Quantitative Bottom-Up and Top-Down LC/FT-ICR MS for Multisite Phosphoproteins and Natural Product Biosynthetic Pathways

by

Wendi Anne Hale

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2015

Doctoral Committee:

Professor Kristina I. Håkansson, Chair
Professor Robert T. Kennedy
Associate Professor Nina Lin
Assistant Professor Brandon T. Ruotolo

To my mom, for her endless support and sacrifice

Acknowledgements

There are so many people to thank after about five years of working toward my thesis. I would first like to wholeheartedly thank my advisor, Professor Kristina Håkansson. Kicki has been incredibly supportive throughout each stage of my PhD. She is a brilliant scientist, an enthusiastic mentor and I have never met a more patient person. Her flexibility and positive attitude have helped me to become a more confident and competent scientist, allowing me to venture out and experience new places and ideas, whether it be traveling for conferences or working with a variety of collaborators.

I would like to thank the rest of my committee next for their invaluable advice and suggestions throughout my studies. Professor Brandon Ruotolo, thank you for the opportunity to rotate in your lab, for insightful advice on my projects, and also valuable advice for postdoctoral studies abroad. Professor Robert Kennedy, thank you for your helpful advice throughout my committee meetings and for being an excellent separations teacher; I frequently look back at my notes and thoroughly enjoyed the way the course was taught. Professor Nina Lin, thank you for being a wonderful committee member and collaborator. Thank you for also welcoming me into and allowing me to have the freedom to work in your lab. Collaborative meetings with you were also extremely constructive and informative.

My thesis is full of collaborations, where I have had the opportunity to expand my knowledge of phosphorylation and natural products. All of my collaborators have been patient, kind, and wonderful to work with. I would like to thank Professor Nina Lin, Dr. Marjan Varedi, Lian Zhu, and all of Prof. Lin's students who have answered questions while I have worked in their lab, especially Scott Scholz and Tatyana Saleski. I would also like to thank Professor David

had confidence in me and have encouraged me to pursue my dreams and to be a strong and independent yet compassionate person. I am so very proud to be your daughter!

Wendi Hale

April 3, 2015

Ann Arbor, MI

Table of Contents

List of Tables

List of Figures

expected heights of a 1:1 ratio of a mixture of the β-ketohexaketide and β-hydroxyhexaketide. The latter squares indicate that a second reduction may have occurred due to excess NADPH. 81

List of Abbreviations

Abbreviations

ACP	Acyl carrier protein
AMT tag	Accurate mass and time tag
AT	Acyl transferase
CID	Collision induced dissociation
Cryo-EM	Cryo-electron microscopy
Cur	Curacin
DDA	Data dependent acquisition
DEBS	6-deoxyerythronolide B synthase
DH	Dehydratase
DIA	Data independent acquisition
ECD	Electron capture dissociation
ECH_1	Dehydratase
ECH_2	Decarboxylase
EDD	Electron detachment dissociation
EIC	Extracted ion chromatogram
EID	Electron induced dissociation
ER	Enoyl reductase
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FDR	False discovery rate
FID	Free induction decay
FT-ICR	Fourier Transform Ion Cyclotron Resonance
Hal	Halogenase
HDX	Hydrogen deuterium exchange
HPLC	High performance liquid chromatography
IMAC	Immobilized metal affinity chromatography
IPSO	Intelligent Pulse Sequence Organizer

IRMPD	Infrared multiphoton dissociation
Jam	Jamaicamide
KR	Ketoreductase
KS	Ketosynthase
MDR	Medium chain dehydrogenase/reductase
mFAS	Mammalian fatty acid synthase
MM	Methyl-malonyl
MOAC	Metal oxide affinity chromatography
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
nETD	Negative electron transfer dissociation
niECD	Negative ion electron capture dissociation
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthetase
OASIS	Orthogonal active site identification system
PAcIFIC	Precursor acquisition independent from ion count
PCP	Peptide carrier protein
PKS	Polyketide synthase
Ppant	Phosphopantetheine
PRISM	Proteomic Investigation of Secondary Metabolism
PTM	Post-translational modification
RF	Radiofrequency
SILAC	Stable isotope labeling by amino acids in cell culture
TE	Thioesterase
UVPD	Ultraviolet photo dissociation

Abstract

Quantitative Bottom-Up and Top-Down LC/FT-ICR MS for Multisite

Phosphoproteins and Natural Product Biosynthetic Pathways

by

Wendi Anne Hale

Chair: Kristina I. Håkansson

Bottom-up (involving proteolysis) and top-down (intact proteins) mass spectrometry (MS)-based proteomics are increasingly important tools for protein identification and quantitation from complex mixtures. A multitude of quantitation techniques are available, each with advantages and disadvantages, from label to label-free and relative to absolute quantitation. The ultra-high resolution and mass accuracy of Fourier transform ion cyclotron resonance (FT-ICR) MS, can improve current quantitation methods. This thesis presents FT-ICR MS-based quantification of proteins and peptides with dynamic post-translational modifications, and a novel data-independent liquid chromatography (LC)/FT-ICR MS approach for targeted discovery of novel natural product biosynthetic pathways.

A novel negative ion mode quantitation method was developed involving standard peptides with Val to Leu or Leu to Val substitutions, resulting in similar ionization efficiencies. This method was applied to Sic1, a yeast protein with a switch-like response when six out of

nine phosphorylation sites are phosphorylated. With this quantitation technique, the phosphorylation kinetics of all nine sites were determined.

Relative FT-ICR MS-based quantitation was applied to marine natural product biosynthetic enzymes. Ketosynthase, acyl transferase, and acyl carrier protein (ACP) active site occupancies were identified and quantified via LC/FT-ICR MS, as PikAIII, module 5 in the pikromycin biosynthetic pathway, completed its catalytic cycle with the addition of pentaketide substrate, methyl-malonyl extender unit, and NADPH. Several protocol changes were made to observe the pentaketide bound to the ketosynthase active site for the first time. The largely homologous jamaicamide (Jam) and curacin (Cur) biosynthetic pathways diverge at an enoyl reductase (ER) domain. Catalytic efficiencies of wildtype and mutant ER domains were measured by quantifying ACP-bound product via top-down FT-ICR MS and collision-induced dissociation. Cyclopropanase activity of a novel chimeric ER, designed by inserting a Cur ER loop into Jam ER was measured via FT-ICR MS.

Finally, a novel data-independent acquisition method was developed for targeted discovery of novel natural product biosynthetic pathways from unsequenced organisms. This technique utilizes the resonant IR absorptivity of a secondary phosphate in the ACP phosphopantetheine modification. Peptides with this modification dissociate with high efficiency upon IR irradiation, whereas unmodified peptides are largely unaffected by IR irradiation.

Chapter 1

Introduction

1.1 Mass Spectrometry for Proteomics

The field of proteomics, the study of the proteome, has exploded in the last few decades. The proteome encompasses all proteins expressed in complex tissues, cells, or whole organisms[1]. Proteomics aims to investigate the function of all proteins, including their structure, interactions with other proteins, and composition[3,4]. Proteomics is affected by many challenges due to differences in expression level, hundreds of post-translational modifications (PTMs), dynamics in interactions, and sample preparation for analysis[3,4]. Many biochemical and biophysical techniques are available to study proteins, including surface plasmon resonance, circular dichroism, fluorescence, and Raman spectroscopy, mass spectrometry (MS), nuclear magnetic resonance spectroscopy, X-ray crystallography, and electron microscopy[5,6]. While each technique has its own set of advantages and disadvantages, mass spectrometry is an ideal tool for proteomics[7]: MS offers high sensitivity for complex and low abundance proteins, including flow rates at picoliters per minute, consuming only zeptomoles of peptide[8] while also being able to detect, quantify, and localize PTMs[9]. In addition to high sensitivity, high mass accuracy and high-throughput are other requirements of proteomics, both achieved by mass spectrometers[10,11]. Mass accuracy can easily reach less than 5 ppm on high performance mass spectrometers such as Orbitraps and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers[12-14].

There are three complementary approaches to performing proteomics by mass spectrometry: bottom-up, top-down, and the more recent middle-down approach.
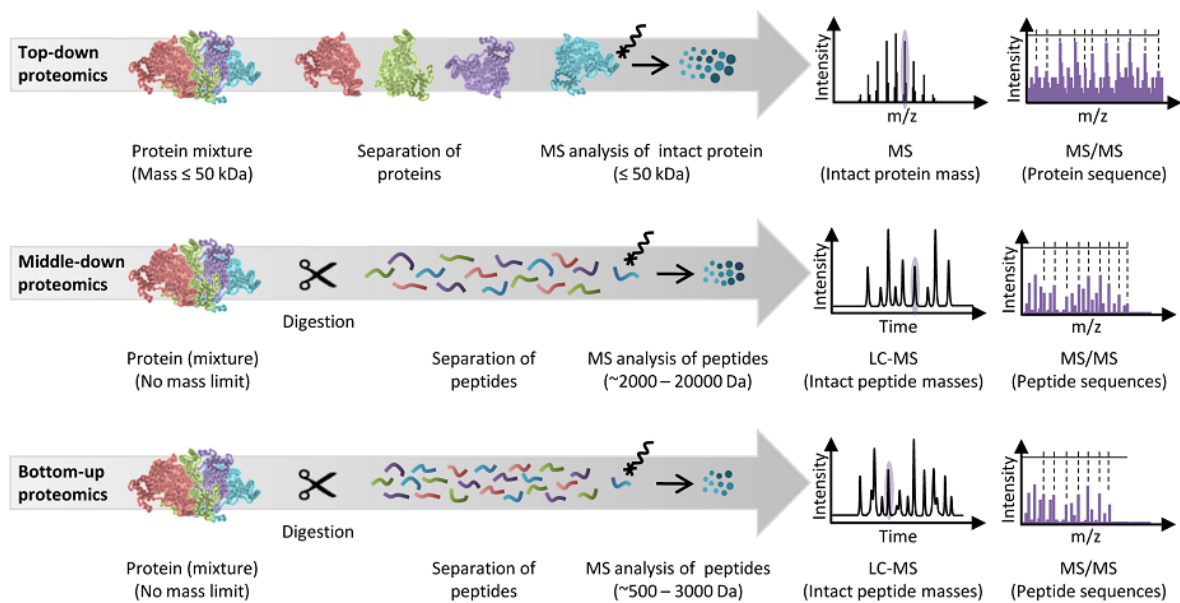
Figure 1.1. Schematic of top-down, middle-down, and bottom-up mass spectrometry for deducing protein primary sequence information. Reprinted with permission from Switzar et al. Copyright 2013 American Chemical Society[15].

Bottom-up mass spectrometry is arguably the most widely utilized approach in proteomics and is ideal for protein identification while also providing localization of PTMs[16]. One or more proteins are subjected to proteolysis by one or more proteases[15,17]. The resulting peptides are usually separated by reverse phase high performance liquid chromatography (HPLC) and then are analyzed by mass spectrometry[18-21]. Multidimensional protein identification technology (MudPIT) directly combines strong cation exchange with reverse phase HPLC[22]. Samples can also be pre-fractionated followed by HPLC/MS analysis of the resulting fractions. Peptides typically undergo fragmentation via tandem mass spectrometry (MS/MS) and the resulting fragment masses are compared to a genome-derived database using algorithms such as MASCOT or Sequest[7,16,23,24]. Although collision induced dissociation (CID)[25,26] is the most common MS/MS activation technique, other methods are becoming increasingly popular, especially for their ability to retain PTMs. Examples of alternative MS/MS techniques include infrared multiphoton dissociation (IRMPD)[27], electron capture dissociation (ECD)[28], electron transfer dissociation (ETD)[29], negative electron transfer dissociation (nETD)[30], negative ion electron capture dissociation (niECD)[31], electron detachment dissociation (EDD)[32], electron induced dissociation (EID)[33-35], and ultraviolet photo dissociation (UVPD)[36,37]. In bottom-up

proteomics, no prior knowledge of a protein's sequence is required for identification as long as it is present in the utilized database[23]. In addition, the fragmentation pathways/bond cleavage preferences of the utilized MS/MS technique must be known. Although bottom-up mass spectrometry is a sensitive technique that works well for protein identification, it is challenging to achieve 100% sequence coverage, due to ion suppression, sample loss during separation, or when the same or similar peptide sequence occurs multiple times in a single protein or occurs in multiple different proteins[19].

Top-down proteomics measures the intact mass of one or more proteins and subsequently dissociates the protein or proteins inside the vacuum of the mass spectrometer[19,38]. Top-down analysis is best applied to protein identification and structural characterization, including identification and localization of PTMs[39]. Combinatorial PTMs, where multiple PTMs affect protein function, e.g., binary switches in histones and phosphorylation, can be more fully understood and characterized if they are located on the same polypeptide chain[40]. Top-down MS analysis is emerging as a technique in structural biology[41]. Protein-protein, protein-peptide, and protein-ligand binding sites and interactions can be determined with, e.g., ECD, as it can cleave the backbone of the protein or peptide without disrupting non-covalent interactions[42-45]. The advantages of top-down proteomics are numerous and counteract the disadvantages of bottom-up proteomics. For example, bottom-up proteomics rarely provides 100% sequence coverage, while this is possible and fairly common in top-down proteomics[19]. As aforementioned, ECD and ETD provide complementary sequence information to CID and they are more efficient with increasing charge, which is readily available on intact proteins. Top-down analysis sample preparation time is shortened without proteolysis[41] although it also comes with its own challenges and shortcomings. Intact proteins have a higher detection limit and have lower sensitivity than peptides due to lower ionization efficiency and dilution of signal over many charge states upon electrospray ionization (ESI), thus proteins must have a high purity for top-down MS to work well[38]. Fragmenting an intact protein results in complex MS/MS spectra and thus high resolution is typically required. It has been historically challenging to detect intact proteins without a time-of-flight mass analyzer, which may lack the resolution required to resolve the protein and high mass

fragments[38]. However, recent advancements have allowed large proteins, above 100 kDa to be detected by other mass spectrometers, including Orbitraps and FT-ICRs[46-48].

Middle-down mass spectrometry is an emerging technique that combines the advantages of bottom-up and top-down mass spectrometry[15]. Middle-down MS analyzes large peptides that range in size from 2 to 20 kDa, whereas bottom-up peptides are typically less than 3 kDa[15]. These large peptides can be produced by incomplete digestion with traditional bottom-up proteases such as Glu-C, or proteases that cleave in few specific areas, such as outer membrane protease T (OmpT) or immunoglobulin G-degrading enzyme from *S. pyogenes* (IdeS)[49-51]. Although the concept of middle-down MS for proteomics is somewhat new, pharmaceutical companies have been performing middle-down analysis in a targeted manner by digesting antibodies with endoproteinase Lys-C to create large peptides for disulfide mapping for years[52,53]. Middle-down MS is also beneficial for untargeted proteomics, as it is able to distinguish between protein isoforms and can detect and identify peptides with modifications, particularly combinatorial PTMs, traditional benefits of top-down MS[50,54,55] while avoiding the ionization efficiency issue associated with top down analysis. Middle-down proteomics has advanced recently, for example by the introduction of OmpT, a protease developed by the Kelleher group to create peptides greater than 2 kDa. Larger peptides can be advantageous in proteomics experiments because electron based methods, such as ECD and ETD, are complementary to CID and are more efficient when the peptide has more charges[50,55]. Larger peptides are able to provide information on combinatorial PTMs, e.g., in histones that have a high number of isoforms with high sequence homology, traditionally an advantage of top-down proteomics[50,55,56]. As the middle-down MS approach is combining the positive attributes of bottom-up and top-down protein MS, it appears that there are few disadvantages to middle-down analysis. However, it is still an emerging technique.

1.2 Protein MS Quantitation Methods

Protein PTMs are highly dynamic and thus change abundance readily, particularly if they are involved in cell regulation. Quantitation is important for determining both the amount of a specific protein in a system and the fold change in response to either diseases or drugs[24,57].

Mass spectrometry-based quantitation is more universal than other biochemical quantitation methods, such as immunoassays, as they may incur non-specific protein binding, and can detect small changes to the protein such as modifications[58]. There are multiple ways to categorize MS-based quantitation methods. The main category is label versus label-free quantitation and, within those two groups, there is global versus targeted and relative versus absolute quantitation.

1.2.1 Protein quantitation involving labeling methods

One of the most popular labeling methods is stable isotope labeling by amino acids in cell culture (SILAC). Since its inception in 2002, over five hundred papers employing or reviewing SILAC have been published[59]. Although it is a relative quantitation method, it is arguably the most accurate technique and can be applied to global proteomics. Amino acids are added into two sets of cell culture; one set has no label, while the other set of amino acids is labeled with heavy stable isotopes[60]. Leucine labels are common, as they are the most common amino acid and label about seventy percent of tryptic peptides, or lysine and arginine, which guarantee at least one label per tryptic peptide[24]. After performing protein expression, the labeled and unlabeled proteins from the two cell cultures are mixed in a 1:1 ratio, digested and analyzed by MS[60]. Labeled and unlabeled peptides elute at the same time during chromatography but differ by a certain mass, depending on the label[57]. Although SILAC requires living cells and can be quite costly, it functions well in mammalian, yeast, and bacterial cell cultures and has low variance because the labels are introduced at the organism level in model organisms[57,60,61]. The versatility of SILAC allows for many applications, including protein turnover, monitoring PTM dynamics such as phosphorylation and histone modifications[62-66]. SILAC's flexibility allows for variations, such as neutron encoding (NeuCode) SILAC. This technique requires ultra-high resolution as it exploits the millidalton differences in different combinations of SILAC isotopologues[67]. NeuCode increases the dynamic range of traditional SILAC without complicating the mass spectrum, as the mass differences are too small to observe with low resolution[67]. Comparable to SILAC, ^{15}N metabolic labeling for plant proteomics works by growing half of the plant cells in medium enriched with over ninety six

percent ^{15}N NH$_4$Cl[68]. An alternative method is feeding the plant with an inorganic nitrogen source enriched in ^{15}N, leading to ^{15}N incorporation into the amino acids and therefore proteins. The advantages of this approach are also comparable to SILAC: it is an accurate technique, applicable to global proteomics, can determine changes in protein expression, and can detect changes in PTM amounts[57,68]. Like SILAC, ^{15}N labeling can be applied to protein turnover calculations, and can interrogate plant metabolism and quantitatively compare multi-protein plant complexes in different strains[69-71]. There are some limitations to this technique; due to the large number of nitrogen atoms in any peptide sequence, it is possible to partially label nitrogens, which can lead to ambiguous data interpretation[24]. Further, the number of nitrogen atoms is dependent on the peptide sequence length, so each labeled peptide has a varying mass difference[24].

Chemical, or *in vitro*, labeling is another common technique for protein and peptide quantitation. Isotope-coded affinity tags (ICATs) have specific reactivity with thiol groups and thus covalently attach to cysteine sidechains. A linker group incorporates stable isotopes for light or heavy labeling with hydrogens or deuteriums, and a biotin tag is included for affinity purification[72]. Two sets of protein samples are labeled with ICAT (light vs. heavy), digested, enter an avidin column, and are then subjected to LC/MS analysis[24]. The resulting light and heavy pairs of peptides have a fixed 8 Da difference, making data analysis simpler than metabolic labeling[72]. Despite the improvements in simplicity, ICAT is only applicable to peptides with cysteine residues, only including about twenty percent of peptides[57]. The ICAT tags are fairly large, which influence MS/MS fragmentation patterns and the deuterium labeling affects coelution of the paired peptides[24]. The latter issue has been solved by creating a cleavable tag that allows coelution[73,74]. Peptide dimethyl labeling is achieved by a reductive amination reaction with formaldehyde and deuterated formaldehyde, respectively, resulting in the dimethyl label on the N-terminus and lysine sidechains of the peptides[75]. The mass difference between labeled and unlabeled peptides is 28 Da, the labeling reaction is inexpensive and takes less than five minutes[75]. The hydrogen and deuterium labeled peptides have similar retention times, however they are different enough such that data analysis is complicated, as peptides are not in the same spectrum, rendering relative quantitation more

difficult[75,76]. Absolute quantification using synthetic standard peptides (AQUA) is one of the few absolute quantitation strategies. Synthesized isotopically labeled peptide standards at a known concentration with identical sequences to the peptides in question function as internal standards[77]. Although this approach has been in use for non-proteomic applications for a few decades prior to the introduction of AQUA, the technique did not become popular until many other peptide quantitation techniques were developed[78]. The concentration of the peptide of interest can be calculated based on the internal standard peptide. Because the peptides are chemically identical, they will have the same retention time[77]. To reduce signal interference from other peptides, selected or multiple reaction monitoring can be performed on the peptide fragments for quantitation[24]. This universal technique can be applied to any peptide, because the isotopic label can be added to any amino acid. AQUA has the potential to give insight into dynamic changes in protein and peptide levels, even if they are modified and of low abundance[77]. Although this technique is robust and very specific, sample preparation may affect the true concentration of the peptide in question[57,76].

Isobaric tags for relative and absolute quantitation (iTRAQ) is another labeling technique that adds isobaric mass labels at the N terminus and lysine side chains of peptides[79]. The labels do not increase spectral complexity, as each labeled peptide is isobaric, however upon MS/MS, reporter ions of different mass are among the product ions and are the peaks used for identification and quantitation[79]. The reporter ions range between 114.1-117.1 Da and the mass balance group ranges between 28 and 31, so that the mass spectra all contain the same added mass, 145.1 Da[79]. The main advantage of this technique is that quantitation for more than two peptides is possible; originally iTRAQ was designed for multiplexing with 4 peptides and is now capable of 8-plexing[79-81]. iTRAQ is applicable for absolute quantitation which can be obtained by adding an iTRAQ tag to a synthetic standard peptide[79]. Although iTRAQ is an overall robust technique, it suffers from isotopic interference and low signal-to-noise, affecting quantitation[82]. However, when a fragment ion is further isolated and fragmented in an MS[3] experiment, isotopic interference is greatly reduced due to the secondary isolation[83]. Tandem mass tags (TMTs) are similar to iTRAQ in that they are also isobaric tags that are revealed upon MS/MS[84]. The tags consist of two or three parts: an amino acid tag group linked to a guanidino

functionality for sensitization and a mass normalization amino acid, while a second generation of tags includes a proline for cleavage enhancement[84]. The tag attaches to the N-terminus or a lysine sidechain of the peptide. Like iTRAQ, tagged peptides co-elute, which simplifies quantitation, allowing for comparison of relative ion abundances in a single spectrum. Overall, both isobaric labeling techniques are less accurate than SILAC, because their accuracy depends on the isolation width of the precursor ions for fragmentation, where all isolated ions will contribute to fragments in the reporter ion mass ranges[57]. iTRAQ is now a trademark by AB Sciex and is sold by AB Sciex and Sigma Aldrich, while TMTs are sold by Thermo Scientific.

