Title
Active site mapping of iterative polyketide synthases and the detection of polyketide intermediates using high- resolution Fourier Transform mass spectrometry

Permalink
https://escholarship.org/uc/item/86317072

Author
Meehan, Michael Joseph

Publication Date
2009

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Active site mapping of iterative polyketide synthases and the detection of polyketide
intermediates using high-resolution Fourier Transform mass spectrometry

A Thesis submitted in partial satisfaction of the requirements for
the degree Master of Science

in

Chemistry

by

Michael Joseph Meehan

Committee in charge:

    Professor Pieter Dorrestein, Chair
    Professor Vineet Bafna
    Professor Michael Burkart

2009

The Thesis of Michael Joseph Meehan is approved, and it is acceptable

in quality and form for publication on microfilm and electronically:

_____

_____

_____
<div align="right">Chair</div>

University of California, San Diego

2009

DEDICATION

For all of her patience and support; for all of her constant encouragement throughout

my 2 years as a student at UCSD, for her unconditional love, and for making my life a

happier one, I dedicate my work to my wonderful wife, Ann.

We have both worked very hard as we have progressed through our studies, in pursuit

of our goals. Despite the challenges and obstacles we have faced, we have never quit

and we will continue to give each other the help and support necessary to succeed.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3,4-DHBA	3,4-Dihydroxybenzoic acid
μg	Microgram
μl/min	Microliter per minute
μM	micro Molar
A (domain)	Adenylation
Acac	Acetoacetyl
ACP	Acyl Carrier Protein
AL	Acyl Ligase
AMP	Adenosine Monophosphate
APS	5'-Adenylylsulfate
Asn	Asparagine
AT	Acyltransferase (Domain)
A-T (didomain)	Adenylation-Thiolation (didomain)
Atd	Actyltranseferase Docking Site
ATP	Adenosine Triphosphate
ATP-PPi	Adeonsine Triphosphate-Pyrophosphate Exchange
BIRD	Blackbody Infrared Radiative Dissociation
C (domain)	Condensation (domain)
CD	Circular Dichroism
CDA	Calcium-Dependent Antibiotic
CID	Collision-Induced Circulat Dichroism
CLC	Claisen-Like Cyclase
CLF	Chain Length Factor
CoA	Coenzyme A
Cys	Cysteine
Da	Dalton
DH (domain)	Dehydratase (domain)
DXP	deoxy-D-xylulose-phosphate
E (domain)	Epimerization (domain)
ER (domain)	Enoyl Reductase (domain)
ESI	Electro Spray Ionization
ESI-FT-ICR-MS	Electro Spray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
ESI-FTMS	Electro Spray Ionization Fourier Transform Mass Spectrometry
ESI-MS	Electro Spray Ionization Mass Spectrometry
ESI-MS	Electro Spray Ionization Mass Spectrometry

ESI-Q-TOF-MS	Electro Spray Ionization Quadrupole Time-Of-Flight Mass Spectrometry
ETD	Electron Transfer Dissociation
FAS	Fatty Acid Synthase
FID	Free Induction Decay
FT-ICR-MS	Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
FTMS / FT-MS	Fourier Transform Mass Spectrometry
GC/EI-MS	Gas Chromatography/Electron Impact Mass Spectrometry
HCS	HMG-CoA Synthase
HIC	Hydroxyl Isocaproate
His	Histadine
HMG	3-Hydroxy-3-Methylglutaryl
HPLC	High Performance Liquid Chromatography
IPKS	Interative Polyketide Synthase
IR	Infrared
IRMPD	Infrared Multiphoton Dissociation
kDa	Kilo-Dalton
KIC	Ketoisocaproate
KR (domain)	Ketoreductase (domain)
KS (domain)	Ketosynthase (domain)
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass Spectrometry
LC-MS-PEA	Liquid Chromatography Mass Spectrometry Phosphopantetheinyl Ejection Assay
LCQ	Three-Dimensional Linear Quadrupole
Linear IT	Linear Ion Trap
LTQ	Two-Dimensional Linear Trap Quadrupole
LTQ-FT-ICR-MS	Hybrid Linear Trap Quadrupole - Fourier Transform Ion Cyclotron Resonance Mass Spectrometry
LTQ-ORBI	Hybrid Linear Trap Quadrupole - Orbitrap
Lys	Lysine
m/z	Mass / Charge
MALDI	Matrix-Assisted Laser Desporption/Ionization
MALD-TOF	Matrix-Assisted Laser Desporption/Ionization - Time-Of-Flight
MAT (domain)	Malonyl-CoA:Acyl Carrier Protein Acyltransferase

MDa	Mega Dalton
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MS2	Tandem Mass Spectrometry (or MS/MS)
MS3	Additional Fragmentation of Ions generated by MS/MS
NAD(P)H	Nicotinamide Adenine Dinucleotide (Phospate) Hydride
NAD+	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide Hydride
NADPH	Nicotinamide Adenine Dinucleotide Phospate Hydride
nl/min	Nanoliters per Minute
NMR	Nuclear Magnetic Resonance
NRP	Non-Ribosomal Peptide
NRPS	Non-Ribosomal Peptide Synthetase
NRPS/PKS	Hybrid Non-Ribosomal Peptide Synthetase / Polyketide Synthase
PEA	Phosphopantetheinyl Ejection Assay
Phe	Phenylalanine
PKS	Polyketide Synthase
PPant	Phosphopantetheine
ppm	Parts Per Million
PPTase	Phosphopantetheinyl Transferase
PQD	Pulsed-Q Dissociation
Q-TOF	Quadrupole Time-Of-Flight
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
RPLC	Reverse Phase Liquid Chromatography
-S-	Bonding Sulfur of Thioester
SAT (domain)	Starter-CoA:Acyl Carrier Protein Acyltransferase
SerT	Seryltransferase
SORI-CAD	Sustained Off-Resonance Irradiation-Collisionally Activated Dissociation
T (domain)	Thiolation (domain)
T (unit)	Tesla
TE (domain)	Thioesterase
TE/CLC (domain)	Thioesterase/Claisen-Like Cyclase
TGH	Transglutaminase Homologue
Tyr	Tyrosine
UV-Vis	Ultraviolet-Visible Spectroscopy
X-Ray	X-Ray Diffraction

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

ABSTRACT OF THE THESIS

Active site mapping of iterative polyketide synthases and the detection of polyketide
intermediates using high-resolution Fourier Transform mass spectrometry

by

Michael Joseph Meehan

Master of Science in Chemistry

University of Califoria, San Diego 2009

Professor Pieter Dorrestein, Chair

In an attempt to better understand the biosynthesis of the enediyne core of the
antitumor compound C-1027 by SgcE and to investigated the iterative steps involved
in the formation of the methylbutyrl side-chain of lovastatin by LovF, high resolution
mass spectrometry was used to identify peptides containing domain active sites of
these polyketide synthases and to determine with very high mass accuracy the mass of
substrates covalently attached to the active site. In particular the phosphopantetheinyl

ejection assay was used to successfully detect the formation of the methylbutyrate by

LovF on the ACP active site. It was found that methylbutyrate is detectable only at

very low substrate concentrations, but complete formation of the product was carried

out by LovF. SgcE was found to be active and capable of catalyzing elongation of a

polyene intermediate. The substrate loading and one iteration of elongation were

detected using mass spectrometry and the phosphopantetheinyl ejection assay. The

results from these experiments showed the power of high-resolution mass

spectrometry in studying the mechanisms of action by which these iterative polyketide

synthases function and form their fully elongated products.

Introduction

The current emergence of drug resistance and higher incidence of diagnosed illnesses, such as cancer, Alzheimer's disease and diabetes, coupled with an increasing world population has resulted in an increased interest in the study of natural products and their biosynthetic machineries. More than 50% of all therapeutics have origins in natural productswith many more currently in clinical trials.[1] The recognition that many organisms have the metabolic capacity to produce a large number of natural products is leading to increased availability of sequenced genomes, which is further resulting in the betterment of the instruments used to study them. There are many tools to characterize natural products and their biosynthetic machinery including NMR, HPLC, UV-Vis, IR, CD, and X-ray. Standing out among these common methods is mass spectrometry. Mass spectrometry has become so essential in the study of these systems that there is not a scientific journal in the world that would accept the characterization of a new natural product in the absence of mass spectrometry data. Nevertheless, unlike NMR where new methods for the characterization of molecules are frequently developed, the use of creative mass spectrometry has been rather limited within the general biosynthetic natural product community. Mass spectrometry is usually done as an afterthought; once one has already obtained activities by other means. This, however, is changing and mass spectrometry is moving to the forefront of many investigations. The reason for this paradigm shift is the ever changing landscape of modern mass spectrometry tools.

The past few years have seen a substantial increase in the capacity of commercial mass spectrometry instruments. These changes are partially driven by the clinical "omics" community, but are also found to be very useful in other areas of

science. Unlike in the past where promising developments in mass spectrometry would take a decade or longer to reach the general public, new developments in mass spectrometry are being quickly commercialized on user friendly instruments. For example, the tandem mass spectrometry method of electron transfer dissociation was originally published in 2004 and the first commercial version came out in 2005.[2] Electron transfer dissociation (ETD) and its, by reactivity only, related cousin electron capture dissociation[3] represent just two of the many recent advances in mass spectrometry. The past 5 years alone have bared witness to an explosion of advances in resolution, sources acquisition speeds, data processing and ionization sources.[4-19] The current rate of development of mass spectrometry tools indicate that this chapter too will have aged by the time this review is published. Therefore, this review will not only serve as a snapshot of widely used mass spectrometry approaches in the biosynthetic investigations of natural products, but also aims to provide a glimpse into the short and long term future capacities of mass spectrometry in the field.

The emphasis of this chapter is placed on two structural classes of natural products: polyketides and non-ribosomal peptides. The mass spectrometry of these biosynthetic pathways is most advanced and will be covered in detail. In the following pages we will describe the current methods and applications used to study the biosynthetic pathways of natural products and provide a glimpse into upcoming techniques. In addition, a brief introduction to experimental design using high-end mass spectrometry to study the biosynthesis of other natural metabolites, such as ribosomally encoded pathways and co-factors, is described.

Chapter 1: Analysis of PKS and NRPS by Mass Spectrometry

1.1. Introduction in NRPS and PKS Biosynthesis

Many important therapeutics, in use in clinics today, are biosynthesized by the non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) paradigm. For example many of the antibiotics (penicillin, cephalosporin, vancomycin, erythromycin etc.), immunosuppressors (cyclosporine, rapamycin), anti-viral agents (luzopeptin A), anti-tumor agents (bleomycin), and toxins (thaxtomin) are NRPS and PKS derived.[20-22] Figure 1 displays a small selection of natural products that are NRPS and PKS derived and illustrates the diversity of molecular structures generated by these biosynthetic paradigms.
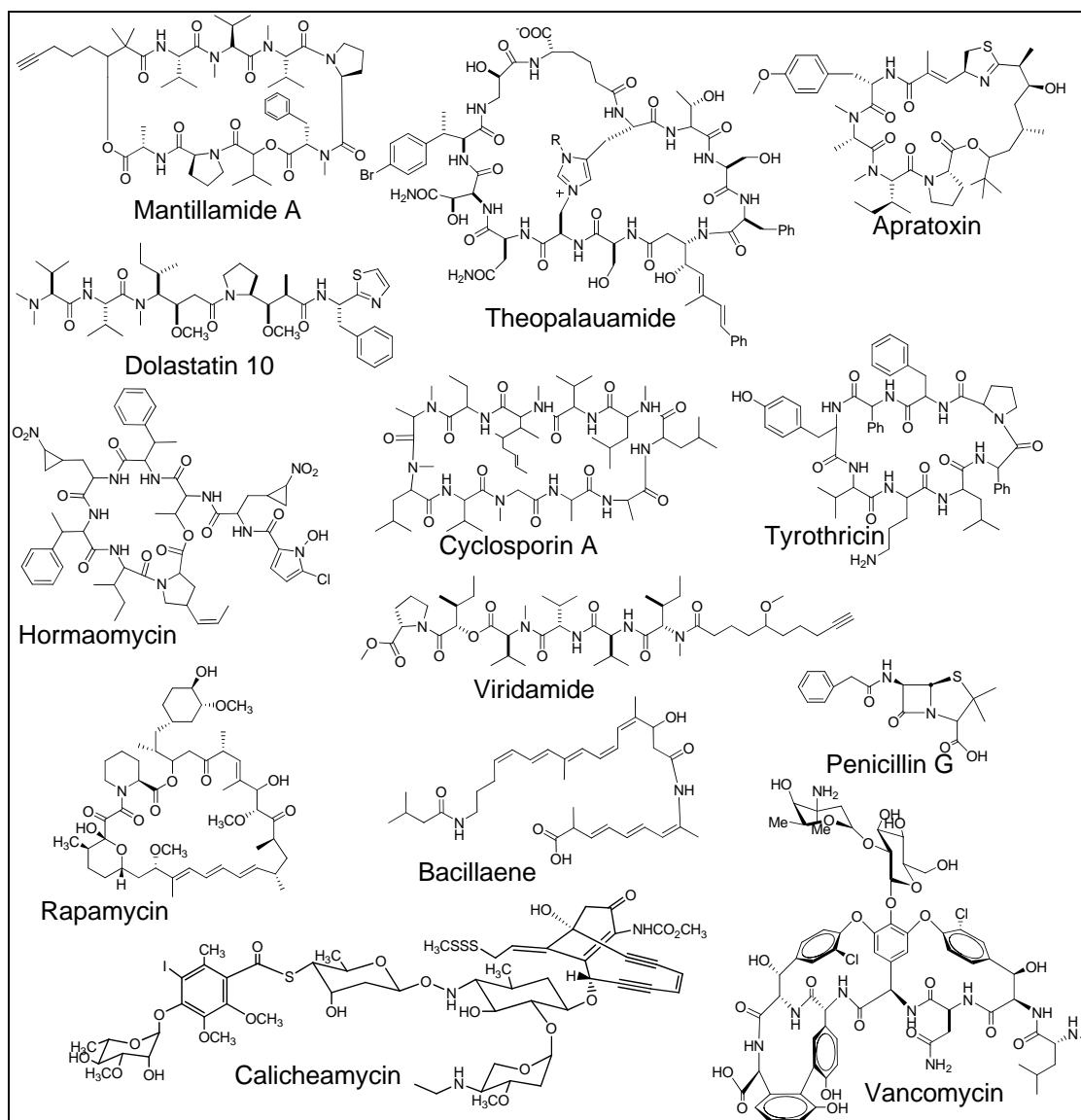
Figure 1: **Select PKS and NRPS natural products.** Some examples showing the wide structural diversity of NRPS and PKS derived natural products.

NRPS biosynthesis differs substantially from ribosomally encoded peptides. In the NRPS biosynthetic strategy, the substrates and intermediates are covalently linked to an active site serine on the thiolation (T) domain of a non-ribosomal peptide synthethase *via* the phosphopantetheinyl arm.[23] This post-translational event is accomplished by a phosphopantetheinyl transferase (PPTase), which primes the active site serine with a portion of coenzyme A (CoA) to generate the holo form of the thiolation domain (Figure 2B).[24] In traditional NRPS biosynthesis, amino acids are activated by ATP. Activation is achieved by the formation of an adenylated carboxylate of an amino acid that is encoded by the first adenylation (A) domain (Figure 2A).[25] The activated amino acid then undergoes nucleophilic displacement by the thiol terminus of the phosphopantetheine arm tethered to the first thiolation domain. During the elongation step, the resulting amino acyl-enzyme is then condensed with a second acyl-enzyme species on a downstream carrier domain to form a linear peptide, in a reaction catalyzed by a condensation domain. Further modifications often take place while the amino acid or the growing peptide is still attached to a carrier domain (*e.g.* oxidation, chlorination, or other types of modifications).[26] Finally, the mature peptide is released from the last carrier domain via cyclization or removed by a termination domain.[27] The most common off-loading domain is a thioesterase,[28] although other termination domains exist. While there are currently 22 known ribosomally encoded amino acids, there are hundreds of non-ribosomally encoded amino acids expanding the array of structures (and therefore pharmacologically active compounds) that can be generated via the NRPS paradigm.[29]
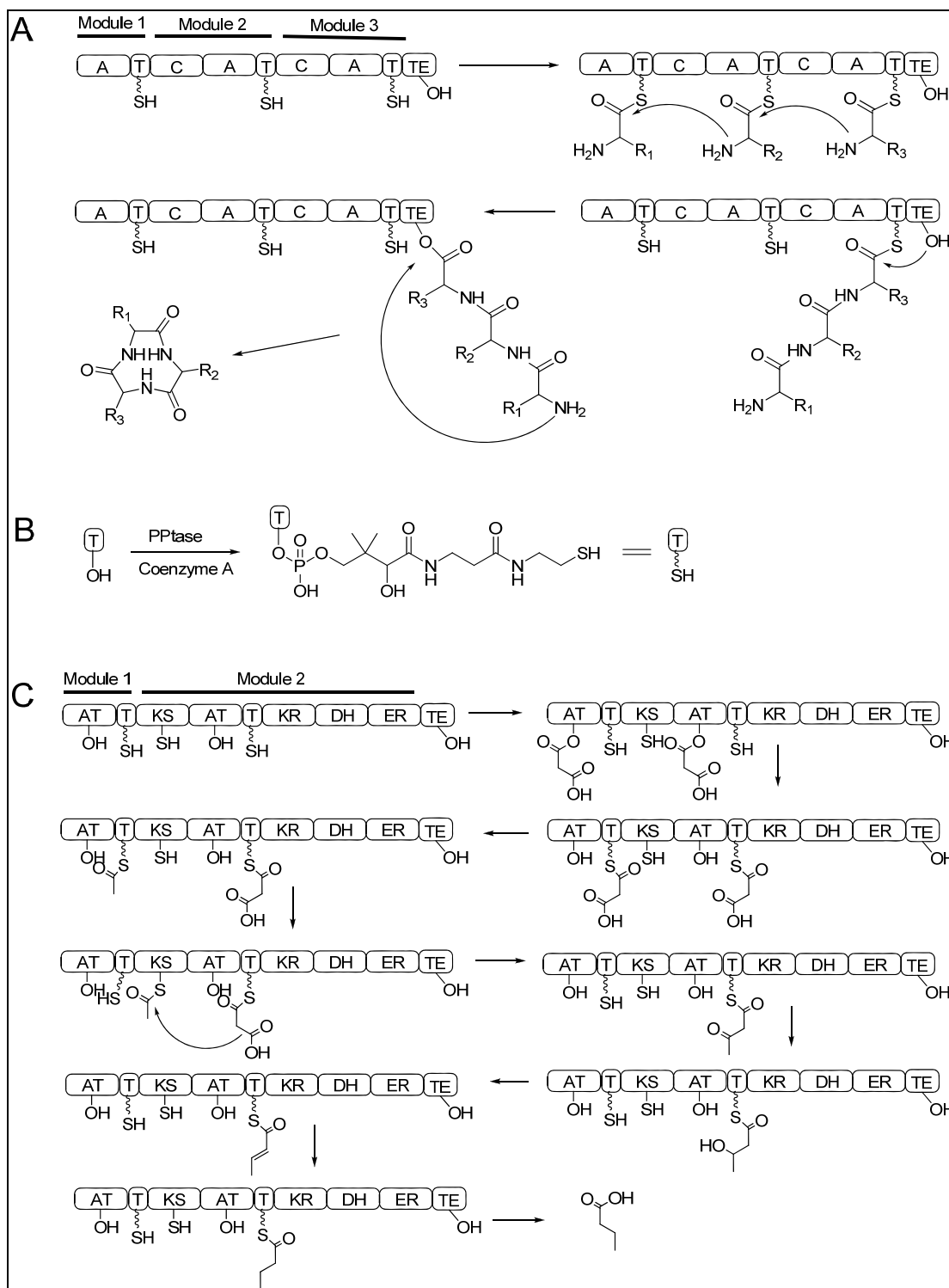
Figure 2: **Overview of NRPS and PKS biosynthesis.** (A) NRPS biosynthesis.
(B) Priming of *apo*-T domain by PPTase. (C) General type I PKS biosynthesis
paradigm.

Just like NRPS biosynthesis, polyketide biosynthesis often takes place on multidomain megasynthases.[23] Polyketide synthases also carry their substrates and intermediates on T domains, also referred to as acyl carrier proteins (ACP). The substrates in PKS biosynthesis are not amino acids or free carboxylic acids as it is observed with NRPS systems, but rather they are coenzyme A activated carboxylic acids. The most common substrate for polyketide biosynthesis is malonyl-CoA. In order to divert a small subfraction of the cellular pool of malonyl-CoA PKSs activate the malonyl by a nucleophilic attack of the active site serine from the acyl transferase (AT) domain forming a covalent oxo-ester linkage (Figure 2C). Subsequently, a transacylation takes place onto the T domain linking the malonyl to the thiol of the phosphopantetheinyl functionality via a thioester bond. While the malonyl is still attached to the T domain, it is decarboxylated by the ketosynthase (KS) domain before a second transacylation takes place to the active site cysteine of the KS domain generating an acetyl-*S*-KS intermediate, something that to date is very hard to capture directly by mass spectrometry. At this point a second malonyl group is loaded onto the upstream T domain, a reaction again catalyzed by an AT domain. As the KS domain decarboxylates the second malonyl, a Claisen condensation takes place with the acetyl group on the KS domain forming acetoacetate, the first unit in an elongating polyketide. Like NRPS biosynthetic pathways, PKS biosynthesis can incorporate a variety of tailoring reactions.[23] Tailoring reactions such as chlorination, oxidation, cyclization may be observed.[30-32] By far the most common tailoring steps in PKS biosynthesis are ketoreduction, dehydration and enoyl-reductions. Ketoreduction is catalyzed by the ketoreductase (KR) domain and uses the co-factor NAD(P)H to

convert the β-carbonyl to an alcohol. This same alcohol can be dehydrated by a dehydratase (DH) domain. The DH domain catalyses the deprotonation of the α-proton and protonates the leaving hydroxyl group in an α,β-unsaturated thioester. This thioester can then be further reduced to a completely saturated bond by the enoylreductase (ER). This reaction utilizes the co-factor NAD(P)H as required in the ketoreduction. While there are many natural products that are either NRPS or PKS derived, there are multidomain megasynthases that contain both NRPS and PKS biosynthetic features. Examples of such biosynthetic motifs are epothilone, jamaicamide, and the endiyne C-1027.[33-35] At NRPS-PKS interfaces in these systems, the condensation of malonyl takes place with an amino acid that was activated by an A domain. Alternatively, an amino acid loaded on a T domain condenses with an elongating polyketide. Since the substrates and intermediates in NRPS, PKS or hybrid NRPS/PKS biosynthesis introduce mass changes on T domains they are ideal candidates for investigation by mass spectrometry.

1.2. New Tools in the Characterization of Multidomain and Phosphopantetheinylated Proteins

For many years, it was very difficult to study multidomain, post-translationally, or transiently modified proteins. These studies required multi-year efforts by mass spectrometry experts. But as instrumentation improved it has enabled additional investigators to carry out this type of research. Many proteins, in particular NRPS and PKS, are of substantial size. Some of the largest single open reading frames are responsible for the generation of MDa polypeptides. An example of this are the 2.5

MDa NRPSs of the syringomycin pathway and the megasynthase of the biosynthetic pathway of a 18-mer peptaibol.[36,37] In the majority of cases, it is not possible to look at intact proteins of this size, so they must be digested before the active sites can be analyzed in terms of tethered biosynthetic substrates and intermediates. However, when these proteins are digested, the complexity increases and the active sites have to be found within a haystack of data. Many modern proteomic programs such as InSpecT, Sequest, OMSSA, Spectumill, Mascott, MassLynx, mainly to find these active site peptides or domains of post-translationally modified megasynthases from complete or limited digests.[38,39] There are three main reasons for this: 1) Despite the surge in the development of proteomic platforms for the analysis of digested proteins, all proteomic programs have a difficult time identifying ions based on the fragmentation data of parent ion that have charges that are greater than 4+. Most of these programs, by default, ignore 4+ ions or larger as this would increase the number of false positives. We estimate that ~30% of all the data in an LC-MS of a complete tryptic digest would belong to these higher charges. In addition because of the poorer nature of 1+ ion fragmentation, these too are often not identified. 2) While many of these programs can identify non-labile modifications, they have to be "trained" for labile modifications should it be possible to find them at all. For example, most of these programs have a specific scoring function for the analysis of phosphopeptides and none of these programs have a scoring function for labile modificiations such as the phosphopantetheinylation found on fatty acid synthase (FAS), NRPS and PKS multidomain proteins. 3) For a given LC-MS/MS run for a digest we typically obtain 10,000-20,000 MS/MS spectra, yet on a good run only 2,000-3,000 MS/MS spectra

are annotated. This begs the question what the remaining spectra are and clearly indicate that we are not yet identifying everything we can from the data. Thus there is an enormous opportunity for the development of programs that can capture the remaining uncharacterized spectra, including 4+ or greater charge ions or ions with different labile modifications.

Even though automated programs to facilitate the analysis of these large multidomain proteins involved in the biosynthesis of natural products do not exist, it has become possible to investigate such systems by mass spectrometry. After a short protease digestion, most of the active sites are often 5-20 kDa. It is not yet reasonable to map the active sites by tandem mass spectrometry data as it is done in proteomic investigations.[40] However, with the emergence of high-resolution instruments that have routine mass accuracies within 10 ppm, it became possible to find the active sites by the intact mass of the peptides alone giving rise to a manageable number of false positive matches. Any false match is eliminated by tandem mass spectrometry. When one performs tandem mass spectrometry, many fragment ions are detected that should match up to expected fragments for the active sites. The false positive active sites matches are eliminated on the basis that their fragments will not match-up to the sequence of the active site, while all of the true positive candidate fragments match quite well. In 1999, Kelleher was the first to recognize the usefulness of Fourier transform-ion cyclotron resonance-mass spectrometry (FT-ICR-MS) resolution and mass accuracy for the investigations of multi-domain NRPS proteins.[41] At that time, FT-ICR-MS instruments were the only instruments with this degree of mass accuracy and the best of these instruments were custom-built and not readily available to the

general public. As instrumentation and the development of efficient strategies to analyze multidomain NRPS/PKS proteins are advancing, more and more laboratories can carry out these type of investigations, including laboratories that only have access to low resolution instruments.[42-46] In the following sections we describe the current and new mass spectrometry approaches to analyze some of these biosynthetic pathways since a 2006 review on this topic.[47]

There are several key advances that have been in use since 2006 for the investigations of NRPS and PKS proteins: 1) The use of larger sized magnets (e.g. 12T) enables active site mapping on an LC-timescale.[48,49] 2) It was discovered that the labile post-translational modification of thiolation domains can be ejected during thermal activation methods and that this ejection can be used to "observe" substrates and intermediates on the active site thiol of a phosphopantetheinyl functionality. This assay, called the phosphopantetheinyl ejection assay (PEA), has found many uses and has been adopted by several other laboratories already.[48-58] 3) The phosphopantetheinyl ejection assay can also be applied to low resolution instruments making this assay accessible to many other researchers that work with phosphopantetheinylated proteins.[43,44] Finally, 4) There are now instances describing other methods for analyzing the phosphopantetheinylated proteins, such as the use of a phosphatase to remove the phopshopantetheinyl functionality so that it may be characterized by mass spectrometry. All of these new mass spectrometry capabilities are highlighted in the next sections.

1.3. A Brief Introduction to FT-ICR-MS

The application of FT-ICR MS analysis to the characterization of NRPS and PKS proteins is critical as this type of analysis has accelerated our understanding of these complex systems over the past decade. It is important to introduce some of the fundamentals of FT-ICR MS in order to understand the advantage of this technique in NRPS/PKS research. Please be aware that this is not meant as an authorative review on the subject but as a simplified introduction to the complexities of FT-ICR MS for a reader that has never been exposed to this type of instrumentation. For a deeper insight, there are several very good FT-ICR-MS reviews on the subject that should be consulted.[59-61] FT-ICR mass spectrometry is an image current based detection strategy.[59-61] This means that ions are detected by a perturbation of a current that is captured as a frequency. To accomplish this, charged ions are introduced into a detection cell about the size of a soda can. Inside of this soda can-sized cell there are excitation plates on which an alternating current is applied to generate a cyclotron motion to the charged ions. The basis of FT-ICR MS is this ion cyclotron motion, which arises from the interaction of an ion with a unidirectional magnetic field. The ion experiences a force, the Lorentz force, which causes this ion to travel in a circular orbit perpendicular to this magnetic field. The radius of the orbital motion is defined by the magnetic field strength. While the charged ion is in an orbital motion it passes detector plates where the ion perturbs the current. This current can be very accurately measured as a time domain. The resulting time domain (FID) can be subjected to Fourier-transform to obtain the frequency of the ions undergoing the orbital motion. This frequency can then be converted to a mass measurement with equation 1.

Equation 1: $f_{cyc} = zB/2\pi m$

where f_{cyc} is the cyclotron frequency, z is the charge of the ion, B is the magnetic field strength and m is the ion mass. The magnetic field of the spectrometer is held constant, provided by an ultra-highly stable superconducting magnet, and the mass-to-charge ratio of the ion (m/z) is determined by measuring its f_{cyc}. Since frequency measurement is inherently accurate and can be measured more accurately than other physical properties[59], the FTICR mass spectrometer offers superb mass resolution and mass accuracy. Equation 1 also shows that by increasing the magnet field strength, the resolving power and scan speed increase in a linear fashion.[59] The ability to collect a good FID is important for high resolution. Truncating the FID, results in an increased scan rate but also in a loss of fine information. This is one of the reasons why FT-ICR-MS instruments needs very high vacuum which is $10E^{-10}$ Torr inside the analyzer cell. Without high vacuum, the ions collide with other ions from air or gas and the FID beads out, essentially truncating the FID. The effect of truncation on the resulting resolution is shown in Figure 3. Many commercial instruments may not use the wording "truncation of FID" and they may simply refer to resolutions of 100,000 or 50,000 or others. To accomplish the different resolutions, these commercial manufacturers truncate the FID or expand the time-domain collection. How is this relevant experimentally? The larger the FID, the greater the resolution but the longer the scan time is for a single scan event. Thus there is a trade-off when collecting data at high resolution. For example, if one collects data at the highest resolution a single scan can take place from seconds to minutes per scan. This would not be optimal to

use on an LC-timescale. Therefore one must weigh the importance of scan rate versus the required resolution carefully when performing LC-MS with an FT-ICR-MS instrument.
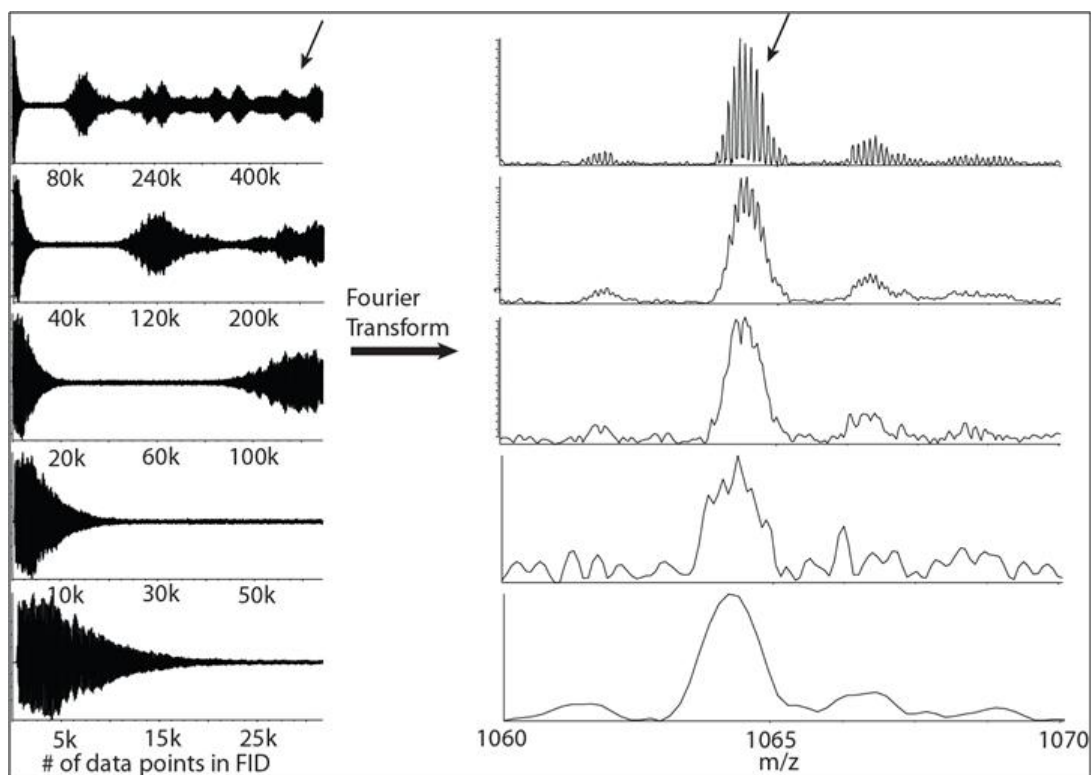
Figure 3: Correlation of image current detection length and mass resolution. The more data points the FID (left) consists of, the higher is the resolution of the mass signals (right) obtained by Fourier-Transform.

1.4.   Interpretation of FT-ICR-Mass Spectra

Due to the fact that most people investigating natural product biosynthesis do not routinely use FT-ICR-MS in their research, the interpretation of a high-resolution mass spectrum of a protein domain is highlighted in this section. The broadband spectrum of a free-standing thiolation protein Pks4 from the bikaverin biosynthetic pathway is shown in Figure 4A as a mixture of its apo and holo form.[62] This protein has a mass of 14,394 Da. Typically. The mass range for FT-ICR-MS analysis of proteins involved in the biosynthesis of natural products must fall between 200-2000 *m/z*. In order to enable the visualization of these ions within this standard *m/z* window, ions of Pks4, or other protein domains that are larger than 2000 Da, need to be multiple charged (*z*) which is experimentally accomplished by application of electrospray ionization. In the case of Pks4 we see that the same protein has multiple charge states ranging from 15 charges on the left side of the spectrum to charge 11 on the right side of the spectrum. The mass is calculated by taking the observed *m/z* and multiplying this value with the observed charge. Details on how to calculate the mass of such a protein or protein domain by manual means are reviewed in reference 47. Pks4 is a relatively small protein and these proteins of a size <20kDa are usually well suited for FT-ICR-MS analysis. Larger proteins pose more challenges to MS characterization. For instance, if a protein of 100 kDa is analyzed it is not uncommon to find between 100-150 charges on the protein and that the charge profile has >100 different charge states as opposed to the 5 observed for Pks4. Figure 4B shows an enlarged picture of a single charge state. When one zooms in on a charge state observed in Figure 4A, an isotopic profile of the ion becomes visible. This type of

characteristic isotopic profile exists because larger proteins, or protein domains, have many carbons, nitrogens, and oxygens. The natural abundance of isotopes other than the monoisotopic masses (e.g. ^{12}C, ^{14}N and ^{16}O) are the reason why an isotopic profile is observed for a protein. The main forms of isotopes that contribute to the isotopic profile of a protein are the natural ^{13}C and ^{15}N that are present at ~1% and 0.36% respectively. FT-ICR-MS has the resolving power that enables the visualization of the isotopes of proteins up to 110kDa in mass.[63] It should be noted however that, in practice, the larger the protein, the more difficult it is to get isotopic resolution and that the theoretical maximum isotopic resolution of large multi-domain proteins with higher field magnets (>12T) has not yet been achieved experimentally.[59]
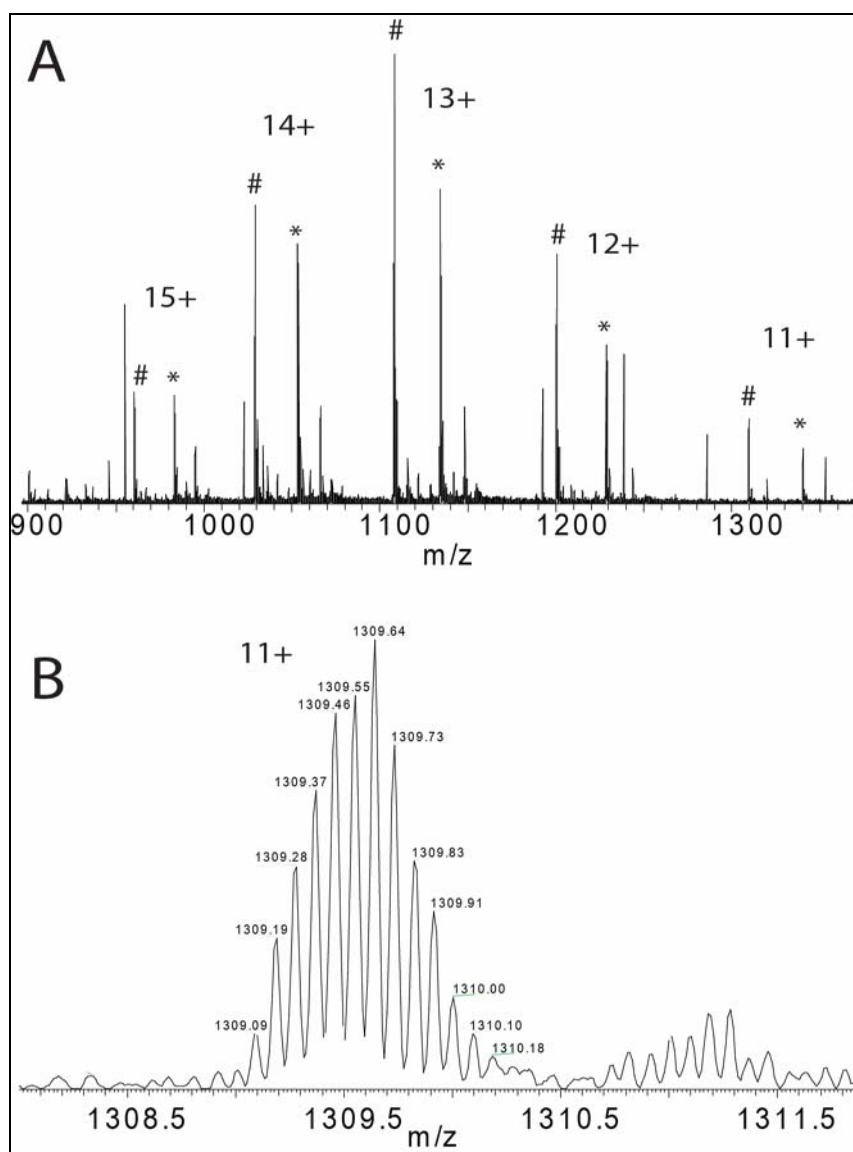
Figure 4: ESI-FT-ICR mass spectrum of freestanding ACP domain of Pks4.
(A) Multiple charge states of apo (#) and holo (*) Pks4. **B** – Single charge state (11+)
of Pks4 shows isotopic resolution obtained by FT-ICR MS analysis.

