. A. Fecik, D. A. Sherman and J. L. Smith, *Nat. Chem. Biol.*, 2006, **2**, 537.
16 S. Canaan, D. Maurin, H. Chahinian, B. Pouilly, C. Durousseau, F. Frassinetti, L. Scappuccini-Calvo, C. Cambillau and Y. Bourne, *Eur. J. Biochem.*, 2004, **271**, 3953.
17 P. Caffrey, B. Green, L. C. Packman, B. Rawlings, J. Staunton and P. F. Leadlay, *Eur. J. Biochem.*, 1991, **195**, 823.
18 S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 14808.
19 K. J. Weissman, C. J. Smith, U. Hanefeld, R. Aggarwal, M. Bycroft, J. Staunton and P. F. Leadlay, *Angew. Chem., Int. Ed.*, 1998, **37**, 1437.
20 R. Aggarwal, P. Caffrey, P. F. Leadlay, C. J. Smith and J. Staunton, *J. Chem. Soc., Chem. Commun.*, 1995, **15**, 1519.
21 H. Lu, S. Tsai, C. Khosla and D. E. Cane, *Biochemistry*, 2002, **41**, 12590.
22 C. N. Boddy, T. L. Schneider, K. Hotta, C. T. Walsh and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 3428.
23 K. K. Sharma and C. N. Boddy, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 3034.
24 S. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, 2002, **41**, 12598.
25 Z. Suo, C. C. Tseng and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 99.
26 J. J. La Clair, T. L. Foley, T. R. Schegg, C. M. Regan and M. D. Burkart, *Chem. Biol.*, 2004, **11**, 195.
27 A. S. Worthington, H. Rivera, J. W. Torpey, M. D. Alexander and M. D. Burkart, *ACS Chem. Biol.*, 2006, **1**, 687.
28 Y. Liu and S. D. Bruner, *ChemBioChem*, 2007, **8**, 617.
29 Z. Y. Yuan and G. G. Hammes, *J. Biol. Chem.*, 1986, **261**, 13643.
30 M. J. Evans and B. F. Cravatt, *Chem. Rev.*, 2006, **106**, 3279.
31 Y. Liu, M. P. Patricelli and B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14694.
32 L. E. N. Quadri, *Infect. Disord. Drug Targets*, 2007, **7**, 230.
33 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117.
34 L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Hakansson, J. L. Smith and D. H. Sherman, *Science*, 2007, **318**, 970.
35 T. P. Stinear, H. Hong, W. Frigui, M. J. Pryor, R. Brosch, T. Garnier, P. F. Leadlay and S. T. Cole, *J. Bacteriol.*, 2005, **187**, 1668.
36 F. Wang, R. Langley, G. Gulten, L. Wang and J. C. Sacchettini, *Chem. Biol.*, 2007, **14**, 543.
37 F. Kopp, C. Mahlert, J. Grunewald and M. A. Marahiel, *J. Am. Chem. Soc.*, 2006, **128**, 16478.

38 T. Liu, D. You, C. Valenzano, Y. Sun, J. Li, Q. Yu, X. Zhou, D. E. Cane and Z. Deng, *Chem. Biol*., 2006, **13**, 945.
39 M. Harvey, H. Hong, M. A. Jones, Z. A. Hughes-Thomas, R. M. Goss, M. L. Heathcote, V. M. Bolanos-Garcia, W. Kroutil, J. Staunton, P. F. Leadlay and J. B. Spencer, *ChemBioChem*, 2006, **7**, 1435.
40 M. D. Alexander, S. D. Fontaine, J. J. La Clair, A. G. Dipasquale, A. L. Rheingold and M. D. Burkart, *Chem. Commun*., 2006, **44**, 4602.
41 F. Feyen, A. Jantsch and K. Altmann, *Synlett*, 2007, **3**, 415.
42 R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins and J. D. Thompson, *Nucleic Acids Res*., 2003, **31**, 3497.
43 L. J. McGuffin, K. Bryson and D. T. Jones, *Bioinformatics*, 2000, **16**, 404.
44 J. A. Cuff, M. E. Clamp, A. S. Siddiqui, M. Finlay and G. J. Barton, *Bioinformatics*, 1998, **14**, 892.
45 B. Morgenstern, *Nucleic Acids Res*., 2004, **32**, W33.
46 K. B. Nicholas, H. B. Nicholas, Jr and D. W. Deerfield II, *EMBNEW.NEWS*, 1997, **4**, 14.
47 M. M. Bradford, *Anal. Biochem*., 1976, **72**, 248.

The Unusual Macrocycle-Forming Thioesterase of Mycolactone

*Tiffany Barrows-Yano, Jordan L. Meier, Timothy L. Foley, Candice Wilkey, and Michael D. Burkart**
Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0358
mburkart@ucsd.edu

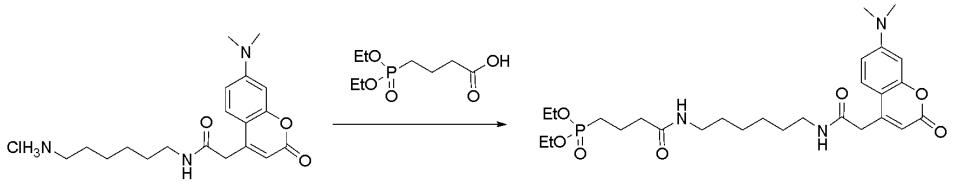
Online Supporting Information

Contents

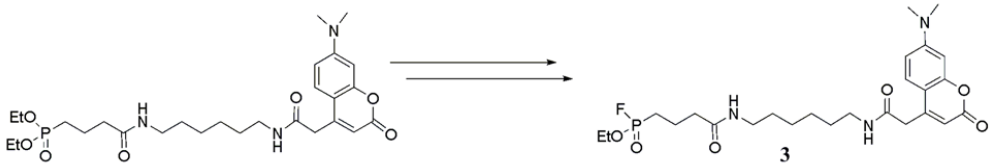
General synthetic procedures and materials

All commercial reagents (Sigma-Aldrich, Spectrum, MP Biomedicals, Alfa Aesar, TCI America, Acros) were used as provided unless otherwise indicated. BODIPY-CoA,[1] 7-dimethylaminocoumarin-hexanediamine HCl,[2] and 4-(diethoxyphosphinyl)-butanoic acid[3] were prepared according to published literature procedures. All reactions were carried out under argon atmosphere in dry solvents with oven-dried glassware and constant magnetic stirring unless otherwise noted. Triethylamine (TEA), N-methyl morpholine (NMM), and ethyl-*N*,*N*-diisopropylamine (DIPEA) were dried over sodium and freshly distilled. ^{1}H-NMR spectra were taken at 300, 400, or 500 MHz and ^{13}C-NMR spectra were taken at 100.6 or 75.5 MHz on Varian NMR spectrometers and standardized to the NMR solvent signal as reported by Gottlieb.[4] Multiplicities are given as s=singlet, d=doublet, t=triplet, q=quartet, p=pentet, dd=doublet of doublets, bs=broad singlet, bt=broad triplet, m=multiplet using integration and coupling constant in Hertz. TLC analysis was performed using Silica Gel 60 F254 plates (EM Scientific) and visualization was accomplished with UV light (λ=254 nm) and/or the appropriate stain (iodine, 2,4-dinitrophenylhydrazine, cerium molybdate, ninhydrin). Silica gel chromatography was carried out with Silicycle 60 Angstrom 230-400 mesh according to the method of Still.[5] TLC prep plate purification was performed with EMD Silica Gel 60 F_{254} pre-coated plates. Electrospray (ESI) and fast atom bombardment (FAB) mass spectra were obtained at the UCSD Mass Spectrometry Facility by Dr. Yongxuan Su using a Finnigan LCQDECA mass spectrometer and a ThermoFinnigan MAT- 900XL mass spectrometer, respectively.

Synthetic Procedures and Spectroscopic Data for Fluorophosphonate 3

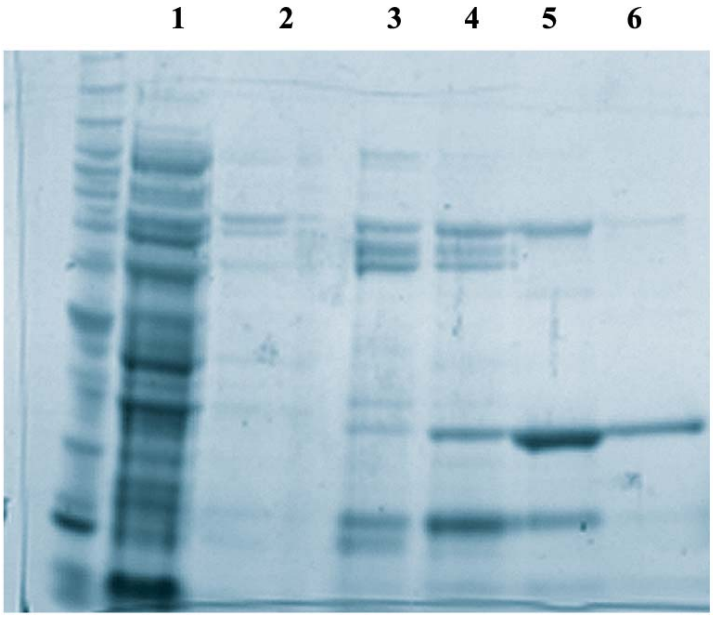


4-(diethoxyphosphinyl)-butanoic acid (220 mg, 0.98 mmol), DIPEA (677 uL, 3.88 mmol), 7-dimethylaminocoumarin-hexanediamine HCl (368 mg, 0.970 mmol), and HOBt (558 mg, 3.71 mmol) were dissolved in DMF (20 mL) with stirring and cooled to 0°C. EDC (558 mg, 2.91 mmol) was added in one portion and the reaction was allowed to slowly warm to RT and stirred overnight. The solvent was removed under reduced pressure to yield a crude oil which was taken up in EtOAc (150 mL) and washed with water (1x100 mL), saturated $NaHCO_3$ (2x125 mL), and brine (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure to yield the diethylphosphonate amide as a fluorescent oil (405 mg, 76%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.48 (d, *J*=8.5 Hz, 1H), 6.95 (bt, *J*=6.0 Hz, 1H), 6.55 (dd, *J*=2.5 Hz, 9.0 Hz, 1H), 6.48 (bt, *J*=6.0 Hz, 1H), 6.39 (d, *J*=2.5 Hz, 1H), 6.03 (s, 1H), 4.01 (q, *J*=7.5 Hz, 4H), 3.57 (s, 2H), 3.13 (m, 4H), 2.99 (s, 6H), 2.26 (t, *J*=7.5 Hz, 2H), 1.86 (p, *J*=7.5 Hz, 2H), 1.73 (m, 2H), 1.37 (m, 4H), 1.26 (t, *J*=7.5 Hz, 6H), 1.21 (m, 4H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 172.4, 168.5, 162.3, 156.1, 153.2, 151.0, 126.1, 110.0, 109.3, 108.8, 98.2, 61.8 (d, *J*=6.8 Hz), 40.5, 40.3, 39.6, 39.1, 36.6 (d, *J*=13.8 Hz), 29.6, 29.3, 26.3, 25.5, 24.1, 19.2 (d, *J*=4.6 Hz), 16.6 (d, *J*=6.1 Hz). HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{21}H_{48}N_3O_7P_1$, 551.2755 found 551.2757.

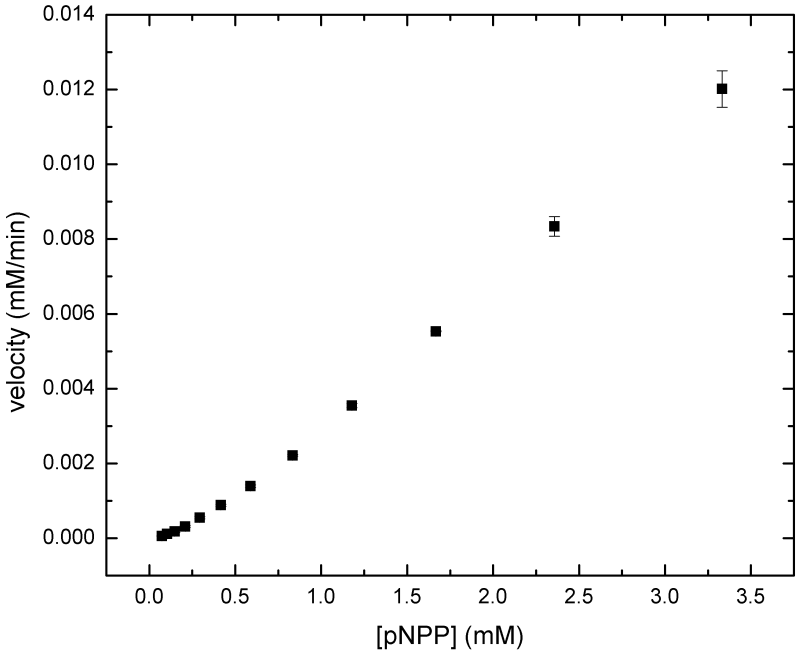


The product of the above reaction (350 mg, 0.635 mmol) was dissolved in DCM (6.5 mL) and bromotrimethylsilane (0.418 mL, 3.24 mmol) was added dropwise at 0°C. The reaction was allowed to slowly attain RT and stir for 2.5 hrs. The reaction was quenched with H_2O and the solvent was removed under reduced pressure via rotovap with an inline base (NaOH) trap. The crude monophosphinic acid was carried forward as a low R_f oil as observed by TLC. This monophosphinic acid residue, was then dissolved in dry DMF (5 mL), transferred to a plastic reaction vial, and cooled to –78°C with stirring. DAST (50 uL, 0.39 mmol) was added dropwise via microsyringe. After 10 minutes the reaction was

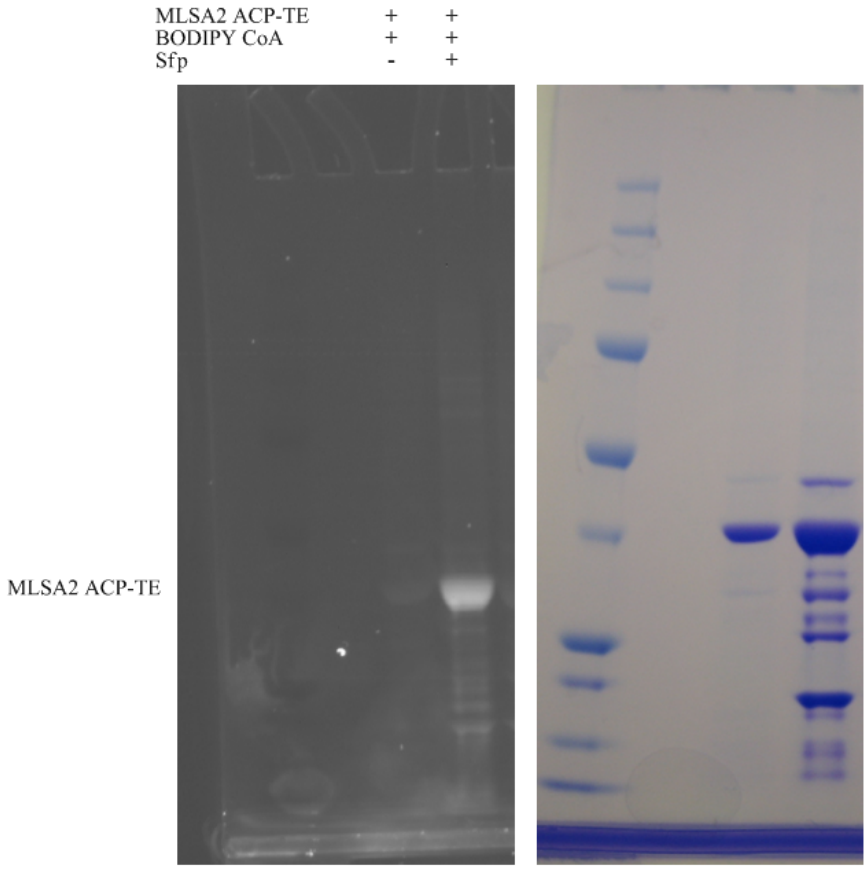
allowed to slowly attain RT and diluted with EtOAc (250 mL). This solution was washed with H_2O (1x100 mL) and brine (1x100 mL), and the H_2O layer saved and lyophilized to recover the crude monophosphinic acid (282 mg, 85%). The organic layer was then dried over Na_2SO_4, filtered, and evaporated under reduced pressure to yield the crude fluorophosphonate. This residue was redissolved in $CHCl_3$ and purified by flash chromatography ($CHCl_3$ to 1:3 $CHCl_3$:acetone to 1:3:0.1 $CHCl_3$:acetone:MeOH) to give **1** as a fluorescent oil. This procedure was repeated on the recovered monophosphinic acid an additional time to provide a total of 22.5 mg (6.5 % yield for two steps) of fluorophosphonate **3** (estimated 90% purity by ^{1}H-NMR). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.47 (m, 1H), 6.60 (m, 2H), 6.48 (m, 1H), 6.23 (bs, 1H), 6.04 (s, 1H), 3.60 (m, 2H), 3.18 (m, 4H), 3.03 (s, 6H), 2.31 (t, *J*=7.0 Hz, 2H), 1.99 (m, 4H), 1.40 (m, 4H), 1.25 (m, 7H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 171.9, 168.1, 160.7, 155.1, 153.3, 153.2, 150.9, 126.1, 113.8, 110.2, 109.5, 98.3, 53.4 (d, *J*=6.9 Hz), 43.5, 41.7, 40.4, 39.5, 36.0 (d, *J*=15.7 Hz), 29.5, 29.2, 26.1, 24.0, 22.6, 18.5 (d, *J*=4.6 Hz), 14.5. ^{19}F-NMR (282 Hz, $CDCl_3$) δ -66.9 (d, J_p=1071 Hz, F-P). ^{31}P-NMR (162 Hz, $CDCl_3$) δ 33.3 (d, J_F=1071 Hz, P-F). HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{25}H_{37}N_3O_6F_1P_1$, 525.2399, found 525.2398.



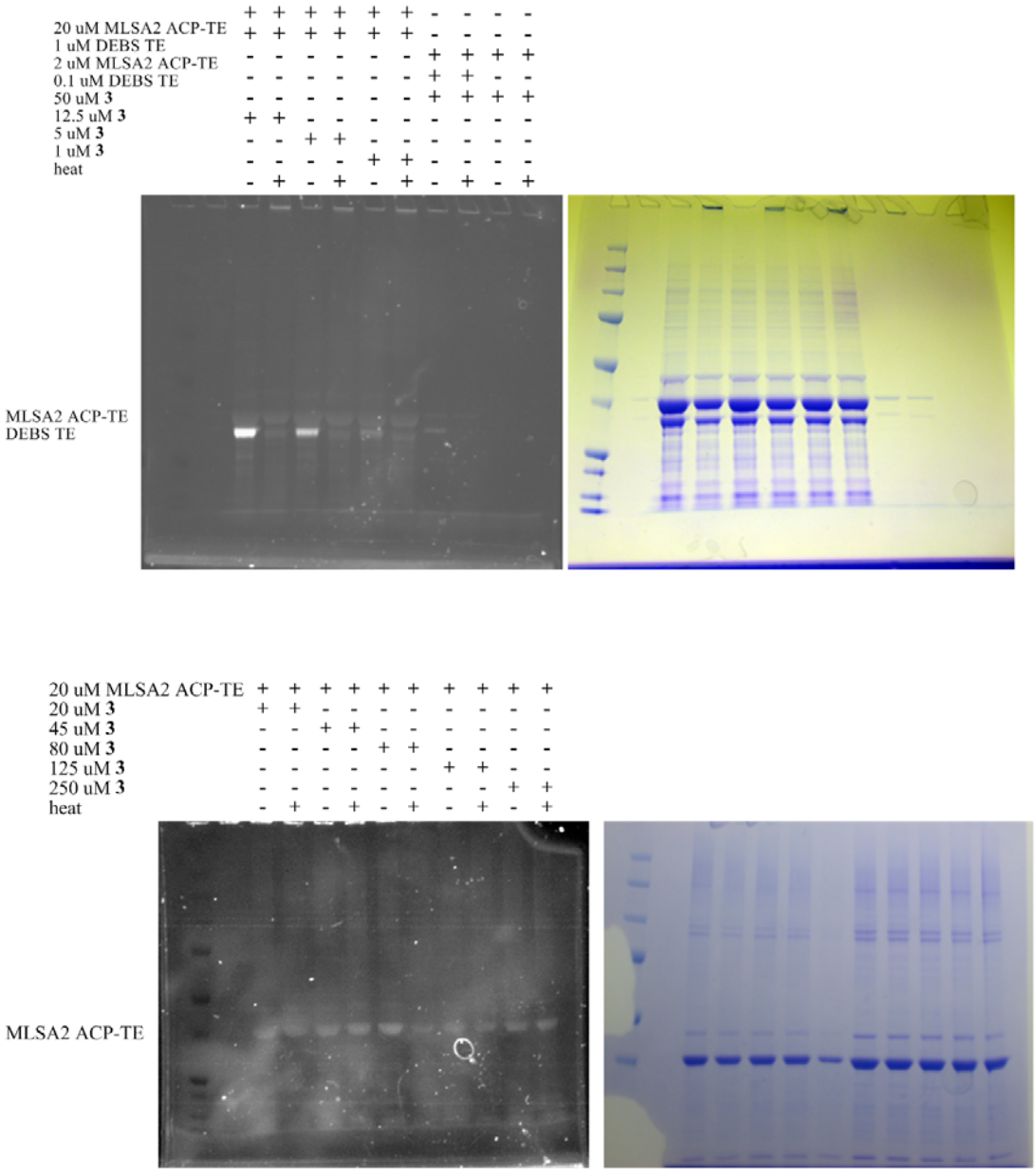
Complete Gel Image for Purification of MLSA2 TE Monodomain. MLSA2 TE was cloned and expressed as detailed in the main text. Gel lanes depict fractions taken during Ni-NTA chromatography and are as follows: 1) flow through, 2) wash, 3) 10 mM imidazole wash, 4) 50 mM imidazole, 5) 100 mM imidazole, 6) 200 mM imidazole.



Kinetic Plot of MLSA2 TE Monodomain: Kinetics of recombinant MLSA2 TE monodomain with pNPP. Initial values allowed estimation of K_{cat}/K_m for the monodomain at 0.16 ± 0.00041 (compared to 0.20 for the didomain).



Complete Gel Data from Figure 4. Sfp dependent labeling of MLSA2 ACP-TE by BODIPY CoA and the PPTase Sfp.



Complete Gel Data from Figure 5. Top: Labeling of DEBS-TE and MLSA2 ACP-TE in one-pot with fluorophosphonate **3**. Heat dependent labeling is observed with DEBS-TE but not MLSA2. Bottom: Concentration dependence of MLSA2 labeling by fluorophosphonate **3**. In contrast to the DEBS labeling in the top gel no clear ± labeling is observed.

Compiled MS Data from fluorophosphonate treated MLSA2 ACP-TE

MLSA2 ACP-TE Didomain – Fluorophosphonate Treatment and Chymotrypsin Digest

1 MKHHHHHHHHHHGGLESTSLYKKAGSTLATL**VAAATATVLGHHTPESISP A**
51 **TAFKDLGIDSLTALELRNTLTHNTGLDLPPTLIFDHPTPT ALTQHLHTRL**
101 **TQSHTPVGPIASLLSHAIDEGKFRAGADLLMAASNLNQSF**SNMAELNQLP
151 **AVTDIADASPDGLLTLICISTSENEYARLAAANIHSTF**AEIAAPGFYDA
201 QLPNSIETSA EALATAITGA YANTSIVL**VA HSIVCELAQA TMTRLQDADI**
251 **DLVGL**VLLDPLEGTNSTEDYVET**VLTRIEHINAPRVGVDGYLAALGRYLQ**
301 **FHEDRRIPIPETRHMTLHSDTKIDRAQTPMNL**LQDEAALAL**KIGNWMND**
351 **TGSIAVTLRD GPVFL**GRARSVNMR

Purple = peptide sequence coverage received for chymotrypsin digest of MLSA2 ACP-TE treated with **1**
Green = peptide sequence coverage received for chymotrypsin digest of untreated MLSA2 ACP-TE
Red = overlapping peptide sequence coverage received from both proteome treatments

MLSA2 ACP-TE Didomain – Fluorophosphonate Treatment and Pepsin Digest

1 MKHHHHHHHHHHGGLESTSLYKKAGSTL**ATLVAAATATVLG HHTPESISPA**
51 **TAFKDLGIDSLTALELRNTLTHNTGLDLPPTLIFDHPTPTALTQHLHTRL**
101 **TQSHTPVGPIASLLSHAIDEGKFRAGADLLMAASNLNQSFSNMAELNQLP**
151 **AVTDIADASPDGLLTLICISTSENEYARLAAANIHSLTFAEIAAPGFYDA**
201 **QLPNSIETSA EALATAITGA Y**ANTSIVL**VA HSIVCELAQA TMTRLQDADI**
251 **DLVGLVLLDPLEGTNSTEDYVETVLTRIEHINAPRVGVDGYLAALGRYL**
301 **FHEDRRIPIPETRHMTLHSDTKIDRAQTPMNL**LQDEAALTAL**KIGNWMND**
351 **TGSIAVTLRD GPVFL**GRARS VNMR

Purple = peptide sequence coverage received for pepsin digest of MLSA2 ACP-TE pretreated with **1**
Green = peptide sequence coverage received for pepsin digest of untreated MLSA2 ACP-TE
Red = overlapping peptide sequence coverage received from both proteome treatments

Peptide sequence coverage received upon chymotyrpsin or pepsin digest of MLSA2 ACP-TE treated with fluorophosphonate **3**. In both digests the proposed catalytic serine (residue 232 following the convention of the report above) is observed unmodified after treatment with 50x fluorophosphonate. No evidence of the fluorophosphonate modified or dehydroalanine eliminated Ser232 was observed. Note: the above report differs in numbering from that of the sequence analysis (in which the above S232 is referred to as S126) due to incorporation of the CP domain in the above analysis.

References:

1. La Clair JJ, Foley TL, Schegg TR, Regan CM, Burkart MD (2004) Manipulation of carrier proteins in antibiotic biosynthesis. Chem. Biol. 11: 195-201.

2. Clarke KC, Mercer AC, LaClair JJ, Burkart MD (2005) In Vivo Reporter Labeling of Proteins via Metabolic Delivery of Coenzyme A Analogues. J. Am. Chem. Soc. 127: 11234 – 11235.

3. Swierczek K, Pandey AS, Peters JW, Hengge ACA (2003) Comparison of Phosphonothioic Acids with Phosphonic Acids as Phosphatase Inhibitors. J. Med Chem. 46: 3703-3708.

4. Gottlieb HE, Kotlyar V, Nudelman A (1997) NMR Chemical Shifts of Common Laboratory Solvents As Trace Impurities. J. Org. Chem. 62: 7512-7515.

5. Still WC, Kahn A, Mitra A (1978) Rapid Chromatographic Technique For Preparative Separations With Moderate Resolution. J. Org. Chem. 43: 2923-2925.

Acknowledgments

The text of chapter 7, in full, is a reprint of material as it appears in: Meier, J.L., Barrows-Yano, T., Foley, T.L., Wike, C. L., and Burkart M.D. (2008). The unusual macrocycle forming thioesterase of mycolactone. *Mol. Biosyst.4*, *663-671*. Permission was obtained from all co-authors. I was the co-primary author of this work. Tiffany Yano expressed and purified the recombinant MLSA2 ACP-TE didomain, with assistance from Timothy Foley, and Candice Wike. I synthesized all compounds, assisted with biological assays, designed mass spectral experiments, and wrote the manuscript. This research was performed under the supervision of Prof. Michael Burkart.

Chapter Eight

A mechanism based protein crosslinker for acyl carrier protein dehydratases

Adapted from:

Meier, J. L.; Haushalter, R.W.; Burkart, M. D. *Submitted* (2009).

A Mechanism Based Protein Crosslinker for Acyl Carrier Protein Dehydratases

Introduction

Fatty acid synthase (FAS) and polyketide synthase (PKS) biosynthetic enzymes are modular, multienzymatic biosynthetic catalysts responsible for the production of a large array of biologically relevant primary and secondary metabolites.[1,2] The modular nature of these enzymes, in which the sequence and identity of the FAS or PKS active sites dictates the final structure of the small molecule formed, has long made them an attractive target for metabolic engineering.[3] However, despite much effort, the full potential of combinatorial biosynthetic approaches to produce polyketides remains to be realized. One major obstacle to rational reengineering of PKS biosynthesis is the scarcity of structural information available for these enzymes, which makes it difficult to predict how changes to the primary sequence of PKS enzymes will effect tertiary and quaternary structure.[4,5]

While much of the difficulty in generating structural data for multienzyme fragments of FAS and PKS enzymes can be attributed to the large size (often > 200 kDa), another major challenge in the structural study of these enzymes is the acyl carrier protein (ACP). ACPs are small (~ 9 kDa), highly acidic proteins to which biosynthetic intermediates are tethered throughout all stages of the FAS and PKS biosynthetic cycles.[6] Since this necessitates interaction of the ACP with multiple enzymatic partners, the ACP is inherently flexible, relying on specific protein-protein interactions to recognize and channel biosynthetic intermediates to the proper partner enzyme. While a better understanding of these ACP-partner protein interactions would greatly aid efforts towards

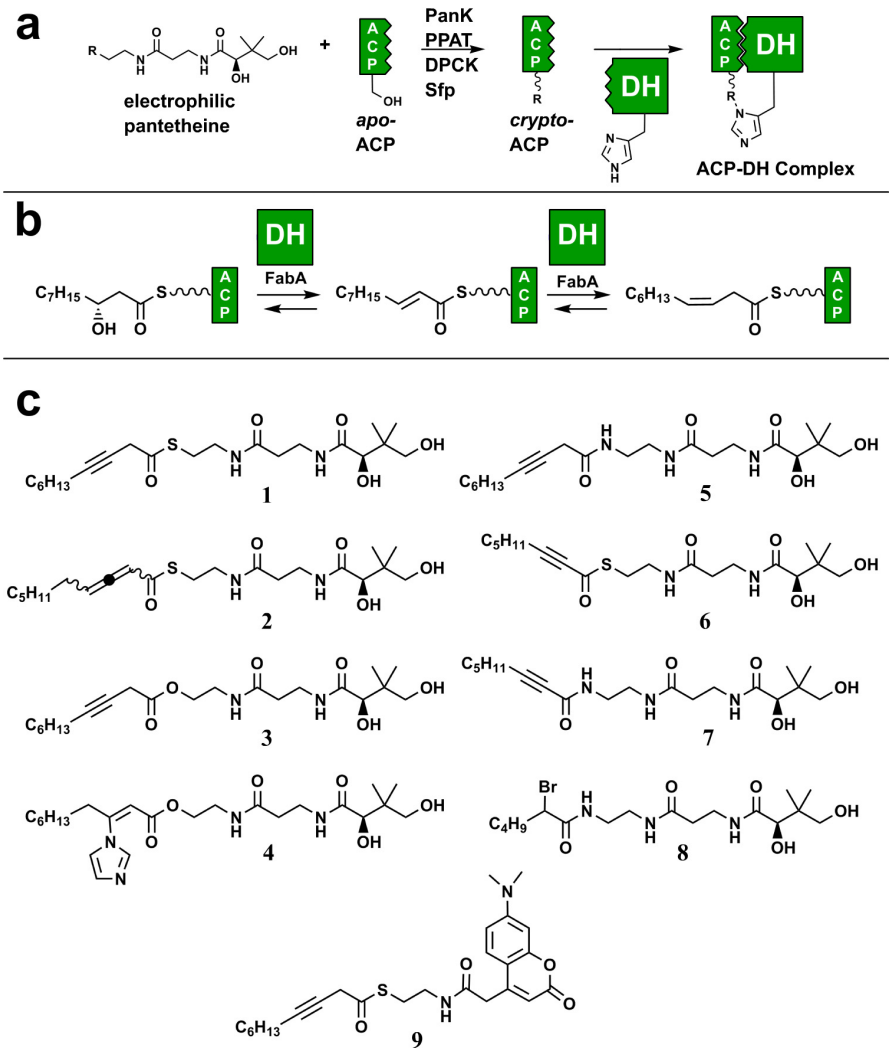


Figure 8.1 General strategy for site-specific mechanism-based crosslinking of ACP and DH domains. (a) *Apo*-ACPs can be modified with electrophilic 4'-phosphopantetheine arms by utilizing the CoA biosynthetic enzymes (PanK, PPAT, and DPCK) and the PPTase Sfp. Upon addition of an ACP partner protein, the native protein-protein interactions of the pair result in a transient interaction which is captured by the electrophile. (b) Dehydration and isomerization reactions catalyzed by the *E. coli* FAS DH enzyme FabA, the subject of the current study. (c) Structures of electrophilic pantetheine analogues (**1**-**8**) and a fluorescent FabA probe (**9**) used in this study.

the rational reengineering of these enzyme systems, their transient nature presents a challenge to traditional methods of protein structure analysis.