1.2.2 Label-free quantitation

Label-free quantitation techniques are emerging as increasingly popular and simpler quantitation approaches, as they do not involve extra chemical reactions, are generally less expensive, have high proteome coverage and can be applied to a large variety of biological samples[85]. There are three main label-free strategies; one measures the abundance of precursor ion peaks, one measures the number of fragment ion peaks, while the last one makes use of indices[76].

Spectral counting is a relative quantitation technique that works by counting the number of MS/MS peaks from a protein, which directly corresponds with protein abundance[61,86]. Proteins are identified via a database based on the MS/MS spectra; next, all of the MS/MS peaks from a single protein are added together based on their peak area, then this peak area is normalized to the peak area of an internal standard protein digest at a specified concentration[87]. An advantage of spectral counting is that the MS/MS data not only provide quantitative information, but also provide a plethora of identification information[76]. Spectral counting is very accurate for measuring large changes in protein abundance but is less accurate with smaller changes[61]. Another disadvantage is that dynamic exclusion during online LC/MS/MS can affect the quantitative accuracy rate, as the basis of spectral counting is that the abundance of a protein correlates with the amount of times a corresponding peptide is isolated and fragmented [76]. Further, LC/MS/MS runs are repeated multiple times for quantitation and there is a possibility that the same peptides are selected in each run, thus biasing the data[24].

Generally, signal abundance from electrospray ionization directly correlates with the concentration of the ion[88]. Extracted ion chromatogram (XIC)-based quantitation involves creation of extracted ion chromatograms for the same peptide across different LC/MS runs[24]. For this technique to successfully relatively quantify a peptide in question, the sample handling must be the same between both runs to reproduce spectral total ion abundance, background peptides and contaminants, thus ensuring similar peptide ionization efficiency[24,57]. Quantitative accuracy depends on signal to noise ratio: if the peptide abundance is near the noise, the accuracy will decrease[24]. There is always some variability between two LC/MS runs, even if they are performed under identical conditions, therefore some sort of normalization is necessary, such as a background peptide that will not change abundance[24,86]. This technique in particular has been applied to determine phosphorylation stoichiometry[89]. XIC-based quantitation works best with high mass accuracy mass spectrometers, so that peptides with similar, but not identical, masses are not chosen for quantitation[76]. Overall, XIC-based methods can suffer from variability due to non-identical conditions between LC/MS runs. In addition, HPLC retention times must be highly reproducible.

Absolute quantitation with label-free techniques makes use of derived indices, such as protein abundance indices, and therefore do not require internal standards. The exponentially modified protein abundance index (emPAI) is an index that shows a logarithmic relationship between the number of peptides observed and the protein amount in a certain sample[90]. emPAI was first demonstrated with 46 proteins in a whole cell lysate from a mouse[90]. An emPAI is calculated by the number of observed spectra for each protein divided by the number of observable peptides[57]. The index is normalized and takes into account that larger proteins will create more peptides than a smaller protein[90]. emPAI is easily calculated from values provided by databases such as Mascot and does not need additional experimentation[86,90]; however, emPAI is less accurate with low resolution mass spectrometers and may become saturated with highly abundant proteins, thus decreasing the sensitivity of the technique[90]. Additionally, emPAI will not produce accurate results if peptides overlap, which will occur more often with larger proteins, presenting a size limitation to this technique[91]. Absolute protein expression measurements (APEX) is another index-type quantitation method that measures the

absolute protein concentration per cell based on the proportion between the protein abundance and the number of peptides observed[86,92]. There are a number of correction factors in APEX that create the proportionality between the observed and expected number of peptides found[86]. The APEX score provides the protein abundance based on the fraction of observed peptide mass spectra associated with a protein, and is corrected by an estimate of the number of unique peptides expected, calculated from online databases[86]. APEX is able to determine the absolute protein expression within a correct order of magnitude and correlates well with other measurements such as 2D gel electrophoresis[76,92]. Issues and biases in bottom-up proteomics such as ionization and solubilization are already corrected for in APEX[92]. APEX may be applied to biomarker discovery or serum profiling, and can also estimate protein degradation rates[92]. The derived indices also suffer from caveats; because they calculate scores from online databases, these methods are limited to positive ion mode.

1.3 Liquid Chromatography and Mass Spectrometry Instrumentation

1.3.1 Liquid chromatography

High performance liquid chromatography is an analytical chemistry technique that separates a wide variety of components. When coupled to a mass spectrometer, it is often used for separation of complex mixtures or desalting and cleanup prior to analysis. The first example of chromatography was shown by Mikhail Tswett in 1903, while the first HPLC was designed by Csaba Horvath in 1965 to separate nanomolar mixtures of small organic molecules, however, the application of LC has greatly expanded and is now essential to the proteomics field[93].

The LC consists of multiple parts, starting with the solvents flowing through a degasser. Each solvent is pushed through its own pump and solvents are then mixed together. The mixed solvents flow through a cooled autosampler, where the sample of interest is stored and injected onto an LC column. The LC column consists of a stationary phase and a mobile phase, typically encased in stainless steel for high flow columns (~mL/min). The most common stationary phase consists of silica particles, although alumina is another potential material[94].

The bonded stationary phase is covalently attached to the silica and interacts with the analytes[94]. For reverse phase chromatography, the bonded stationary phase is non-polar and either has an octadecyl (C18), octyl (C8), or phenyl (C4) group, while the mobile phase solvent is more polar[94]. Analytes are detected as they are eluted from the column. Although, there are a number of detection methods, the HPLC in the Hakansson lab has a UV detector, normally set to 214 nm, where the peptide amide bond absorbs. UV detectors are among the most common detectors in LC because many analytes absorb ultraviolet light[94]. After UV detection, the analytes are further analyzed in the mass spectrometer. Several factors can contribute to separation and its resolution. Solvent choices and the shape of the LC gradient are two of these factors. In reverse-phase chromatography, solvent A is generally water, while solvent B is usually acetonitrile or methanol. Solvents typically have about 0.1% of an organic acid, such as formic acid, which aids with positive ion mode ionization, which will be discussed in the next section. A gradient, starting at low organic solvent and increasing to high organic solvent can greatly aid in separation of analytes. A shallower gradient can separate peaks well, although this action will increase the time, another factor that should be considered. A lower flow rate, a longer column, and a smaller particle size generally improves resolution[94]. A column's efficiency is measured by its number of theoretical plates, N= L/H, in which L is the length of the column packing and H is the plate height. The van Deemter equation illustrates the efficiency of a column as function of flow rate:

(1) $\quad H = A + B/u + C_S u + C_M u$

in which H is the plate height, the A term is Eddy diffusion, the B term is the longitudinal diffusion coefficient, and the C terms are the mass-transfer coefficients of the stationary and mobile phases, respectively[95]. Eddy diffusion is affected by the multiple ways an analyte can travel through a packed column; because the same type of molecule will travel through the column by different paths, the molecules will arrive at the detector at different times, which causes zone broadening, also referred to as the Random Walk Model[95,96]. The B term is affected by the tendency of a solute to migrate from a concentrated to a dilute section of a band, both with and against the flow of the solvent[95]. The C terms consider the time that the

analytes take to be adsorbed and desorbed into/off from a porous particle, as well as stagnant mobile phase stuck in the particle[95].

Advancements in separations have led to smaller columns and smaller stationary particles, resulting in an increased pressure. To accommodate for these innovations, ultra high performance liquid chromatography (UHPLC) and nanoflow-LC have been developed. UHPLC utilizes packing material particles less than 2 μm and LC pressures up to 1200 Bar to achieve increased separation efficiency and shorter analysis time[97-99]. Smaller sized particles reduce the mobile phase C term from the van Deemter equation, which in turn reduces the plate height[97]. Karlsson and Novotny first discovered that a smaller inner diameter correlates with a lower plate height for slurry-packed capillary columns,[100] the first nano-LC, although this improvement was first predicted by Jorgenson[101]. Due to a decreased flow rate, less sample and solvent is consumed and the sensitivity is increased[102,103]. UHPLC systems and nano-LC systems are commercially available and popular in both industry and academic laboratories.

1.3.2 FT-ICR mass spectrometry

Fourier transform ion cyclotron resonance mass spectrometry is one of the most powerful mass analysis methods that exists today, due to its ultra high resolution, unmatched mass accuracy, and its ability to execute many MS/MS fragmentation techniques. The following description of FT-ICR MS specifically applies to the Apex and SolariX models manufactured by Bruker Daltonics. However, the general principles of each part may apply to more than one model or instrument.
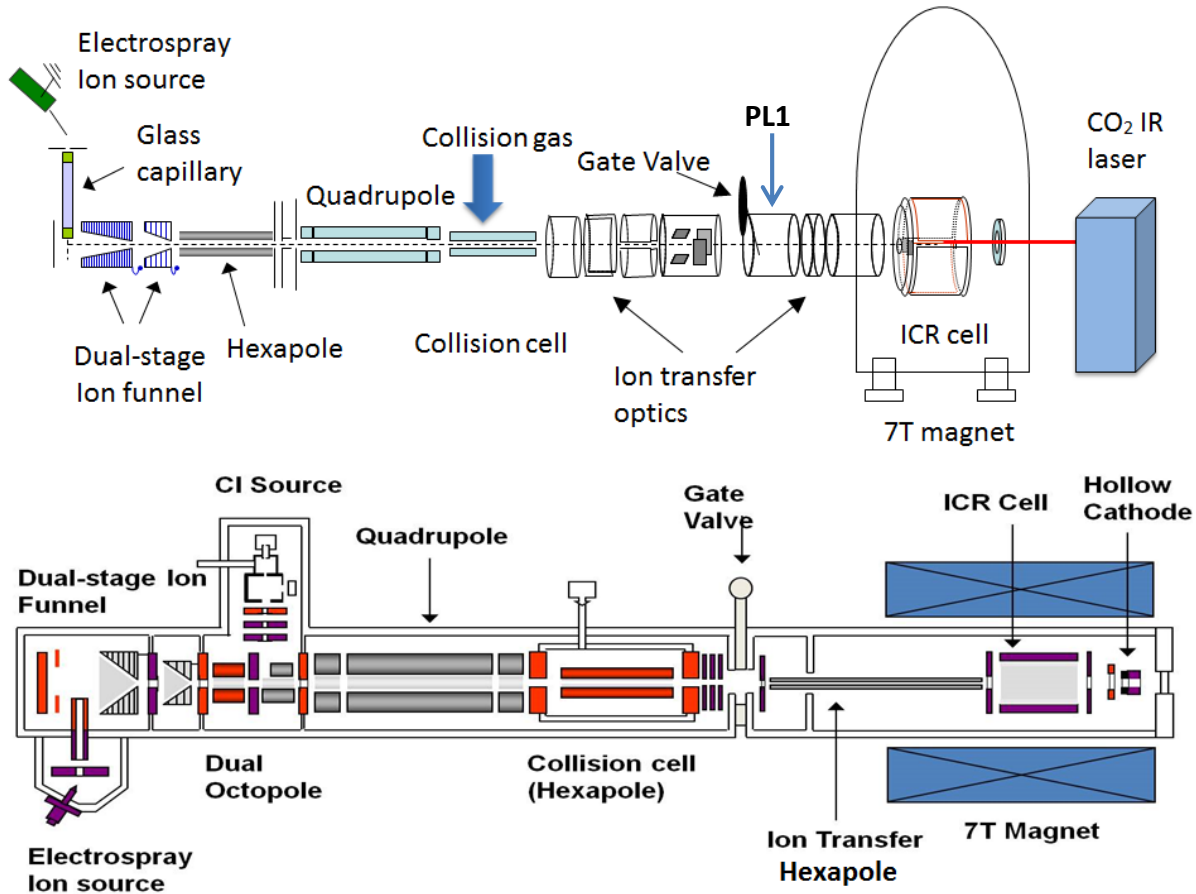
Figure 1.2. Schematic diagrams of Apex (top) and SolariX (bottom) FT-ICR mass spectrometers.

Molecules of interest enter the mass spectrometer via an ionization and volatilization technique. One of the most common ionization techniques for large biomolecules is electrospray ionization. To facilitate ESI, the sample should be dissolved in water, an organic solvent, and either a small percentage of acid or base, although this solvent combination may disrupt the native structure of the analyte. The organic solvent improves solvent volatility, aiding in solvent evaporation, and decreases the surface tension of the sample while the acid or base promotes protonation or deprotonation of the analyte, respectively. The two classic mechanisms for ESI are the charge residue mechanism[104,105] and the ion evaporation mechanism[106]. In ESI, the aqueous sample is infused through a capillary needle into the source region which, in our instruments, is biased at ~4000 volts for negative ion mode and ~-4000

volts for positive ion mode at atmospheric pressure. Typically, the capillary needle is connected to high voltage; however, on our instruments, the spray needles are grounded for safety purposes and the end of a glass capillary, constituting the entrance to the mass spectrometer, is supplied with high voltage instead. The glass capillary is a hollow, glass tube with an inner diameter of about 0.75 mm, coated with platinum on both ends, to provide separate conductivity at the two ends. Due to the high electric field between the ESI capillary needle and the glass capillary inlet to the mass spectrometer, there is a buildup of charge at the surface of the ESI capillary needle that faces the electric field. The charge buildup causes destabilization of the meniscus and draws it out into a Taylor cone. Nitrogen nebulizing gas aids in breaking the meniscus of the surface, and also helps the Taylor cone form. The liquid Taylor cone exists due to the competition between the high electric field and the surface tension of the charged liquid[104]. Charged droplets are formed at the tip of the Taylor cone and are expelled from the cone due to charge repulsion. The charged droplets shrink in size as they evaporate with the aid of heated nitrogen drying gas. As the droplet size shrinks, the charge density in the droplets increases and the droplets undergo Coulomb fission as they reach the Rayleigh limit, defined in Equation 2:

(2) $q_R = 8\pi\varepsilon_0^{1/2}\ \sigma^{1/2}\ r^{3/2}$

in which q_R is the droplet charge, r is the droplet radius, ε_0 is the vacuum permittivity, and σ is the surface tension of the droplet. When the Rayleigh limit occurs, the droplets explode and form smaller 'progeny' droplets[107]. In the charged residue mechanism this process continues and the droplets become increasingly smaller, down to a few nanometers in radius[108]. Once the droplets have reached this size, they are completely desolvated and analytes are in the gas phase.

In the ion evaporation mechanism, the charged droplets are formed in the same manner as the charge residue mechanism. However, the transformation of analytes from droplets into the gas phase is quite different. In the ion evaporation mechanism, the droplets undergo fission to a lesser extent, thus involving larger droplets. The buildup of charge in the droplets

causes the surface electric field to increase until it directly desorbs analyte ions from the surface of the droplets[107]. Both of these mechanisms can account for several theories on ion suppression. Many investigators agree that substances such as salts or other biological entities mixed with the analyte of interest increase the surface tension of the droplets formed from the Taylor cone, preventing droplets from shrinking and evaporating at the same rate and thus decreasing the overall formation efficiency of gas phase ions[109-112]. Another theory states that charged substances compete with the analyte of interest for space at the surface of the droplet[112]. Ion suppression occurs as contaminants and other charged substances enter the gas phase along with the analytes of interest and dominate the signal from the mass spectrometer.

After the ions exit the glass capillary, they enter the first ion funnel. There is a deflection plate that is at a ninety degree angle to the capillary and faces the ion funnels, thus directing ions away from the plate and toward the funnels. The main purpose of the ion funnels is to increase sensitivity by compressing ions together to form a more focused ion beam. The ion funnel is made of ring electrodes that decrease in diameter further into the instrument[113]. There is a DC potential gradient along the ring electrodes thus forming an electric field which pushes the ions into the instrument. An added RF field focuses the ions radially toward the center of the funnel[113], so that the ion beam fits through the end of the funnel, corresponding to Skimmer 1. The purpose of the skimmers is to maintain the pressure difference between the source region and the lower pressure regions closer to the ICR cell[114]. The difference in the skimmer voltages affects ion transmission efficiency[114]. There is an additional ion funnel and skimmer following Skimmer 1. However, the second ion funnel is significantly shorter in length than the first ion funnel in the SolariX instrument and slightly shorter in the Apex instrument. In general, the voltages from the source through the second skimmer decrease to allow the ions to gently make their way through the mass spectrometer.

After the ion funnels come a series of multipoles: a hexapole, a quadrupole and a second hexapole in the Apex instrument. The first hexapole continues the work of the ion funnels, keeping the ions together in a packet[115], so they are transmitted efficiently into the infinity ICR cell. The first hexapole is vital in converting the continuous ion beam into an ion

packet. In the SolariX instrument, a split octopole replaces the first hexapole. The following quadrupole can be utilized for mass selection and the second hexapole acts as the collision cell.

Another important point regarding the multipoles is the pressure change compared to the source. The collision cell, which is the third multipole, has a local higher pressure, closer to 1×10^{-3} mbar to ensure efficient ion cooling in this area, although the gauge reads a lower pressure because of its placement on the instrument. The pressure then continues to decrease closer to the ICR cell. The increased pressure in the third multipole is achieved by adding argon to this area.

A greater number of poles decreases the oscillating radius of ions travelling through a multipole, therefore a hexapole is more efficient for ion transmission than a quadrupole[115].   In both of our mass spectrometers, hexapoles or octopoles are utilized for ion transmission, while the quadrupole is utilized for both ion transmission and isolation, which will be described later. This advantage is shown in Equation 3[115]:

(3) $\qquad R_{k,i} = \rho \left[\dfrac{\sum_{j=1}^{i+k-1} Q_j}{q_{Ni}Q_{max}}\right] \wedge \dfrac{1}{2(N-1)} \; ; \text{k} = 0,1; \; i = 1,\dots, N_{spec} \; ; q_{Ni} = 2N\,(N-1)\dfrac{ez_iV_{RF}}{m_i\omega^2\rho^2}$

 in which N is the number of poles, R is the radial range, Q is the linear (axial) charge density, k is the inner and outer radius of the ion cloud, ρ is the inscribed radius of the multipole, V_{RF} is the radiofrequency (RF) amplitude, and ω is its angular frequency, q_n is the space charge density, m is the ion mass, e is the elementary charge, and Q_{max} is the axial space charge limit for a multipole. i represents each ion species, while k represents an index, designating the inner and outer radius, in which k= 0 and k= 1, respectively. If the radius of the ion cloud is too large, the ions may collide with the poles and lose their charge. RF-only quadrupoles are more likely to 'overfill' and reach storage capacity than other multipoles, because RF confinement in a quadrupole is less efficient for more energetic ions compared with higher order multipoles. This action tends to discriminate against higher m/z species. A greater number of poles reduces the radius of the ion packet, therefore combatting the effects of space charge.

The middle multipole, the quadrupole, consists of 4 poles with the same RF potential except that adjacent poles are 180° out of phase with each other and thus have opposing polarities. The poles across from each other have the same DC and RF potentials, including the

16

same polarity and magnitude. If an RF only potential is applied, the quadrupole will not act as a mass filter and all ions will pass through except for ions that fall below the low mass cutoff. Ions of very low mass are not stable, because they do not fall within the limits of the ion stability diagram. These ions will have an unstable trajectory, quickly increasing their oscillation amplitude and thus eventually hitting a pole or exiting the electric field[116]. When there is superimposed DC potential, the Mathieu equation can be applied to demonstrate a quadrupole acting as a mass filter:

(4) $\qquad \dfrac{d^2 u}{d\xi^2} + (a_u - 2q_u \cos 2\xi)u = 0$,

in which u is the x- or y-coordinates, a and q are trapping parameters (see below) , and ξ is equal to Ωt/2, in which Ω is the RF frequency[117]. The Mathieu equation may be rearranged and reduced to the Mathieu parameters, shown below, in which a and q act as parameters for the mass filter[118].

(5) $\qquad q_{u=} \dfrac{4eV}{mr_0^2 \Omega^2} \qquad a_u = \dfrac{8eU}{mr_0^2 \Omega^2}$

In these parameters, e is the elementary charge, m is the mass, r_0 is the inscribed radius of the quadrupole (8.3 mm in our FT-ICR), Ω is the RF frequency (880 kHz), U is the DC voltage, and V represents the RF amplitude. When these two parameters are plotted against each other, an ion stability diagram is created. The electric field is hyperbolic in shape and ions lying within the a and q region are filtered through, while ions outside of this range collide with a rod in the quadrupole, or escape between the rods and do not make it through. An ideal quadrupole forms a symmetric hyperbolic shape, whereas an ideal hexapole's electric field is more complex and non-hyperbolic in shape. The quadrupole's symmetrical electric field is ideal for a mass filter. The quadrupoles on our mass spectrometers are equipped with a pre- and post-filter which are shorter quadrupoles that sandwich the main quadrupole[118]. Both filters are RF-only quadrupoles and their purpose is to transition the ions in and out of the quadrupole because of the potentially disruptive DC potentials[118]. The filters act as buffers between each multipole and simulate an infinitely long quadrupole.

The hexapole following the quadrupole serves as both the collision cell and an ion trap to store and accumulate ions before they enter the ICR cell. The entrance and exit lenses are placed at either end of this hexapole so that they can act as an axial trap. The ion accumulation time is a tunable parameter, which stores the ions in the hexapole for a given amount of time. The ions exit the hexapole by pulsing the entrance and exit lenses, as well as the DC offset on the hexapole. When this hexapole acts as the collision cell, the MS/MS technique CID occurs. Quadrupole-selected precursor ions undergo higher energy collisions with the added argon gas as a result of increasing the potential difference between the first and third multipole. The kinetic energy from these collisions is converted to precursor ion internal energy[25]. This imparted energy vibrationally excites the precursor ion and the energy is distributed over all of the bonds in the ion. The most labile bonds are subsequently cleaved and the precursor ion is depleted into fragments. For example, in a peptide, the higher vibrational energy results in migration of mobile protons to backbone amide nitrogens, weakening and cleaving peptide backbone bonds to yield N-terminal *b*- and C-terminal *y*-type ions[119]. Neutral losses, such as water and ammonia are also common in CID of peptides.