1.5. LC-FT-ICR-MS Analysis of NRPS and PKS Proteins

Before 2006, all NRPS and PKS proteins were investigated via direct infusion of a protein or protein domain that was purified off-line by HPLC, or by other C_{18}, C_8 or C_4 forms of peptide/protein purification such as Ziptips or traps. The best results were obtained using nanospray infusion, such as an Advion nanospray robot or similar nanospray devices. The advantage of a nanospray over direct microspray infusion via a syringe, a commonly used infusion method, is two-fold. First, nanospray creates finer droplets than traditional forms of electrospray, making the desolvation of the droplets emitted from the spray needle or nozzle easier, and many more ions enter into the gas phase improving the detection. Secondly, with just 5-10µl of a purified sample, one can analyze this sample for over an hour, sometimes up to three hours. With a syringe infusion approach, the infusion rate is usually 2 µl/min. This is a significant limitation in situations where one has limited sample or a low concentration sample. A single FT-ICR-MS scan may take 8-120 seconds depending on the selected settings, and typically, 20-200 scans are required to attain a good spectrum. Therefore nanospray becomes a critical component in the analytical platform for the biosynthetic investigations of NRPS and PKS systems. The reason why so many scans are required is illustrated in Figure 5. Figure 5 shows that even when the sample ionizes well, a single scan is not sufficient for accurate data. As shown for Pks4, it takes ~10 scans to generate an accurate isotopic profile for the protein. Many more scans may be required if the sample's concentration is low, does not ionize well, or contains competing ions. Depending on the conditions, acquisition of data may take minutes to hours, and is unlike NMR where some experiments run overnight. While dependent on many

factors, in practice, for every two-fold signal-to-noise improvement needed in order to collect the data, the data acquisition time needs to be increased four-fold. However, as the magnet size of an FT-ICR instrument increases, the sensitivity, and therefore the scan rate, increases compared to when the data is obtained at the same resolution on an instrument with a smaller magnet.
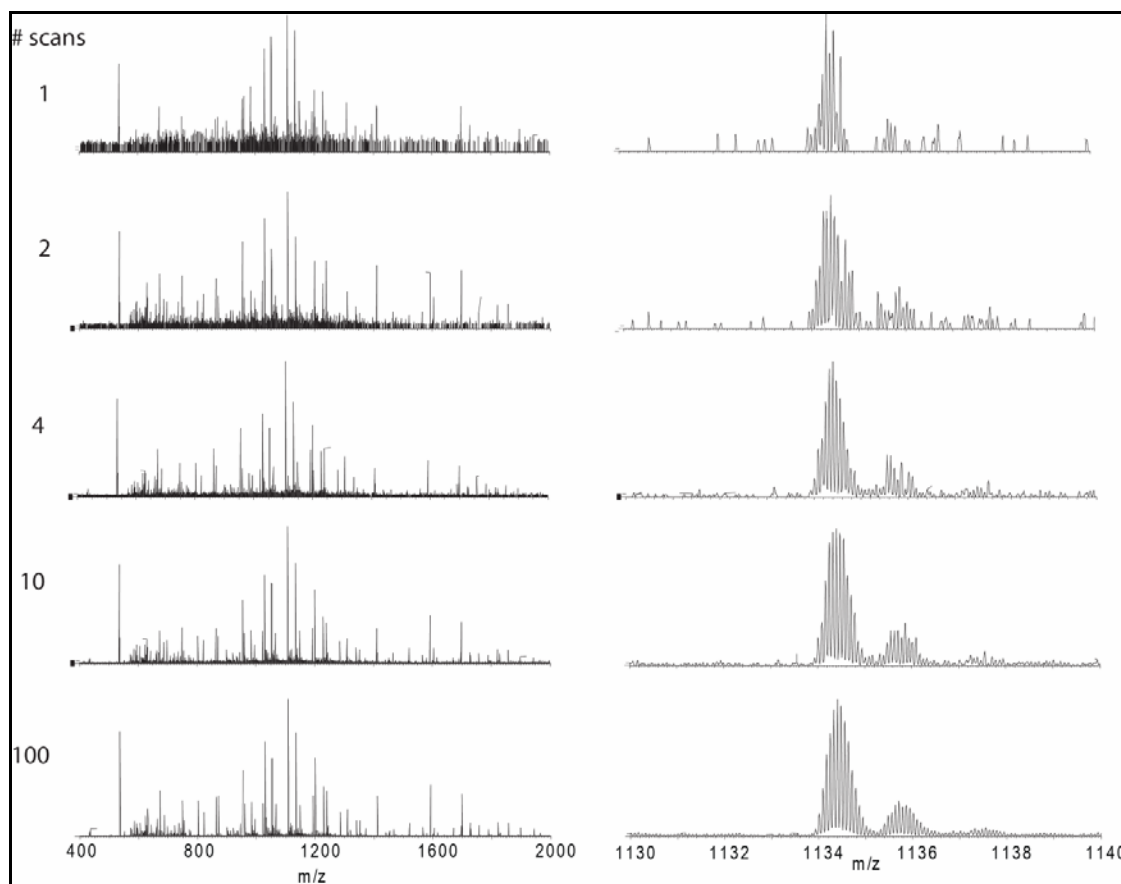
Figure 5: **Correlation of FT-ICR MS scan number and mass signal quality.** The more MS scans are acquired (left), the higher the signal-to-noise ratio in the corresponding mass spectrum (right).

NRPS/PKS active site mapping on the LC-timescale is possible by on-line LC-FT-ICR-MS. While this approach is still difficult to perform with a 7T magnet, the

most common magnet size at this time, it becomes routine with a 12T magnet. The challenges with on-line active site mapping are the inherent signal limitation of the electrospray and the long accumulation times of FT-ICR-MS instruments. The first online LC-FT-ICR-MS analysis of a thiolation active site was described by the Marahiel group in 2006.[42] In their LC-trace, they were able to observe a 1097.995 Da ion that corresponded to the post-translationally phosphopantetheinylated form of the T domain of tyrocidine synthase B.[42] In addition, they were able to observe the amino acid Phe and dipeptide Phe-Phe loaded onto the active site thiol of the phosphopantetheinyl group. If the digestion had resulted in a 15kDa protein domain containing the active site, this data would have been most likely much more difficult to obtain on an LC-timescale. Since then, the Kelleher lab has build a 12T FT-ICR-MS instrument enabling routine on-line LC-FT-ICR-MS analysis and detection of NRPS and PKS active sites.[58,64] Hopefully instruments with such high sensitivities and with this type of resolving power will become available to others who are tackling these types of systems for biosynthetic interrogation. Yet, as our understanding of the gas phase fragmentation behavior of active site-tethered substrates and intermediates advances, it may not even be necessary to have high resolution instruments for these types of experiments. One such advance is the phosphopantetheinyl ejection assay (PEA) for the characterization of phopshopantetheinylated proteins.[50]

1.6.   The Phosphopantetheinyl Ejection Assay (PEA)

Thiolation domains of many biosynthetic pathways are phospho-pantetheinylated. The phosphopantetheinyl post-translational modification of serine is

in many ways a novel phosphopepetide. It has been well recognized that the C-O connection of the phosphodiester bond in phosphopeptides is preferentially broken when they are subjected to thermal fragmentation methods (described in section 1.7).[65-69] This C-O bond is energetically the most labile connection in such a peptide. Two main mechanisms are forwarded for the ejection of a neutral loss phosphate (98 Da) from a phosphopeptide and are shown in Figure 6A.[2,68] The first mechanism results in the end-products that would be expected for a McLafferty-type rearrangement, the dehydoroalanine and the uncharged phosphate. In this mechanism the phosphate deprotonates the α-proton on the serine, ejecting the phosphate as a neutral ion. It is not yet known if the rearrangement is a McLafferty–type homolytic rearrangement or a heterolytic cleavage as drawn in Figure 6A. The second mechanism for neutral ion loss observed in phosphopeptides was forwarded by Hunt and co-workers.[2] in 2004. In this mechanism depicted in Figure 6B, a five-membered oxazoline is formed, a reaction that is promoted by the formation of protonated phosphate in the gas phase leading to the cleavage of the C-O connection and release of phosphoric acid as a neutral ion. Support that both neutral loss mechanisms are operational has been provided by isotope labeling studies, although these studies did show that the oxazoline mechanism was favored.[68] A phosphopantetheinyl serine modification has a related C-O phosphoester connection but it also has a second C-O connection bearing the active site thiol of the thiolation domain. These C-O linkages on the phosphodiester are the atom connections that most readily fragment in a fashion similar to phosphopepetide neutral losses observed with phosphopeptides. In addition, these C-O linkages will be preferentially fragmented over the fragmentation of amide

linkages normally expected of peptides. While the phosphoester ejection from a phosphopeptide results in a neutral loss, the ejection of the phosphodiester from a phosphopantetheinylated peptide or protein results in a loss of a charge on the peptide, resulting in the formation of the charge-reduced apo protein minus water in the case of the McLafferty-type ejection (Figure 6C), the oxazoline ejection (Figure 6D), and the thiazoline mechanism, or the charge-reduced apo protein plus phosphate when an iminolactone is ejected (Figure 6E).[50] At this time the iminolactone phosphopantetheinyl ejection appears to be the ion that is the most abundant and therefore the most useful in the investigation of substrates and intermediates tethered to the active site thiol of the phosphopantetheine. Figure 6F and G show examples of phosphopantetheinyl ejection of a peptide obtained via digestion of the T protein CouN5 from the coumermycin biosynthetic pathway. The intact ion corresponding to GILNSLNTAILVAH was subjected to CID and clearly showed all the different ejected PPant ions (obs 261.16 *m/z* and 359.12 *m/z*). This figure shows that these ions are the most prevalent ones generated in this spectrum. While they are the most abundant ions in ~70% of the phosphopantetheinylated proteins or domains investigated to date, these ejected ions are not always this abundant. Nonetheless, the PEA is a welcome addition to the arsenal of tools used to interrogate NRPS and PKS proteins.
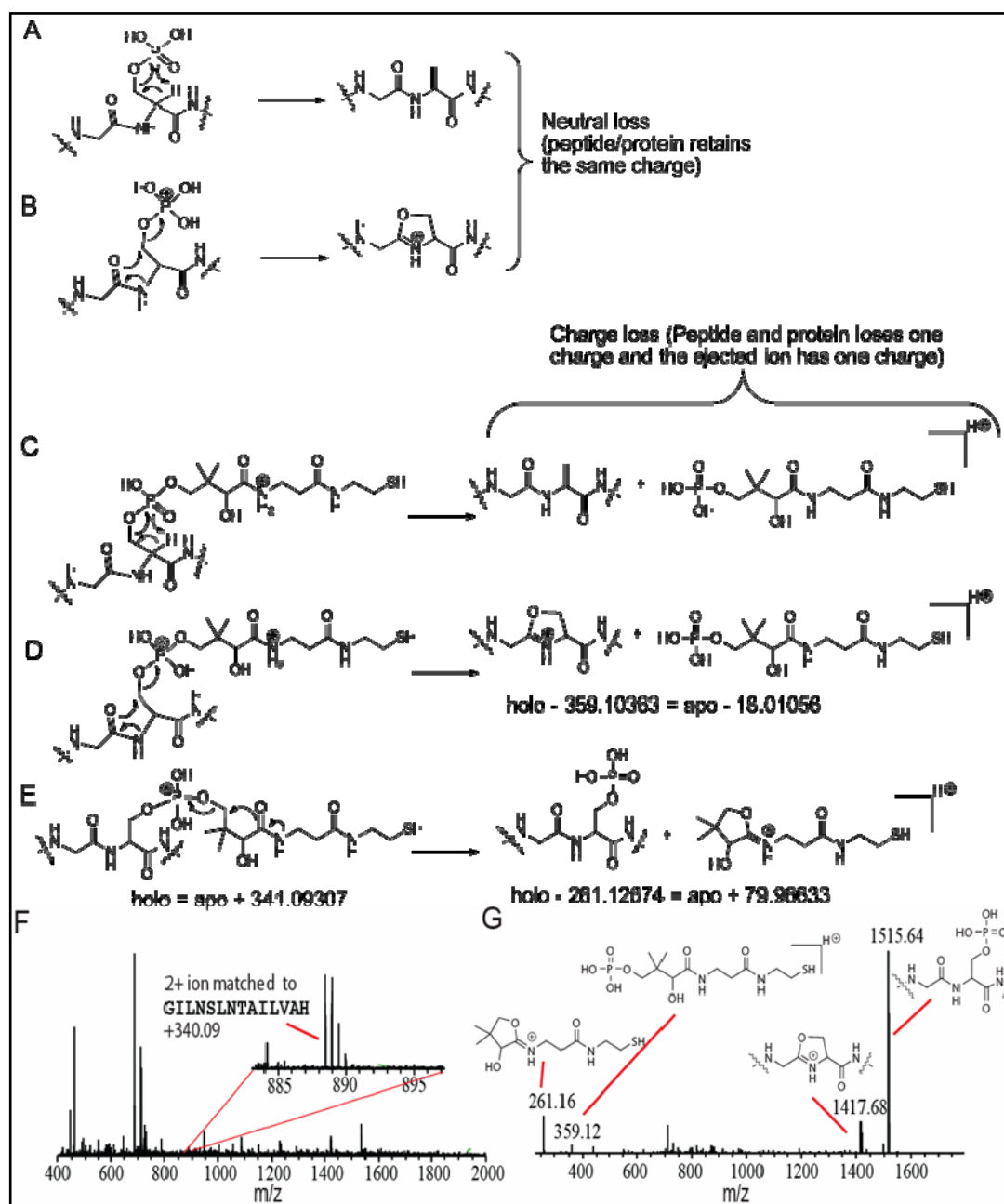
Figure 6: The phosphopantetheinyl ejection assay (PEA). (A) McLafferty-type
ejection mechanism. (B) Oxazoline ejection mechanism. (C,D) McLafferty-type
ejection (C) and oxazoline ejection (D) on phosphopantetheinylated protein yield
charge-reduced protein minus water and phospho-PPant ejection ion. (E) Iminolactone
PPant ejection on phosphopantetheinylated protein yields charge-reduced phospho-apo
protein and PPant ion. (F,G) PEA on CouN5 fragment in broadband FT mass
spectrum (F), MS^2 spectrum (G) shows PPant ion (261.16 *m/z)* and phospho-PPant ion
(359.12 *m/z)*.

PEA can also be carried out to monitor time courses as shown in Figure 7. In this reaction, the condensation of malonyl-S-PigH with pyrroly-S-PigG catalyzed by PigJ was directly monitored by the PEA of both of the two T domains of PigH in a time dependent manner and shows that changes can be monitored using PEA. This implies that kinetic information may be obtained from monitoring the ejected ion only. However, caution should be taken to prevent the over-interpretation of the kinetics from the observed ejected ions. While changes in a time dependent manner can be observed using PEA, there are three concerns that should be considered if one wants to obtain true kinetics using this method. First, the ejected ions may have different ionization efficiencies. Altering the ratios of two forms of a phosphopantetheinylated protein and then analyzing this by PEA may not respond in a linear behavior. Some of the discrepancy of what the ejected ions report is evidenced in Figure 7, as the intact spectrum on the left of Figure 7 does not correlate with the PEA relative ratios. In addition to the differences in ionization efficiencies, the substrate is often partially eliminated from the ejected ion as well. Therefore, the strength of the thioester of the substrate or biosynthetic intermediate will effect the ratios of the ejected ions observed. Because PEA is usually performed on a single charge state of the protein or peptide, the ratio of the ejected ion will vary as well, depending on the ionization efficiency of the parent ion. That change in ratios between ions of different charge states is directly evident from the Pks4 data in Figure 4A. Looking at the apo vs holo ratio of each charge state of Pks4, the ratio of apo (annotated with # in Figure 4A) vs the phosphopantetheinylated form (annotated with * in Figure 4A) of the protein

changes from 0.95 to 0.8. This same phenomenon is observed with different substrate or biosynthetic intermediate loaded forms of T domains. While the ratios of different charge states will affect the ejected ions, the manner in which the ions are accelerated by CID also affect the PEA ratios, especially when a small parent ion isolation window is applied. Therefore, it is recommended that the isolation window width for CID, when multiple species are analyzed, is greater than the isotopic profile width of any of the ions that one is interested in fragmenting. Otherwise, different fragmentation energies are applied to the different parent ions and differential ejection can be observed. Despite these inherent caveats, PEA has been used to provide some kinetic information[55,69] and follows the generally accepted accuracy of 10-20% that can be obtained with the investigations of NRPS and PKS protein domains.[70]
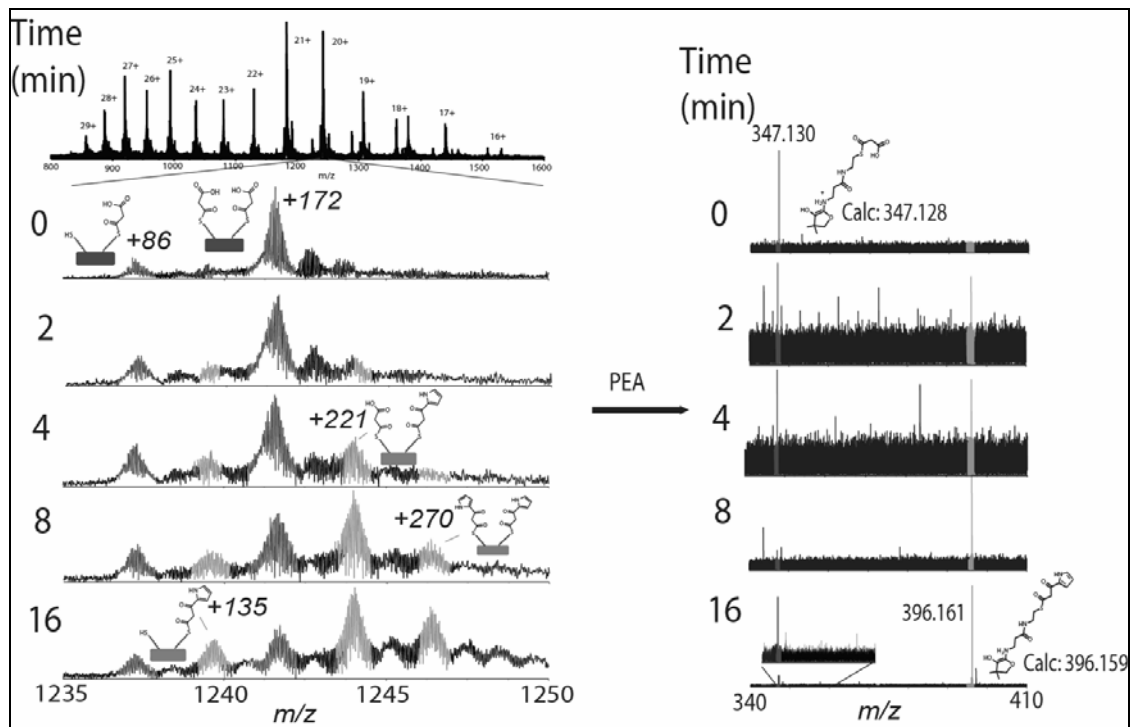
Figure 7: **Monitoring time courses by PEA.** The condensation of malonyl-*S*-PigH with pyrrolyl-*S*-PigG catalyzed by PigJ is monitored in a time course by broadband FTMS (left) and PEA (right) on the two active sites of PigH.

1.7. How PEA is Accomplished

PEA is accomplished using thermal activation methods. There are a large number of thermal activation methods that could be used for this. Source fragmentation, blackbody infrared radiation dissociation (BIRD), sustained off-resonance irradiation-collisionally activated dissociation (SORI-CAD), infrared radiation multi-photon dissociation (IRMPD) and collisionally induced dissociation (CID) are examples of such thermal tandem mass spectrometry methods.[71-76] Currently, source fragmentation, CID and IRMPD are the only methods that have been utilized for the PEA, and because of this, these are the only three that are covered in

this section of this chapter. Before we begin to look at how and when one should use these methods, it is important to describe the common instrumentation configurations used in NRPS and PKS studies. Since most of PEAs have been performed on FT-ICR-MS instruments, the two main commercially FT-ICR-MS configurations are described in this section. In the first configuration, after the sample is introduced, the ions pass through a heated capillary inlet to a quadrupole where ions can be isolated and then passed on to a linear ion trap. The ions can then be detected in the ion trap but at low resolution. Alternatively, the ions can be passed to the ICR cell and the ions can be detected with high resolution (Figure 8). Thermo Finnigan hybrid instruments are typically configured in this fashion. In this instrument configuration, thermal fragmentation of ions can take place by exciting the ions at the source by increasing the voltage and colliding the ions with air. On the other hand, CID can be accomplished via excitation of the ions in the ion trap resulting in helium gas collisions. Finally, the instrument can also be equipped with an optional IRMPD and the ions will be fragmented directly in the cell of the instrument. Other commercial FT-ICR instruments from Varian (formerly IonSpec) and Bruker often have a different configuration. Both of these instruments have a sample inlet, usually with a heated capillary, an optional ion funnel (not depicted) to capture as many ions as possible, followed by an isolation quadrupole and an accumulation octupole before passing the ions on to the cell of the ICR instrument. Thermal activation can be accomplished by three means: 1) at the source via collisions with air, 2) in the accumulation octupole via collisions with helium or another inert gas, and 3) inside the cell using a pulsed

laser (IRMPD). Below we describe the types of experiments that can be performed by each set-up.
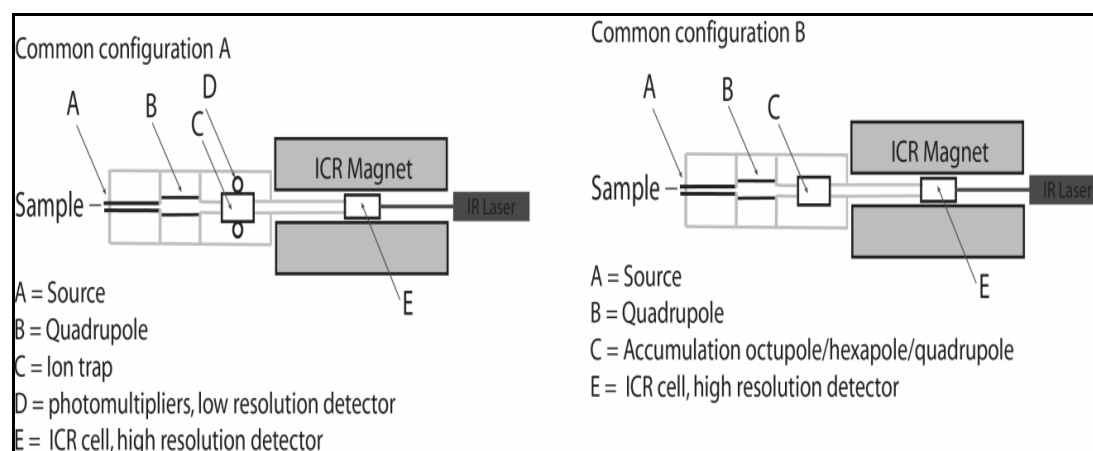
Figure 8: **Two main FT-ICR-MS configurations.** The LTQ-FT-ICR-MS configuration (left) and the accumulation multipole FT-ICR-MS configuration (right) of FT-ICR mass spectrometers.

1.7.1. The LTQ-FT-ICR-MS Configuration for PEA (common configuration A)

A hybrid instrument such as the LTQ-FT-ICR-MS configuration (Figure 8) allows one to perform some very interesting experiments with respect to NRPS and PKS multi-domain proteins. First, we can fragment at the source of the instrument, resulting in phosphopantetheinyl ejection for any ion that enters the instrument. The advantage of this approach is that we are not just isolating a single charge state for fragmentation but all the charge states from a protein, thereby increasing the signal intensity of the ejected ion. The disadvantage of this approach is that one is likely to get many more signals in the low *m/z* region, requiring additional confirmation of the signals observed. This confirmation can be accomplished using substrates with stable isotopes or an additional round of fragmentation in a data dependent manner as done for proteomics experiments. This is of particular importance when the ejection of the

phosphopantetheine is performed on low resolution instruments as described in section 1.8. An additional disadvantage of using source fragmentation for PEA is that the ejected ion cannot be correlated with the parent ion. This is important when one wants to map active sites. While source fragmentation is one of the approaches to obtain phosphopantetheine ejection, the most common approach is collisionally induced dissociation (CID). In CID, ions of interest are accelerated and collide with gas ions (helium being the most commonly used). During this collision, most of the kinetic energy is subsequently converted to thermal energy and if the phosphopantetheinylated ion undergoing the collisions has enough vibrational energy amassed which likely requires multiple collisions it will dissociate into two ions. CID can be accomplished in an inert gas-filled quadrupole or in an ion trap. In the case of a Thermo Finnigan instrument (configuration A), CID is accomplished in the linear ion-trap portion of the instrument. Once dissociated, the fragment ions can be observed using a photomultiplier or the ions can be passed on to ICR-cell for high resolution analysis. Unfortunately, ion traps suffer from a major but well documented limitation that is defined as the 1/3rd rule.[77,78] When the activation q, i.e. the energy that is responsible for accelerating all of the ions to be fragmented, is raised, low *m/z* product ions start to lose stable trajectories causing them to be ejected from the trap.[77,78] For example when an activation q is set to 0.25 and the parent ion is isolated at 1200 *m/z* one cannot observe fragment ions below 400 *m/z*. Some improvements in terms of the detectability of the low *m/z* scanning range can be gained by lowering the activation q, but this change necessitates an increase in the time for activation from 30 ms to 100 ms in order to obtain significant fragmentation. On the LTQ-FT-ICR-MS

instrumentin the Dorrestein laboratory, when the activation q is set to 0.2, the 1/3rd rule becomes a 1/4th rule thereby increasing the size of the ions that can be fragmented, while still detecting fragments with an *m/z* low enough for the PEA assay. A partial solution for this limitation of ion traps is a related software-controlled mechanism called Pulsed-Q Dissociation (PQD). Although this is not yet available for commercial LTQ-FT-ICR-MS instruments, it is available for LTQ's, LCQ's and LTQ-ORBI's.[79,80] Most ion traps in existence today are not equipped with a PQD software upgrade so that its utility is limited at this time. While it is possible to equip an LTQ-FT-ICR-MS instrument with IRMPD, it has not yet been applied towards PEA on such instruments.

1.7.2. The Accumulation Multipole Set-up for PEA (common configuration B)

The other common set-up for FT-ICR-MS instrumentations is configuration B (Figure 8) which differs from the ion trap configuration by having an accumulation multipole instead of a linear ion trap. Such an instrument is also capable of PEA by conducting fragmentation at the source of the instrument. The combination of the quadrupole-accumulation multipole set-up is very useful in PEA. The advantage is that one can select out ions of interest in the quadrupole and then accumulate ions in the accumulation octupole as shown in Figure 10. The ions can then be excited for CID inside the accumulation multipole or they can be passed on to the cell of the instrument. Inside the cell of the instrument, the isolated ions can be subjected to PEA using infrared radiation (IRMPD). While both CID and IRMPD are thermal activation methods and the resulting fragmentation are quite similar it is not yet investigated if

one results in improved phosphopantetheine ejection over the other. The advantage of CID or IRMPD on an accumulated ion signal is that the intensity of the ejected ion is much more intense. The disadvantage is that it can take many seconds, sometimes up to 60 seconds, to collect a single scan on an 8.4T instrument. Therefore, this configuration is typically not amenable to the LC-timescale.
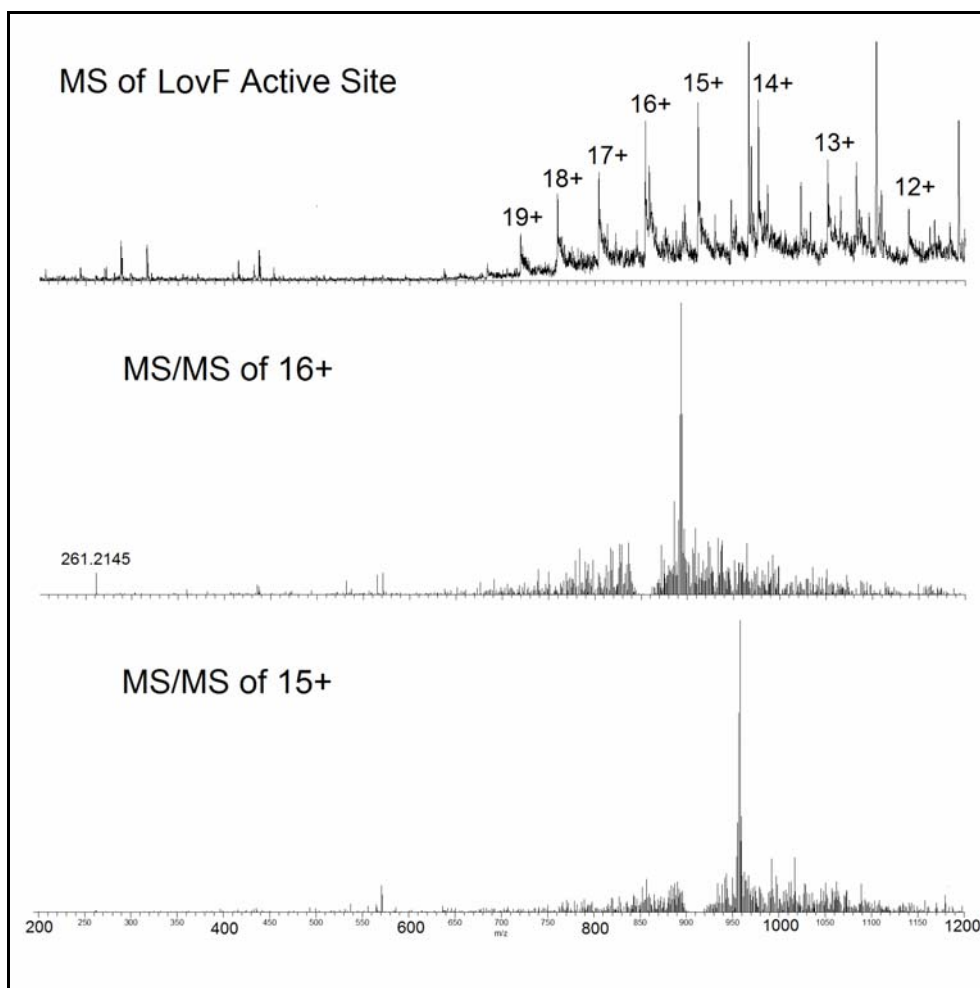
Figure 9: Illustration of the 1/3rd rule. The top panel shows multiple charge states of the phosphopantetheinylated active site containing peptide from PKS LovF. The 16+ ion at 854.636 *m/z* and the 15+ ion at 911.636 *m/z* of the peptide were subjected to CID. The middle panel shows that PPant ejection (261 *m/z*) is clearly detected in the MS/MS spectrum resulting from CID of the 16+ ion. The bottom panel shows that

PPant ejection resulting from fragmentation of the 15+ ion cannot be detected due to limitations defined by the 1/3rd rule.
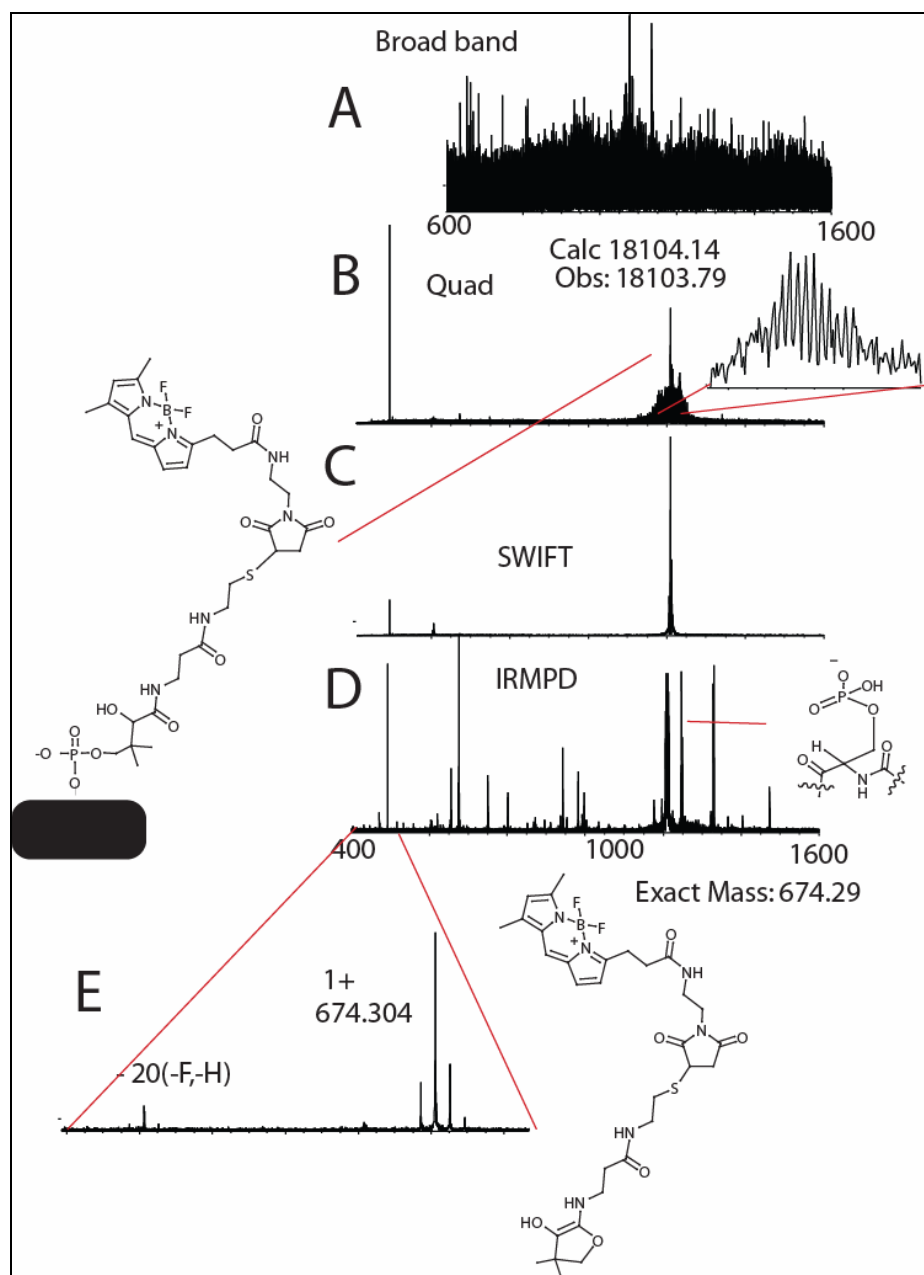
Figure 10: **The accumulation multipole set-up for PEA.** Ions of interest are selected by the quadropole (A,B), accumulated in the accumulation octupole (C) and subjected to PEA by IRMPD (D,E) in the analyzer cell

1.8. PEA on Non-ICR Instruments: Low Resolution PPant Ejection

PEA is an important discovery in the characterization of phosphopante-theinylated proteins. It allows one to analyze non-radioactively labeled substrates thereby making many more substrates accessible to biosynthetic studies of these types of proteins. In addition, this approach often eliminates the need for preparing synthetic standards for comparison of hydrolyzed thioester intermediates, which may be challenging and time consuming to accomplish. Moreover, PEA immediately shows that the substrate or intermediate is connected to the phosphopantetheine post-translational modification and not elsewhere on the protein. This is not possible to detect with radioactivity or any other conventional assay used to study these types of systems. An additional advantage that PEA has over radioactive assays is that it provides a mass signature and therefore unexpected intermediates can be observed. Even though PEA is such a useful tool in the investigation of substrates and biosynthetic intermediates, it has not been widely used by other investigators because it requires costly FT-ICR-MS instrumentation and highly skilled researchers to use them. Because most mass spectrometry instruments are able to perform source fragmentation, CID, or both, this method should be readily be applicable to low resolution instruments. Since properly calibrated low resolution instruments have mass accuracies well within 0.5 Da, they can be used to differentiate molecular species within 1 Da for the 1+ ejected phosphopantetheine ions, making such an instrument a very useful tool for the characterization of phosphopantetheinylated proteins. As PEA can be applied to low resolution instruments, it is an assay that is accessible to most

researchers studying phosphopantetheinylated proteins. The challenge of this assay is that even though the ejected ion in the low *m/z* range is often the most abundant one, there are other ions in this region of the spectrum. Due to these other ions and to noisier low resolution detectors which are not based on image current measurement, it can be more difficult to confirm the candidate ejected ion than anticipated. This confirmation can be carried out using labeling studies or directly by PPant fragmentation, a new MS^3 method, if an instrument has the ability to perform MS^3. When this MS^3 method was originally reported, detection of 12 diagnostic ions were reported for the second round of MS/MS on the ejected ion, but there are many more fragments with lower abundance, including an important and abundant ejected ion fragment that reports on the substrates loaded onto the thiol of the phosphopantetheinyl functionality (Figure 11, Dorrestein unpublished observations).[81] Thusfar, the confirmation of the ejected ions has only been accomplished on an ion trap instrument in two different modes. In the first mode, the phosphopantetheinylated protein is observed by mass spectrometry, the phosphopantetheinylated ions are isolated in the ion trap and subsequently subjected to CID (MS^2). Then the ejected ion is isolated and fragmented again using CID (MS^3). This will result in the diagnostic fragment pattern for the phosphopantetheinylated peptide ejection ions. The second mode relies on the initial use of source fragmentation followed by CID on the ejected ions in the low *m/z* region (MS^2). In this case, the method can be performed in a data dependent manner, as it is done for proteomic experiments, to find the characteristic phosphopantetheinyl "MS^3" signature from any of the ions that enter the instrument. The main disadvantage of the source fragmentation method, as mentioned earlier, is

that all the information of the precursor ion is lost, but it should still enable the analysis of phosphopantetheinylated proteins on non-ion trap, low resolution instruments that only have one stage of MS capabilities by additional confirmation of the PEA ions.
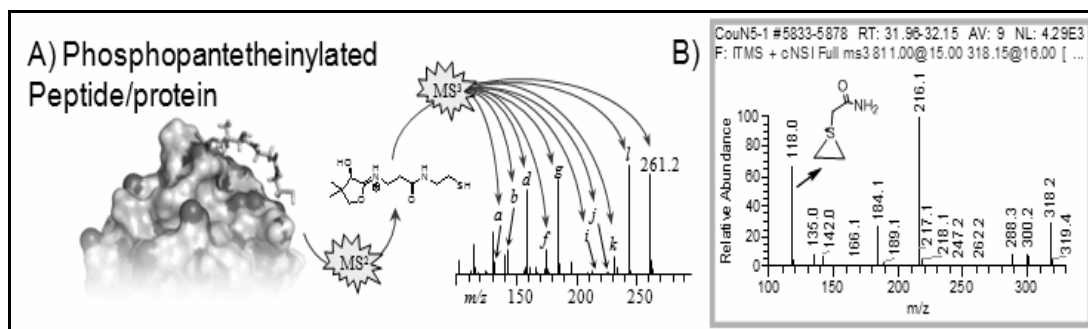
Figure 11: The PPant fragmentation method. (A) A phosphopantetheinylated protein is first subjected to the PEA and, subsequently, the PPant ejected ion is fragmented again to yield diagnostic MS/MS peaks (B).