To date, the only multimodular FAS or PKS crystal structure in which electron density corresponding to the ACP has been clearly observed has been in the yeast FAS.[7] While this study relied on the fortuitous formation of a disulfide bond between the terminal thiol of the ACP 4'-phosphopantetheine (4'-PPant) and active site cysteine of the ketosynthase (KS) domain to stall the ACP, a more ideal approach would be to develop a toolbox of molecular probes enabling predictable, reproducible crosslinking of the ACP to any FAS or PKS partner domain. With this approach in mind, we first reported a strategy for site-specific, mechanism-based crosslinking of ACP and KS domains.[8] This strategy utilized the promiscuous PPTase Sfp together with electrophilic CoA analogues to install mechanism-based inhibitor moieties on the 4'-PPant arm of the ACP.[9,10] When these electrophile containing *crypto*-ACPs were treated with KS domains possessing protein-protein interactions complementary to the ACP, the modified 4'-PPant arm was found to irreversibly modify the KS active-site, resulting in formation of an ACP-KS complex (Figure 8.1a).[8] This approach was first applied to the ACP and KS domains of type II bacterial fatty acid biosynthesis, and has since been extended to both multimodular (type I) and trans-acting (type II) FAS and PKS enzymes.[8,11,12] In addition to its potential utility in facilitating structural analysis of FAS and PKS enzymes, such crosslinking reactions provide a direct gel-based readout of the complementarity of ACP-partner protein interactions, providing a powerful tool for the determination of suitable partner proteins for use in combinatorial biosynthetic approaches.[12] Similar technologies have also been developed to probe the interactions of peptidyl carrier proteins (PCPs)

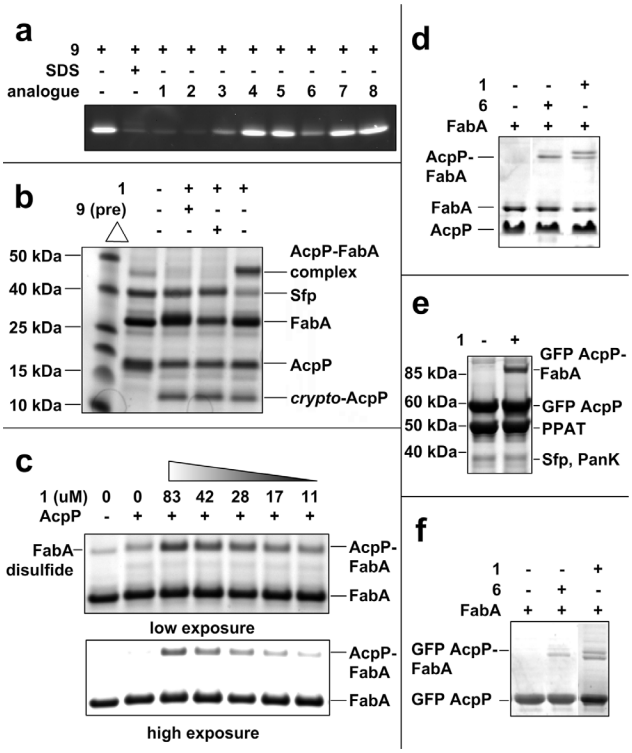


Figure 8.2 Evaluation of ACP-DH crosslinking reagents **1**-**8**. (a) Thioesters **1**-**2** and **6** and oxo-ester **3** demonstrate the ability to block the FabA active site from labeling by FabA probe **9**, indicating potential utility in crosslinking applications (b) ACP-DH crosslinking of the *E. coli* FAS enzymes AcpP and FabA. The crosslinked species is sensitive to inactivation of FabA by **9** as well as pre-denaturation of the DH domain by heat. (c) ACP-DH crosslinking is dependent on the presence of **1**, with decreasing amounts of crosslinking observed as the amount of **1** is decreased. A persistent FabA disulfide is observed in the negative controls (lanes 1 and 2) which is not observed using higher film exposure times. (d) Reaction of FabA with purified *apo*-ACP (lane 1), 2-octynoyl-ACP (lane 2), and 3-decynoyl-ACP (lane 3). (e) ACP-DH crosslinking of the GFP AcpP fusion and FabA. (f) Reaction of FabA with purified *apo*-GFP AcpP (lane 1), 2-octynoyl GFP AcpP (lane 2), and 3-decynoyl GFP AcpP (lane 3).

with partner domains in nonribsomal peptide synthetase (NRPS) biosynthetic systems.[13-15] In this report we extend the repetoire of chemoenzymatic ACP crosslinking to study the interactions of ACP and acyl carrier protein dehydratase (DH) domains in FAS and PKS systems. DH domains catalyze the dehydration of β-hydroxyacyl-ACP substrates to α,β-unsaturated acyl-ACPs during the reductive steps of fatty acid and polyketide biosynthesis (Figure 8.1b).[16,17] In addition, DH domains play a key role in initiation of bacterial unsaturated fatty acid biosynthesis by isomerizing *trans*-2-decenoyl-ACP to *cis*-3-decenoyl-ACP. These reactions were first studied in the context of the *E. coli* FAS, where they are catalyzed by the prototypical DH enzyme FabA. (Figure 8.1b).[18,19] These studies also yielded the discovery of the first mechanism-based inhibitor of a fatty acid biosynthetic enzyme, 3-decynoyl N-acetyl-cystamine (NAC), a suicide substrate that undergoes α-deprotonation in the FabA active site to form an electrophilic allene, which then modifies the active site histidine of FabA to irreversibly inactivate the enzyme.[20-22]

To explore the possibility of using this historical inhibitor scaffold together with more recently developed ACP modification techniques to develop an ACP-DH crosslinking strategy, a small panel of pantetheine analogues was synthesized incorporating well-known inhibitor scaffolds of DH and other α-deprotonation utilizing enzymes (Figure 8.1c, **1**-**8**). These pantetheine analogues can be transformed into CoA analogues and site-specifically incorporated into ACPs using the one-pot chemoenzymatic method depicted in Figure 8.1a.[10] In addition to the well-known 3-decynoyl and 2,3-decadienoyl thioester inhibitors (**1-2**), we also examined the utility of hydrolysis resistant 3-deycnoyl oxoester and amides (**3-4**), a transition state analogue (**5**), 2-octynoyl thioesters and amides capable of forming reactive allenes upon γ-

deprotonation (**6-7**), and a simple histidine reactive acyl-bromoacetamide affinity tag (**8**).[23-25] These pantetheine analogues were quickly assayed for their relative abilities to modify the active site of recombinantally expressed FabA by testing their ability to block labeling by fluorescent probe **9**, a 3-decynoyl-NAC derivative which reacts with FabA in an active-site dependent manner. As expected, pre-incubation of FabA with the denaturing agent SDS or known inhibitor scaffolds **1** or **2** each efficiently blocked labeling by **9** (Figure 8.2a). 3-decynoyl-oxoester **3** and 2-octynoyl thioester **6** showed decreased active site modification, blocking fluorescent labeling by **9** to a lesser degree, while pantetheine analogues **4-5** and **7**-**8** showed no effect. This is consistent with previous studies on mechanism-based inhibition of FabA by 3-decynoic acid analogues, which found enzyme inactivation to be strongly dependent on the pKa of the α-proton of the suicide substrate (thioester > oxoester >> acid/amide). This lead us to focus on the use of thioesters **1**, **2**, and **6** as ACP-DH crosslinking reagents.[20,26]

First, the ability of analogues **1**, **2**, and **6** to modify the *E. coli* FAS ACP (AcpP) was demonstrated. Using the CoA biosynthetic enzymes PanK, PPAT, and DPCK along with the permissive PPTase Sfp, modification of AcpP by **1**, **2**, and **6** could be observed by MALDI-TOF or SDS-PAGE, in which a characteristic gel shift to lower molecular weight was observed upon AcpP-incorporation of fatty acyl pantetheines **1**, **2**, or **6** (Figure 8.2b).[27,28] Upon addition of FabA to AcpP modified by **1**, a faint new band appearing at ~ 45 kDa corresponding to a putative AcpP-FabA complex was observed (Figure 8.2b). While this band co-migrated with a persistent FabA disulfide, it could be clearly visualized using strongly reducing SDS-PAGE conditions and/or high film exposure times. This putative AcpP-FabA complex was observed to be dependent upon

the presence as well as amount of **1** added to the reaction mix (Figure 8.2c). ACP, PanK, and Sfp were each also judged to be necessary components for this crosslinking to occur (Figure 8.S1). In addition, complex formation was highly sensitive to the integrity of the FabA active site, and was not observed in reactions in which FabA had been pre-denatured by boiling or inactivated by high concentrations of **9** (Figure 8.2b). Replacement of pantetheine analogue **1** with **2** resulted in approximately equivalent results, while analogue **6** produced a noticeably fainter crosslinked species (Figure 8.S2).

For the unambiguous identification of the crosslinked species we first applied an orthogonal purification strategy to isolate *crypto*-AcpP modified by **1** or **6** from the CoA biosynthetic enzymes and Sfp.[29] These *crypto*-ACPs could then be incubated with FabA, allowing a clear gel-shift to be observed in each case which was not observed when **1** or **6** was omitted (Figure 8.2d). A similar result was seen upon repetition of the same experiment, this time substituting a GFP-tagged AcpP construct for AcpP.[30] The GFP AcpP construct appears to maintain the necessary interactions for both ACP-KS and ACP-DH crosslinking, and removes the crosslinked species from the molecular weight region of the earlier noted FabA disulfide, facilitating visualization of the AcpP-FabA complex (Figure 8.2e-f). Both the GFP AcpP and native AcpP-FabA complexes were excised and subjected to tryptic digest, and identified by MALDI-TOF/TOF and peptide mass fingerprinting (Figure 8.S3).

One consideration when applying Michael acceptors **2** and **6** to ACP-DH crosslinking of multidomain synthases is their potential reactivity with KS domains. Indeed, when c*rypto*-AcpPs generated from **1**, **2**, and **6** were incubated with the *E. coli* KS enzyme FabB each demonstrated ACP-KS crosslinking activity (Figure 8.S3).

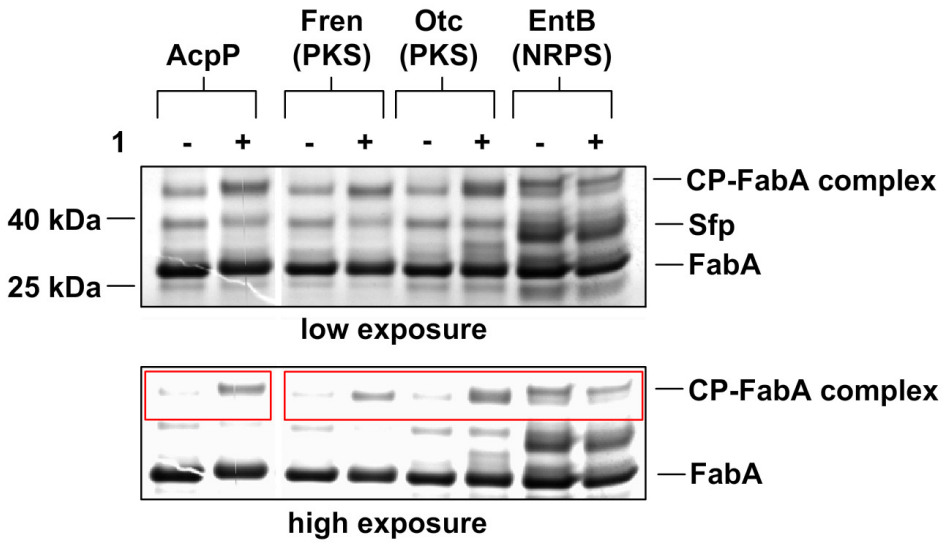


Figure 8.3 Effect of carrier protein-dehydratase protein-protein interactions on CP-FabA crosslinking. 3-decynoyl pantetheine **1** was used to modify the carrier proteins AcpP (FAS), Fren (PKS), Otc (PKS), and EntB (NRPS). Upon addition of FabA, crosslinking is observed only with AcpP, Fren, and Otc, indicating the preferential interaction of FabA with 3-decynoyl ACPs.

This could be abrogated by pre-incubation of FabB with the KS-selective reagent cerulenin, a strategy which may prove similarly useful for achieving ACP-DH specific crosslinking in multidomain synthases. Additionally, the finding that 3-decynoyl thioester **1** promotes ACP-KS crosslinking suggests some non-enzymatic allenic isomerization of this reagent occurs during AcpP loading of **1**.

Finally, we tested the abiilty of the ACP-DH crosslinking reaction to discriminate between native and non-native carrier protein-DH pairs.[8,12] In addition to the FAS carrier protein AcpP, the PKS carrier proteins Otc and Fren, as well as the NRPS carrier protein EntB were posttranslationally modified by **1** and incubated with FabA. While the limited progress of these crosslinking reactions necessitated high exposure times to discriminate CP-DH complexes from background, FabA appears to crosslink preferentially with 3-decynoyl ACPs from FAS and PKS systems compared to the PCP EntB. This is strong evidence that, as with ACP-KS crosslinking, the formation of ACP-DH complexes is driven by the inherent compatability of the protein-protein interactions of the ACP-DH pair.

One limitation to this ACP-DH crosslinking approach is the modest degree of AcpP-FabA complex observed. The finding that crosslinking yield was not dramatically improved by purification 3-decynoyl *crypto*-ACP (Figure 8.1d) could be an indication the inhibitor moiety of **1** is inactivated, most likely through non-enzymatic allenic isomerization and nucleophilic addition to the 2,3-didecenoic thioester. This is supported by the observation that purification of 3-decynoyl-AcpP prior to FabB addition resulted in less ACP-KS crosslinking compared to FabB addition to the corresponding crude reactions (Figure 8.S3). Alternately, the observed crosslinking may reflect the inherent

equilibrium of the Acp-FabA pair. The extension of this technique to alternate ACP-DH pairs will be necessary to further probe this phenomenon.

This study demonstrates the first site-selective protein crosslinking of ACP and DH domains from a FAS system. The crosslinking requires an intact FabA active site, and is dependent on complementary protein-protein interactions of the 3-decynoyl carrier protein-FabA pair, with FAS and PKS carrier proteins being favored over NRPS carrier proteins. The AcpP-FabA complex formed by **1** is stable, with little hydrolysis observed after 24 hrs or upon treatment with hydroxylamine; this may be explained by the crystal structure of FabA in complex with 3-decynoyl-NAC, which found the fatty acyl inhibitor was efficiently sequestered from bulk solvent in the hydrophobic FabA active site.[31] Through their skillful application in concert with refined methods for the structural analysis of FAS and PKS enzymes,[32] such carrier protein crosslinking techniques have the potential to provide snapshots of these elusive and fascinating biosynthetic catalysts at work.

Supporting Information

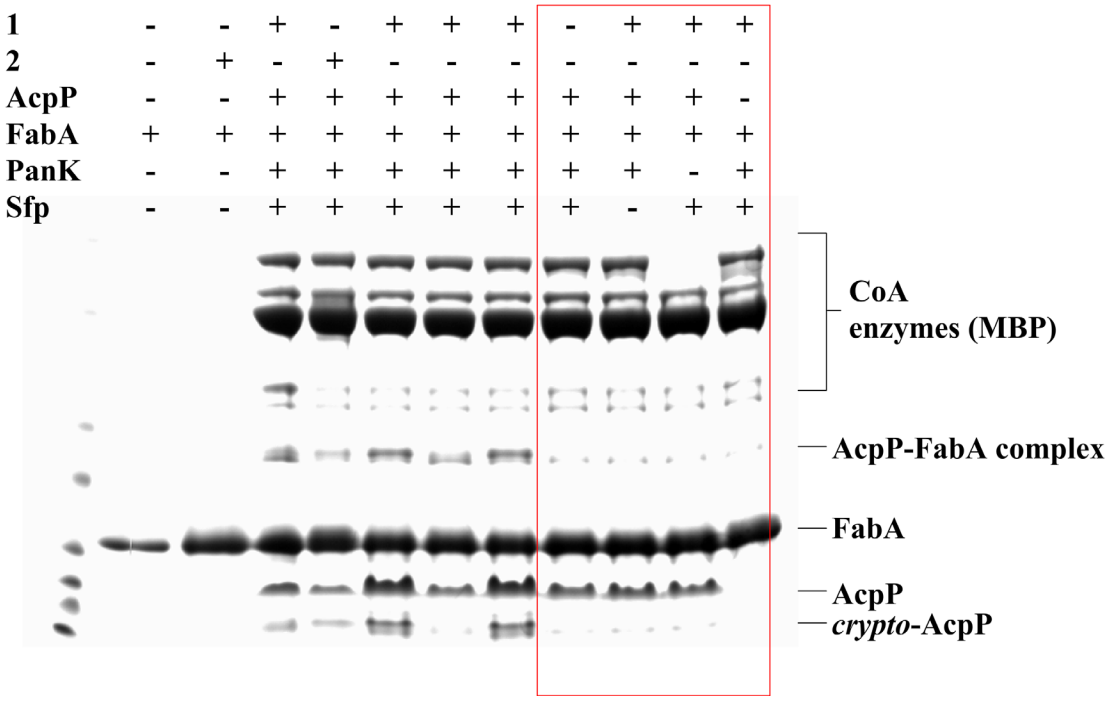


Figure 8.S1 Evaluating necessity of reaction components (AcpP, PanK, Sfp) for AcpP-FabA (ACP-DH) crosslinking. Each of these reaction components is necessary to formation of crosslinked complex. Lanes 2-3: FabA. Lanes 4-8: positive crosslinking. Lane 9: pantetheine analogue **1** omitted. Lane 10: Sfp omitted. Lane 11: PanK omitted. Lane 12: AcpP omitted.

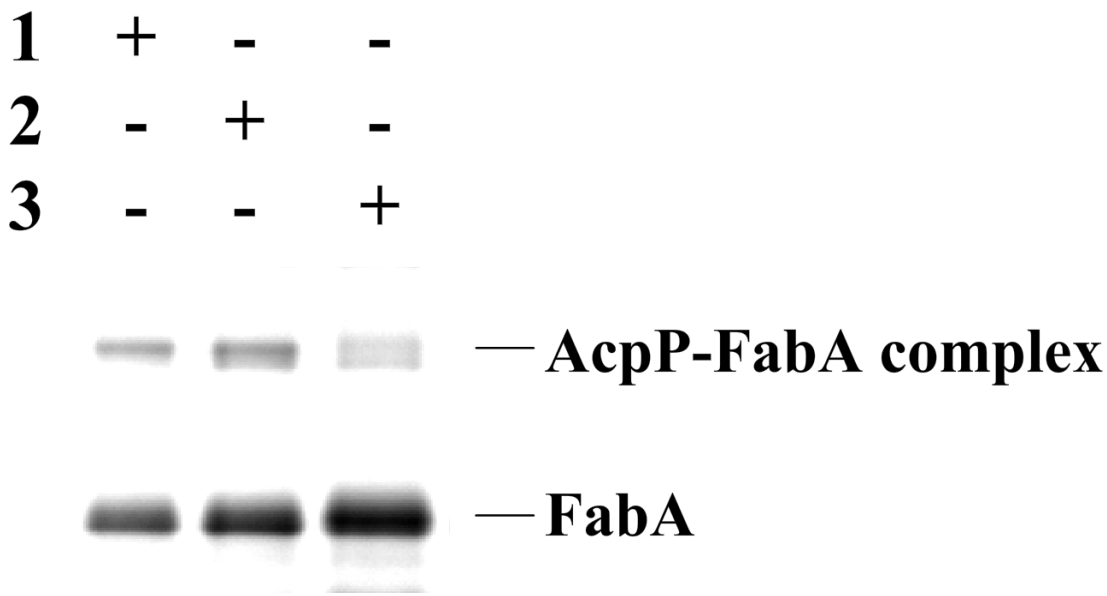


Figure 8.S2 Comparison of AcpP-FabA crosslinking using chemoenzymatic labels **1**, **2**, and **6**. Pantetheine analogues **1** and **2** show approximately equivalent crosslinking (based on density of crosslinked band) while analogue **6** produces a crosslinked species that is notably more diffuse.

Figure 8.S3 MALDI-TOF analysis of crosslinked complex. Peptides from both AcpP and FabA were identified in the crosslinked sample, while peptides containing the known active site residues of these two enzymes were *not* observed, consistent with their covalent crosslinking. (a) Data for digestion of AcpP-FabA crosslinked complex. Highlighted residues were identified by in gel tryptic digest and analysis by MALDI-TOF/TOF of the SDS-PAGE gel slice specified. (b) Because of the low number of peptides formed upon tryptic digest of AcpP, the GFP AcpP-FabA crosslinked complex was subjected to an identical analysis. Highlighted residues were identified by tryptic digest and analysis by MALDI-TOF/TOF of the SDS-PAGE gel slice specified. All highlighted residues were found in peptides identified with

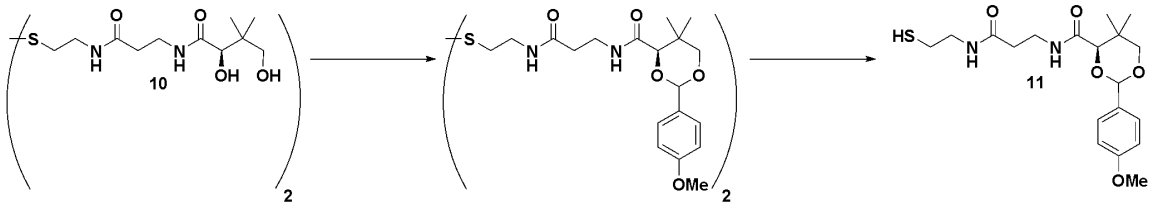
General synthetic procedures and materials

All commercial reagents (Sigma-Aldrich, Spectrum, MP Biomedicals, Alfa Aesar, TCI America, Acros) were used as provided unless otherwise indicated. 2-octynoyl-N-pantetheine **7**,[12] 2-bromo-hexanoyl-N-pantetheine **8**,[12] 3-decynoic acid **12**,[33] 3-decynoyl chloride **14**,[26] acetonide protected-S-pantetheine **13**,[34] *p*-methoxybenzylidene (PMB) protected pantothenic acid **16**,[28,35] PMB-protected N-pantetheine **20**,[28] and 7-dimethyl-aminocoumarin-4-acetic acid-cystamine **23**[36]were each prepared according to published literature procedures. 2-octynoyl chloride **21** was prepared from the commercially available 2-octynoic acid using oxalyl chloride in a procedure analogous to that for preparation of 3-decynoyl chloride.[26] All reactions were carried out under argon atmosphere in dry solvents with oven-dried glassware and constant magnetic stirring unless otherwise noted. Triethylamine (TEA), N-methyl morpholine (NMM), and ethyl-*N,N*-diisopropylamine (DIPEA) were dried over sodium and freshly distilled prior to use. ^{1}H-NMR spectra were taken at 300, 400, or 500 MHz and ^{13}C-NMR spectra were taken at 100.6 or 75.5 MHz on Varian NMR spectrometers and standardized to the NMR solvent signal as reported by Gottlieb[37]. Multiplicities are given as s=singlet, d=doublet, t=triplet, q=quartet, p=pentet, dd=doublet of doublets, bs=broad singlet, bt=broad triplet, m=multiplet using integration and coupling constant in Hertz. TLC analysis was performed using Silica Gel 60 F254 plates (EM Scientific) and visualization was accomplished with UV light (λ=254 nm) and/or the appropriate stain (iodine, 2,4-dinitrophenylhydrazine, cerium molybdate, ninhydrin). Silica gel chromatography was carried out with Silicycle 60 Angstrom 230-400 mesh according to the method of Still.[38]

Electrospray (ESI) and fast atom bombardment (FAB) mass spectra were obtained at the UCSD Mass Spectrometry Facility using a Finnigan LCQDECA mass spectrometer and a ThermoFinnigan MAT- 900XL mass spectrometer, respectively.

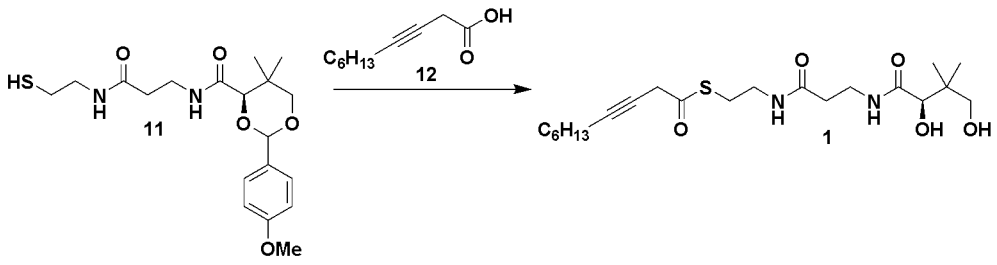
Synthetic Procedures and Spectroscopic Data for Chemoenzymatic Crosslinking Probes 1-9

Synthesis of 3-decynoyl-S-pantetheine (1)



D-pantethine monohydrate **10** (1144 mg, 2.0 mmol was dissolved in DMF (20 mL) with stirring at room temperature. To this mixture was added *p*-anisaldehyde dimethyl acetal (0.68 mL, 4.0 mmol) followed by a solution of 4M HCl in dioxanes (0.1 mL). After stirring overnight the solvent was removed under reduced pressure to yield a crude oil which was further purified by flash chromatography (CH_2Cl_2 to 5% MeOH/ CH_2Cl_2) to yield PMB-pantethine-disulfide (948 mg, 60% for first step) as a white foam. In order to access the reduced thiol, this product was dissolved in distilled MeOH (50 mL) and cooled to 0^oC. Sodium borohydride (1.0 g, 27 mmol) was then carefully added in four portions producing vigorous bubbling. After this initial stage of the reaction had subsided, the reaction vessel was placed under argon and stirred for two hours at room

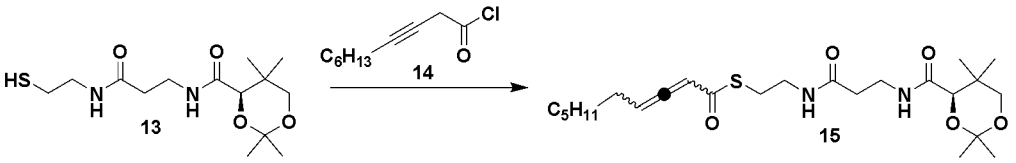
temperature. The reaction was quenched by dropwise addition of water using an addition funnel, and neutralized with HCl. The reaction mixture was then combined with brine (100 mL) and extracted three times with ethyl acetate (200 mL). The extracts were combined, washed again with brine, and the solvent removed under reduced pressure giving the free thiol **11** as a white foam (929 mg, 59% over two steps) with a characteristic faint, putrid odor. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.40 (d, *J*= 8.4 Hz, 2H), 7.06 (bt, 1H), 7.00 (bt, 1H), 6.88 (d, *J*=8.4 Hz, 2H), 5.42 (s, 1H), 4.03 (s, 1H), 3.78 (s, 3H), 3.67 (d, *J*=11.2 Hz, 1H), 3.61 (d, *J*=11.2 Hz, 1H), 3.47 (m, 4H), 2.71 (t, *J*=6.0 Hz, 2H), 2.41 (t, *J*=6.4 Hz, 2H), 1.06 (s, 3H), 1.04 (s, 3H). ^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 172.0, 169.9, 160.5, 130.2, 127.6, 113.7, 101.3, 83.8, 78.4, 55.4, 42.0, 36.2, 34.4, 33.2, 24.5, 22.0, 19.4. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{19}H_{28}O_5N_2S_1$, 396.1719, found 396.1714.



3-decynoic acid **12** (218 mg, 1.3 mmol) was dissolved in THF (4 mL) and cooled to 0^{o}C with stirring. N,N'-dicyclohexylcarbodiimide (267 mg, 1.3 mmol) was then added and stirred for 10 minutes before addition of thiol **11** (126 mg, 0.32 mmol) The reaction was allowed to slowly warm to room temperature and stirred for 15 hours. The reaction was filtered twice to remove urea byproducts and the solvent removed under reduced pressure. Resuspension in CH_2Cl_2 followed by flash chromatography (CH_2Cl_2 to 2%

MeOH/ CH_2Cl_2) yielded the coupled, PMB-protected thioester as a brown oil contaminated with 3-decynoic acid (3:1 product:acid as judged by ^{1}H-NMR). This material was dissolved in 1N HCl:THF, 1:8 (30 ml) and stirred until the starting material was consumed as shown by TLC (2,4-dinitrophenyl-hydrazine visualization). The reaction was then neutralized by addition of AG-1-X8 Strong Basic anionic exchange resin. After filtration, the solvent was removed under reduced pressure, and the reaction mixture was resuspended and purified by flash chromatography (1:6 EtOAc/hexanes to 2:1 EtOAc/hexanes) to afford compound **1** as an oil (120 mg, 88% over two steps). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.40 (bt, 1H), 6.40 (bt, 1H), 3.99 (s, 1H), 3.57-3.37 (m, 6H), 3.03 (m, 2H), 2.42 (t, *J*=5.0 Hz, 2H), 2.21 (t, *J*=5.0 Hz, 2H), 1.53-1.24 (m, 8H), 1.00 (s, 3H), 0.91 (s, 3H), 0.88 (t, *J*=7.0 Hz, 3H).^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 196.4, 173.9, 171.9, 87.0, 77.8, 76.9, 71.1, 39.6, 35.9, 35.4, 35.1, 31.5, 29.0, 28.8, 28.7, 22.8, 22.8, 21.8, 20.6, 19.0, 14.3. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{21}H_{36}O_5N_2S_1$, 428.2339, found 428.2342.