The ion transfer optics following the collision cell are an additional important feature and, in the Apex instrument, are comprised of a series of lenses that transport the ions through the magnetic field gradient. Because the superconducting magnet surrounding the ICR cell is actively shielded for safety purposes this gradient is steep. The ion optics accelerate the ions axially so that they are efficiently transferred to the ICR cell for analysis. The first lens, PL1 (Figs. 1.2 and 1.3), has a high DC voltage that keeps the ions in the center of the lens. It also contributes to the axial acceleration[120]. The second and third lens, PL2 and PL4 (Fig. 1.3), perform vertical and horizontal ion packet deflection, respectively, and help direct the ions towards the ICR cell. DPL2 and DPL4 (Fig. 1.3) are differential parameters between these lenses. Their purpose is for low energy beam steering and these parameters are quite sensitive[120]. These two parameters are generally kept at or very close to 0 V so that the ions remain on axis. The next lens is the HVO (Fig. 1.3), which is kept at high voltage, generally -2900 V for positive ion mode. The HVO voltage is distributed over the entire set of transfer optics lenses to create a potential well (Fig. 1.3). There are deflection plates in the lens that

steer the ion beam in both the x-and y-directions[120]. The voltages on the deflection plates are fairly low, generally between 50 and 70 V. The FOCL1 and FOCL2 lenses (Fig. 1.3) as well as the PL9 lens (Fig. 1.3) focus the ions further. The latter lenses also slightly decelerate the ions, so that the voltage differential between the HVO and the low voltage ICR cell trapping potentials is minimized. The PL9 lens is an Einzel lens which helps focus the ion beam without changing its energy[120]. The SolariX instrument has a single transfer hexapole in lieu of the high voltage ion transfer optics, accounting for an order of magnitude increase in sensitivity for SolariX.
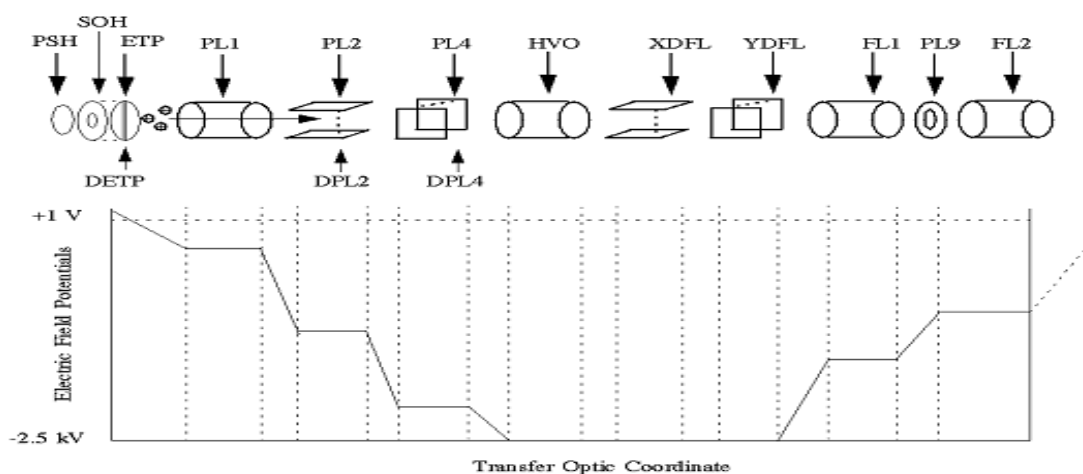
Figure 1.3. Diagram of the ion transfer optics lenses and the electric potential well in the Apex FT-ICR mass spectrometer[120].

In FT-ICR MS, the ICR cell is the mass analyzer and detector. The cell is kept at ultrahigh vacuum, with the pressure in the 10^{-10} mbar range. The cell is held in the bore of the superconducting magnet; both of our magnets have a field strength of 7 Tesla. Although the bore of the magnet is at room temperature, the coil is kept at 4 K to keep the magnet superconducting. On the Apex instrument, a liquid nitrogen region surrounds a liquid helium region to prevent the helium from boiling off too quickly. A vacuum can surrounds both regions. On the SolariX instrument, there is no liquid nitrogen, however, a liquid helium condenser allows the liquid helium to be recycled and therefore boil off at a lower rate. A superconducting magnet is important in FT-ICR MS, because it has no electrical resistance, which means there is no heat produced from the magnet, therefore it does not lose energy as it continually conducts electricity and the magnetic drift is minimum[121]. The ion cyclotron

frequency induced by the magnetic field is directly proportional to the strength of the magnetic field and inversely proportional to the mass to charge ratio. The kinetic theory of the mean free path of ions explains why the ICR cell pressure must be kept very low[116]. The mean free path is the distance that ions can travel without undergoing collisions with gas molecules. The equation for the mean free path, L, is:

(6) $\quad L = \dfrac{kT}{\sqrt{2}p\sigma}$,

in which L is given in meters, k is the Boltzmann constant, T is temperature in K, p is the pressure and σ is the collision cross section. Due to the small size of the cell (~4" diameter) and the relatively lengthy time (~1 s) ions spend in the cell, corresponding to an extensive ion path traveled, the mean free path must be very long, about forty miles at typical cell pressures. Any residual gas in the cell may result in collisions, although the likelihood of this situation decreases with an increasing mean free path. Collisions reduce the velocity and radius of the cyclotron orbit, which may result in losing ions from the cell[12]. Space charge effects also may cause Coulomb peak broadening[12].

There are several options for the geometry of an analyzer cell, including open, closed, cylindrical and cubic[122]. The Infinity cells in the Hakansson lab are closed cylindrical cells. The purpose of the Infinity cell is to minimize z-axis ejection parallel to the magnetic field lines[123]. Eliminating these ejection events optimizes the excitation stage in the mass analyzer and provides more consistent signal abundancies[123]. It also increases the signal to noise ratio by allowing for an increased excitation radius which in turn leads to more abundant image currents for detection[123]. The trapping plate geometry of the Infinity cell simulates RF mirrors and a cell that is infinitely long[123], which would lack z-axis ejection events. The cell is divided into two pairs of electrodes located across from each other, one for excitation and the other for detection.

The three steps occurring in the cell in an MS experiment are trapping, excitation, and detection of ions. The plates that assist these steps are seen in Figure 1.4. The trapping plates axially confine the ions in the cell and allow them to move back and forth in a simple harmonic motion along the lines of the magnetic field[124]. Ions are confined radially by their cyclotron motion[12]. The overall ion motion in the xy-plane is a spiral-like movement with superimposed

magnetron motion, where the center of the cyclotron orbit is precessing around the cell[116]. The cyclotron motion is induced by the Lorentz force (equation 7), which causes the ions to have an orbit perpendicular to the magnetic field, **B**, and the ion velocity **v**.[12]. Other relevant factors include the ion's mass, m and the charge q, which equals the elementary charge, e, multiplied by the number of elementary charges, z. This outward centrifugal force may also be expressed as equation 8, in which r is the orbit's radius. Equating equations 7 and 8, and substituting ω, the angular frequency for v/r, with some rearrangement yields equation 9. The cyclotron frequency can be measured very accurately because the frequency is measured and averaged over hundreds of rotations and therefore the mass to charge ratio may also be measured very accurately, in accordance with equation 9.

(7) $\quad$ F=$ze\boldsymbol{v} \ x \ \boldsymbol{B}$ or F= $q\boldsymbol{v} \ x \ \boldsymbol{B}$

(8) $\quad$ $F = \dfrac{mv^2}{r}$

$\quad$**(9)** $\quad$ $\omega = \dfrac{qB}{m}$

On the other hand, the magnetron motion is governed by the magnetic field in combination with the electric field from the trapping plates[12,122]. The magnetic field moves ions away from the cell wall, while the electric field moves ions away from the center of the cell[12].

In order for detection to occur, the ions are first excited with an RF signal across the excitation plates containing frequencies equal to the cyclotron frequencies of the ions[124]. As the ions are being excited, they move according to the applied RF chirp into coherent ion packets with increased cyclotron radius at their m/z-dependent frequencies. These coherently rotating ion packets induce an alternating signal in the detection plates[124]. To minimize dephasing, ions must be excited for a short amount of time (typically ~50 microseconds)[116]. The process of dephasing is comparable to T_2 (spin-spin relaxation) dephasing in NMR spectroscopy except Coulomb repulsion occurs for the charged ions. The analogy of T_1 (spin-lattice relaxation) in NMR is charge neutralization due to collisions[95].

Ions are detected by their image current: coherently excited ions generate an alternating current between the two detection plates in the cell[12,124]. This current is converted to an RF voltage, which is displayed as a free induction decay (FID) or a transient with sine like

properties, providing time domain information about the detected ions[124]. The time domain signal is converted to the frequency domain via Fourier transformation and the frequency domain spectrum is in turn converted to a mass spectrum via equation (9). A unique aspect of image current detection is that it is non-destructive, so that the ions can be detected for a long time[12]. As for any time domain spectroscopy, a longer transient results in higher resolution[12]. A long mean free path is essential for preventing collisions in the cell, thus resulting in longer transients and therefore higher resolution.
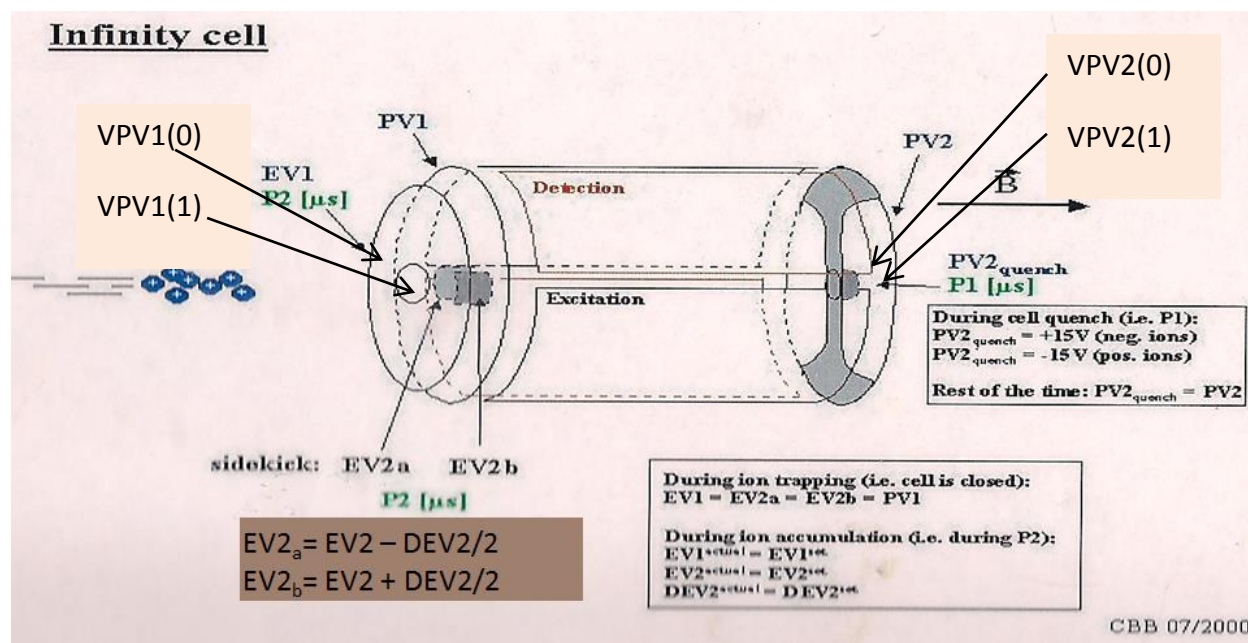
Figure 1.4. Schematic diagram of the ICR cell parameters, particularly displaying Sidekick trapping voltages.

The analyzer entrance voltage, EV1 (applied on the outside of the trapping plate facing the ion source, see Figure 1.4), gently attracts the ions into the cell and acts as a fine adjustor of the ion axial energy[120]. The voltages PV1 and PV2 are applied to the inside of the trapping plates, causing ions to oscillate along the magnetic field lines. The PV1 and PV2 voltages are set to the same polarity and confine the ions axially, so that they do not escape from the cell. The VPV1(0), VPV2(0), VPV1(1), and VPV2(1) voltages (Figure 1.4), also applied to the inside of the trapping plates, are tuned for gated or dynamic trapping. In dynamic trapping, the trapping potentials are gradually changed between two preset values, (0) and (1), for each trapping plate. Without this gradual change, the trapped ions will have a wider energy range which, in

22

turn, decreases the m/z range[125]. In the absence of dynamic trapping, Sidekick trapping is employed. This kind of trapping is accomplished by utilizing an additional element, EV2 (Fig. 1.4), at the entrance to the ICR cell. A potential difference, DEV2, is applied between the two halves of the EV2 element, creating a voltage gradient across the cell entrance and thus facilitates ion cyclotron motion[120]. P3 and PL3 are very important parameters because they control excitation of the ion packets. PL3 controls the amplitude of the excitation waveform and P3 is the excitation time. These 'excite' parameters control the ion cyclotron radius prior to detection. If the ions are too close to the detection plates, mass spectral peaks may split, which will also happen if there are too many ions in the cell[126]. If ions hit the detection plates, signal will quench. On the other hand, a too small cyclotron radius leads to loss of signal. On the Apex instrument, the PL3 parameter acts as a damper, however on SolariX, it corresponds directly to excitation amplitude. Overall, the different cell voltages depend on one another and are therefore very sensitive to tuning.

There have been several recent advancements in FT-ICR MS, including multiple cell designs. One attractive design is the dynamically harmonized cell, put into production by Bruker Daltonics based on a design by Professor Eugene Nikolaev who has shown a resolving power of about twelve million in a 4.7 T magnet[127]. The electric field applied to an ICR cell is necessary to trap ions axially, however, this necessary field interrupts a potentially perfect ion motion, assuming a perfectly homogenous magnetic field and an absence of ion interactions with other ions or neutrals. The electric field causes magnetron motion, which disturbs the cyclotron motion. The z-axis amplitude also disturbs the cyclotron motion as an increase in z-amplitude increases the electric field gradient[128]. High z-amplitude ions rotate with lower velocities, which eventually results in de-phasing of ion packets as they are analyzed and detected[128]. One way to lessen the effects of the electric field is to create an electric field with a hyperbolic shape, which allows the ion cyclotron frequency to be independent of axial ion motion in the cell[128]. To create a hyperbolic electric field, Professor Nikolaev proposed to create a potential distribution that would be harmonic once the cyclotron field is averaged, rather than creating a harmonized field, a previous idea for other cell designs[129-133]. To accomplish this feat, the Nikolaev cell is segmented with concave and convex parabolic shaped

slits, with the number of segments varying, in which a larger number of segments corresponds to a higher efficacy of averaging the potential[128]. The National High Magnetic Field Laboratory (NHMFL) at Florida State University and the Pacific Northwest National Laboratory have been working separately on record 21 Tesla FT-ICR instruments[134,135]. Recently, the NHMFL has successfully brought the 21 T magnet to field and it remains at field. There has been a large push for higher magnetic fields due to their many advantages: Increasing magnetic fields have a linear relationship with increased resolving power and acquisition speed and have a quadratic relationship with increasing dynamic range, ion trapping time, and ion kinetic energy[136].

1.4 Protein Post-Translational Modifications

One of the primary reasons for the complexity of the proteome is the large number of post-translational modifications, which exponentially increase the size of the proteome[137-139].

Figure 1.5. The proteome is more complex than the genome or transcriptome, primarily due to an estimate of about one million PTMs[2]. Reprinted with permission from Munoz et al. Copyright 2014 Angewandte Chemie[2].
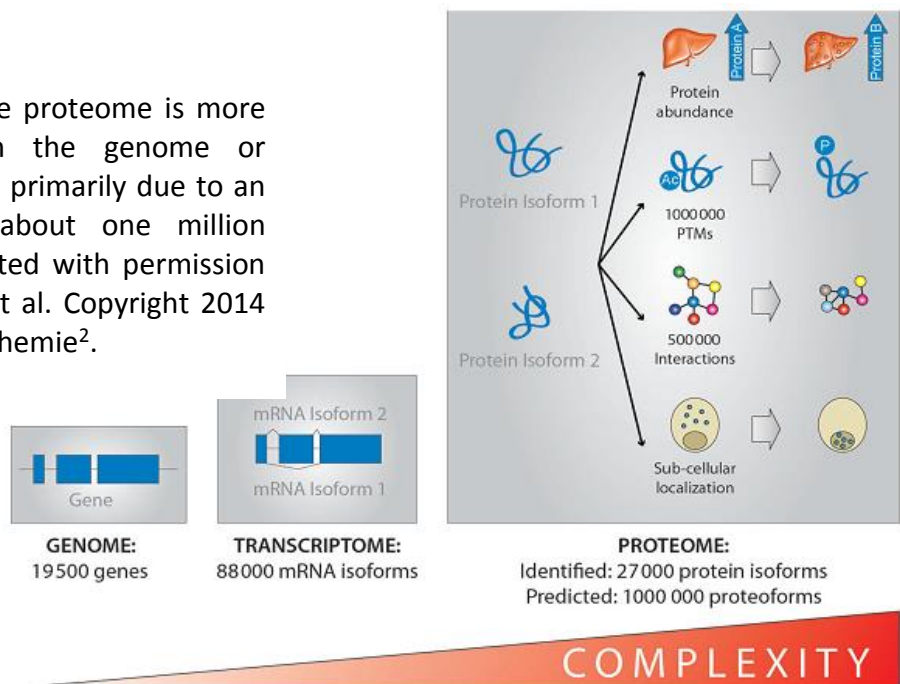
A PTM is a covalent chemical modification to a protein and though the numbers vary, it is estimated that at least 80% of proteins in eukaryotes are modified with PTMs[140]. PTMs play a large variety of roles in proteins and regulate protein function, as well as play a significant role in protein structure[138]. PTMs significantly influence protein-protein interactions, cell signaling,

DNA repair, cell division, enzyme activity, activation and inactivation of proteins, and regulate cellular pathways.[138,139,141]. Malfunctioning PTMs have been linked to a number of disorders and diseases[126]. PTMs are generally of low abundance, dynamic, and labile, thus presenting numerous challenges[18]. Mass spectrometry is an ideal tool for detection, identification, and characterization of PTMs due to its high sensitivity and versatility, including multiple MS/MS techniques. Although there are over 300 different PTMs,[137,142] this thesis will only focus on two: phosphorylation and phosphopantetheinylation.

1.4.1 Protein phosphorylation

According to the PTM statistics curator, which pulls updated statistics from the Swiss-Prot knowledgebase, phosphorylation is the most frequent PTM[143]. Phosphorylation is a prominent PTM in eukaryotes; estimates of phosphorylated proteins reach up to one third of all proteins in cells[144-147]. Phosphorylation is found on serine, threonine, tyrosine, and occasionally histidine and involves the addition of a phosphate functional group to the side chain of these amino acids.
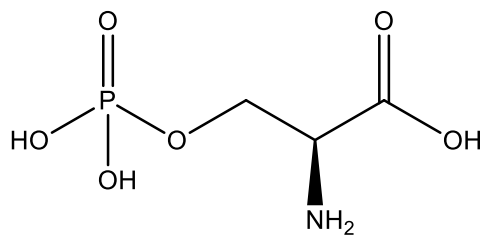
Figure 1.6. Structure of a phosphorylated serine.

phosphoserine

Phosphorylation is thought to be the most understood PTM. Phosphorylation is reversible; proteins are phosphorylated with kinases and dephosphorylated with phosphatases and there are hundreds of known kinases and phosphatases[18]. Phosphorylation often changes protein structure, and thus function[148,149]. Phosphorylation is vital because it is involved in many cellular processes, including signal transduction in different pathways and is responsible for regulation of different intracellular occurrences, such as gene expression and protein synthesis[145,147,150-152]. The prominence and frequency of phosphorylation has opened up the field of phosphorylation enrichment. Immobilized metal affinity chromatography (IMAC) utilizes metal ions such as Ga^{3+} or Fe^{3+} in a resin of either iminodiacetic acid or nitrilotriacetic

acid. Negatively charged phosphate groups bind strongly to these metals and phosphoproteins or phosphopeptides can thus be isolated and later eluted under basic conditions[153-157]. Although IMAC is a popular technique, its overall specificity is somewhat low. Metal oxide affinity chromatography (MOAC) is a similar method to IMAC, except that metal oxides, such as ZrO_2 and TiO_2, constitute the matrix[158-160]. MOAC has higher specificity to phosphorylated peptides than IMAC. More recently, a method has been developed that exploits monodisperse microsphere IMAC beads that contain a linker to phosphonate groups that chelate Ti^{4+} [161]. This approach combines IMAC and MOAC, has high specificity and is robust due to the structure of the materials that comprise this technique. The microspheres are uniform and their surface is hydrophilic, thus minimizing non-specific adsorption. The overall structure of the linkers and phosphonate group optimize the spatial orientation for phosphopeptide enrichment[161].

1.4.2 Protein phosphopantetheinylation

Phosphopantetheine (Ppant) is a much less common PTM than phosphorylation; yet it is just as vital to the proteins it is found on. Ppant is covalently bound to acyl carrier proteins (ACPs) and peptidyl carrier proteins (PCPs) that are found in fatty acid synthases (FASs), polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs)[162,163]. Ppant chemical moieties are bound to serine residues and are long (about 2 nm) and flexible. The Ppant terminal thiol group covalently binds substrates or natural product intermediates created in the FAS, PKS, or NRPS biosynthetic pathways[164].
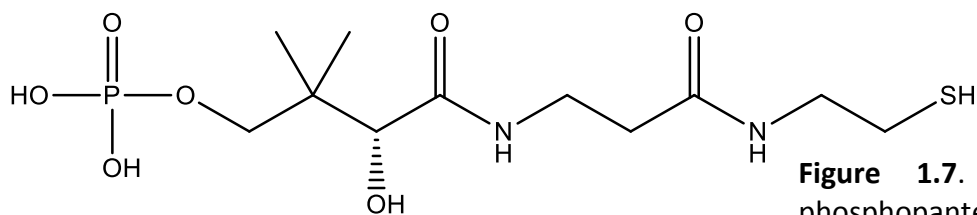
Figure 1.7. Structure of a phosphopantetheine arm.