1.9. Low Resolution Capillary LC-MS on Ion Traps

To date, FT-ICR-MS instruments are the main instruments used for PEA but not the best instruments to analyze NRPS/PKS on an LC-timescale because their scan rates are very long at the present time. The scan rates of most low resolution instruments are shorter and therefore much better suited for interfacing with LC-chromatography. In proteomics, it has become standard to use small-bore columns (30-100 µM inner diameter) with nanoflow (200-500 nl/min) gradients, but this has not yet found much use in the investigations of biosynthetic pathways. More recently, with the ability to perform PEA, these LC-MS approaches have emerged as useful tools in the investigation of phosphopantetheinylated peptides and proteins. The advantage of LC with a 100µM column as opposed to traditional 1, 2.1 or 4.6 mm HPLC columns is that it uses much less material. In the case of a 4.6 mm column,

injection of >100 μg of sample is not uncommon to attain good signals, while in the case of a 100μM diameter column, one typically uses 0.1-1 μg of material. This approach has recently been used to observe the phosphopantetheinylation of a carrier domain from the hemolytic toxin pathway from *Streptococcus agalactiae* and its corresponding phosphopantetheinyl ejection.[81]

1.10. Mass Spectrometry of Intact NRPS and PKS Multidomain Proteins

There have only been a few reports on the mass spectrometry of intact multi-domain NRPS and PKS proteins.[82] The NRPS proteins GrsA and NikP1 have been analyzed by FT-ICR-MS but not with isotopic resolution. In addition, the DEBS PKS didomain has been investigated by mass spectrometry.[82] In most cases, a mass shift can be observed upon phosphopantetheinylation or substrate loading. The main challenge with intact NRPS and PKS protein analysis is to obtain sufficient quantities of very pure proteins. A small amount of contamination by a small peptide or small domain will compete for the signal very efficiently, making it nearly impossible to observe the larger ion. While the examples of GrsA, NikP1 and DEBS illustrate that it is possible to interrogate these larger proteins, it should even be possible to analyze entire NRPS and PKS muli-protein complexes such as the ones observed for bacilleane.[83] These NRPS/PKS multiprotein complexes are similar in size to intact viral particles or intact ribosomes that have been investigated by mass spectrometry.[84,85] The mass spectrometry of intact protein and multiprotein complexes is an area that remains wide open for exploration.

1.11   Mass Spectrometry of Phosphopantetheinylated but Non-NRPS and Non-PKS Proteins

In many respects, the phosphopantetheinyl functionality is similar to a phosphoserine and treatments that work with phosphorylated proteinaceous materials will often work with phosphopantetheinylated materials as well. For instance, phosphatases are responsible for the removal of phosphates but were recently shown to remove the phosphopantetheinyl functionality as well.[86] While this assay has not yet been utilized on the biosynthesis of natural products, this assay has been applied to the investigation of the phosphopantetheinylated protein 10-formyltetrahydrofolate dehydrogenase which is involved in the formation of formyltetrahydrofolate.[86] Donato *et al.* used the phosphatase assay to confirm the presence of the phosphopantetheinylation. The characterized phosphopantetheinyl modification suggests that there may be many other phosphopantetheinylated proteins that have not been identified. This assumption has recently been confirmed using a phage display approach.[87]   In the study by Donato *et al.* a common MALDI-TOF instrument was used for analysis. One caution about their interpretation of the spectra should be noted and it is a direct result that the experiments were performed by a novice and not a mass spectrometry expert. In their spectra these researchers observed a mass of 360.08 Da and in their text they report that the mass of the ion should be 358.33 Da, which is not correct. We recalculated the mass of the hydrolyzed ion and anticipate that the mass of this hydrolyzed species should be 359.104 Da in positive mode while in negative mode it is 357.089 Da. In addition to the monomer unit, the authors saw a dimer with a mass of 550.60 Da where both phosphates had been removed. By our

calculations, this dimer should have a mass of 555.252 Da for a 1+ ion or 553.237 Da for the 1- ion, while they reported a calculated mass of 552 Da for this ion. Their measured mass error is likely a result of poor calibration or relying on old calibration files, as most people who use MALDI-TOF instruments in core facilities do. However, it is unclear why their calculated masses were not spot-on. In the end, while it is important for people to realize the mass errors in this particular report if they wish to repeat the experiments, this point of caution does not invalidate the overall conclusion provided in this paper. The mass spectrometry still supports that 10-formyltetrahydrofolate dehydrogenase is a phosphopantetheinylated protein and therefore is an addition to the ever expanding population of phosphopantetheinylated proteins such as the citrate synthase GcvT in the glycine cleavage system of select organisms. Furthermore, this paper described a new approach to the characterization of phosphopantetheinylated proteins by mass spectrometry. It remains to be determined if this approach is widely applicable to other phosphopantetheinylated proteins and if it can be applied to substrate- and intermediate-loaded phosphopantetheine arms, e.g. of NRPS and PKS systems. The major concern in this case is the lability of the thioester and the long phosphatase incubation times. A typical half-life of a thioester is 200-400 min^{-1} even under stabilizing acidic conditions.[88]

1.12   The Development of Recognition Software for PEA on an LC-Timescale

While there are many proteomics programs designed to find peptides or neutral ion losses such as phosphoric acid from phosphopeptides, there are no programs or

algorithms developed that can identify phosphopantetheinylated peptides. The main challenges in the annotation of phosphopantetheinylated peptides by such programs is the observation that the ejected ions are often very abundant and therefore limiting to the intensity and number of normal fragment ions typically encountered with peptides. In the case of phosphopeptides, there were enough examples available that software could be trained to recognize such peptides.[89] In the case of phosphopantetheinylated proteins, there are a limited number of such data training sets with which to train new software. In addition, phosphopantetheinylated peptides have two phosphoester linkages while phosphopeptides only have one, making it more likely that this ion is ejected instead of amide cleavage. Finally, the thiol of the phosphopantetheinyl functionality is modified with substrates and intermediates making it even harder to identify and find phosphopantetheinylated peptides. Such modifications to phosphopantetheinylated peptides will need to be taken into account when conducting searches for active sites. Current efforts are underway to overcome some of these limitations. Once solutions to this problem are obtained, it will become possible to study the biosynthesis of therapeutics that are biosynthesized on phosphopantetheinylated proteins, at their native le vels, using proteomic approaches as a major genome mining tool.

Chapter 2: Applications of Mass Spectrometric Analysis of PKS Systems

Polyketides represent a source of numerous pharmacologically and commercially useful compounds.[90] The appeal of polyketides is that they are a structurally diverse class of compounds, yet these complex molecules can be synthesized from much smaller, simpler acyl-CoAs. Polyketides are synthesized by large multidomain megasynthase polyketide synthases (PKS).[23,91] These megaenzymes efficiently carry out the addition of the acyl-CoAs to form elongated intermediates that undergo a variety of different enzymatic tailoring steps. Polyketide synthases are classified as either type I, type II, or type III. Type I PKSs are a single protein consisting of a linear arrangement of the various catalytic and carrier domains. Type II PKSs are comprised of the various domains that exist as individual proteins that interact with each other. Type III PKSs function without the use of a phosphopantetheinylated carrier domain. In addition, polyketides can be classified as either being modular or iterative. Modular PKSs have multiple domains that function in an assembly-line fashion in which the substrate is bound to a carrier protein of the first domain, undergoes modifications, and then is transferred to a carrier protein within the next domain. The growing intermediate is passed from one catalytic domain to the next and undergoes elongation and additional modifications at each domain until a full-length intermediate is released. Iterative PKSs possess only a single module consisting of a limited number of catalytic and carrier domains that are reused over multiple iterations during the elongation of the intermediate. The intermediate undergoes cycles of addition and modification. Type I PKSs and type II PKSs have been the subject of recent investigations and the fact that they possess the

phosphopantetheinyl functionality on their acyl carrier protein (ACP) domains make them suitable targets for studies involving mass spectrometry. Using high resolution FTMS, the intermediates of PKS biosynthetic pathways can be detected while still covalently bound to active site of the T domain via phosphopantetheine. Confirmation of the exact mass of PPant-bound intermediates can be confirmed by subsequent PPant ejection assay.

2.1. Bacillaene Biosynthesis

2.1.1. Orphan Gene Cluster *pksX*

The *pksX* gene cluster of the *Bacillus subtilus* encodes a hybrid PKS-NRPS that produces a previously unknown secondary metabolite. This orphan gene cluster was found to share many similarities with gene clusters involved in the biosynthesis of curacin, jamacamide, pederin, as well as others. Of particular interest were several biosynthetic tailoring enzymes expressed such as a trans-acting acyltransferase (PksC). Also of interest were several free-standing proteins, such as a thiolation domain (AcpK), a ketosynthase (PksF), a 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase (PksG), two enoyl-CoA hydratases (PksH and PksI), and a larger protein containing multiple thiolation domains (PksL, Figure 12A). In addition to various biochemical techniques, high resolution mass spectrometry was applied to designate functional roles to these various proteins encoded by the *pksX* gene cluster. The genes for PksC, AcpK, PksF, PksG, PksH, PksI, and region of the gene encoding the pair of tandem thiolation domains PksL were heterologously expressed in *E coli* and then purified. The AcpK protein and the tandem thiolation domains (PksL-T$_2$) were phosphopantetheinylated *in vivo* by coexpressed Sfp.[57]

Substrate loading onto AcpK and subsequent alterations to the phosphopantetheine-tethered intermediate that were hypothesized to be carried out by the array of proteins encoded by *pksX* were determined by FTMS. PksC, when incubated with holo AcpK and malonyl-CoA, resulted in an +86 Da shift of the AcpK protein, corresponding to an 8.6 m/z shift of the 10+ of AcpK ion. This mass difference was consistent with the formation of malonyl-*S*-AcpK. This observation, coupled with detection of an +86 Da mass shift of PksC itself when incubated with malonyl-CoA, confirmed the malonyl-acyltransferase function of PksC. The function of PksF was determined by incubating the protein with malonyl-S-AcpK. The resulting 44 Da loss in mass, representing decarboxylation of malonyl-*S*-AcpK to acetoacetyl-*S*-AcpK, was detected with FTMS and helped verify the function of PksF as that of a ketosynthase. To probe the possible function of PksG as that of an HMG-CoA synthase, acetoacetyl-*S*-PksL-T2 was incubated with acetyl-*S*-AcpK and PksG. The addition of PksG facilitated the formation HMG-*S*-PksL-T$_2$, at one or both tandem thiolation domain active sites, observed as a +60 Da mass shift by FTMS. Infrared multiphoton dissociation resulted in a PPant ejection ion with a mass of 503.152 Da, 60.022 Da larger than the mass of acetoacetyl-loaded PPant (Figure 12B). The functional characterization of PksH and PksI activity was again facilitated by high resolution MS and confirmed by PPant ejection analysis. HMG-*S*-PksL-T$_2$ was incubated first with PksH alone, then PksI alone. On its own, PksH did not yield any change in the phophopantetheine-tethered intermediate. Incubation of HMG-*S*-PksL-T$_2$ with PksI, on the other hand resulted in a loss of 18 Da representing dehydration of the intermediate. Incubation of HMG-*S*-PksL-T$_2$ with both PksH and PksI resulted in a mass shift of 62 Da. This corresponds

with the dehydration (-18 Da) and subsequent decarboxylation (-44 Da) of the HMG-intermediate. These findings were further validated by observation of mass shifts in the PPant ejection ions generated by IRMPD of the various PksL-T_2-bound intermediates (Figure 12C). These studies highlighted the utility of FTMS and the PPant ejection assay as means to elucidate the functions of the various products of an orphan PKS gene cluster such as *pksX* by *in vitro* reconstitution.

2.1.2. Trans-enoylreductase and α- and β-ketone Reduction

More recently, the ultimate biosynthetic product of the *pksX* gene cluster was identified and its structure was elucidated.[92] This product was identified as dihydrobacillaene, which is later converted to bacillaene (Figure 13A) by PksS.[93] Dihydrobacillaene is produced by numerous enzymes and tailoring domains discussed previously, as well as an additional 4 megasynthase complexes. The first of these megasynthases, PksJ, contains two NRPS modules followed by two PKS modules. Recent investigation into the dihydrobacillaene biosynthetic pathway involved the determination of the origin of the α-hydroxyacyl N-cap.[58] While the first two modules of PksJ seemed capable of accepting α-hydroxyisocaproate (α-HIC) directly, before being transferred to the third PksJ module, researchers found that PksJ preferentially loads α-ketoisocaproate (α-KIC) first, then transfers the intermediate to the third module. Module 3 of PksJ possesses a pair of tandem thiolation domains which receive the α-KIC containing intermediate. The presence of a ketoreductase domain and the tandem-T domains within the third PksJ module led to the hypothesis that the α-KIC containing intermediate is reduced to α-HIC within the third module by the

single KR domain acting on both the α-keto and β-keto groups of the α-KIC intermediate (Figure 13B).

To test the functions of the KR domain, investigators heterologously overexpressed the region of PksJ containing the ketoreductase domain and the tandem T domains, PksJ(KR-T_3-T_4). To assess the typical function of the PksJ-KR, as a β-ketoreductase, apo PksJ(KR-T_3-T_4) was incubated with acetoacetyl (Acac)-CoA and Sfp in order to generate Acac-*S*-PksJ(KR-T_3-T_4). Reduction was carried out using either NADH or NADPH, followed by ArgC digestion. The T domain active site fragments incubated with NAD(P)H were compared to controls. A small shift in mass of the active site was detected in samples treated with NADH/NADPH, and the mass shift of +2.0166 Da was verified by carrying out source-induced dissociation for PPant ejection (Figure 13C). This mass shift is consistent with β-ketoreduction of Acac-*S*-PksJ(KR-T_3-T_4) to β-hydroxybutyrl-*S*-PksJ(T_3T_4), thus confirming the function of the PksJ module 3-KR as a β-ketoreductase. While a β-ketoreduction represents the rule for KR function, α-ketoreduction represents the exception. In order to test the ability of PksJ module 3-KR to reduce the distant α-ketone of α-KIC, researchers utilized α-ketoisocaproyl-γ-aminobutyrate (α-KIC-GABA) as a model substrate. Reduction of α-KIC-GABA-*S*-PksJ(KR-T_3-T_4) occurred in an NAD(P)H-dependent manner and was detected by protease digestion followed by FTMS and PPant ejection analysis (Figure 13C).
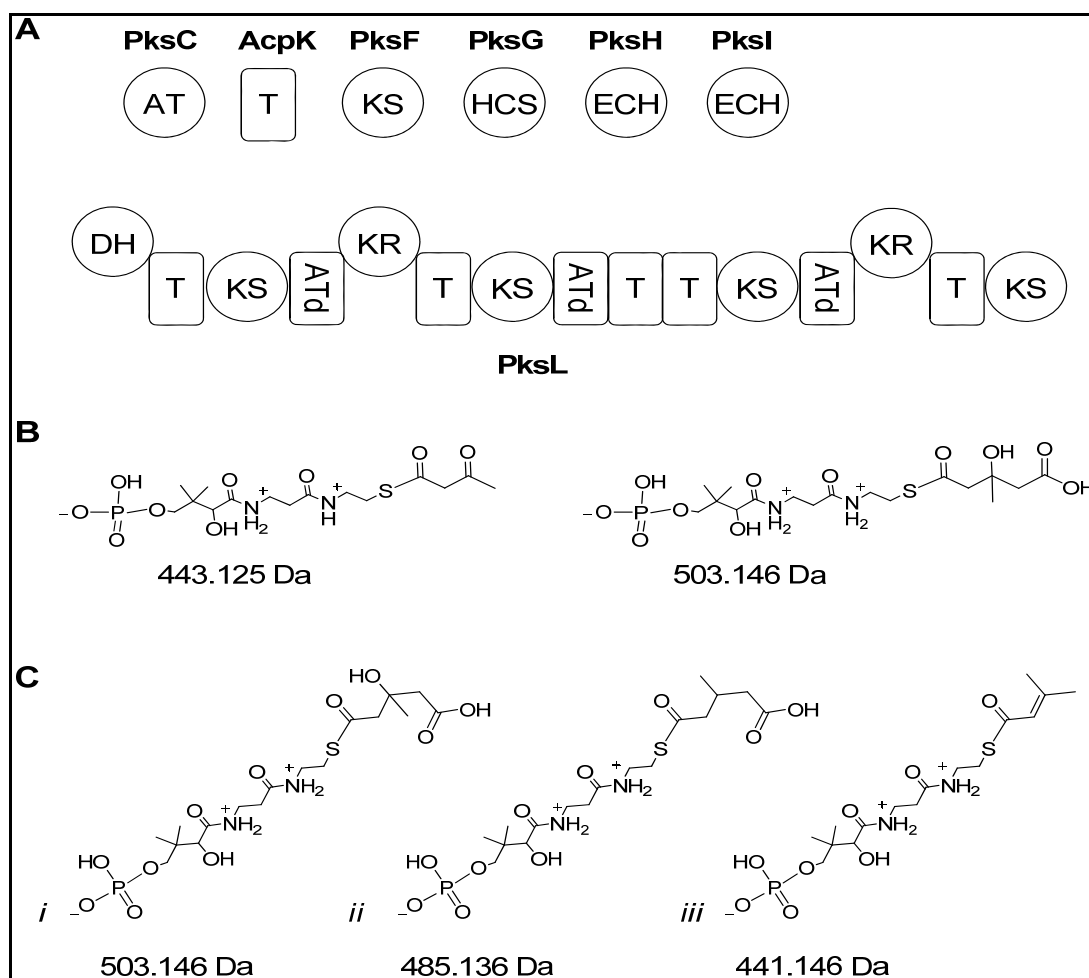
Figure 12: Bacillaene biosynthesis I. (A) Annotation of the function of multiple proteins encoded by the *pksX* gene. (B) Determination of PksG (HCS) function. Acetoacetyl (Acac)-*S*-AcpK phosphopanthetheine ejection ion was observed to have a mass of 443.125 Da. Incubation of Acac-*S*-AcpK with PksG results in an increase of 60.02 Da, consistent with formation of HMG-*S*-AcpK and confirmation of HCS activity by PksG. (C) Functional determination of PksH and PksI. HMG-*S*-PksL-T_2 was incubated with PksI and PksI & PksH. Phosphopantetheinyl ejection ions were generated using IRMPD. (i) The ejection ion from HMG-*S*-PksL-T_2 has a mass of 503.146 Da. (ii) Incubation with PksI results in a loss of 18 Da consistent with dehydration. (iii) Incubation with PksI and PksH results in a loss of 62 Da, corresponding to dehydration followed by decarboxylation. Incubation of HMG-*S*-PksL-T_2 did not result in any mass changes.
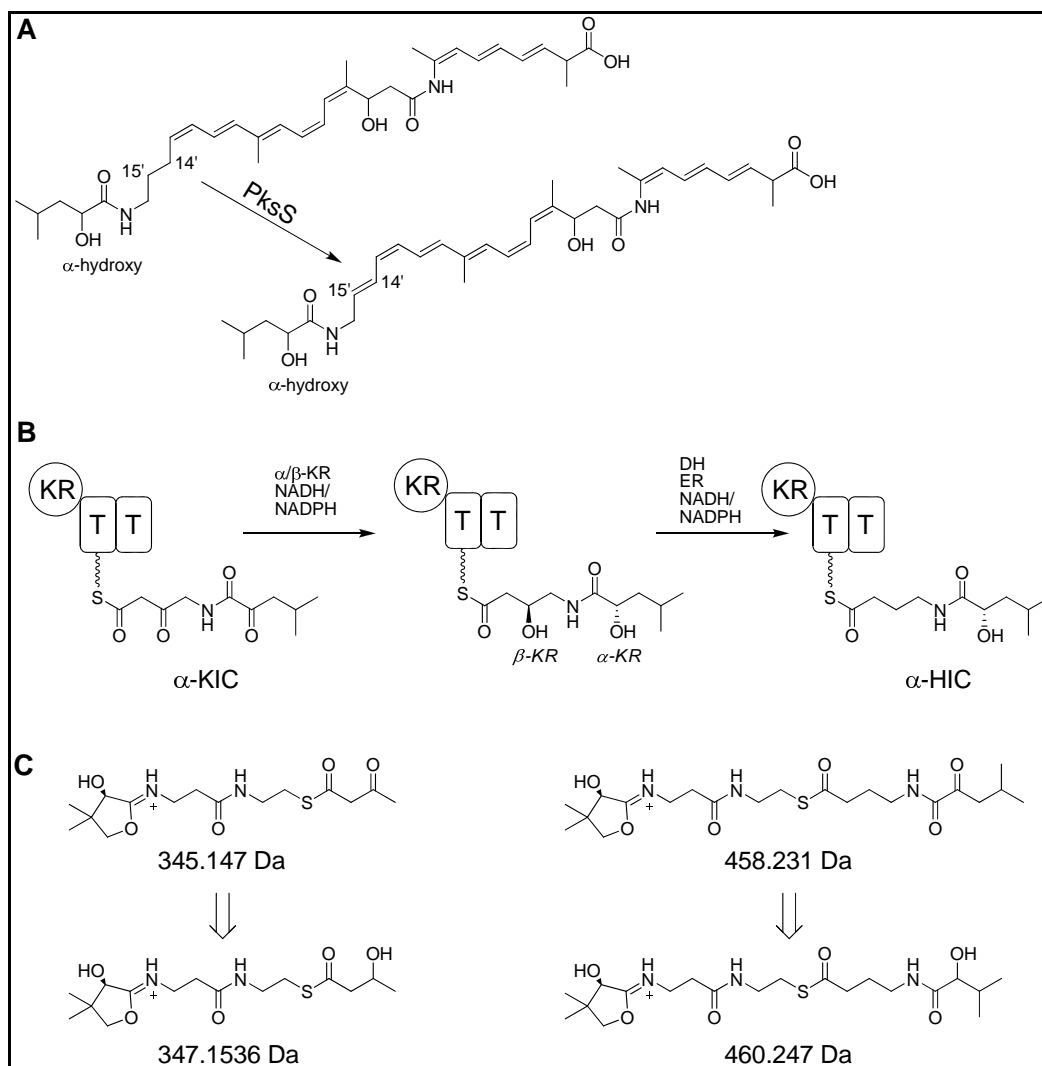
Figure 13: Bacillaene biosynthesis II. (A) Structure of dihydrobacillaene and bacillaene. PksJ oxidizes the C14'-C15' bond after dihydrobacillaene has been synthesized. Also, note the α-hydroxyacyl N-cap. This particular N-capping has been reported very rarely. B - α- and β-ketoreduction of α-KIC to α-HIC. The KR domain of the first PKS module in PksJ is capable of reducing both the α-KIC amide and the β-ketone in an NAD(P)H-dependent fashion. The order in which these two reductions occur is unknown. Ultimately, keto-reduction is followed by dehydration and enoyl reduction. C -Theoretical structure of PPant ejection ions used to analyze PksJ ketoreduction. (Right) PPant ejection ion resulting from IRMPD of Acac-*S*-PksJ(T₃-T₄) incubated with PksJ. Mass shift of +2.017 Da corresponds with reduction of the β-ketone. (Left) PPant ejection ion resulting from IRMPD of α-KIC-GABA-*S*-PksJ(T₃-T₄) incubated with PksJ. Shift of +2.015 Da is observed in PPant ejection ions.

As previously noted, there were several trans-acting elements of the dihydrobacillaene biosynthetic pathway. The exact function of one such element from the *pksX* gene cluster, PksE, was analyzed by FTMS.[49] PksE was proposed to act as enoyl reductase (ER), reducing the C14'-C15' bond during dihydrobacillaene biosynthesis by PksJ. To test this activity, 2-butenoyl-*S*-Pks(T_3T_4) was generated, incubated with NAD(P)H, digested and analyzed by LC-FTMS. An NAD(P)H-dependent reduction of the alkene bond was detected by a +2 Da shift of the active-site containing peptide of 2-butenoyl-*S*-Pks(T_3T_4) and could be confirmed by PPant ejection. The *pksX* gene cluster possesses many atypical features. Initially, FTMS proved to be useful in determining the function of several major components of the dihydrobacillaene biosynthetic pathway. More recent research has highlighted the usefulness of FTMS and the PPant ejection assay as a means of exploring noncanonical features of complex NRPS/PKS systems, in particular the function of ketoreductases and enoylreductases that impart small (2 Da) changes in the intermediate. The PPant ejection assay is a reliable method for verifying these slight changes.

2.2. Curacin A Biosynthesis: ECH1 & ECH2

The application of high resolution mass spectrometry has allowed characterization of enzymes involved in the biosynthesis of Curacin A, a mixed polyketide-nonribosomal peptide produced by *Lyngbya majuscule*, that possesses cytotoxic properties.[94] Researchers were intrigued by the unusual structure of this

molecule (Figure 14A) and during their investigations of the enzymes involved in Curacin A biosynthesis, an HMG-CoA synthase (HCS)-like gene cassette was identified, similar to that of the *pksX* gene cluster in *Bacillus subtilus*, as well as others. The HCS-like cassette, involved in Curacin A biosynthesis, encodes five separate enzymes that include a T domain, a ketosynthase, an HMG-CoA synthase, and two separate enoyl-CoA synthases.[95] This set of enzymes was hypothesized to be responsible for the formation of the cyclopropanyl ring of Curacin A. In particular, the functions of the two enoyl-CoA synthases ECH1 (CurE) and ECH2 (CurF) were determined using high resolution mass spectrometry.

As members of the functionally diverse crotonase superfamily, ECH1 and ECH2 were expected to have different roles in the formation of the cyclopropanyl ring precursor of Curacin A.[96] To probe the function of ECH1 and ECH2, the T domain (CurB) was first incubated with (S)-HMG-CoA, in order to covalently attach HMG to the phophopantetheine arm of the T domain active site. Next, the individual activities of each enzyme were determined by incubating ECH1, ECH2, or both enzymes with HMG-*S*-T. ESI-FTMS was used to analyze the products of these reactions by detecting the mass differences observed in the various reactions compared to HMG-*S*-T alone. The 12+ ion was used to detect these mass differences. The most abundant mass of HMG-*S*-T was determined to be 11325.8 Da. Incubation with ECH2 alone did not result in any new products. However, incubation of HMG-*S*-T with ECH1 resulted in the detection of a new product with a mass of 11307.8Da, and incubation of HMG-*S*-T with ECH1 and ECH2 yielded two products with masses of 11307.8Da and 11264.8Da. These differences observed by FTMS corresponded to losses of 18 and 62

Da. This provided evidence that ECH1 functions to dehydrate HMG-*S*-T (-18Da) to form 3-methylglutaconyl-*S*-T. The 3-methylglutaconyl-*S*-T undergoes subsequent decarboxylation, catalyzed by ECH2, to form 3-methylcrotonyl-*S*-T (Figure 32B). After identifying the exact functions of ECH1 and ECH2, researchers determined the crystal structure of the N-terminal domain ECH2 (CurF).[97] Structural alignments of CurF ECH2 with other members of the crotonase superfamily revealed several key features of the enzyme active site. Crystallization of ECH2 complexed with product analogues was not successful, so computational modeling was used to identify three polar side chains within the active site chamber that possessed potential catalytic function; Tyr82, Lys86, His240. The previously established ECH1/ECH2 enzymatic assay was carried out using ECH2 mutants containing Y82F, K86A, K86Q, H240A, and H240Q. The wildtype and mutant ECH2 enzymes were incubated with ECH1 and (S)-HMG-*S*-T (CurA-S-T(II)), then the different incubated reaction mixtures were run on a C4 column and eluted with acetonitrile. The samples were analyzed by ESI-FTMS after being redissolved in an electrospray solution (55% Acetonitrile: 45%Water, with 0.05% Formic Acid and 0.05%TFA). FTMS was used to detect the presence of 12+ charged (S)-HMG-*S*-T, 3-methylglutaconyl-*S*-T, and 3-methylcrotonyl-*S*-T as a means of confirming the effect of the site directed mutagenesis of the various residues. Using this ECH1/ECH2 assay and FTMS, it was determined that substitution of Tyr82 resulted in only minimal reduction in production of 3-methylcrotonyl-*S*-ACP from HMG-*S*-T, while substitutions of Lys86 and His240 resulted in drastic decreases in product formation. The identification of essential active site residues allowed researchers to propose a mechanism of action for ECH2 (CurF,

Figure 14C). The work carried out with the Curacin A biosynthetic pathway highlights the importance of FT-ICR-MS in both the determination of intermediates in NRPS/PKS pathways as well as evaluation of PKS domain functions.
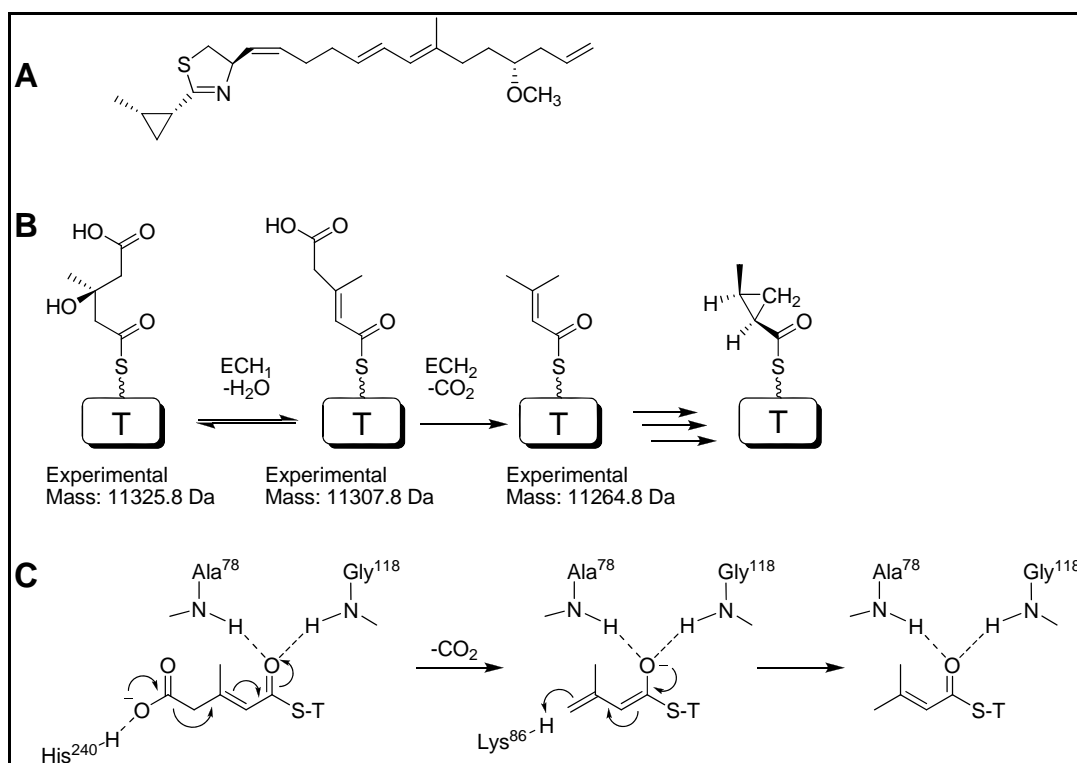
Figure 14: Curacin biosynthesis. (A) Curacin A structure. (B) T domain bound intermediates involved in cyclopropyl ring formation. 3-hydroxy-3-methylglutaryl (HMG)-*S*-T undergoes dehydration catalyzed by ECH1 to produce 3-methylglutaconyl-*S*-T. ECH2 catalyzes the subsequent decarboxylation to yield 3-methylcrotonyl-*S*-T. (C) Proposed mechanism of decarboxylation of 3-methylglutaconyl-*S*-T by ECH2. The His240 residue of ECH2 acts to position the substrate and prime its decarboxylation. Lys86 donate a proton to the enolate anion. Point mutations of these two residues substantially diminished the production of 3-methylcrotonyl-*S*-T, as detected by FT-ICR-MS.

2.3 Deconstructive Analysis of PksA

Recent efforts have been made to better understand eukaryotic iterative polyketide synthases (IPKS) and how these biosynthetic pathways ensure the formation of specific products. As opposed to much larger modular polyketide synthases that carry out the biosynthesis and tailoring of their products in an assembly-line fashion, iterative polyketide synthases contain many fewer catalytic domains and the same domains are used multiple times prior to yielding the final product. It has been unclear what exactly determines the number of catalytic cycles the growing intermediate of an IPKS undergoes before it is released. To clarify the "global division of labor" between the domains of an IPKS, researchers studied PksA by dissecting out the individual domains and reassembling them *in vitro*.[48] By expressing the various domains of the PksA as smaller units and recombining them in different combinations, the hope was to paint a better picture of the function of each domain in the biosynthesis of aflatoxin B_1. Various mono-, di-, and tridomains were expressed. The various domains of PksA were then mixed back with the PksA SAT-KS-malonyl-CoA:T-transferase (MAT) domain, and the different products of the reactions between the substrate, SAT-KS-MAT, and the other domains were analyzed. In theory, the full complement of the 6 PksA domains (Figure 15A) would be required in order to efficiently produce norsolorinic acid, an isolable precursor of aflatoxin B_1, from a starter hexanoyl-CoA and seven malonyl-CoAs.

The products formed by the various combinations of the deconstructed PksA domains, incubated with substrate, were monitored by HPLC. It was found that the

combination of SAT-KS-MAT with the putative product template (PT) domain and the T domain yielded significant quantities of a product that was not norsolorinic acid. Addition of the thioesterase/Claisen cyclase (TE/CLC) to the mixture resulted in substantial formation of norsolorinic acid. These findings revealed that product formation was not occurring to an appreciable extent in the absence of the PT domain. Therefore it was hypothesized that the PT domain acts as a cyclase/aromatase and catalyzes the formation of the first two rings of a putative intermediate. This putative intermediate can then undergo spontaneously C-O ring closure in the absence of the TE/CLC domain to form the shunt product napthopyrone, or it can undergo C-C ring closure in the presence of the TE/CLC domain to from the norsolorinic acid anthrone precursor (Figure 15B).