Synthesis of 2,3-decadienoyl-S-pantetheine (2)



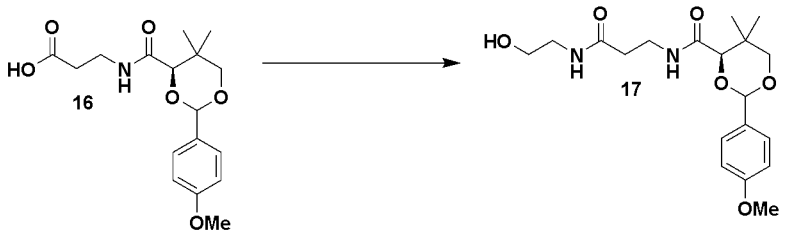
3-decynoyl chloride **14** (41 mg, 0.22 mmol) was added dropwise to a stirring solution of acetonide protected-S-pantetheine **13** (55 mg, 0.17 mmol) and triethylamine (0.082 mL, 0.6 mmol) in THF (5 mL) prechilled to 0^{o}C. After stirring for 1 hour, the

reaction was diluted in EtOAc (50 mL) and washed with citric acid (1x50 mL), saturated $NaHCO_3$ (2x50 mL), and brine (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. Resuspension and purification by flash chromatography (1:6 EtOAc/hexanes to 2:1 EtOAc/hexanes) afforded the protected allene **15** (50 mg, 63%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.03 (bt, 1H), 6.56 (bt, 1H), 5.79 (m, 1H), 5.71 (q, *J*=6.8 Hz, 1H), 4.01 (s, 1H), 3.62 (d, *J*=11.6 Hz, 1H), 3.52-3.35 (m, 4H), 3.21 (d, *J*=11.6 Hz, 1H), 2.96 (q, *J*=6.4 Hz, 2H), 2.38 (t, *J*=6.0 Hz, 2H), 2.12 (m, 2H), 1.45-1.18 (m, 14H), 0.97 (s, 3H), 0.91 (s, 3H), 0.83 (bt, 3H). $[M+H]^+$ calcd for $C_{24}H_{40}O_5N_2S_1$, 468.2658, found 468.2652.

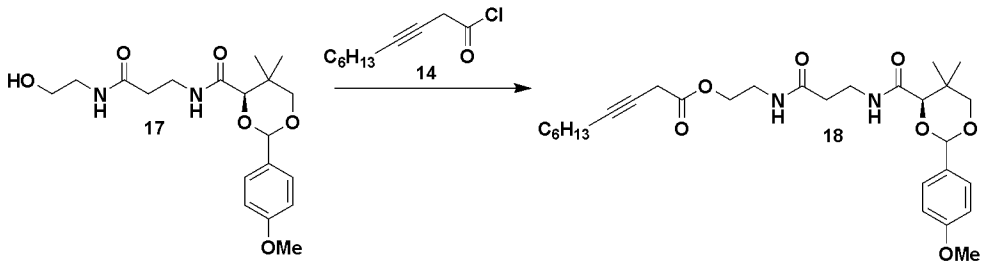
Allenic acetonide **15** was dissolved in 1N HCl:THF, 3:8 (6 ml) and stirred until the benzylidene protected starting material was consumed as shown by TLC (2,4-dinitrophenylhydrazine visualization). AG-1-X8 Strong Basic anionic exchange resin was added to neutralize the solution. After removal of solvent under reduced pressure, the reaction mixture was purified by flash chromatography (1:1 EtOAc/hexanes to EtOAc to 2%MeOH/EtOAc) to afford **2** as a mixture of diasteromeric allenes (38 mg, 63% over two steps). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.44 (bt, 1H), 6.57 (bt, 1H), 5.83 (m, 1H), 5.76 (q, *J*=5.2 Hz, 1H), 3.97 (s, 1H), 3.52-3.32 (m, 6H), 2.99 (m, 2H), 2.41 (m, 2H), 2.17 (m, 2H), 1.49-1.26 (m, 8H), 0.97 (s, 3H), 0.89 (s, 3H), 0.862 (bt, 3H). ^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 211.6, 191.2, 191.1, 174.1, 172.0, 98.8, 96.5, 77.7, 71.0, 40.0, 39.5,

35.9, 35.5, 31.7, 29.9, 28.9, 28.8, 28.3, 27.9, 22.8, 21.7, 21.6, 20.6, 19.0, 14.3. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{21}H_{36}O_5N_2S_1$ 428.2345, found 428.2348.

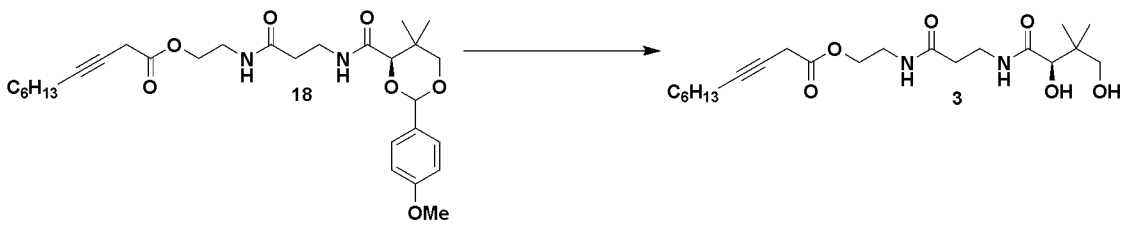
Synthesis of 3-decynoyl-O-pantetheine (3)



PMB-pantothenic acid **16** (2.36 g, 7.0 mmol), ethanolamine (0.843 mL, 14.0 mmol), DIPEA (16.3 mL, 14.0 mmol), and HOBt (2.63 g, 17.5 mmol) were dissolved in DMF with stirring and cooled to 0 ^oC. EDC (2.67 g 14 mmol) was added in one portion and the reaction was allowed to slowly warm to RT and stir for 24 hrs. The solvent was removed under reduced pressure, taken up in EtOAc (100 mL), and extracted w/ saturated $NaHCO_3$ (3x75 mL), water (1x50 mL), and brine (1x75 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent removed under reduced pressure.to provide PMB-oxypantetheine **17** (2.37 g, 89%) as a white foam. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.41 (d, *J*= 9.0 Hz, 2H), 7.09 (bt, 1H), 6.90 (d, *J*=8.5 Hz, 2H), 6.84 (bt, 1H), 5.44 (s, 1H), 4.05 (s. 1H), 3.79 (s, 3H), 3.68 (d, *J*=11.5 Hz, 1H) 3.64-3.60 (m, 3H), 3.52-3.49 (m, 2H), 3.33 (t, *J*=5.5 Hz, 2H), 2.42 (t, *J*=6.5 Hz), 1.08 (s, 3H), 1.05 (s, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 172.1, 170.1, 160.5, 130.3, 127.8, 113.9, 101.5, 84.0, 78.6, 61.9, 55.6, 42.6, 36.2, 35.4, 33.3, 22.1, 19.4. HRMS (EI) (*m/z*) $[M]^+$ calcd for $C_{19}H_{28}O_6N_2$, 380.1942, found 380.1947.

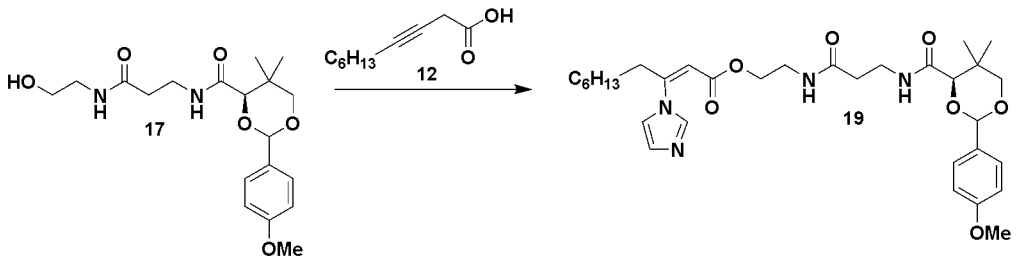


A solution of PMB-oxypantetheine **17** (200 mg, 0.53 mmol) and pyridine (0.05 mL, 0.54 mmol) in CH_2Cl_2 (5 mL) was added dropwise to a stirring solution of 3-decynoyl chloride **14** (50 mg, 0.27 mmol) in CH_2Cl_2 (5 mL) prechilled to 0^oC. The addition step was visibly exothermic and caused the reaction to turn a deep brown/black color. After stirring for 1.5 hours, the reaction was quenched with MeOH (1 mL), diluted in EtOAc (75 mL) and washed with brine (1x100 mL), saturated $NaHCO_3$ (2x75 mL), and brine again (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. Resuspension and purification by flash chromatography (1:1 EtOAc/hexanes to EtOAc) afforded a mixture the protected pantetheine analogue **18** (18.5 mg, 13%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.42 (d, *J*= 8.8 Hz, 2H), 7.08 (bt, 1H), 6.91 (d, *J*=8.8 Hz, 2H), 6.18 (bt, 1H), 4.16 (t, *J*=4.8 Hz, 2H), 4.07 (s, 1H), 3.82 (s, 3H), 3.70 (d, *J*=11.2 Hz, 1H), 3.66 (d, *J*=11.2 Hz, 1H), 3.54-3.51 (m, 4H), 3.27 (t, *J*=2.8 Hz, 2H), 2.44 (t, *J*=6.2 Hz, 2H), 2.18 (m, 2H), 1.61-1.25 (m, 8H), 1.09 (s, 6H), 0.88 (t, *J*=7.2 Hz, 3H). ^{13}C-NMR (100.6 MHz, $CDCl_3$) δ 171.3, 169.8, 168.3, 160.4, 130.3, 127.7, 113.9, 101.5, 84.0, 78.6, 64.3, 63.8, 55.5, 38.9, 38.7, 36.2, 35.1, 33.3, 31.8, 31.5, 29.2, 28.1, 26.2, 22.8, 22.1, 19.0, 14.3. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{29}H_{42}O_7N_2$, 530.2987, found 530.2993.

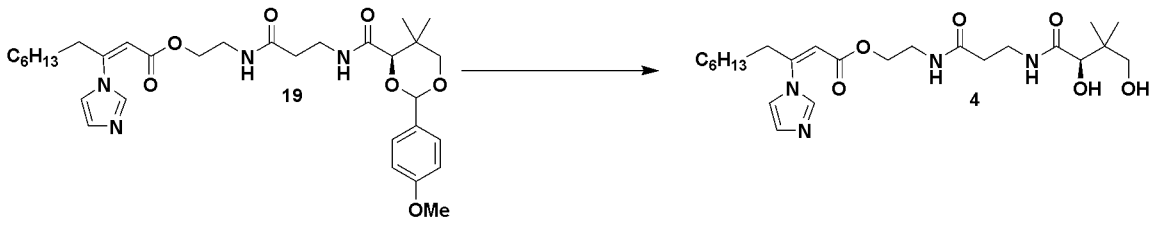


Product **18** (18.5 mg, 0.03 mmol) was dissolved in 1N HCl:THF, 1:3 (6 ml) and stirred until the benzylidene protected starting material was consumed as shown by TLC (2,4-dinitrophenylhydrazine visualization). The reaction was neutralized by dropwise addition of 1M NaOH, diluted in brine (50 mL) and extracted with CH_2Cl_2 (3x50 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. Resuspension and purification by flash chromatography (EtOAc to 5% MeOH/EtOAc) afforded **3** as an oil (8.5 mg, 59%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.36 (bt, 1H), 6.29 (bt, 1H), 4.27-4.17 (m, 2H), 4.00 (s, 1H), 3.63-3.41 (m, 6H), 3.28 (s, 2H), 2.49-2.43 (m, 2H), 2.22-2.18 (m, 2H), 1.52-1.25 (m, 8H), 1.02 (s, 3H), 0.89 (s, 3H), 0.88 (t, *J*=7.5 Hz, 3H). ^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 173.9, 171.9, 169.7, 84.6, 77.9, 71.2, 64.3, 39.6, 39.0, 36.0, 35.3, 31.6, 29.9, 28.9, 26.2, 22.8, 21.9, 20.6, 19.0, 14.3. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{21}H_{35}O_5N_2$, 412.2568, found 412.2575.

Synthesis of (*Z*)-2-decenoyl-3-imidazole-O-pantetheine (4)

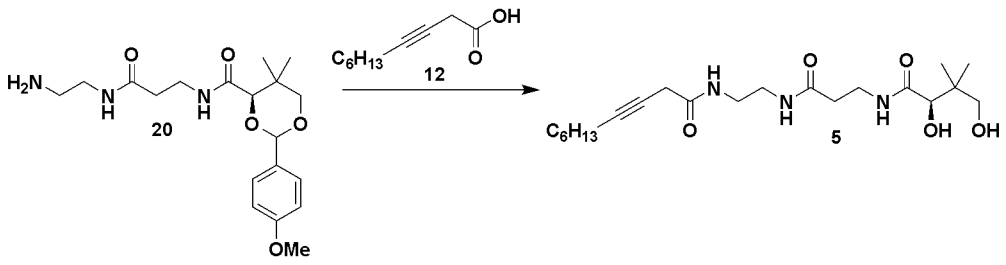


Carbonyl diimidazole (136 mg, 1 mmol) was added to a stirring solution of 3-decynoic acid **12** (169 mg, 1 mmol) dissolved in dry DMF (10 mL). The reaction was heated at 50^{o}C for 20 minutes, during which time the reaction mixture turned a pale orange. To this mixture was added dropwise a solution of PMB-oxypantetheine **17** (760 mg, 2 mmol) and 1,8-diazabicyclo[5.4.0]undec-7-ene (101 mg, 1 mmol) in DMF (1 mL). After stirring overnight the solvent was removed under reduced pressure, and the reaction mixture resuspended and purified by flash chromatography (2:1 EtOAc/hexanes to 5% MeOH/EtOAc) to afford recovered PMB-oxypantetheine starting material (590 mg) as well as compound **19** as an oil (221 mg, 42%).^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.82 (s, 1H), 7.41 (d, *J*=8.1 Hz, 2H), 7.18 (s, 1H), 7.13 (s, 1H), 7.03 (bt, 1H), 6.90 (d, *J*=8.0 Hz, 2H), 6.41 (bt, 1H), 6.92 (s, 1H), 5.44 (s, 1H), 4.19 (t, *J*=5.5 Hz, 2H), 4.05 (s, 1H), 3.80 (s, 3H), 3.66 (d, *J*=15.5 Hz, 1H), 3.63 (d, *J*=15.5 Hz, 1H), 3.57-3.49 (m, 4H), 3.11 (t, *J*=8.0 Hz, 2H), 2.44 (t, *J*=6.0 Hz, 2H), 1.51 (p, *J*=8.0 Hz, 2H), 1.37-1.22 (m, 6H), 1.21 (s, 3H), 1.06 (s, 3H), 0.85 (t, *J*=6.5 Hz, 3H). ^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 171.2, 169.8, 165.7, 160.5, 153.9, 135.5, 131.2, 130.3, 127.7, 117.0, 114.0, 107.1, 101.6, 84.0, 78.7, 63.4, 55.6, 38.8, 36.3, 35.1, 33.3, 31.9, 29.9, 29.6, 29.2, 28.5, 22.8, 22.1, 19.3, 14.3. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{32}H_{46}O_7N_4$, 598.3367, found 598.3364.



Product **19** (28 mg, 0.05 mmol) was dissolved in 1N HCl:THF, 1:4 (10 ml) and stirred until the starting material was consumed as shown by TLC (2,4-dinitrophenylhydrazine visualization). The reaction was then neutralized by addition of AG-1-X8 Strong Basic anionic exchange resin. After filtration, the solvent was removed under reduced pressure, and the reaction mixture was resuspended and purified by flash chromatography (EtOAc to 5% MeOH/EtOAc) to afford compound **4** as an oil (11 mg, 49%). ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.86 (s, 1H), 7.46 (bt, 1H), 7.20 (s, 1H), 7.11 (s, 1H), 6.87 (bt, 1H), 5.96 (s, 1H), 4.20-4.19 (m, 2H), 3.97 (s, 1H), 3.58-3.43 (m, 6H), 3.10 (t, *J*=8.0 Hz, 2H), 2.45 (t, *J*=6.5 Hz, 2H), 1.50 (p, *J*=7.5 Hz, 2H), 1.37-1.24 (m, 6H), 0.97 (s, 3H), 0.88 (s, 3H), 0.85 (t, *J*=6.5 Hz, 3H).^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 174.2, 171.9, 165.8, 153.8, 136.2, 130.9, 117.8, 107.4, 77.7, 71.1, 63.4, 39.5, 38.9, 36.1, 35.4, 31.9, 30.0, 29.6, 29.1, 28.5, 22.8, 21.6, 20.7, 14.3. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{24}H_{40}O_6N_4$, 480.2942, found 480.2941.

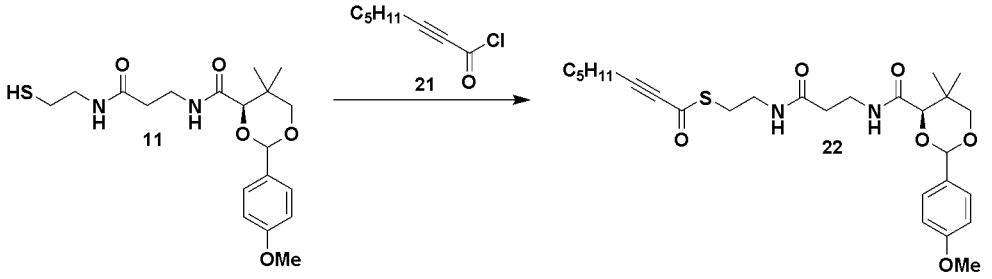
Synthesis of 3-decynoyl-N-pantetheine (5)



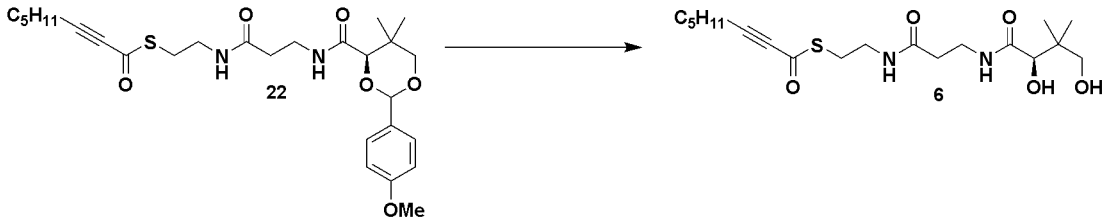
PMB-pantetheine amine **20** (89 mg, 0.23 mmol), 3-decynoic acid **12** (116 mg, 0.69 mmol), DIPEA (0.08 mL, 0.47 mmol), and HOBt (89 mg, 0.58 mmol) were dissolved in DMF (4 mL) with stirring and cooled to 0 ^{o}C. EDC (89 mg 0.47 mmol) was added in one portion and the reaction was allowed to slowly warm to RT and stir for 24

hrs. The solvent was removed under reduced pressure, taken up in EtOAc (75 mL), and extracted w/ water (1x50 mL), saturated $NaHCO_3$ (3x75 mL), and brine (1x50 mL). The organic layer was dried over Na_2SO_4, filtered, and the solvent removed under reduced pressure. This material was resuspended and run through a silica gel plug (CH_2Cl_2 to 3% MeOH/ CH_2Cl_2) to afford the crude coupled material. compound **19** as an oil (78 mg). The coupled material (78 mg) was then dissolved in 1N HCl:THF, 1:1 (8 ml) and stirred until the starting material was consumed as shown by TLC (2,4-dinitrophenylhydrazine visualization). The reaction was then neutralized by addition of AG-1-X8 Strong Basic anionic exchange resin. After filtration, the solvent was removed under reduced pressure, and the reaction mixture was resuspended and purified by flash chromatography (CH_2Cl_2 to 3% MeOH/ CH_2Cl_2) to afford compound **5** as an oil (59 mg, 62% over two steps). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.45 (bt, 1H), 7.2 (bt, 1H), 7.15 (bt, 1H), 3.97 (s, 1H), 3.57-3.27 (m, 6H), 3.15 (t, *J*=2.4 Hz, 2H), 2.41 (m, 2H), 2.20 (t, *J*=6.8 Hz, 2H), 1.49 (p, *J*=7.2 Hz, 2H), 1.38-1.24 (m, 6H), 0.96 (s, 3H), 0.89 (s, 3H), 0.87 (t, *J*=6.8 Hz, 3H). ^{13}C-NMR (100.5 MHz, ($CDCl_3$) δ 174.4, 172.5, 169.7, 86.9, 77.6, 72.6, 70.9, 40.2, 39.9, 39.5, 36.2, 35.5, 31.5, 28.9, 28.8, 27.9, 22.8, 21.6, 20.8, 19.0, 14.3. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{21}H_{37}O_5N_3$, 411.2728, found 411.2728.

Synthesis of 2-octynoyl-S-pantetheine (6)

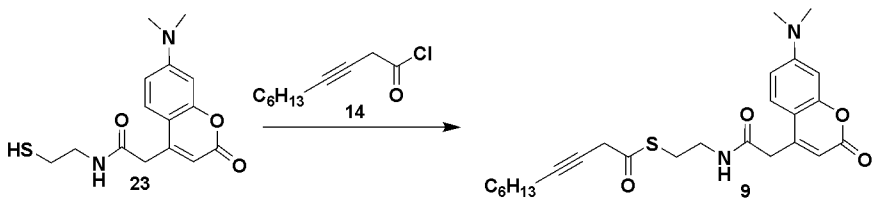


A solution of reduced thiol **11** (300 mg, 0.76 mmol) and triethylamine (0.51 mL, 3.7 mmol) in CH_2Cl_2 (5 mL) was added dropwise to a stirring solution of 2-octynoyl chloride **21** (150 mg, 0.94 mmol) in CH_2Cl_2 (5 mL) prechilled to 0^oC. The addition step was visibly exothermic and caused the reaction to turn a pale orange/red color. After stirring for 1 hour, the reaction was quenched with MeOH (1 mL), diluted in EtOAc (75 mL) and washed with brine (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. Resuspension and purification by flash chromatography (1:1 EtOAc/hexanes to 2:1 EtOAc/hexanes) resulted in recovery of PMB-pantetheine **22** (200 mg, 0.51 mmol) as well as isolation of **21** (40 mg, 10%) as a crude oil. ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.41 (d, *J*=8.8 Hz, 2H), 7.02 (bt, 1H), 6.90 (d, *J*=8.4 Hz, 2H), 6.35 (bt, 1H), 5.45 (s, 1H), 4.07 (s, 1H), 3.81 (s, 3H), 3.69 (d, *J*=11.6 Hz, 1H), 3.64 (d, *J*=12.0 Hz, 1H), 3.54-3.36 (m, 4H), 3.02 (t, *J*=6.4 Hz, 2H), 2.42-2.35 (m, 4H), 1.58 (p, *J*=7.2 Hz, 2H), 1.39-1.29 (m, 4H), 1.09 (s, 3H), 1.08 (s, 3H), 0.89 (t, *J*=7.6 Hz, 3H). ^{13}C-NMR (100.5 MHz, $(CD_3)_2SO$) δ 176.6, 171.4, 169.8, 160.4, 130.4, 127.7, 113.9, 101.6, 97.0, 84.1, 78.8, 78.7, 55.6, 39.4, 36.1, 35.1, 33.3, 31.2, 29.5, 27.4, 22.3, 22.1, 19.4, 19.3, 14.1. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{27}H_{38}O_6N_2S_1$, 518.2451, found 518.2455.



Protected pantetheine **22** (28 mg, 0.06 mmol) was dissolved in 1N HCl:THF, 1:4 (5 ml) and stirred until the starting material was consumed as shown by TLC (2,4-dinitrophenylhydrazine visualization). The reaction was then neutralized by addition of saturated $NaHCO_3$ (10 mL), and extracted with EtOAc (3x25 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. Resuspension and purification by flash chromatography (1:1 EtOAc/hexanes to 1% MeOH/EtOAc) resulted in isolation of **6** as an oil (20 mg, 91%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.48 (bt, 1H), 6.69 (bt, 1H), 4.1-4.03 (m, 1H), 3.99 (s, 1H), 3.57-3.36 (m, 6H), 3.13-3.04 (m, 2H), 2.43 (t, *J*=6.0 Hz, 2H), 2.37 (t, *J*=6.8 Hz, 2H), 1.62-1.55 (m, 2H), 1.41-1.29 (m, 4H), 0.98 (s, 3H), 0.94-0.88 (m, 6H). ^{13}C-NMR (75 MHz, ($CDCl_3$) δ 176.9, 174.2, 172.2, 97.5, 78.7, 78.5, 39.5, 39.4, 35.8, 35.5, 31.2, 29.4, 29.3, 27.4, 22.3, 21.7, 20.6, 19.3, 14.1. HRMS (EI) (*m/z*): $[M+H]^+$ calcd for $C_{19}H_{32}O_5N_2S_1$, 400.2032, found 400.2026.

Synthesis of 3-decynoyl-N-DMC-cystamine (9)



A solution of 7-dimethylamino-coumarin thiol **23** (23 mg, 0.08 mmol) and triethylamine (0.011 mL, 0.08 mmol) in DMF (1 mL) was added to a stirring solution of 3-decynoyl chloride **14** (70 mg, 0.375 mmol) in CH_2Cl_2 (5 mL) prechilled to 0^oC. Care must be taken that only a small amount of triethylamine is added to the reaction, to prevent chemical isomerization to the allenic thioester. After stirring for 1 hour, the reaction was quenched with MeOH (1 mL), diluted in EtOAc (75 mL) and washed with saturated Na_2HCO_3 (1x25 mL) and brine (1x100 mL). The organic layer was dried over Na_2SO_4, filtered, and evaporated under reduced pressure. Resuspension and purification by flash chromatography (1:6 EtOAc/hexanes to 2:1 EtOAc/hexanes) resulted in isolation of fluorescent oil **9** (10 mg, 29%) as well as recovery of the disulfide of **22** (10 mg). Compound **9** can alternatively be prepared by coupling of 3-decynoic acid **12** and **22** by DCC without the addition of base (see synthetic notes for compound **1**).^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.43 (d, *J*= 8.8 Hz, 1H), 6.61 (d, *J*= 9.2 Hz, 1H), 6.51 (s, 1H), 6.09 (bt, 1H), 6.03 (s, 1H), 3.61 (s, 2H), 3.43 (t, *J*=5.6 Hz, 2H), 3.31 (s, 2H), 3.06 (s, 6H), 2.97 (t, *J*=6.0 Hz, 2H), 2.20 (m, 2H), 1.39-1.25 (m, 8H), 0.88 (m, 3H). ^{13}C-NMR (100.5 MHz, $(CD_3)_2SO$) δ 195.3, 168.7, 161.4, 156.1, 153.5, 151.7, 126.7, 110.2, 109.7, 108.9, 98.2, 86.2, 73.0, 45.1, 40.8, 40.2, 38.9, 31.7, 31.4, 28.9, 28.5, 22.7, 18.8, 14.6. HRMS (EI) (*m/z*): $[M]^+$ calcd for $C_{25}H_{32}O_4N_2S_1$, 456.2077, found 456.2073.

Protein Expression and Materials

Recombinant *E. coli* CoA biosynthetic enzymes (pantothenate kinase, phosphopantetheine-adenyltransferase, dephospho-CoA kinase) were expressed and purified as MBP-fusions as described previously.[29] Recombinantally expressed Sfp (PPTase; *B. subtilis*) was overexpressed in *E. coli* as the native protein (untagged) and purified as described previously.[29] The carrier proteins AcpP (*E. coli* FAS), Fren (frenolicin PKS; *Streptomyces roseofulvus*), Otc (oxytetracycline PKS; *Streptomyces rimosus*), and EntB (enterobactin NRPS; *E. coli*) were overexpressed and purified as the 6xHis tagged constructs as described previously.[8,39] FabA (*E. coli* DH domain) and the GFP-AcpP fusion were produced and isolated as the 6xHis tagged constructs using the respective *E. coli* overexpression strains from the ASKA collection, kindly provided by Dr. Hirotada Mori at Keio University, Japan.[30] Briefly, proteins were expressed by growth of *E. coli* (K12, strain AG1) ASKA clones harboring the desired plasmid in Luria Bertani medium supplemented with 30 μg/ml chloramphenicol at 37°C. At OD 0.6 (600 nm) protein expression was induced using 1mM isopropyl β-D-1 thiogalactopyranoside, and cells were cultured with shaking for approximately 16 hr. Cells were pelleted and resuspended in lysis buffer (100mM potassium phosphate, 300mM NaCl, pH 8.0) and incubated with lysozyme (1 mg/mL) for 1hr, then lysed by passage through a French pressure cell. After treatment with DNAse cellular debris was separated from overexpressed enzyme by centrifugation and the supernatant was loaded onto a Ni^{2+}-NTA column and eluted with a gradient of imidazole (5-300 mM). Ni-NTA affinity chromatography and all steps following growth were performed at 4°C. Eluted proteins

were visualized by SDS-PAGE with Blue-silver Coomassie stain[40] and quantitated by the method of Bradford.[41]

Procedures for ACP-DH Crosslinking

Crosslinking reactions consisted of two sequential steps. *First*, electrophilic pantetheine analogues (**1**, **2**, or **6**) were loaded onto AcpP through a one-pot chemoenzymatic reaction protocol which utilized the CoA biosynthetic enzymes (PanK, PPAT, DPCK) to form the corresponding CoA analogues in situ, followed by loading onto *apo*-AcpP by the promiscuous PPTase Sfp. Reagents and enzymes were added to 27 μL reactions in sequential order as follows. Final reaction concentrations: 55 mM potassium phosphate (pH 7.0), 27 mM ATP, 47 mM $MgCl_2$, 0.1 μg/μL MBP-PanK, 0.1 μg/μL MBP-PPAT, 0.1 μg/μL MBP-DPCK, 0.08 μg/μL Sfp (native), 0.37 μg/μL AcpP, 185 μM pantetheine analogue (**1**, **2**, or **6**). In negative controls pantetheine analogues **1-3** were replaced with vehicle DMSO. In all reactions DMSO concentrations were maintained at a level ≤ 1 %. For affinity purification of crosslinked complex, reactions were performed similarly on a 570 μL scale. For crosslinking studies of alternate carrier proteins (GFP AcpP, Fren, Otc, EntB), the corresponding carrier protein was substituted for native AcpP. After addition of all components reactions were vortexed and incubated at 37°C for 1-3 hours. In the *second* step of the reaction FabA was added, and reactions were further incubated for 3-12 hours before quenching with 5x SDS loading buffer (strong reducing) and analysis by SDS-PAGE. Gels were fixed in destain solution (50:43:7 MeOH:H_2O:AcOH) and washed twice with water prior to staining with blue-silver Coomassie stain.