Ppant's length and flexibility allow abound reaction intermediates to interact with various protein domains and their active sites, so that elongation can occur to approach the structure of the final product. The crystal structure of a Ppant transferase/PCP complex was recently solved and the structural basis for Ppant transfer was shown to be highly conserved through

organisms[165]. Although there are no direct enrichment strategies for Ppant like there are for phosphorylation, there are several methods that utilize Ppant for discovery of novel natural products and biosynthetic pathways, however details about these techniques will be discussed in further detail in section 1.5 and Chapter 5[166,167].

1.5 Natural Products and their Biosynthetic Pathways

1.5.1 Natural products as medicines

Forty to over fifty percent of drugs on the market today derive from natural products[168,169]. Nature continues to be a rich source for antibiotics, vaccines, anti-cancer drugs, as well as other drugs[170,171]. Some well-known examples of natural products are penicillin from the skin of an orange or morphine, derived from opium, which is extracted from poppy seeds[168]. One route for creating natural products for medicinal use is total synthesis, however, it is costly and challenging to synthesize complex and diverse structures. An alternative route is chemoenzymatic synthesis, which is accomplished by feeding a biosynthetic enzyme and allowing it to create the natural product. This process is much simpler than total synthesis, provided the biosynthetic pathway is known. As a result, a field has developed containing biochemists and medicinal chemists who are focused on learning how biosynthetic pathways work, discovering novel biosynthetic pathways and novel natural products as well as the design and engineering of these pathways.

1.5.2 Polyketide synthases and non-ribosomal peptide synthetases

Polyketides constitute one class of natural products, biosynthesized in an assembly line-like process by proteins called type I and type II polyketide synthases. Polyketide synthases are megadaltons in size and consist of several polypeptides (modules), each containing multiple protein domains of varying function[162]. Each module acts as an independent step in the assembly line-like chain elongation process, adding a two or three carbon unit to the growing natural product intermediate. A PKS module consists of at least a ketosynthase (KS), an acyl transferase (AT), and an ACP domain, although they may have additional domains to promote structural diversity in the final natural products[172]. PKS natural product diversity stems from

the order of each domain, and the combinations of modules and domains. The assembly line starts with an acyl CoA, usually malonyl-CoA or methylmalonyl-CoA[162]. The KS domain is about 45 kDa and catalyzes a C-C bond formation through a Claisen condensation and results in two or three carbons from malonyl or methylmalonyl respectively being added to the growing intermediate[162,173]. The AT is a 50 kDa domain that selects either malonyl or methylmalonyl and covalently transfers it to the Ppant of the ACP[162]. The ACP is a smaller protein, about 12 kDa, whose role is to tether and transfer the natural product intermediate to other domains and modules via the terminal thiol residue on the Ppant arm, which was previously described in section 1.4[162,174,175]. The ACP is considered active or 'holo' when it is modified with Ppant and 'apo' without the Ppant modification[176,177]. There are several other domains which add diversity to the natural product intermediate such as ketoreductases, which reduce a β-ketone. Dehydratase domains dehydrate a β-hydroxy group to an α-β alkene, while enoyl reductases reduce the α-β alkene to an alkane[162,174]. Finally, thioesterase (TE) domains hydrolyze and release the finished natural product, often as a cyclic molecule[162,178].

Non-ribosomal peptides are a related class of natural products, biosynthesized byNRPSs. Many natural products are hybrids of polyketides and non-ribosomal peptides, constructed from biosynthetic pathways with both PKS and NRPS enzymes, each enacting their part in the assembly line[163]. Although non-ribosomal peptides are formed in a similar manner to polyketides, they show some key differences, including their building blocks. NRPSs utilize activated amino acids or aryl acids rather than malonyl/methylmalonyl, thus adding diversity as all 20 amino acids can act as building blocks. The NRPS domains are also different, although the required domains are comparable with those in PKSs: Condensation (C), adenylation (A), and PCP are the three required domains in each module. The C domain is similar to a KS in that it elongates the natural product intermediate, however it catalyzes C-N bond formation through transamidation rather than C-C bond formation[163]. The intermediate acts as a nucleophile and is attacked by the free amino group from the upstream module intermediate[163]. The A domain is comparable to the AT domain, as it selects an amino acid, activates the carboxylate end with ATP to make an aminoacyl-AMP group and then transfers it to the PCP domain[163,176]. The PCP domain has the same function as the ACP and has a Ppant PTM. Like PKSs, NRPSs have several

28

other domains that add diversity to the final product. Epimerization domains epimerize amino acids from D to L or vice versa to control stereochemistry[163,179]. *N*-methylation domains transfer *S*-methyl groups from *S*-adenosylmethionine to an amino group and thus strengthen the product by methylating all of the nitrogens on the peptide backbone, preventing proteolytic breakdown[179]. Oxazoline and thiazoline rings may be reduced or oxidized by domains as well[163]. An additional, yet rare, modification is formylation of the peptide N-terminus by a formylation domain[163]. NRPSs also have a TE domain for hydrolysis and potential cyclization; however, they also have the option of a reduction domain for releasing the non-ribosomal peptide[163]. The reduction domain reduces the C-terminus of the natural product to an aldehyde or alcohol[163].

1.5.3 Biochemical and biophysical methods for natural product characterization and discovery

Structural biology techniques such as X-ray crystallography and NMR spectroscopy have provided some insight into the partial structure of PKSs and NRPSs. X-ray crystallography is a high resolution technique, but it cannot provide dynamic information and flexible proteins are difficult to crystallize. Despite these limitations, X-ray crystallography has been successful at characterizing domain structures of natural product biosynthetic pathways or partial biosynthetic pathways, and a full structure of a fatty acid synthase is available[180-185]. NMR spectroscopy is size limited, but useful for characterizing the natural products alone, including details such as stereochemistry[186,187]. 2D NMR can also determine the skeletal structure of natural products[187,188] and may be employed to characterize the structure of acyl and peptidyl carrier proteins[189-193]. Recently, vibrational spectroscopy has elucidated Ppant dynamics and determined its solution exposure in the 6-deoxyerythronolide B synthase (DEBS) biosynthetic pathway through installation of a thiocyanate vibrational spectroscopy probe[194]. Cryo-electron microscopy (cryo-EM) is another emerging technique for characterizing large protein complex structures and their dynamics. The power of this technique will be demonstrated in Chapter 3.

Mass spectrometry is an effective method in characterizing natural products and natural product intermediates, including enzyme-substrate intermediates. MS can be utilized for high-throughput screening and for developing libraries of natural products[195]. Other MS methods

utilize the presence of the Ppant modification. For example, CID or IRMPD will eject the Ppant modification from the ACP, creating a diagnostic ion of 261.1 Da (as long as the ACP is solely in its 'holo' state). Meluzzi et al. further employed MS^3 on a low resolution ion trap to create a high confidence method to measure the Ppant modification and any bound substrates[164]. Professor Neil Kelleher has developed several methods, including a Ppant ejection assay[169] and proteomic investigation of secondary metabolism (PrISM), which targets the Ppant modification following MS/MS in unsequenced genomes[167]. Another targeted approach for natural product biosynthetic pathway discovery is orthogonal active site identification system (OASIS), which targets TEs or ACPs with chemical probes[166]. However, OASIS has not yet been applied to complex mixtures and uses a protein database to identify natural products, thus biasing the data towards known pathways. The Hakansson lab also utilizes MS methods for characterization and discovery of natural products and biosynthetic pathways. Top-down MS and IRMPD were utilized to characterize parts of the biosynthetic pathways for two natural products, curacin A and jamaicamide[196,197]. Ultra high resolution has been employed to monitor active site occupancy for natural and unnatural building blocks, characterization of enzyme-substrate intermediates and products, and mining for the presence of an unsequenced biosynthetic pathway[198-200].

1.6 Dissertation Overview

The research in this dissertation focuses on utilizing ultra high resolution bottom-up as well as top-down mass spectrometry coupled to LC to discern kinetics for the multisite phosphoprotein Sic1, to characterize several natural product biosynthetic pathways, and to develop a method to selectively identify unsequenced biosynthetic pathways. Chapter 2 determines the phosphorylation kinetics of Sic1, a protein with nine phosphorylation sites. This chapter was a collaborative project with Professor Xiaoxia Nina Lin in the Chemical Engineering Department. Mass spectrometry data, which will be compared to mathematical models, reveal that each site has a different phosphorylation rate, therefore supporting hypothesized switch-like response of the protein. In order to perform this work, a novel negative ion mode quantitative method was developed. Chapter 3 determines and confirms the structural states

of PikAIII, the third module in the pikromycin biosynthetic pathway. This chapter involved a collaboration with Professor Janet Smith, Professor David Sherman, and Professor Georgios Skiniotis of the Life Sciences Institute. In this chapter, I improved and applied the previous method utilized in the Hakansson lab for confirming active sites and bound substrates. My findings on the precise covalent protein state agreed with cryo-EM data collected in the Skiniotis laboratory. This work has been published in a back-to-back article and letter in *Nature*[201,202]. Chapter 4 further characterizes parts of the curacin and jamaicamide biosynthetic pathways via phosphopantetheine ejection. This project is a collaboration with Professor Janet Smith in the Biological Chemistry Department/Life Sciences Institute. Mass spectrometry confirmed the specific amino acids that are crucial for the structures of the unique natural product intermediates. Mass spectrometry also confirmed the activity of a unique cyclopropanase, created by my collaborators. Chapter 5 investigates discovery of novel biosynthetic pathways. This chapter highlights a novel method which selectively detects and fragments peptides containing Ppant, more specifically the secondary phosphate in Ppant via IRMPD at 10.6 μm. In addition, Chapter 6 summarizes all results and includes a section on future directions.

1.7 References

(1)   Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J.-C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Nat. Biotechnol.* **1996**, *14*, 61.

(2)   Munoz, J.; Heck, A. J. From the human genome to the human proteome. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10864.

(3)   Tyers, M.; Mann, M. From genomics to proteomics. *Nature* **2003**, *422*, 193.

(4)   Altelaar, A. F.; Munoz, J.; Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2013**, *14*, 35.

(5)   Protein Characterization Using Modern Biophysical TechniquesMalik, S. S.; Shrivastava, T. In *Advances in Protein Chemistry*; Ashraf, G., Sheikh, I. A., Eds.; OMICS Group eBooks: Foster City, CA, 2013.

(6)   Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Part 2. Techniques for the Study of Biological Structure and Function*; W.H. Freeman & Company: New York, NY, 1980.

(7)   Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198.

(8)   Marginean, I.; Tang, K.; Smith, R. D.; Kelly, R. T. Picoelectrospray Ionization Mass Spectrometry Using Narrow-Bore Chemically Etched Emitters. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 30.

(9)   Chalmers, M. J.; Kolch, W.; Emmett, M. R.; Marshall, A. G.; Mischak, H. Identification and analysis of phosphopeptides. *J. Chromatogr. B* **2004**, *803*, 111.

(10)   Desiere, F.; Deutsch, E. W.; Nesvizhskii, A. I.; Mallick, P.; King, N. L.; Eng, J. K.; Aderem, A.; Boyle, R.; Brunner, E.; Donohoe, S.; Fausto, N.; Hafen, E.; Hood, L.; Katze, M. G.; Kennedy, K. A.; Kregenow, F.; Lee, H.; Lin, B.; Martin, D.; Ranish, J. A.; Rawlings, D. J.; Samelson, L., E; Shiio, Y.; Watts, J. D.; Wollscheid, B.; Wright, M. E.; Yan, W.; Yang, L.; Yi, E. C.; Zhang, H.; Aebersold, R. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.* **2004**, *6*, R9.1.

(11)   Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J. R., 3rd; Bairoch, A.; Bergeron, J. J. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Meth.* **2010**, *7*, 681.

(12)   Amster, I. J. Fourier Transform Mass Spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1325.

(13)   Perry, R. H.; Cooks, R. G.; Noll, R. J. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom. Rev.* **2008**, *27*, 661.

(14)   Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. High-Performance Mass Spectrometry: Fourier Transform Ion Cyclotron Resonance at 14.5 Tesla. *Anal. Chem.* **2008**, *80*, 3985.

(15)   Switzar, L.; Giera, M.; Niessen, W. M. Protein digestion: an overview of the available techniques and recent developments. *J. Proteome Res.* **2013**, *12*, 1067.

(16)   Chait, B. T. Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314*, 65.

(17)   Brownridge, P.; Beynon, R. J. The importance of the digest: Proteolysis and absolute quantification in proteomics. *Methods* **2011**, *54*, 351.

(18)   Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269.

(19)   Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806.

(20)   Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1998**, *33*, 1.

(21)   Yates, J. R. Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 297.

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

(23)   Bensimon, A.; Heck, A. J.; Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **2012**, *81*, 379.

(24)   Ong, S. E.; Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, *1*, 252.

(25)   McLuckey, S. A. Principles of Collisional Activation in Analytical Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 599.

(26)   Senko, M. W.; Speir, J. P.; McLafferty, F. W. Collisional Activation of Large Multiply Charged Ions Using Fourier Transform Mass Spectrometry. *Anal. Chem.* **1994**, *66*, 2801.

(27)   Little, D. P.; Speir, P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994**, *66*, 2809.

(28)   Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120*, 3265.

(29)   Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528.

(30)   Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. Electron Transfer Dissociation of Peptide Anions. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 880.

(31)   Yoo, H. J.; Wang, N.; Zhuang, S.; Song, H.; Hakansson, K. Negative-Ion Electron Capture Dissociation: Radical-Driven Fragmentation of Charge-Increased Gaseous Peptide Anions. *J. Am. Chem. Soc.* **2011**, *133*, 16790.

(32)   Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. Electron detachment dissociation of peptide di-anions: an electron-hole recombination phenomenon. *Chem. Phys. Lett.* **2001**, *342*, 299.

(33)   Cody, R. B.; Freiser, B. S. Electron impact excitation of ions in Fourier transform mass spectrometry. *Anal. Chem.* **1987**, *59*, 1054.

(34)   Cody, R. B.; Freiser, B. S. Electron impact excitation of ions from organics: an alternative to collision induced dissociation. *Anal. Chem.* **1979**, *51*, 547.

(35)   Gord, J. R.; Horning, S. R.; Wood, J. M.; Cooks, R. G.; Freiser, B. S. Energy deposition during electron-induced dissociation. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 145.

(36)   Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little, D. P.; McLafferty, F. W. 193 nm photodissociation of larger multiply-charged biomolecules. *Int. J. Mass Spectrom.* **1996**, *157–158*, 357.

(37)   Gimon-Kinsel, M. E.; Kinsel, G. R.; Edmondson, R. D.; Russell, D. H. Photodissociation of High Molecular Weight Peptides and Proteins in a Two-Stage Linear Time-Of-Flight Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 578.

(38)   Zhang, H.; Ge, Y. Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry. *Circ. Cardiovasc. Genet.* **2011**, *4*, 711.

(39)   McLafferty, F. W.; Breuker, K.; Jin, M.; Han, X.; Infusini, G.; Jiang, H.; Kong, X.; Begley, T. P. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* **2007**, *274*, 6256.

(40)   Lanucara, F.; Eyers, C. E. Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrom. Rev.* **2013**, *32*, 27.

(41)   Bogdanov, B.; Smith, R. D. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* **2005**, *24*, 168.

(42)   Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A. Top-Down ESI-ECD-FT-ICR Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites. *J. Am. Chem. Soc.* **2006**, *128*, 14432.

(43)   Clarke, D.; Murray, E.; Hupp, T.; Mackay, C. L.; Langridge-Smith, P. R. Mapping a Noncovalent Protein–Peptide Interface by Top-Down FTICR Mass Spectrometry Using Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1432.

(44)   Loo, J. A.; Benchaar, S. A.; Zhang, J. Integrating Native Mass Spectrometry and Top-Down MS for Defining Protein Interactions Important in Biology and Medicine. *Mass Spectrom. (Tokyo)* **2013**, *2*, S0013.

(45)   Zhang, H.; Cui, W.; Wen, J.; Blankenship, R.; Gross, M. Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1966.

(46)   Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011**, *480*, 254.

(47)   Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. Native Top-Down Electrospray Ionization-Mass Spectrometry of 158 kDa Protein Complex by High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2013**, *86*, 317.

(48)   Snijder, J.; van de Waterbeemd, M.; Damoc, E.; Denisov, E.; Grinfeld, D.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A.; Heck, A. J. R. Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136*, 7295.

(49)   She, Y.-M.; Xu, X.; Yakunin, A. F.; Dhe-Paganon, S.; Donald, L. J.; Standing, K. G.; Lee, D. C.; Jia, Z.; Cyr, T. D. Mass Spectrometry Following Mild Enzymatic Digestion Reveals Phosphorylation of Recombinant Proteins in Escherichia coli Through Mechanisms Involving Direct Nucleotide Binding. *J. Proteome Res.* **2010**, *9*, 3311.

(50)   Wu, C.; Tran, J. C.; Zamdborg, L.; Durbin, K. R.; Li, M.; Ahlf, D. R.; Early, B. P.; Thomas, P. M.; Sweedler, J. V.; Kelleher, N. L. A protease for 'middle-down' proteomics. *Nat. Meth.* **2012**, *9*, 822.

(51)   von Pawel-Rammingen, U.; Johansson, B. P.; Bjork, L. IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *EMBO J.* **2002**, *21*, 1607.

(52)   Tsai, P. L.; Chen, S.-F.; Huang, S. Y. Mass spectrometry-based strategies for protein disulfide bond identification. *Rev. Anal. Chem.* **2013**, *32*.

(53)   Zhang, W.; Marzilli, L. A.; Rouse, J. C.; Czupryn, M. J. Complete disulfide bond assignment of a recombinant immunoglobulin G4 monoclonal antibody. *Anal. Biochem.* **2002**, *311*, 1.

(54)   Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-throughput middle-down analysis using an orbitrap. *J. Proteome Res.* **2010**, *9*, 3886.

(55)   Kalli, A.; Sweredoski, M. J.; Hess, S. Data-Dependent Middle-Down Nano-Liquid Chromatography–Electron Capture Dissociation-Tandem Mass Spectrometry: An Application for the Analysis of Unfractionated Histones. *Anal. Chem.* **2013**, *85*, 3501.

(56)   Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. *Mol. Cell Proteomics* **2009**, *8*, 2266.

(57)   Schulze, W. X.; Usadel, B. Quantitation in Mass-Spectrometry-Based Proteomics. *Annu. Rev. Plant Biol.* **2010**, *61*, 491.

(58)   Liebler, D. C.; Zimmerman, L. J. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry* **2013**, *52*, 3797.

(59)   Pandey, A. SILAC. [Online Early Access]. Published Online: 2013. http://silac.org/research_articles.

(60)   Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell Proteomics* **2002**, *1*, 376.

(61)   Gouw, J. W.; Krijgsveld, J.; Heck, A. J. R. Quantitative Proteomics by Metabolic Labeling of Model Organisms. *Mol. Cell Proteomics* **2010**, *9*, 11.

(62)   Ong, S.-E. The expanding field of SILAC. *Anal. Bioanal. Chem.* **2012**, *404*, 967.

(63)   Lanucara, F.; Eyers, C. E. In *Methods in Enzymology*; Jameson, D., Hans, M. V., Westerhoff, V., Eds.; Academic Press: 2011; Vol. Volume 500, p 133.

(64)   Krüger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fässler, R.; Mann, M. SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell*, *134*, 353.

(65)   Doherty, M. K.; Hammond, D. E.; Clague, M. J.; Gaskell, S. J.; Beynon, R. J. Turnover of the Human Proteome: Determination of Protein Intracellular Stability by Dynamic SILAC. *J. Proteome Res.* **2008**, *8*, 104.

(66)   Bonenfant, D.; Towbin, H.; Coulot, M.; Schindler, P.; Mueller, D. R.; van Oostrum, J. Analysis of Dynamic Changes in Post-translational Modifications of Human Histones during Cell Cycle by Mass Spectrometry. *Mol. Cell Proteomics* **2007**, *6*, 1917.

(67)   Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat. Meth.* **2013**, *10*, 332.

(68)   Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6591.

(69)   Gaudin, Z.; Cerveau, D.; Marnet, N.; Bouchereau, A.; Delavault, P.; Simier, P.; Pouvreau, J.-B. Robust Method for Investigating Nitrogen Metabolism of 15N Labeled Amino Acids Using AccQ•Tag Ultra Performance Liquid Chromatography-Photodiode Array-Electrospray Ionization-Mass Spectrometry: Application to a Parasitic Plant–Plant Interaction. *Anal. Chem.* **2013**, *86*, 1138.

(70)   Trompelt, K.; Steinbeck, J.; Terashima, M.; Hippler, M. A new approach for the comparative analysis of multiprotein complexes based on 15N metabolic labeling and quantitative mass spectrometry. *J. Vis. Exp.* **2014**, *85*, e51103.

(71)   Lyon, D.; Castillejo, M. A.; Staudinger, C.; Weckwerth, W.; Wienkoop, S.; Egelhofer, V. Automated Protein Turnover Calculations from 15N Partial Metabolic Labeling LC/MS Shotgun Proteomics Data. *PLoS One* **2014**, *9*, 1.

(72)   Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994.

(73)   Li, J.; Steen, H.; Gygi, S. P. Protein Profiling with Cleavable Isotope-coded Affinity Tag (cICAT) Reagents: The Yeast Salinity Stress Response. *Mol. Cell Proteomics* **2003**, *2*, 1198.

(74)   Hansen, K. C.; Schmitt-Ulms, G.; Chalkley, R. J.; Hirsch, J.; Baldwin, M. A.; Burlingame, A. L. Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable 13C-Isotope-coded Affinity Tag and Multidimensional Chromatography. *Mol. Cell Proteomics* **2003**, *2*, 299.

(75)   Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal. Chem.* **2003**, *75*, 6843.

(76)   Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **2007**, *389*, 1017.

(77)   Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6940.

(78)   Desiderio, D. M.; Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biol. Mass Spectrom.* **1983**, *10*, 471.

(79)   Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell Proteomics* **2004**, *3*, 1154.

(80)   Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; Pribil, P.; Lee, K. H. 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **2007**, *7*, 3651.

(81)   Pierce, A.; Unwin, R. D.; Evans, C. A.; Griffiths, S.; Carney, L.; Zhang, L.; Jaworska, E.; Lee, C.-F.; Blinco, D.; Okoniewski, M. J.; Miller, C. J.; Bitton, D. A.; Spooncer, E.; Whetton, A. D. Eight-channel iTRAQ Enables Comparison of the Activity of Six Leukemogenic Tyrosine Kinases. *Mol. Cell Proteomics* **2008**, *7*, 853.