In order to verify the proposed structures of the phosphopantetheinyl-bound intermediates, an LC-FTMS assay was employed to detect PPant ejection ions containing bound intermediate. Reactions with SAT-KS-MAT and either the T domain or a PT-T didomain were carried out with both hexanoyl-CoA and malonyl-CoA. The reactions were carried out for varying times, followed by limited trypsin digestion (15 min). The digested reactions were run separated on a C4 column over a 1 hour water/acetonitrile gradient and injected directly into an FTMS. The phosphopantetheinylated active site peptides possessing various intermediates were observed and the structures were confirmed by MS/MS and PPant ejection. The fully extended intermediate generated by a single starter hexanoyl unit and seven malonyl units was detected bound to the T domain active site. In addition to this PPant bound intermediate, the intermediate was detected as a singly-dehydrated with a signal

aromatic ring or it was detected as a double dehydrated compound with two closed rings. The MS/MS PPant ejection ions reveal mass shifts with less than 1.5 ppm mass accuracy. These ejection ions were subjected to an additional round of fragementation and the resulting ions were mapped to the structure of the intermediate (Figure 15C). The T domain was observed to accept both hexanoyl-CoA and malonyl-CoA, but PksA preferentially loads hexanoyl-CoA as the starter unit. The MS/MS PPant ejection ions reveal mass shifts with less than 1.5 ppm mass accuracy. The high mass accuracy of the FTMS allowed for the accurate determination of the mass of PPant bound intermediates, even for relatively large intermediates such as the full extended octa-ketide intermediate of norsolorinic acid.
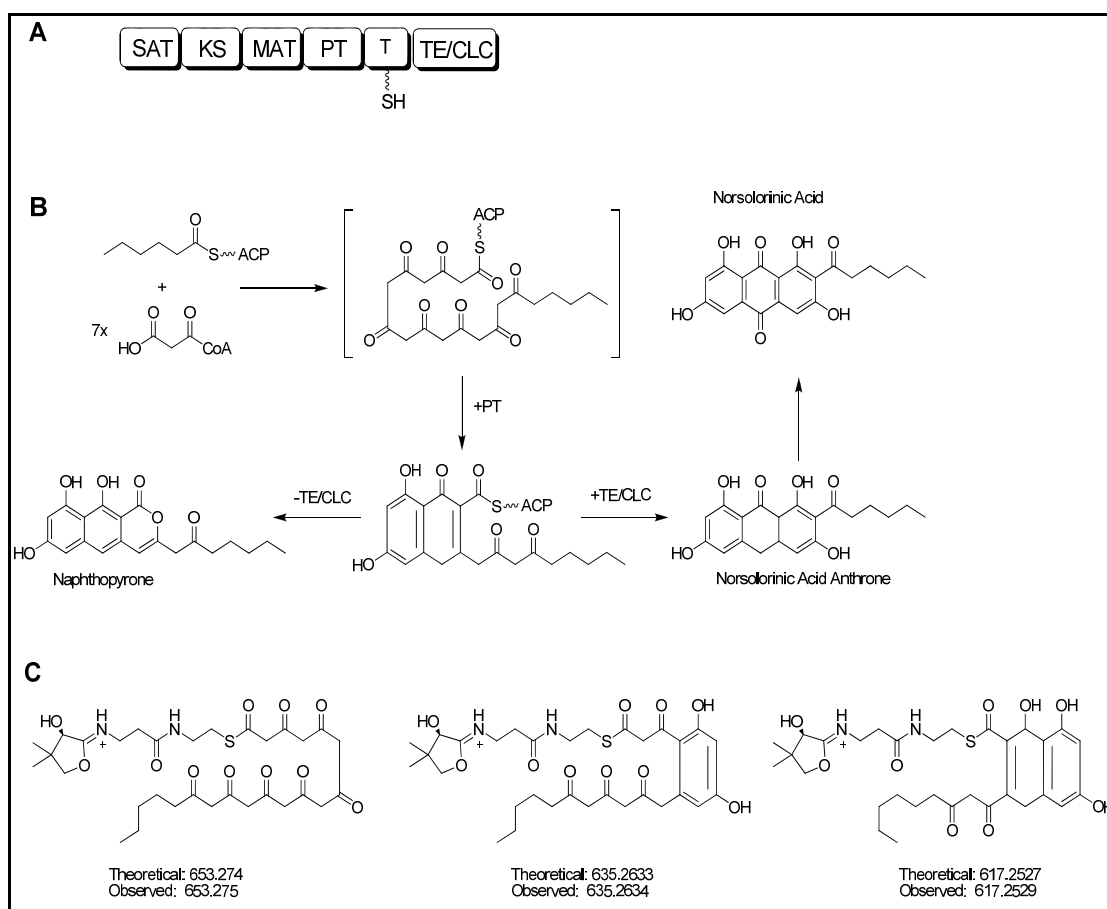
Figure 15: PksA deconstructive domain function analysis. (A) Enzymatic domain architecture of PksA. (B) PksA utilizes a starter hexanoyl-CoA and 7 malonyl-CoAs to produce the covalently linked intermediate (brackets). The PT domain acts as an aromatase/cyclase facilitating the closure of the first two rings on the intermediate. In the absence of the TE/CLC domain the intermediate undergoes C-O cyclization to spontaneously form the naphthopyrone. In the presence of the TE/CLC domain, the intermediate undergoes C-C cyclization to from the norsolorinic acid anthrone, which auto-oxidizes to form norsolorinic acid. (C) Observed PPant ejection ions confirming the structures of the proposed intermediates bound to the active site of the PksA T domain. The first intermediate (left) was detected on the T domain active site after incubation of SAT-KS-MAT with T domain alone. Incubation of SAT-KS-MAT with PT-T results in the formation of the intermediates containing first a single-cyclization product (middle) followed by a double-cyclization product (right).

Chapter 3: Investigations of C-1027 Enediyne Core Biosynthesis by the Type I

Iterative Polyketide Synthase SgcE

3.1. Introduction to Enediynes

The enediyne family of antibiotics is comprised of some of the most potent antitumor natural products currently known. The potency of these compounds can be attributed to the unique chemical structure of their highly enediyne core. The core chromophore unit contains two acetylenic groups conjugated to a double bond within a 9-membered ring (e.g. maduropeptin, or neocarzinostatin) or a 10-membered ring (e.g. calicheamicin) (Figure 16). The enediyne chromophore is capable of inducing single-strand or double-strand cleavages of a target cell's DNA.[98] The extremely cytotoxic characteristic of enediynes has generated a considerable amount of interest in harnessing these natural products as potential anticancer drugs.[99] In particular, a neocarzinostatin-polymer conjugate[100] and a calicheamicin-antibody, in which the antibody specifically targets cancer cells[101], are currently in clinical use. While synthesis of some endiynes has been carried out, there has been substantial research into the biosynthetic origin of this class of compound. The enediyne core bears virtually no structural similarity to other known polyketides, however efforts to clone and characterize the biosynthetic gene clusters that code for various enediynes revealed that all of these compounds are, in fact, synthesized by dedicated iterative polyketide synthases.[35,102,103] To date, five endiyne biosynthetic gene clusters have been sequenced[104], all of which encode an iterative polyketide synthase (PKSE), with high sequence homology, which is responsible for the *de novo* synthesis of the enediyne core. Analysis of the domain architecture of each sequenced PKSE indentified that each one contains a acyl carrier protein, ketosynthase, acyltransferase, ketoreductase, and dehydratase domain. In addition, the enediyne PKS has two unique

features that differentiate it from other iterative polyketide synthases. First, the ACP domain of PKSE bears no homology to any other proteins, and the position of the ACP between the acyltransferase and ketoreductase domains resembles the domain arrangement of polyunsaturated fatty acid (PUFA) synthases.[105] However, the most striking feature that sets PKSE is the presence of a C-terminal phosphopantetheinyl transferase (PPTase) which the PKSE utilizes to phosphopantetheinylate its own ACP active site, which is quite unusual[106], as virtually all other characterized PKSs are known to utilize discrete PPTases.
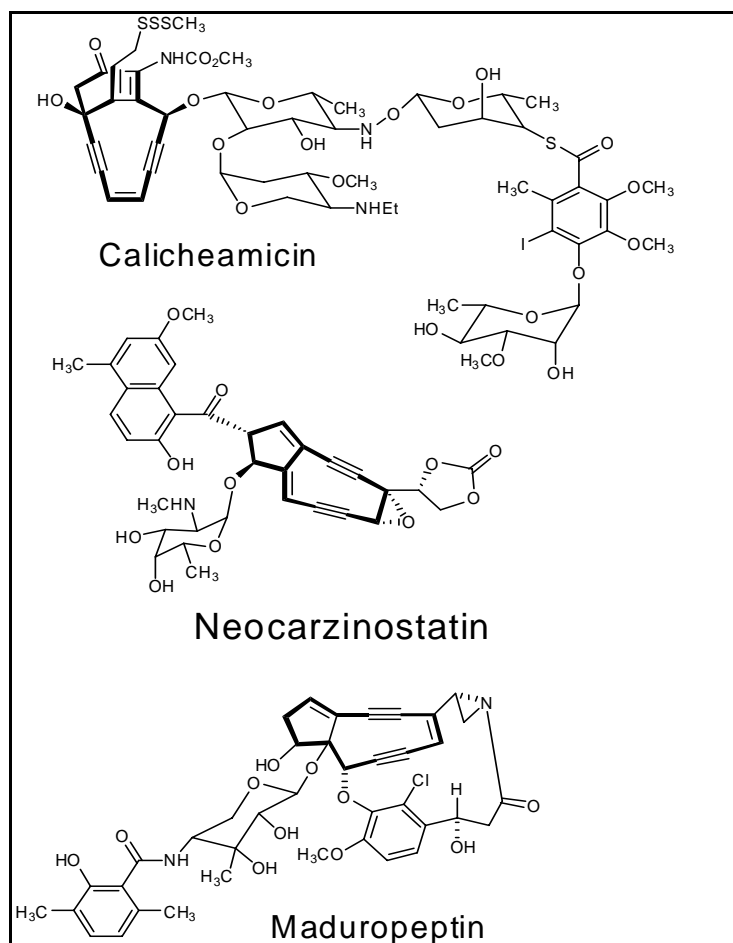
Figure 16: Selected enediyne structures. Shown are the structures of 10-membered enediyne calicheamicin and 9-membered enediynes neocarzinostatin and maduropeptin

The enediyne natural product, C-1027, is a produced by *Streptomyces globisporous* as a chromoprotein complex consisting of the reactive chromophore, which of four distinct chemical units, bound non-covalently to an apoprotein, CagA. The biosynthesis of the C-1027 chromophore is results from the union of a deoxy aminosugar, β-amino acid, benzoxazolinate, and the enediye core (Figure 17). While the function of each enzyme involved in the biosynthetic pathways for each of the four components are not fully characterized, it is known that each component is ultimately synthesized from the basic metabolites *D*-glucose-1-phosphate, chorismic acid, L-α-

tyrosine, and acyl-CoAs[69].   As with all other known enediynes, the antitumor properties of C-1027 have been attributed to the enediyne core and its ability to induce double-stranded breaks in DNA.[107]   Once delivered to the host cell's DNA in the nucleus, C-1027 is released from CagA and the enediyne core will spontaneously undergo Bergman cycloaromatization[108] which generates a highly reactive 1,4 benzenoid diradical (Figure 18) that sits in the minor grove of DNA and strips hydrogen atoms from the deoxyribose backbone of DNA, resulting in double-stranded breaks or interstrand cross-links.   The already potent C-1027 demonstrates an even higher cytotoxicity in hypoxic cells[109], unlike other enediyne antitumor compounds whose cytotoxicity decreases in low oxygen environments, such as those in solid tumors.   While these findings indicate that the power of C-1027 can potentially be harnessed for chemotherapeutic use, in its natural state C-1027 is far too cytotoxic to all cells to be used clinically.
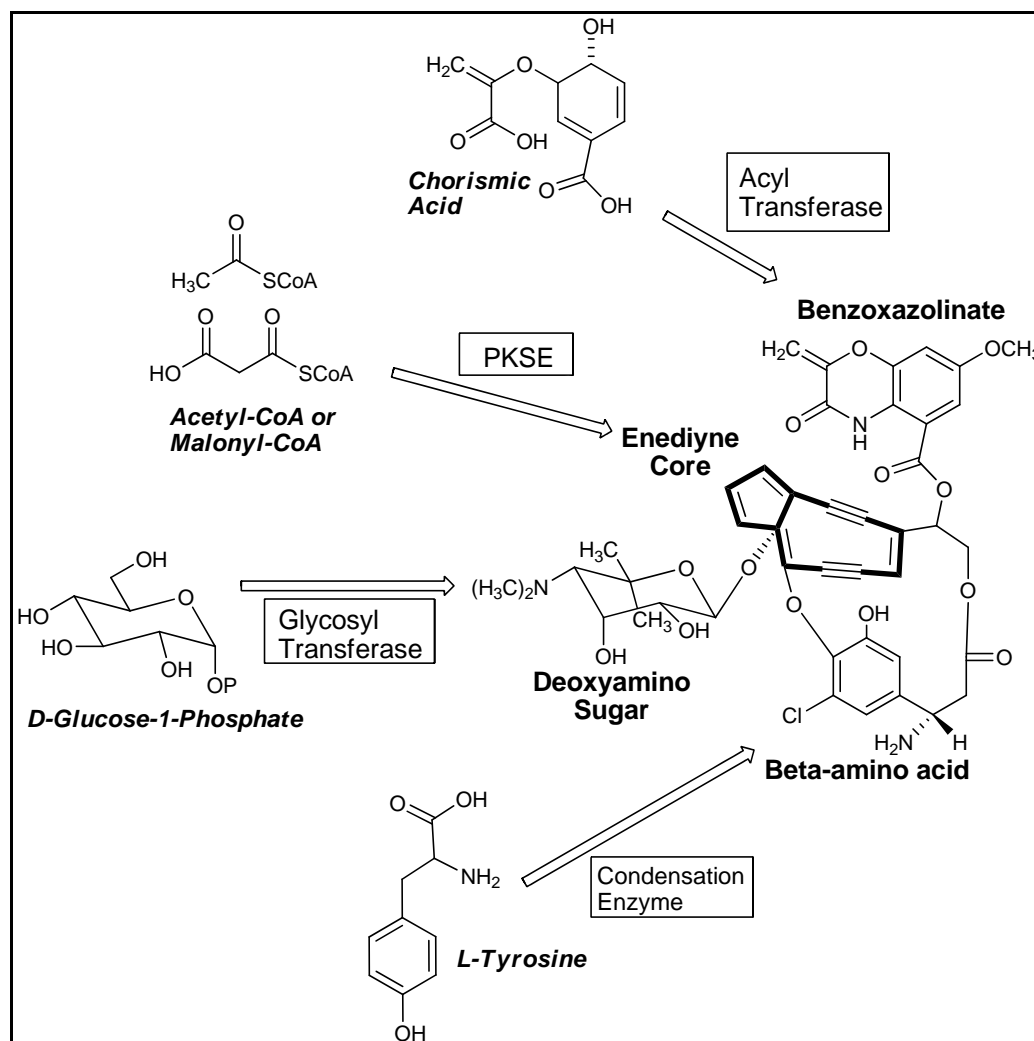
Figure 17: Biosynthesis of C-1027 chromophore. The convergent biosynthesis of C-1027 chromophore from four separate subunits. The core endiyne is in bold. Each subunit is biosynthesized by multiple enzymes.

While the enediyne core acts as the "warhead" of C-1027, the other components of the compound are important for its biological activity. The benzoxazolinate component of the chromophore serves dual purposes: it anchors the chromophore to its carrier apoprotein CagA[110], and once the chromophore is released the benzoxazoline is critical for situating the enediyne core in the minor grove of DNA.[111] Structural variations of the benzoxazoline moiety and the β-amino acid

moiety have been found to significantly reduce the cytotoxicity of C-1027 analogues.[112] Novel C-1027 analogues have been the result from manipulations of the biosynthetic pathways that carry out the synthesis of each of the four subunits. Characterization of the set enzymes involved in the step-wise biosynthesis of the β-amino acid moiety has allowed investigators to induce mutations to inactivate specific enzymes in the pathway, resulting in 20-deschloro-C-1027 and 20-deschloro-22-deshydroxy-C-1027 analogues[113], in addition to the 22-deshydroxy-C-1027 analogue first identified.[35] As significant progress has been made in recent years in characterizing the biosynthesis of the β-amino acid moiety, as well as the benzoxazolinate moiety[114], less is known about the assembly of the enediyne core itself. A more detailed knowledge of the biosynthesis of the enediyne core of C-1027 by its PKSE will perhaps lead to additional avenues by which the potency of C-1027 can be altered in order to make it a clinically useful antitumor antibiotic.
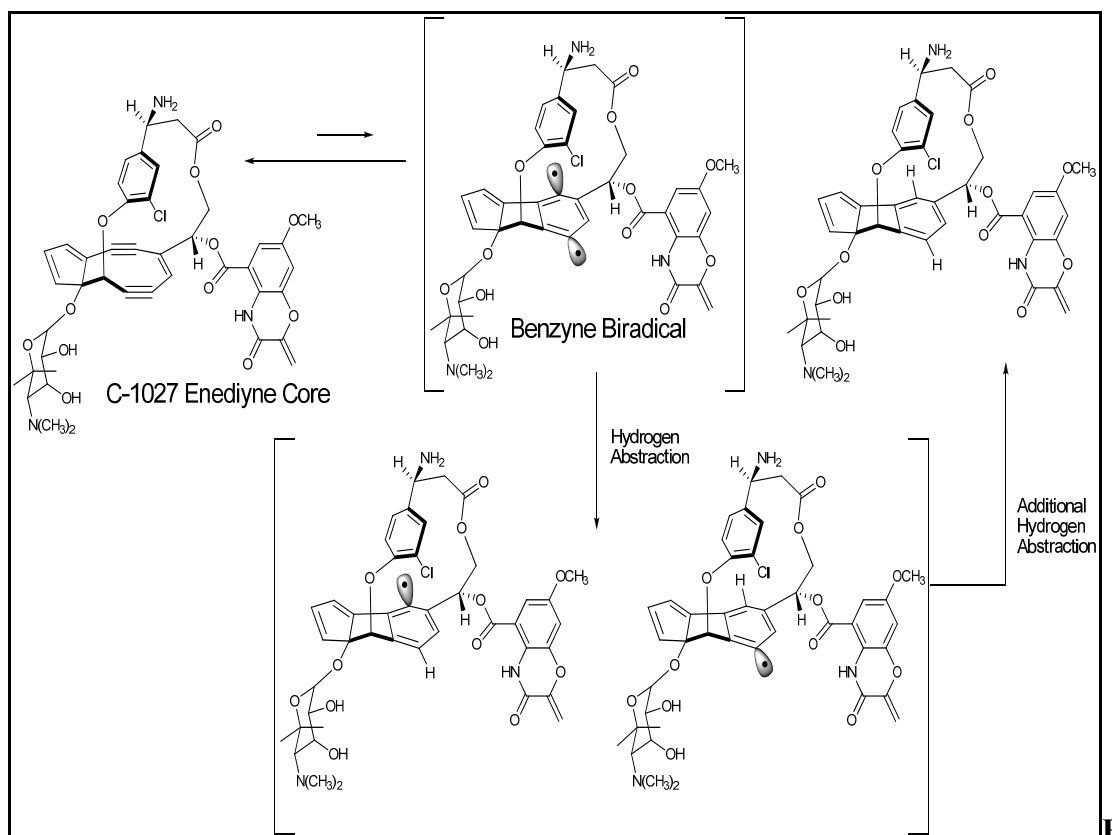
Figure 18: C-1027 di-benzyne radical formation. The 9-membered enediyne core of C-1027 spontaneously undergoes Bergman cycloaromatization to form a highly reactive benzyne biradical capable of abstracting hydrogen atoms from DNA.

In 2002, the biosynthetic machinery for the production of C-1027 and calicheamicin were first presented in the literature, and it was revealed that the gene clusters for each contained a single gene encoding a polyketide synthase. At the time, the product of this gene was a polyketide synthase unlike any other known bacterial PKS in that it only possessed five domains: KS, AT, KR, DH, and then-unidentified C-terminal domain, classifying the PKSE as an iterative type I PKS. At the time iterative type I polyketide synthases were the expected the paradigm for polyketide biosynthesis in fungi, while a type I iterative PKS in bacteria was considered much more rare.[115] In *Streptomyces globisporus*, researchers designated the PKSE as SgcE,

and determined it to have the characteristic domain architecture of a type I iterative polyketide synthase.[99] Prior to cloning and characterization of the C-1027 and calicheamicin gene clusters and, it was unknown whether or not the enediye core resulted from de novo biosynthesis or was the result of degradation of a fatty acid precursor.[116] ^{13}C-labeled feeding studies had provided confirmation that the enediyne core of neocarzinostatin, dynemicin A, and esperamicin A all originated from acetate precursors but did not conclusively prove that the enediyne core was in fact directly produced by PKS.

To verify that SgcE was responsible for the *de novo* biosynthesis of the enediyne core of C-1027, researchers sought to first inactivate SgcE by mutating the *sgce* gene so as to inactivate the KS active site of the enzyme. This *Δsgce* mutant form of *S. globisporus* produced an inactive form of SgcE that was incapable of producing C-1027. When a plasmid containing a functional copy of the *sgce* was introduced into the *Δsgce* mutant, the normal synthesis of C-1027 was restored. This experiment provided strong evidence that the SgcE was critical for the synthesis of C-1027, therefore the enediyne core was shown to be of PKS origin. Subsequent similar research involving other enediyne gene clusters have given results that helped confirmed that all enediynes utilize a dedicated PKSE to synthesize their 9- and 10-membered core.

After the functional role of SgcE was confirmed to be that of a PKS, additional work was carried out to map the active sites of the enzyme *in vivo.* Investigators individually introduced point mutations at each of the proposed active site residues of SgcE, and then introduced these SgcE-mutant plasmids, under the control of a

constitutive promoter, into the *ΔsgcE* mutant *S. globisporus* and looked at the ability of these mutant plasmids to restore C-1027 production.[54] From these experiments, the active site residue of each of the catalytic domains was confirmed when a given point mutation resulted in the mutant plasmid being unable to restore C-1027 production. Mutation (to alanine) of the active-site cysteine (Cys211) of the KS domain and mutation of the active-site serine (Ser659) of the AT domain resulted in plasmids that were unable to restore C-1027 in the *ΔsgcE* mutant. Prior to these plasmid complementation experiments, the catalytic role of the domain located downstream of the AT domain was hypothesized to be that of an ACP[35] and the location of the 4'-phosphopantetheinyl attachment site. Mutation of one particular conserved serine (Ser974) generated a mutant plasmid that was incapable of restoring C-1027 production in the *ΔsgcE* mutant. Having confirmed that SgcE contained an ACP domain that was necessary to function, it was logical then to assume that the C-terminal domain, that shared similarities to the phosphopantetheinyl transferase Sfp from *Bacillus subtilis*, did in fact function as a PPTase. Subsequent plasmid complementation experiments confirmed that mutation of the predicted sites of Mg^{2+} coordination, Asp1827 and Glu1829, eliminated biosynthesis of C-1027, indicating that this domain was functional *in vivo* and required for phosphopantetheinylation of the ACP of SgcE.

Site-directed point mutations provided strong evidence allowing investigators to map the active site of SgcE *in vivo*. In order to provide even more conclusive evidence that the domain designated as the ACP of SgcE was in fact that, recombinant SgcE was heterologously expressed in E. coli and analyzed by mass spectrometry[54].

The protein was subjected to digestion by trypsin, HPLC fractionation of the proteolytic fragments, then analysis by ESI FT-MS. A peptide fragment was identified by mass spectrometry as having a mass consistent with that of the peptide containing Ala953 through Arg987 (3,793.90 Da) plus the additional mass of phophopanteteine (340.1 Da), giving a final observed mass of 4,134.00 Da (calculated mass: 4,134.08 Da). Collisionally induced dissociation of the peptide localized the site of the modification to five residues, one of which was Ser974 of SgcE. Subjecting this particular peptide to infrared multiphoton dissociation (IRMPD) resulted in observation of the PPant ejection ion with a mass of 261.129 Da and the two resulting forms of the parent peptide; the dehydroalanine (3,775.82 Da) peptide resulting from elimination of the entire PPant moiety and subsequent loss of water, and the phosphoserine (3,873.78 Da) peptide, which retains the phosphate portion of the phosphopantetheinyl modification. This data clearly demonstrated that the active-site residue of the ACP domain of SgcE is phosphopantetheinylated, consistent with the proposed function of SgcE as a PKS responsible for the biosynthesis of the enediyne core of C-1027. In addition to verification of the presence of the phosphopantetheinylated active site of the ACP domain of SgcE, mass spectrometry was used to confirm the function of the C-terminal domain as a PPTase by monitoring that PPant mass shift of the ACP active site fragment did not occur when the Mg^{2+} coordination residues of the PPTase domain were mutated. Clearly, mass spectrometry is a valuable means for verifying the catalytic function of SgcE and its involvement in biosynthesis of the enediyne core of C-1027.

As it was demonstrated that the heterologously expressed SgcE contained both a functional PPTase and an ACP domain capable of being phosophopantetheinylated, investigators assumed that the enzyme should be functional in its heterologous host. Indeed, the heterologously expressed SgcE when purified was bright yellow in color, something which could not be removed by protein denaturation, indicating that the source of the color was a convalently bound to the protein. UV/Vis spectroscopy showed multiple absorption maxima between 350 and 450 nm, which is a profile similar to that of a polyene.[117] Initially, investigators were unable to isolate sufficient quantities of SgcE-bound product, which was predicted to be 3-hydroxy-4,6,8,10,12,14-hexadecahexaene, for spectroscopic analysis. Therefore, *sgcE* was co-expressed in *E coli* with the gene *sgcE10*, which was identified as a thioesterase (TE) within the C-1027 gene cluster. The TE was expected to hydrolyze the ACP-bound polyketide intermediate allowing isolation and characterization of the SgcE/E10 product. In fact, the SgcE/E10 product was successfully isolated and its structure determined to most likely be 1,3,5,7,9,11,13-pentadecaheptaene based on a combination of HPLC-APCI-MS, ^{1}H NMR, ^{1}H COSY, and ^{13}C NMR experiments.

3.2. SgcE Research Goals

While the structure of the pentadecaheptaene isolable SgcE/E10 intermediate was elucidated by a variety of methods, the SgcE ACP-bound polyene intermediate was not isolated in sufficient quantity or quality for further characterization. However, since the intermediate is covalently tethered to the 4'-phosphopantetheinyl moiety of the ACP, it is an attractive target for analysis by high resolution mass

spectrometry by implementing the PPant ejection assay. In addition, the pantetheinylated ACP active-site containing tryptic peptide had been previously identified. Previously mass spectrometric analysis resulted identification of the 3+ charge state of the ACP active-site peptide. The goal of subsequent mass spectrometric analysis carried out in the Dorrestein Laboratory was to search specifically for the 4+ charge state of the 4,134.08 Da peptide ACP active-site peptide, plus mass shifts of this peptide corresponding to the addition of the predicted masses various intermediates in the biosynthesis of the enediyne core of C-1027. Using this hypothesis driven approach, in which a peptide of a known mass and predicted charge state, additional information regarding was the biosynthesis of the 9-membered enediyne core by the PKSE SgcE was collected.

3.3. SgcE Materials and Methods

3.3.1. Sequence of Information for Investigated Protein(s)/Peptide(s)

Heterologously expressed, purified SgcE was generously provided by Ben Shen of the University of Wisconsin. The exact methods for the generation of the E. coli expression constructs can be found in the supplemental information of reference 54. The gene for *sgcE* was amplified by PCR using pBS1006[118] as a template. Reactions were performed using the following primer for sgcE:

(forward) 5'-GGTATTGAGGGTCGCATGAGCCGCATAGCCATCGT-3'

(reverse) 5'-AGAGGAGAGTTAGAGTCACGCGCGGGCGCTCC-3'

The gel-purified PCR product was inserted into pET-30 Xa/LIC using ligation-independent cloning as described by Novagen (Madison, WI) affording plasmid pBS1050

Purified Holo-SgcE with His-tag – (MW 208,243.14 Da)

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGGGSGIEGRMSRI
AIVGVACTYPDATTPRELWENAVAGRRAFRRLPDVRMRLDDYWNPDPTVPD
TFYARNAAVLEGWEFDRVAHRIAGSTFRSTDLTHWLALDTATRALADAGFPA
GEGLPTERTGVVVGNTLTGEFSRANGLRLRWPYVRRILADALQEQEWDDDRL
GAFLRGVEEAYKKPFPAVDEDTLAGGLSNTIAGRICNHFDLNGGGYTVDGAC
SSSLLSITTAATSLQSGDLDVAVAGGVDLSIDPFEIIGFAKTGALARKEMRLYD
RGSNGFWPGEGCGMVVLMREEDAVASGHRIYASIAGWGISSDGQGGITRPEV
SGYQLALSRAYDRAGFGIETVPLFEGHGTGTAVGDATELRAIMSARAAADPH
APSAVITSIKGMIGHTKAAAGIAGLIKAVMALDSGVLPPAIGCVDPHDLLTDES
ANLRVLRKAESWPENAPLRAGITAMGFGGINTHVVLDRSDASGRRPAVNRRT
TLLANSLQDSELLLLDGESPAALARRLTQVADFAAQVSYAQLGDLAATLQRE
LRDLPHRAAVVATSPEDAELRLRGLAETAGGRAPDDGPVFSQDGRAFLGTAA
EGARVGFLFPGQGSGTSTAGGALARRFTEAAEVYARAGLPTAGDMVATHVA
QPRIVTGSTAGLRVLEALGIEADIALGHSLGELSALHWAGALDETTLLEAART
RGAAMAAHSASGTMASLTATPEEAVRLVEGLPVVISGYNGPRQTVVAGTVE
AVESVGERAAAAEIAFTRLAVSHAFHSPLVAPAAESFGDWLAKAPLGGLGRR
VVSTVTGAELERDTDLAKLLRQQITDPVLFTQAVRAAAAEVDLFVEVGPGRV
LSVLAAETAGKPAVALNTDDESLRGLLQVVGAAFVIGAPIIHERLFNDRLTRPL
EVGKEFLFLSSPCEQAPEFTLPAAAREPLVQEHDAPTTAGAGDTAEESALDVL
RALVAERAELPSELIDENSSLLDDLHMSSITVGQIVNQTAVRLGLAPSSIPTNFA
TATLAELASALTTLVETGADPTAAPVVTGSAAWARPFSVDLDELPLPPAVADE
KDGTWELFTSADHPFAEEVRRALQDAAVGSGVLVCLPAGCSPDQLELALDGA
RSALAGSQEGRFVLVQHDRGAAGLAKTLHLEAPHLRTTVVHTPVADGAADR
VAAEVAATTHFSEVHLDRDGTRRVPVLRALPFAPDRTDQVLGPDDVLLVTGG
GKGITAECALAVAERTGAALAVLGRSDPGSDQDLAANLGRMRESGIRVAYA
AADVTDPVRVAGAVAELTGALGSVTAVLHGAGRNEPTALGGLDMAAVRSTL
APKVDGLRHVLDVVGEQNLRLLVTFGSIIGRAGLRGEAHYATANEWLAGLTE
DVARRNPDCRALCMEWSVWSGVGMGEKLSVVESLSREGIVPVSPDQGIEILL
RLISDPDAPVVTVISGRTEGIGTVRREQPPLPLLRFTGEPLVRYHGVELVTEAEL
NAGTDLYLTDHMLDGNLLLPAVIGMEAMVQVGSAVTGRRDVPVIEDARFLR
PIVVPPGGTTRIRIAATVTGTDRVDVAVHAQDTGFAAEHFRARLVYGGAAIPD
GAPDQVGPKVPTAPLDPATDLYGGVLFQGERFQRLRRFHRAAARHVDAEVA
LDTASGWFAGFLPGTLLLSDPGMRDALMHGNQVCVPDATLLPSGIERLYPMA
AGEDLPELVRYCATERHRDGDTYVYDIAVRTPDGSVVERWDGLTLHAVRKS
DGSGPWVAPLLGSYLERTLEEVLGTHVDVAVEPVPADSGGSVADRRKATAR
AVQRALGESVKVRYRPDGRPELDGVRRLSAAHGPGVTLGVVGTTTVACDIEA
VTARGAQEWEGLLGEHGNLAALVAKETGETPDHAATRVWTAVECLKKAGL
PAGAPLTLEPQVRSGWIVLTAGGLRIATFATTLRHVEEPVVLAFLTAGTDDAA
PGSARA

Expected SgcE Acyl Carrier Protein (ACP) Active-Site Containing Typtic Peptide:

AELPSELIDENSSLLDDLHMS**S**ITVGQIVNQTAVR

Expected SgcE Acyltransferase (AT) Active-Site Containing Tryptic Peptide:

VLEALGIEADIALGH**S**LGELSALHWAGALDETTLLEAAR

3.3.2. SgcE Sample Preparations and Tryptic Digestion Conditions

SgcE samples provided were determined by Bradford to have a concentration of approximately 3.73 mg/ml (17.8 μM). 25 μL of soluble SgcE was diluted in 25 μL of 200mM Tris-HCl pH 8.0 (TEKnova Inc - Hollister, CA) to bring the final sample volume of 50 μL containing ~93.3 μg of *holo*-SgcE. For samples incubated with malonyl-CoA (Sigma-Aldrich - St. Louis, MO), 25 μL of soluble *holo*-SgcE was diluted in 15 μL buffer (5 μL of 1 mM Tris-HCl pH 8.0 and 10 μL Milli-Q H_2O). SgcE was incubated for 30 minutes with 10 μL of 5mM malonyl-CoA (Sigma-Aldrich), giving a final concentration of 8.9 μM SgcE mixed with 1mM malonyl-CoA. 50 μL of these samples were then digested using Trypsin Singles™ containing 1 μg of proteomics-grade trypsin (between 1:20 to 1:100 trypsin:protein is optimal) in 49 μL of 20mM NH_4HCO_3 (pH 8.2) reaction buffer provided with Sigma-Aldrich Trypsin Singles™ kit. Digestions were quenched after 10 minutes with an equal volume (100 μL) of 10% formic acid (Acros Organics- Belgium) diluted in HPLC grade water (Sigma-Aldrich).

Additional samples containing malonyl-CoA and NAD(P)H were prepared in with the aim of attempting to increase the quantity of fully-elongated polyketide intermediate covalently-bound to SgcE. Mixtures of 200 μL were prepared from which 50 μL was taken and digested at 3 hr (sampled twice), 6 hrs, and 12 hrs of incubation. The 200 μL reaction mixtures were prepared to contain 160 μg of soluble *holo*-SgcE, 600 μM NAD(P)H (Sigma-Aldrich), 1 mM malonyl-CoA, and 50mM Tris-HCl pH 8.0. The 50 μL samples withdrawn were subjected to the same digestion conditions noted above.

3.3.3. Capillary Column Preparation and Packing

Deactivated fused silica (Agilient, San Jose, CA) with an internal diameter of .100mm was used to make the capillary columns for capillary LC-MS. The fused silica was cut using a P-2000 CO_2 Laser Based Micropipette Puller (Sutter Inst. Co. - Novato, CA) to produce pointed tips at one end of the pre-but fused silica with an aperture of in sufficient diameter to allow for nanoflow (100-1000 nL/min). A custom-made steel packing bomb is used to pack reverse phase C-18 chromatographic 200Å resin-coated 5μm beads (Michrome Biosources – Auburn, CA) into the pulled capillary column. The C-18 resin is suspended in 100% methanol then pumped into the capillary at 700 PSI until approximately 10 cm of C-18 coated beads are firmly packed into a 15 cm pulled capillary column.

3.3.4. Active Site Mapping and Substrate Screening By Capillary LC-MS

Each digested sample was first diluted to a concentration of 100 ng/μL, and then loaded into 96-well plates. During each run, 10 μL of each sample was automatically aspirated from the 96-well plate and injected into a 20 μL sample loop for LC-MS analysis using a Thermo Surveyor Autosampler/MS Pump stack (Thermo Finnigan -San Jose, CA). Samples were first loaded onto an LC-MS capillary column (described above) by pumping 95% water/5% acetonitrile through the sample loop (non-split flow) for 5 min at a rate of 20 μL/min. Proteins/peptides were then eluted from the LC-MS capillary column by applying a water/acetonitrile gradient, as outlined in Table 1 below, while applying a voltage of 1.4 - 1.8kV to the capillary in order to generate an the electrospray ion stream, in which the analyte was completely ionized into the gas phase, that was then introduced directly into the mass spectrometer. The digested SgcE mixtures were analyzed using a Thermo Finnigan 6.43T LTQ-FT mass spectrometer equipped with a dual front-end quadrupole octupole linear trap. Samples were analyzed with methods that were programmed within the Thermo's Xcalibur software, in which peptides/proteins eluted from the capillary column were first detected by FT-MS set at a resolution of 50,000 to first collect high-resolution spectra of the intact peptides/proteins. Next the 5 most abundant ions from each FT-MS scan were isolated in the +/-2.5 *m/z* windows and all ions in that range were fragmented by CID, with an activation Q of 0.200 to generate the MS/MS spectra. Those precursor peptides/proteins that were selected for isolation for fragmentation were then added to an exclusion list that tracked the previous 25 that

were fragmented during the 30 seconds prior to each FT-MS scan and excluded these ions from repeated rounds of fragmentation. All spectra were analyzed using Qualbrowser (Thermo).

Table 1: Capillary LC-MS gradient for SgcE analysis.

Min	0	3	5	8	20	21	51	54	57	57.1	90
%A	95	95	95	95	95	95	40	20	20	95	95
%B	5	5	5	5	5	5	60	80	80	5	5
Flow Rate (ul/min)	20	15	10	1	150	150	150	150	150	150	150

Solvent A is water with 0.1% TFA and solvent B is acetonitrile with 0.1%. TFA.