Procedure for Affinity Purification of ACP-DH Crosslinked Complex

AcpP-FabA and GFP AcpP-FabA crosslinking reactions were performed on a 570 μL scale using pantetheine analogues **1** or **6** as described above. FabA was added after 3 hours and reactions were incubated overnight. Reactions were then added to 60 μL Ni^{2+}-NTA resin (pre-washed with 50 mM sodium phosphate, pH 7.0) and rotated at room temperature for 1 hour. Reactions were centrifuged, supernatant removed, and 1000 μL washing buffer 1 (50 mM sodium phosphate, pH 7.0) added. After gentle mixing for 5 minutes, reactions were centrifuged, the supernatant removed, and 1000 μL washing buffer 2 (5 mM imidazole, 50 mM sodium phosphate, pH 7.0) added. After gentle mixing for 5 minutes, reactions were centrifuged, the supernatant removed, and 100 μL elution buffer 1 (50 mM imidazole, 50 mM sodium phosphate, pH 7.0) added. After gentle mixing for 5 minutes, reactions were centrifuged, the supernatant removed, and 100 μL elution buffer 2 (400 mM imidazole, 50 mM sodium phosphate, pH 7.0) added. After gentle mixing for 5 minutes, reactions were centrifuged, the supernatant removed, and 100 μL rinsing buffer (2500 mM imidazole, 50 mM sodium phosphate, pH 7.0) added. After gentle mixing for 5 minutes, reactions were centrifuged, the supernatant removed. All fractions were analyzed by SDS-PAGE, fixed, and visualized by staining with Blue-silver Coomassie stain.

Procedures for ACP-KS Crosslinking

ACP-KS crosslinking of AcpP and FabB was performed in a similar fashion as above, using a slightly modified version of the previously reported protocol. Briefly, to a

buffered solution consisting of 55 mM potassium phosphate (pH 7.0), 50 mM $MgCl_2$, and 27 mM ATP was added 6xHis-PanK (0.1 μg/μL), 6xHis-PPAT (0.1 μg/μL), 6xHis-DPCK (0.1 μg/μL), Sfp (native) (0.08 μg/μL), and AcpP (0.37 μg/μL) to bring the total reaction volume to 27 μL. Note that 6xHis tagged CoA biosynthetic proteins are used in this protocol (in contrast to the MBP-tagged proteins of the ACP-DH crosslinking reaction) to facilitate visualization of the AcpP-FabB crosslinked complex (~ 70 kDa). Electrophilic pantetheine analogues (**1**, **2**, or **6**) were then added to a final concentration of 185 μM. In negative controls pantetheine analogues **1-3** were replaced with vehicle DMSO. In all reactions DMSO concentrations were maintained at a level ≤ 1 %. Reactions were incubated for 30 minutes at 37 °C, followed by addition of ketosynthase enzyme FabB (0.1 μg/μL). Reactions were further incubated for another hour before quenching with 5x SDS loading buffer (strong reducing) and analysis by SDS-PAGE. Gels were fixed in destain solution (50:43:7 MeOH:H_2O:AcOH) and washed twice with water prior to staining with blue-silver Coomassie stain.

Procedures for Activity Based Labeling of DH Domains by Fluorescent Suicide Substrate 9

For FabA (DH) activity based labeling experiments, fluorescent probe **9** (25 μM) was added to a 27 uL reaction containing FabA (0.1 μg/μL), Tris-HCl pH 6.8 (100 mM), and DTT (10 mM). Controls were performed by pre-denaturation of FabA by addition of 0.5% SDS. To assess the FabA active site modification properties of pantetheine analogues **1-8**, FabA was preincubated with each analogue for 30 minutes at 1 mM before addition of **9**. In all reactions DMSO concentrations were maintained at a level ≤

3%. After reaction for 30 minutes FabA labeling reactions were quenched with 5x SDS-loading buffer (strong reducing) and subjected to SDS-PAGE. Probe **9** showed high off-site reactivity when heated prior to SDS-PAGE, so boiling of these samples was avoided. Fluorescent gel visualization was performed using a BioRad Fluor-S Gel Doc equipped with a 460 nm emission filter.

Procedures for In Gel Digest of Crosslinked Complex

For MS analysis, proteins were separated by SDS-PAGE using 10% Bis-Tris NuPAGE Gels (Invitrogen). Gels were fixed in destain solution (50:43:7 MeOH:H_2O:AcOH) and washed twice with water prior to overnight staining with blue-silver Coomassie stain. Subsequently, the gel was washed twice with water (30 min each). Excised gel bands were cut into 1 mm cubes and washed for 10 min with vortexing with 1:1 H_2O/CH_3CN (40 µL) followed by CH_3CN (40 µL). After removal of the supernatant, 100 mM NH_4CO_3 (40 µL) was added and shaken for 5 min, followed by addition of CH_3CN (40 µL) and shaking for 10 min. The gel slice was then dried on a speed vacuum concentrator system. The gel slice was then resuspended in 100 mM NH_4CO_3 (100 µL) and TCEP added to a concentration of 10 mM and incubated at 55°C for 30 minutes. The supernatant was removed and the gel slice resuspended in 100 mM NH_4CO_3 (100 µL) and iodoacetamide added to a concentration of 15 mM. The reaction was covered with foil and shaken at room temperature for 30 minutes. After removal of the supernatant, 100 mM NH_4CO_3 (40 µL) was added and shaken for 5 min, followed by addition of CH_3CN (40 µL) and shaking for 10 min. The gel slice was then dried on a speed vacuum concentrator system. The gel slice was then resuspended by addition of 20

μL 10 ng/μL trypsin (Promega) in a 5 mM DTT, 25 mM NH_4CO_3, 5 mM $CaCl_2$ solution and incubated on ice 45 minutes. Additional trypsin was added if necessary, and the digest allowed to continue at 37°C overnight. The supernatant was then extracted and the gel slice washed multiple times (1:1 25 mM NH_4CO_3:CH_3CN, 1:1 5% formic acid: CH_3CN) with vortexing. Combining each wash provided ~ 100 μL post-digest supernatant. This material was then dried on a speed vacuum concentrator system and resuspended in 15 μL 5% formic acid for MALDI TOF/TOF analysis.

MALDI TOF/TOF Analysis of Crosslinked Bands

Each sample from the in gel digest (prepared above) was mixed 1:1 with α-cyano-4-hydroxycinnamic acid (Agilent) and spotted on a MALDI target plate. Peptides were identified by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) on a 4800 MALDI tandem time-of-flight mass spectrometer (Applied Biosystems). Briefly, peptide mass fingerprints were acquired with 500 shots in reflector positive mode from *m/z* 1000 – 3000. Peptide tandem mass spectra were acquired with 2000 shots in 2 kV positive MSMS mode. Peaks occurring with Minimim Chromatogram Peak Width of 2 fractions, a Minimum S/N filter of 40, and a Fraction-to-Fraction Precursor Mass Tolerance of 100 ppm were MSMS acquired in the order of Strongest Precursors First with a Max Precursors/Fraction at 8. Sodium and potassium adducts were excluded with an Adduct Tolerance at +/- 0.03 Da, and precursors < 200 resolution were excluded. A Plate Model and Default Calibration were performed for reflector positive mode as well as a Default calibration for 2 kV positive mode. MSMS spectra were assigned by database searching using Mascot 2.1 (Matrix Science) using Global Proteomics Sever 3.1

(Applied Biosystems). A custom database containing 57 sequences including the *E. coli* proteins AcpP, FabA, PanK, PPAT, DPCK, *B. subtilis* Sfp, the fusion tags GFPuv4 and MBP, and a number of commonly occurring protein contaminants (e.g. keratin) was searched. The MSMS ion search parameters identified tryptic peptides with up to 2 missed cleavages and used mass tolerances of 100 ppm (MS) and 0.10 Da (MSMS), with the constant modification carbamidomethylation (C) due to reductive alkylation by iodoacetamide. Common non-specific variable modifications that occur as a result of the in-gel digest assay were also included in the search parameters: deamidation (NQ), oxidation (M), propionamide (C), and pyro-glu (N-term Q). The search results indicated that individual ion scores > 31 indicate identity or extensive homology ($P < 0.05$).

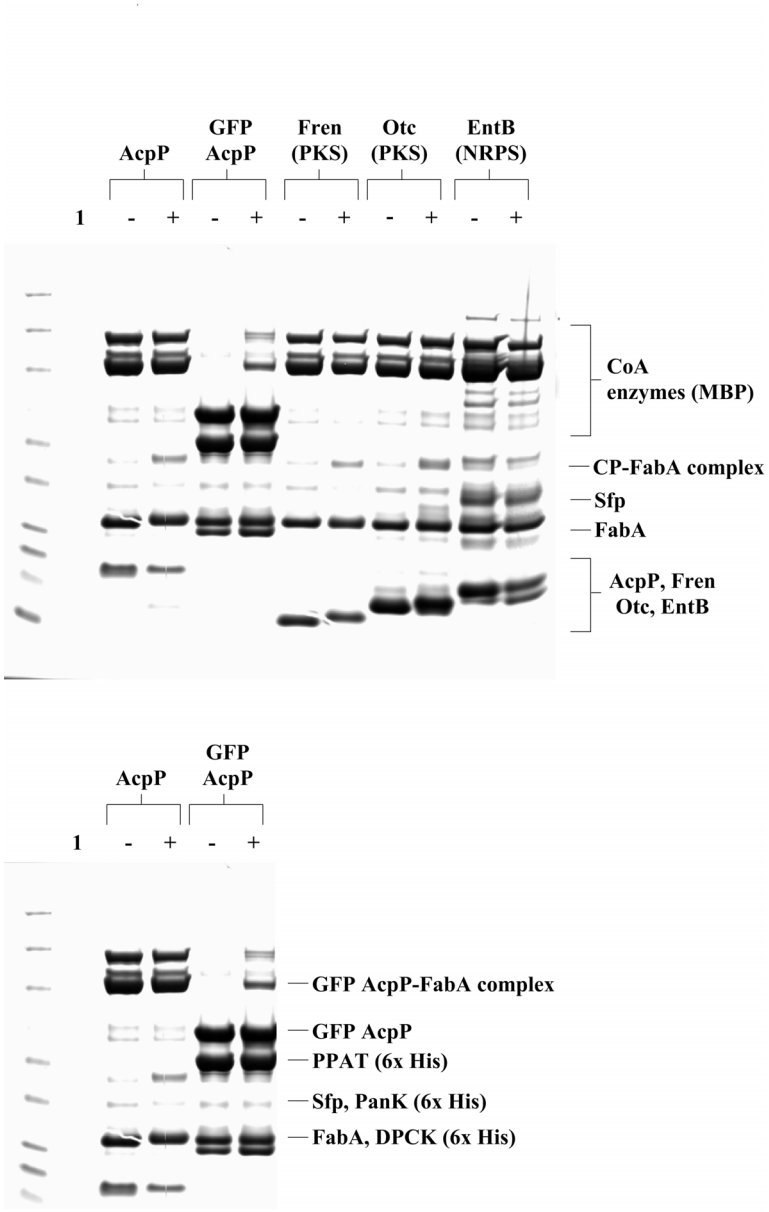


Figure 8.S4 Full gel data for figure 8.3. Effect of carrier protein-dehydratase protein-protein interactions on CP-FabA crosslinking. 3-decynoyl pantetheine **1** was used to modify the carrier proteins AcpP (FAS), Fren (PKS), Otc (PKS), and EntB (NRPS). Upon addition of FabA, crosslinking is observed only with AcpP, Fren, and Otc, indicating the preferential interaction of FabA with 3-decynoyl ACPs.

References

1. Smith, S.; Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **24**, 1041-72 (2007).

2. Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106**, 3468-96 (2006).

3. Weissman, K. J.; Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**, 925-36 (2005).

4. Weissman, K. J.; Muller, R. Protein-protein interactions in multienzyme megasynthetases. *Chembiochem* **9**, 826-48 (2008).

5. Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu Rev Biochem* **76**, 195-221 (2007).

6. Mercer, A. C.; Burkart, M. D. The ubiquitous carrier protein--a window to metabolite biosynthesis. *Nat Prod Rep* **24**, 750-73 (2007).

7. Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288-90 (2007).

8. 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* **1**, 687-91 (2006).

9. 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* **37**, 1585-95 (1998).

10. Worthington, A. S.; Burkart, M. D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org Biomol Chem* **4**, 44-6 (2006).

11. Kapur, S.; Worthington, A.; Tang, Y.; Cane, D. E.; Burkart, M. D.; Khosla, C. Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase. *Bioorg Med Chem Lett* **18**, 3034-8 (2008).

12. Worthington, A. S.; Hur, G. H.; Meier, J. L.; Cheng, Q.; Moore, B. S.; Burkart, M. D. Probing the compatibility of type II ketosynthase-carrier protein partners. *Chembiochem* **9**, 2096-103 (2008).

13. Hur, G. H.; Meier, J. L.; Baskin, J.; Codelli, J. A.; Bertozzi, C. R.; Marahiel, M. A.; Burkart, M. D. Crosslinking studies of protein-protein interactions in nonribosomal peptide biosynthesis. *Chem Biol* **16**, 372-81 (2009).

14. Liu, Y.; Bruner, S. D. Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases. *Chembiochem* **8**, 617-21 (2007).

15. Qiao, C.; Wilson, D. J.; Bennett, E. M.; Aldrich, C. C. A mechanism-based aryl carrier protein/thiolation domain affinity probe. *J Am Chem Soc* **129**, 6350-1 (2007).

16. Heath, R. J.; Rock, C. O. Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in Escherichia coli fatty acid biosynthesis. *J Biol Chem* **271**, 27795-801 (1996).

17. Wu, J.; Zaleski, T. J.; Valenzano, C.; Khosla, C.; Cane, D. E. Polyketide double bond biosynthesis. Mechanistic analysis of the dehydratase-containing module 2 of the picromycin/methymycin polyketide synthase. *J Am Chem Soc* **127**, 17393-404 (2005).

18. Brock, D. J.; Kass, L. R.; Bloch, K. Beta-hydroxydecanoyl thioester dehydrase. II. Mode of action. *J Biol Chem* **242**, 4432-40 (1967).

19. Kass, L. R.; Brock, D. J.; Bloch, K. Beta-hydroxydecanoyl thioester dehydrase. I. Purification and properties. *J Biol Chem* **242**, 4418-31 (1967).

20. Endo, K.; Helmkamp, G. M., Jr.; Bloch, K. Mode of inhibition of beta-hydroxydecanoyl thioester dehydrase by 3-decynoyl-N-acetylcysteamine. *J Biol Chem* **245**, 4293-6 (1970).

21. Helmkamp, G. M., Jr.; Brock, D. J.; Bloch, K. Beta-hydroxydecanoly thioester dehydrase. Specificity of substrates and acetylenic inhibitors. *J Biol Chem* **243**, 3229-31 (1968).

22. Schwab, J. M.; Ho, C. K.; Li, W. B.; Townsend, C. A.; Salituro, G. M. Beta-hydroxydecanoyl Thioester Dehydratase - Complete Characterization of The Fate of the Suicide Substrate 3-Decynoyl-NAC. *J Am Chem Soc* **108**, 5309-5316 (1986).

23. Powell, P. J.; Thorpe, C. 2-octynoyl coenzyme A is a mechanism-based inhibitor of pig kidney medium-chain acyl coenzyme A dehydrogenase: isolation of the target peptide. *Biochemistry* **27**, 8022-8 (1988).

24. Freund, K.; Mizzer, J.; Dick, W.; Thorpe, C. Inactivation of general acyl-CoA dehydrogenase from pig kidney by 2-alkynoyl coenzyme A derivatives: initial aspects. *Biochemistry* **24**, 5996-6002 (1985).

25. Haeffner-Gormley, L.; Cummings, J. G.; Thorpe, C. S-2-bromo-acyl-CoA analogues are affinity labels for the medium-chain acyl-CoA dehydrogenase from pig kidney. *Arch Biochem Biophys* **317**, 479-86 (1995).

26. Morisaki, M.; Bloch, K. On the mode of interaction of -hydroxydecanoyl thioester dehydrase with allenic acid derivatives. *Biochemistry* **11**, 309-14 (1972).

27. Rock, C. O.; Cronan, J. E., Jr. Acyl carrier protein from Escherichia coli. *Methods Enzymol* **71 Pt C**, 341-51 (1981).

28. Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J Am Chem Soc* **128**, 12174-84 (2006).

29. Haushalter, R. W.; Worthington, A. S.; Hur, G. H.; Burkart, M. D. An orthogonal purification strategy for isolating crosslinked domains of modular synthases. *Bioorg Med Chem Lett* **18**, 3039-42 (2008).

30. Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res* **12**, 291-9 (2005).

31. Leesong, M.; Henderson, B. S.; Gillig, J. R.; Schwab, J. M.; Smith, J. L. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**, 253-64 (1996).

32. Mueller, M.; Jenni, S.; Ban, N. Strategies for crystallization and structure determination of very large macromolecular assemblies. *Curr Opin Struct Biol* **17**, 572-9 (2007).

33. Schmieder-van de Vondervoort, L.; Bouttemy, S.; Padrón, J.; Bras, J. L.; Muzart, J.; Alsters, P. Chromium Catalyzed Oxidation of (Homo-)Allylic and (Homo-)Propargylic Alcohols with Sodium Periodate to Ketones or Carboxylic Acids. *Synlett* (2002).

34. Bloom, J. D.; Dutia, M. D., 1991, Simplified thioester and isostere analogues of oleoyl coenzyme A as hypochloesterolemic agents. United States March 29 1990, Issued, October 1 1991.

35. Meier, J. L.; Burkart, M. D. Chapter 9. Synthetic probes for polyketide and nonribosomal peptide biosynthetic enzymes. *Methods Enzymol* **458**, 219-54 (2009).

36. Alexander, M. D.; Burkart, M. D.; Leonard, M. S.; Portonovo, P.; Liang, B.; Ding, X.; Joullie, M. M.; Gulledge, B. M.; Aggen, J. B.; Chamberlin, A. R.; Sandler, J.; Fenical, W.; Cui, J.; Gharpure, S. J.; Polosukhin, A.; Zhang, H. R.; Evans, P. A.; Richardson, A. D.; Harper, M. K.; Ireland, C. M.; Vong, B. G.; Brady, T. P.; Theodorakis, E. A.; La Clair, J. J. A central strategy for converting natural products into fluorescent probes. *Chembiochem* **7**, 409-16 (2006).

37. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J Org Chem* **62**, 7512-7515 (1997).

38. Still, W. C.; Kahn, M. Rapid chromatographic technique for preparative separations with moderate resolution. *Journal of Organic Chemistry* **43**, 2923-2925 (1978).

39. Mercer, A. C.; La Clair, J. J.; Burkart, M. D. Fluorescent multiplex analysis of carrier protein post-translational modification. *Chembiochem* **6**, 1335-7 (2005).

40. Candiano, G.; Bruschi, M.; Musante, L.; Santucci, L.; Ghiggeri, G. M.; Carnemolla, B.; Orecchia, P.; Zardi, L.; Righetti, P. G. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327-33 (2004).

41. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-54 (1976).

Acknowledgments

The text of chapter 8, in full, has been submitted for publication. I am the primary author of the manuscript. Robert Haushalter was responsible for the expression and purification of recombinant CoA biosynthetic enzymes. I designed and performed the experiments. All research was performed under the supervision of Professor Mike Burkart.

Chapter Nine

Global proteomic profiling of polyketide and nonribosomal peptide biosynthesis by orthogonal active site probes

Adapted from:

Meier, J. L.; Niessen, S.; Hoover, H.; Foley, T. F.; Cravatt, B. F..; Burkart, M. D. *In preparation* (2009).

Global proteomic profiling of polyketide and nonribosomal peptide biosynthesis by orthogonal active site probes

Introduction

Polyketides and nonribosomal peptides constitute two classes of natural products which are well known for their therapeutic applications as antibiotic, antiparasitic, antifungal, and anticancer agents. Due to the significant role many of these compounds play in promoting human health and combating disease, the catalysts responsible for their biosynthesis, known as polyketide synthase (PKS), and nonribosomal peptide synthetase (NRPS) enzymes, have been the subject of intense biochemical, genetic, and structural characterization over the past three decades.[1,2] These studies have lead to revolutionary advances in natural products research, ranging from the engineering of PKS pathways in heterologous hosts, to the small-scale production of polyketide and nonribosomal peptide analogues through combinatorial biosynthetic approaches.[3,4] However, one area in which our knowledge and capabilities for studying these systems remain limited is at the proteomic level. Methods for the functional proteomic analyses of PKS and NRPS enzymes would advance our knowledge of natural product biosynthesis by providing insight into the dynamics and regulation of PKS and NRPS expression in the context of wild-type producer organisms. In unsequenced organisms such studies could facilitate discovery of natural product gene clusters on an LC-MS/MS time scale, while in sequenced natural product producers such technology would have applications in monitoring the expression of orphan gene clusters, improving fermentative

microorganisms, and analyzing pathogenic bacteria for the production of small molecule virulence factors. [5-7]

The many unique properties of PKS and NRPS enzymes present a technical challenge to traditional methods of proteomic analysis. For example, in type I PKS and NRPS systems, the multiple domains which catalyze each step of metabolite biosynthesis are found on a single massive polypeptide, presumably produced at tremendous metabolic cost to the cell.[1,8] Accordingly such proteins are high molecular weight (>100 kDa) and often low abundance, two properties which impede their analysis by gel-based methods such as two-dimensional polyacrylamide gel electrophoresis (2DE).[9] Similarly underrepresented in such analyses are the low molecular weight, highly acidic (pI ~ 3.0-4.0) carrier protein (CP) domains which are central to type II PKS biosynthesis.[10] Finally, the activity of these enzymes is dependent on posttranslational modification, requiring modification of their CP domains by a 4'-phosphopantetheinyltransferase (PPTase) to convert *apo* to *holo*-enzyme.[11] Thus, an ideal proteomic method for the global analysis of PKS and NRPS enzymes would incorporate both a functional enrichment method, allowing separation of *holo*-PKS and NRPS enzymes from their more abundant proteomic background, as well as a gel-free analysis platform, unbiased against the biochemical properties of these biosynthetic catalysts.

As a first step towards achieving the potential of natural product proteomics, here we investigate the utility of functional proteomic probes which directly target PKS and NRPS active sites for the analysis of modular biosynthetic enzymes from the model bacterium *Bacillus subtilis* (Figure 9.1). When applied in combination with the gel-free LC-MS/MS platform termed multidimensional protein identification technology

(MudPIT),[12] these methods allow for isolation and detection of enzymes encoded by all four known PKS and NRPS gene clusters found in this organism, greatly expanding the dynamic range of analysis for these enzymes. The active site probes applied are complementary, resulting in isolation of both distinct and common PKS and NRPS proteins, and can be used for the relative quantitation of natural product synthase expression between strains and identification of specific PKS and NRPS peptides in complex proteomic mixtures. By assessing and comparing the utility of each of these methods, this study will guide future efforts at the proteomic study and characterization of these important biosynthetic assembly lines.

Results

Design of a high-content proteomic platform for PKS/NRPS proteomics

PKS and NRPS systems produce their seemingly disparate chemical structures by a similar modular biosynthetic logic, in which multidomain enzymes activate, condense, and tailor a series of monomer units to produce the final natural product (Figure 9.1). The structural and functional homologies of the discrete enzymatic domains responsible for each biosynthetic step make it possible to establish a universal strategy for functional enrichment of both enzyme classes based on labeling of their commonly utilized active sites. In this initial study we have chosen to focus on the functional proteomic analysis of PKS and NRPS enzymes through their carrier protein (CP) and thioesterase (TE) domains. CPs are small enzymatic domains which serve to covalently tether biosynthetic

intermediates to the PKS or NRPS during biosynthesis.[10] The site of tethering is the terminal thiol of a CoA derived 4'-phosphopantetheine prosthetic group, whose posttranslational introduction by a PPTase converts *apo*-PKS and NRPS enzymes to *holo*-enzyme (Figure 9.1).[11,13] Substrate tethering through CP domains is a central feature of type I PKS and NRPS enzymes, with multiple CPs often found within a single polypeptide, making them attractive targets for proteomic enrichment. In contrast to the ubiquitous CP domain, TE domains are canonical serine hydrolases found in many PKS and NRPS termination modules which catalyze the hydrolysis or macrocyclization of CP-bound polyketides and nonribosomal peptides at the end of the biosynthetic cycle (Figure 9.1).[14] Due to their more specialized activity, proteomic enrichment of PKS and NRPS TE domains may be used to target a subset of PKS and NRPS enzymes, in particular those responsible for the synthesis of macrocycles and cyclic peptides.

Chemical probes capable of the isolation and functional analysis of each of these domains have been previously reported, and thus are only briefly introduced here (Figure 9.1).[15-18] CP domains can be labeled through metabolic incorporation of bioorthogonal CoA precursor **1**. This compound is added as a component of the growth media of cultured bacteria, whereupon it exhibits uptake and transfer to CP domains through the action of the organism's endogenous CoA biosynthetic pathway and PPTase enzymes (Supporting Information, Figure 9.S2).[15,19] Modified proteins can then be visualized or enriched through chemoselective ligation of the azide group of **1** to a conjugate reporter alkyne using a Cu(I)-catalyzed [3+2] cycloaddition reaction (Figure 9.1).[20] Alternately, biotinylated CoA analogue **2** can be used to label CP domains in proteomic samples by

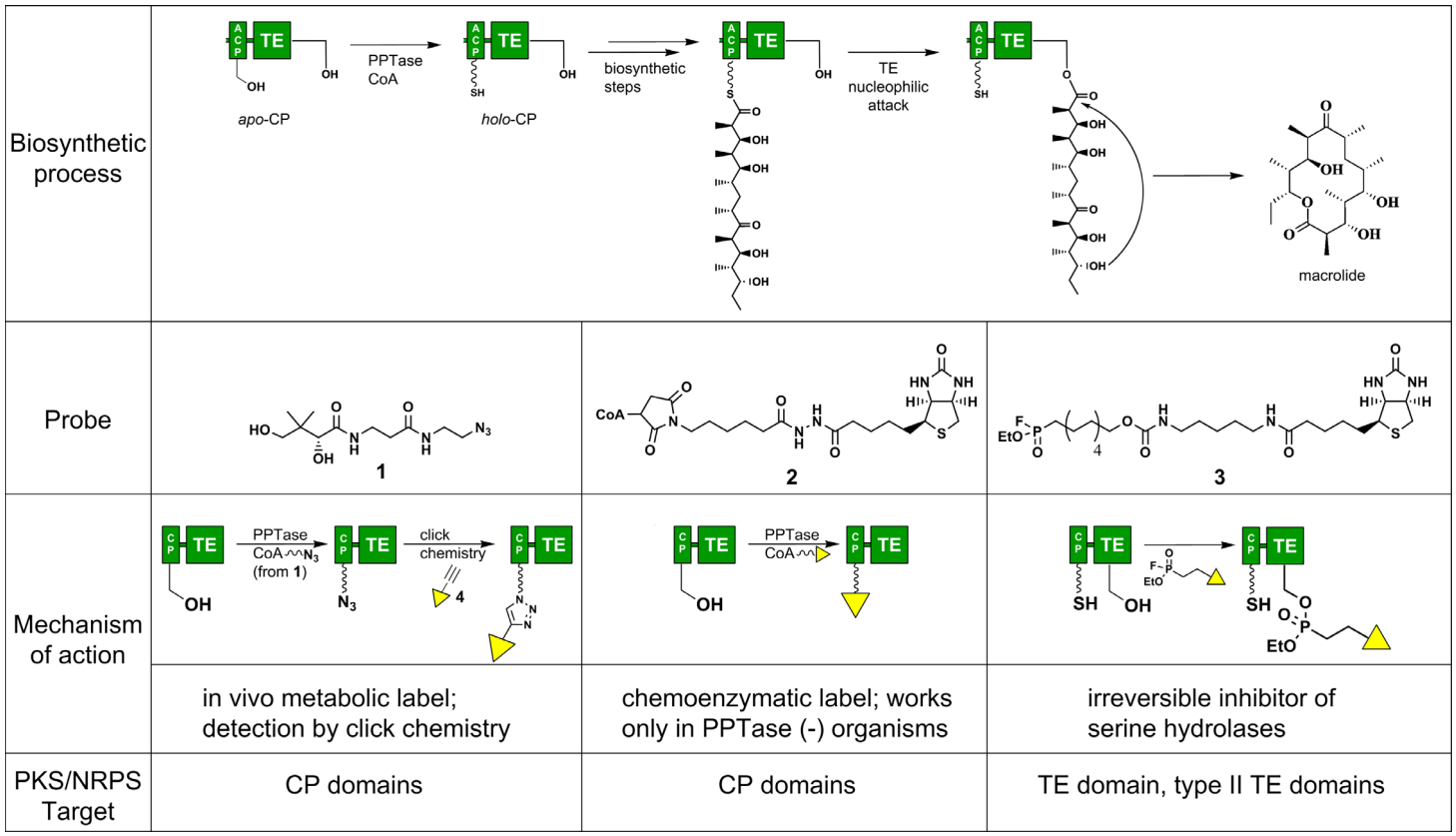


Figure 9.1 Functional proteomic probes for PKS and NRPS enzymes used in this study. The natural posttranslational modification and activity of CP and TE domains during PKS biosynthesis are depicted. These same processes can be exploited for functional enrichment and proteomic identification of PKS and NRPS biosynthetic enzymes using probes **1**-**4**. (Left) Bioorthogonal CoA precursor **1** is added as a component of the bacterial growth media and is metabolically incorporated into CP domains in vivo through the action of the endogenous CoA biosynthetic pathway and 4'-phosphopantetheinyl-transferase (PPTase) enzymes. The azide-moiety then allows for enrichment of labeled proteins via Cu(I)-catalyzed cycloaddition with biotin alkyne **2** followed by affinity chromatography. See Supplementary Figure 9.S2 for full mechanism. (b) Biotin-CoA **2** can be used for quantitative labeling of PKS and NRPS CP domains but only in PPTase deficient organisms such as *B. subtilis* strain 168. (c) Fluorophosphonate-biotin **3** labels TE domains through irreversible inhibition of the catalytic serine residue, and can be used directly for enrichment without further derivatization.

application in combination with exogenously added promiscuous PPTase enzyme Sfp.[16] Notably, this chemoenzymatic approach is most effective in proteomes isolated from organisms deficient in PPTase activity. Finally, TE domains can be functionally enriched using fluorophosphonate **3**, which modifies its catalytic serine residue through irreversible inhibition of the conserved α,β-hydrolase motif (Figure 9.1).[17,21] This activity based protein profiling (ABPP) probe is well known for its class wide reactivity with serine hydrolase enzymes, and thus enriches not only terminal PKS and NRPS modules, but also the larger pool of cellular serine hydrolases.[22,23]

In order to circumvent the limitations of gel-based strategies and expand the dynamic range of analysis for low abundance PKS and NRPS enzymes, we applied each of these probes in tandem with the gel-free protein identification platform known as MudPIT (Figure 9.2).[12,24] This involves a two dimensional (strong cation and reverse phase) LC separation of peptides produced by proteolytic digest of enzymes enriched by probes **1**-**3**, followed by identification by MS/MS analysis. The increased separation of this LC-MS/MS detection method facilitates both protein detection and relative quantitation between samples using spectral counting.[25] Importantly, serine hydrolase probe **3** has been previously incorporated into this gel-free proteomics platform for the analysis of human samples in an approach known as ABPP-MudPIT.[26] This is crucial, as it allows comparison of CP domain enrichment by probes **1** and **2** with a validated proteomic method.