(82)   Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly". *J. Proteome Res.* **2009**, *8*, 5347.

(83)   Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Meth.* **2011**, *8*, 937.

(84)   Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895.

(85)   Patel, V. J.; Thalassinos, K.; Slade, S. E.; Connolly, J. B.; Crombie, A.; Murrell, J. C.; Scrivens, J. H. A Comparison of Labeling and Label-Free Mass Spectrometry-Based Proteomics Approaches. *J. Proteome Res.* **2009**, *8*, 3752.

(86)   Zhu, W.; Smith, J. W.; Huang, C. M. Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* **2010**, *2010*, 840518.

(87)   Bondarenko, P. V.; Chelius, D.; Shaler, T. A. Identification and Relative Quantitation of Protein Mixtures by Enzymatic Digestion Followed by Capillary Reversed-Phase Liquid Chromatography–Tandem Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 4741.

(88)   Chelius, D.; Bondarenko, P. V. Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *J. Proteome Res.* **2002**, *1*, 317.

(89)   Steen, H.; Jebanathirajah, J. A.; Springer, M.; Kirschner, M. W. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3948.

(90)   Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell Proteomics* **2005**, *4*, 1265.

(91)   Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A* **2008**, *105*, 18132.

(92)   Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **2007**, *25*, 117.

(93)   Horvath, C. G.; Preiss, B. A.; Lipsky, S. R. Fast liquid chromatography. Investigation of operating parameters and the separation of nucleotides on pellicular ion exchangers. *Anal. Chem.* **1967**, *39*, 1422.

(94)   Harris, D. C. In *Quantitative Chemical Analysis*; W.H. Freeman and Co.: New York, NY, 2007, p 556.

(95)   Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*; 6 ed.; Cengage Learning: Independence, KY, 2006.

(96)   Giddings, J. C.; Byring, H. A Molecular Dynamic Theory of Chromatography. *J. Phys. Chem.* **1955**, *59*, 416.

(97)   MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. Ultrahigh-Pressure Reversed-Phase Liquid Chromatography in Packed Capillary Columns. *Anal. Chem.* **1997**, *69*, 983.

(98)   MacNair, J. E.; Patel, K. D.; Jorgenson, J. W. Ultrahigh-Pressure Reversed-Phase Capillary Liquid Chromatography: Isocratic and Gradient Elution Using Columns Packed with 1.0-μm Particles. *Anal. Chem.* **1999**, *71*, 700.

(99)   Tolley, L.; Jorgenson, J. W.; Moseley, M. A. Very High Pressure Gradient LC/MS/MS. *Anal. Chem.* **2001**, *73*, 2985.

(100)   Karlsson, K.-E.; Novotny, M. Separation Efficiency of Slurry-Packed Liquid Chromatography Microcolumns with Very Small Inner Diameters. *Anal. Chem.* **1988**, *60*, 1662.

(101)   Kennedy, R. T.; St. Claire III, R. L.; White, J. G.; Jorgenson, J. W. Chemical Analysis of Single Neurons by Open Tubular Liquid Chromatography. *Mikrochim. Acta* **1987**, *11*, 37.

(102)   Hernandez-Borges, J.; Aturki, Z.; Rocco, A.; Fanali, S. Recent applications in nanoliquid chromatography. *J. Sep. Sci.* **2007**, *30*, 1589.

(103)   Saz, J. M.; Marina, M. L. Application of micro- and nano-HPLC to the determination and characterization of bioactive and biomarker peptides. *J. Sep. Sci.* **2008**, *31*, 446.

(104)   Kebarle, P.; Tang, L. From Ions in Solution to Ions in the Gas Phase: The Mechanism of Electrospray Mass Spectrometry. *Anal. Chem.* **1993**, *65*, 972.

(105)   Dole, M.; Mack, L. L.; Hines, R. C.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49*, 2240.

(106)   Iribarne, J. V.; Thomson, B. A. On the Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976**, *64*, 2287.

(107)   Grimm, R. L.; Beauchamp, J. L. Evaporation and Discharge Dynamics of Highly Charged Multicomponent Droplets Generated by Electrospray Ionization. *J. Phys. Chem. A* **2010**, *114*, 1411.

(108)   Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85*.

(109)   Fu, I.; Woolf, E.; Matuszewski, B. Effect of the sample matrix on the determination of indinavir in human urine by HPLC with turbo ion spray tandem mass spectrometric detection. *J. Pharm. Biomed. Anal.* **1998**, *18*, 347.

(110)   Annesley, T. Ion suppression in mass spectrometry. *Clin. Chem.* **2003**, *49*, 1041.

(111)   Mallet, C.; Lu, Z.; Mazzeo, J. A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 49.

(112)   Jessome, L. L.; Volmer, D. A. Ion suppression: a major concern in mass spectrometry. *Advanstar Communications Inc.* **2006**.

(113)   Kim, T.; Tolmachev, A. V.; Harkewicz, R.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. Design and Implementation of a New Electrodynamic Ion Funnel. *Anal. Chem.* **2000**, *72*, 2247.

(114)   Murata, H.; Takao, T.; Shimonishi, Y. Optimization of Skimmer Voltages of an Electrospray Ion Source Coupled with a Magnetic Sector Instrument. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 205.

(115)   Tolmachev, A. V.; Udseth, H. R.; Smith, R. D. Radial stratification of ions as a function of mass to charge ratio in collisional cooling radio frequency multipoles used as ion guides or ion traps. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1907.

(116)   de Hoffmann, E.; Stroobant, V. *Mass Spectrometry Principles and Applications*; 3 ed.; John Wiley and Sons: West Sussex, England, 2007.

(117)   March, R. E. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *J. Mass Spectrom.* **1997**, *32*, 351.

(118)   Q-Interface User and Service ManualRMB; 001 ed.; Daltonics, B., Ed. Bilerica, MA, 2003; Vol. AQHU01UM.

(119)   Qin, J.; Chait, B. T. Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino Acid Residues. *J. Am. Chem. Soc.* **1995**, *117*, 5411.

(120)   Daltonics, B. *Acquisition- The Basics*.

(121)   Dixon, I. R. Superconducting Magnet. [Online Early Access]. Published Online: 2013. http://www.magnet.fsu.edu/education/tutorials/magnetminute/superconductingmagnet-transcript.html (accessed 6/21/13).

(122)   Marshall, A. G. Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass Spectrom.* **2000**, *200*, 331.

(123)   Caravatti, P.; Allemann, M. The 'Infinity Cell': a New Trapped-ion Cell With Radiofrequency Covered Trapping Electrodes for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Org. Mass Spectrom.* **1991**, *26*, 514.

(124)   Daltonics, B. *SolariX User Manual*, 2010.

(125)   Gorshkvov, M. V.; Masselon, C. D.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. Dynamically assisted gated trapping for Fourier transform ion cyclotron mass spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1558.

(126)   Wang, Y.-C.; Peterson, S. E.; Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **2014**, *24*, 143.

(127)   Popov, I. A.; Nagornov, K.; Vladimirov, G. N.; Kostyukevich, Y. I.; Nikolaev, E. N. Twelve million resolving power on 4.7 T Fourier transform ion cyclotron resonance instrument with dynamically harmonized cell-observation of fine structure in peptide mass spectra. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 790.

(128)   Boldin, I. A.; Nikolaev, E. N. Fourier transform ion cyclotron resonance cell with dynamic harmonization of the electric field in the whole volume by shaping of the excitation and detection electrode assembly. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 122.

(129)   Jackson, G. S.; White, F. M.; Guan, S.; Marshall, A. G. Matrix-shimmed ion cyclotron resonance ion trap simultaneously optimized for excitation, detection, quadrupolar axialization, and trapping. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 759.

(130)   Bruce, J. E.; Anderson, G. A.; Lin, C.-Y.; Gorshkov, M.; Rockwood, A. L.; Smith, R. D. A novel high-performance Fourier transform ion cyclotron resonance cell for improved biopolymer characterization. *J. Mass Spectrom.* **2000**, *35*, 85.

(131)   Tolmachev, A. V.; Robinson, E. W.; Wu, S.; Kang, H.; Lourette, N. M.; Pasa-Tolic, L.; Smith, R. D. Trapped-ion cell with improved DC potential harmonicity for FT-ICR MS. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 586.

(132)   Gabrielse, G.; Haarsma, L.; Rolston, S. L. Open-Endcap Penning Traps for High Precision Experiments. *Int. J. Mass Spectrom.* **1989**, *88*, 319.

(133)   Brustkern, A. M.; Rempel, D. L.; Gross, M. L. An Electrically Compensated Trap Designed to Eight Order for FT-ICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1281.

(134)   Development of an FT-ICR mass spectrometer in preparation for 21 teslaKaiser, N.; Weisbrod, C.; Quinn, J.; Blakney, G. T.; Beu, S.; Chen, T.; Hendrickson, C. L.; Marshall, A. G. In *62nd ASMS Conference on Mass Spectrometry and Allied Topics* Baltimore, MD, 2014.

(135)   Comprehensive Evaluation of Preamplifier Configurations and Performance Parameters for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry at 21 TeslaLin, T.-Y.; Anderson, G. A.; Norheim, R. V.; Leach III, F. E.; Shaw, J. B.; Tolmachev, A. V.; Roscioli, K. M.; O'Connor, P. B.; Koppenaal, D. W.; Robinson, E. W.; Pasa-Tolic, L. In *62nd ASMS Conference on Mass Spectrometry and Allied Topics* Baltimore, MD, 2014.

(136)   Marshall, A. G.; Guan, S. Advantages of High Magnetic Field for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1819.

(137)   Zhao, Y.; Jensen, O. N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **2009**, *9*, 4632.

(138)   Karve, T. M.; Cheema, A. K. Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease. *J. Amino Acids* **2011**, *2011*.

(139)   Wang, X.; Pattison, J. S.; Su, H. Posttranslational Modification and Quality Control. *Circ. Res.* **2013**, *112*, 367.

(140)   Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*; 4 ed.; W.H. Freeman: New York, 2000.

(141)   Hitosugi, T.; Chen, J. Post-translational modifications and the Warburg effect. *Oncogene* **2014**, *33*, 4279.

(142)   Walsh, C. T. *Posttranslational modification of proteins : Expanding nature's inventory.*; Roberts and Co. Publishers: Englewood, CO, 2006.

(143)   Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **2011**, *1*.

(144)   Mann, M.; Ong, S.-E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **2002**, *20*, 261.

(145)   Yan, J. X.; Packer, N. H.; Gooley, A. A.; Williams, K. L. Protein Phosphorylation: technologies for the identification of phosphoamino acids. *J. Chromatogr. A* **1998**, *808*, 23.

(146)   Blume-Jensen, P.; Hunter, T. Oncogenic kinase signaling. *Nature* **2001**, *411*, 355.

(147)   Chalmers, M. J.; Quinn, J. P.; Blakney, G. T.; Emmett, M. R.; Mischak, H.; Gaskell, S. J.; Marshall, A. G. Liquid Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometric Characterization of Protein Kinase C Phosphorylation. *J. Proteome Res.* **2003**, *2*, 373.

(148)   Duclos, B.; Marcandier, S.; Cozzone, A. J. In *Methods in Enzymology*; Tony Hunter, B. M. S., Ed.; Academic Press: 1991; Vol. Volume 201, p 10.

(149)   Hunter, T. A thousand and one protein kinases. *Cell* **1987**, *50*, 823.

(150)   Krebs, E. G. The growth of research on protein phosphorylation. *Trends Biochem. Sci.* **1994**, *19*, 439.

(151)   Faux, M. C.; Scott, J. D. More on target with protein phosphorylation: conferring specificity by location. *Trends Biochem. Sci.* **1996**, *21*, 312.

(152)   Hunter, T. Protein Kinases and Phosphatases: The Yin and Yang of Protein Phosphorylation and Signaling. *Cell* **1995**, *80*, 225.

(153)   Choi, H.; Lee, S.; Jun, C.-D.; Park, Z.-Y. Development of an off-line capillary column IMAC phosphopeptide enrichment method for label-free phosphorylation relative quantification. *J. Chromatogr. B* **2011**, *879*, 2991.

(154)   Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C. Large Scale Analysis of in Vivo Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry. *Mol. Cell Proteomics* **2003**, *2*, 1234.

(155)   Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* **2002**, *19*, 301.

(156)   Fila, J.; Honys, D. Enrichment techniques employed in phosphoproteomics. *Amino acids* **2012**, *43*, 1025.

(157)   Andersson, L.; Porath, J. Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. *Anal. Biochem.* **1986**, *154*, 250.

(158)   Sano, A.; Nakamura, H. Chemo-affinity of Titania for the Column-switching HPLC Analysis of Phosphopeptides. *Anal. Sci.* **2004**, *20*, 565.

(159)   Kweon, H. K.; Hakansson, K. Selective Zirconium Dioxide-Based Enrichment of Phosphorylated Peptides for Mass Spectrometric Analysis. *Anal. Chem.* **2006**, *78*, 1743.

(160)   Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal. Chem.* **2004**, *76*, 3935.

(161)   Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J.; Zou, H.; Mohammed, S. Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* **2013**, *8*, 461.

(162)   Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106*, 3468.

(163)   Sieber, S. A.; Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chem. Rev.* **2005**, *105*, 715.

(164)   Meluzzi, D.; Zheng, W. H.; Hensler, M.; Nizet, V.; Dorrestein, P. C. Top-down mass spectrometry on low-resolution instruments: characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3107.

(165)   Tufar, P.; Rahighi, S.; Kraas, Femke I.; Kirchner, Donata K.; Löhr, F.; Henrich, E.; Köpke, J.; Dikic, I.; Güntert, P.; Marahiel, Mohamed A.; Dötsch, V. Crystal Structure of a PCP/Sfp Complex Reveals the Structural Basis for Carrier Protein Posttranslational Modification. *Chem. Biol.* **2014**, *21*, 552.

(166)   Meier, J. L.; Niessen, S.; Hoover, H. S.; Foley, T. L.; Cravatt, B. F.; Burkart, M. D. An Orthogonal Active Site Identification System (OASIS) for Proteomic Profiling of Natural Product Biosynthesis. *ACS Chem. Biol.* **2009**, *4*, 948.

(167)   Bumpus, S. B.; Evans, B. S.; Thomas, P. M.; Ntai, I.; Kelleher, N. L. A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* **2009**, *27*, 951.

(168)   Lahlou, M. The Success of Natural Products in Drug Discovery. *Pharmacol. Pharm.* **2013**, *04*, 17.

(169)   Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry* **2006**, *45*, 12756.

(170)   Li, J. W.; Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **2009**, *325*, 161.

(171)   Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311.

(172)   Hutchinson, C. R. Microbial polyketide synthases: More and more prolific. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3336.

(173)   Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7*, 285.

(174)   Staunton, J.; Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **2001**, *18*, 380.

(175)   Mercer, A. C.; Burkart, M. D. The ubiquitous carrier protein-a window to metabolitebiosynthesis. *Nat. Prod. Rep.* **2007**, *24*, 750.

(176)   Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, *97*, 2651.

(177)   Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A new enzyme superfamily — the phosphopantetheinyl transferases. *Chem. Biol.* **1996**, *3*, 923.

(178)   Walsh, C. T. Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science* **2004**, *303*, 1805.

(179)   Schwarzer, D.; Finking, R.; Marahiel, M. A. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **2003**, *20*, 275.

(180)   Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315.

(181)   Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11124.

(182)   Zheng, J.; Piasecki, S. K.; Keatinge-Clay, A. T. Structural Studies of an A2-Type Modular Polyketide Synthase Ketoreductase Reveal Features Controlling α-Substituent Stereochemistry. *ACS Chem. Biol.* **2013**, *8*, 1964.

(183)   Ramelot, T. A.; Rossi, P.; Forouhar, F.; Lee, H. W.; Yang, Y.; Ni, S.; Unser, S.; Lew, S.; Seetharaman, J.; Xiao, R.; Acton, T. B.; Everett, J. K.; Prestegard, J. H.; Hunt, J. F.; Montelione, G. T.; Kennedy, M. A. Structure of a specialized acyl carrier protein essential for lipid A biosynthesis with very long-chain fatty acids in open and closed conformations. *Biochemistry* **2012**, *51*, 7239.

(184)   Raman, M. C. C.; Johnson, K. A.; Clarke, D. J.; Naismith, J. H.; Campopiano, D. J. The serine palmitoyltransferase from Sphingomonas wittichii RW1: An interesting link to an unusual acyl carrier protein. *Biopolymers* **2010**, *93*, 811.

(185)   Lohman, J. R.; Ma, M.; Cuff, M. E.; Bigelow, L.; Bearden, J.; Babnigg, G.; Joachimiak, A.; Phillips Jr., G. N.; Shen, B. The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. *Proteins* **2014**, *80*, 1210.

(186)   Lang, G.; Mayhudin, N. A.; Mitova, M. I.; Sun, L.; van der Sar, S.; Blunt, J. W.; Cole, A. L. J.; Ellis, G.; Laatsch, H.; Munro, M. H. G. Evolving Trends in the Dereplication of Natural Product Extracts: New Methodology for Rapid, Small-Scale Investigation of Natural Product Extracts. *J. Nat. Prod.* **2008**, *71*, 1595.

(187)   Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S. NMR in Metabolomics and Natural Products Research: Two Sides of the Same Coin. *Acc. Chem. Res.* **2011**, *45*, 288.

(188)   Breton, R. C.; Reynolds, W. F. Using NMR to identify and characterize natural products. *Nat. Prod. Rep.* **2013**, *30*, 501.

(189)   Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **2007**, *16*, 2093.

(190)   Zhou, Z.; Lai, J. R.; Walsh, C. T. Directed evolution of aryl carrier proteins in the enterobactin synthetase. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11621.

(191)   Evans, S. E.; Williams, C.; Arthur, C. J.; Burston, S. G.; Simpson, T. J.; Crosby, J.; Crump, M. P. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem.* **2008**, *9*, 2424.

(192)   Lim, J.; Kong, R.; Murugan, E.; Ho, C. L.; Liang, Z.-X.; Yang, D. Solution Structures of the Acyl Carrier Protein Domain from the Highly Reducing Type I Iterative Polyketide Synthase CalE8. *PLoS One* **2011**, *6*, e20549.

(193)   Li, Q.; Khosla, C.; Puglisi, J. D.; Liu, C. W. Solution Structure and Backbone Dynamics of the Holo Form of the Frenolicin Acyl Carrier Protein†,◊. *Biochemistry* **2003**, *42*, 4648.

(194)   Johnson, M. N. R.; Londergan, C. H.; Charkoudian, L. K. Probing the Phosphopantetheine Arm Conformations of Acyl Carrier Proteins Using Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2014**, *136*, 11240.

(195)   Ito, T.; Masubuchi, M. Dereplication of microbial extracts and related analytical technologies. *J. Antibiot.* **2014**, *67*, 353.

(196)   Gu, L.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Hakansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459*, 731.

(197)   Khare, D.; Wang, B.; Gu, L.; Razelun, J.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L. Conformational switch triggered by α-ketoglutarate in a halogenase of curacin A biosynthesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14099.

(198)   Bonnett, S. A.; Rath, C. M.; Shareef, A.-R.; Joels, J. R.; Chemler, J. A.; Hakansson, K.; Reynolds, K.; Sherman, D. H. Acyl-CoA Subunit Selectivity in the Pikromycin Polyketide Synthase PikAIV: Steady State Kinetics and Active-Site Occupancy Analysis by FTICR-MS. *Chem. Biol.* **2011**, *18*, 1075.

(199)   Rath, C. M.; Janto, B.; Earl, J.; Ahmed, A.; Hu, F. Z.; Hiller, L.; Dahlgren, M.; Kreft, R.; Yu, F.; Wolff, J. J.; Kweon, H. K.; Christiansen, M. A.; Håkansson, K.; Williams, R. M.; Ehrlich, G. D.; Sherman, D. H. Meta-omic Characterization of the Marine Invertebrate Microbial Consortium That Produces the Chemotherapeutic Natural Product ET-743. *ACS Chem. Biol.* **2011**, *6*, 1244.

(200)   Ding, Y.; Rath, C. M.; Bolduc, K. L.; Hakansson, K.; Sherman, D. H. Chemoenzymatic synthesis of cryptophycin anticancer agents by an ester bond-forming non-ribosomal peptide synthetase module. *J. Am. Chem. Soc.* **2011**, *133*, 14492.

(201)   Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **2014**, *510*, 512.

(202)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014**, *510*, 560.

Chapter 2

Site-Specific Kinetic Analysis of Mutant and Wildtype Sic1 Phosphorylation

2.1 Introduction

Phosphorylation is a prominent protein post-translational modification (PTM) in eukaryotes; estimates of phosphorylated proteins reach up to one third of all proteins in cells[1-4]. Phosphorylation is significant because of its involvement in many cellular processes, including signal transduction in numerous pathways, and it is also responsible for regulation of various intracellular occurrences, such as gene expression and protein synthesis[2,4-7]. Many signaling proteins have multiple phosphorylation sites and many biological systems show switch-like responses to external signals. It has been hypothesized that multisite phosphorylation in proteins may be one mechanism underlying switch-like responses[9-11]. Experimental observations[12] suggest that multisite phosphorylation can generate switch-like biological responses[13,14], however, the mechanism for triggering such behavior is relatively unknown.