3.4. SgcE Analysis Results

Based on the domains present within SgcE one can predict the initial steps of enediyne biosynthesis and the structures of the intermediates covalently bound to the ACP active-site PPant moiety. As with most known type I iterative polyketide synthases, SgcE contains the minimal catalytic domains: ketosynthase, acyl transferase, and acyl carrier protein. In addition, the presence of both ketoreductase and dehydratase domains indicate that the polyketide intermediate is highly reduced and dehydrated. One can envision initial loading of acetate (or malonate followed by decarboxylation) followed by Claisen condensation to yield an acetoacetate intermediate, which is reduced in the presence in NADH or NADPH, and dehydrated. Six additional rounds of Claisen condensation follow in an iterative fashion with concomitant ketoreduction and dehydration to yield the hypothetical 3-hydroxy-4,6,8,10,12,14 hexadecahexaene intermediate (Figure 19).
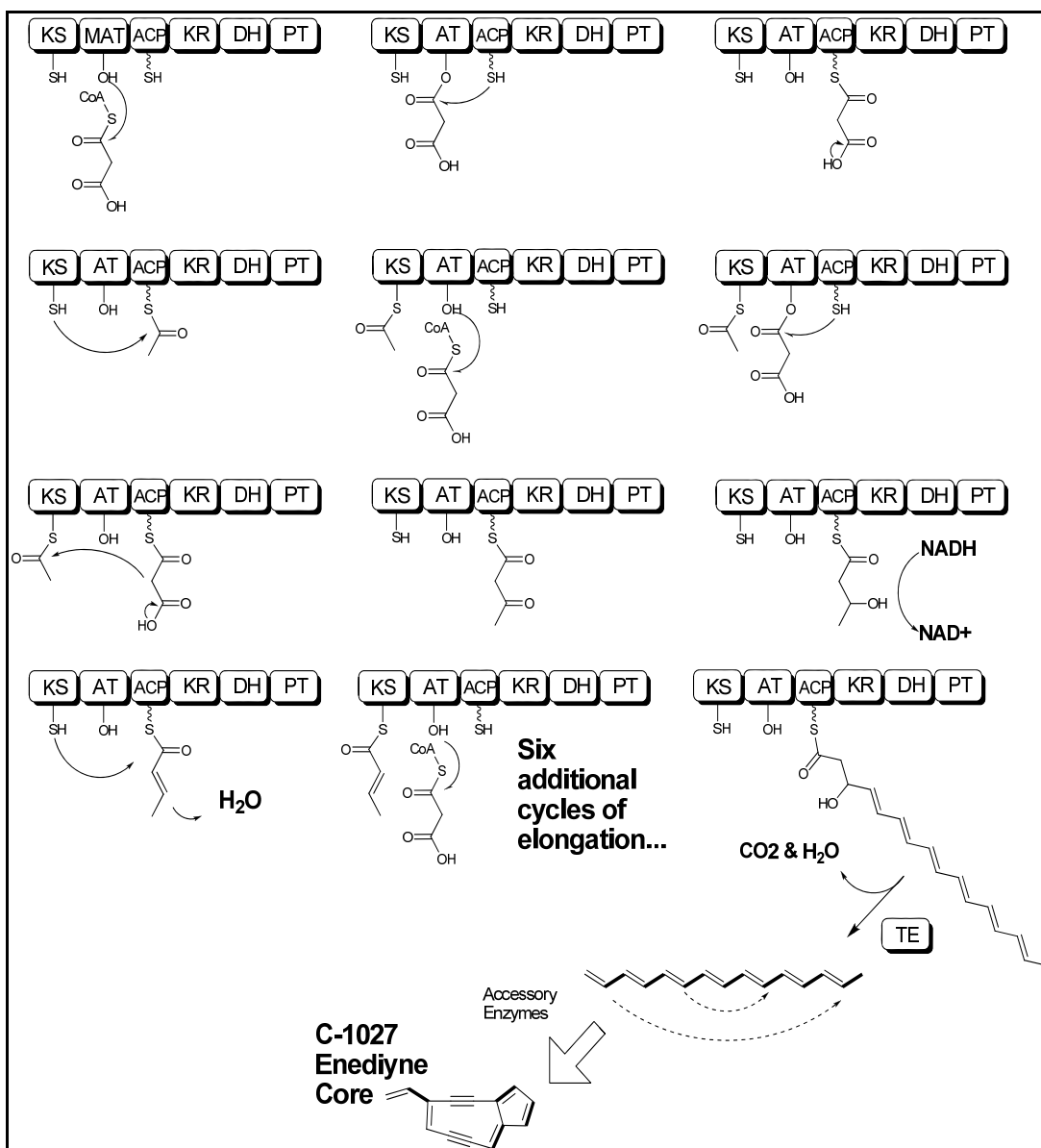
Figure 19: Mechanism of C-1027 enediyne core biosynthesis by SgcE. SgcE can accept malonyl-CoA or acetyl-CoA as a starter unit. The enzyme then carries out seven iterative cycles of decarboxylative condenstations between acyl units and followed by successive reduction of the β-keto position and dehydration until the fully-elongated 3-hydroxy-4,6,8,10,12,14 hexadecahexaene intermediate is formed. A dissociated thioesterase domain catalyzes release of the intermediate to yield an isolable intermediate 1,3,5,7,9,11,13-pentadecaheptaene. Subsequent formation of the cyclized enediyne may be carried out by accessory enzymes.

Initial analysis of the capillary LC-MS spectral data was aimed at attempting to locate 3+ and 4+ charge states of the ACP active-site fragment previously identified by mass spectrometric analysis. Within the entire mixture of eluted peptide both the *apo-* and *holo-* forms of the active-site peptide, 3793.90 Da and 4133.00 Da respectively, were first searched by their calculated *m/z* (Table 2). Using Qualbrowser, the entire peptide elution profile could be filtered to show only the peptides in the FT-MS scans that had an *m/z* close to that of the peptides of interest. The advantage to using FT-MS for the precursor scans during LC-MS was that the high resolution spectra obtained showed the individual isotopes of each of the candidate peptides. By calculating the inverse function of the m/z difference between two adjacent isotopes within the isotopic distribution of a peptide the charge state of that peptide can quickly be determined. This eliminated any candidate ions that did not have the charge state matching that of the peptide ions being searched for using this hypothesis driven approach. Ions in the FT-MS that had similar *m/z* and charge state to those predicted were then evaluated based on their MS/MS spectra. The MS/MS spectra of phosphopantetheinylated peptides should reveal a signal corresponding to the pantetheine ejection ion and the resulting dehydroalanine and phosophoserine forms of the peptide subjected to CID. Therefore, the MS/MS spectra of candidate ions could be evaluated based on a combination of their b- and y- ion profiles, and/or PPant elimination ions. Using this approach, the *apo-* form of the ACP active-site peptide was identified (Figure 20). Interestingly, the *holo-* form of this peptide could not be located in the spectra.
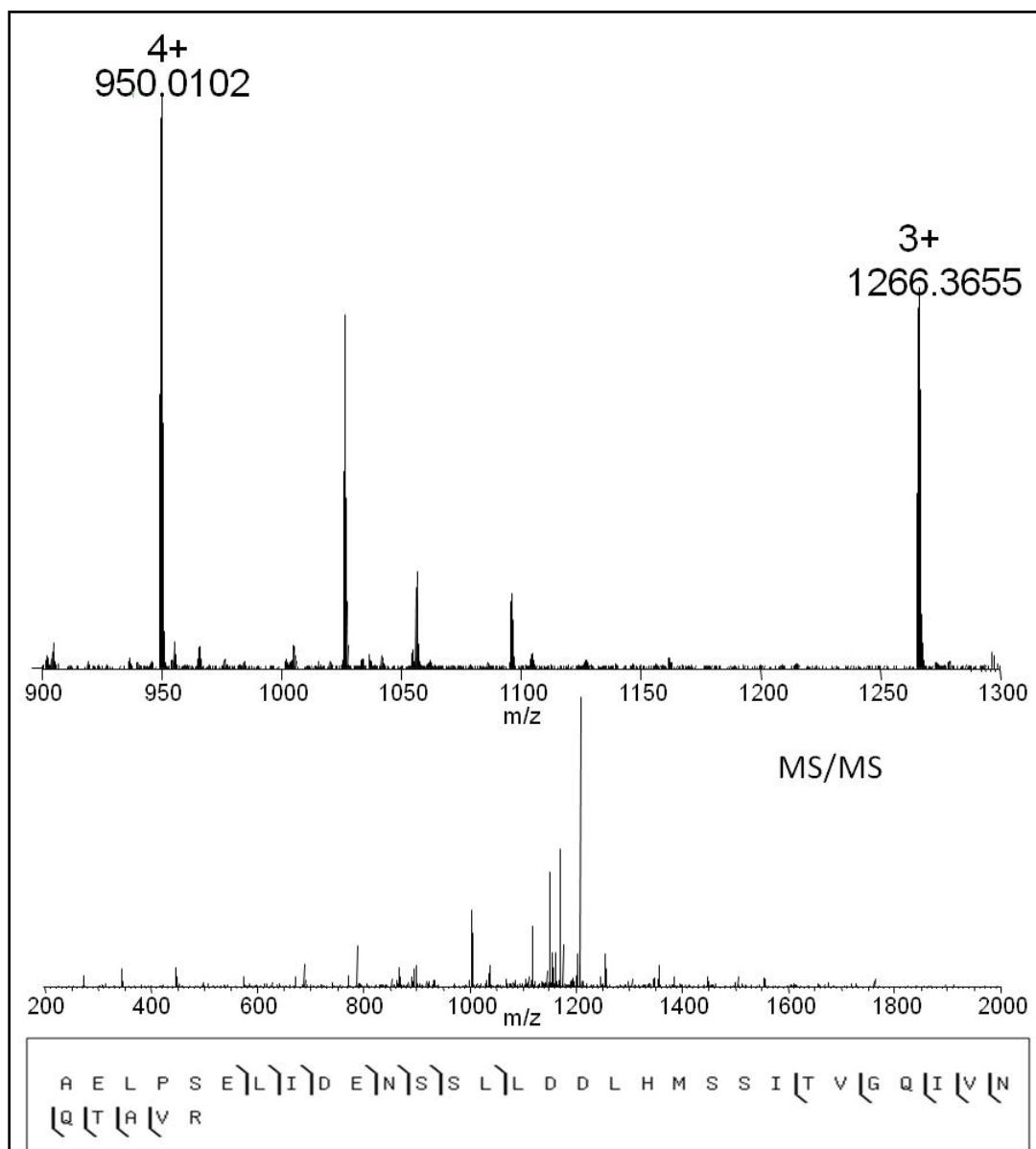
Figure 20: Mass Spectra of SgcE Apo-ACP active-site peptide fragment.
High-resolution mass spectra of the Apo-ACP active site peptide fragment shows the
presence of both the 3+ and the 4+ charge states of the ion (top). The *m/z* values
match those calculated for the active site fragment based on the average mass of the
predicted peptide. (Bottom) The MS/MS spectrum of the peptide confirming the
identity of the peptide based on its fragmentation pattern.

Table 2: Calculated m/z values for SgcE ACP active site fragment

ACP FRAGMENT

Active Site Fragment Sequence: AELPSELIDENSSLLDDLHMSSITVGQIVNQTAVR

		m/z Based on Average Mass				
		No Substrate		Substrate		
		Apo	Holo	Acetyl (+42)	Malonyl (+86)	Acetoacetate (+84)
Charge (+)	1	3797.18296	4137.2688	4179.2688	4223.2692	4221.2688
	2	1899.09545	2069.1384	2090.1384	2112.1386	2111.1384
	3	1266.39961	1379.7615	1393.7615	1408.4283	1407.7615
	4	950.05170	1035.0731	1045.5731	1056.5732	1056.0731
	5	760.24294	828.2601	836.6601	845.4602	845.0601
		m/z Based on Monoisotopic Mass				
		No Substrate		Substrate		
		Apo	Holo	Acetyl (+42)	Malonyl (+86)	Acetoacetate (+84)
Charge (+)	1	3794.91179	4134.9976	4177.00819	4220.99799	4219.01869
	2	1897.95986	2068.0028	2089.00806	2111.00296	2110.01331
	3	1265.64256	1379.0045	1393.00802	1407.67129	1407.01152
	4	949.48390	1034.5054	1045.00800	1056.00545	1055.51063
	5	759.78871	827.8059	836.20799	845.00595	844.61009

Digested samples of SgcE that were incubated with malonyl-CoA were subjected to the same capillary LC-MS analysis methods as the samples that were only digested, and they were analyzed using the same methods. When these samples were first run, it was unknown to what degree the molecules of SgcE had unoccupied *holo*-ACP active sites to which malonate could be loaded. As noted previously, the heterologously expressed SgcE possessed a yellow color when purified, indicating the presence of covalently bound polyene which is not removed in the absence of the thioesterase SgcE10. The samples provided to the Dorrestein Laboratory for analysis had only a subtle yellow appearance, as opposed to a bright yellow appearace (noted in reference 54 - supplemental information section), which seemed to indicate that this particular sample of SgcE contained some SgcE with unoccupied *holo*-ACP active-sites. Therefore, by incubating a sample of SgcE with its substrate, such as acetyl-

CoA or malonyl-CoA, prior to tryptic digestion then it was possible to evaluate the initial steps in the biosynthesis of the enediyne core of C-1027. In unpublished work carried out by Dr. Pieter Dorrestein, it was observed that malonyl-CoA was more efficiently loaded onto the phosphopante-theinyl moiety of SgcE than acetyl-CoA by the SgcE PPTase. Based on these results, it was decided to first incubate some of the SgcE sample with malonyl-CoA in the absence of NAD(P)H prior to digestion and FT-MS analysis.

In these experiments, both the 3+ and 4+ charge states of the malonate-loaded (+86.001 Da) ACP active site were identified based on their intact *m/z,* and identity was verified by detection of the signature PPant elimination ions (Figure 21). The malonate-loaded PPant ejection ion was present at 347.112 m/z (calculated m/z is 347.127), and the dehydroalanine and phosphoserine forms of the corresponding active-site fragments were detected with an observed average mass of 3777.263 Da and 3875.292 Da respectively. In the samples not incubated with malonyl-CoA, there were no detectable ions in the samples of both a charge state and *m/z* that matched to that of the malonate-loaded ACP active site. In addition to malonate loading, clear evidence of acetate-loaded (+42.011) ACP active site was detected (Figure 22). The acetate-loaded PPant ejection ion was observed at 303.165 *m/z* (calculated *m/z* is 303.137) in MS/MS spectra of the 4+ charge state of the +42 Da shifted ACP active-site peptide. The MS/MS spectra of the 3+ charge state of this same peptide did not contain the PPant ejection ion at 303 m/z due to signal low intensity on account of the 1/3 rule. However, the dehydroalanine and phosphoserine active-site fragments were present in the MS/MS spectra of both charge states of the acetate-loaded active-site

peptide. This showed clear evidence that malonate loading of the SgcE active site was occurring, providing strong evidence that malonyl-CoA can act as a primary substrate in C-1027 enediyne biosynthesis, then the malonate is subsequently decarboxylated to acetate.
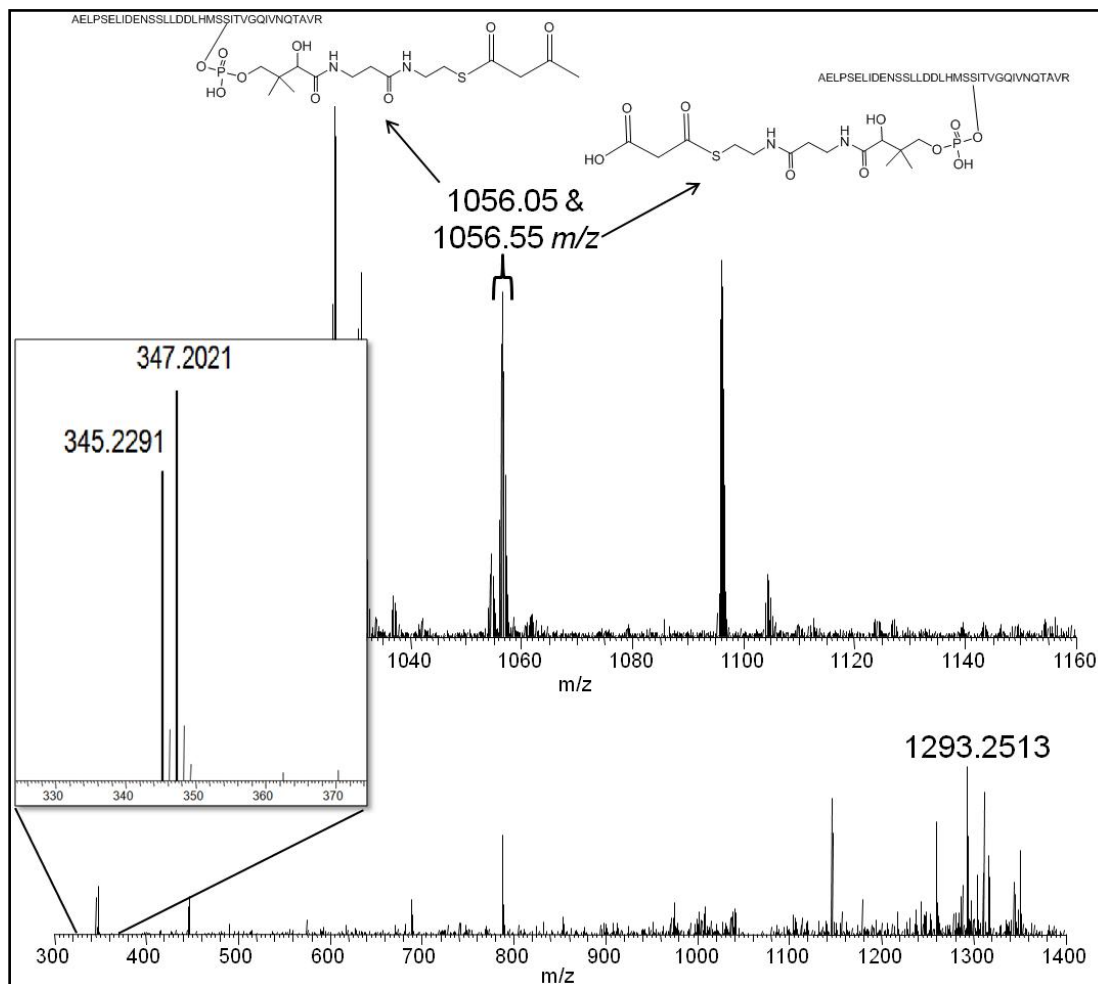
Figure 21: Malonyl- and acetoactyl-loaded SgcE ACP active site fragment. The mass spectrum at the top shows the 4+ ions of the ACP active site loaded with the malonate substrate as well as the product of the condensation of two decarboxylated units of malonate, acetoacetate. The subsequent MS/MS spectrum below shows the presence of the charge-loss phosphoserine form of the active site (3+ 1293.3 *m/z*). PPant ejection by isolation of the region contain the ACP active site reveals two ions (inset) with masses consistent with the ejection of PPant+84Da (345 *m/z*) and PPant+86Da (347 *m/z*).
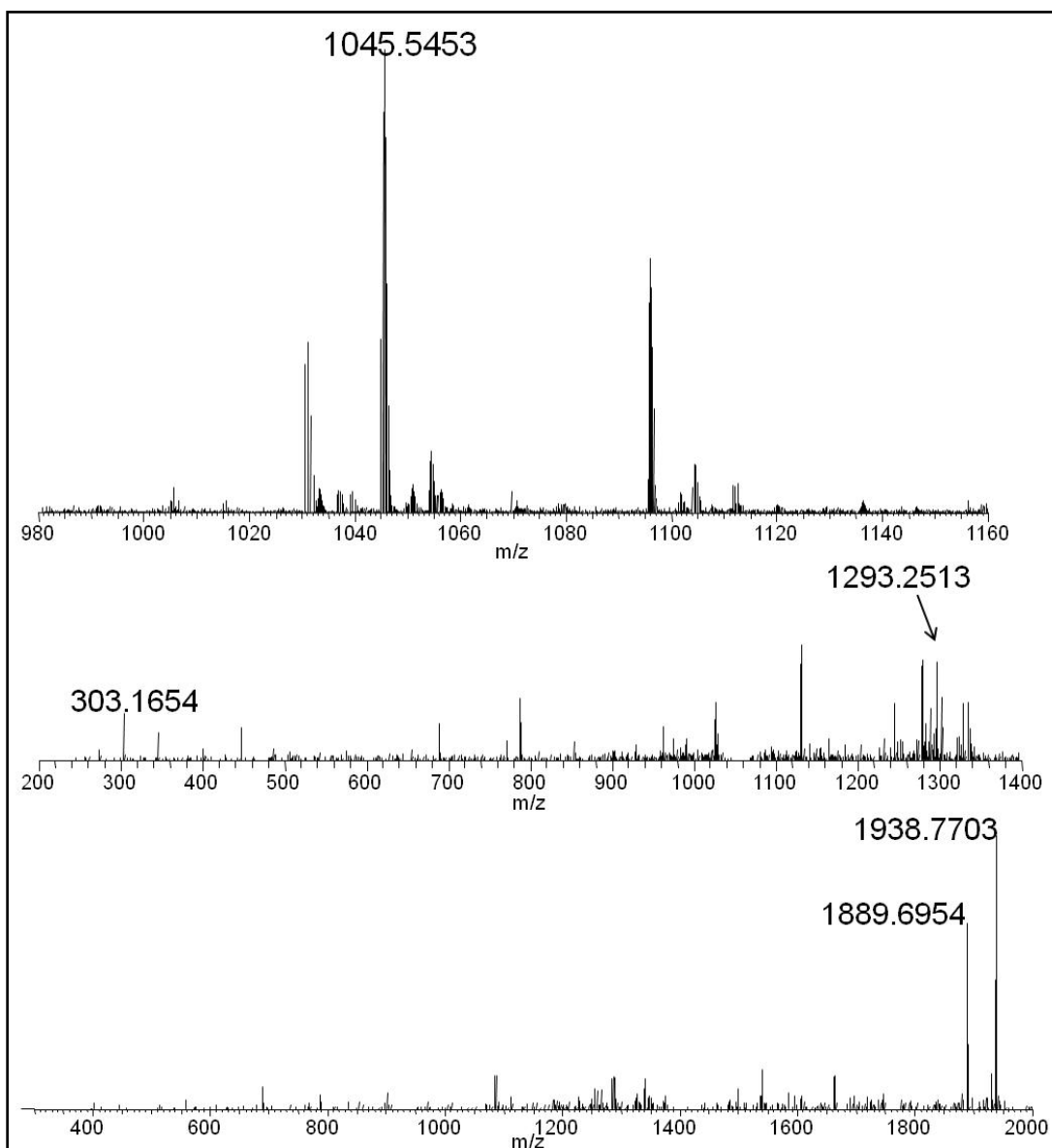
Figure 22: Acetyl-intermediate confirmed by PPant ejection. Mass spectrum reveals a 4+ peptide (top) with an *m/z* corresponding to that of acetate after decarboxylation of malonate bound to to the phosphopantetheinyl moiety. Below are the two MS/MS spectra that verified the identity of the peptide. The MS/MS of the 4+ ion (middle) shows an ion with a mass of 303.1654 corresponding to the pantetheinyl ejection ion plus the addition of acetate. Also seen is the charge-loss phosphoserine of the active site containing peptide. The MS/MS at the bottom is of the 3+ charge state of the same peptide. Both the charge-loss phosphoserine and the dehydroalanine forms of the active site containing peptide are seen. The pantetheinyl ejection ion is not seen due to the 1/3 rule.

The first unit acetate unit is proposed to reside on the KS domain, which catalyzes its condensation with a second acetate extender unit, provided by decarboxylation of a second malonate from the recently unoccupied ACP. In the absence of NAD(P)H, it is expected that the enediyne polyketide intermediate will not undergo additional elongation and will remain tethered to the SgcE ACP active-site phophopantetheine as the un-reduced diketide, acetoacetate (Acac). Based on these assumptions, the malonyl-CoA incubated samples of SgcE were analyzed for the presence of the acetoacetate intermediate (+84.0211 Da) loaded onto the ACP active-site. In the FT-MS spectra, an ion with an average *m/z* corresponding to that of the 4+ charge state of the Acac-loaded active-site peptide was detected at approximately the same time during the gradient as the malonate-loaded peptide. As the isotopic distributions of these two peptides overlapped, and they both eluted between 63 to 63.4 minutes, there was not distinct separation in the FT-MS spectra. During CID fragmentation, all ions within +/- 2.5*m/z* of the ion selected for fragmentation are isolated and fragmented as well. The average *m/z* of 4+ charge state of the malonate-loaded peptide was located at 1056.5 *m/z* and that of the acetoacetate-loaded peptide was 1056.1 *m/z*, thus resulting of simultaneous isolation and fragmentation of the two peptides. The resulting MS/MS spectrum revealed two PPant ejection ions at 345.13 *m/z* and 347.22 *m/z* corresponding to the mass of the pantetheine ejection ion (261.13 Da) plus the mass of acetoacetate (84.02 Da) and malonate (86.00 Da) respectively (Figure 21). These results were in agreement with those of Dr. Dorrestein's in additional unpublished work. In these samples, there was no evidence of additional

intermediates already bound to the ACP active site of SgcE when it was purified from *E coli*.

In addition to the ACP active site, Dr. Dorrestein had previously identified a peptide containing the ACP-transacylase (AT) domain of SgcE. The peptide was determined to have a mass of 4,025.11 Da. The spectra collected from the same samples containing the ACP active-site fragment were searched and the AT active site was detected for 5+ charge state (806.03 *m/z*), the 4+ charge state (1007.89 *m/z*), and the 3+ charge state (1342.71 *m/z*). In the samples of SgcE incubated with malonyl-CoA, this peptide was detected to have an *m/z* shift corresponding to the mass addition of malonate from the malonyl-CoA (Figure 23). This was therefore the second active site from SgcE that was mapped and substrate loading can be monitored in future experiments that are aimed at determining the substrate specificity of SgcE.

Based on the successful detection of the initial ACP-bound intermediates in enediyne core biosynthesis, samples of SgcE were co-incubated with malonyl-CoA and either NADH or NADPH. The goal was to detect additional extended intermediates and to perhaps enrich for the fully-elongated 3-hydroxy-4,6,8,10,12,14 hexadecahexaene intermediate. Because of the iterative nature of SgcE, it was possible that numerous intermediates of various lengths would be detected in these samples. Therefore, *m/z* values were calculated for the 4+ and 3+ charge states of the ACP active-site peptide fragment plus the mass of 21 possible intermediates. The FT-MS spectra revealed several potential matches. In addition to analyzing the MS/MS spectra of candidate peptides based on the mass from the FT-MS specta (Table 3), all data-dependent MS/MS spectra were searched for evidence of the dehydroalanine and

phosphoserine active-site fragments that result from PPant ejection (1888.93 m/z and 1937.93 resulting from 3+ charged peptide, and 1259.62 m/z and 1292.28 resulting from the 4+ charged peptide). This additional method of searching the MS/MS spectra allowed for the possibility of identifying spectra that contained PPant ejection ions containing intermediates of unpredicted mass. Also, this method compensated for possibly missing candidate peptides on account of mass-inaccuracy in the FT-MS spectra. Unfortunately, no additional ACP-bound intermediates of the enediyne core of C-1027 were detected using this capillary LC-FTMS method.
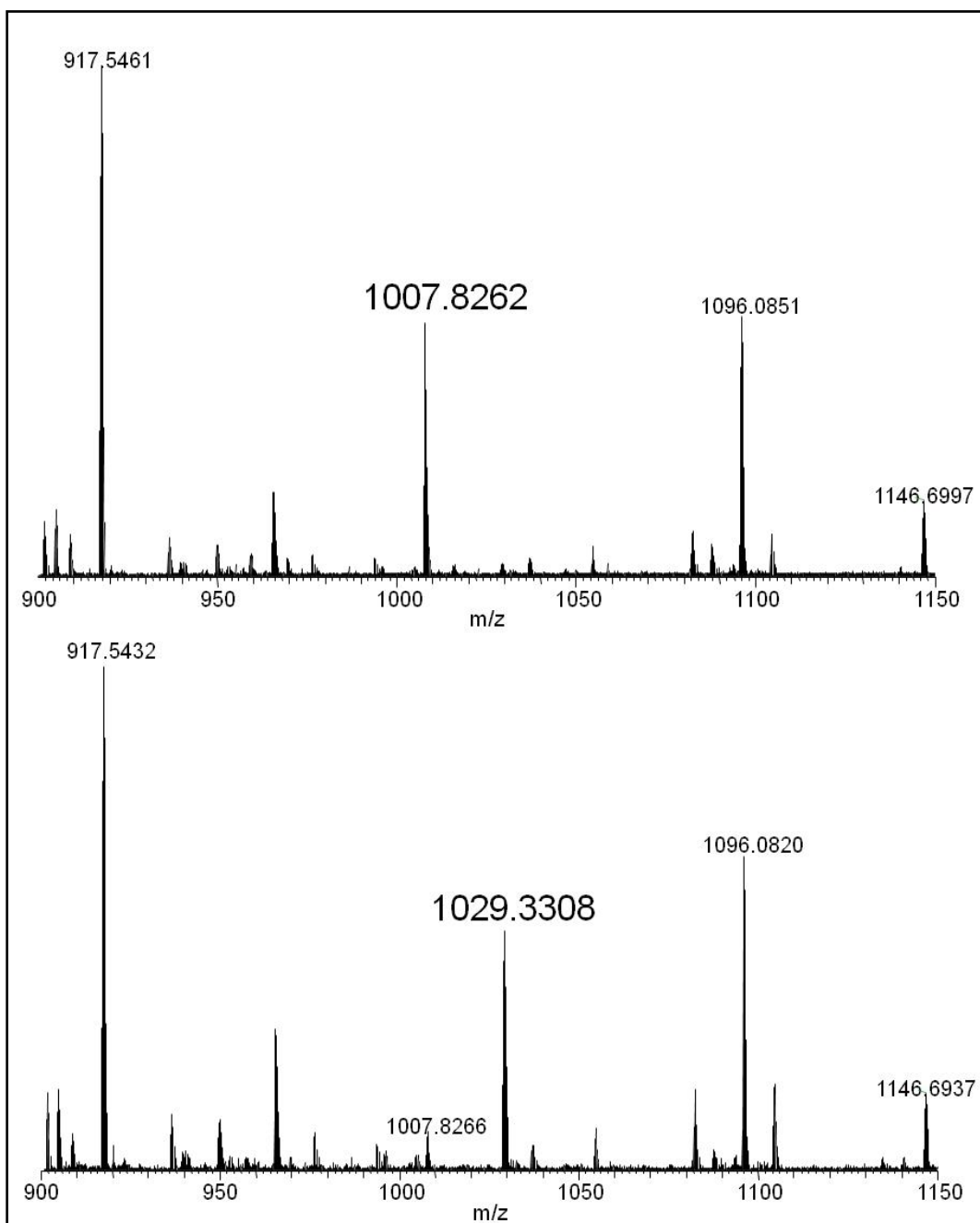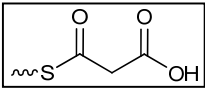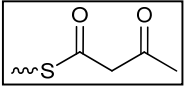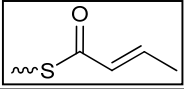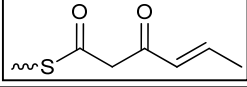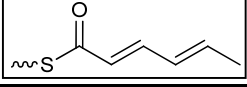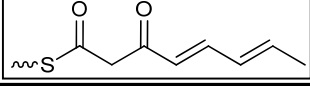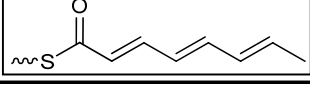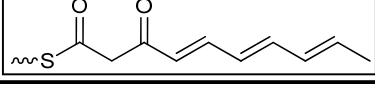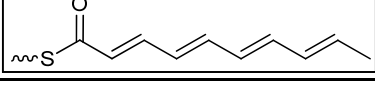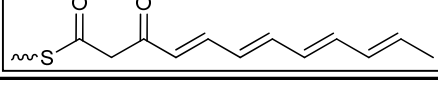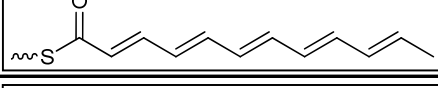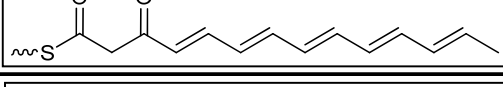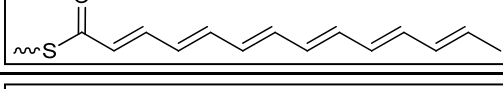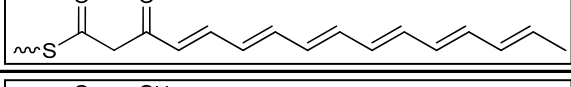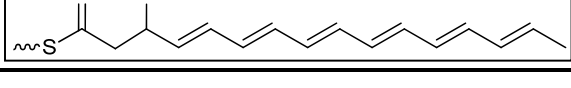
Figure 23: Acyl-transferase (AT) domain loading with malonate. The predicted AT domain active-site containing fragment was identified having a 4+ charge state at 1007.82 *m/z*. After incubation with malonyl-CoA, it can be seen this peptide has a mass shifts of 86 Da, the mass of malonate, and is seen then at 1029.33 *m/z*. The same trend was observed for both the 3+ and the 5+ charge states of the ion as well.

Table 3: Predicted Enediyne Intermediate Structures

Intermediate Mass Addition	Calc'd 4+ *m/z*	Predicted Intermediate Structure
86	1056.60	
84	1056.10	
68	1052.10	
110	1062.60	
94	1058.60	
136	1069.11	
120	1065.11	
162	1075.61	
146	1071.61	
188	1082.11	
172	1078.11	
214	1088.61	
198	1084.62	
240	1095.12	
242	1095.63	

3.5. SgcE Discussion

These results have provided strong evidence in support of the role of SgcE as a polyketide synthase involved in the *de novo* biosynthesis of the enediyne core of C-1027. Using a hypothesis-driven approach to direct the identification of intermediates bound to the phosphopantetheinyl arm of the SgcE active-site containing fragment, detection of the malonate, acetate, and acetoacetate intermediates was accomplished. The additional detection of high levels of malonate-loading to the AT active-site containing peptide paints a complete picture of the initiation of polyketide biosynthesis by mass spectrometric methods. Given that SgcE possesses its own PPTase domain, it is conceivable that the enzyme is capable of utilizing the malonyl-CoA provided during the experiments as the substrate for phosphopantetheinyl transfer, by the PPTase, to the apo-ACP active-sites, which might represent the source of the observed intermediates. However, the observation of AT active-site loading with malonate confirms that this domain is active in SgcE and that initiation of enediyne biosynthesis follows the conventional mechanism of polyketide initiation catalyzed by a malonyl-acetyl-transferase (MAT) domain. While there is evidence of a small amount of malonate-loading that is likely provided from the metabolism of *E coli* in which SgcE was heterologously expressed, it is clear that the malonyl-CoA provided during the experiment was, in fact, the source of the substrate utilized by SgcE.

A prominent issue that stands out from this set of experiments is that the unloaded *holo*-ACP active site containing fragment was never detected in any of the samples analyzed by capillary LC-MS. It is possible that there was simply a low

abundance of *holo*-ACP active-site fragments due to a lack of phosphopantetheinyl transfer; a fair assumption considering that the abundance of the *apo*-ACP active-site peptides was quite high when compared to all peptides in the LC-MS spectra of these experiments. Also, the spectrum published in Figure 2 of reference 54 shows that the intensity of the *apo*-peptide is at least equal to that of the *holo*-peptide, if not greater. However, an additional explanation for this phenomenon could be attributed to the method of MS analysis used in these experiments carried out in the Dorrestein Lab. The peptides analyzed[54] by Dr. Dorrestein in unpublished work in the Kelleher Lab were purified and separated by HPLC prior to direct injection of each fraction into the mass spectrometer. In the experiments conducted more recently, the peptides analyzed were chromatographically separated in the same fused silica capillary column that acts as the electrospray emitter. An applied voltage of 1.5kV produces the charged droplets necessary for ionization, yet the same voltage is also applied to the liquid in the column. It is possible that this method of electrospray ionization results in elimination of the labile pantetheinyl moiety from some peptides prior to detection by the mass spectrometer, thereby the resulting ion appears to be merely a phosophorylated peptide rather than a phosphopantetheinylated peptide. This possible explanation for the absence of the *holo*-peptide is supported by the observation of a 3+ ion with an *m/z* of 1292.70 which possesses an MS/MS spectrum similar to the MS/MS spectra of the *apo*- active-site peptides.

The success of the capillary-LC-MS method used in this experiment demonstrates the potential for using this method to analyze other phosphopantetheinylated proteins from PKS or NRPS systems by the PPant ejection

assay. The advantage of LC-MS is that the time and amount of sample preparation is significantly less when compared with off-line HPLC fractionation, and LC-MS is well suited for analysis of more complex protein samples.   In these experiments, peptides were analyzed using both FT-MS detector and the low-resolution linear trap detector, both followed by data-dependent MS/MS analysis. Signature PPant ejection ions were detectable in the samples analyzed with the low-resolution detector. This is an exciting prospect as it opens up the opportunity to study PKS and NRPS biosynthesis to laboratories that only have access to low resolution mass spectrometers.
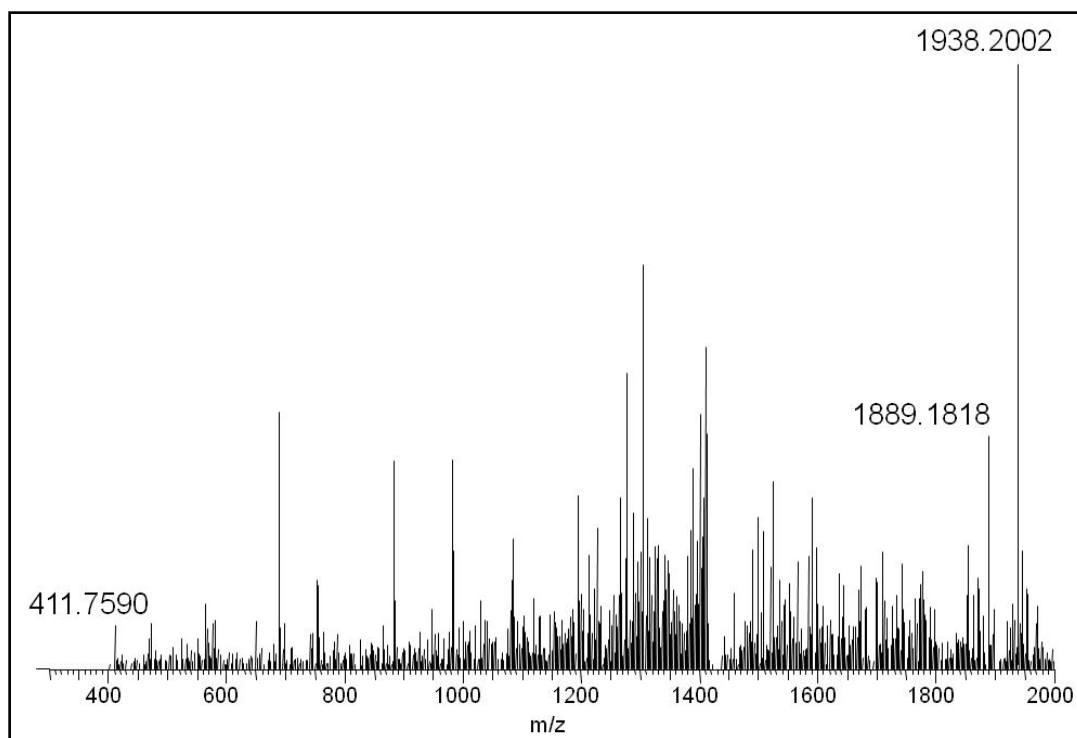
Figure 24: MS/MS of unidentified intermediate on SgcE ACP fragment. This spectrum resulted from CID of a low intensity ion detected using the low-resolution linear trap. Based on the precursor m/z of the ion and the presence of the 2+ charge-loss phosphoserine and dehydroalanine forms of the active-site peptide, the intermediate was determined to have a mass of 150 Da. Low mass accuracy and inability to isolate the ion in more than one spectrum prevent final identification.