Finally, as a subject for the comparative evaluation of each of these methods for the global analysis of PKS and NRPS production we chose the gram-positive bacterium

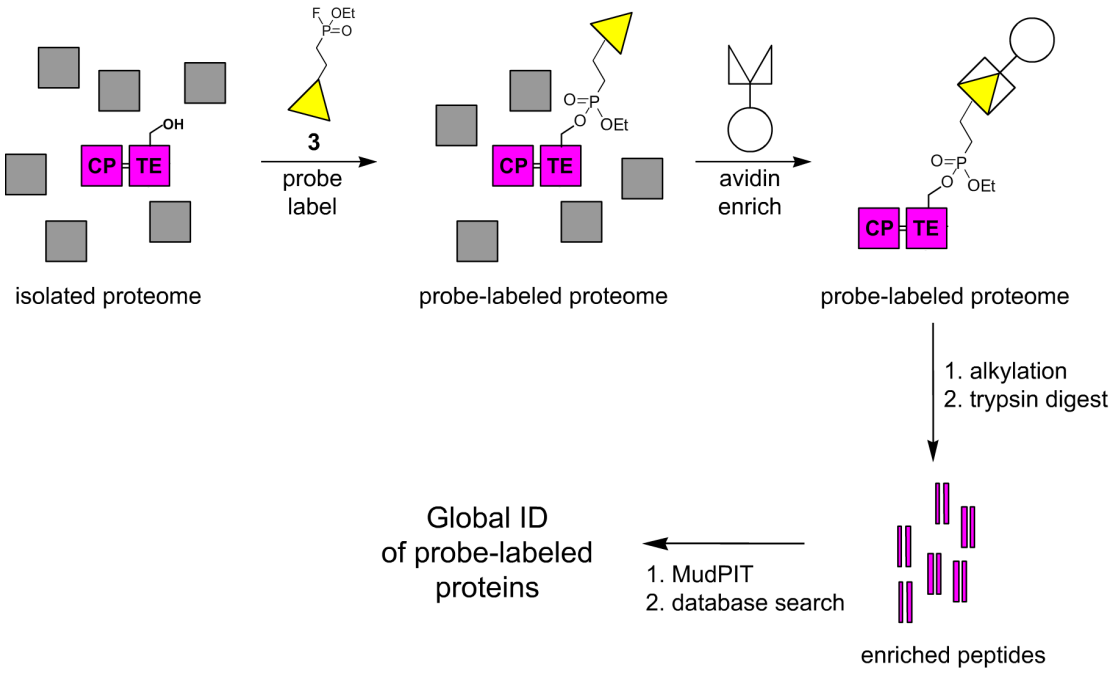


Figure 9.2 Schematic for functional proteomic enrichment and MudPIT analysis of PKS and NRPS peptides. Natural product producer proteomes are labeled with a biotin through metabolic labeling of CP domains followed by click chemistry (probes **1**, **4**), chemoenzymatic labeling of CP domains (probe **2**), or activity-based labeling by fluorophosphonate (probe **3**, depicted). Enrichment by avidin affinity chromatography followed by on-bead tryptic digest results in tryptic peptides of PKS and NRPS enriched samples. Proteins are then analyzed by multidimensional LC-MS (MudPIT) and the levels of enzyme activity estimated by spectral counting analysis.

Bacillus subtilis. This organism is well known for its production of the nonribosomal peptides surfactin, bacillibactin, and plipastatin, as well as the hybrid polyketide-nonribosomal peptide natural product bacillaene.[27,28] Notably, the terminal domains of the PKS and NRPS enzyme responsible for production of these compounds contain both CP and TE domains, allowing a direct comparison of PKS and NRPS labeling by orthogonal active site probes (Figure 9.3). We chose two strains of *B. subtilis* for profiling in this study, strain 6051 and 168.[29] Strain 6051 is a wild type strain known for its sporulation and formation of surface-associated biofilms, while 168 is a domesticated laboratory strain of the organism in which multicellularity has been lost.[30,31] The secondary biosynthetic capabilities of strain 168 also differ from wild type, due to its known incorporation of an in frame deletion in the gene encoding its secondary PPTase Sfp. This mutation halts antibiotic production by allowing for only the production of *apo*-PKS and NRPS enzymes, and in turn facilitates labeling of CP domains by probe **2**.[17,32] Finally, due to its role as the prototypical gram-positive endospore-forming bacteria, *B. subtilis* is one of the few natural product producers which has been well studied proteomically, providing a foundation for comparison of our functional proteomics approaches with previous studies of this organism.[33]

Active site labeling, enrichment, and MudPIT of *B. subtilis* proteomes

Due to our interest in natural product biosynthesis, we focused on examination of stationary phase proteomes isolated from *B. subtilis*, as earlier studies had observed PKS and NRPS secondary metabolism at this growth stage of growth.[8,34] The wild-type strain

6051 was cultured in the presence of CoA precursor **1** for metabolic labeling of CP domains, while strain 168 was grown under identical conditions without **1**. After initial verification of probe labeling by gel-based fluorescence profiling[17,35] (Supplementary Figure 9.S3) isolated proteomes were biotinylated by probes **2**, **3**, and **4** for enrichment. For CP-domain enrichment, metabolically labeled proteomes of strain 6051 were subjected to Cu(I)-catalyzed [3+2] cycloaddition reaction with biotin alkyne **4**, while proteomes of strain 168 were chemoenzymatically labeled with biotin CoA **2** using the PPTase enzyme Sfp.[36] For global enrichment of serine hydrolases, including TE domains, both proteomes were separately incubated with fluorophosphonate **3**. Each proteomic sample was then enriched on avidin-agarose, followed by tryptic digest and MudPIT analysis (Figure 9.2).[26] Multiple replicate analyses were performed for each enrichment method ($n \geq 4$), as well as for probe-free controls in which **2**, **3**, and **4** were not added to proteomes ($n \geq 2$). The number of tandem mass spectra assigned to each identified protein (spectral counts) was used as a measure of relative abundance,[25] and comparison of spectral counts of enzymes from enriched samples compared to probe-free controls was used to distinguish proteins specifically identified by CP and TE probes from endogenously biotinylated proteins. In addition, an *ad hoc* cutoff value of ≥ 15 spectral counts was employed to separate proteins showing probe-specific enrichment from background proteins.[26] This value is similar to those used in previous studies, and was chosen based on the observation that with these criteria, >85% of fluorophosphonate **3** enriched from *B. subtilis* showed annotation or strong homology to serine hydrolases. Replicate MudPIT data sets were averaged, and each method was assessed for number of PKS/NRPS proteins identified, total number of proteins observed,

Table 9.1 Summary of proteins specifically enriched by probes **1**, **2**, and **3** in *B. subtilis*. Specific enrichment is based on the abundance of each protein in enriched compared to non-enriched samples, as indicated by spectral count analysis (total spectrum counts ≥ 15, ratio of enriched: probe-free control ≥ 10). S.E.M. = standard error of mean. [a]Abbreviations denotes PKS/NRPS gene cluster of origin for identified proteins. Parentheses indicate total number of CP and TE containing open reading frames for each PKS and NRPS locus.

Enrichment Probe	*B. subtlis* Strain	PKS/NRPS Proteins Identified					total proteins identified	number SEM ≥ 30%
		srf (4)[a]	*dhb* (2)	*pps* (5)	*pks* (6)	total (16)		
1	6051	2	2	3	5	12	58	1
2	168	3	2	4	2	11	45	15
3	6051	1	1	1	1	4	29	0
3	168	3	1	1	0	5	36	0

and reproducibility. The results of these analyses are summarized in Table 9.1 (for individual protein identification data see Supporting Information, Tables 9.S1-S6).

Comparative analysis of CP and TE probes for proteomic identification of PKS and NRPS enzymes

We first examined each method for global identification of PKS and NRPS enzymes. Metabolic CP-probe **1** proved most effective in this regard, identifying 12 of the possible 16 PKS and NRPS CP-containing enzymes found in the *B. subtilis* genome, including hits from all four PKS and NRPS gene clusters (Table 9.1). This is especially remarkable given the mode of action of **1**, which requires it to cross the cell membrane and compete with endogenous CoA for labeling of CP-domains in vivo. Chemoenzymatic CP-probe **2** proved similarly effective, enriching a distinct but overlapping set of modular biosynthetic enzymes. However, individual inspection of the proteins identified by each method reveals a common underrepresentation of the discrete, low molecular weight (~9 kDa) CP domains of bacterial FAS and type II PKS biosynthesis. This is likely due to the low number of peptides suitable for LC-MS/MS sampling which these CPs produce upon tryptic digest, effectively biasing the analysis method (spectral counting) against them. Indeed, our spectral counting criteria identified only a single type II CP, DltC (involved in cell wall biosynthesis), enriched only by metabolic probe **1** (Table 9.S1).[37] For these reasons, MudPIT analyses using CP-enrichment will likely be most useful for the functional proteomic study of type I PKS and NRPS systems.

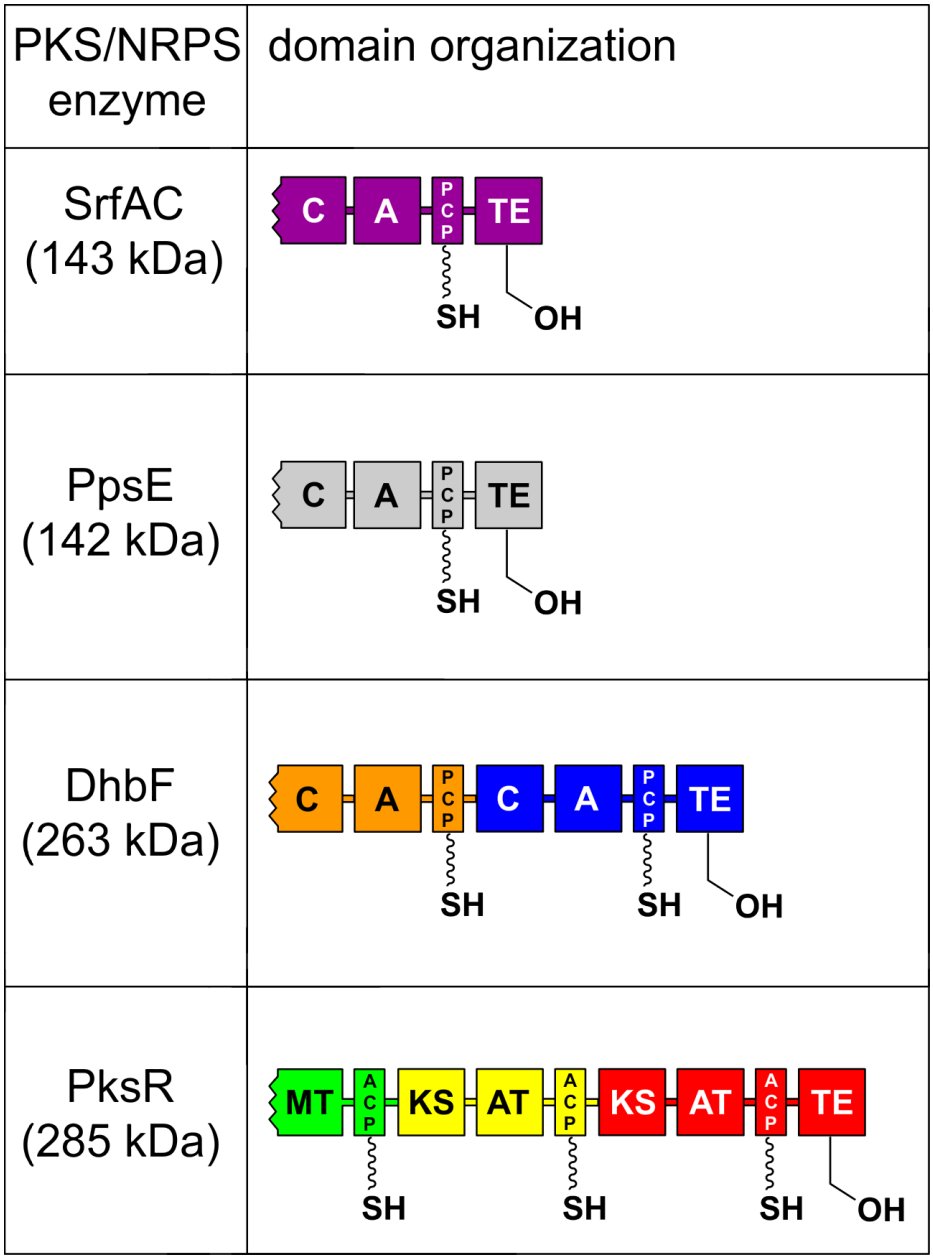


Figure 9.3 Domain organization for terminal modules of type I modular biosynthetic enzymes in *B. subtilis*. The depicted enzymes encode surfactin (SrfAD - NRPS), plipastatin (PpsE - NRPS), and bacillibactin (DhbF - NRPS), and bacillaene (PksR – hybrid PKS/NRPS). Each CP-containing module (responsible for a biosynthetic step) is color coded for clarity. Abbreviations: C, condensation domain; A, adenylation domain; PCP, peptidyl carrier protein; ACP, acyl carrier protein; TE, thioesterase; MT, methyltransferase; KS, ketosynthase; AT; acyltransferase.

Hydrolase probe **3** also proved well-suited to the analysis of modular biosynthetic enzymes, identifying the TE-containing termination modules of 3 out of 4 PKS and NRPS enzymes in each *B. subtilis* strain, as well as SrfAD, a type II TE domain which is involved in editing of misprimed CP domains (Table 9.1).[38] In strain 168 a non-TE containing NRPS, SrfAA, also showed specific enrichment by fluorophosphonate **3**, but at greatly decreased spectral counts compared to CP-probe **2**. This disparity suggests SrfAA is not a covalent target of **3**, but rather is enriched due to non-covalent interaction with a partner enzyme biotinylated by **3**, possibly SrfAC. This is remarkable, given the harsh denaturation and washing steps involved in our MudPIT enrichment protocol. Previous studies have use co-immunoprecipitation of modular biosynthetic enzymes in *B. subtilis* as evidence for the macromolecular organization of these enzymes.[8] With further refinement, the functional proteomic probes may be able to provide similar insights into the affinity partners of PKS and NRPS enzymes in native organisms, an additional beneficial aspect of proteomic studies of natural product producers.

Comparative analysis of active site labeling by metabolic CP and activity-based TE probes

One potential caveat to the use of CoA precursor **1** for visualization of PKS and NRPS enzymes is that it must compete with endogenous intracellular CoA for labeling of PKS and NRPS enzymes (Figure 9.S2).[19] Furthermore, the non-toxic nature of metabolic probe **1** implies that it labels only a percentage of total cellular CP-domains, as inhibition of fatty acid biosynthesis (which would result from quantitative loading of fatty acid CP

domains by probe **1**) is a validated antibiotic target. In contrast, activity-based probes such as fluorophosphonate **3** are applied post-lysis and are expected to label their cognate targets stochiometrically, provided the proteomic sample retains activity.[22] In order to better gauge the relative labeling efficiencies of these two methods, we further examined the data obtained from the MudPIT analyses of **1** and **3**-enriched enzymes from *B. subtilis* 6051. Previous studies have suggested spectral count based quantitation methods are most well-suited to analysis of higher abundance proteins, so we focused on the PKS termination module PksR (Figure 9.3) due to its position as the highest abundance PKS/NRPS target of probe **3** in *B. subtilis* 6051.[39,40]

First, we compared the relative spectral counts observed for PksR upon enrichment of *B. subtilis* 6051 lysate with metabolic CP probe **1** (through click chemistry ligation – Figure 9.1) and activity-based probe **3**. This analysis indicates approximately equal enrichment by each method (Figure 9.4a); however, under closer scrutiny it is observed that much greater sequence coverage of PksR is observed after enrichment by CP-probe **1** than with TE-probe **3** (Figure 9.4b). Sequence coverage arising from enrichment with TE-probe **3** is concentrated around the TE active site (Figure 9.4b, blue sequences), an indication of proteolytic cleavage of some PksR enzymes in the sample. This is not entirely surprising, given that the labeling mechanisms of click chemistry probe **1** and hydrolase probe **3** prohibit the treatment of these proteomic samples with common protease inhibitors such as EDTA or PMSF. Our gel-based fluorescence analysis of probe-labeling also was suggestive of PKS/NRPS cleavage (Supporting Information, Supplementary Figure 9.S3).[17] Correcting the spectral count data for sequence coverage (corrected = [spectral counts]/[sequence coverage]) leads to the

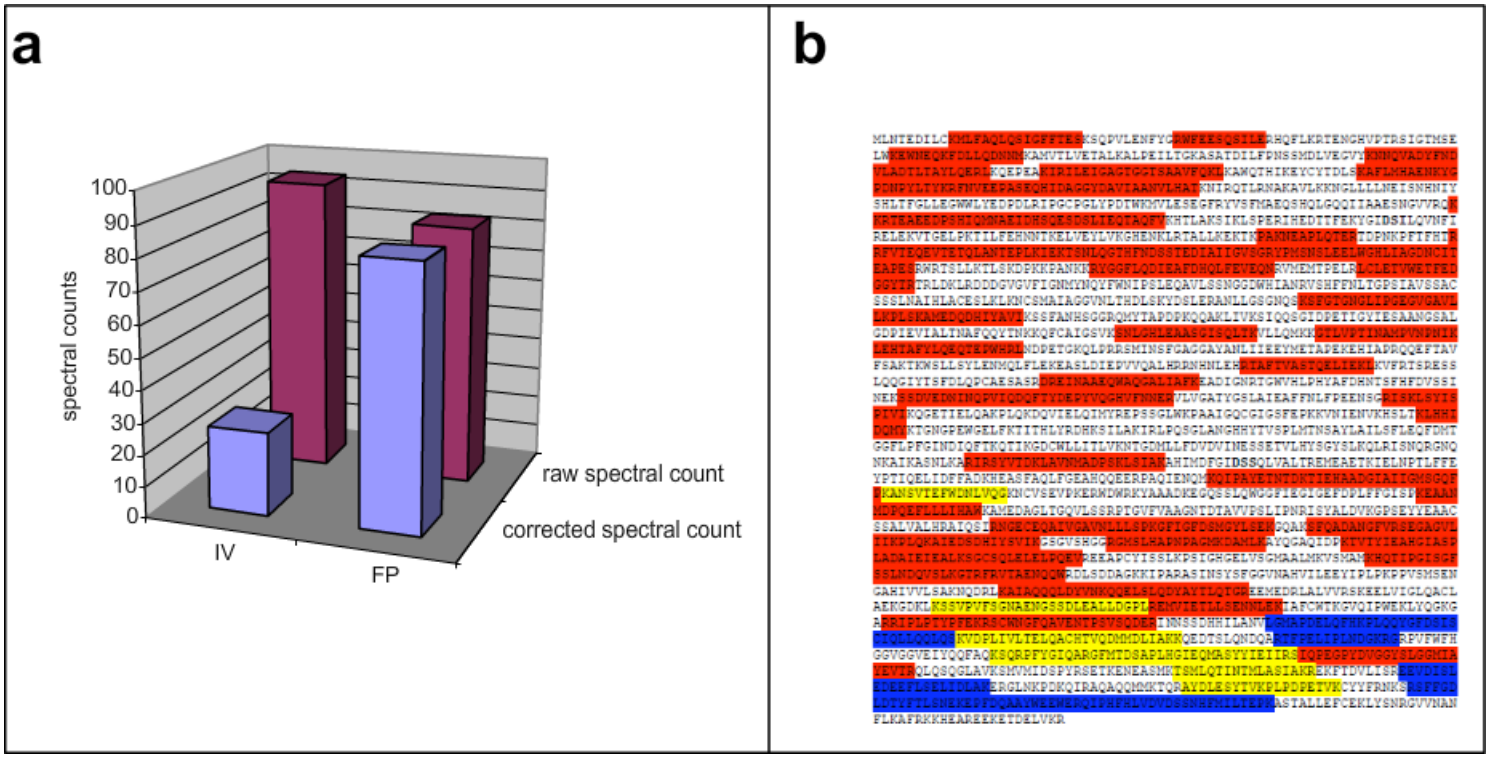


Figure 9.4 Analysis of relative efficiencies of functional proteomics probes **1** and **3** for labeling PksR in *B. subtilis* 6051. (a) Relative enrichment of PksR (terminal module of bacillaene synthase) by active site probes **1** and **3** based on spectral counting. CP-probe **1** labels PksR active sites less efficiently than TE-probe **3** (corrected spectral counts) but is equally competent in PKS identification (raw spectral counts) due to the greater number of CP-active sites per PKS/NRPS enzyme. (b) Sequence coverage for PksR (terminal module of bacillaene synthase) generated by enrichment of *B. subtilis* 6051 using in vivo probe **1** (red) and activity-based probe **3**. Red = sequences enriched by CP-probe, Blue = sequences enriched by TE-probe, Yellow = sequences enriched by both probes.

estimate that in vivo CP-labeling is about 25-30% as effective as TE-labeling for quantitative labeling of PksR active sites. However, the presence of three CP-active sites on PksR compensate for this overall labeling efficiency, resulting in the approximately equivalent spectral counts observed. Thus, despite its lower overall efficiency, the abundance of CP-domains in type I PKS and NRPS biosynthetic enzymes (> 1 CP domain per module) make metabolic labeling by **1** a comparably efficient method for enrichment and identification of many PKS and NRPS enzymes.

Comparative analysis of specificity and reproducibility of CP and TE probes for functional enrichment PKS and NRPS enzymes

In addition to enrichment of low abundance enzymes, one major advantage of metabolic and activity-based probes **1**-**3** in proteomic analyses is their ability to provide a functional readout of posttranslational modification and hydrolytic activity in PKS and NRPS biosynthetic pathways.[41] Their utility in functional proteomics is therefore related to their ability to enrich targets based on a designated mode of action, in the cases of probes **1**-**2** reaction with *apo*-CP domains, in the case of probe **3** reaction with a nucleophilic serine residue. In order to assess the relative functional proteomics capabilities of each labeling method we inspected the targets of probes **1**-**3** to provide insight into the proteomic selectivity of each. ABPP probe **3** enriched the smallest total number of proteins, followed by chemoenzymatic probe **2**, and metabolic label **1** (Table 9.1). The majority of enzymes labeled by **3** in both samples are annotated or homologous to serine hydrolases, consistent with its mode of action.[21,22] Analysis of sequence

coverage for the PKS and NRPS enzymes enriched by both CP and TE probes showed a lack of coverage for peptides corresponding to the known CP and TE active sites of these enzymes, a finding also in accord with the specific labeling of these proteins through their enzymatic activities. However, only ~20-25 % of the total proteins identified by CP-probes **1** and **2** are annotated as CP-containing enzymes, indicating that these probes have substantial off-target reactivity. This is not unexpected, as the Cu(I)-catalyzed chemoselective ligation reaction used for detection of enzymes labeled by metabolic probe **1** is well-known to show some nonselective labeling of proteomic samples,[20] while excess biotin-CoA **2** could result in enrichment of non CP-proteins which strongly bind the CoA or adenylate moiety.[42] These non-CP protein hits also may represent authentic targets of the 4'-phosphopantetheine posttranslational modification which are not currently annotated as such, a phenomena previously observed in *B. subtilis*.[43,44]

Finally, because probe targets were identified based on averaging of spectral count data generated from multiple data sets, we also performed a statistical analysis to assess the reproducibility of protein identifications between each individual MudPIT run for each enrichment method. Fluorophosphonate **3** showed reproducible spectral count measurements of a consistent pool of enzymes with a standard error of the means (S.E.M.) below 30%, consistent with previous reports (Table 9.1).[26] Metabolic CP-probe **1** showed similarly high reproducibility. In contrast, chemoenzymatic CP-probe **2** showed variability in the identification of one third (15/45) of its identified targets, possibly due to loss of activity of recombinant Sfp between sample preparations. This indicates probes **1** and **3** may be better suited for spectral counting-based quantitation of PKS and NRPS enzymes in proteomic samples.

Spectral count-based quantitation of PKS and NRPS enzymes in Bacillus subtilis

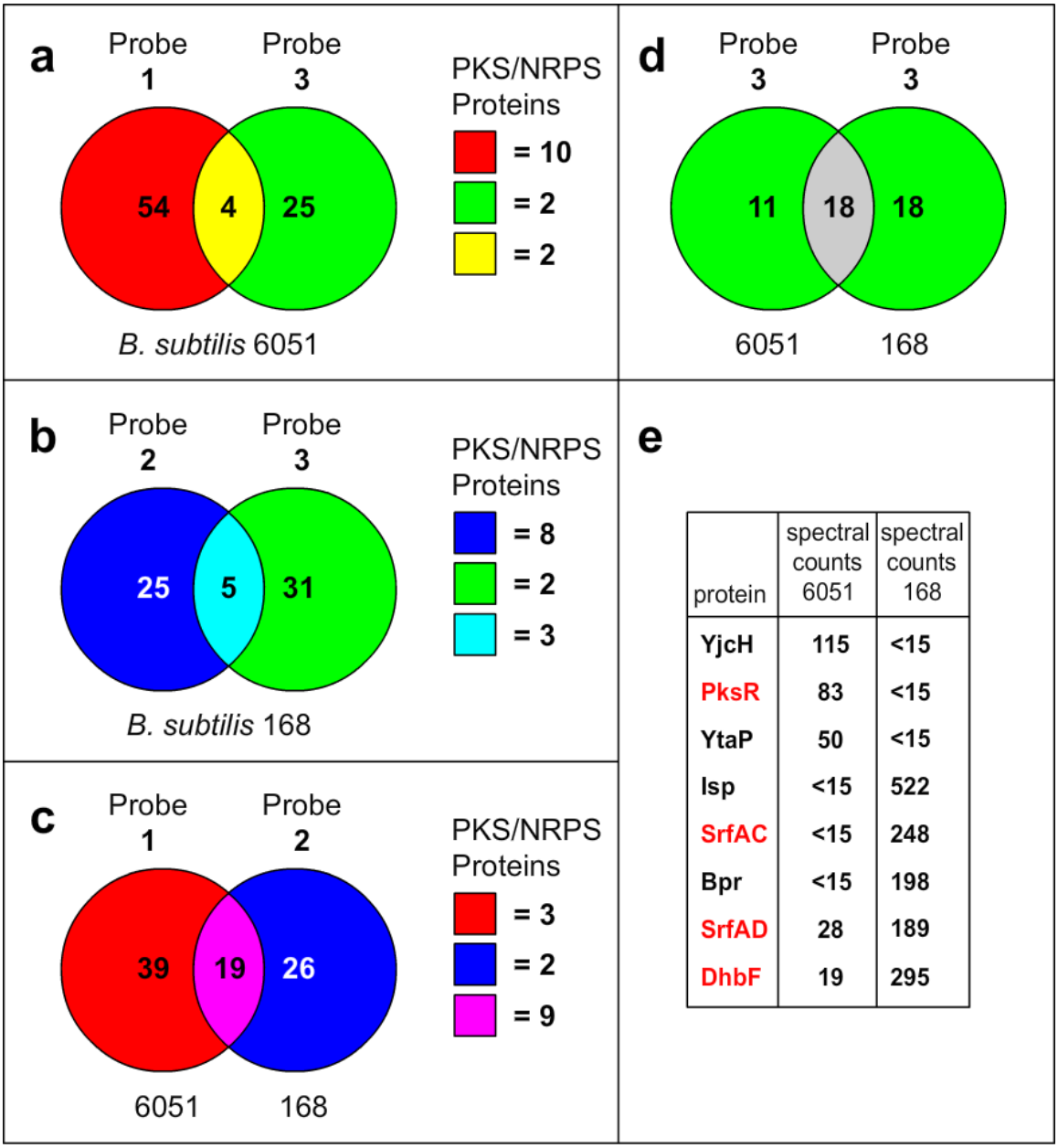
Next we applied these methods to analyze the relative expression of PKS and NRPS enzymes in *B. subtilis* strains 6051 and 168. Because CP probes **1** and **2** label their cognate PKS and NRPS enzymes by different mechanisms in these two strains, we focused on comparison of data generated by ABPP-MudPIT analysis using fluorophosphonate **3**.[26] As can be seen in Figures 9.5e and 9.6a, the terminal NRPS enzymes SrfAC and DhbF show strong upregulation in strain 168, demonstrating enrichment at levels 5-10 times observed in strain 6051 as estimated by spectral count analysis. This finding is consistent with the higher levels of SrfAC observed in strain 168 during our previous gel-based proteomic studies of this organism.[17] The overall upregulation of the Srf operon is also evidenced by the greater spectral counts observed for the type II TE domain SrfAD in strain 168. PksR, the PKS module responsible for chain termination in the biosynthesis of bacillaene, demonstrated the opposite relationship, with higher levels found in strain 6051 (Figure 9.5e).

In addition to providing a readout of PKS and NRPS enzymatic activity in proteomic samples, enrichment by fluorophosphonate **3** also provides insight into the functional activity of ancillary hydrolases in this organism. In particular, strain 168 showed higher expression levels of several serine hydrolases, including Isp and Bpr, which were not functionally enriched in strain 6051 despite growth under equivalent conditions (Figure 9.5e).[45-47] Bpr is known to be ribosomally produced as a zymogen, and its enrichment by fluorophosphonate **3** provides strong evidence that this enzyme has been proteolytically processed to its *holo*-form in late stationary phase cultures of strain

168.[48] In contrast strain 6051 showed identification of an overlapping but distinct pool of hydrolases after enrichment by **3**, including several unannotated hydrolases (YjcH, YtaP) which were enriched at high abundance.[49] In future studies these secondary hydrolase identifications may be helpful in delineating the regulatory pathways of secondary metabolism, or alternately may be used as biomarkers associated with production of specific natural products.