Sic1, a protein found in budding yeast, has nine reported phosphorylation sites, which serve to trigger switch-like behavior during the cell cycle[9,10]. Sic1 functions by inhibiting Clb5,6-Cdc28 kinase complexes, which initiate DNA replication[15-17]. Sic1 is phosphorylated by the Cdc28 kinase and phosphorylation of precisely six of the nine sites results in ubiquitination and subsequent degradation of the protein, a key step in regulating the cell cycle, specifically the transition from the G1 to S phase, as indicated in Figure 2.1[9,18,19].
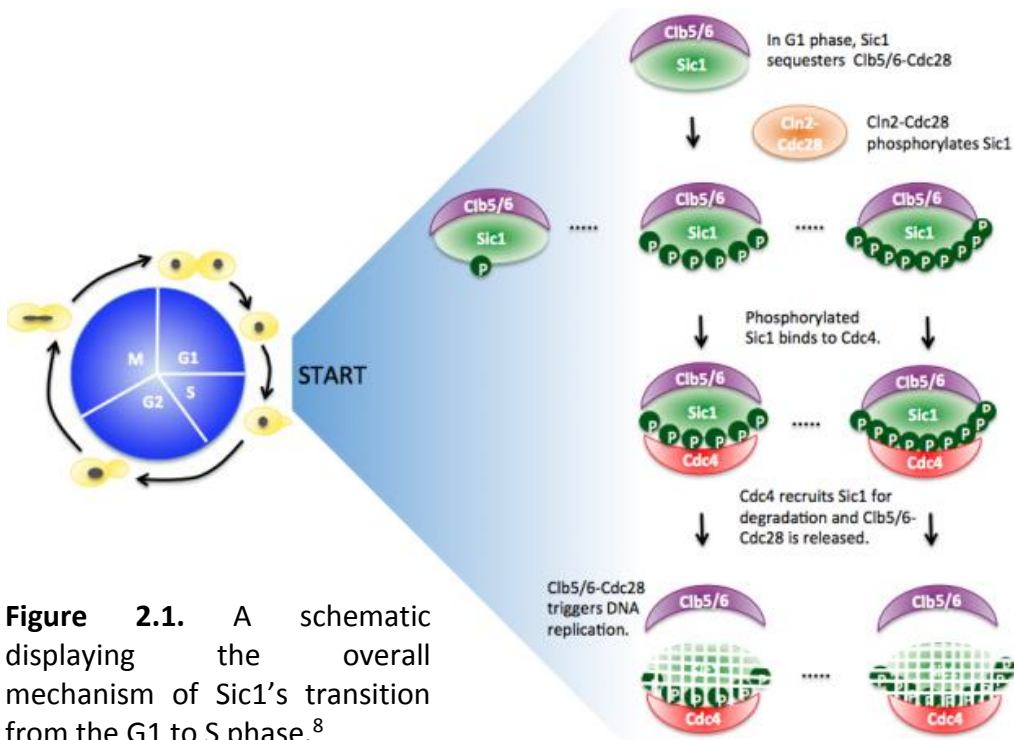
Figure 2.1. A schematic displaying the overall mechanism of Sic1's transition from the G1 to S phase.[8]

One hypothesis is that the mechanism of the switch-like response is a chain-like reaction[20,21]. The chain-like reaction is triggered by different phosphorylation states of the protein. Sic1's homologue, p27[Kip1], is a tumor suppressor found in mammalian cells, where low levels of the protein are linked to various diseases, including cancers[22,23]. Although the phosphorylation-dependent degradation of Sic1 and p27[Kip1] appears to be conserved, details regarding the degradation mechanism is yet to be discovered[24].

Coupling liquid chromatography (LC) to a mass spectrometer is a long standing technique and is customary in proteomics applications, including quantitation and detection of PTMs. In complex mixtures, this practice decreases ion suppression in the mass spectrometer, causing the signal to noise ratio to increase and also increases peak capacity. Mass spectrometry is an established technique for analyzing phosphopeptides and proteins, including quantitation, because it is accurate and can be done without chemical derivatization or labeling, which are employed in antibody recognition and Edman sequencing, two alternative techniques[2,25]. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is a highly powerful technique because of its high resolving power and mass accuracy, arguably, the highest of all mass analyzers[26-28]. The most common way of analyzing molecules in a mass spectrometer is in

positive ion mode for several reasons: Standard LC columns have an upper pH limit around 7 but work best at lower pH, lending itself to peptide protonation. In addition, collision induced dissociation tandem mass spectrometry (CID MS/MS) yields predictable fragmentation patterns for peptide cations and online databases and other informatics software therefore are based on positive ion mode mass spectra. Most MS detectors are electron multipliers, which contribute a high noise level in negative ion mode due to the presence of free electrons, furthering the preferred choice of positive ion mode. However, FT-ICR MS utilizes image current detection in which noise level is unaffected by polarity. Additionally, acidic molecules, such as phosphorylated peptides and proteins, can be difficult to ionize in positive ion mode[29,30]. Analyzing acidic analytes in negative ion mode increases the sensitivity and lowers the limit of detection dramatically[31,32]. High sensitivity and resolution are necessary to detect the peptide isotopes for label-free quantitation; therefore, all of the mass spectrometry work in this Chapter is conducted in negative ion mode, a novel aspect of this work. Corona discharge can be a major issue in negative ion mode electrospray ionization, preventing stable ion signal[33]. However, with source tuning, including the use of nebulizing gas, corona discharge was not a problem in this work.

Here, we aim to quantitatively determine site-specific Sic1 phosphorylation kinetics to shed light on the mechanism behind the switch-like response. All nine phosphorylation sites are characterized over six tryptic peptides and their phosphorylation rates are determined. A new quantitation method is utilized with internal standards showing similar ionization efficiencies as the peptides of interest.

2.2 Experimental

Mutant (2p) and wildtype Sic1 with an N-terminal six residue histidine tag were overexpressed in *E.coli* and affinity purified via the His tag with a Ni-NTA column (Qiagen) following dialysis with Slide-a-Lyzer Dialysis Cassettes (Pierce). Sic1 was phosphorylated with the kinase Cdc28 (gift from the Protein Expression Center of CalTech) and phosphorylation was quenched after 0, 5, 10, 20, 40, 80, and 180 minutes, respectively, with EDTA (Life Technologies). Each timepoint utilized 5 μL of 4 mg/mL kinase and 75 μL of 0.22 to 0.32 mg/mL Sic1 protein, depending on the replicate. Sic1 mutant internal standard peptides (see Table 2.1) were

purchased from Anaspec, except for the singly phosphorylated 169-177 peptide, which was purchased from GenicBio. Sic1 was denatured by heat at 95 degrees Celsius for 30 minutes, digested with trypsin (Promega) at 37 degrees Celsius overnight and quenched with 1.5 µL of formic acid (ThermoFisher Scientific). Each Sic1 internal standard was dissolved in water to a final concentration of 2 mg/mL followed by 5:100 dilution. Each sample for LC/MS analysis contained 2 µL of each peptide internal standard and 8 µL of digested and phosphorylated Sic1.

Fifteen microliters of digested Sic1 with internal standards were injected onto a Synergi Hydro C18 hydrophilically endcapped 1 × 150 mm column with 4 µm particles (Phenomenex). A gradient of 2-60% acetonitrile with 0.1% formic acid over seventy minutes was generated with an Agilent 1100 HPLC coupled to a quadrupole FTICR-MS (Apex or SolariX with 7T magnet, Bruker Daltonics). Flow was at 50 µl min^{-1} and was diverted for the first 5 min of the run. Buffer A was 0.1% formic acid in HPLC-grade water (ThermoFisher Scientific), and buffer B was 0.1% formic acid in acetonitrile (ThermoFisher Scientific). Data were gathered from m/z 200–2000 in negative ion mode. Electrospray was conducted at 4,200 V with four scans per spectrum and a 256k transient. External ion accumulation in a hexapole was 0.2 s and there was 1 ICR fill before excitation and detection. External calibration was performed with HP tuning mix (Agilent). Each experiment was run in triplicate.

Peaks and peak areas were chosen and integrated with Data Analysis (Bruker Daltonics). Isotope percentages were calculated with Protein Prospector (http://prospector.ucsf.edu/prospector/mshome.htm). All other calculations were carried out in Microsoft Excel. The 2p mutant Sic1 protein has seven out of nine phosphorylation sites mutated to an alanine residue.

Standard peptide	Peptide Sequence	Peptide Molecular Weight (monoisotopic mass in Da)
2p mutant: 33-50-T33A V42L	*A*PQKPSQN*L*LPV**T**PSTTK	1906.052/ 1986.019
2p mutant: 54-79-L57V S69A	NAP*V*LAPPNSNMGMT*A*PFNGLT**S**PQR	2681.305/ 2761.272
wt: 33-50-V42L	**T**PQKPSQN*L*LPV**T**PSTTK	1936.063/2016.029/2095.996

wt: 54-79-L57V	NAP*V*LAPPNSNMGMT**S**PFNGLT**S**PQR	2697.3/ 2777.267/ 2857.233
wt: His tag+ 1-8- L13V	MHHHHHHSSGVD*V*GTENLYFQSNA**T**PS **T**PPR	3476.582/3556.548/ 3636.514
wt: 80-84 *^{13}C ^{15}N	**S**P*FPK	580.1/ 660.0663
wt: 169-177-V170L	D*L*PG**T**PSDK	928.4502/ 1008.416
wt: 186-193 *^{13}C ^{15}N	NWNNN**S**P*K	978.1/ 1058.066

Table 2.1. List of each standard peptide for the 2p and wt Sic1 phosphorylated peptides. Red bolded residues indicate phosphorylation sites whereas italicized residues indicate substitutions in the internal standards. Starred proline residues indicate ^{13}C and ^{15}N labels.

2.3 Results and Discussion

2.3.1 Label-free method development

A label-free method was developed specifically to quantify Sic1 phosphorylated peptides, however, this method is applicable to nearly all peptides. Internal standard peptides must have comparable ionization efficiencies to the peptides that are being quantified. Valine and leucine residues only differ by a methyl group, thus Val to Leu or Leu to Val substitutions likely yield similar ionization efficiencies[34]. Such substitutions are more cost-effective than isotopic labeling. The Sic1 2p mutant yields two phosphopeptides after trypsin digestion, both containing a valine or leucine residue. Out of the six phosphopeptides (three with one phosphorylation site and three with two phosphorylation sites) from trypsin digestion of wt Sic1, the four larger peptides all contain leucine or valine. For the remaining two wt phosphopeptides that lack leucine or valine, a proline is present and internal standard peptides were instead doubly labeled with ^{13}C and ^{15}N at this residue. The Leu to Val and Val to Leu substitutions were validated for differences in ionization efficiencies. The 33-50 peptide ionization efficiency difference was 7 ± 4%, the 54-79 peptide ionization efficiency difference was 9 ± 23%, the His1-8 peptide ionization efficiency difference was 39 ± 3%, and the 169-177 peptide ionization efficiency difference was 9 ± 6%. Example extracted ion chromatograms (EICs) used to calculate the ionization efficiency difference for the 33-50 peptide are displayed in Figure 2.2. The majority of these ionization

efficiency differences are minor; therefore, the Val to Leu and Leu to Val substitutions are an acceptable method of quantitation.
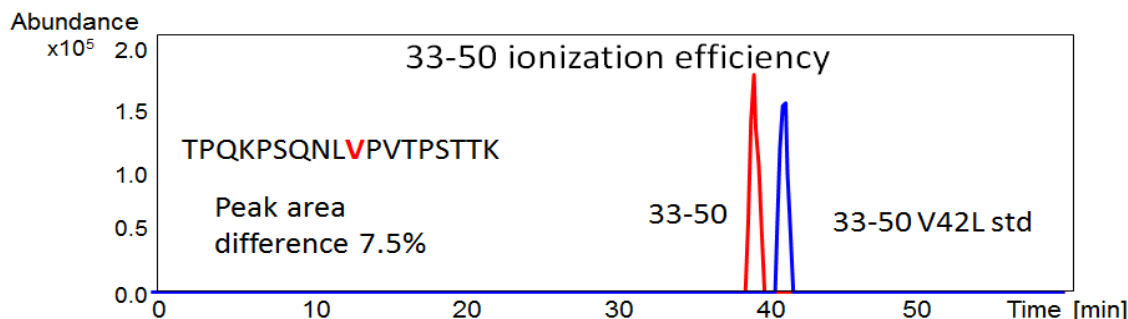
Figure 2.2. Comparable EICs were seen for the 33-50 peptide and the 33-50 V42L substitution internal standard.

Many conventional methods quantify a single peak height or peak area per peptide. However, ultra-high resolution mass spectrometry allows for the resolution of multiple isotopic peaks. Therefore, in this study, peak areas of the first three isotopes are included to improve quantitative accuracy. The first three isotopes represent different percentages of the total number of observable isotopes for each peptide as this number is mass dependent. These percentages were calculated with Protein Prospector's MS Isotope function. For example, a 600 Da peptide's first three isotopes represent ~100% of the total isotope distribution; however, a 2800 Da peptide's first three isotopes only represent ~72% of the total isotope distribution. To calculate a peptide's concentration, the sum of the three peak areas is multiplied by this percentage. This number is then divided by the peptide standard's first three isotopes, with percentage taken into account, and is then multiplied by the standard peptide's concentration. Additionally, all mass spectrometry data were collected in negative ion mode. Although negative ion mode is unusual for protein analysis, phosphorylated peptides and other acidic peptides ionize more readily in negative ion mode than in positive ion mode. Traditionally, a base is added to the spray solvent in negative ion mode to promote deprotonation. In our developed method, 0.1% formic acid was added to the LC solvents, because of the pH limit of the LC column. The pH of 0.1% formic acid in water is about 2.7, whereas acidic side chains with a phosphorylation modification can have a pKa lower than 1.5; therefore, phosphorylated peptides may still

deprotonate in acidic solvents and mobile phases[35]. The advantage of negative ion mode over positive ion mode is shown in Figure 2.3 for one of the peptides of interest. Similar abundance improvement was shown for all Sic1 phosphopeptides. The developed novel method is more accurate than utilizing a single peak area, because it includes more of the data available and thus more accurately represents the data.
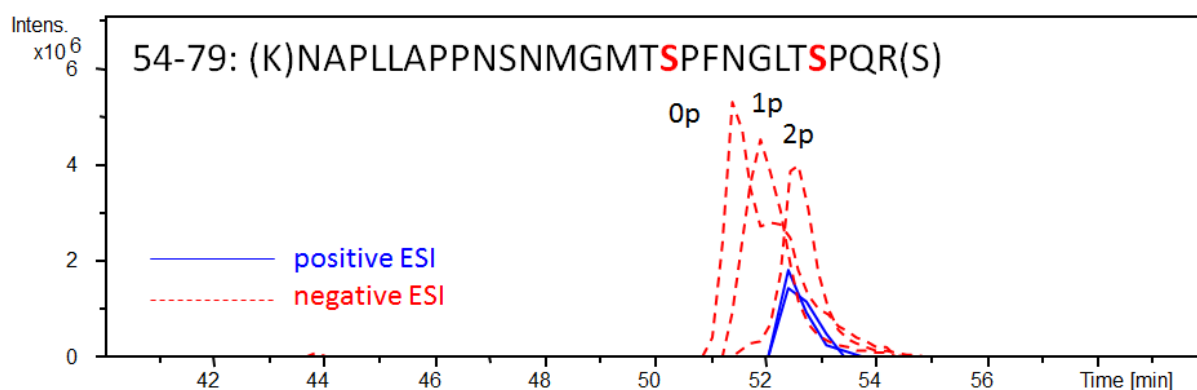
Figure 2.3. EICs from positive (in blue) and negative (in red) ion mode LC/MS of the Sic1 54-79 peptide. The serines in bold and red constitute the phosphorylation sites The doubly phosphorylated (2p) peptide was only detected in negative ion mode and both the unmodified and singly phosphorylated (1p) peptides ionize more efficiently in negative ion mode. Positive ion mode EICs were collected by Dr. Hangtian Song.

2.3.2 Site-specific phosphorylation kinetics for the Sic1 2p mutant

The Sic1 2p mutant has two phosphorylation sites: T45 and S76; the other seven phosphorylation sites are mutated to alanine. The purpose of determining the 2p kinetics was to establish a protocol for quantitative negative ion mode LC/MS. The concentration of each unmodified and phosphorylated peptide was determined over seven timepoints. Figure 2.4 illustrates the kinetic curves for the 33-50 T33A peptide and the 54-79 S69A peptide, respectively.
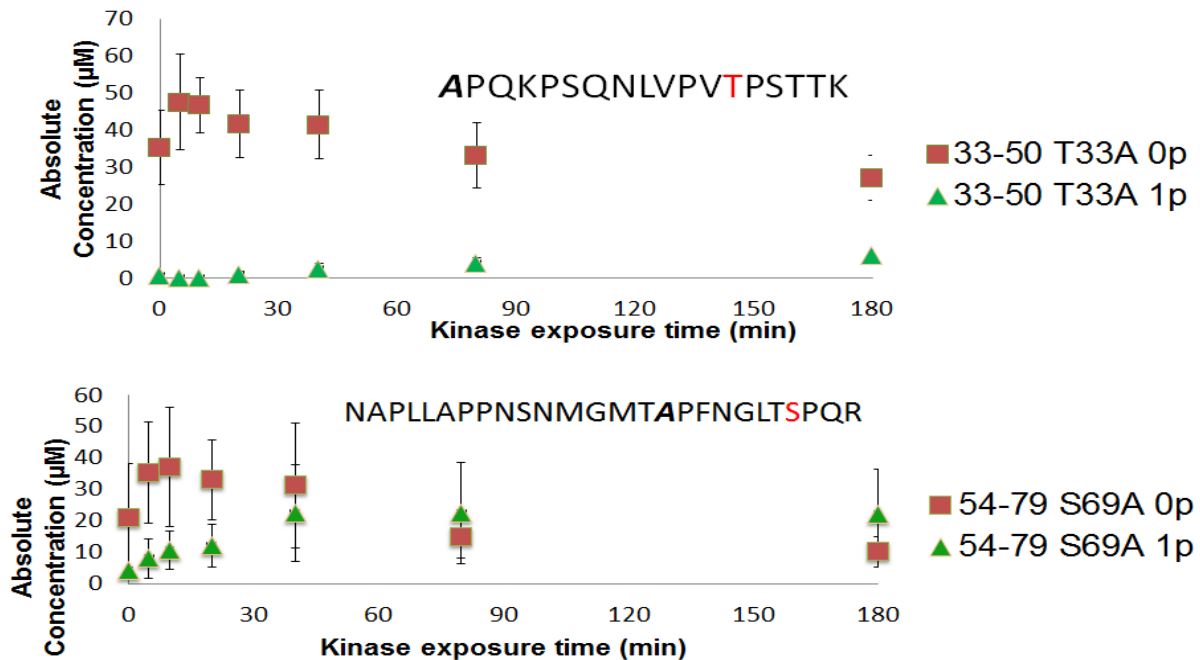
Figure 2.4. Phosphorylation kinetics of the Sic1 2p mutant for the 33-50 peptide (top) and the 54-79 peptide (bottom). The absolute concentration of the unmodified and singly phosphorylated peptides is shown as a function of kinase exposure time. The 2p mutant was not subjected to denaturation prior to trypsin digestion.

Phosphorylation occurs at a faster rate on S76 than on T45. Each phosphorylation event does not occur at the same rate, which was not necessarily expected. A switch-like response could indicate all phosphorylation events occurring simultaneously or occurring at different times, including a chain-like reaction. However, the 2p mutant provides limited data, merely two rates; therefore, it was necessary to collect further data with the Sic1 wildtype protein.

2.3.3 Protein denaturation effects

An unexpected result for the 2p data shown in Fig. 2.4 was the observed initial increase in abundance of the non-phosphorylated peptide. We hypothesized that the trypsin digestion efficiency may be different for the different phosphostates of the protein due to structural changes and differences in flexibility occurring at the onset of multisite phosphorylation[36]: A protein with little or no exposure to kinase may be less flexible with a tighter fold than the same protein with more exposure to kinase, thus influencing trypsin digestion efficiency. Similar kinetic curves were initially also observed for wildtype Sic1. Following these observations, we

experimented with different protein denaturation methods prior to trypsin digestion. A comparison between no denaturation, urea-based denaturation, and heat denaturation prior to trypsin digestion is shown in Figure 2.5 for the 33-50 peptide from the wildtype protein. With denaturation, the overall concentration of the peptides increased, although there was a larger effect for the shorter kinase exposure times.
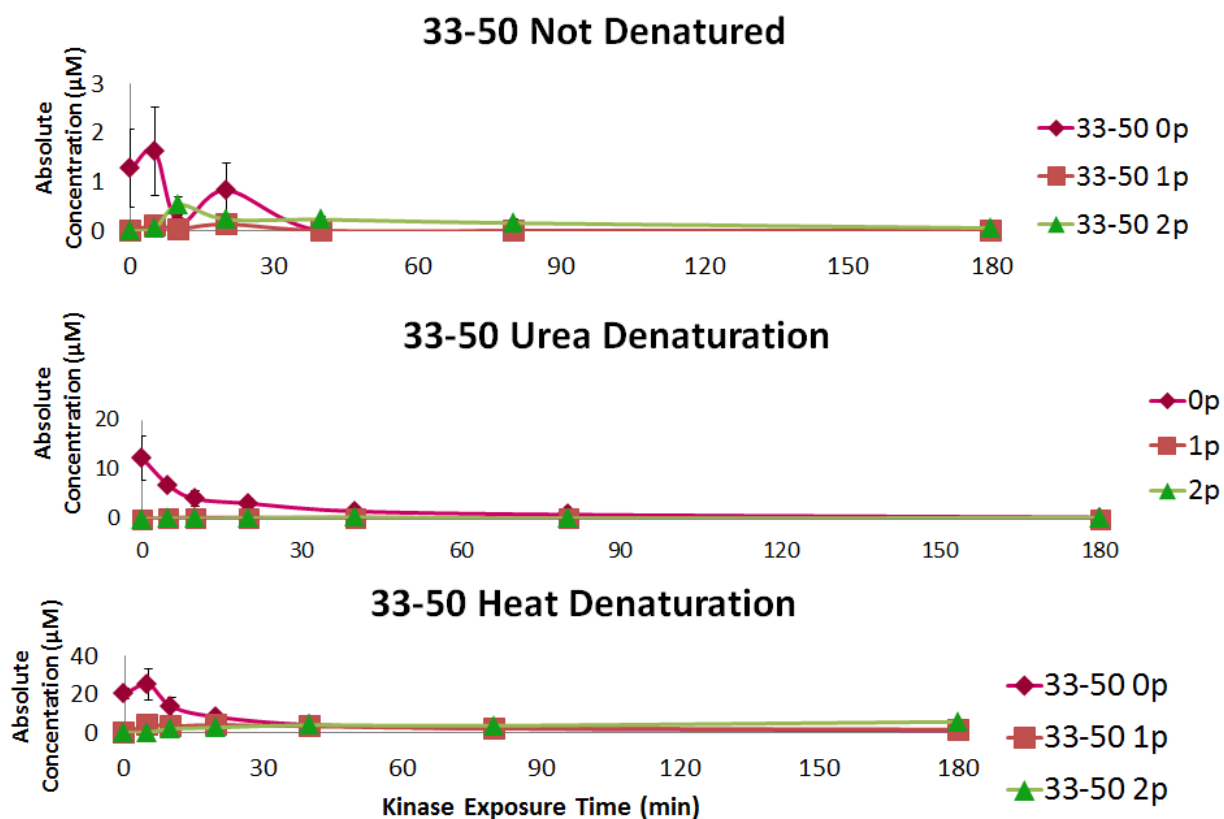
Figure 2.5. A comparison between three protein denaturation options prior to trypsin digestion. Heat denaturation yields a lower error and higher overall concentration of Sic1 peptides.