The main drawback of using the LC-MS method of analysis, as compared to off-line separation followed by direct infusion, is that all peptides that enter the mass spectrometer are analyzed by the detector within a narrow span of time. Most of the active-site containing peptides from SgcE eluted from the capillary and into the mass spectrometer in less than 30 seconds. This narrow window of time allows only a limited number of MS scans of all peptides and a limited amount of CID fragmentation. In particular, analysis by FT-MS is limited because long scan times are often needed in order to achieve high resolution and high mass accuracy for a given scan, and detection of relatively low abundance peptides of interest requires averaging of multiple scans over the course of a few minutes. In the SgcE samples analyzed by the low resolution linear trap, an "interesting" peptide with an *m/z* of 1429.79 was fragmented and the dehydroalanine and phosphoserine forms of the ACP active site fragment were present in the resulting MS/MS spectrum at 1889.18 *m/z* and 1938.20 *m/z* (Figure 24). This information indicates that this peptide has a 3+ charge and a mass of 4286.37 Da, which suggests that it is the 4136.26 Da phosphopantetheinylated ACP active-site peptide plus an additional 150.11 Da. This mass does not match any of the proposed intermediates, and may represent an unexpected intermediate or a combination of an intermediate and a non-covalently bound adduct to the ACP active-site peptide. Unfortunately, the intensity of this peptide was not high enough to give a detectable signal when analyzed by LC-FTMS, and the peptide was not detected in subsequent samples incubated with NAD(P)H. Also, the corresponding pantetheinyl ejection ion should be expected at 411.23 *m/z*. There is, in fact a very low intensity

ion close to the predicted *m/z*, which is encouraging that some elongated intermediate was detected, but without addition MS^3 data the identity of this intermediate cannot be determined. Therefore, further confirmation of the mass, charge, and identity of this peptide was not achieved using LC-MS. In future work, off-line HPLC separation will likely need to be attempted in order to detect additional intermediates or the fully elongated intermediate.

While this series of capillary-LC-FTMS experiment have not yet led to isolation of the proposed 3-hydroxy-4,6,8,10,12,14-hexadecahexaene intermediate bound to the phosphopantetheinyl arm of the SgcE active-site peptide fragment, the results provided additional insight into studying iterative polyketide synthases, such as PKSEs, and has led to the development of a valuable new software tool for rapid detection of PPant ejection signatures from LC-MS spectra. As discussed previously in the introduction, proteomics programs such as InSpect are excellent tools for the identification of post-translational modifications, but are not designed to identify ions resulting from parent peptides with a charge greater than 3+, and it is difficult search for labile PTMs, such as phosphopantetheine. In addition, the fact that the PPant ejection ion carries a charge with it, results in a charge loss of the precursor peptide, making it difficult for most proteomics programs to properly annotate the MS/MS spectra. These limitations, and the encouraging PEA results from SgcE, resulted in a new collaborative project between the Dorrestein Lab and the Bafna Lab of the UCSD Computer Science Dept.

In order to more effectively analyze mass spectra for the presence of phosphopantetheinyl ejection ions, a new software tool was developed, currently

named PPantFinder, by Anand Patel of the Bafna Lab. At the time that the SgcE spectra containing intermediate-loaded PPant ejection ions was collected, all identification of PPant ejection was done by manual inspection of MS/MS spectra. Manual inspection of LC-MS samples posed a challenge as they often contained several thousand spectra. After identifying PPant ejection ions from SgcE, the goal was set to develop a program that was capable of rapidly searching though large mass spectra files for spectra containing sets of characteristic peaks that represent PPant elimination ions and the PPant ejection ion itself. As part of this first goal, we also sought to have the program be capable of searching for predefined intermediates, which would aid in detecting intermediates or analyzing NRPS/PKS substrate specificity by mass spectrometry. The second goal of the new tool, and something that is currently in development, is to annotate the y and b fragment ions present in spectra with the PPant signatures, in order confirm the sequence of the fragmented peptide.

The development of the program was aimed initially developing an algorithm to positively identify the spectra from the SgcE files based on which spectra had the most intense PPant ejection ion signals. However, it was quickly recognized that a more well-developed method for discriminating between true-positives and false-positives was needed. Therefore, additional sample spectra containing PPant ejection signatures were collected and used as a training set for the development of the program. Spectra previously collected by members of the Dorrestein Lab were from from analysis of the following phosphopantetheinylated proteins: YbbR, a protein of unknown function from *Bacillus subtilis*, the peptidyl carrier protein CouN5 from the coumermycin gene cluster in *Streptomyces rishiriensis*, the NRPS

adenylation/thiolation didomain Strop_4416 from *Salinispora tropica*, and the indivudally expressed ACP from fungal polyketide synthase Pks4 from *Gibberella fujikuroi*. All spectra from this training set were evaluated for the presence of the iminolactone PPant ion (261.16 *m/z)*, the phospho-PPant ion (359.12 *m/z*), the dehydroalanine active site (charge-reduced protein minus water), and the phosphoserine active-site (charge-reduced phospho-apo protein). In addition, spectra were analyzed based on the presence of peaks that correspond to charge-neutral dehydroalanine and phosphoserine. Finally, peaks corresponding to the both the charge-reduced and neutral proteins that undergo no loss of water were considered. Based on these features, the spectra that the program singled out from the training set were identified as either true-positives or false-positives. The PPantFinder program was then programmed by Anand Patel to apply the discriminatory model, defined by the training set, when searching other spectra. Despite this model being developed from phosphopantetheinylated peptide/protein data, it was successful in correctly identifying MS/MS spectra in the SgcE data set that contained substrate/intermediate-loaded phophophopantetheinyl ejection signatures.

Chapter 4: Analysis of Linear Diketide Formation by LovF in the Lovastatin

4.1. LovF Introduction

Of the numerous pharmaceutical drugs that have been derived from polyketide natural products, simvastatin (trade name Zocor), a member of the cholesterol-lowering statin class of drugs, has been one of the top-selling pharmaceuticals worldwide, accounting for over five billion dollars in sales annually[119] from 2002-2004, reflecting its popularity and therapeutic value. Currently, atorvastatin (trade name Lipitor), another cholesterol-lowering medication, holds the top position as the best-selling drug in the U.S., but statins overall are the most prescribed medications in the country.[120] In 2008, there were 20 times more prescriptions issued for statins in the U.S. than for the second-most prescribed drug class. The statin-class of drugs is used for the treatment of hypercholesterolemia by lowering levels of low-density lipoprotein (LDL) and very low-density lipoprotein (VLDL) circulating in the blood.[121] Statins are competitive inhibitors of 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase, the rate-limiting enzyme within the metabolic pathway leading to production of cholesterol in humans.[122,123] Inhibition of HMG-CoA reductase results in a decrease of cholesterol synthesis and up-regulation of specific hepatic (liver) cell-surface receptors in response to the lower serum cholesterol levels. These receptors binds to and take-up LDL and VLDL into hepatic cells, leading to their digestion. The effectiveness of HMG-CoA inhibition and the relatively low risk of side-effects are what make statins, such as simvastatin and atrovastatin, a very attractive therapy for hypercholesterolemia.

Currently, Simvastatin is available as a generic medication and remains a popular, less-expensive alternative to atrovastatin. Simvastatin, is a semisynthetic

derivative[124] of the compound lovastatin (tradenames Mevacor, Altocor, Altoprev). Lovastatin was one of the first specific inhibitors of HMG-CoA reductase discovered and isolated[125, 126, 127] from the fungus *Aspergillus terreus*. Natural statins, such as lovastatin, which are isolated directly as fermentation products, and the semi-synthetic statin, simvastatin, share a very similar chemical structure (Figure 25). Simvastatin possesses a substitution of an α-dimethylbutyrate in place of the α-methylbutyrate side chain present in lovastatin. This fairly simple substitution significantly increases the HMG-CoA inhibition of simvastatin over lovastatin. Various methods have been developed for the synthesis of simvastatin using lovastatin as the precursor. One semi-synthetic method in involves ester hydrolysis to remove the α-methylbutyrate, followed by selective silylation to protect the nearby hydroxyl group, then re-esterification of the C8 hydroxyl group using dimethylbutryl chloride, and de-protection of the silylated C13 hydroxl group.[128] However, due to the regioselective esterification and the need to protect nearby alcohols contribute to a relatively low yield. Because methods for the synthesizing simvastatin require lovastatin, attention has been directed towards better understanding the biosynthetic machinery involved in lovastatin production by *Aspergillus terreus.*
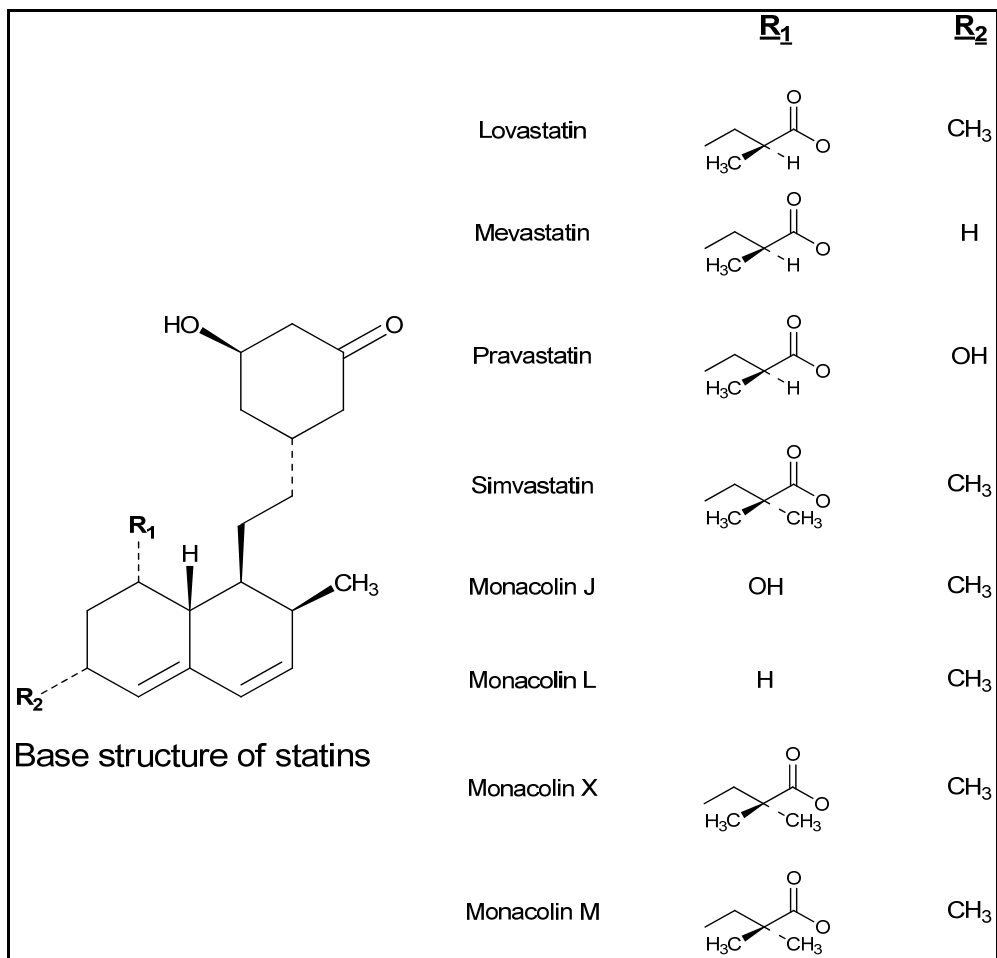
Figure 25: Base structure of statins. The statins are composed of a naphthalene ring and B-hydroxylactone. Substitutions at the C8 (R_1) side chain and C6 (R_2) are shown to the right the result in the various lovastatin analogues.

In the late 1990's researchers identified genes in *A. terreus* that were required for the proper biosynthesis of lovastatin, using gene inactivation techniques.[129] From this research, the genes for lovastatin were cloned and characterized to possess two separate type I iterative polyketide megasynthases, and what was proposed at the time to be a trans-esterase.[130] It was determined that a PKS, nonaketide synthase LovB, along with dissociated enoylreductase (ER) LovC, produces dihydromonacolin L. Dihydromonacolin L is oxidized by cytochrome P450 to form monacolin J. The α-

methylbutyrate side chain is synthesized by a second PKS, the diketide synthase LovF (Figure 26). Both LovB and LovF posses the minimal catalytic domains of a PKS, such as the ketosynthase (KS) domain that carries out decarboxylative claisen condensation for polyketide chain elongation, the malonyl-CoA:ACP acyltransferase (MAT) that selects malonyl-CoA and transfers it as a malonic ester to be used as an extender unit, and the acyl carrier protein (ACP), that serves as the anchoring point for the flexible 4'-phosphopantetheine arm that tethers the growing polyketide chain. In addition, LovB and LovF possess a ketoreductase (KR) domain, which reduces the β-keto group to a β-hydroxy group, and a dehydratase domain (DH) which removes water to form an α-β alkene. LovF posseses an enoyl reductase domain (ER) that fully reduces the α-β double bond to a single bond, whereas LovB relies on the dissociated ER LovC for enoyl-reduction. LovB preferentially utilizes nine malonyl-CoA units[131] for the synthesis of dihydromonacolin L, whereas LovF utilizes only two malonyl-CoA units to produces the α-methylbutyrate side-chain.

The unique feature of the lovastatin biosynthetic pathway in *Aspergillus terreus* is that neither LovB nor LovF possess a dedicated thioesterase (TE) domain that is typically found on the C-terminal domain of polyketide synthases and catalyzes the cyclization and/or releases the fully elongated polyketides by hydrolysis.[132] While the nonaketide release mechanism for LovB remains unknown, evidence has supported the role of an additional enzyme, LovD, as an acyltransferase that catalyzes the transfer of the α-methylbutyl acyl product from LovF to the C8 hydroxyl of monacolin J (Figure 25). When the lovastatin biosynthetic pathway was first characterized, it was found that inactivation of either the gene for LovF or the gene for

LovD resulted in accumulation of the monacolin J precursor.[129] More recent research in the laboratory of Yi Tang at UCLA has focused on characterizing the interaction of LovD with both LovF and monacolin J.[133] Their laboratory has successfully expressed LovB, LovF, and LovD from *A. terreus* in a heterologous yeast strain, allowing for isolation and *in vitro* characterization of these enzymes. Their research has revealed that LovD has broad substrate specificity, preferentially accepting 3 to 6 carbon acyl groups from various acyl-CoAs, and is capable of transferring them to monacolin J. In addition, it was demonstrated that at high concentrations, monacolin J inhibits LovD, most likely by binding to LovD and preventing the entrance of acyl-CoA substrates to the enzyme active site.

4.2. Research Goals

Although there has been significant progress made in the in vitro characterization of LovB and LovD, the *in vitro* interaction of LovF and LovD has been studied to a lesser extent. Soluble LovF was heterologouly expressed by members of the Tang laboratory at UCLA and generously provided to the Dorrestein laboratory for characterization. It was unknown whether the heterologously expressed LovF was in fact phosphopantetheinylated, as this information could not be determined using low-resolution mass spectrometry. The initial aim was to use tryptic digestion, followed by off-line HPLC separation and high-resolution mass spectrometry to attempt to detect and map the LovF ACP active site to the expected site of phosphopantetheinylation. In addition, the second goal was to characterize the

activity of LovF and to analyze diketide synethsis using high-resolution mass spectrometry.
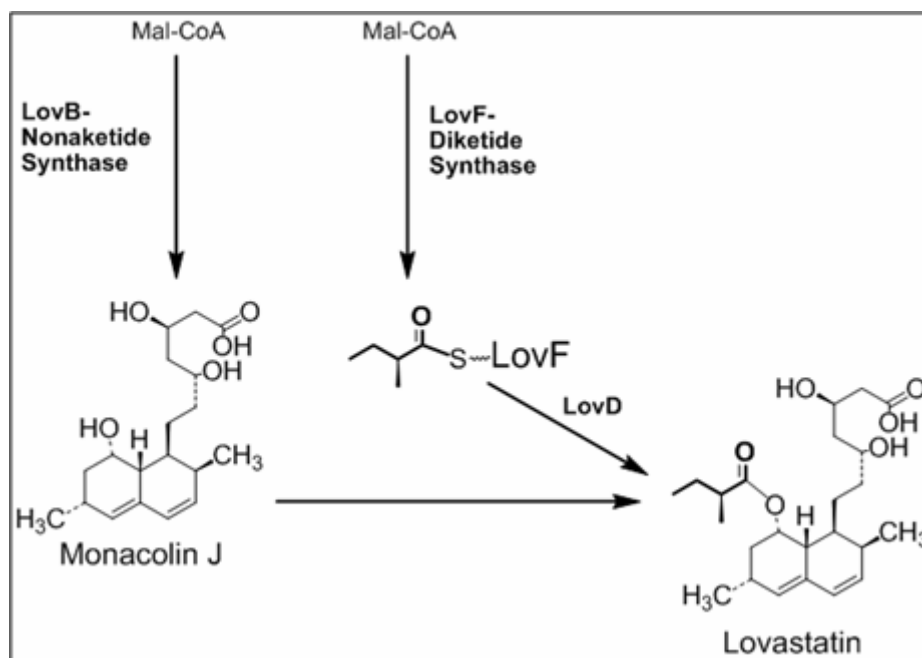
Figure 26: Lovastatin Biosynthesis by LovB/LovF. The schematic shows the contribution of the nonaketide by LovB, that forms Monacolin J, and the diketide side-chain contributed by LovF. Transfer of the α-methylbutyrate side chain to the C8 hydroxyl- group of Monacolin J is catalyzed by LovD.

Initially, it was unknown how efficiently the soluble LovF would be able to produce the diketide α-methylbutyrate side-chain prior to transferring to LovD. The expected mechanism by which LovF synthesizes the α-methylbutyrate product requires two molecules of malonyl-CoA, and an additional methyl donor, such as S-adenosyl methionine (SAM) and two molecules of a reducing agent such as NADH or NADPH. Therefore, the full complement of substrates and cofactors, incubated with active *holo*-LovF was expected to yield α-methylbutyrate tethered to the phosphopantetheinyl (PPant) moiety of the LovF ACP domain. Exclusion of

NAD(P)H, SAM, or both was expected to produce different PPant-bound product profiles. The final goal of this research was to attempt to utilize the phosphopantetheinyl ejection assay (PEA) to evaluate substrate-loading effiency, α-methylbutyrate synthesis, and the formation of any alternative products by LovF.
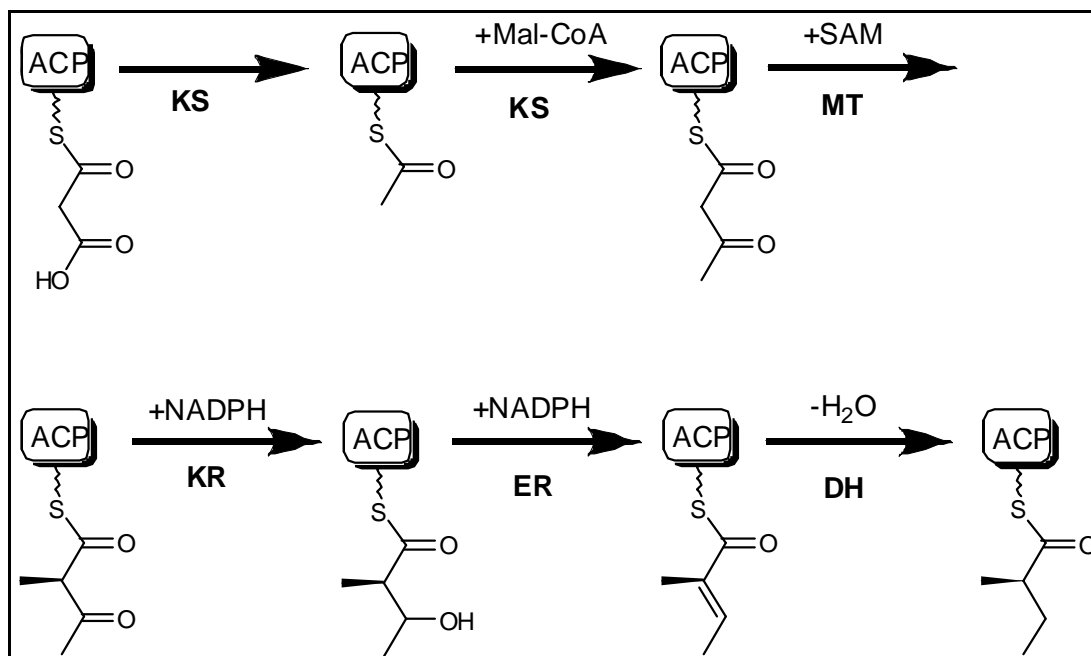
Figure 27: Mechanism of α-methylbutyrate biosynthesis by LovF. Malonate is decarboxylated and condenses with an extender unit. Cofactors NADPH and SAM are utilized in tailoring reactions.

4.3. Materials and Methods

4.3.1. Sequence of Information for Investigated Protein(s)/Peptide(s):

Soluble LovF expressed and purified by Xinkai Xie of the Tang laboratory at UCLA. To obtain soluble LovF megasynthase for in vitro studies, the *lovF* gene from *Aspergillus terreus* was amplified and put into pET28a vector inserted into a yeast 2μ expression vector under the control of the ADH2 promoter. The resulting vector was transformed into *Saccharomyces cerevisiae* BJ5464-NpgA, which contained a

chromosomal copy of the phosphopantetheinyl (ppant) transferase NpgA. Specific

details can be found in reference 133.

Purified Holo-SgcE with His-tag – (MW 277,627.43 Da)
MTPLDAPGAPAPIAMVGMGCRFGGGATDPQKLWKLLEEGGSAWSKIPPSRFNVGGVYHPNGQ
RVGSMHVRGGHFLDEDPALFDASFFNMSTEVASCMDPQYRLILEVVYEALEAAGIPLEQVSGS
KTGVFAGTMYHDYQGSFQRQPEALPRYFITGNAGTMLANRVSHFYDLRGPSVSIDTACSTTLT
ALHLAIQSLRAGESDMAIVAGANLLLNPDVFTTMSNLGFLSSDGISYSFDSRADGYGRGEGVA
AIVLKTLPDAVRDGDPIRLIVRETAINQDGRTPAISTPSGEAQECLIQDCYQKAQLDPKQTSYVE
AHGTGTRAGDPLELAVISAAFPGQQIQVGSVKANIGHTEAVSGLASLIKVALAVEKGVIPPNAR
FLQPSKKLLKDTHIQIPLCSQSWIPTDGVRRASINNFGFGGANAHAIVEQYGPFAETSICPPNGYS
GNYDGNLGTDQAHIYVLSAKDENSCMRMVSRLCDYATHARPADDLQLLANIAYTLGSRRSNF
RWKAVCTAHSLTGLAQNLAGEGMRPSKSADQVRLGWVFTGQGAQWFAMGRELIEMYPVFK
EALLECDGYIKEMGSTWSIIEELSRPETESRVDQAEFSLPLSTALQIALVRLLWSWNIQPVAVTS
HSSGEAAAAYAIGALTARSAIGISYIRGALTARDRLASVHKGGMLAVGLSRSEVGIYIRQVPLQ
SEECLVVGCVNSPSSVTVSGDLSAIAKLEELLHADRIFARRLKVTQAFHSSHMNSMTDAFRAGL
TELFGADPSDAANASKDVIYASPRTGARLHDMNRLRDPIHWVECMLHPVEFESAFRRMCLDE
NDHMPKVDRVIEIGPHGALGGPIKQIMQLPELATCDIPYLSCLSRGKSSLSTLRLLASELIRAGFP
VDLNAINFPRGCEAARVQVLSDLPPYPWNHETRYWKEPRISQSARQRKGPVHDLIGLQEPLNL
PLARSWHNVLRVSDLPWLRDHVVGSHIVFPGAGFVCMAVMGISTLCSSDHESDDISYILRDVN
FAQALILPADGEEGIDLRLTICAPDQSLGSQDWQRFLVHSITADKNDWTEHCTGLVRAEMDQP
PSSLSNQQRIDPRPWSRKTAPQELWDSLHRVGIRHGPFFRNITCIESDGRGSWCTFAIADTASAM
PHAYESQHIVHPTTLDSAVQAAYTTLPFAGSRIKSAMVPARVGCMKISSRLADLEARDMLRAQ
AKMHSQSPSALVTDVAVFDEADPVGGPVMELEGLVFQSLGASLGTSDRDSTDPGNTCSSWHW
APDISLVNPGWLEKTLGTGIQEHEISLILELRRCSVHFIQEAMESLSVGDVERLSGHLAKFYAW
MQKQLACAQNGELGPESSSWTRDSEQARCSLRSRVVAGSTNGEMICRLGSVLPAILRREVDPL
EVMMDGHLLSRYYVDALKWSRSNAQASELVRLCCHKNPRARILEIGGGTGGCTQLVVDSLGP
NPPVGRYDFTDVSAGFFEAARKRFAGWQNVMDFRKLDIEDDPEAQGFVCGSYDVVLACQVL
HATSNMQRTLTNVRKLLKPGGKLILVETTRDELDLFFTGLLPGWWLSEEPERQSTPSLSPTM
WRSMLHTTGFNGVEVEARDCDSHEFYMISTMMSTAVQATPMSCSVKLPEVLLVYVDSSTPMS
WISDLQGEIRGRNCSVTSLQALRQVPPTEGQICVFLGEVEHSMLGSVTNDDFTLLTSMLQLAGG
TLWVTQGATMKSDDPLKALHLGLLRTMRNESHGKRFVSLDLDPSRNPWTGDSRDAIVSVLDL
ISMSDEKEFDYAERDGVIHVPRAFSDSINGGEEDGYALEPFQDSQHLLRLDIQTPGLLDSLHFTK
RNVDTYEPDKLPDDWVEIEPRAFGLNFRDIMVAMGQLESNVMGFECAGVVTSLSETARTIAPG
LAVGDRVCALMNGHWASRVTTSRTNVVRIPETLSFPHAASIPLAFTTAYISLYTVARILPGETVL
IHAGAGGVGQAAIILAQLTGAEVFTTAGSETKRNLLIDKFHLDPDHVFSSRDSSFVDGIKTRTRG
KGVDVVLNSLAGPLLQKSFDCLARFGRFVEIGKKDLEQNSRLDMSTFVRNVSFSSVDILYWQQ
AKPAEIFQAMSEVILLWERTAIGLIHPISEYPMSALEKAFRTMQSGQHVGKIVVTVAPDDAVLV
RQERMPLFLKPNVSYLVAGGLGGIGRRICEWLVDRGARYLIILSRTARVDPVVTSLQERGCTVS
VQACDVADESQLEAALQQCRAEEMPPIRGVIQGAMVLKDALVSQMTADGFHAALRPKVQGS
WNLHRIASDVDFFVMLSSLVGVMGGAGQANYAAAGAFQDALAEHRMAHNQPAVTIDLGMV
QSIGYVAETDSAVAERLQRIGYQPLHEEEVLDVLEQAISPVCSPAAPTRPAVIVTGINTRPGPHW
AHADWMQEARFAGIKYRDPLRDNHGALSLTPAEDDNLHARLNRAISQQESIAVIMEAMSCKLI
SMFGLTDSEMSATQTLAGIGVDSLVAIELRNWITAKFNVDISVFELMEGRTIAKVAEVVLQRYK
AHHHHHH

4.3.2. *Holo*-LovF Sample Preparation for Comparison Tryptic Digestions of *holo*-LovF Incubated With & Without Malonyl-CoA

The first samples of *holo*-LovF to be analyzed were prepared by using equivalent incubation conditions. Samples of *holo*-LovF provided by the Tang laboratory had a concentration of 200 µM (55.6 mg/ml). The protein was provided in a Tris-HCl buffer with a pH of 7.8. *Holo*-LovF was then diluted with HPLC grade water (J.T. Baker, Netherlands) to reduce the concentration to 2.86 mg/ml. The final 50 µL incubation mixtures were prepared to contain 100 µg of soluble *holo*-LovF with 1 mM TCEP and 100mM Tris-HCl pH 8.0 (TEKnova Inc, Hollister, CA). Samples incubated with malonyl-CoA (Sigma-Aldrich, St. Louis, MO) contained 500 µM of the acyl substrate. Samples were incubated at ambient temperature for 30 minutes, and then treated to tryptic digestion followed by HPLC separation.

4.3.3. Initial *Holo*-LovF Sample Preparation for the Evaluation of α-methylbutyrate Formation and Alternate Product Formation

After it was determined that HPLC peak intensity improved and there was no detectable difference in the tryptic digestion patters, double quantity of *holo*-LovF was used. Therefore, *holo*-LovF was diluted with HPLC grade water (J.T. Baker) to reduce the concentration to 11.11 mg/ml. The final 50 µL incubation mixtures were prepared to contain 200 µg of soluble *holo*-LovF in 1 mM TCEP and 100mM Tris-HCl pH 7.5 (TEKnova). All samples were incubated with 500 µM malonyl-CoA. To test for the production of α-methylbutyrate, samples were also incubated with 500 µM

of buffered (25 mM Tris, pH 7.5) S-adenosyl-L-methionine (Sigma) and 500 μM of NADPH (Sigma). To test for the production of butyrate by LovF, the same conditions were repeated without SAM. To test for the production of 2-methyl-3-keto butyrate by LovF, the same conditions were repeated except without NADPH. Finally, samples were incubated without SAM and NADPH, only malonyl-CoA. The sample preparation lacking SAM and NADPH cofactors had been tested previously, and no 3-keto-butyrate (acetoacetate) was detected, so its purpose was simply to aid HPLC analysis.

Incubations were typically allowed to proceed for 20 minutes. In addition, Xinkai Xie in the Tang laboratory at UCLA had noted substantial hydrolysis of α-methylbutyrate-load stand-alone *holo*-ACP constructs when left for more than 30 minutes. Therefore, the same incubation conditions to evaluate the production of α-methylbutyrate were repeated, but with incubation times of 1, 2, 8, and 16 minutes. Each sample was then treated to tryptic digestion followed by HPLC separation.

4.3.4. *Holo*-LovF Sample Preparation for the Evaluation of α-methylbutyrate Formation Using Different Quantities of Malonyl-CoA

Samples of *holo*-LovF were diluted with HPLC grade water (J.T. Baker) to reduce the concentration to 11.11 mg/ml. The final 50 μL incubation mixtures were prepared to contain 200 μg of soluble *holo*-LovF in 1 mM TCEP and 100mM Tris-HCl pH 7.5 (TEKnova). Samples were prepared so that the quantity of the cofactors SAM and NADPH were held constant at 500 μM SAM and 500 μM of NADPH. Different samples were prepared with variable concentrations of malonyl-CoA we

varied so that the samples contained 1, 2, 4, or 8 molar equivalents of malonyl-CoA.
Samples were incubated for 15 minutes. Each sample was then treated to tryptic
digestion followed by HPLC separation.

4.3.5. *Holo*-LovF Sample Preparation for the Evaluation of α-methylbutyrate and Alternate Product Formation Using Substoicheometric Quantities of Malonyl-CoA

Again, *holo*-LovF was diluted with HPLC grade water to reduce the
concentration to 11.11 mg/ml. The final 50 µL incubation mixtures were prepared to
contain 200 µg of soluble *holo*-LovF (14.4 µM) in 1 mM TCEP and 100mM Tris-HCl
pH 7.5 (TEKnova). All samples were incubated with 500 µM malonyl-CoA. To test
for the production of α-methylbutyrate, samples were also incubated with 500 µM S-
adenosyl-L-methionine (Sigma) and 500 µM of NADPH (Sigma). Each sample was
then treated to tryptic digestion followed by HPLC separation.

4.3.6. *Apo*-LovF Sample Preparation for the Evaluation of Acetoacetyl-CoA Loading Using PPTase *sfp* and Alternate Product Formation

Samples of *apo*-LovF (300 µM) were also provided by the Tang laboratory. A
larger sample volume was used in order to allow for subsequent assay using SAM and
NADPH after loading of acetoacetate (3-keto-butyrate) was evaluated. Samples
mixtures of 160 µL in volume contained 100 µM *apo*-LovF, 25mM Tris-HCl, pH
7.5mM, 1mM TCEP. Samples were then incubated with 20 µM *sfp*[24], 8mM MgCl,

and 1mM Acac-CoA. Additional samples were prepared using 2mM MgCl after it was noted that 8mM resulted in a very cloudy solution, and possible precipitation of soluble *apo*-LovF.

4.3.7. Offline HPLC Purification of Trypsin-Digested LovF Samples

Samples to be analyzed by FTMS were prepared by digesting 50μL holo-LovF (final concentrations of either 2mg/ml or 4mg/ml) using 1μg of proteomics grade trypsin (Sigma Trypsin Singles™) in 50 μL of 40mM Ammonium bicarbonate reaction buffer, pH 8.2, at 37°C for 10 minutes. After 10 minutes, 100μL of 100% formic acid (TEKnova, Hollister, CA) was added to quench the digestion. The reaction mixture was separated by HPLC using a Jupiter 5μ C4 300A 150 x 4.6 mm reverse phase column (Phenomenex - Torrance, CA) on a water-acetonitrile gradient running at 1ml/min as shown below:

Table 4: HPLC gradient used for PKS/NRPS fractionation

min	0.0	10.0	15.0	55.0	60.0	60.1	60.2	62.6	63.0	65.0	66.0	73.0
% A	90	90	70	30	10	10	95	95	5	5	90	90
% B	10	10	30	70	90	90	5	5	95	95	10	10

Solvent A is water with 0.1% TFA and solvent B is acetonitrile with 0.1%. TFA.

1 mL fractions were collected using an auto-sampler then frozen at -80°C Fractions were then lyophilized for 3-6 hours until solvent was completely evaporated.

4.3.8. Electrospray Fourier-Transform Mass Spectrometric Analysis of Digested LovF Fractions

The lyophilized HPLC fractions were re-dissolved in 30-60μL of 49% methanol, 49% H2O, 2% formic acid, vortexed vigorously for 1-2 minutes, then centrifuged at 14,000 RPM to spin down any insoluble debris. For mass spectrometric analysis samples were analyzed using a Thermo Finnigan 6.43T LTQ-FT mass spectrometer equipped with a dual quadrupole octupole linear trap. Samples were introduced into the mass spectrometer using a Nanomate 100 apparatus for automated nanospray (Advion Biosciences - Ithica, NY) operating a with an electrospray voltage of 1.4-1.8 kV and a spray pressure of 0.3-0.7 psi. For high-resolution FTMS analysis, typically an 8000 millisecond ion accumulation time per scan was used, and at least 200 scans were acquired per spectrum and averaged together. Spectra collection was carried between the ranges of 200 m/z to 2000 m/z. Precursor ions to be fragmented were first isolated from other ions in the sample by isolating ions within narrow m/z regions (5-50 m/z). These isolation windows could also be used without and CID to enrich the signal of lower intensity ions in a given spectrum. MS/MS spectra acquisition was carried out using the same ion accumulation times and fragmentation was accomplished using collisionally induced dissociation (CID) with normalized collision energy of 20 and an activation Q of 0.200. Phosphopanteheinylated ions with an m/z above 1300 required and even lower activation Q in order to detect ions in the lower m/z ranges where PPant ejection ions are expected.

Samples analyzed for presence of the α-methylbutyrate formation and other alternate products by LovF were first analyzed by the methods described above. The

intensity of the α-methylbutyrate-loaded PPant ejection ion resulting from CID was detectable using the linear trap in the subsequent MS/MS spectra. However, the intensity of this ion was too low to be detected by FT-ICR after CID. In order to try to determine the identity of the 345 m/z species with high mass accuracy, the sample was fragmented at the source by setting fragmentation energy of 75.0V then isolating ions in a 3 m/z window centered on 345 m/z. In addition MS/MS/MS (MS3) spectra were collected for PPant ejection ions. The ejection ion resulting from collision of the precursor peptide could be isolated and fragmented using a second round of CID. To accomplish that, PPant ejection ions were isolated in a 3-5 *m/z* window and fragmented by CID with an ion accumulation time of 1000 milliseconds, an activation Q of 0.250, and normalized collision energy of 20.