Assessing the orthogonal nature of CP and TE active site probes for detection of PKS and NRPS enzymes

One of the most unique properties of PKS and NRPS modular biosynthetic enzymes is their utilization of multiple active sites within the context of a single polypeptide. Potentially, this property could be used to facilitate enrichment of specific subsets of PKS and NRPS enzymes (i.e. those containing *both* a CP and TE domains) from complex proteomic mixtures based on the application of orthogonal active site probes (i.e. **1** and **3**) followed by purification using orthogonal affinity purification methods. As a first step towards developing such a method, we assessed the orthogonal active site labeling properties of CP (**1**-**2**) and TE (**3**) probes by comparing the proteins identified from *B. subtilis* proteomic samples using each (Figure 9.5a and 9.5b). In *B. subtilis* strain 6051, four proteins (PksR, DhbF, HutI, and GatA) show enrichment by both probes **1** and **3** under the criteria applied. Probes **2** and **3** show similar orthogonal enrichment properties in strain 168, with five proteins (DhbF, SrfAA, SrfAC, AhbF, and Tuf), also being identified. Compared to the total pool of enzymes enriched by any single



protein	spectral counts 6051	spectral counts 168
YjcH	115	<15
PksR	83	<15
YtaP	50	<15
Isp	<15	522
SrfAC	<15	248
Bpr	<15	198
SrfAD	28	189
DhbF	19	295

Figure 9.5 Comparison of enzyme identifications and functional proteomics relationships revealed by by **1**-**3**. (a-d) Venn diagrams comparing protein identifications made by CP-probes **1**-**2** and TE-probe **3**. Areas of overlap indicate proteins identified by both analyses, non-overlapping areas indicate distinct identifications. (e) Truncated list of PKS/NRPS and non-PKS/NRPS proteases showing differential enrichment between strains 6051 and 168 using probe **3**. For full list of identified proteins by each method see Supplementary Tables 9.S1-6.

probe, these doubly identified enzyme pools are extremely rich in PKS and NRPS proteins, particularly those containing both CP and TE domains (Figure 9.5a-b). This observation highlights the unique susceptibility of modular biosynthetic enzymes to such bifunctional enrichment methods.

We further explored this phenomenon on the peptide level. Identification of PKS and NRPS peptides may be used to facilitate cloning of natural product biosynthetic gene clusters, an approach we have previously demonstrated in studies focusing on labeling of fatty acid biosynthetic enzymes by **1**.[15] Here looked at individual rather than averaged MudPIT datasets, and analyzed the number of unique peptides which were identified in every dataset generated by MudPIT of the *B. subtilis* proteome following enrichment by probes **1** and **3**, but were identified in none of the datasets generated by MudPIT of the *B. subtilis* proteome following enrichment using an identical protocol but with no probe applied.

We first examined the impact of both the enrichment methods used (i.e. CP or TE targeting) as well as the number of data sets collected/contrasted on the total number of unique peptides identified. As can be seen in Table 9.2, collection of an increasing number of data sets reduces the size of the pool (total peptides, Table 9.2). As the peptide pool shrinks, the overall percentage represented by PKS and NRPS (% PKS/NRPS) increases, indicating reproducible isolation of a subset of PKS and NRPS peptides. Examining the proteins to which these peptides are assigned to indicates their enrichment accurately reflect the mechanism of the functional proteomic probes (**1** and **3**) utilized, as every PKS and NRPS peptide identified in the final two table entries originates from a

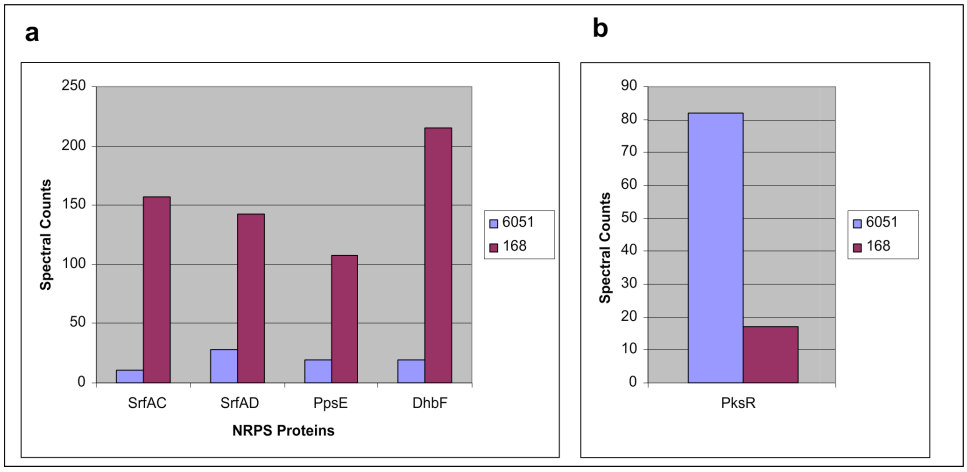


Figure 9.6 Comparison of NRPS levels in *B. subtilis* 168 and *B. subtilis* 6051 using hydrolase (TE) probe **3**. (a) Comparison of NRPS levels in *B. subtilis* 168 and *B. subtilis* 6051 using TE-probe **3**. All three NRPS termination modules of *B. subtilis* as well as the type II thioesterase encoded in the *srf* gene cluster (SrfAD) are highly upregulated in strain 168. (b) Comparison of PksR levels in *B. subtilis* 168 and *B. subtilis* 6051 using TE-probe **3**. Wild-type organism 6051 shows higher levels of PksR than does 168 (<15 spectral counts). For complete tabulated data see Supplementary Tables 9.S3 and 9.S4.

Table 9.2 Analysis of the orthogonal enrichment properties of active site probes **1** and **3** on a peptide level. Collection of a greater number of MudPIT data sets using each enrichment method (**1** or **3**) decreases the total number of peptides (total peptides) uniquely identified in probe-labeled samples. The use of orthogonal active site probes biases the identified peptides towards PKS/NRPS enzymes (% PKS/NRPS increasing), particularly those containing CP-TE domains (% CP-TE). A list of peptides identified from the bottom row (consisting of 40 unique peptide identifications observed in all 9 probe **1**/**3** enriched samples but not in no probe datasets) is provided in Table 9.S7 (Supporting Information).

# MudPIT Data Sets Collected			peptides enriched		
Probe 1 (CP)	Probe **3** (TE)	no probe	total peptides	% PKS/NRPS (total)	% CP-TE (total)
1	1	1	599	7.5% (45)	51% (23)
2	2	2	216	8.3% (18)	72% (13)
3	3	3	66	13.6% (9)	88% (8)
4	4	4	47	17% (8)	100% (8)
5	4	5	40	20% (8)	100% (8)

CP- and TE-containing terminal PKS or NRPS module. Unique peptides are observed from the terminal domains SrfAC, DhbF, PpsE, and PksR, which together make up every PKS and NRPS coded for in the *B. subtilis* genome.[49] This is especially remarkable given the previously observed low abundance of the SrfAC and PpsE enzymes in *B. subtilis* 6051 samples, as detailed above. Analysis of each of the 40 unique peptides observed in the final dataset (Table 9.2, bottom row; a full list of peptides is given in Supplementary Table 9.S7) using the error tolerant Basic Local Alignment Search Tool MS-BLAST showed PKS/NRPS homology only for annotated peptides from PKS and NRPS gene clusters among the top ten hits.[50] While this is an extremely crude analysis given the presence of several other species of Bacilli in the searched database, this does suggest that homology based approaches may be useful in distinguishing PKS and NRPS peptides enriched by **1** and **3** from their non-specific background targets. An additional obstacle that must be addressed before this approach can be extended to unsequenced producer organisms of interest is that this preliminary study made mass spectral assignments based on database searching,[51] while in unsequenced organisms no such database exists. For this reason we are currently in the process of assessing the utility of making peptide identifications using alternate assignment methods such as error tolerant database searching[52] and de novo sequencing[53] to perform a similar analysis. Here our *B. subtilis* analyses provide valuable datasets to further train and assess each of these bioinformatics methods. These studies are currently ongoing.

Discussion

Almost 40 years ago, the first investigations into the biochemistry of polyketide and nonribosomal peptide biosynthesis were performed utilizing PKS and NRPS enzymes purified from natural product producers such as *Penicillium patulum* and *Brevibacillus brevis*.[54,55] These studies were essential in building the foundational hypotheses of what we now know as the modular biosynthetic paradigm, but were hindered by the laborious nature of traditional protein fractionation methods, often resulting in isolation of PKS and NRPS enzymes of varied activity. Since the advent of modern molecular biology, PKS and NRPS enzymes have been the subject of detailed enzymological study in vitro using recombinantally expressed enzyme preparations.[1,56] However, while such studies have greatly enhanced our understanding of the mechanisms, posttranslational modification, and substrate specificity of these enzymes, there remains a fundamental gap between our detailed understanding of these organisms in vitro and our less developed knowledge of their regulation, interaction, and activity in native natural product producers. Targeted proteomic study of natural product biosynthetic processes could help fill this gap, but require the development of new methods for their analysis, the subject of this report.

The proteomic study of PKS and NRPS biosynthetic enzymes presents several technical challenges, even in a well-characterized organism such as *B. subtilis*. These enzymes incorporate multiple posttranslational modifications and utilize varied enzymatic architectures with biochemical properties ranging the extremes of molecular weight and pI. Furthermore, while virtually no data exists investigating the relative levels of PKS and NRPS enzymes in natural product producers, their natural abundance is expected to be low. A survey of previous two-dimensional gel electrophoresis (2DE)

based proteomic studies of the *B. subtilis* cytosolic fraction finds only a single NRPS CP domain identification, the natural carrier protein-isochorismate lyase fusion protein DhbB.[57] Not surprisingly, this enzyme is of intermediate size (~40 kDa) relative to most PKS and NRPS carrier proteins, a property which may have facilitated its identification.

To circumvent these obstacles and guide future efforts at the proteomic analysis of natural product producing organisms, we have performed a full study of the use of recently reported PKS and NRPS active site probes in combination with the 2D-LC-MS/MS platform MudPIT for the identification of modular biosynthetic enzymes in *B. subtilis*. PKS/NRPS active site probes **1** and **2** label CP domains through their endogenous mechanism of posttranslational modification, while fluorophosphonate **3** is an activity-based probe which provides global enrichment of serine hydrolases including TE domains (Figure 9.1). In this initial study we sought to answer a variety of questions regarding the sensitivity, efficiency, specificity, and reproducibility of these techniques in the study of PKS and NRPS proteins. Metabolic CP-probe **1** provided the best overall coverage of PKS and NRPS enzymes of any probe used, facilitating identification of multiple proteins from all four PKS/NRPS gene clusters (*srf*, *dhb*, *pps*, *pks*) in this organism (Table 9.1). The increase in PKS/NRPS dynamic range provided by enrichment of CP domains using probe **1** prior to MudPIT analysis is emphasized by comparison to a previous gel-free study, which used MudPIT to analyze the unfractionated cytosolic proteome of *B. subtilis* 168.[58] This previous study resulted in identification of only the Srf biosynthetic enzymes, despite the use of laboratory strain 168, which we have shown here to produce considerably higher levels of NRPS enzymes than wild type *B. subtilis* 6051. Promisingly, we have previously demonstrated the utility of **1** in the labeling of CP

domains of a number of bacterial organisms, including *Escherichia coli*, *Schewanella oneidensis*, and *Brevibacillus brevis*,[15] and a survey of genomes shows that a number of organisms, including several species of marine and terrestrial actinomycetes, contain CoA biosynthetic pathways compatible with metabolic labeling by **1**. Future testing in our laboratory will aim to further outline the diversity of organisms and pathways which may be studied using this metabolic labeling strategy.

Chemoenzymatic CP label **2** showed similar coverage of PKS and NRPS enzymes with slightly decreased background labeling (Table 9.1). However, a subset of enriched enzymes showed high inter-dataset variation, a characteristic not observed with probes **1** and **3** (Supporting Information, Table 9.S2). Another limitation to this method is also of narrower utility than metabolic label **1**, as previous studies have shown that chemoenzymatic labeling by biotin-CoA and Sfp is effectively blocked by endogenous PPTase activity.[16] Indeed, when applying this method to wild-type *B. subtilis* 6051 proteomes we saw no evidence of CP-labeling above background. Therefore, enrichment of PKS and NRPS enzymes using chemoenzymatic label **2** may be most effective in combination with PPTase inhibitors or in mutant PPTase knockout organisms.

Both CP probes **1** and **2** showed enrichment of many *B. subtilis* enzymes not currently annotated as polyketide, nonribosomal peptide, or fatty acid biosynthetic enzymes (Table 9.5c). While even the most reliable of chemical proteomic labeling methods are well-known to have instances of off-site reactivity (i.e. labeling of the human proteasome by **3**), an alternate explanation for these enzyme identifications is their representation of non-canonical, 4'-phosphopantetheinylated targets of PPTase activity.[59] This phenomenon was previously observed in *B. subtilis* during genomic

profiling of PKS and NRPS enzymes by phage display.[43] Because such proteins would be expected to be identified by both probes **1** and **2**, we compared non-PKS and NRPS proteins identified in these analyses, and found ten such enzymes (Supporting Information, Table 9.S5 a). These enzymes are excellent candidates for expression and future biochemical study, which will be necessary to validate these enzymes as true targets of the PPTase posttranslational modification. More important form the perspective of this study is the ability of such a functional proteomics approach to generate new hypotheses even in a well-studied organism such as *B. subtilis*.

The utility of activity-based probe **3** in the proteomic study of eukaryotic systems has been well-demonstrated in the past, and it proved a similarly privileged scaffold for analysis of natural product biosynthetic enzymes. Enrichment by **3** allowed for the expression-level dependent identification of the terminal TE-containing modules of all four type I PKS/NRPS in *B. subtilis*. Also identified was SrfAD, the discrete type II TE involved in editing of misprimed CP domains during PKS/NRPS biosynthesis, and YtpA, a hydrolase involved in biosynthesis of the phospholipid antibiotic bacilysocin (Tables 9.S3-4).[60] Taken together these data indicate the principal role hydrolases play in natural product biosynthesis, and the utility of profiling their varied activities with activity-based probes such as **3**.

In addition to comparing the sensitivity, specificity, and reproducibility of each labeling method, we also used activity-based probe **3** to report on the relative PKS and NRPS activity between strains, comparing natural product biosynthetic enzyme production in strains 6051 and 168 of *B. subtilis*. Using this approach we observed strong upregulation of the surfactin and bacillibactin pathways in strain 168, while

undomesticated strain 6051 produced higher levels of the biosynthetic enzymes responsible for production of the hybrid PKS-NRPS natural product bacillaene (Figure 9.6). The high levels of SrfAC observed in *B. subtilis* 168 are consistent with the known properties of this strain, which is highly competent relative to wild type strains such as 6051 (one reason for its routine use in molecular biology applications).[61] Competence, in turn, is under control of the transcription factor ComA, which is known to drive expression of the *srf* operon. Transcription of the *dhb* operon is iron regulated, and the lack of a functional PPTase required for biosynthesis of the siderophore bacillibactin is likely responsible for slight iron deficiency, leading to the observed upregulation in *B. subtilis* 168. In contrast, little is known about regulation of bacillaene production, and further study will likely be necessary to decipher the physiological basis for the higher levels of this enzyme observed in strain 6051.

Additional content provided by proteomic profiling of *B. subtilis* using activity-based probe **3** was the identification of non-PKS/NRPS hydrolases differentially expressed in the two strains. Interestingly, Isp and Bpr, two proteases which showed strong expression in strain 168, Isp and Bpr, are under control of the DegU regulon, which has been found to antagonize ComA-mediated transcription of the *srf* gene cluster under conditions of high protease production.[62,63]. Two uncharacterized hydrolases, YjcH and YtaP, were identified at high levels in strain 6051 but not in 168 cells grown under equivalent conditions (Figure 9.5e). While in this initial study we have focused mainly on method development rather than the physiological significance of individual protein identifications, the continued generation of such differential hydrolytic activity profiles have potential to provide proteomic biomarkers for the pathway-specific production of

natural products, and thus could be useful in the monitoring and facilitating the functional improvement of fermentative microorganisms.

Finally we assessed the orthogonal enrichment properties of CP and TE probes by comparing the protein and peptide identifications made by each. Analysis at the protein level provided strong evidence for the distinct mechanisms of action of these two probes, with very few proteins demonstrating enrichment by both probe sets (overlapping area, Figure 9.5a and 9.5b). Similar qualities were observed by contrasting peptide identifications made after enrichment with both probes **1** and **3** with those enriched in corresponding no probe controls. This subtractive approach facilitated identification of a small pool of peptides which showed consistent enrichment by both the CP probe **1** and TE probe **3**, and in combination with homology-based analysis allowed discrimination of peptides originating from multifunctional CP-TE containing PKS and NRPS enzymes from the background targets of each individual probe. A similar approach may be useful in accelerating the identification of natural product gene clusters in unsequenced organisms, as knowledge of PKS and NRPS peptide sequences could be applied to the design of degenerate PCR-primers in a reverse-genetics approach to gene discovery. Further studies addressing the specific bioinformatic and technical challenges involved in this approach are currently underway.

Conclusion

In this work we have performed the first detailed study of functional proteomic methods for the global analysis of bacterial PKS and NRPS biosynthetic enzymes, and

demonstrated the analytical power of the MudPIT platform when applied in combination with PKS and NRPS directed active site probes. The CP and TE-directed labeling strategies examined can be used for the highly sensitive detection and analysis of PKS and NRPS enzymes from wild-type natural product producer organisms, and each has corresponding advantages and limitations which have been outlined here. These methods are complementary to existing biochemical, genetic, and structural methods of analysis, and thus stand to complement and expand our knowledge of these molecular assembly lines.

Supporting Information

Supplementary Figures

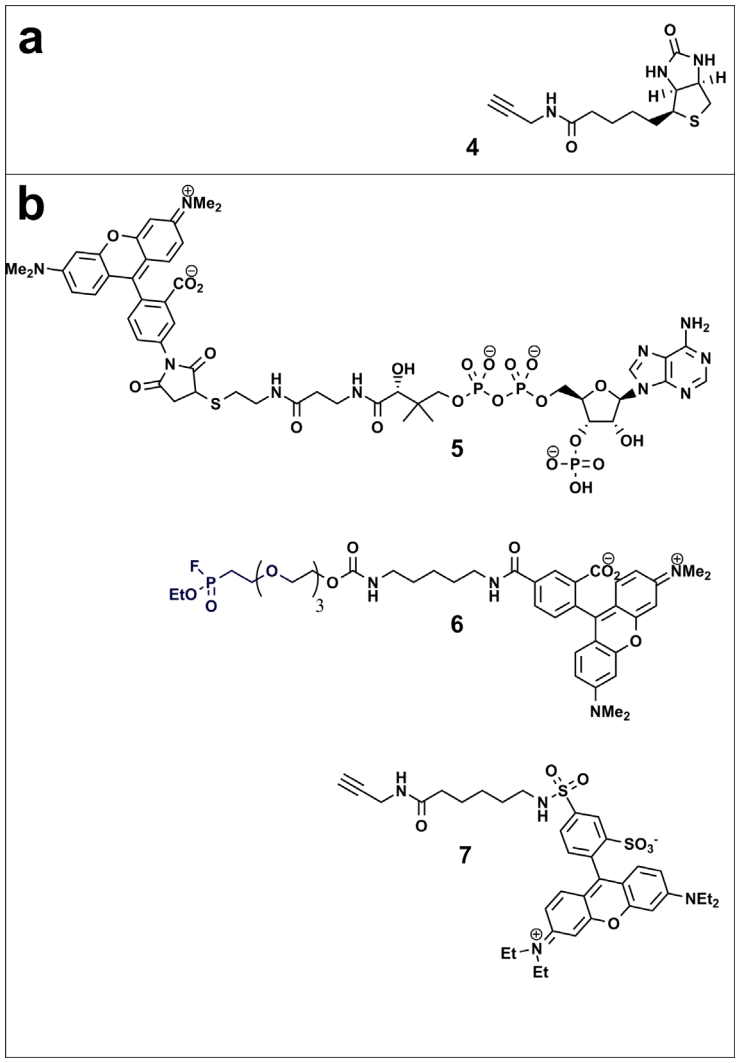


Figure 9.S1 Completed list of probe structures used for profiling PKS and NRPS enzymes in this study. Structures for probes **1**-**3** are given in Figure 9.1. (a) Probe **4** is a biotin alkyne conjugate which can be appended to proteins metabolically labeled by azide **1** using Cu(I)-catalyzed [3+2] cycloaddition (click) chemistry. Biotinylated proteomes are then enriched by standard avidin affinity chromatography. On-bead tryptic digest followed by MudPIT and spectral count analysis is used for protein identification (b) Probes **5**-**7** incorporate fluorescence which can be used for visualization of probe-labeled proteins by in gel fluorescence scanning following SDS-PAGE. This can be used as a facile check of probe-labeling prior to enrichment and MudPIT analysis (see Figure 9S3).

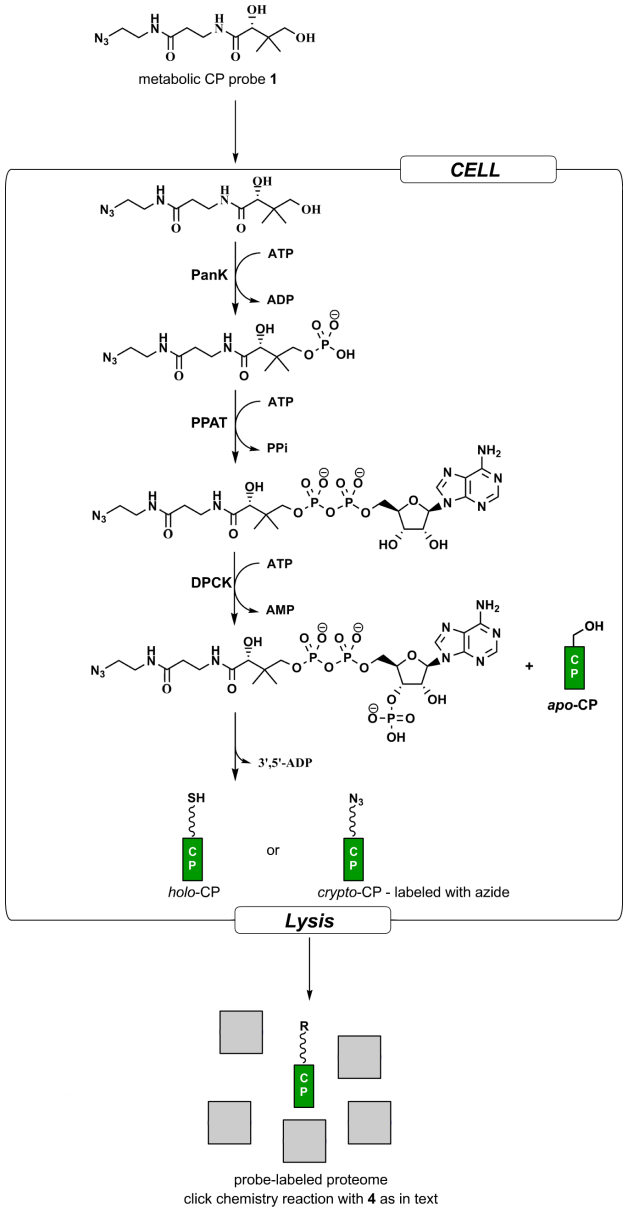


Figure 9.S2 Detailed mechanism of labeling by metabolic CP-probe **1** (process is abbreviated in Figure 1). In vivo labeling strategy. Cells are grown in the presence of azido-CoA precursor **1**. After uptake, the native CoA biosynthetic enzymes (PanK, PPAT, and DPCK) convert **1** to the azido-CoA analogue (depicted in Figure 9.1), which is then transferred to PKS and NRPS CP domains by the endogenous PPTase enzyme. After cell lysis azide-modified CPs can be detected by click chemistry reaction with alkyne probes **4** (affinity) or **7** (fluorescence – see Figure 9S 1 for structures).

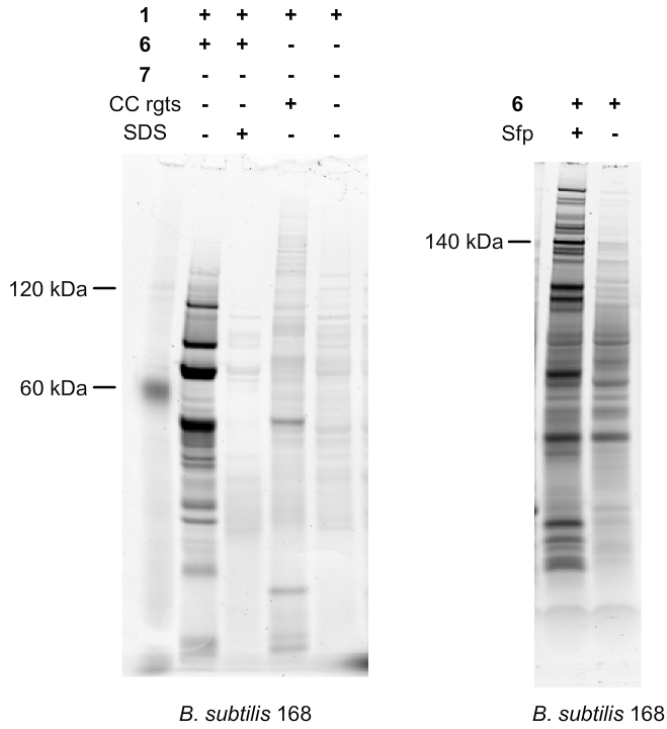


Figure 9.S3 Fluorescent gel images for *B. subtilis* 6051 and 168 samples labeled with probes **5**-**7**. The number of bands labeled by CP probes **1** and **6** is considerably larger than the number of annotated CP containing PKS/NRPS/FAS enzymes, suggesting substantial proteolysis of these proteins occurs following cell lysis. For additional fluorescent profiling data for these two strains see reference 17.

Supplementary Tables

Supplementary Table 9.S1 Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 6051 lysates following enrichment with metabolic CP probe **1**. The generated tandem mass spectra were searched using SEQUEST 3.0 against a concatenated target/decoy[64] non-redundant database incorporating the *B. subtilis* 168 database (NCBI),[49] the human IPI database version 3.33, as well as the corresponding reverse databases. Probe enriched samples represent avidin enrichment of strain 6051 lysates grown with **1** following click chemistry with biotin alkyne **4**, while no probe controls represent avidin enrichment of strain 6051 lysates grown with **1** using an identical protocol, except with omission of biotin alkyne **4**. Listed proteins have spectral count averages ≥ 15 and show enrichment over control reactions at a spectral count ratio of ≥ 10. Protein identifications in red represent known CP-containing proteins identified by this analysis. Protein identifications highlighted in yellow represent protein identifications exhibiting high inter-dataset variability (≥ 30%), as indicated by statistical analysis of standard error of the mean (S.E.M.).

				Probe **1** Enriched					No probe					
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Probe Average	Control Average	Ratio	%SEM
1	AhpF	NP_391890.1	NADH dehydrogenase	63	48	61	70	76	3	3	64	3	21	7
2	AmhX	NP_388183.1	Amidohydrolase	17	9	12	19	22	0	0	16	0	#DIV/0!	15
3	BfmBAA	NP_390285.1	2-oxoisovalerate dehydrogenase subunit alpha	23	12	15	11	17	0	0	16	0	#DIV/0!	14
4	Csd	NP_391148.1	Probable cysteine desulfurase	17	19	22	16	22	0	0	19	0	#DIV/0!	6
5	Dat	NP_388848.1	D-alanine aminotransferase	22	15	39	21	28	0	0	25	0	#DIV/0!	16
6	DeoD	NP_389844.1	Purine nucleoside phosphorylase II	18	12	17	14	18	0	0	16	0	#DIV/0!	8
7	DhbB	NP_391077.1	Isochorismatase (bacillibactin NRPS)	44	35	23	38	30	0	0	34	0	#DIV/0!	10
8	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	40	41	30	40	35	0	0	37	0	#DIV/0!	6
9	DltC	NP_391731.1	D-alanyl carrier protein	33	20	38	18	24	0	0	27	0	#DIV/0!	14
10	DnaK	NP_390425.1	Class I heat-shock protein (molecular chaperone	23	17	15	24	23	0	0	20	0	#DIV/0!	9
11	DnaN	NP_387883.1	DNA polymerase III subunit beta	14	17	20	14	27	0	0	18	0	#DIV/0!	13
12	FbaA	NP_391593.1	Fructose-bisphosphate aldolase	18	20	8	26	32	0	0	21	0	#DIV/0!	19
13	FusA	NP_387993.1	Elongation factor G	76	52	103	62	70	6	5	73	6	13	12

Supplementary Table 9.S1 (continued): Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 6051 following enrichment with metabolic CP probe **1**. Listed proteins have spectral count averages ≥ 15 and ratio of ≥ 10.

				Probe **1** Enriched					No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Probe Average	Control Average	Ratio	%SEM
14	GatA	NP_388550.1	Glutamyl-tRNA amidotransferase subunit A	17	19	8	15	16	0	0	15	0	#DIV/0!	12
15	GcvT	NP_390337.1	Glycine cleavage system T	20	21	12	21	20	0	0	19	0	#DIV/0!	9
16	GlnA	NP_389628.1	Glutamine synthetase	38	29	79	71	77	0	0	59	0	#DIV/0!	18
17	GlyA	NP_391571.1	Serine hydroxymethyltransferase	23	10	13	30	35	3	0	22	2	15	22
18	GuaA	NP_388517.1	GMP synthase/glutamine amidotransferase	21	13	7	21	21	0	0	17	0	#DIV/0!	17
19	HemL	NP_390690.1	Glutamate-1-semialdehyde 2,1-aminomutase	27	17	41	35	42	0	2	32	1	32	14
20	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	14	14	31	33	23	0	0	23	0	#DIV/0!	18
21	HutU	NP_391815.1	Urocanate hydratase	28	18	23	15	27	0	0	22	0	#DIV/0!	11
22	IlvC	NP_390707.1	Ketol-acid reductoisomerase	17	6	12	18	23	0	0	15	0	#DIV/0!	19
23	MntB	NP_390954.1	Manganese transport system ATP-binding prote	8	12	21	25	16	0	0	16	0	#DIV/0!	19
24	MsrA	NP_390052.1	Methionine sulfoxide reductase A	23	19	22	21	20	0	0	21	0	#DIV/0!	3
25	OdhA	NP_389819.2	2-oxoglutarate dehydrogenase E1 component	61	54	61	58	53	2	4	57	3	19	3
26	PdhC	NP_389343.1	Dihydrolipoamide acetyltransferase component (	55	46	62	70	72	3	2	61	3	24	8
27	PepT	NP_391771.1	Tripeptide aminopeptidase	43	30	27	40	38	0	0	36	0	#DIV/0!	9
28	PksJ	NP_389598.2	Bacillaene PKS-NRPS subunit J	65	57	28	51	71	0	0	54	0	#DIV/0!	14
29	PksL	NP_389600.2	Bacillaene PKS-NRPS subunit L	81	55	31	51	60	0	0	56	0	#DIV/0!	14
30	PksM	NP_389601.2	Bacillaene PKS-NRPS subunit M	67	58	29	49	59	0	0	52	0	#DIV/0!	12
31	PksN	NP_389602.2	Bacillaene PKS-NRPS subunit N	54	43	28	53	45	0	0	45	0	#DIV/0!	10
32	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit R	96	99	85	100	88	0	0	94	0	#DIV/0!	3
33	PpaC	NP_391935.1	Manganese-dependent inorganic pyrophosphata	26	19	23	16	21	0	0	21	0	#DIV/0!	8
34	PpsA	NP_389716.1	Plipastatin NRPS subunit A	31	39	23	39	36	0	0	34	0	#DIV/0!	9
35	PpsB	NP_389715.1	Plipastatin NRPS subunit B	17	13	11	29	24	0	0	19	0	#DIV/0!	18
36	PpsD	NP_389713.1	Plipastatin NRPS subunit D	14	20	8	17	18	0	0	15	0	#DIV/0!	14
37	RnjA	NP_389336.1	Ribonuclease J	19	19	12	20	19	0	0	18	0	#DIV/0!	8
38	RocA	NP_391658.1	1-pyrroline-5-carboxylate dehydrogenase	38	26	29	35	41	0	2	34	1	34	8

Supplementary Table 9.S1 (continued): Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 6051 following enrichment with metabolic CP probe **1**. Listed proteins have spectral count averages ≥ 15 and ratio of ≥ 10.