Denaturation with heat yielded higher peptide concentrations than denaturation with urea. Even after dilution and desalting via LC, the urea concentration is still high, thus lowering peptide signal to noise ratios. Therefore, heat denaturation appears to be the most effective Sic 1 denaturation method and we utilized this approach for the data shown below.

2.3.4 Site-specific wildtype Sic1 phosphorylation kinetics

The rate of phosphorylation for each tryptic phosphopeptide from wt Sic1 was determined in the same manner as the 2p mutant. Interestingly, phosphorylation of peptides with two phosphorylation sites showed faster kinetics than phosphorylation of peptides with only one site. T33 and T45 in the 33-50 peptide both show relatively fast phosphorylation rate and have very similar rates of phosphorylation. For the 1-8 and 54-79 peptides (both with two phosphorylation sites), one site showed faster kinetics than the other. Although it has not yet been determined which site is fast and which is slow, previous studies showed that T5 and S76 contribute more to Sic1's degradation than T2 and S69[10]. An electron capture dissociation (ECD) experiment would be able to determine the identification of the phosphorylation site for the 1p curve for both the 54-79 and 1-8 peptide. The experiment would also be able to determine if it was a mixture of both phosphorylation sites. Fraction collection and ECD have been attempted; however, the peptides of interest were not abundant enough for a successful ECD experiment. S191 in the 186-193 peptide shows a faster phosphorylation rate than the other peptides with one phosphorylation site, but S191 kinetics is still slower than the kinetics of one of the phosphorylation sites in the 33-50 and the 1-8 peptide, both are peptides with two phosphorylation sites. S80 in the 80-84 peptide and T173 in the 169-177 peptide showed low phosphorylation rates. Each kinetic curve is illustrated in Figure 2.6. These results imply that there is a specific order to the phosphorylation and that certain sites are more significant to Sic1's degradation. The 2p mutant and the wt are comparable in that not all sites phosphorylate at the same rate. Although this observation does not prove the hypothesis that a chain-like reaction leads to the degradation of Sic1, it supports the possibility that the phosphorylation of one site triggers the phosphorylation of the next site. Mathematical modeling to fit the kinetic curves will lend further support or negation of this hypothesis. The mathematical modeling will be performed to measure the ultrasensitivity response of Sic1, as multisite phosphorylation may lead to ultrasensitive responses[10,11,37]. An ultrasensitive response in a system is more sensitive to stimuli change than a standard system and is characterized by the Hill coefficient, n_H, producing sigmoidal curves instead of Michaelis-Menten hyperbolic curves. For a standard hyperbolic curve, $n_H = 1$, and the number increases for ultrasensitive responses. Previous data demonstrated that Sic1 binding to Cdc4 (Figure 2.1) results in an ultrasensitive response[8].

Although our wildtype kinetic data have not yet been fitted, mathematical modeling will be performed in the future to further elucidate the switch-like response of Sic1.

33-50 Heat Denaturation
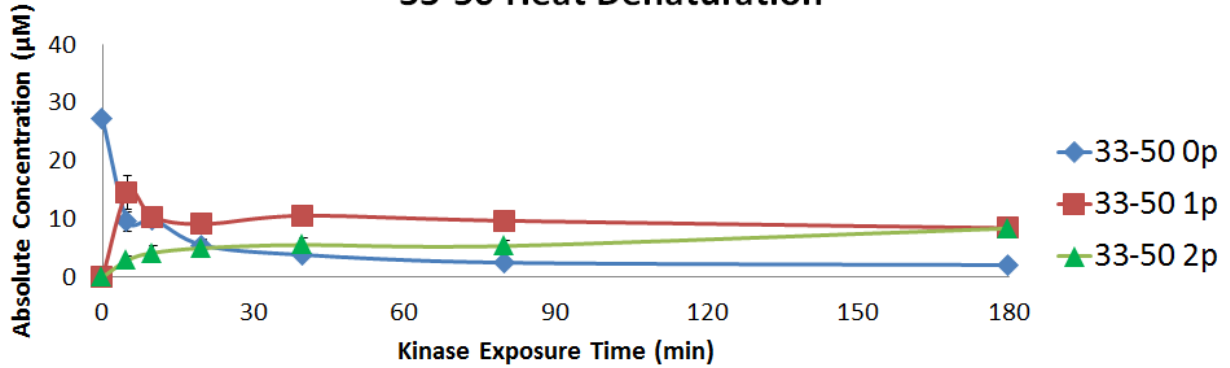
54-79 Heat Denaturation
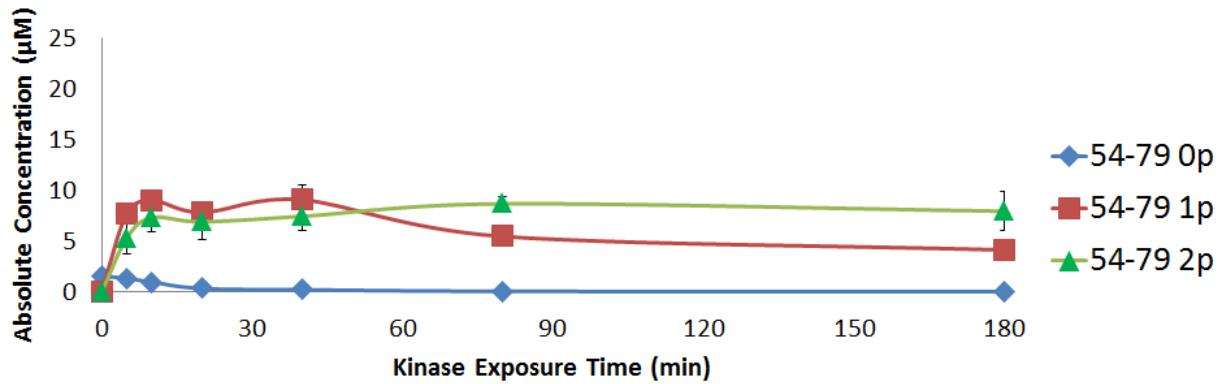
1-8 Heat Denaturation
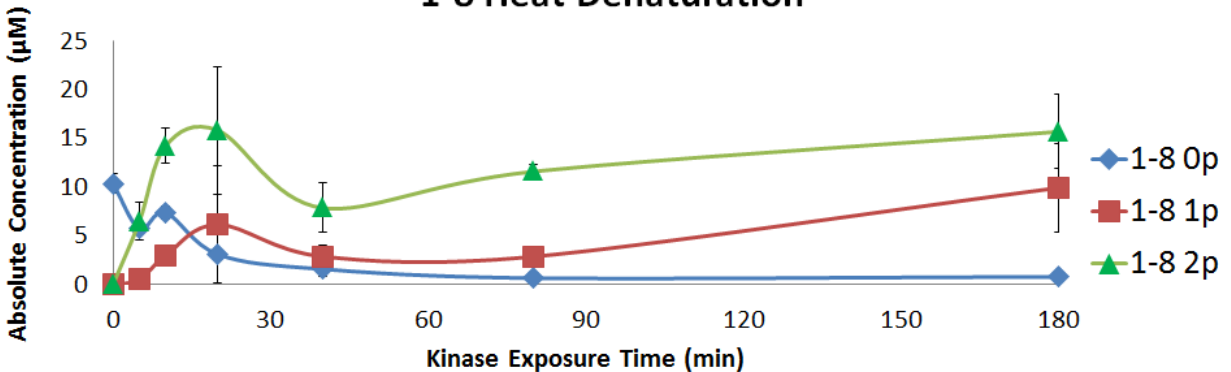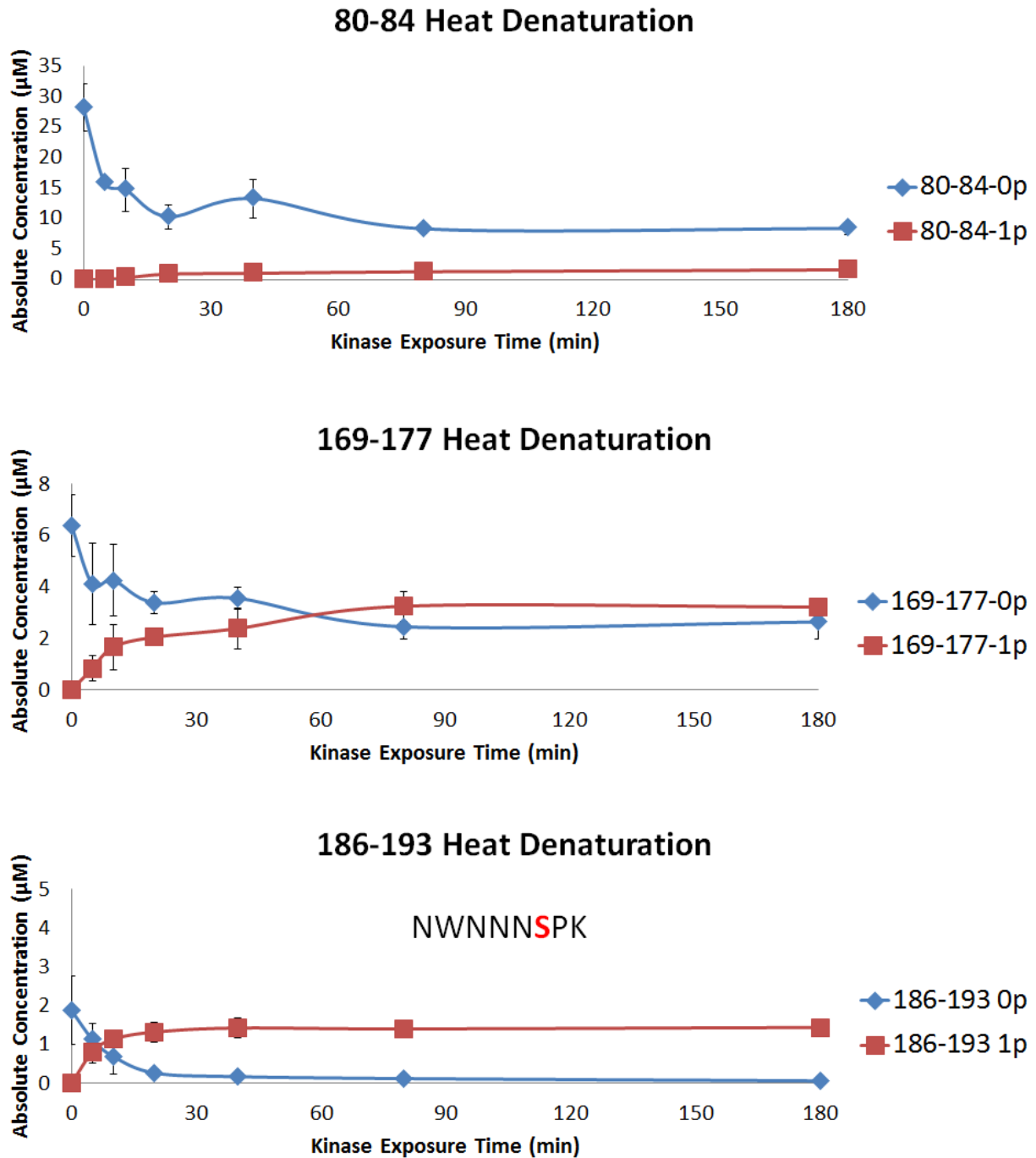
Figure 2.6. Phosphorylation kinetics of all nine phosphorylation sites in wt Sic1. The absolute concentration of the unmodified, singly, and doubly phosphorylated peptides are shown as a function of kinase exposure time.

Another observation to note is that the overall total amount (sum of the amount of non-phosphorylated, singly- and doubly-phosphorylated peptide) for the 54-79 peptide decreases

with kinase exposure time and that there is a dramatic decrease in the total amount of the 33-50 peptide. One hypothesis is that a third, non-specific, phosphorylation event occurs, because both peptides have multiple serine and threonine residues in addition to the two known (proline-directed) sites. A third phosphorylation event was not observed for the 54-79 peptide, thus the overall loss of this peptide is unaccounted for. The LC autosampler was chilled and the time points were collected in a random order, so there is no evidence of peptide degradation with time. A third phosphorylation event was, however, observed for the 33-50 peptide through precise addition of the mass corresponding to HPO_3 (with less than 1 ppm error). The abundance of a species with this specific mass increases with kinase exposure time. The mass of this presumably triply phosphorylated 33-50 peptide does not correspond to any other possible tryptic peptide from Sic1.
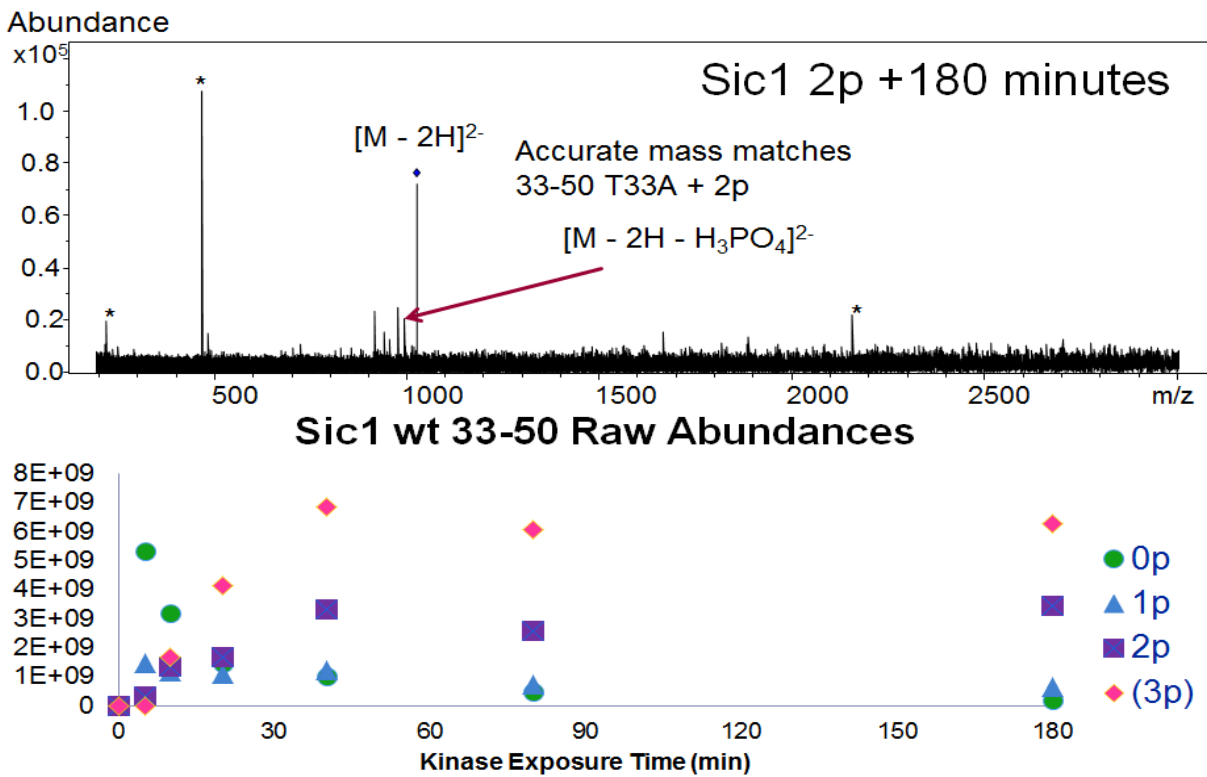
Figure 2.7. Evidence demonstrating an additional, non-specific phosphorylation event on both the wildtype 33-50 peptide (total of three phosphorylations), and the 33-50 2p mutant peptide (total of two phosphorylations). CID MS/MS (top, 2p peptide after 180 minutes of kinase exposure) showed loss of phosphoric acid, as expected. Raw data abundances (without normalizing to internal standard as no 3p standard was available) illustrate how the triply phosphorylated wildtype peptide abundance increases with kinase exposure time (bottom).

Although the low abundance of the putative triply phosphorylated peptide was challenging for tandem mass spectrometry, further identity evidence was provided through the observation of phosphoric acid loss upon collision induced dissociation (CID) for an analogous doubly phosphorylated peptide for the Sic1 2p mutant (with only one phosphorylation site), as seen in Figure 2.7 (top). Lack of additional peptide sequence information may be attributed to CID being performed in negative ion mode, which is known to primarily provide neutral losses and unpredictable sequence information[38]. The putative triply phosphorylated peptide was collected from off-line LC for an ECD experiment. A successful ECD experiment would yield fragments localizing the phosphorylation sites. However, the collected LC fraction was a mixture with additional abundant peptides, rendering the peptide of interest challenging to isolate and fragment. Additionally, the peptide of interest's abundance was too low for a successful ECD experiment.

In order to ensure reproducible results, the entire set of experiments, including protein preparation and phosphorylation, was repeated a second time for the wildtype protein with the same protocol. The protein concentration was higher when measured with the Nanodrop, 0.33 mg/mL to the above data's 0.22 mg/mL, but overall the curves are very similar, demonstrating reproducibility. Notably, throughout both replicates, there are points where the concentrations dip and increase again. Comparing one replicate to another, the dips occur at different points, indicating these unexpected changes in concentration are due to experimental error, perhaps a poor signal to noise ratio.

54-79 Heat Denaturation

1-8 Heat Denaturation

MHHHHHHSSGVDLGTENLYFQSNA**TPST**PPR
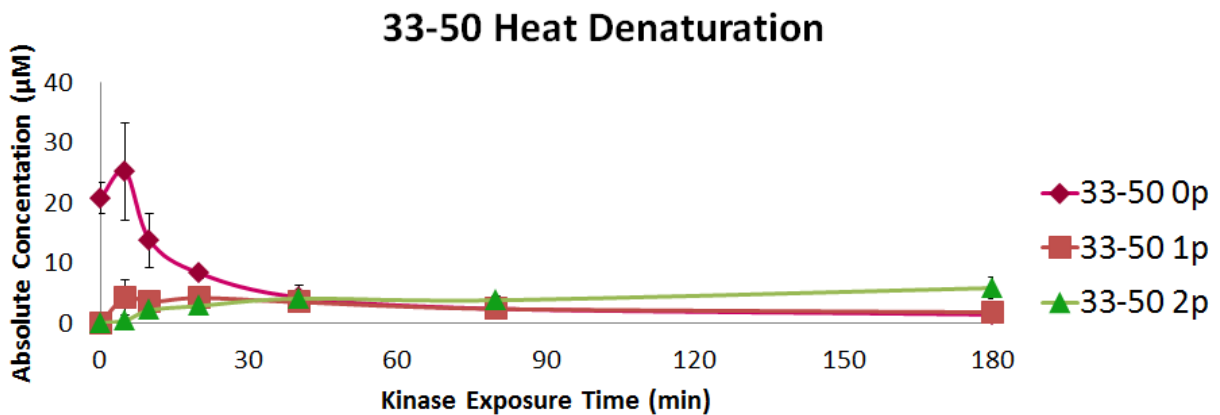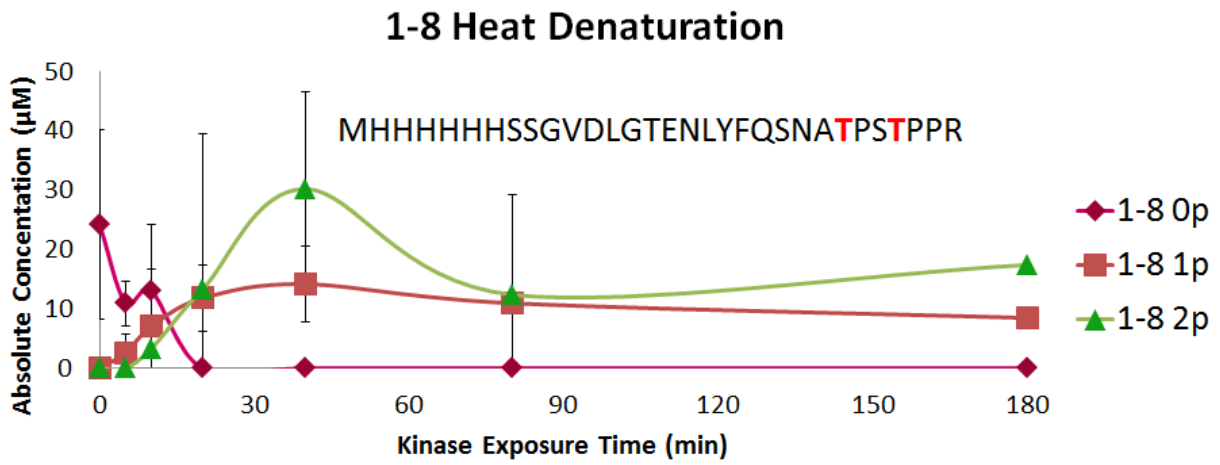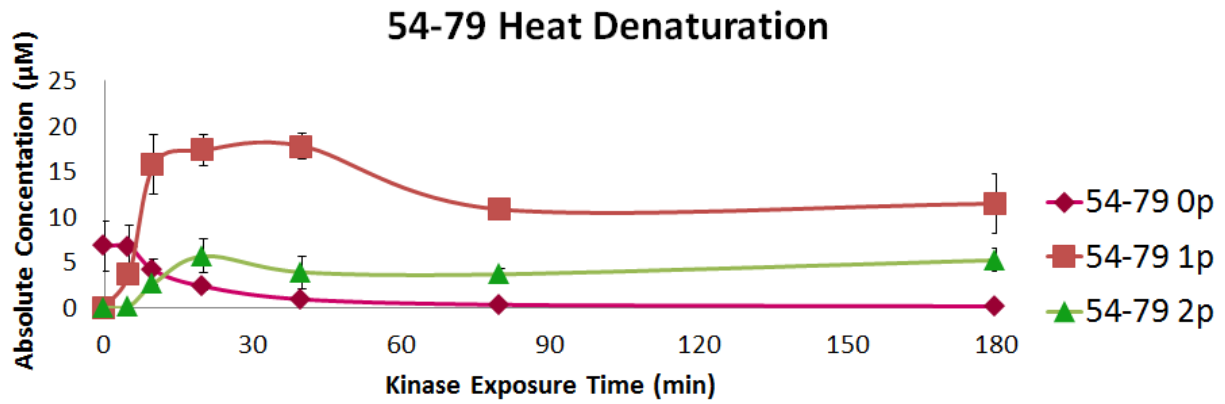
80-84 Heat Denaturation