4.4. LovF Results

A key strategy relying on HPLC migration differences for the rapid detection of PKS and NRPS active sites in protease digested sample was outlined in reference 47. If great care is taken to prepare two equivalent samples of a PKS, such as LovF, with identical conditions, then the migration differences of active-site tryptic fragments can be observed as peak shifts when one sample is treated with a different substrate or co-factor than the sample it is being compared to. Differently loaded active sites may have different retention times when separated by identical gradients. Using this method, one can compare the HPLC traces of a digest of an *apo-* form of the protein to a digest of the *holo-* form of the protein. Also, a digest of the same holo- protein can be compared to a digest of the *holo-* protein incubated with acyl-

CoAs and/or additional co-factors. As long as sample incubation/digestion times and conditions are the same, different samples are separated using the same HPLC gradients and solvents, peak shifts should be detectable. This method of detecting PKS/NRPS active sites served as the basis for analysis of the digested samples of *holo*-LovF and for identification of the phosphopantetheinylated ACP active-site containing peptides.
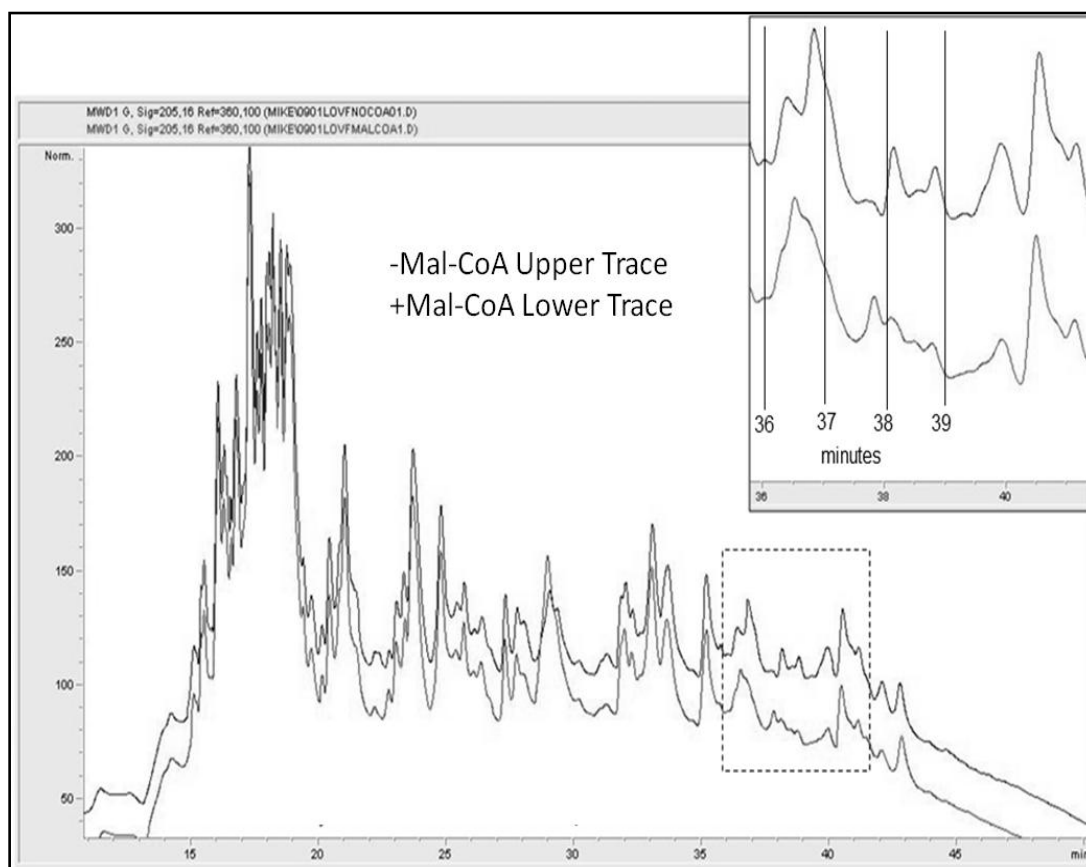
Figure 28: HPLC trace alignment of LovF digests. Comparison of the peptide fragmentation patterns of a typical tryptic digests of *holo*-LovF. Samples compared were either incubated with malonyl-CoA or not prior to treatment with trypsin. The inset is a magnified view of the peak profile between 36- 41 minutes highlighting the peak shifting that correspond to the ACP active-site containing peptides when loading malonate to the phosphopantetheinyl moiety.

The first samples analyzed by HPLC were a tryptic digest of 100 μg of *holo*-LovF compared to a tryptic digest of *holo*-LovF incubated with 0.5 mM malonyl-CoA. When comparing the two HPLC trace patterns of these two samples, the peaks aligned very well overall (Figure 28). It was found that there were two sets of noticeable peak shifts; the first seen between 36-37.5 minutes and the second set between 37.75-38.5 minutes. As the LovF digest was separated by HPLC, each 1 mL fraction that was collected consisted of peptides eluted over the span of 1 minute (e.g. from 34min to 35 min). Therefore, because some of the sets of peak shifting occurred over the span of slightly over a minute, all HPLC fractions collected between 34 mins and 40 minutes were frozen then lyophilized prior to analysis by mass spectrometry. Each lyophilized fraction was re-dissolved in a water/methanol/formic acid mixture to facilitate protonation of the peptides and desolvation during electrospray ionization. When samples were introduced into the mass spectrometer numerous peptides resulting from the digestion with trypsin were observed. One approach to identifying the active site containing a complex mixture of peptides would involve collecting high-resolution/high-mass-accuracy spectra, then averaging the spectra could be carried out using Thermo-Finnigan Qualbrowser software to improve the mass accuracy, and finally generate a mass list based on the monoisotopic masses of each peptide would be generated from the *m/z* and charge. The monoisotopic masses on the mass list can then be input into peptide-mapping program, along with the aminio acid sequence, to map peptides to different regions of the protein.
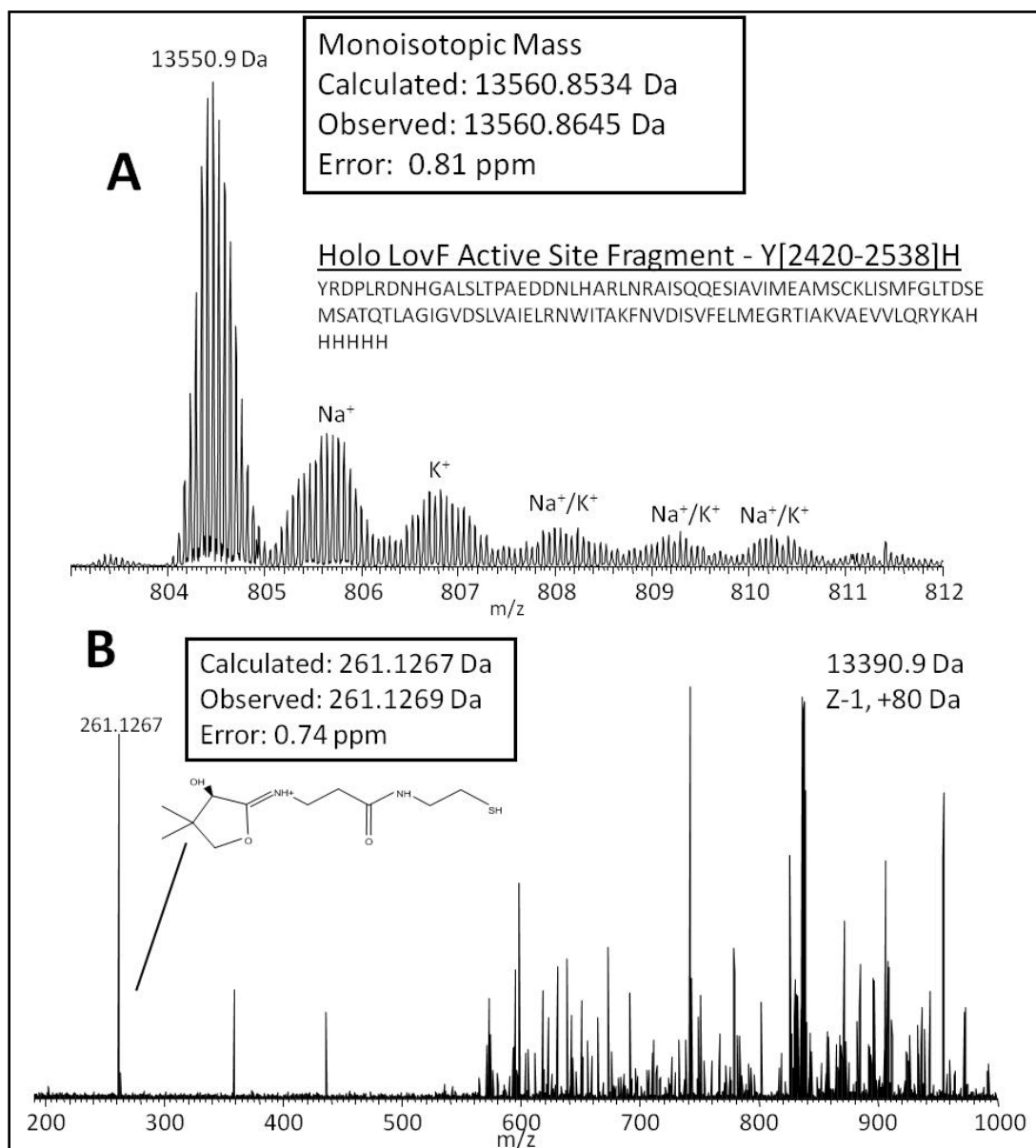
Figure 29: LovF holo-ACP active-site fragment and PPant ejection. (A) The high-resolution mass spectrum of the larger-sized ACP active site fragment. The observed mass is in close agreement to the calculated mass of this peptide. (B) High-resolution MS/MS of the pantetheinyl ejection ion at 261.1267 *m/z* confirms the identity of this peptide as the ACP active site of LovF. The resulting phosphoserine form of the active site peptide was detected with high mass accuracy.

In the case of the LovF digests, a slightly different approach was taken. Each fraction was analyzed using FT-MS to first collect high-res/high-mass-accuracy spectra of all peptides between 200 m/z – 2000 m/z. Next, the same fraction would be analyzed using the lower resolution front-end linear trap as the detector, and a MS/MS fragmentation method was programmed to fragment the most abundant ions in the spectra, then add those to a large exclusion list, so that the 250 most abundant ions in fraction were fragmented with CID. In real-time, MS/MS spectra were surveyed for the presence of a phosphopantetheinyl-ejection ion at approximately 261.1 m/z. In essence, the PPant ejection assay was used to initially identify ions as ACP active-site candidates, as opposed to using the monoisotopic mass and peptide mapping. The precursor ion that yielded each MS/MS spectra containing an ion signal near 261.1 +/- 0.5 m/z was located and evaluated the previously acquired FT-MS spectrum of the same fraction being analyzed. That method, coupled with the fact that HPLC peak shifts were clearly detected, allowed the efficient identification ions that contained the phosphopantetheinylated ACP active-site.

In the 36-37 minute fraction it was determined that there multiple charge states of one particular ion that consistently yielded a signal at 261.1 m/z in the MS/MS spectra. The 37-38 minute fraction also contained a second ion with multiple charge states that consistently yielded a peak near 261.1 m/z. The ion in the 36-37 minute fraction was determined to have a mass of 13,650.85 Da, which matched the mass of C-terminal peptide resulting from cleavage after Lys2419 plus the additional of 340.1 Da phosphopantetheinyl moiety (Figure 29A). The ion identified in the 37-38 minute

fraction was determined to have a mass of 10,440.25 Da, corresponding to a shorter C-terminal peptide resulting from cleavage after Arg2447 plus the additional of 340.1 Da phosphopantetheinyl moiety. These two peptides both contained the Ser2490 which is the highly conserved site of phosphopantetheinyl attachment.   Both ions were subjected to CID and the resulting MS/MS spectra were collected using the FT-MS to identify mass the proposed PPant ejection ion with high mass accuracy. In fact, the ion was found to have a mass of 261.1269 Da, which is accurate to with 0.74 parts per million of the calculated mass of the pantetheinyl ejection ion (261.1267 Da) (Figure 29B).   In addition, the dehydroalanine and phosophoserine forms of the precursor active-site containing peptide that result after PPant ejection by CID were identified in the MS/MS spectra. This data combined clearly showed that the soluble LovF was successfully phosphopantetheinylated.

Digested samples of *holo*-LovF incubated with malonyl-CoA were already prepared to serve as a means of identifying which HPLC fractions showed mass shifts corresponding to loading of the ACP active site PPant moiety. Therefore, it was a relatively simple task to identify which fractions should be analyzed by mass spectrometry to evaluate whether or not malonate was bound to PPant. The 36-37 minute and 37-38 minute fractions were analyzed using the same means as described above for analyzing the unloaded ACP active-site containing peptides, with the difference being that the ejected PPant ion was expected to have an addition of +86 Da, corresponding to the mass of malonate. Both the 13,650.85 Da and the 10,440.25 Da peptides were found to have a +86 Da mass shift, resulting in peptides with masses of 13736.86 Da and 10526.25 Da respectively. There was virtually 100% conversion

of the holo-LovF active-site peptides to the malonate-loaded form. Fragmentation of these peptides by CID resulted in the ejection of a prominent ion with a mass of 347.1277 Da which agrees with the calculated mass of the ejected pantetheine plus the mass of malonate (347.1271 Da) (Figure30).
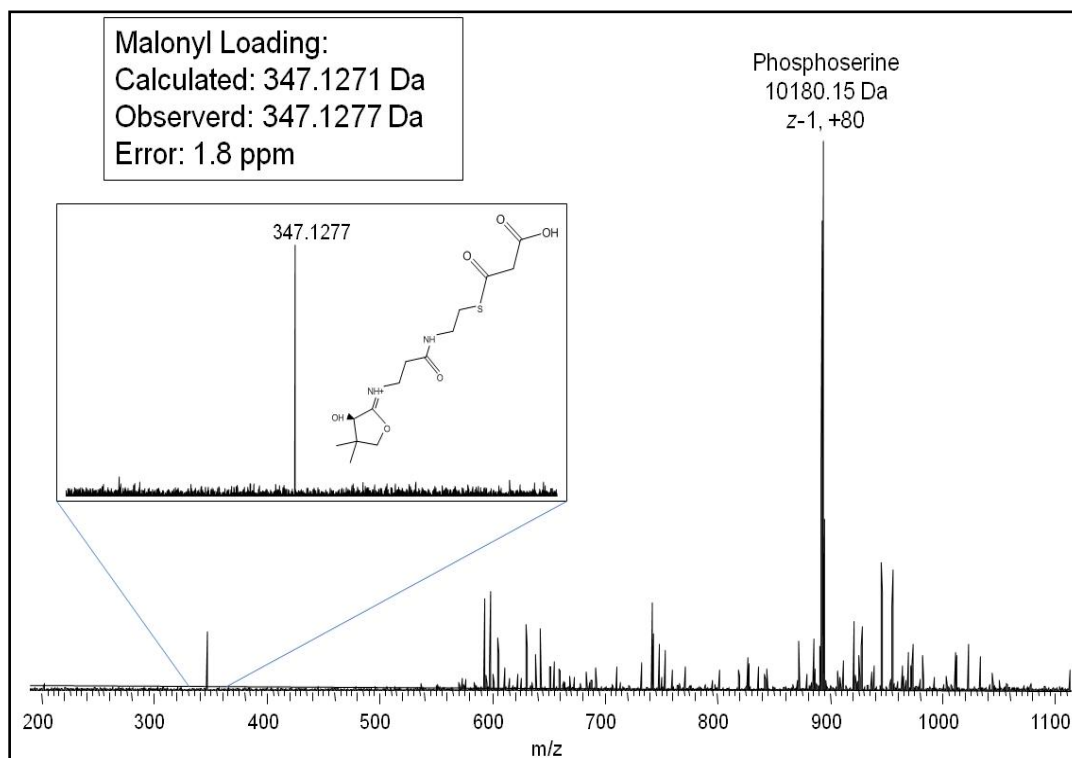
Figure 30: Malonate-loaded PPant Ejection Ion. Detection of the malonate-loaded (+86 Da) PPant ejection ion was achieved by CID of the ACP active-site containing precursor peptide. The resulting MS/MS spectra were collected by FT-MS to provide high mass-accuracy detection and verification of the identity of the intermediate. The resulting charge-loss phosphoserine form of the precursor ion provides confirmation that the 347.1277 *m/z* is the product of pantetheinyl ejection. Malonate loading was detected on both the 13,650.85 Da and the 10,440.25 Da ACP active-site peptides.

It was clear from these experiments that the samples of holo-LovF provided by the Tang laboratory were indeed phosphopantetheinylated and that substrate (malonyl-CoA) loading was detectable by mass spectrometry. Therefore, research was focused towards attempting generate and detect the phosphopantetheine-bound product, α-

methylbutyrate, in an attempt to evaluate the efficiency of conversion. Initially, reaction mixtures were prepared containing 0.5 mM malonyl-CoA, 0.5 mM SAM, and 0.5 NADPH, but there were no indications that α-methylbutyrate formation occurred when analyzed by mass spectrometry. Various reaction times and conditions were attempted, but the end result in each case was that α-methylbutyrate formation could not be detected by FT-MS, and the end result was only detection of malonate bound to the LovF ACP. At the time, the hypothesis that α-methylbutyrate formation was not detected on the ACP active-site because of a combination of possible hydrolysis of the final product and slowing of the elongation steps catalyzed by the ketosynthase (KS) domain. It was likely that at steady state, the decarboxylative condensation step between acetate on the KS domain and the malonate on the ACP to form acetoacetate occurs very slowly. Therefore, the concentrations of malonyl-CoA incubated with *holo*-LovF were adjusted so that the enzyme was incubated with 1, 2, 4 and 8 molar equivalents of malonyl-CoA. In these series of reactions, α-methylbutyrate formation was still not detected. As the concentration of malonyl-CoA used in the various reactions with LovF was increased, a detectable conversion of the *holo*-ACP active site peptide to malonyl-S-ACP was detectable by mass spectrometry (Figure 31). When LovF was incubated with 1 molar equivalent of malonyl-CoA, low levels of malonte-loading were seen. Evidence of subsequent decarboxylation catalyzed by the KS domain was also present in the form of acetyl-S-ACP. The acetate loaded onto the ACP active site was clearly detectable. However, as the concentration of malonyl-CoA was increased a corresponding increase in the amount of malonate loading was observed, without a corresponding conversion to acetate. This confirms that there is

an accumulation of acetate on the KS domain and the subsequent condensation step is occurse very slowly at steady-state. This is consistent with the rate of secondary metabolite biosynthesis.

As previously noted, detectable in these samples was a conversion of the *holo*-ACP active site peptide to acetyl-S-ACP. The observed mass of the loaded active site peptide was within 1 ppm of the calculated mass plus the addition of acetate. The identity of the decarboxylated acetate-intermediate was then verified by PPant ejection assy. By isolating the peak that corresponded to the ACP active site peptide +42 Da and subsequent fragmentation resulted in the ejection of an ion with mass of 303.1382 Da (Figure 32), consistent with the theoretical mass of the acetyl-loaded pantetheinyl ejection ion (303.1373 Da). As the concentration of malonyl-CoA was increased, the ACP active site was found to contain significantly more of the malonate intermediate and less of the acetate-intermediate.
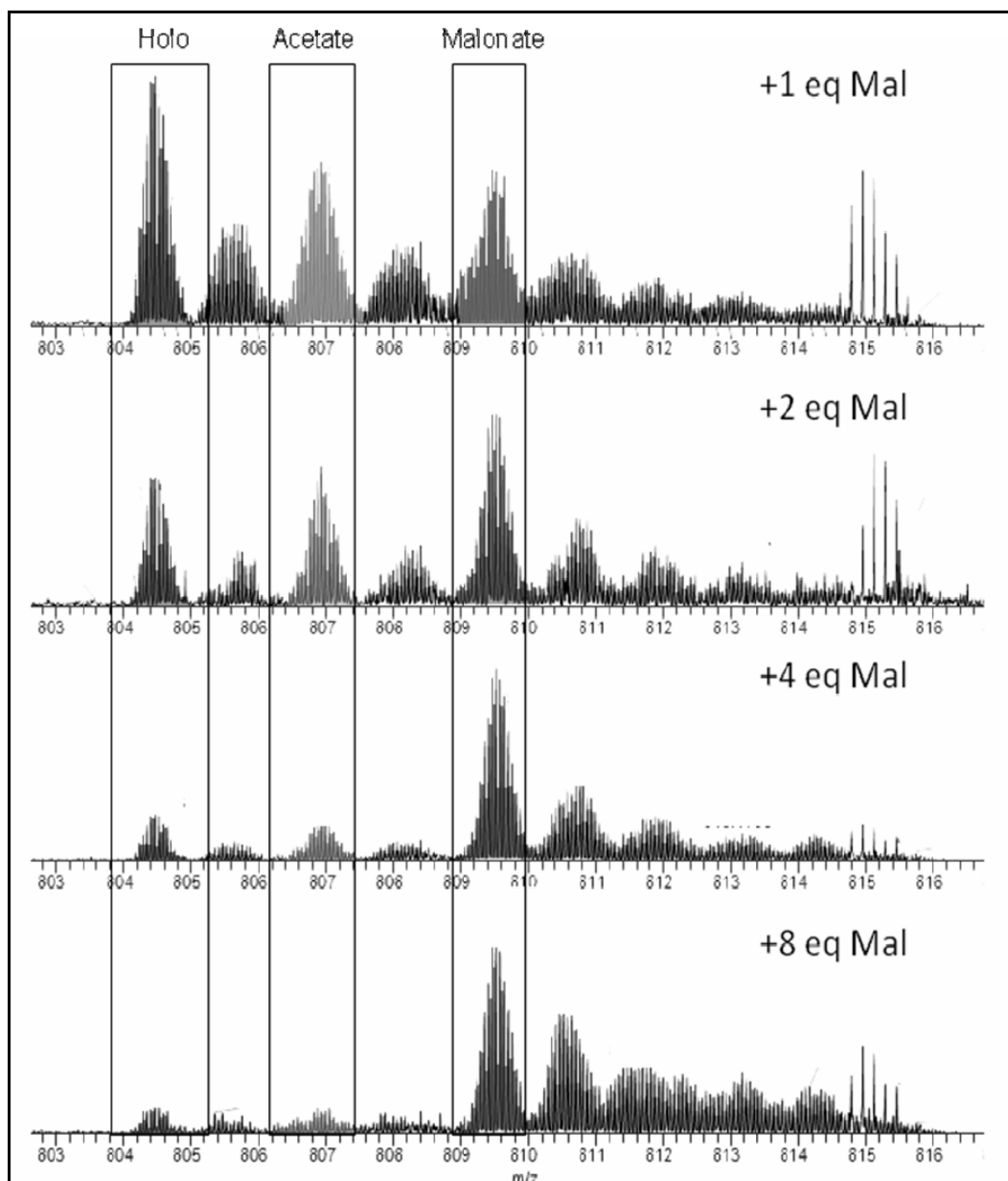
Figure 31: Concentration Dependent Substrate Loading of LovF. Samples of *holo*-LovF were incubated with 1, 2, 4, and 8 molar equivalents of malonyl-CoA. At low concentrations of malonyl-CoA, acetate from the decarboxylation of malonate were observed bound to the ACP active-site containing tryptic peptide. Samples that were incubated with higher concentrations of malonyl-CoA show a shift towards being predominantly loaded with malonate. The lack of conversion of the malonate to acetate at higher concentration of malonyl-CoA show that there is likely an accumulation of acetate at the KS active site and that condensation between acyl units is slow. Additional peaks seen to the right of each of the primary peaks represent the presence of nonconvalent adducts of sodium from the buffers used.
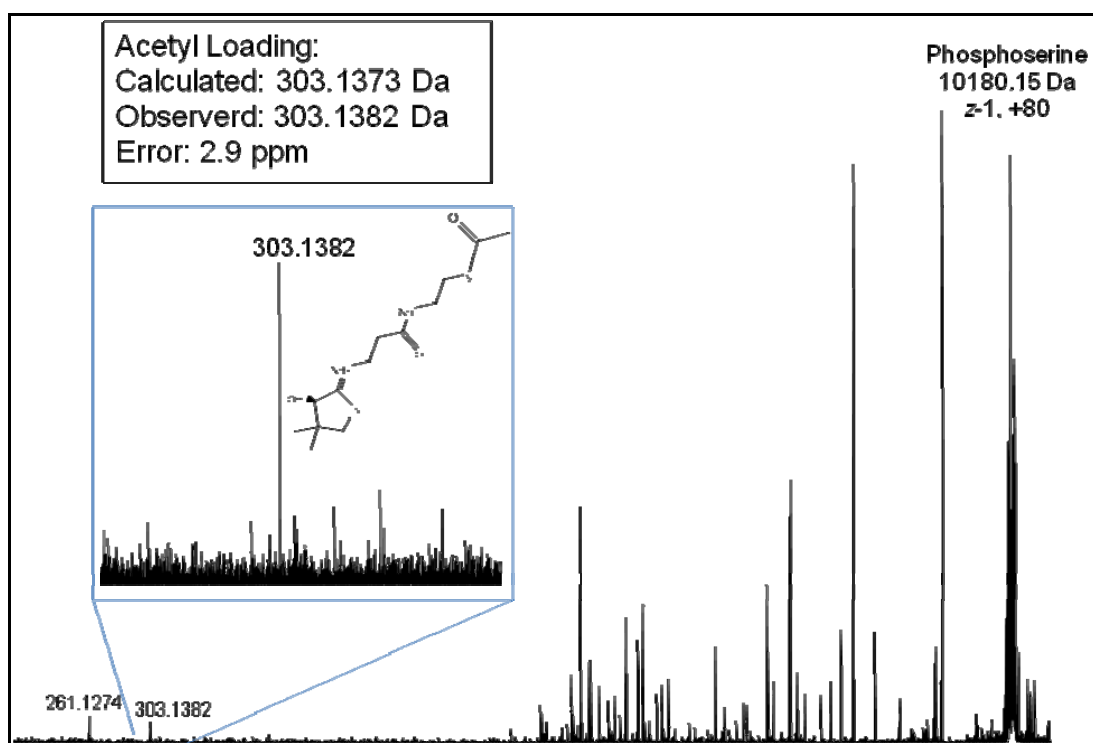
Figure 32: Acetate-loaded PPant Ejection Ion. Detection of the acetate-loaded (+42 Da) PPant ejection ion was achieved by CID of the ACP active-site containing precursor peptide. The resulting MS/MS spectra were collected by FT-MS to provide high mass-accuracy detection and verification of the identity of the intermediate. The resulting charge-loss phosphoserine form of the precursor ion provides confirmation that the 303.1382 *m/z* is the product of pantetheinyl ejection.

For every two moles of malonyl-CoA made available to LovF, one mole of α-methylbutyrate should be produced. Therefore, it was surprising that samples of LovF incubated with a 1-to-1, or 2-to-1 concentration of 1 molee malonyl-CoA to LovF don't yield the final product. During one of the early preparation of the LovF reactions with differing molar equivalents of malonyl-CoA, a simple calculation error was made that resulted in the addition of 1/10 dilution of the intended quantities of malonyl-coA. For example, instead of adding 1 meq, or 14.4 µM, of malonyl-CoA to

LovF, only 1.44 µM of malonyl-CoA was added. This calculation error was not recognized until after mass spectrometric analysis was carried out. From the digested sample of LovF that was incubated with 1/10 of 4 molar equivalents, it was found that the HPLC fraction collected between 38-39 minutes contained mostly unloaded *holo*-ACP active sites. However, isolation of a relatively wide *m/z* range around the ACP active-site peptides followed by CID resulted in a very low abundance ion with a mass near 345 *m/z* in the MS/MS spectra (Figure33A). The higher sensitivity of the linear trap for low m/z ions resulting from fragmentation by CID allowed for the detection of this ion, whereas the FT-MS detector was not able to detect such a low *m/z* ion that resulted from CID on account of the 1/3 rule and the low intensity of the higher charge states of the precursor ACP active-site peptides. The drawback to using the linear trap quadrupole detector is that the mass accuracy was not high enough to allow definitive identification of the 345 *m/z* ion. It was possible that this ion could be the PPant ejection ion (261.1267 Da) loaded with the 3-keto-butyrate (acetoacetate) intermediate (+84.0211Da) that has a mass of 345.1479 Da, or the PPant ejection ion loaded with the α-methylbutyrate (+84.0575) that has a total mass of 345.1843 Da. In order to determine the identity of this ion, MS/MS spectra of the 38-39 minute fraction that contained 1.44 µM of malonyl-CoA, plus SAM and NADPH cofactors, was fragmented using source fragmentation. This method involves applying a voltage to a sample being ionized by ESI to induce fragmentation just before it enters the mass spectrometer, and all ions in a sample are fragmented simultaneously, resulting in a very large quantity of ions present in the spectrum. Because the ions in the sample are not fragmented in the collisions cell by CID, FT-MS can be used to obtain high mass-

accuracy spectra without the limitations of the 1/3 rule. Using this method, the mass of the previously observed 345 *m/z* species was determined to be 345.1843 Da (Figure33B), corresponding to the pantetheinyl ejection ion plus the mass of α-methylbutyrate. The ion at 345.1843 *m/z* was subjected to a subsequent round of CID fragmentation in order to generate low-resolution MS^3 fragmentation spectra of the α-methylbutyrate–bound PPant ejection ion (Figure33C). Six of the diagnostic peaks[81] were shifted by the mass of the covalently bound final product (84 Da), while one of the three substrate-independent ions was detected. These ions provide additional confirmation of the identity of the ppant ejection ion. The * indicates the α-methylbutyrate loaded phosphopantetheinyl ejected ion at 345.1 m/z.
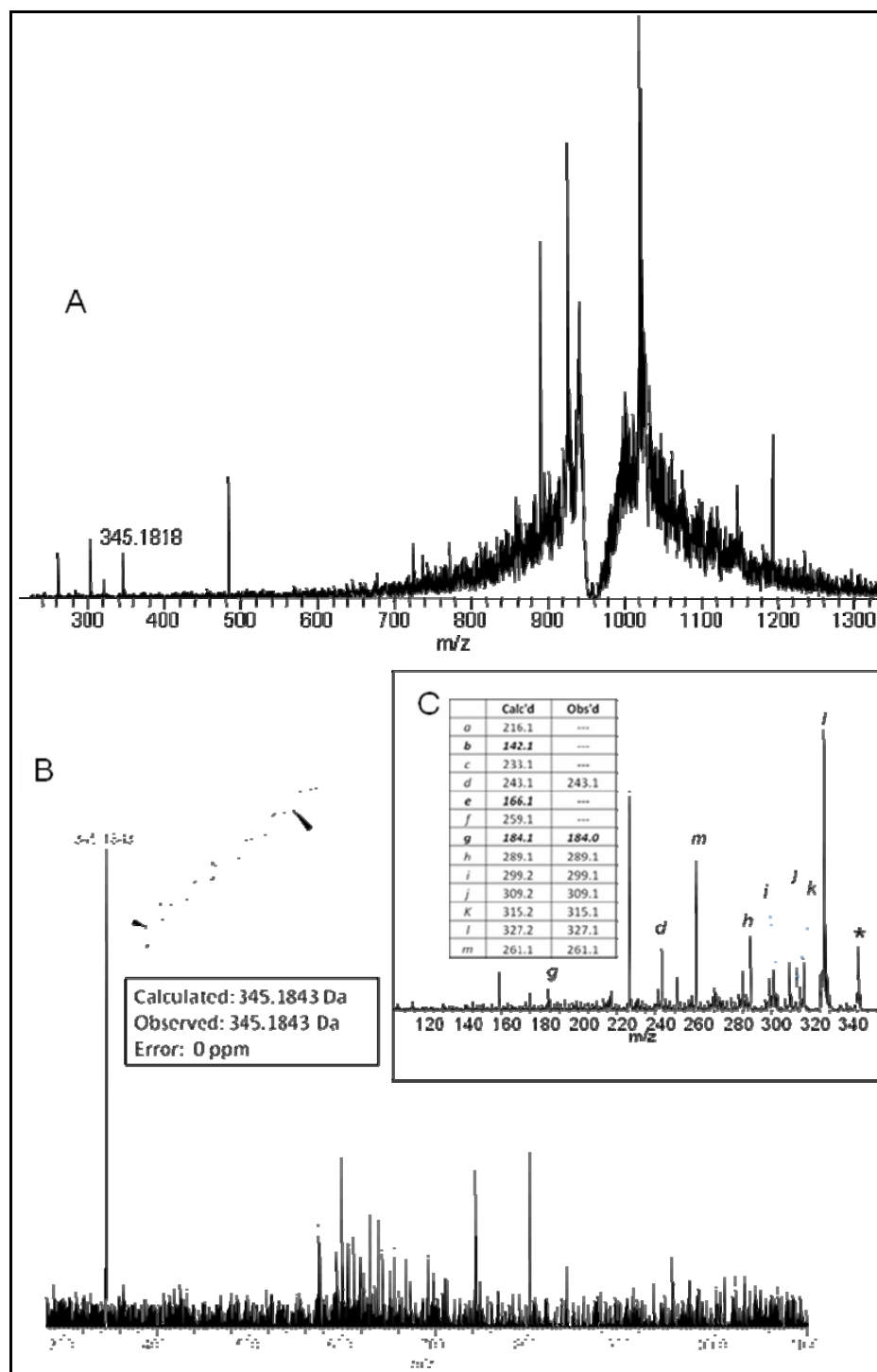
Figure 33: Detection of α-methylbutyrate by PPant Ejection Assay. (A) Detection of unverified species using high-sensitivity/low-resolution detector. (B) Isolated α-methylbutyrate-loaded PPant ejection ion generated by source fragmentation. (C) Subsequent CID of the 345 m/z ion reveals fragmentation pattern consistent with that of phosphopantetheine, as some of the thiol-containing fragments also showed a mass shift of 84 Da corresponding to the presence of α-methylbutyrate.

4.5. LovF Discussion

LovF presented an interesting opportunity to utilize and develop off-line HPLC analysis methods. The off-line method requires significantly more preparatory work prior to mass spectrometric analysis, in the form of HPLC separation, fraction collection, and sample lyophilization, than an analagous liquid chromatography (LC)-MS technique. The advantage of the off-line method was that it allows for more thorough characterization of individual fractions. This pays off when tryptic digests are analyzed by mass spectrometry in which the active-site containing peptides are quite large, have a high charge state, or are in low abundance. With a 6-7 Tesla magnet, LC-FT-MS simply would not have been sensitive enough to allow the identification of the α-methylbutyrate product on the LovF active site fragment. The techniques carried out in the analysis of LovF are powerful means of detecting intermediates in the biosynthesis of natural products by PKS and NRPS enzymes.

From these experiments, it was clear that the heterologously express LovF make in the Tang laboratory was a fully-functional phosphopantetheinylated PKS. The activity of this enzyme was successfully monitored by mass spectrometry. While LovF could efficiently take the malonyl-CoA substrate and load it onto the PPant moiety of the ACP active site, it was determined that detection of the fully elongated product could only be detected at very low concentrations. This was most likely due to accumulation of acetate on the KS domain. At higher concentrations of malonyl-CoA, the ability to clearly detect the final product, α-methylbutyrate, by mass spectrometry was not possible, as the malonyl-S-ACP signal intensity overwhelmed

the α-methylbutyrl-S-ACP signal. Even at low concentrations of malonyl-CoA, the full-length product was present at very low levels on the ACP active-site containing peptide and creative use of source fragmentation was the only way to confirm the identity of the PPant-bound species. Additional experiments aimed at detecting the catalytic activity of the KR and MT domains have been unsuccessful. Most likely, the converstion of acetoacetate to α-methylbutyrate, catalyzed by the KR and MT domains, occurs very quickly. On going work will be aimed at first priming *apo*-LovF with acetoacetyl-CoA, and then attempting to detect KR and/or MT catalyzed conversion to new intermediates.

LovF was a good PKS to begin with when developing the off-line HPLC/ FT-MS methods in the Dorrestein lab. The enzyme was very amenable to tryptic digestion and the HPLC peak shifts were very pronounced. When analyzing the digested fractions by high-resolution mass spectrometry, the ACP active-site containing peptide ionized very well and has strong signal intensity. Most likely, the presence of the C-terminal His-tag, containing several basic histadine residues, facilitated the extensive protonation of the peptide by the acidic solution used for electrospray ionization. Finally, LovF digestion patterns resulted in the formation of two different sizes of active-site containing peptides. The presence of only two tryptic peptides greatly simplifies the process of mapping peptides. Others using the off-line HPLC/FT-MS methods to study NRPS have found upwards of 6 different sizes of tryptic peptides containing phosphopantetheinylated active sites.

The iterative nature of fungal polyketides, such as LovF and LovB, make these megasynthases more challenging to study than NRPS or type I noniterative PKS

systems. The domains of iterative PKSs are used for multiple iterations of chain elongation, and thus their function cannot be probed by mutational inactivation of individual domains. Each domain is indispensible for the production of the fully elongated product. An iterative PKS such as LovF is more easily characterized as it synthesizes a relative short diketide side-chain of lovastatin which results from only a single condensation between two acyl-units. While there are still additional avenues to try to probe for the other intermediates covalently bound to the ACP of LovF, it will be interesting to investigate side-chain transfer mediated by LovD using high-resolution mass spectrometry. It will be interesting to evaluate how efficiently the off-loading of the α-methylbutyrate product from LovF onto LovD proceeds, by detecting for mass shifts in LovD corresponding to the loaded protein. Soluble LovD has already been isolated and purified by the Tang laboratory. Tryptic digestion may not even need to be employed when analyzing the side-chain transfer because LovD is small enough (46kDa) to be analyzed intact by mass spectrometry. The methods outlined in reference 47, coupled with the work carried out with LovF, have lead to the development of versatile experimental platform that applied to the study of other iterative polyketide synthases in the Dorrestein laboratory.

Chapter 5: Future Investigations into Linear Polyketide Biosynthesis by Fungal

Megasynthases

5.1. Future Outlook

The use of high-resolution mass spectrometry in the investigation of natural products biosynthesis will continue to be implemented as future work is carried out to better understand the step-wise construction of compounds of polyketide and non-ribosomal peptide synthase origin. The presence of the phosphopantetheinyl moiety that serves as an "anchor" for the growing intermediate in both biosynthetic paradigms, coupled with the propensity for PPant ejection by CID, allows mass spectrometry to be a very useful investigative tool. While high-resolution mass spectrometry provides the means necessary for detecting small changes in mass to PKS and NRPS intermediate, PPant ejection detected on a high-resolution FT-ICR instrument can allow for sub-part-per-million mass accuracy when determining the structure or identity of intermediates. In addition to targeting the phosphopantetheinylated active site, mass spectrometric techniques can be used to analyze substrate specificity or starter unit selection of the starter-unit ACP-transacylase (SAT) and additional AT domains of PKSs[47] or adenylation domains of NRPSs[134]. On-going efforts involving high-resolution mass spectrometry are being carried out to investigate the biosynthetic programming of iterative polyketides synthases of fungal origin Pks4 from *Gibberrella fujikuroi*.