				Probe **1** Enriched					No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Probe Average	Control Average	Ratio	%SEM
39	RocD	NP_391914.1	Ornithine aminotransferase	27	22	20	18	17	0	4	21	2	10	9
40	RocF	NP_391912.1	Arginase	74	52	90	83	115	0	0	83	0	#DIV/0!	12
41	RpsB	NP_389531.1	30S ribosomal protein S2	22	21	19	22	25	0	0	22	0	#DIV/0!	4
42	SdhA	NP_390722.1	Succinate dehydrogenase flavoprotein subunit	112	103	141	80	107	5	5	109	5	22	9
43	SrfAA	NP_388230.1	Surfactin NRPS module 1	54	49	32	56	62	0	0	51	0	#DIV/0!	10
44	SrfAB	NP_388231.1	Surfactin NRPS module 2	116	111	78	103	104	0	0	102	0	#DIV/0!	6
45	SucC	NP_389491.1	Succinyl-CoA ligase subunit beta	14	15	23	14	18	0	0	17	0	#DIV/0!	10
46	SucD	NP_389492.1	Succinyl-CoA ligase subunit alpha	18	28	27	40	52	0	0	33	0	#DIV/0!	18
47	SufB	NP_391146.1	FeS cluster formation protein	32	25	41	29	43	0	3	34	2	23	10
48	Tkt	NP_389672.1	Transketolase	142	107	119	152	196	2	6	143	4	36	11
49	Tsf	NP_389532.1	Elongation factor Ts	13	9	10	17	47	0	0	19	0	#DIV/0!	**37**
50	YcnH	NP_388273.1	Succinate-semialdehyde dehydrogenase	11	20	13	28	29	0	2	20	1	20	18
51	YdjL	NP_388505.1	Acetoin reductase/2,3-butanediol dehydrogenas	27	29	22	38	45	0	0	32	0	#DIV/0!	13
52	YpwA	NP_390090.1	Putative metalloprotease	33	26	34	22	34	0	0	30	0	#DIV/0!	8
53	YqhT	NP_390326.1	Uncharacterized peptidase	24	20	35	28	29	0	0	27	0	#DIV/0!	9
54	YqjE	NP_390271.1	Hypothetical protein (homology to metalloprotea	182	100	158	186	277	0	0	181	0	#DIV/0!	16
55	YqjI	NP_390267.2	6-phosphogluconate dehydrogenase	24	30	13	35	33	0	2	27	1	27	15
56	YtjP	NP_390876.1	Putative dipeptidase	37	24	18	39	38	0	0	31	0	#DIV/0!	14
57	YvgN	NP_391220.1	Glyoxal reductase	12	10	25	21	12	0	0	16	0	#DIV/0!	18
58	Zwf	NP_390266.1	Glucose-6-phosphate 1-dehydrogenase	95	77	129	89	126	0	0	103	0	#DIV/0!	10

Supplementary Table 9.S2 Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 168 lysates following enrichment with chemoenzymatic CP probe **2**. The generated tandem mass spectra were searched using SEQUEST 3.0 against a concatenated target/decoy non-redundant database incorporating the *B. subtilis* 168 database (NCBI),[49] the human IPI database version 3.33, as well as the corresponding reverse databases. Probe enriched samples represent avidin enrichment of strain 168 proteomes after treatment with biotin CoA **2** and Sfp, while no probe controls represent avidin enrichment of strain 168 proteomes treated with Sfp where biotin CoA **2** was omitted. Listed proteins have spectral count averages ≥ 15 and show enrichment over control reactions at a spectral count ratio of ≥ 10. Protein identifications in red represent known CP-containing proteins identified by this analysis. Protein identifications highlighted in yellow represent protein identifications exhibiting high inter-dataset variability (≥ 30%), as indicated by statistical analysis of standard error of the mean (S.E.M.).

				Probe **2** Enriched					No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
1	AbfA	NP_390750.1	Alpha-N-arabinofuranosidase	16	14	24	67	26	2	0	0	29	1	44	**33**
2	AhpC	NP_391889.1	Alkyl hydroperoxide reductase subunit C	283	226	97	254	73	4	11	13	187	9	20	23
3	AhpF	NP_391890.1	NADH dehydrogenase	45	31	44	74	33	0	4	6	45	3	14	17
4	AlsD	NP_391481.1	Alpha-acetolactate decarboxylase	22	6	10	39	7	0	0	0	17	0	#DIV/0!	**37**
5	DhbB	NP_391077.1	Isochorismatase (bacillibactin NRPS)	1200	925	413	712	555	0	0	0	761	0	#DIV/0!	18
6	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	264	205	183	541	242	0	0	0	287	0	#DIV/0!	23
7	FbaA	NP_391593.1	Fructose-bisphosphate aldolase	24	26	32	106	31	0	0	0	44	0	#DIV/0!	**36**
8	FusA	NP_387993.1	Elongation factor G	53	48	78	132	43	0	13	8	71	7	10	23
9	GlmS	NP_388059.1	Glucosamine--fructose-6-phosphate aminotransferase	4	7	19	54	18	0	0	0	20	0	#DIV/0!	**44**
10	GlnA	NP_389628.1	Glutamine synthetase	384	353	90	347	136	7	13	7	262	9	29	24
11	GlvA	NP_388699.1	Maltose-6'-phosphate glucosidase	52	41	56	103	32	0	9	7	57	5	11	22
12	GlyA	NP_391571.1	Serine hydroxymethyltransferase	17	17	15	57	30	5	2	0	27	2	12	29
13	GuaA	NP_388517.1	GMP synthase/glutamine amidotransferase	23	36	32	83	32	0	2	0	41	1	62	26
14	LeuS	NP_390910.1	Leucyl-tRNA synthetase	14	9	4	39	10	0	0	0	15	0	#DIV/0!	**41**
15	MurF	NP_388338.1	UDP-MurNAc-pentapeptide synthetase	7	4	26	49	4	0	0	4	18	1	14	**49**

Supplementary Table 9.S2 (continued): Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 168 following enrichment with chemoenzymatic CP probe **2**. Listed proteins have spectral count averages ≥ 15 and ratios ≥ 10.

				Probe **2** Enriched					No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
16	NamA	NP_390263.1	NADPH dehydrogenase	5	6	23	45	13	0	0	0	18	0	#DIV/0!	**40**
17	OdhA	NP_389819.2	2-oxoglutarate dehydrogenase E1 component	10	12	18	38	15	4	0	0	19	1	14	27
18	PdhC	NP_389343.1	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex	16	16	20	45	32	0	0	0	26	0	#DIV/0!	22
19	PepT	NP_391771.1	Tripeptide aminopeptidase	21	22	21	40	22	0	0	0	25	0	#DIV/0!	15
20	PksJ	NP_389598.2	Bacillaene PKS-NRPS subunit J	15	21	16	59	33	0	0	0	29	0	#DIV/0!	28
21	PksL	NP_389600.2	Bacillaene PKS-NRPS subunit L	23	21	24	41	24	0	0	0	27	0	#DIV/0!	14
22	Pnp	NP_389551.1	Polyribonucleotide nucleotidyltransferase	13	9	24	50	11	0	0	0	21	0	#DIV/0!	**36**
23	PpaC	NP_391935.1	Manganese-dependent inorganic pyrophosphatase	10	7	16	38	8	0	0	0	16	0	#DIV/0!	**36**
24	PpsA	NP_389716.1	Plipastatin NRPS subunit A	100	99	112	229	105	0	0	0	129	0	#DIV/0!	19
25	PpsB	NP_389715.1	Plipastatin NRPS subunit B	266	228	250	397	153	0	0	0	259	0	#DIV/0!	15
26	PpsC	NP_389714.1	Plipastatin NRPS subunit C	92	88	89	225	91	0	0	0	117	0	#DIV/0!	23
27	PpsD	NP_389713.1	Plipastatin NRPS subunit D	106	97	114	178	112	0	0	0	121	0	#DIV/0!	12
28	Pta	NP_391646.1	Phosphotransacetylase	20	10	16	34	10	0	0	0	18	0	#DIV/0!	25
29	PurB	NP_388526.1	Adenylosuccinate lyase	53	49	16	58	42	12	0	0	44	4	11	17
30	SrfAA	NP_388230.1	Surfactin NRPS module 1	562	488	566	900	398	2	0	0	583	1	874	15
31	SrfAB	NP_388231.1	Surfactin NRPS module 2	1857	1409	960	1902	944	0	0	0	1414	0	#DIV/0!	15
32	SrfAC	NP_388233.1	Surfactin NRPS module 3	369	301	114	389	191	0	0	0	273	0	#DIV/0!	19
33	SufB	NP_391146.1	FeS cluster formation protein	31	32	41	71	37	0	5	0	42	2	25	17
34	Tkt	NP_389672.1	Transketolase	90	76	41	201	40	0	0	2	90	1	134	**33**
35	TpiA	NP_391272.1	Triosephosphate isomerase	69	102	136	295	72	0	5	2	135	2	58	**31**
36	Tuf	NP_387994.1	Elongation factor Tu	13	10	14	36	10	0	0	0	17	0	#DIV/0!	**30**
37	YjbG	NP_389036.1	Oligoendopeptidase F homolog	11	15	7	59	24	3	0	3	23	2	12	**40**

Supplementary Table 9.S2 (continued): Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 168 following enrichment with chemoenzymatic CP probe **2**. Listed proteins have spectral count averages ≥ 15 and ratios ≥ 10.

				Probe **2** Enriched						No probe					
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
38	YjiB	NP_389103.1	Putative cytochrome P450	9	15	11	28	12	0	0	0	15	0	#DIV/0!	23
39	YjiC	NP_389104.1	Uncharacterized UDP-glucosyltransferase	5	8	15	35	13	0	0	0	15	0	#DIV/0!	**35**
40	YqiT	NP_390288.1	Leucine dehydrogenase	5	5	15	46	14	0	0	0	17	0	#DIV/0!	**44**
41	YqjE	NP_390271.1	Hypothetical protein (homology to metalloprotease)	6	5	18	40	16	0	0	0	17	0	#DIV/0!	**37**
42	YqjI	NP_390267.2	6-phosphogluconate dehydrogenase	88	64	93	205	71	3	11	6	104	7	16	25
43	YugJ	NP_391015.1	Probable NADH-dependent butanol dehydrogenase 1	12	13	12	29	16	0	0	0	16	0	#DIV/0!	20
44	YumC	NP_391091.1	Ferredoxin-NADP reductase 2	19	17	20	41	15	0	0	3	22	1	22	21
45	YurY	NP_391150.1	Vegetative protein 296 (homology to ATP binding)	44	32	43	43	34	0	0	0	39	0	#DIV/0!	7

Supplementary Table 9.S3 Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 6051 lysates following enrichment with ABPP probe **3**. The generated tandem mass spectra were searched using SEQUEST 3.0 against a concatenated target/decoy non-redundant database incorporating the *B. subtilis* 6051 database (NCBI),[49] the human IPI database version 3.33, as well as the corresponding reverse databases. Probe enriched samples represent avidin enrichment of strain 6051 proteomes following treatment with fluorophosphonate **3**, while no probe controls represent avidin enrichment of strain 6051 proteomes untreated with fluorophosphonate **3**. Listed proteins have spectral count averages ≥ 15 and show enrichment over control reactions at a spectral count ratio of ≥ 10. Protein identifications in red represent known TE-containing proteins identified by this analysis.

				Probe **3** Enriched				No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
1	AspB	NP_390118.1	Aspartate aminotransferase	19	16	11	23	0	0	0	17	0	#DIV/0!	15
2	AtpD	NP_391562.1	ATP synthase subunit beta (annotated hydr	24	26	22	11	0	0	0	21	0	#DIV/0!	16
3	Cah	NP_388200.1	Cephalosporin C deacetylase	104	134	77	110	0	0	0	106	0	#DIV/0!	11
4	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	24	18	13	22	0	0	0	19	0	#DIV/0!	13
5	Est	NP_391242.1	Carboxylesterase	145	92	79	112	0	0	0	107	0	#DIV/0!	13
6	GatA	NP_388550.1	Glutamyl-tRNA amidotransferase subunit A	22	17	20	24	0	0	0	21	0	#DIV/0!	7
7	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	29	31	35	43	0	5	4	35	3	12	9
8	Nap	NP_388425.1	Uncharacterized carboxylesterase	102	71	57	98	0	0	0	82	0	#DIV/0!	13
9	PbpE	NP_391324.1	Penicillin-binding protein 4	68	58	49	78	0	0	0	63	0	#DIV/0!	10
10	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit R	73	99	72	87	2	3	0	83	2	50	8
11	PnbA	NP_391319.1	Para-nitrobenzyl esterase	358	280	217	312	2	12	0	292	5	63	10
12	PpsE	NP_389712.1	Plipastatin NRPS subunit E	19	17	21	20	0	0	0	19	0	#DIV/0!	4
13	RsbQ	NP_391290.1	Sigma factor sigB regulation protein (alpha/b	46	21	18	41	0	0	0	32	0	#DIV/0!	22
14	SppA	NP_390831.1	Signal peptide peptidase	20	17	19	23	0	0	0	20	0	#DIV/0!	6
15	SrfAD	NP_388234.1	Surfactin NRPS type II TE domain	29	28	29	27	0	0	0	28	0	#DIV/0!	2

Supplementary Table 9.S3 (continued): Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 6051 following enrichment with activity-based probe **3**. Listed proteins have spectral count averages ≥ 15 and ratios ≥ 10.

				Probe **3** Enriched				No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
16	WprA	NP_388958.1	Cell wall-associated protease	693	279	213	653	0	0	0	460	0	#DIV/0!	27
17	YclE	NP_388248.1	AB hydrolase superfamily protein	99	48	49	94	0	0	0	73	0	#DIV/0!	19
18	YdeN	NP_388407.1	Putative hydrolase	28	26	20	22	0	0	0	24	0	#DIV/0!	8
19	YesY	NP_388588.1	Probable rhamnogalacturonan acetylesteras	36	37	28	34	0	0	0	34	0	#DIV/0!	6
20	YfmJ	NP_388626.1	Putative oxidoreductase	57	46	48	42	0	0	0	48	0	#DIV/0!	7
21	YitV	NP_388996.1	Putative esterase	77	83	70	89	0	0	0	80	0	#DIV/0!	5
22	YjcH	NP_389068.1	Putative hydrolase	149	101	74	135	0	0	0	115	0	#DIV/0!	15
23	YodC	NP_389836.1	Putative NAD(P)H nitroreductase	70	32	24	56	0	0	0	46	0	#DIV/0!	23
24	YolF	NP_389837.1	Putative hydrolase	104	63	62	99	0	0	0	82	0	#DIV/0!	14
25	YqiT	NP_390288.1	Leucine dehydrogenase	25	15	11	21	0	0	4	18	1	14	17
26	YqkD	NP_390245.1	Hypothetical protein	17	22	11	14	0	0	0	16	0	#DIV/0!	15
27	YtaP	NP_390903.1	Putative hydrolase	68	50	33	47	0	0	0	50	0	#DIV/0!	15
28	YtpA	NP_390929.1	Phospholipase component of bacilysocin sy	63	25	24	43	0	0	0	39	0	#DIV/0!	24
29	YuxL	NP_391103.1	Uncharacterized peptidase	452	196	159	450	0	0	0	314	0	#DIV/0!	25

Supplementary Table 9.S4 Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 168 lysates following enrichment with ABPP probe **3**. The generated tandem mass spectra were searched using SEQUEST 3.0 against a concatenated target/decoy non-redundant database incorporating the *B. subtilis* 168 database (NCBI),[49] the human IPI database version 3.33, as well as the corresponding reverse databases. Probe enriched samples represent avidin enrichment of strain 168 proteomes following treatment with fluorophosphonate **3**, while no probe controls represent avidin enrichment of strain 168 proteomes untreated with fluorophosphonate **3**. Listed proteins have spectral count averages ≥ 15 and show enrichment over control reactions at a spectral count ratio of ≥ 10. Protein identifications in red represent known TE-containing proteins identified by this analysis.

				Probe **3** Enriched					No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
1	AhpF	NP_391890.1	NADH dehydrogenase	59	104	52	54	45	0	0	4	63	1	47	17
2	AlsS	NP_391482.1	Acetolactate synthase	20	30	15	24	28	0	0	0	23	0	#DIV/0!	12
3	AsnB	NP_390932.1	Asparagine synthetase	31	36	31	38	48	0	0	0	37	0	#DIV/0!	8
4	Bpr	NP_389413.1	Bacillopeptidase F	222	330	152	145	145	0	0	0	199	0	#DIV/0!	18
5	Cah	NP_388200.1	Cephalosporin C deacetylase	561	526	391	324	264	0	0	0	413	0	#DIV/0!	14
6	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	269	379	262	245	319	0	0	5	295	2	177	8
7	Est	NP_391242.1	Carboxylesterase	217	340	241	159	236	0	0	0	239	0	#DIV/0!	12
8	Isp	NP_389202.1	Major intracellular serine protease	577	941	422	337	336	0	0	0	523	0	#DIV/0!	22
9	KatA	NP_388762.1	Vegetative catalase 1	53	83	32	39	40	0	0	3	49	1	49	18
10	Mdh	NP_390790.1	Malate dehydrogenase	16	27	11	15	14	0	0	2	17	1	25	16
11	MetG	NP_387919.1	Methionyl-tRNA synthetase	17	27	8	15	17	0	0	0	17	0	#DIV/0!	18
12	Nap	NP_388425.1	Uncharacterized carboxylesterase	162	323	166	97	163	0	0	0	182	0	#DIV/0!	21
13	NarH	NP_391608.1	Nitrate reductase beta chain	21	26	42	17	20	0	0	4	25	1	19	18
14	PbpE	NP_391324.1	Penicillin-binding protein 4	59	123	39	36	48	0	0	0	61	0	#DIV/0!	26
15	PdhB	NP_389342.1	Pyruvate dehydrogenase E1 component subunit beta	14	33	21	20	17	0	0	0	21	0	#DIV/0!	15
16	PnbA	NP_391319.1	Para-nitrobenzyl esterase	1073	1317	1312	743	1070	34	20	41	1103	32	35	10
17	PpsE	NP_389712.1	Plipastatin NRPS subunit E	65	83	94	52	76	0	0	0	74	0	#DIV/0!	10
18	PyrAB	NP_389435.1	Carbamoyl phosphate synthase large subunit	13	22	9	18	16	0	0	0	16	0	#DIV/0!	14

Supplementary Table 9.S4 (continued): Spectral count data of proteins identified by MudPIT analysis of *B. subtilis* 168 following enrichment with activity-based probe **3**. Listed proteins have spectral count averages ≥ 15 and ratios ≥ 10.

				Probe **3** Enriched					No probe						
No.	Protein	Accession Number	Description	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Control 1	Control 2	Control 3	Probe Average	Control Average	Ratio	%SEM
19	RsbQ	NP_391290.1	Sigma factor sigB regulation protein (alpha/beta hydrolase)	115	128	52	60	83	0	0	0	88	0	#DIV/0!	17
20	SppA	NP_390831.1	Signal peptide peptidase	54	57	79	52	49	0	0	0	58	0	#DIV/0!	9
21	SrfAA	NP_388230.1	Surfactin NRPS module A	15	14	22	23	39	0	0	5	23	2	14	20
22	SrfAC	NP_388233.1	Surfactin NRPS module C	312	249	251	186	243	0	0	0	248	0	#DIV/0!	8
23	SrfAD	NP_388234.1	Surfactin NRPS type II TE domain	217	251	185	129	162	0	0	0	189	0	#DIV/0!	11
24	Tuf	NP_387994.1	Elongation factor Tu	39	36	40	35	31	0	0	6	36	2	18	4
25	WprA	NP_388958.1	Cell wall-associated protease	852	969	784	842	609	0	0	0	811	0	#DIV/0!	7
26	YclE	NP_388248.1	AB hydrolase superfamily protein	49	52	35	37	54	0	0	0	45	0	#DIV/0!	9
27	YcnH	NP_388273.1	Succinate-semialdehyde dehydrogenase	19	22	19	35	31	0	0	0	25	0	#DIV/0!	13
28	YdeN	NP_388407.1	Putative hydrolase	67	108	76	19	21	0	0	0	58	0	#DIV/0!	29
29	YdjL	NP_388505.1	Acetoin reductase/2,3-butanediol dehydrogenase	14	7	16	15	23	0	0	3	15	1	15	17
30	YfkO	NP_388664.1	Putative NAD(P)H nitroreductase	30	30	21	21	33	0	0	0	27	0	#DIV/0!	9
31	YitV	NP_388996.1	Putative esterase	105	123	196	49	155	0	0	0	126	0	#DIV/0!	20
32	YocD	NP_389798.1	Hypothetical protein (homology to carboxypeptidase)	51	88	66	35	59	0	0	0	60	0	#DIV/0!	15
33	YodC	NP_389836.1	Putative NAD(P)H nitroreductase	122	109	109	106	91	0	0	0	107	0	#DIV/0!	5
34	YolF	NP_389837.1	Putative hydrolase	90	199	248	70	65	0	0	0	134	0	#DIV/0!	28
35	YtpA	NP_390929.1	Phospholipase component of bacilysocin synthesis or export	54	90	40	40	30	0	0	0	51	0	#DIV/0!	21
36	YuxL	NP_391103.1	Uncharacterized peptidase	391	485	346	310	283	0	0	0	363	0	#DIV/0!	10

Supplementary Table 9.S5 Compiled list of proteins showing functional enrichment in multiple proteomic samples (spectral count averages ≥ 15, probe:no probe ratio of ≥ 10, S.E.M. ≤ 30%). (a) Proteins commonly identified by orthogonal active site probes **1** and **3** in *B. subtilis* 6051. (b) Proteins commonly identified by orthogonal active site probes **2** and **3** in *B. subtilis* 168. (c) Proteins commonly identified in strains 168 and 6051 by CP probes **1** and **2**. See tables 1-4 for complete spectral count data.

(a)

No.	Protein	Accession Number	Description	Probe **1** Average strain 6051	Probe **3** Average strain 6051
1	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	37	19
2	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit R	94	83
3	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	23	35
4	GatA	NP_388550.1	Glutamyl-tRNA amidotransferase subunit A	15	21

(b)

No.	Protein	Accession Number	Description	Probe **2** Average strain 6051	Probe **3** Average strain 168
1	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	287	295
2	SrfAA	NP_388230.1	Surfactin NRPS module A	583	23
3	SrfAC	NP_388233.1	Surfactin NRPS module C	273	248
4	AhpF	NP_391890.1	NADH dehydrogenase	45	63

Supplementary Table 9.S5 (continued from previous page).

(c)

No.	Protein	Accession Number	Description	Probe **1** Average strain 6051	Probe **2** Average strain 168
1	AhpF	NP_391890.1	NADH dehydrogenase	64	45
2	FusA	NP_387993.1	Elongation factor G	73	71
3	GlnA	NP_389628.1	Glutamine synthetase	59	262
4	GlyA	NP_391571.1	Serine hydroxymethyltransferase	22	27
5	GuaA	NP_388517.1	GMP synthase/glutamine amidotransferase	17	41
6	OdhA	NP_389819.2	2-oxoglutarate dehydrogenase E1 component	57	19
7	PdhC	NP_389343.1	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex	61	26
8	PepT	NP_391771.1	Tripeptide aminopeptidase	36	25
9	SufB	NP_391146.1	FeS cluster formation protein	34	42
10	YqjI	NP_390267.2	6-phosphogluconate dehydrogenase	27	104

Supplementary Table 9.S6 Proteins enriched by fluorophosphonate **3** showing differential activity levels in *B. subtilis* strains 6051 and 168. Spectral counts given in red indicate enzyme activity was observed at greater levels in that proteomes isolated from that strain. See tables 3-4 for complete spectral count data.

No.	Protein	Accession Number	Description	Probe **3** Average (6051)	Probe **3** Average (168)
1	YjcH	NP_389068.1	Putative hydrolase	115	<15
2	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit R	83	<15
3	YtaP	NP_390903.1	Putative hydrolase	50	<15
4	YfmJ	NP_388626.1	Putative oxidoreductase	48	<15
5	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	35	<15
6	YesY	NP_388588.1	Probable rhamnogalacturonan acetylesterase	34	<15
7	GatA	NP_388550.1	Glutamyl-tRNA amidotransferase subunit A	21	<15
8	AtpD	NP_391562.1	ATP synthase subunit beta (annotated hydrolase)	21	<15
9	YqiT	NP_390288.1	Leucine dehydrogenase	18	<15
10	AspB	NP_390118.1	Aspartate aminotransferase	17	<15
11	YqkD	NP_390245.1	Hypothetical protein	16	<15
12	Isp	NP_389202.1	Major intracellular serine protease	<15	522.6
13	SrfAC	NP_388233.1	Surfactin NRPS module C	<15	248.2
14	Bpr	NP_389413.1	Bacillopeptidase F	<15	198.8
15	AhpF	NP_391890.1	NADH dehydrogenase	<15	62.8
16	YocD	NP_389798.1	Hypothetical protein (homology to carboxypeptidase)	<15	59.8
17	KatA	NP_388762.1	Vegetative catalase 1	<15	49.4
18	AsnB	NP_390932.1	Asparagine synthetase	<15	36.8
19	Tuf	NP_387994.1	Elongation factor Tu	<15	36.2
20	YfkO	NP_388664.1	Putative NAD(P)H nitroreductase	<15	27
21	NarH	NP_391608.1	Nitrate reductase beta chain	<15	25.2
22	YcnH	NP_388273.1	Succinate-semialdehyde dehydrogenase	<15	25.2
23	AlsS	NP_391482.1	Acetolactate synthase	<15	23.4
24	SrfAA	NP_388230.1	Surfactin NRPS module A	<15	22.6
25	PdhB	NP_389342.1	Pyruvate dehydrogenase E1 component subunit beta	<15	21

Supplementary Table 9.S6 (continued): Proteins enriched by fluorophosphonate **3** showing differential activity levels in *B. subtilis* strains 6051 and 168. Spectral counts given in red indicate enzyme activity was observed at greater levels in that proteomes isolated from that strain. See tables 3-4 for complete spectral count data.

No.	Protein	Accession Number	Description	Probe **3** Average (6051)	Probe **3** Average (168)
26	MetG	NP_387919.1	Methionyl-tRNA synthetase	<15	16.8
27	Mdh	NP_390790.1	Malate dehydrogenase	<15	16.6
28	PyrAB	NP_389435.1	Carbamoyl phosphate synthase large subunit	<15	15.6
29	YdjL	NP_388505.1	Acetoin reductase/2,3-butanediol dehydrogenase	<15	15
30	Cah	NP_388200.1	Cephalosporin C deacetylase	106	413
31	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	19	295
32	Est	NP_391242.1	Carboxylesterase	107	239
33	Nap	NP_388425.1	Uncharacterized carboxylesterase	82	182
34	PbpE	NP_391324.1	Penicillin-binding protein 4	63	61
35	PnbA	NP_391319.1	Para-nitrobenzyl esterase	292	1103
36	PpsE	NP_389712.1	Plipastatin NRPS subunit E	19	74
37	RsbQ	NP_391290.1	Sigma factor sigB regulation protein (alpha/beta hydrol	32	88
38	SppA	NP_390831.1	Signal peptide peptidase	20	58
39	SrfAD	NP_388234.1	Surfactin NRPS type II TE domain	28	189
40	WprA	NP_388958.1	Cell wall-associated protease	460	811
41	YclE	NP_388248.1	AB hydrolase superfamily protein	73	45
42	YdeN	NP_388407.1	Putative hydrolase	24	58
43	YitV	NP_388996.1	Putative esterase	80	126
44	YodC	NP_389836.1	Putative NAD(P)H nitroreductase	46	107
45	YolF	NP_389837.1	Putative hydrolase	82	134
46	YtpA	NP_390929.1	Phospholipase component of bacilysocin synthesis or	39	51
47	YuxL	NP_391103.1	Uncharacterized peptidase	314	363

Supplementary Table 9.S7 Assessing the orthogonal nature of CP and TE probes on the peptide level. Peptides listed were identified in all probe positive samples (in which enrichment was used) and in no probe negative samples (no probe controls, in which **3** or click chemistry with **4** were not used to enrich proteins). Numbers indicate spectral count for that peptide (number of times sampled per MudPIT LC-MS/MS run).

No.	Peptide	Protein	Accession number	Description	Probe **1** Enriched					Probe **3** Enriched				No probe				
					Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 1	Sample 2	Sample 3	Sample 4	Control 1	Control 2	Control 3	Control 4	Control 5
1	K.MGSGTDASEFADKEPFDVLVVGGGPAGASAAIYTAR.K	AhpF	NP_391890.1	NADH dehydrogenase	1	2	1	1	4	2	2	2	1	0	0	0	0	0
2	K.DMLALVDELASMSSK.I	AhpF	NP_391890.1	NADH dehydrogenase	4	5	5	4	5	2	1	2	1	0	0	0	0	0
3	R.FGGQVLDTMSIENFISVK.A	AhpF	NP_391890.1	NADH dehydrogenase	3	7	12	11	8	4	3	7	3	0	0	0	0	0
4	K.VSAGEDDTSKDMLALVDELASMSSK.I	AhpF	NP_391890.1	NADH dehydrogenase	11	11	1	21	25	2	1	3	1	0	0	0	0	0
5	R.LSYATSLDLLEEAIER.I	AspB	NP_390118.1	Aspartate aminotransferase	4	3	1	4	3	5	3	4	5	0	0	0	0	0
6	K.AIDALIPIGR.G	AtpA	NP_391564.1	ATP synthase subunit alpha	1	1	3	1	1	1	2	1	1	0	0	0	0	0
7	K.MLTEIGEVENAEPYIR.A	CitZ	NP_390792.1	Citrate synthase 2	2	1	3	1	1	1	1	1	1	0	0	0	0	0
8	K.IAHDNGAVIVVDGAQSTPHMK.I	Csd	NP_391148.1	cysteine desulfurase [Bacillus subtilis subsp. subtilis str. 168]	3	6	2	4	4	1	2	5	1	0	0	0	0	0
9	K.EGVTVLNQTPSAFYQFMQAER.E	DhbF	NP_391076.2	Dimodular bacillibactin NRPS	5	2	2	1	1	1	1	1	1	0	0	0	0	0
10	K.MDSTSANVDVPDFLSSLTGDIK.G	GatA	NP_388550.1	Glutamyl-tRNA amidotransferase subunit A	5	3	2	3	3	2	3	4	2	0	0	0	0	0
11	R.YGLVAFASSLDQIGPITR.T	GatA	NP_388550.1	Glutamyl-tRNA amidotransferase subunit A	3	3	1	4	5	6	2	3	4	0	0	0	0	0
12	K.SMFIVSANPLALGLLTPPGK.F	GcvPA	NP_390336.1	Glycine dehydrogenase subunit 1	1	1	4	3	4	1	1	2	1	0	0	0	0	0
13	R.LIVGAAVGVTGDTMTR.V	GuaB	NP_387890.1	Inosine-5'-monophosphate dehydrogenase	4	4	16	3	9	3	2	2	2	0	0	0	0	0
14	K.ALAAGGHAVMLGSLLAGTSESPGETEIYQGR.R	GuaB	NP_387890.1	Inosine-5'-monophosphate dehydrogenase	7	5	3	7	9	9	5	4	5	0	0	0	0	0
15	K.AVSADHLVGTSDEGIKK.L	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	1	2	1	1	1	1	1	2	2	0	0	0	0	0
16	K.IHADEIDPLGGAELAGK.L	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	1	1	1	2	1	1	1	4	3	0	0	0	0	0
17	K.LAEAGTIAVLLPGTTFYLGK.S	HutI	NP_391816.1	Imidazolone-5-propionate hydrolase	1	1	3	3	1	2	4	3	3	0	0	0	0	0
18	K.FDGIATNAVFAWVDER.A	IolH	NP_391848.1	Inositol utilization protein H	2	2	4	3	2	1	5	3	2	0	0	0	0	0
19	R.TFPELIPLNDGKR.G	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit L	2	1	2	2	1	2	1	2	2	0	0	0	0	0
20	R.AYDLESYTVKPLPDPETVK.C	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit L	3	2	2	2	1	2	1	2	2	0	0	0	0	0
21	R.SFFGDLDTYFTLSNEK.E	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit L	2	4	5	4	3	1	4	3	5	0	0	0	0	0
22	K.SQRPFYGIQAR.G	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit L	1	1	5	1	1	5	7	7	5	0	0	0	0	0

Supplementary Table 9.S7 (continued):

No.	Peptide	Protein	Accession number	Description	Probe **1** Enriched						Probe **3** Enriched					No probe				
					Sample 1	Sample 2	Sample 3	Sample 4	Sample 5		Sample 1	Sample 2	Sample 3	Sample 4		Control 1	Control 2	Control 3	Control 4	Control 5
23	K.SSVPVFSGNAENGSSDLEALLDGPLR.E	PksR	NP_389604.1	Bacillaene PKS-NRPS subunit L	7	7	4	6	3		3	2	3	3		0	0	0	0	0
24	R.NLVEEELANIWK.Q	PpsE	NP_389712.1	Plipastatin NRPS subunit E	3	1	1	1	1		1	1	1	1		0	0	0	0	0
25	K.IKPENVVIIGAR.S	RocF	NP_391912.1	Arginase	4	3	5	7	21		1	1	2	3		0	0	0	0	0
26	K.TAVELVESLLGK.K	RocF	NP_391912.1	Arginase	21	3	19	9	7		1	1	1	1		0	0	0	0	0
27	R.GIVAQNLTNMQIESFR.S	SdhA	NP_390722.1	Succinate dehydrogenase flavoprotein subunit	4	1	18	1	1		1	1	1	1		0	0	0	0	0
28	K.YVNGLESSAEDMSSSLFDAHVK.K	SdhA	NP_390722.1	Succinate dehydrogenase flavoprotein subunit	11	6	6	3	6		1	4	3	1		0	0	0	0	0
29	K.QLPAYMVPQTFTFLDELPLTTNGK.V	SrfAC	NP_388233.1	Surfactin NRPS module 3	1	1	1	1	1		1	1	1	1		0	0	0	0	0
30	K.EELTELLDLATK.A	YodC	NP_389836.1	Putative NAD(P)H nitroreductase	2	1	1	1	1		4	3	2	4		0	0	0	0	0
31	K.QIVESSAVVAILGDLK.A	YodC	NP_389836.1	Putative NAD(P)H nitroreductase	1	2	2	2	2		8	6	8	4		0	0	0	0	0
32	K.ANENGEEVYAELASQGYITDEIK.Q	YodC	NP_389836.1	Putative NAD(P)H nitroreductase	5	3	1	4	3		23	6	1	32		0	0	0	0	0
33	R.HLHEEGANLIVTDINK.Q	YqiT	NP_390288.1	Leucine dehydrogenase	6	2	2	3	6		9	4	4	4		0	0	0	0	0
34	K.DSKPHHVSADHPLVK.T	YtjP	NP_390876.1	Putative dipeptidase	2	2	5	8	3		1	1	1	1		0	0	0	0	0
35	R.INSVMDETTAGPGKPFGEGVNASLTSLLELGEK.E	YtjP	NP_390876.1	Putative dipeptidase	4	1	1	8	10		1	1	1	2		0	0	0	0	0
36	K.AGATVHELAGVEPNPR.V	YugJ	NP_391015.1	Probable NADH-dependent butanol dehydrogenase 1	1	1	11	1	1		2	1	1	1		0	0	0	0	0
37	K.DYDDVMQAVDEAIKR.D	YuxL	NP_391103.1	Uncharacterized peptidase	1	2	1	2	1		6	4	1	7		0	0	0	0	0
38	K.LITADDITAIVSVTDPQYAPDGTR.A	YuxL	NP_391103.1	Uncharacterized peptidase	2	3	1	2	2		102	39	35	81		0	0	0	0	0
39	K.IAGGANVDDFNWTDVLMK.K	Xsa	NP_390729.1	Alpha-N-arabinofuranosidase	3	2	1	1	2		1	1	2	1		0	0	0	0	0

Materials

Pantetheine azide **1**,[18] Biotin CoA **2**,[16] fluorophosphonate biotin **3**,[21] Biotin alkyne **4**,[19] TAMRA CoA **5**, [16] fluorophosphonate TAMRA **6**,[35] and rhodamine alkyne **7** [15] were synthesized as previously described. Recombinant Sfp was expressed and purified as previously described.[16] Luria-Bertani (LB) media was purchased from Aldrich. Bis-Tris NuPage gels (4-12%) were purchased from Invitrogen. PD10 desalting columns were purchased from GE Healthcare. Avidin-agarose was purchased from Aldrich. Capillary columns were prepared by drawing 100 μm inner diameter deactivated, fused silica tubing (Agilent) with a Model P-2000 laser puller (Sutter Instruments Co.) and packed at ~ 600 psi with the appropriate chromatography resin (Aqua C18 reverse phase resin [Phenomex] or Partisphere strong cation exchange resin [Whatman]) suspended in methanol. Desalting columns were packed with 3 cm C18 resin, while biphasic MudPIT columns were packed with 10 cm C18 and 3 cm strong cation exchange (SCX) resin. Fluorescent visualization probe-labeled proteins by SDS-PAGE was performed using a Typhoon Laser Flatbed Scanner (GE Healthcare) with excitation at 532 nm. Total protein content was visualized by Blue-silver coomassie stain made according the published procedures.[65] LC-MS analysis was performed using an LTQ ion trap mass spectrometer (ThermoFisher) coupled to an Agilent 1100 series HPLC.

Culture Conditions and Proteome Preparation

B. subtilis strains 168 and 6051 were streaked on LB-agar and incubated overnight at 37 °C. A single colony of each strain was picked and used to inoculate individual 5 mL liquid LB starter cultures and rotated overnight at 37 °C. These starter cultures (2 mL each) were then used to inoculate 1 L of autoclaved LB media. Media was supplemented with **1** (1 mM; 1000 mM stock solution in DMSO) prior to inoculation of *B. subtilis* 6051, while an equivalent amount of vehicle DMSO (0.1%) was added to *B. subtilis* 168 cultures. These cultures were then grown aerobically at 37 °C with vigorous agitation and their growth curves plotted by analyzing optical density at 600 nm. Cells were harvested in stationary growth phase at an OD_{600} of 1.3. After centrifugation (8000xg for 20 min at 4 °C) cell pellets were washed twice with lysis buffer (25 mM potassium phosphate, pH 7.0, 100 mM NaCl) and again centrifuged. After resuspension in lysis buffer (50-100 mL), cell lysis is performed by two passes through a French pressure cell, followed by treatment with DNase I for 30 minutes at 0 °C and clearing of cell debris by centrifugation (20,000xg for 30 min at 4 °C). Protein concentration was determined by BCA assay, resulting in isolation of unfractionated proteomes of ~5-15 mg/mL. For long-term storage of proteomes for gel-based experiments, proteomic preparations are diluted with glycerol to a final concentration of 10% and stored -80 °C. Under these conditions proteomic proteins were found to retain activity over multiple (>10) freeze thaw cycles. For MudPIT analyses 1 mg aliquots of proteomes are stored at -80 °C without glycerol and thawed immediately prior to enrichment with probes **2**-**4**, as the presence of glycerol was found to severely impede downstream analysis.

Fluorescent SDS-PAGE analysis of probe labeling

Metabolic labeling by **1** was initially confirmed by SDS-PAGE and in-gel fluorescence scanning as reported previously.[15,19] Briefly, to a 30 μL reaction mixture of *B. subtilis* 6051 proteome (isolated from bacteria grown with **1**; 1 mg/mL) in sodium phosphate, pH 7.5 (50 mM) was added sequentially rhodamine-alkyne **7** (50 μM; 5 mM stock solution in DMSO), TCEP (1 mM; 100 mM stock in H_2O), tris-(benzyltriazolylmethyl)amine ligand (100 μM; 1.7 mM stock in DMSO:*t*-butanol 1:4), and $CuSO_4$ (1 mM; 50 mM stock in H_2O). Samples were vortexed and incubated at room temperature for 1 hr. Cycloaddition reactions were quenched by addition of EDTA (5 mM; 0.5 M stock in H_2O) and 5x SDS-loading buffer (strong reducing) and subjected to SDS-PAGE (22 μL per well). Excess probe fluorescence was removed by destaining in a solution of 50% MeOH / 40% H_2O / 10% AcOH overnight. Gels were then washed with water and fluorescently visualized using a Typhoon Laser Flatbed Scanner (GE Healthcare) with excitation at 532 nm. A characteristic intense low molecular weight fluorescence signal ~ 20 kDa (previously identified as the fatty acid acyl carrier protein AcpP) is indicative of probe labeling.

Chemoenzymatic labeling by **2** was initially confirmed by SDS-PAGE and in gel fluorescence scanning as reported previously.[16,17] Briefly, to a 30 μL reaction mixture of *B. subtilis* 6051 proteome (isolated from bacteria grown with **1**; 1 mg/mL) in sodium phosphate, pH 7.5 (50 mM) was added sequentially TAMRA-CoA **5** (10uM), $MgCl_2$ (40 mM), and Sfp. Samples were vortexed and incubated at room temperature for 1 hr. The

reaction was quenched with 5x SDS- loading buffer (strong reducing) and analyzed by SDS-PAGE with in-gel fluorescence scanning as above.

Activity-based labeling by fluorophosphonate **3** was initially confirmed by SDS-PAGE and in gel fluorescence scanning as reported previously.[17,35] Briefly, to a 30 μL reaction mixture of *B. subtilis* 6051 proteome (isolated from bacteria grown with **1**; 1 mg/mL) in Tris-Cl, pH 8.0 (50 mM) was added fluorophosphonate-TAMRA **6** (2 μM). Samples were vortexed and incubated at room temperature for 1 hr. The reaction was quenched with 5x SDS- loading buffer (strong reducing) and analyzed by SDS-PAGE with in-gel fluorescence scanning as above.

Metabolic labeling, click chemistry, and protein enrichment protocol for probe 1

Whole cell proteomes of *B. subtilis* 6051 grown in the presence of metabolic probe **1**were adjusted to a final protein concentration of 1 mg/mL and subjected to Cu(I)-catalzyed [3+2] cycloaddition with biotin alkyne **4** using a procedure analogous to previous reports.[15,20] Briefly, to a 1000 μL reaction mixture of *B. subtilis* 6051 proteome (1 mg/mL in 50 mM sodium phosphate, pH 7.5) was added sequentially SDS (0.1%, 10% stock in H_2O), sodium acetate (43 mM, 1 M stock in H_2O), biotin-alkyne **4** (29 μM; 56.5 mM stock solution in DMSO), TCEP (1 mM; 100 mM stock in water), tris-(benzyltriazolylmethyl)amine ligand (90 μM; 1.7 mM stock in DMSO:*t*-butanol 1:4), and $CuSO_4$ (0.9 mM; 50 mM stock in H_2O). Samples were vortexed and incubated at room temperature for 2 hr. Cycloaddition reactions were quenched by addition of EDTA (5 mM; 0.5 M stock in H_2O) and Triton-X added (1%, 10% stock in H_2O) and the reactions

rotated at 4 °C for 1 hour.[26] Reactions were loaded onto a pre-equilibrated PD10 Desalting column (GE Healthcare) to remove excess biotin alkyne **4**, collected, and denatured by addition of SDS to 0.5% and heating at 90 °C for 10 minutes. Samples were diluted to an SDS concentration of ~ 0.2% and allowed to cool to room temperature before addition of 50 μL pre-washed avidin-agarose, whereupon samples were rotated at 4 °C for 1 hour to facilitate avidin binding of biotinylated proteins. Avidin-agarose bound samples were then washed sequentially with 1% SDS, 6 M urea, and 50 mM Tris-Cl pH 8.0 (two washes each), and resuspended in 200 μL 8M urea, 50 mM Tris-Cl pH 8.0. Samples were then prepared for on-bead digest by reduction with 10 mM tris(2-carboxyethyl)phosphine (TCEP) and alkylation with 12 mM iodoacetamide. Samples were diluted to 2 M urea with 50 mM Tris-Cl pH 8.0 (400 μL total volume), followed by addition of trypsin and 2 mM $CaCl_2$. Digests were allowed to proceed overnight at 37°C overnight. After extraction, tryptic peptide samples were acidified to a final concentration of 5% formic acid and frozen at -80°C for MudPIT analysis.

Chemoenzymatic labeling and protein enrichment protocol for probe 2

Whole cell proteomes of *B. subtilis* 168 were adjusted to a final protein concentration of 2 mg/mL and labeled with chemoenzymatic CP probe **2** using a procedure analogous to previous reports.[16,17] Briefly, to a 500 μL reaction mixture of *B. subtilis* 168 proteome (2 mg/mL in 50 mM Tris-Cl, pH 8.0) was added biotin-CoA **2** (25 μM; 1 mM stock in H_2O), $MgCl_2$ (10 mM; 0.5 M stock in H_2O), and Sfp (8.8 μg). Samples were vortexed and incubated at room temperature for 1 hr, before addition of Triton-X and rotation at 4

°C for 1 hour. Samples were then desalted, denatured, avidin-affinity purified, and prepared for MudPIT as detailed in the protocol for metabolic probe **1** (above).[26]

Activity-based labeling and protein enrichment protocol for probe 3

Whole cell proteomes of *B. subtilis* 6051 grown in the presence of metabolic probe **1** were adjusted to a final protein concentration of 1 mg/mL and labeled with activity-based probe **3** using a procedure identical to previous reports.[22,26] Briefly, to a 1000 μL reaction mixture of *B. subtilis* 6051 proteome (1 mg/mL in 50 mM Tris-Cl, pH 8.0) was added fluorophosphonate-biotin **3** (5 μM; 1 mM stock in DMSO). Samples were vortexed and incubated at room temperature for 2 hr, before addition of Triton-X and rotation at 4 °C for 1 hour. Samples were then desalted, denatured, avidin-affinity purified, and prepared for MudPIT as detailed in the protocol for metabolic probe **1** (above). An identical procedure was employed for labeling of *B. subtilis* 168.

Liquid chromatography – mass spectrometry (MudPIT) analysis

Tryptic peptides enriched by each probe (**1**-**3**) were loaded onto a biphasic (strong cation exchange/reverse phase) capillary column and analyzed by 2D-LC separation in combination with tandem MS as previously described.[26] Peptides were eluted in a five-step MudPIT experiment and data were collected in an ion trap mass spectrometer (Finnigan LTQ) set in a data-dependent acquisition mode with dynamic exclusion turned on (60 s). Each full MS survey scan was followed by 7 MS/MS scans. Spray voltage was set to 2.75 kV and the flow rate through the column was 0.25 μL/min.

Analysis of MS Data

RAW files were generated from mass spectra using XCalibur version 1.4, and ms2 spectra data extracted using RAW Xtractor (version 1.9.1) which is publicly available (http://fields.scripps.edu /?q=content/download). Ms2 spectral data were searched using the SEQUEST algorithm[66] (Version 3.0) against a custom made database containing 72,177 sequences that were concatenated to a decoy database in which the sequences for each entry in the original database was reversed. In total the search database contained 144,354 protein sequence entries (72,177 real sequences and 72,177 decoy sequences). SEQUEST searches allowed for oxidation of methionine residues (16 Da), static modification of cysteine residues (57 Da-due to alkylation), no enzyme specificity, and a mass tolerance set to ±1.5 Da for precursor mass and ±0.5 Da for product ion masses. The resulting ms2 spectra matches were assembled and filtered using DTASelect2 (version 2.0.27). For this analysis, trypic, half-tryptic and fully-tryptic peptides were each individually evaluated using the DTASelect2 software. In each of these sub-groups the distribution of Xcorr and DeltaCN values for a direct (to the direct database) and decoy (reversed database) were separated by quadratic discriminant analysis. Outlier hits in the two distributions were removed. Spectral matches were retained with Xcorr and deltaCN values that produced a maximum peptide false positive rate of 1% which was derived from the frequency of matches to the decoy reverse database (number of decoy database hits/number of filtered peptides identified× 100).[67,68] This value is calculated by the DTASelect2 software. In addition, a minimum peptide length of seven amino acids residues was imposed and protein identification required the matching of at least two peptides per protein. Such criteria resulted in the elimination of most decoy database

hits. In our dataset the identification non-tryptic peptides included half-tryptic peptides from the N- and C- terminus of the identified proteins. Other non-tryptic peptides which were identified may represent endogenous activities of cellular proteases or peptides generated by in-source fragmentation.

Total spectral counts, protein sequences, and peptides were compiled into an Excel spreadsheet and manually sorted to filter proteins according to the designated spectral counting criteria and unique peptide criteria detailed in the text. Proteins identified at spectral counts greater than 15 and ratios greater than 10 in probe enriched versus non-enriched samples were designated functionally enriched probe targets. Manual comparison of these targets was used to identify proteins found in multiple datasets (Supplementary tables 9S 5-6). The Conserved Domains tool (NCBI) was used to estimate the percentage of *B. subtilis* proteins enriched by probe **3** which showed homology to serine hydrolases.

References

1. Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106**, 3468-96 (2006).

2. Sieber, S. A.; Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105**, 715-38 (2005).

3. Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640-2 (2000).

4. Weissman, K. J.; Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**, 925-36 (2005).

5. Gross, H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* **75**, 267-77 (2007).

6. Cropp, A.; Chen, S.; Liu, H.; Zhang, W.; Reynolds, K. A. Genetic approaches for controlling ratios of related polyketide products in fermentation processes. *J Ind Microbiol Biotechnol* **27**, 368-77 (2001).

7. Cisar, J. S.; Tan, D. S. Small molecule inhibition of microbial natural product biosynthesis-an emerging antibiotic strategy. *Chem Soc Rev* **37**, 1320-9 (2008).

8. Straight, P. D.; Fischbach, M. A.; Walsh, C. T.; Rudner, D. Z.; Kolter, R. A singular enzymatic megacomplex from Bacillus subtilis. *Proc Natl Acad Sci U S A* **104**, 305-10 (2007).

9. Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* **97**, 9390-5 (2000).

10. Mercer, A. C.; Burkart, M. D. The ubiquitous carrier protein--a window to metabolite biosynthesis. *Nat Prod Rep* **24**, 750-73 (2007).

11. Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **3**, 923-36 (1996).

12. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7 (2001).

13. Flugel, R. S.; Hwangbo, Y.; Lambalot, R. H.; Cronan, J. E., Jr.; Walsh, C. T. Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in Escherichia coli. *J Biol Chem* **275**, 959-68 (2000).

14. Kopp, F.; Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat Prod Rep* **24**, 735-49 (2007).

15. Mercer, A. C.; Meier, J. L.; Torpey, J. W.; Burkart, M. D. In vivo modification of native carrier protein domains. *Chembiochem* **10**, 1091-100 (2009).

16. La Clair, J. J.; Foley, T. L.; Schegg, T. R.; Regan, C. M.; Burkart, M. D. Manipulation of carrier proteins in antibiotic biosynthesis. *Chem Biol* **11**, 195-201 (2004).

17. Meier, J. L.; Mercer, A. C.; Burkart, M. D. Fluorescent profiling of modular biosynthetic enzymes by complementary metabolic and activity based probes. *J Am Chem Soc* **130**, 5443-5 (2008).

18. Meier, J. L.; Burkart, M. D. Chapter 9. Synthetic probes for polyketide and nonribosomal peptide biosynthetic enzymes. *Methods Enzymol* **458**, 219-54 (2009).

19. Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J Am Chem Soc* **128**, 12174-84 (2006).

20. Speers, A. E.; Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol* **11**, 535-46 (2004).

21. Liu, Y.; Patricelli, M. P.; Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* **96**, 14694-9 (1999).

22. Kidd, D.; Liu, Y.; Cravatt, B. F. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005-15 (2001).

23. Evans, M. J.; Cravatt, B. F. Mechanism-based profiling of enzyme families. *Chem Rev* **106**, 3279-301 (2006).

24. Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683-90 (2001).

25. Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-201 (2004).

26. Jessani, N.; Niessen, S.; Wei, B. Q.; Nicolau, M.; Humphrey, M.; Ji, Y.; Han, W.; Noh, D. Y.; Yates, J. R., 3rd; Jeffrey, S. S.; Cravatt, B. F. A streamlined platform for high-content functional proteomics of primary human specimens. *Nat Methods* **2**, 691-7 (2005).

27. Stein, T. Bacillus subtilis antibiotics: structures, syntheses and specific functions. *Mol Microbiol* **56**, 845-57 (2005).

28. Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. *Proc Natl Acad Sci U S A* **104**, 1506-9 (2007).

29. Zeigler, D. R.; Pragai, Z.; Rodriguez, S.; Chevreux, B.; Muffler, A.; Albert, T.; Bai, R.; Wyss, M.; Perkins, J. B. The origins of 168, W23, and other Bacillus subtilis legacy strains. *J Bacteriol* **190**, 6983-95 (2008).

30. Lemon, K. P.; Earl, A. M.; Vlamakis, H. C.; Aguilar, C.; Kolter, R. Biofilm development with an emphasis on Bacillus subtilis. *Curr Top Microbiol Immunol* **322**, 1-16 (2008).

31. Lopez, D.; Vlamakis, H.; Kolter, R. Generation of multiple cell types in Bacillus subtilis. *FEMS Microbiol Rev* **33**, 152-63 (2009).

32. Nakano, M. M.; Marahiel, M. A.; Zuber, P. Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in Bacillus subtilis. *J Bacteriol* **170**, 5662-8 (1988).

33. Wolff, S.; Antelmann, H.; Albrecht, D.; Becher, D.; Bernhardt, J.; Bron, S.; Buttner, K.; van Dijl, J. M.; Eymann, C.; Otto, A.; Tam le, T.; Hecker, M. Towards the entire proteome of the model bacterium Bacillus subtilis by gel-based and gel-free approaches. *J Chromatogr B Analyt Technol Biomed Life Sci* **849**, 129-40 (2007).

34. Barr, J. G. Changes in the extracellular accumulation of antibiotics during growth and sporulation of Bacillus subtilis in liquid culture. *J Appl Bacteriol* **39**, 1-13 (1975).

35. Patricelli, M. P.; Giang, D. K.; Stamp, L. M.; Burbaum, J. J. Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. *Proteomics* **1**, 1067-71 (2001).

36. 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* **37**, 1585-95 (1998).

37. Perego, M.; Glaser, P.; Minutello, A.; Strauch, M. A.; Leopold, K.; Fischer, W. Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in Bacillus subtilis. Identification of genes and regulation. *J Biol Chem* **270**, 15598-606 (1995).

38. Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci U S A* **99**, 14083-8 (2002).

39. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **4**, 1487-502 (2005).

40. Nesvizhskii, A. I.; Vitek, O.; Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787-97 (2007).

41. Kocher, T.; Superti-Furga, G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* **4**, 807-15 (2007).

42. Hwang, Y.; Thompson, P. R.; Wang, L.; Jiang, L.; Kelleher, N. L.; Cole, P. A. A selective chemical probe for coenzyme A-requiring enzymes. *Angew Chem Int Ed Engl* **46**, 7621-4 (2007).

43. Yin, J.; Straight, P. D.; Hrvatin, S.; Dorrestein, P. C.; Bumpus, S. B.; Jao, C.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genome-wide high-throughput mining of natural-product biosynthetic gene clusters by phage display. *Chem Biol* **14**, 303-12 (2007).

44. Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A* **102**, 15815-20 (2005).

45. Sheehan, S. M.; Switzer, R. L. Intracellular serine protease 1 of Bacillus subtilis is formed in vivo as an unprocessed, active protease in stationary cells. *J Bacteriol* **172**, 473-6 (1990).

46. Lee, A. Y.; Goo Park, S.; Kho, C. W.; Young Park, S.; Cho, S.; Lee, S. C.; Lee, D. H.; Myung, P. K.; Park, B. C. Identification of the degradome of Isp-1, a major intracellular serine protease of Bacillus subtilis, by two-dimensional gel electrophoresis and matrix- assisted laser desorption/ionization-time of flight analysis. *Proteomics* **4**, 3437-45 (2004).

47. Sloma, A.; Rufo, G. A., Jr.; Rudolph, C. F.; Sullivan, B. J.; Theriault, K. A.; Pero, J. Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning of the gene. *J Bacteriol* **172**, 5520-1 (1990).

48. Roitsch, C. A.; Hageman, J. H. Bacillopeptidase F: two forms of a glycoprotein serine protease from Bacillus subtilis 168. *J Bacteriol* **155**, 145-52 (1983).

49. Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A. M.; Alloni, G.; Azevedo, V.; Bertero, M. G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Borriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S. C.; Bron, S.; Brouillet, S.; Bruschi, C. V.; Caldwell, B.; Capuano, V.; Carter, N. M.; Choi, S. K.; Codani, J. J.; Connerton, I. F.; Danchin, A.; et al. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**, 249-56 (1997).

50. Shevchenko, A.; Sunyaev, S.; Loboda, A.; Shevchenko, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* **73**, 1917-26 (2001).

51. Ducret, A.; Van Oostveen, I.; Eng, J. K.; Yates, J. R., 3rd; Aebersold, R. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci* **7**, 706-19 (1998).

52. Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A. Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol* **23**, 1562-7 (2005).

53. Frank, A.; Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **77**, 964-73 (2005).

54. Dimroth, P.; Walter, H.; Lynen, F. [Biosynthesis of 6-methylsalicylic acid]. *Eur J Biochem* **13**, 98-110 (1970).

55. Roskoski, R., Jr.; Gevers, W.; Kleinkauf, H.; Lipmann, F. Tyrocidine biosynthesis by three complementary fractions from Bacillus brevis (ATCC 8185). *Biochemistry* **9**, 4839-45 (1970).

56. Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem Rev* **97**, 2465-2498 (1997).

57. Buttner, K.; Bernhardt, J.; Scharf, C.; Schmid, R.; Mader, U.; Eymann, C.; Antelmann, H.; Volker, A.; Volker, U.; Hecker, M. A comprehensive two-dimensional map of cytosolic proteins of Bacillus subtilis. *Electrophoresis* **22**, 2908-35 (2001).

58. Wolff, S.; Otto, A.; Albrecht, D.; Zeng, J. S.; Buttner, K.; Gluckmann, M.; Hecker, M.; Becher, D. Gel-free and gel-based proteomics in Bacillus subtilis: a comparative study. *Mol Cell Proteomics* **5**, 1183-92 (2006).

59. Donato, H.; Krupenko, N. I.; Tsybovsky, Y.; Krupenko, S. A. 10-formyltetrahydrofolate dehydrogenase requires a 4'-phosphopantetheine prosthetic group for catalysis. *J Biol Chem* **282**, 34159-66 (2007).

60. Tamehiro, N.; Okamoto-Hosoya, Y.; Okamoto, S.; Ubukata, M.; Hamada, M.; Naganawa, H.; Ochi, K. Bacilysocin, a novel phospholipid antibiotic produced by Bacillus subtilis 168. *Antimicrob Agents Chemother* **46**, 315-20 (2002).

61. Hamoen, L. W.; Venema, G.; Kuipers, O. P. Controlling competence in Bacillus subtilis: shared use of regulators. *Microbiology* **149**, 9-17 (2003).

62. Kunst, F.; Msadek, T.; Bignon, J.; Rapoport, G. The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in Bacillus subtilis. *Res Microbiol* **145**, 393-402 (1994).

63. Ogura, M.; Yamaguchi, H.; Yoshida, K.; Fujita, Y.; Tanaka, T. DNA microarray analysis of Bacillus subtilis DegU, ComA and PhoP regulons: an approach to comprehensive analysis of B.subtilis two-component regulatory systems. *Nucleic Acids Res* **29**, 3804-13 (2001).

64. Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-14 (2007).

65. Candiano, G.; Bruschi, M.; Musante, L.; Santucci, L.; Ghiggeri, G. M.; Carnemolla, B.; Orecchia, P.; Zardi, L.; Righetti, P. G. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327-33 (2004).

66. Eng, J. K.; McCormack, A. L.; Yates, J. R., 3rd. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).

67. Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**, 21-6 (2002).

68. Cociorva, D.; D, L. T.; Yates, J. R. Validation of tandem mass spectrometry database search results using DTASelect. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 4 (2007).

Acknowledgments

The text of chapter 9, in part, is in preparation for publication. I am the primary author of the manuscript. Sherry Niessen and Heather Hoover performed MudPIT analyses and aided in aided in protein enrichment and data analyses. Professor Benjamin Cravatt supervised mass spectrometry experiments and provided fluorophosphonate-biotin. Tim Foley provided biotin-CoA. I designed and performed the experiments. All research was performed under the supervision of Professor Mike Burkart.