5.2. Pks4 – Fungal Iterative Polyketide Synthase

Pks4 from *G. fujikuroi* is a nonaketide synthase that produces the anticancer compound bikaverin.[135] The Tang laboratory has been able to successfully express and purify functional Pks4 from *E. coli* and demonstrate that it was fully functional

and capable of synthesizing a polyketide product (SMA76a) with a mass consistent to that of a cyclized nonaketide that has undergone two dehydrations.[135]   SMA76a is normally further modified in *G. fujikuroi* to produce bikaverin. Also demonstrated by the Tang laboratory, Pks4 is capable of accepting both malonyl-CoA and octanoyl-CoA as starter acyl-units.   Surprisingly, Pks4 preferentially selected octanoyl-CoA over malonyl-CoA as starter unit, and inclusion of octanoyl-CoA as the started unit followed by subsequent elongation using malonyl-CoA resulted in the formation of new products, SMA76b and SMA76c. The SAT domain of Pks4 was determined to lack the typically highly-conserved GXCXG[136] active-site sequence typically found in PKS SAT domains.   This led to the hypothesis that the SAT domain of PKS is possibly inactivated and that the octanoyl-unit is loaded directly onto the KS domain[135] and that the MAT domain catalyzes extender-unit selection.   Then, direct thioester exchange can occur between the KS and ACP domains of Pks4.

Heterologously expressed *holo*-Pks4 was generously provided by the Tang laboratory to be used for the biochemical characterization of Pks4. The first sets of experiments involved limited-tryptic digestion of the enzyme followed by capillary-LC-MS for the purposes of active site mapping. Digested samples of Pks4 were analyzed in a data-dependent fashion in the top five most abundant ions in each MS spectrum were selected for data dependent fragmentation.   The resulting MS/MS spectra were analyzed using the peptide sequencing the software InSpect.[38]   In addition, the spectra were analyzed for evidence of phosphopantetheinyl ejection, specifically the ejection ion signature at 261.1 *m/z*.   However, this method only provided tentative identification of the thioesterase/Claisen-like condensation domain

of Pks4, not the SAT, MAT, KS, or ACP domains which would be the expected domains to show mass shifts due to bound intermediates of SMA76a. However, without being able to map any of these active sites, it would be highly unlikely that covalently bound intermediates could be detected. Therefore, trypsin digestion followed by off-line HPLC separation then direct infusion of HPLC fractions for analysis by FT-MS was conducted (identical to the method used for analysis of LovF). To expedite the process of potentially identifying active sites, samples of Pks4 were pre-incubated with malonyl-CoA prior to digestion for comparison to samples that were not pre-incubated with any acyl-CoA. The goal was to identify any HPLC profile differences that might suggest which fraction either the ACP or an SAT/MAT active site fragment was located.
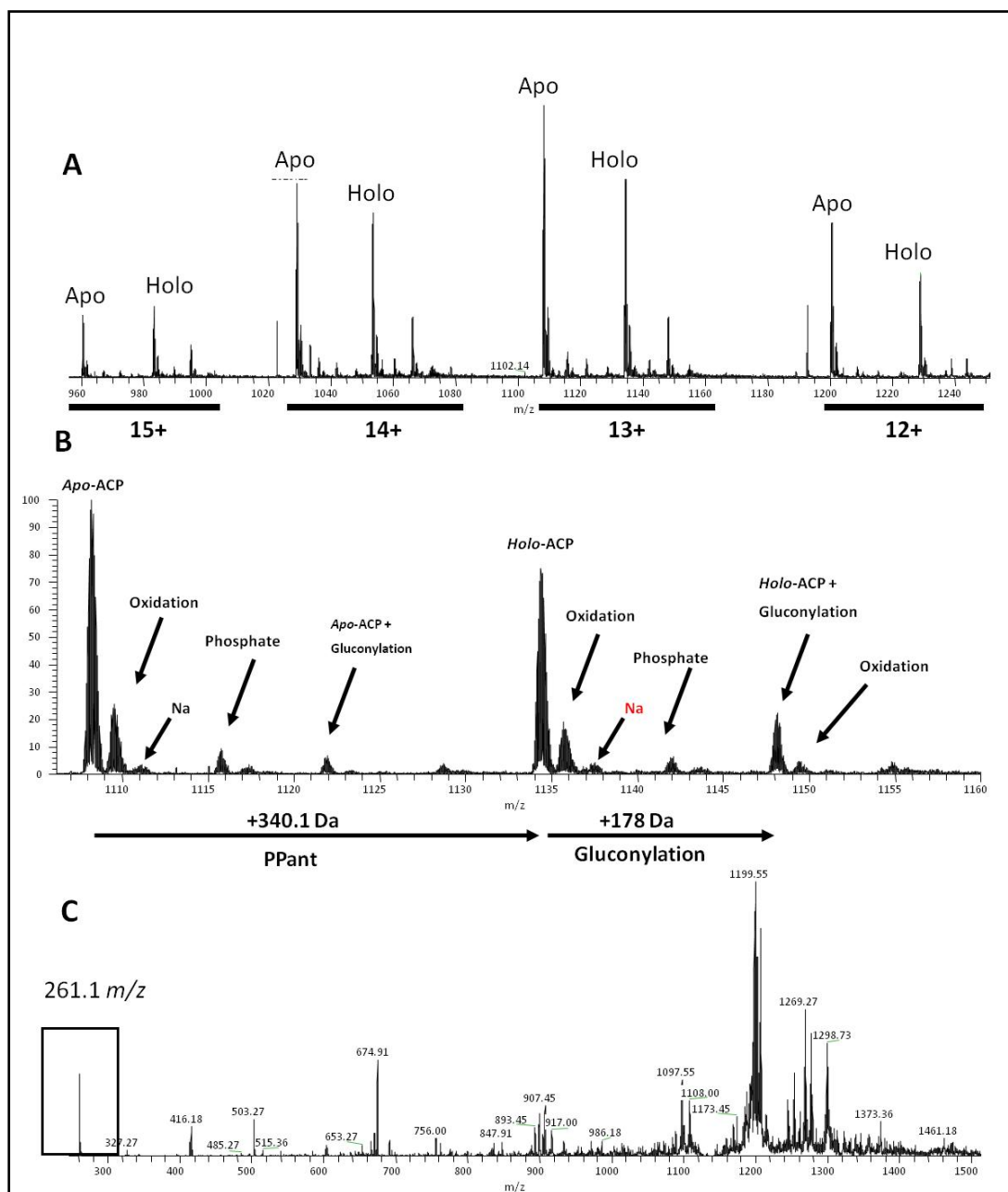
Figure 34: High-resolution MS analysis of Pks4 ACP monodomain protein. (A) Multiple charge states of the intact ACP stand-alone protein detected. (B) A detailed view of the 13+ charge state reveals the extent of phosphopantetheinylation of the ACP protein. In addition to the *apo-* and *holo-* enzymes detected, gluconylation of the protein is observed. (C) MS/MS of the ACP domain yields a pantetheinyl ejection ion with excellent intensity.

The HPLC traces showed two regions of clear peak differences in the fractions collected between 31-32 minutes and then from 41-43 minutes. The spectrum of the fraction at 31 minutes clearly showed an ion with a mass of 11,030.9 Da in both treated samples. However, this did not match up to either the ACP active site or the MAT active site, nor was PPant ejection detected when it was fragmented. The peak shifts seen after 44 minutes were difficult to characterize because the ion intensities were too low. Therefore, no conclusions can be drawn at this time from those fractions. However, considering earlier data collected it is possible that this peak present on the ACP that is diminished by the inclusion of substrate may represent the ACP active-site, which when digested is poorly ionizable and thus not detectable by mass spectrometry. A slightly larger peptide with a mass of 11143.9 Da was detected which does not match the mass of any peptides generated by partial or complete tryptic digestion. Using peptide mapping software PAWS the mass of the peptide was determined to correspond to a non-tryptically generated peptide (V933-G1037) that did contain the MAT domain active-site at Ser_{970}. This data alone is not enough to make any solid conclusions, rather it presents some of the challenges one may face by following general limited-digestion mass-spectrometric method.

An alternative strategy for investigating catalytic steps of the Pks4 biosynthetic pathway may involve a similar approach to that taken by the Tang laboratory in which they expressed a stand-alone KS-MAT didomain along with a stand-alone ACP domain which were determined to be functional.[137] The Tang laboratory has also expressed soluble Pks4 ACP and SAT-KS-MAT didomain. In light of the challenges of analyzing the intact Pks4, an alternate strategy for the evaluation of starter unit

selection and subsequent catalytic steps such as elongation cyclization would involve what has been called a "deconstructive approach"[48] by Crawford et al. In this approach, steps such as starter unit selection can be analyzed my mass spectrometry. For instance, the SAT-KS-MAT tridomain, if functional can be incubated with either octanoyl-CoA or malonyl-CoA, then digested with trypsin, separated by HPLC and analyzed by high-resolution mass spectrometry to map the location of covalent modifications such as the attachment of malonyl to the active-site of the MAT domain, or to determine the location of octonyl loading. In Crawford's paper, confirmation of the functional role of each catalytic domain was evaluated by analyzing the different intermediates bound to the phosphopantetheinyl moiety of the ACP monodomain when it was incubated with substrate and different combinations of the various catalytic domains involved in biosynthesis of aflatoxin B_1 by PksA.

The heterologously express Pks4 ACP monodomain has been analyzed by mass spectrometry both as an intact protein and after trypsin digestion. Samples of the ACP protein were prepared using C18 and C4 zip-tips (Millipore – Billerica, MA) and ionized with ESI and analyzed with FTMS. Prepared samples of ACP yielded several visible groups of peaks that represented the ACP protein at various charge states (Figure 34A). We could clearly distinguish ionized ACP with charges of 11+ through 16+. The most abundant charged species was the 13+ ion. Closer inspection of the 13+ charge state alone revealed three prominent species with m/z values of 1108.2, 1134.4 and 1148.2. The respective most abundant masses of these species were determined to be 14394.0 Da, 14734.1 Da, and 14913.1 Da. Based on the amino acid sequence provided by the Tang laboratory, the estimated weight of the holo-ACP

protein was calculated to be 14865.9 Da, equal to the mass of the ACP monodomain (14525.8 Da) plus the addition of PPant (340.1 Da). The discrepancy between this mass with any of those observed can most likely be explained by truncation of the starting methionine at the N-terminal, resulting in a loss of 131 Da. With this in mind, the calculated mass of holo-ACP would be 14734.9 Da. The species at m/z 1134.4 with an observed most abundant mass of 14734.1 Da (monoisotopic mass 14727.2 Da) was identified to be the holo-ACP protein. The species observed at 1108.2 m/z with an observed mass of 14394.0 Da was determined to be the apo-ACP. The third species at m/z 1148.2 was determined to have a mass of 14913.1 Da. This represents the holo-ACP that has undergone both N-terminal truncation of methionine (-131 Da) and N-terminal gluconylation (+178 Da) (Figure 34B), a known occurrence observed by in heterologously over-expressed proteins with His-tags.[138] When the holo-peptides were subjected to CID, PPant ejection could be detected (Figure 34C).

Tryptic digestion of the ACP monodomain was carried out prior to analysis by mass spec. The resulting spectra revealed only limited tryptic peptides from the Pks4 ACP. In a significantly smaller pool of peptides, a phosphopantetheinylated ACP active-site peptide could not be found. This supports the hypothesis that this particular active site may not be amenable to analysis to mass spectrometric analysis due to poor ionization or a low positive charge preventing the peptide from falling within the instruments detectable range of 200-2000 *m/z*. This also provides support for using intact ACP for mass spectrometric analysis of Pks4 catalytic domain function in future.

References

135

1.	Natural products as sources of new drugs over the last 25 years. Newman DJ, Cragg GM, J Nat Prod (2007); 70(3):461-77

2.	Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF, Proc Natl Acad Sci U S A (2004); 101(26):9528-33

3.	Electron capture dissociation for structural characterization of multiply charged protein cations. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW, Anal Chem (2000); 72(3):563-73

4.	The role of electron capture dissociation in biomolecular analysis. Cooper HJ, Håkansson K, Marshall AG, Mass Spectrom Rev (2005); 24(2):201-22

5.	Application of silicon nanowires and indium tin oxide surfaces in desorption electrospray ionization. Pól J, Novák P, Volný M, Kruppa GH, Kostiainen R, Lemr K, Havlíček V,Eur J Mass Spectrom (Chichester, Eng) (2008); 14(6):391-399

6.	Fundamentals of Traveling Wave Ion Mobility Spectrometry. Shvartsburg AA, Smith RD, Anal Chem (2008): 80(24), 9689-9699

7.	High-throughput quantitative analysis by desorption electrospray ionization mass spectrometry. Manicke NE, Kistler T, Ifa DR, Cooks RG, Ouyang Z., J Am Soc Mass Spectrom (2009); 20(2):321-325

8.	Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. Takáts Z, Wiseman JM, Gologan B, Cooks RG, Science (2004); 306(5695):471-3

9.	Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. Udeshi ND, Compton PD, Shabanowitz J, Hunt DF, Rose KL, Nat Protoc (2008); 3(11):1709-17

10.	A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V, Coon JJ, J Proteome Res (2008); 7(8):3127-36

11.	A nanostructure-initiator mass spectrometry-based enzyme activity assay. Northen TR, Lee JC, Hoang L, Raymond J, Hwang DR, Yannone SM, Wong CH, Siuzdak G, Proc Natl Acad Sci U S A (2008); 105(10):3678-83

12.	Clathrate nanostructures for mass spectrometry. Northen TR, Yanes O, Northen MT, Marrinucci D, Uritboonthai W, Apon J, Golledge SL, Nordström A, Siuzdak G, Nature (2007); 449(7165):1033-1036

13.   Rapidly alternating transmission mode electron-transfer dissociation and collisional activation for the characterization of polypeptide ions. Han H, Xia Y, Yang M, McLuckey SA, Anal Chem (2008); 80(9):3492-3497

14.   Matrix-free high-resolution imaging mass spectrometry with high-energy ion projectiles. Nakata Y, Honda Y, Ninomiya S, Seki T, Aoki T, Matsuo J, J Mass Spectrom (2009); 44(1):128-36

15.   Lipid imaging with cluster time-of-flight secondary ion mass spectrometry. Brunelle A, Laprévote O, Anal Bioanal Chem (2009); 393(1):31-5

16.   Construction of a novel stigmatic MALDI imaging mass spectrometer. Hazama, Hisanao; Aoki, Jun; Nagao, Hirofumi; Suzuki, Ren; Tashima, Toshio; Fujii, Ken-ichi; Masuda, Katsuyoshi; Awazu, Kunio; Toyoda, Michisato; Naito, Yasuhide, Applied Surface Science (2008); 255(4):1257-1263

17.   Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. Chaurand P, Schwartz SA, Caprioli RM, Curr Opin Chem Biol (2002); 6(5):676-81

18.   The Orbitrap: a new mass spectrometer. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R, J Mass Spectrom (2005); 40(4):430-43

19.   The role of mass spectrometry in systems biology: Data processing and identification strategies in metabolomics. Van der Greef, Jan; van der Heijden, Rob; Verheij, Elwin R, Advances in Mass Spectrometry (2004); 16:145-165

20.   Pharmaceutically active secondary metabolites of microorganisms. A. L. Demain, Appl.Microbiol Biotechnol. (1999); 52, 455-463

21.   Recent developments in antitumour taxoids. Dubois, D. Guenard, F. Gueritte, Expert Opinion on Therapeutic Patents (2003); 13(12), 1809-1823.4

22.   Biosynthesis of hybrid peptide-polyketide natural products. L Du, B Shen, Current Opinion in Drug Discovery & Development (2001); 4:215-228

23.   Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Fischbach MA, Walsh CT, Chem Rev (2006); 106(8):3468-96

24.   Genetically Encoded Short Peptide Tags for Orthogonal Protein Labeling by Sfp and AcpS Phosphopantetheinyl Transferases. Zhou, Zhe, Cironi, Pablo, Lin, Alison J., Xu, Yangqing, Hrvatin, Sinisa, Golan, David E.,Silver, Pamela A., Walsh, Christopher T., Yin, Jun, Chem Biol (2007); 2(5):337-346

25.   Aminoacyl-coenzyme A synthesis catalyzed by adenylation domains. Linne U, Schäfer A, Stubbs MT, Marahiel MA, FEBS Lett (2007); 581(5):905-10

26.   Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM, Curr Opin Chem Biol (2001); 5(5):525-34

27.   Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. Kopp F, Marahiel MA, Curr Opin Biotechnol (2007); 18(6):513-20

28.   The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization. Hoyer KM, Mahlert C, Marahiel MA, Chem Biol (2007); 14(1):13-22

29.   Benzodiazepine biosynthesis in Streptomyces refuineus. Hu Y, Phelan V, Ntai I, Farnet CM, Zazopoulos E, Bachmann BO, Chem Biol (2007); 14(6):691-701

30.   Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL, Nature (2006); 440(7082):368-71

31.   Ablation of the otcC gene encoding a post-polyketide hydroxylase from the oxytetracyline biosynthetic pathway in Streptomyces rimosus results in novel polyketides with altered chain length. Peric-Concha N, Borovicka B, Long PF, Hranueli D, Waterman PG, Hunter IS, J Biol Chem (2005); 280(45):37455-60

32.   A gene cluster from a marine Streptomyces encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A. Li A, Piel J, Chem Biol (2002); 9(9):1017-26

33.   Cloning and heterologous expression of the epothilone gene cluster. Tang L, Shah S, Chung L, Carney J, Katz L, Khosla C, Julien B, Science (2000); 287(5453):640-2

34.   Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH, Chem Biol (2004); 11(6):817-33

35.   Biosynthesis of the enediyne antitumor antibiotic C-1027. Liu W, Christenson SD, Standage S, Shen B, Science (2002); 297(5584):1170-3

36.   Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. Guenzi E, Galli G, Grgurina I, Gross DC, Grandi G, J Biol Chem (1998); 273(49):32857-63

37.   Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in Trichoderma virens. Wei X, Yang F, Straney DC, Can J Microbiol (2005); 51(5):423-9

38.   InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V, Anal Chem (2005); 77(14):4626-39

39.   Computational methods for protein identification from mass spectrometry data. McHugh L, Arthur JW, PLoS Comput Biol (2008); 4(2):e12

40.   Quality assessment of peptide tandem mass spectra. Wu FX, Gagné P, Droit A, Poirier GG, BMC Bioinformatics (2008); 9(Suppl 6):S13

41.   Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C, Walsh CT, Chem Biol (1999); 6(6):385-400

42.   Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis. Stein DB, Linne U, Hahn M, Marahiel MA, Chembiochem (2006); 7(11):1807-14

43.   A PLP-Dependent Polyketide Chain Releasing Mechanism in the Biosynthesis of Mycotoxin Fumonisins in Fusarium verticillioides. Gerber R, Lou L, Du L, J Am Chem Soc (2009); Article ASAP

44.   A mechanism-based aryl carrier protein/thiolation domain affinity probe. Qiao C, Wilson DJ, Bennett EM, Aldrich CC, J Am Chem Soc (2007); 129(20):6350-1

45.   GNAT-like strategy for polyketide chain initiation. Gu L, Geders TW, Wang B, Gerwick WH, Håkansson K, Smith JL, Sherman DH. Science (2007); 318(5852):970-4

46.   Mechanistic and structural basis of stereospecific Cbeta-hydroxylation in calcium-dependent antibiotic, a daptomycin-type lipopeptide. Strieker M, Kopp F, Mahlert C, Essen LO, Marahiel MA, ACS Chem Biol (2007); 2(3):187-96

47.   Dissecting non-ribosomal and polyketide biosynthetic machineries using electrospray ionization Fourier-Transform mass spectrometry. Dorrestein PC, Kelleher NL, Nat Prod Rep (2006); 23(6):893-918

48.   Deconstruction of iterative multidomain polyketide synthase function. Crawford JM, Thomas PM, Scheerer JR, Vagstad AL, Kelleher NL, Townsend CA, Science (2008); 320(5873):243-6

49.   Polyunsaturated fatty-acid-like trans-enoyl reductases utilized in polyketide biosynthesis. Bumpus SB, Magarvey NA, Kelleher NL, Walsh CT, Calderone CT, J Am Chem Soc (2008); 130(35):11614-6

50.   Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. Dorrestein PC, Bumpus SB, Calderone CT, Garneau-Tsodikova S, Aron ZD, Straight PD, Kolter R, Walsh CT, Kelleher NL, Biochemistry (2006); 45(42):12756-66

51.   The loading module of mycosubtilin: an adenylation domain with fatty acid selectivity. Hansen DB, Bumpus SB, Aron ZD, Kelleher NL, Walsh CT, J Am Chem Soc (2007); 129(20):6366-7

52.   Harnessing the chemical activation inherent to carrier protein-bound thioesters for the characterization of lipopeptide fatty acid tailoring enzymes. Kopp F, Linne U, Oberthür M, Marahiel MA, J Am Chem Soc (2008); 130(8):2656-66

53.   Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units. Chan YA, Boyne MT 2nd, Podevels AM, Klimowicz AK, Handelsman J, Kelleher NL, Thomas MG, Proc Natl Acad Sci U S A (2006); 103(39):14349-54

54.   A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. Zhang J, Van Lanen SG, Ju J, Liu W, Dorrestein PC, Li W, Kelleher NL, Shen B, Proc Natl Acad Sci U S A (2008); 105(5):1460-5

55.   Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. Kelly WL, Boyne MT 2nd, Yeh E, Vosburg DA, Galonić DP, Kelleher NL, Walsh CT, Biochemistry (2007); 46(2):359-68

56.   Incorporation of nonmethyl branches by isoprenoid-like logic: multiple beta-alkylation events in the biosynthesis of myxovirescin A1. Calderone CT, Iwig DF, Dorrestein PC, Kelleher NL, Walsh CT, Chem Biol (2007); 14(7):835-46

57.   Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC, Proc Natl Acad Sci U S A (2006); 103(24):8977-82

58. A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. Calderone CT, Bumpus SB, Kelleher NL, Walsh CT, Magarvey NA, Proc Natl Acad Sci U S A (2008); 105(35):12809-14

59. High-resolution mass spectrometers. Marshall, AG; Hendrickson, CL, Annual Review of Analytical Chemistry (2008); 1:579-599

60. Fourier transform ion cyclotron resonance mass spectrometry: a primer. Marshall AG, Hendrickson CL, Jackson GS, Mass Spectrom Rev (1998); 17(1):1-35

61. Accurate mass measurements in proteomics. Liu T, Belov ME, Jaitly N, Qian WJ, Smith RD, Chem Rev (2007); 107(8):3621-53

62. Engineered biosynthesis of bacterial aromatic polyketides in Escherichia coli. Zhang W, Li Y, Tang Y, Proc Natl Acad Sci U S A (2008); 105(52):20683-8

63. Unit resolution mass spectra of 112 kDa molecules with 3 Da accuracy. Kelleher NL, Senko MW, Siegel MM, McLafferty FW, Journal of the American Society for Mass Spectrometry (1997); 8(4), 380-383

64. The construction of this hybrid 12T LTQ-FT-ICR-MS instrument has not been published.

65. Comparison of MS(2)-Only, MSA, and MS(2)/MS(3) Methodologies for Phosphopeptide Identification. Ulintz PJ, Yocum AK, Bodenmiller B, Aebersold R, Andrews PC, Nesvizhskii AI, J Proteome Res (2009); 8(2):887-899

66. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. Tholey A, Reed J, Lehmann WD, J Mass Spectrom (1999); 34(2):117-23

67. Phosphate group-driven fragmentation of multiply charged phosphopeptide anions. Improved recognition of peptides phosphorylated at serine, threonine, or tyrosine by negative ion electrospray tandem mass spectrometry. Edelson-Averbukh M, Pipkorn R, Lehmann WD, Anal Chem (2006); 78(4):1249-56

68. Mechanistic insights into the multistage gas-phase fragmentation behavior of phosphoserine- and phosphothreonine-containing peptides. Palumbo AM, Tepe JJ, Reid GE, J Proteome Res (2008); 7(2):771-9

69. Substrate specificity of the adenylation enzyme SgcC1 involved in the biosynthesis of the enediyne antitumor antibiotic C-1027. Van Lanen SG, Lin S, Dorrestein PC, Kelleher NL, Shen B, J Biol Chem (2006); 281(40):29633-40

70.   Principles of Fourier transform ion cyclotron resonance mass spectrometry and its application in structural biology. Barrow MP, Burkitt WI, Derrick PJ, Analyst (2005); 130(1):18-28

71.   Ion activation methods for tandem mass spectrometry. Sleno L, Volmer DA, J Mass Spectrom (2004); 39(10):1091-112

72.   Tandem mass spectrometry of large biomolecule ions by blackbody infrared radiative dissociation. Price WD, Schnier PD, Williams ER, Anal Chem (1996); 68: 859

73.   Sustained off-resonance irradiation for collision-activated dissociation involving Fourier transform mass spectrometry: Collision-activated dissociation technique that emulates infrared multiphoton dissociation. Gauthier JW, Trautman TR, Jacobson DB, Anal Chim Acta (1991); 246:211

74.   Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing. Little DP, Speir JP, Senko MW, O'Connor PB, McLafferty FW, Anal Chem (1994); 66(18):2809-15

75.   Identification of intact proteins in mixtures by alternated capillary liquid chromatography electrospray ionization and LC ESI infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry. Li W, Hendrickson CL, Emmett MR, Marshall AG, Anal Chem (1999); 71(19):4397-402

76.   Collisional activation of large multiply charged ions using Fourier transform mass spectrometry. Senko MW, Speir JP, McLafferty FW, Anal Chem (1994); 66(18):2801-8

77.   Thermally assisted infrared multiphoton photodissociation in a quadrupole ion trap. Payne AH, Glish GL, Anal Chem (2001); 73(15):3542-8

78.   Thermally assisted collision-induced dissociation in a quadrupole ion trap mass spectrometer. Racine AH, Payne AH, Remes PM, Glish GL, Anal Chem (2006); 78(13):4609-14

79.   Schwartz JC, Syka JE, Quarmby ST, The 53rd ASMS Conference on Mass Spectrometry and Allied Topics. (2005)

80.   Schlabach T, Zhang T, Miller K, Kiyonami R, The 2006 ABRF Conference (2006)

81.   Top-down mass spectrometry on low-resolution instruments: characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. Meluzzi D, Zheng WH, Hensler M, Nizet V, Dorrestein PC, Bioorg Med Chem Lett (2008); 18(10):3107-11

82.   Investigating nonribosomal peptide and polyketide biosynthesis by direct detection of intermediates on >70 kDa polypeptides by using Fourier-transform mass spectrometry. Hicks LM, Mazur MT, Miller LM, Dorrestein PC, Schnarr NA, Khosla C, Kelleher NL, Chembiochem (2006); 7(6):904-7

83.   A singular enzymatic megacomplex from Bacillus subtilis. Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R, Proc Natl Acad Sci U S A (2007); 104(1):305-10

84.   Electrospray ionization of a whole virus: analyzing mass, structure, and viability. Bothner B, Siuzdak G, Chembiochem (2004); 5(3):258-60

85.   Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. McKay AR, Ruotolo BT, Ilag LL, Robinson CV, J Am Chem Soc (2006); 128(35):11433-42

86.   10-formyltetrahydrofolate dehydrogenase requires a 4'-phosphopantetheine prosthetic group for catalysis. Donato H, Krupenko NI, Tsybovsky Y, Krupenko SA, J Biol Chem (2007); 282(47):34159-66

87.   Genome-wide high-throughput mining of natural-product biosynthetic gene clusters by phage display. Yin J, Straight PD, Hrvatin S, Dorrestein PC, Bumpus SB, Jao C, Kelleher NL, Kolter R, Walsh CT, Chem Biol (2007); 14(3):303-12

88.   Intramolecular catalysis of thiol ester hydrolysis by a tertiary amine and a carboxylate. Brown RS, Aman A, J Org Chem (1997); 62, 4816-4820

89.   Phosphorylation-specific MS/MS scoring for rapid and accurate phosphoproteome analysis. Payne SH, Yau M, Smolka MB, Tanner S, Zhou H, Bafna V, J Proteome Res (2008); 7(8):3373-81

90.   Natural products of filamentous fungi: enzymes, genes, and their regulation. Hoffmeister D, Keller NP, Nat Prod Rep (2007); 24(2):393-416

91.   Advances in polyketide synthase structure and function. Van Lanen SG, Shen B, Curr Opin Drug Discov Devel (2008); 11(2):186-95

92.   The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J, Proc Natl Acad Sci U S A (2007); 104(5):1506-9.

93.   PksS from Bacillus subtilis is a cytochrome P450 involved in bacillaene metabolism. Reddick JJ, Antolak SA, Raner GM, Biochem Biophys Res Commun (2007); 358(1):363-7

94.   Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. Verdier-Pinard P, Lai JY, Yoo HD, Yu J, Marquez B, Nagle DG, Nambu M, White JD, Falck JR, Gerwick WH, Day BW, Hamel E, Mol Pharmacol (1998); 53(1):62-76

95.   The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. Holden HM, Benning MM, Haller T, Gerlt JA, Acc Chem Res (2001); 34(2):145-57

96.   Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair. Gu L, Jia J, Liu H, Håkansson K, Gerwick WH, Sherman DH, J Am Chem Soc (2006); 128(28):9014-5

97.   Crystal structure of the ECH2 catalytic domain of CurF from Lyngbya majuscula. Insights into a decarboxylase involved in polyketide chain beta-branching. Geders TW, Gu L, Mowers JC, Liu H, Gerwick WH, Håkansson K, Sherman DH, Smith JL, J Biol Chem (2007); 282(49):35954-63

98.   The enediyne antibiotics. Smith AL, Nicolaou KC, J Med Chem (1996); 39 (11):2103-2117

99.   Towards bioengineering anticancer drugs. Kerwin S, Chemistry & Biology (2002); 9:956-958

100.   SMANCS and polymer-conjugated macromolecular drugs: Advantages in cancer chemotherapy. Maeda H, Adv Drug Deliv Rev (2001); 46:169–185.

101.   Mylotarg: Antibody-targeted chemotherapy comes of age. Sievers EL, Linenberger M, Curr Opin Oncol (2001); 13:522–527.

102.   The calichemicin gene cluster and its interative type I enediyne PKS. Ahlert J, Shepard E, Lomovskaya N, Zazopoulos E, Staffa A, Bachmann BO, Huang K, Fonstein L, Czisny A, Whitwam RE, Farnet CM, Thorson JS, Science (2002); 297(5584):1173-1176

103.   Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. Van Lanen SG, Oh TJ, Liu W, Wendt-Pienkowski E., and Shen B, J Am Chem Soc (2007); 129:13082–13094

104.   Iterative type I polyketide synthases for enediyne core biosynthesis.   Horsman GP, Van Lanen SG, Shen B, Methods in Enzymology (2009); Volume 459: 98-112

105.   Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. Metz JM, Science (2001) 293:290-293

106.   A novel function of yeast fatty acid synthase—Subunit alpha is capable of self-pantetheinylation. Fichtlscherer F, Wellein C, Mittag M, Schweizer E, Eur J Biochem (2000); 267:2666–2671.

107.   C1027 chromophore, a potent new enediyne antitumor antibiotic, induces sequence-specific double-strand DNA cleavage. Xu, YJ, Zhen YS, Goldberg IH, Biochemistry (1994); 33:5947-5954

108.   Reactive 1,4-dehydroaromatics. Bergman RG, Acc Chem Res (1973); 6:25-31.

109.   C-1027, A radiomimetic enediyne anticancer drug, preferentially targets hypoxic cells. Beerman TA, Gawron LS, Shin S, Shen B, McHugh MM, Cancer Res (2009); 69(2):593-598.

110.   The benzoxazolinate of C-1027 confers intercalative DNA binding. Yu, L, Mah S, Otani T, Dedon P, J Am Chem Soc (1995); 117:8877-8878

111.   Antitumor enediyne chromoprotein C-1027: Mechanistic investigation of the chromophore-mediated self-decomposition pathway. Inoue M, Usuki T, Lee N, Hirama M, Tanaka T, Hosoi F, Ohie S, Otani T, J Am Chem Soc (2006); 128(24):7896-7903

112.   Single Chemical Modifications of the C-1027 Enediyne Core, a Radiomimetic Antitumor Drug, Affect Both Drug Potency and the Role of Ataxia-Telangiectasia Mutated in Cellular Responses to DNA Double-Strand Breaks. Kennedy DR, Gawron LS, Ju J, Liu W, Shen B, BeermanTA Cancer Res (2007); 67 (2): 773-781

113.   Biosynthesis of the β-Amino Acid Moiety of the Enediyne Antitumor Antibiotic C-1027 Featuring β-Amino Acyl-S-carrier Protein Intermediates Van Lanen SG, Dorrestein PC, Christenson SD, Liu W, Ju J, Kelleher NL, Shen B, J. Am. Chem. Soc. (2005); 127(33): 11594-11595

114.   Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism. Van Lanen SG, Lin S, and Shen B, Proc Nat Acad Sci (2008); 105(2):494-499

115.   Cloning of an avilamycin biosynthetic gene cluster from Streptomyces viridochromogenes Tu57. Gaisser S, Trefzer A, Stockert S, Kirschning A, Bechthold A, J Bacteriol (1997); 179:6271-6278.

116.   Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Shen B, Curr Opinion in Chem Bio (2003); 7:285–295

117.   Molecular and supramolecular chemistry of natural products and their model compounds. Fuhrop JH, Endisch C, (2000) CRC Press, New York:13-17

118.   Genes for production of the enediyne antitumor antibiotic C-1027 in Streptomyces globisporus are clustered with the cagA gene that encodes the C-1027 apoprotein. Liu W, Shen B, Antimicrob Agents Chemother (2000); 44:382-392.

119.   Best-selling human medicines 2002-2004. Maggon, K, Drug Disc Today (2005); 10(11):739-742

120.   IMS National Prescription Audit (2008). IMS Health (http://www.imshealth.com)

121.   Lovastatin: a new cholesterol-lowering agent. McKenny JM, Clinical Pharmacology (1988); 7(1):21-36

122.   3-Hydroxy-3-methylglutaryl-CoA reductase inhibitors. Endo A, Methods Enzymol (1981); 72:684-9.

123.   Structural mechanism for statin inhibition of HMG-CoA reductase. Istvan ES, Deisenhofer J, Science (2001); 292:1160–64.

124.   Biosynthesis and biotechnological production of statins by filamentous fungi and application of these cholesterol-lowering drugs. Manzoni M, Rollini M, Appl Microbiol Biotechnol (2002); 58: 555–564.

125.   Mevinolin: a highly potent competetive inhibitor of hydroxymethyl-glutaryl-coenzyme A reductase and a cholesterol-lowering agent. Alberts AW, Chen J, Kuron G, Hunt V, Huff J, Hoffman C, Rothrock J, Lopez M, Joshua H, Harris E, Patchett A, Monaghan R, Currie S, Stapley E, Albers-Schonberg G, Hensens O, Hirshfield J, Hoogsteen K, Liesch J, Springer, J Proc Nat Acad Sci (1980); 77: 3957-3961.

126.   Competitive inhibition of 3-hydroxy-3-methylglutaryl coenzyme A reductase by ML-236A and ML-236B fungal metabolites, having hypocholesterolemic activity. Endo A, Kuroda M, Tanzawa K, FEBS Lett (1976); 72(2):323-6.

127.   Monacolin K, a new hypocholesterolemic agent that specifically inhibits 3-hydroxy-3-methylglutaryl coenzyme A reductase. Endo A, J Antibiot (Tokyo). (1980); 33(3):334-6.

128.   3-Hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. Four side chain ester derivatives of mevinolin. Hoffman WF, Alberts AW, Anderson PS, Chen JS, Smith RL, Willard AK, J Med Chem (1986); 29:849–852.

129.   Lovastatin biosynthesis in Aspergillus terreus: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene. Hendrickson L, Davis CR, Roach C, Nguyen DK, Aldrich T, McAda PC, Reeves CD, Chem Biol (1999); 6: 429–439

130.   Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. Kennedy J, Auclair K, Kendrew SG, Park C, Vederas JC, Hutchinson CR. Science (1999); 284:1368–1372

131.   Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB. Ma S, Tang Y, FEBS Journal (2007); 274:2854–2864

132.   Polyketide biosynthesis: a millennium review. Staunton J, Weissman KJ, Nat Prod Rep (2001); 18:380–416

133.   Acyltransferase mediated polyketide release from a fungal megasynthase. Xie X, Meehan MJ, Xu W, Dorrestein PC, Tang Y   J Am Chem Soc (*accepted for publication 2009*)

134.   Chain initiation in the leinamycin-producing hybrid nonribosomal peptide/polyketide synthetase from Streptomyces atroolivaceus S-140. Discrete,

monofunctional adenylation enzyme and peptidyl carrier protein that directly load D-alanine. Tang GL, Cheng YQ, Shen B, J Biol Chem (2007); 282(28):20273-82

135.   Enzymatic Synthesis of Aromatic Polyketides Using PKS4 from Gibberella fujikuroi. Ma SM, Zhan J, Watanabe K, Xie X, Zhang W, Wang CC, Tang Y, J Am Chem Soc (2007); 129(35):10642-10643

136.   Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. Crawford, Dancy BCR, Hill EA, Udwary DW, Townsend CA, Proc Natl Acad Sci (2006); 103:16728-33

137.   Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB. Ma SM, Tang Y, Febs J (2007); 274:2854-2864

138.   Spontaneous α-N-6-Phosphogluconoylation of a "His Tag" in Escherichia coli: The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. Geoghegan KF, Dixon HBF, Rosner PJ, Hoth LR, Lanzetti AJ, Borzilleri KA, Marr ES, Pezzullo LH, Martin LB, LeMotte PK, McColla AS, Kamatha AV, Stroha JG, Anal Biochem (1999); 267(1):169-184