SPFPK
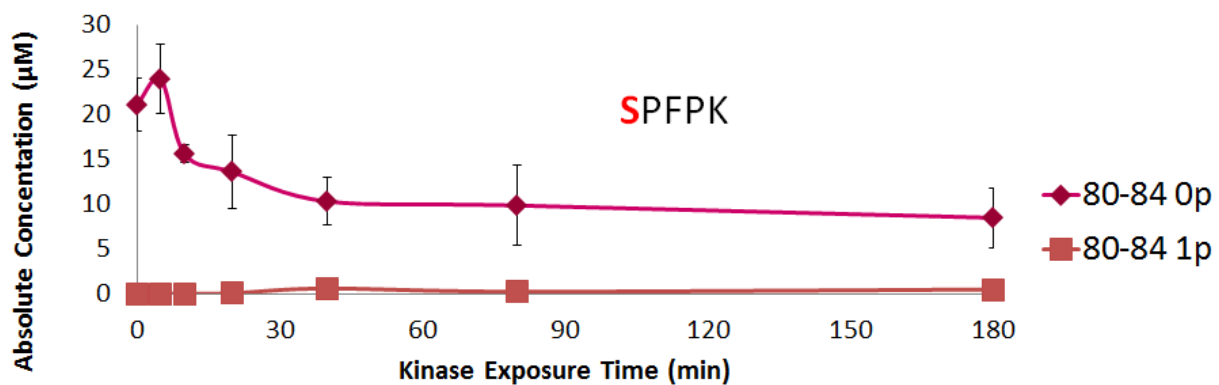
169-177 Heat Denaturation

DVPG**T**PSDK
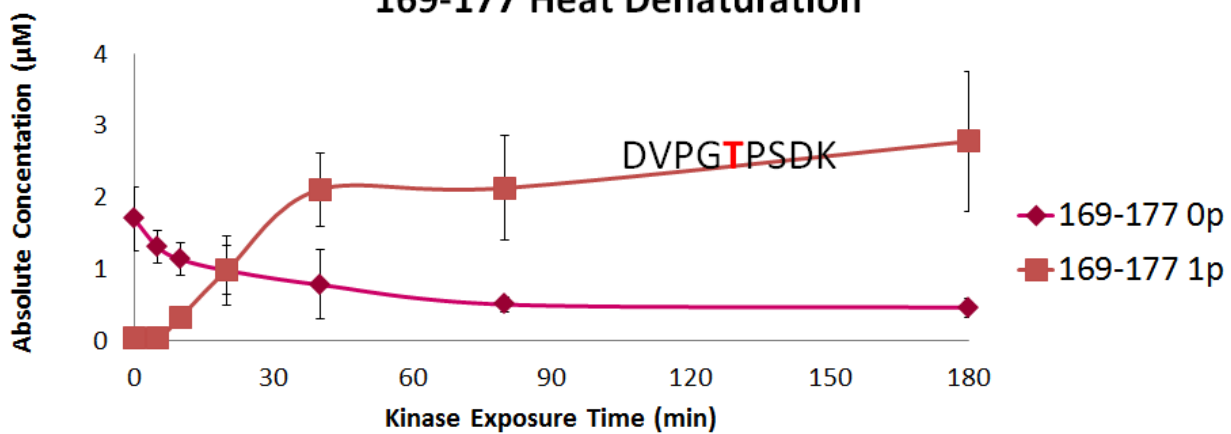
186-193 Heat Denaturation
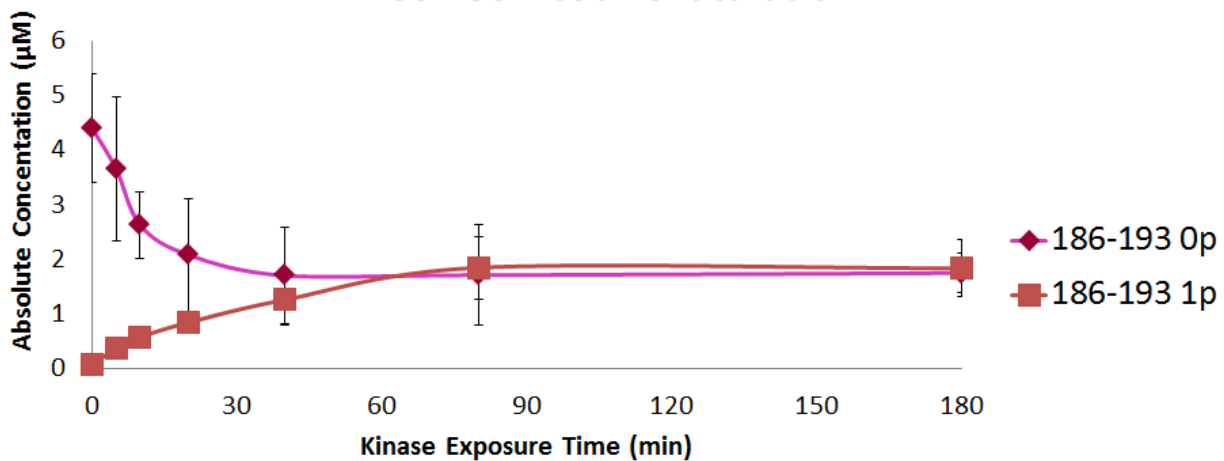
Figure 2.8. A second replicate of the phosphorylation kinetics of all nine phosphorylation sites in wt Sic1. The absolute concentration of the unmodified, singly, and doubly phosphorylated peptides are shown as a function of kinase exposure time.

2.4 Conclusions

A negative ion mode quantitative method was developed to investigate the phosphorylation kinetics and mechanism of the yeast protein Sic1. Valine to leucine or leucine to valine substitutions allowed for a set of internal standard peptides with a 14 Da mass difference without significantly affecting the ionization efficiency of the standard peptides. To establish the quantitative technique, site-specific phosphorylation was quantified for the Sic1 2p mutant and phosphorylation kinetics were established. The two phosphorylation sites in the 2p mutant have very different rates of phosphorylation. The wildtype protein has nine phosphorylation sites; kinetics were determined for each phosphopeptide. Each phosphorylation site showed a different kinetic rate and the fastest rates show good agreement with previous studies[10] that declared T45, S76, T5, and T33 were most significant for Sic1's degradation. Future work includes determining the site of the third putative phosphorylation site on the 33-50 peptide, determining the 1p phosphorylation site(s) on the 54-79 and 1-8 peptides, and comparing the phosphorylation rates with mathematical models. The models will determine the ultrasensitive response of Sic1, which will clarify the mechanism of the switch-like response of this multisite protein.

2.5 References

(1) Blume-Jensen, P.; Hunter, T. Oncogenic kinase signaling. *Nature* **2001**, *411*, 355.

(2) Yan, J. X.; Packer, N. H.; Gooley, A. A.; Williams, K. L. Protein Phosphorylation: technologies for the identification of phosphoamino acids. *J. Chromatogr. A* **1998**, *808*, 23.

(3) Mann, M.; Ong, S.-E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends. Biotechnol.* **2002**, *20*, 261.

(4) Chalmers, M. J.; Quinn, J. P.; Blakney, G. T.; Emmett, M. R.; Mischak, H.; Gaskell, S. J.; Marshall, A. G. Liquid Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometric Characterization of Protein Kinase C Phosphorylation. *J. Proteome Res.* **2003**, *2*, 373.

(5) Krebs, E. G. The growth of research on protein phosphorylation. *Trends Biochem. Sci.* **1994**, *19*, 439.

(6) Faux, M. C.; Scott, J. D. More on target with protein phosphorylation: conferring specificity by location. *Trends Biochem. Sci.* **1996**, *21*, 312.

(7) Hunter, T. Protein Kinases and Phosphatases: The Yin and Yang of Protein Phosphorylation and Signaling. *Cell* **1995**, *80*, 225.

(8) Varedi, K., SM; Song, H.; Hale, W. A.; Hakansson, K.; Lin, X. Multisite Phosphorylation Undergoes Ultrasensitive Binding of Sic1 to Cdc4 in Response to Cln2-Cdc28. *Mol. Cell* **2014**, *In revision.*

(9) Ferrell, J. E. J. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* **1996**, *21*, 460.

(10) Nash, P.; Tang, X.; Orlicky, S.; Chen, Q.; Gertler, F. B.; Mendenhall, M. D.; Sicheri, F.; Pawson, T.; Tyers, M. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **2001**, *414*, 514.

(11) Gunawardena, J. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14617.

(12) Cohen, P. The regulation of protein function by multisite phosphorylation- a 25 year update. *Trends Biochem. Sci.* **2000**, *25*, 596.

(13) Bumpus, S. B.; Evans, B. S.; Thomas, P. M.; Ntai, I.; Kelleher, N. L. A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* **2009**, *27*, 951.

(14) Blackburn, K.; Goshe, M. B. Challenges and strategies for targeted phosphorylation site identification and quantification using mass spectrometry analysis. *Brief. Funct. Genomic. Proteomic.* **2008**, *8*, 90.

(15) Weinreich, M.; Liang, C.; Chen, H.-H.; Stillman, B. Binding of Cyclin-Dependent Kinases to ORC and Cdc6p Regulates the Chromosome Replication Cycle. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11211.

(16) Cross, F. R.; Schroeder, L.; Bean, J. M. Phosphorylation of the Sic1 Inhibitor of B-Type Cyclins in Saccharomyces cerevisiae Is Not Essential but Contributes to Cell Cycle Robustness. *Genetics* **2007**, *176*, 1541.

(17) Schwob, E.; Bohm, T.; Mendenhall, M. D.; Nasmyth, K. The B-Type Cyclin Kinase Inhibitor p40Sic1 Controls the G1 to S Transition in S. cerevisiae. *Cell* **1994**, *79*, 233.

(18)   Verma, R.; Annan, R. S.; Huddleston, M. J.; Carr, S. A.; Reynard, G.; Deshaies, R. J. Phosphorylation of Sic1p by G1 Cdk Required for Its Degradation and Entry into S Phase. *Science* **1997**, *278*, 455.

(19)   Verma, R.; Feldman, R. M. R.; Deshaies, R. J. Sic1 is Ubiquinated In Vitro by a Pathway that Requires CDC4, CDC34, and Cyclin/CDK Activities. *Mol. Biol. Cell* **1997**, *8*, 1427.

(20)   Koivomagi, M.; Valk, E.; Venta, R.; Iofik, A.; Lepiku, M.; Balog, E. R. M.; Rubin, S. M.; Morgan, D. O.; Loog, M. Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase. *Nature* **2011**, *480*, 128.

(21)   Lin, X.; Gao, Y.; Shi, Y.; Church, G. M. Biological switches arising from single protein multisite modifications. **In Revision**.

(22)   Chu, I.; Sun, J.; Arnaout, A.; Kahn, H.; Hanna, W.; Narod, S.; Sun, P.; Tan, C.-K.; Hengst, L.; Slingerland, J. p27 Phosphorylation by Src Regulates Inhibition of Cyclin E-Cdk2. *Cell* **2007**, *128*, 281.

(23)   Nitti, D.; Belluco, C.; Mammano, E.; Marchet, A.; Ambrosi, A.; Mencarelli, R.; Segato, P.; Lise, M. Low Level of p27 (Kip1) Protein Expression in Gastric Adenocarcinoma Is Associated With Disease Progression and Poor Outcome. *J. Surg. Oncol.* **2002**, *81*, 167.

(24)   Vlach, J.; Hennecke, S.; Amati, B. Phosphorylation-dependent degradation of the cyclin-dependent kinase inhibitor p27Kip1. *EMBO J.* **1997**, *16*, 5334.

(25)   Zappacosta, F.; Collingwood, T. S.; Huddleston, M. J.; Annan, R. S. A Quantitative Results-driven Approach to Analyzing Multisite Protein Phosphorylation. *Mol. Cell Proteomics* **2006**, *5*, 2019.

(26)   Chalmers, M. J.; Kolch, W.; Emmett, M. R.; Marshall, A. G.; Mischak, H. Identification and analysis of phosphopeptides. *J. Chromatogr. B* **2004**, *803*, 111.

(27)   Marshall, A. G. Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass Spectrom.* **2000**, *200*, 331.

(28)   Marshall, A. G.; Hendrickson, C. L.; Shi, S. D.-H. Scaling MS Plateaus with High-Resolution FT-ICRMS. *Anal. Chem.* **2002**, *74*, 253A.

(29)   Ewing, N. P.; Cassady, C. J. Dissociation of Multiply Charged Negative Ions for Hirudin (54 – 65), Fibrinopeptide B, and Insulin A (Oxidized). *J. Am. Soc. Mass Spectrom.* **2000**, *12*, 105.

(30)   Janek, K.; Wenschuh, H.; Bienert, M.; Krause, E. Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1593.

(31)   Wilm, M.; Neubauer, G.; Mann, M. Parent Ion Scans of Unseparated Peptide Mixtures. *Anal. Chem.* **1996**, *68*, 527.

(32)   Carr, S. A.; Huddleston, M. J.; Annan, R. S. Selective Detection and Sequencing of Phosphopeptides at the Femtomole Level by Mass Spectrometry. *Anal. Biochem.* **1996**, *239*, 180.

(33)   McAlister, G. C.; Russell, J. D.; Rumachik, N. G.; Hebert, A. S.; Syka, J. E.; Geer, L. Y.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. Analysis of the acidic proteome with negative electron-transfer dissociation mass spectrometry. *Anal. Chem.* **2012**, *84*, 2875.

(34)   Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat. Prod. Rep.* **2009**, *26*, 90.

    (35)   Hung, C.-W.; Kübler, D.; Lehmann, W. D. pI-based phosphopeptide enrichment combined with nanoESI-MS. *Electrophoresis* **2007**, *28*, 2044.

    (36)   Swaminathan, K.; Adamczak, R.; Porollo, A.; Meller, J. Enhanced prediction of conformational flexibility and phosphorylation in proteins. *Adv. Exp. Med. Biol.* **2010**, *680*, 307.

    (37)   Salazar, C.; Brümmer, A.; Alberghina, L.; Höfer, T. Timing control in regulatory networks by multisite protein modifications. *Trends Cell Biol.* **2010**, *20*, 634.

    (38)   Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-induced fragmentations of the (M-H)− parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. *Mass Spectrom. Rev.* **2002**, *21*, 87.

Chapter 3

Improved LC/FT-ICR MS for Interrogation of Natural Product Biosynthetic Enzymes "In Action"

3.1 Introduction

Nature is a rich source of antibiotics, vaccines, anti-cancer drugs, as well as other medicines[1,2]. Many of these natural products are synthesized by protein complexes called polyketide synthases (PKSs). The first modular polyketide synthase gene was sequenced in the early 1990s[3,4]. *Streptomyces venezuelae* is a bacterium found in soil and is known to express multiple PKSs. The pikromycin (PikA) biosynthetic pathway is an example of a type I PKS in *S. venezuelae*. Type I PKSs are modular in construction, whereas type II PKSs are composed of separate domains, and type III PKSs have a similar construction to type II PKSs, except that they are ACP-independent[5,6]. The PikA biosynthetic pathway generates three natural products, 10-deoxymethynolide, narbonolide, and pikromycin, the first macrolide antibiotic to be isolated[7,8]. Although total synthesis of pikromycin is possible, it involves many steps and yields less than 5 mg[9]. Conversion of narbonolide to pikromycin is possible with fewer steps (five compared with over twenty), however, neither product is biologically useful[10,11]. As an alternative to total synthesis for creating biologically relevant compounds, chemoenzymatic synthesis can produce a higher yield and is a simpler process when a biosynthetic pathway is known[12]. The pikromycin system is reportedly the best characterized PKS and is known as a model system since its initial isolation and biochemical characterization of the biosynthetic proteins[13].

The PikA pathway (Figure 3.1) consists of four genes expressed into seven polypeptide modules. Like other type I PKSs, a module minimally consists of three protein domains: a

ketosynthase (KS), an acyl transferase (AT) and an acyl carrier protein (ACP)[14]. The KS domain catalyzes C-C bond formation through a Claisen condensation and results in addition of two or three carbons from malonyl or methylmalonyl (MM), respectively, to the growing natural product intermediate[14,15]. The AT selects either malonyl or methylmalonyl from their corresponding coenzyme As and transfers it to the ACP[14]. The ACP's role is to tether and transfer the natural product intermediate to other domains and modules via a thiol residue on a phosphopantetheine (Ppant) post-translational modification arm[6,14,16]. Structural and mechanistic insights into individual PikA domains have been gained, including the thioesterase[17-19] and the docking domains between the modules PikAIII and PikAIV (Fig. 3.1)[20]. Further, docking domains,[21] ACP,[22] and ketoreductase (KR)[23] domains have been structurally characterized in the 6-deoxyerythronolide B synthase (DEBS), a similar PKS, pathway that produces the antibiotic erythromycin.
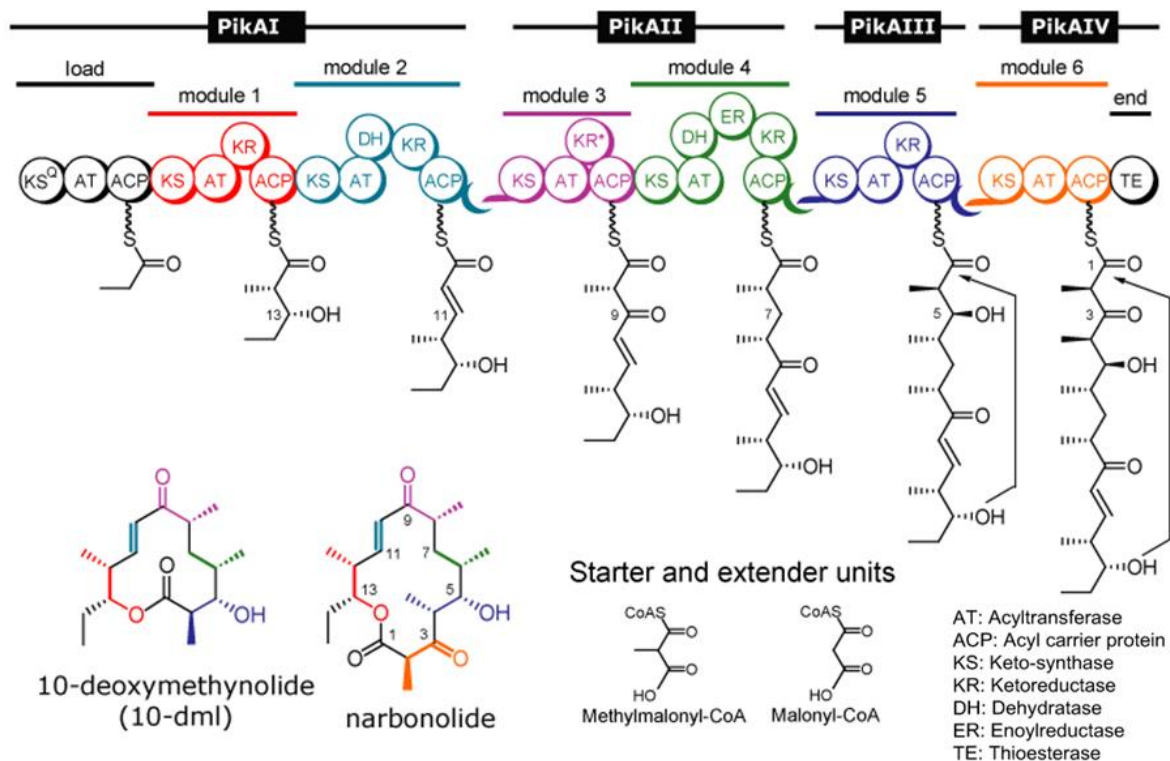
Figure 3.1. The layout of the pikromycin biosynthetic pathway, including 12 and 14-membered products and substrate units[24].

PikA, as well as other biosynthetic pathways, are being explored and investigated for the potential to create novel antibiotics[11]. Although the products naturally generated from the

PikA system are not medically useful in humans, their general structures are an ideal starting point for other useful antibiotics and PikA is unique in that it creates 12 and 14-membered rings[11]. Additionally, the PikA system has significant flexibility in both the substrate choice and enzyme order, including acceptance of domains from other systems such as the DEBS biosynthetic pathway, furthering its utility to generate novel natural products[11,25]. Site-directed mutagenesis in ATs has been shown to alter the AT substrate specificity[26,27]. Further work with mixing and matching DEBS and PikA substrates and domains has provided insight into specificity differences between individual modules[25,28-30]. For example, module five of the PikA system is thought to be a 'gatekeeper' because of its strict specificity, whereas module 6 is much less specific and is thought to be more amenable to the biosynthesis of novel natural products[25,31]. The domain order is very important for the product's structure; therefore, combinatorial biosynthesis with rational design is vital to novel drug design[4,32-36]. Mixing and matching domains and modules from different biosynthetic pathways has resulted in hundreds of novel natural products; however, the failures far outweigh the successes[12,37-39]. Many of these failures arise because certain domains have affinities for other domains, and each domain has a degree of substrate specificity[40-43]. The relationship between each domain that allows for sequential steps in the creation of natural products has been unknown[41]. Further insight into these systems is needed to create more successful antibiotics and other medicines. Determining the structure of PKSs would allow for more thoughtful and intelligent design of substrates and domain and module rearrangement. There have been multiple PKS structural models; however, until recently, the models have not been proven. The head-to-tail and the head-to-head models were based on the mammalian fatty acid synthase (mFAS) structure, in the belief that mFAS had similar architecture to PKSs[44,45]. The double helix model, proposed by Professor Peter Leadlay, challenged the previous models, accounted for the dimeric structure, and placed the ACP in the center, so that it was accessible to all domains[46].
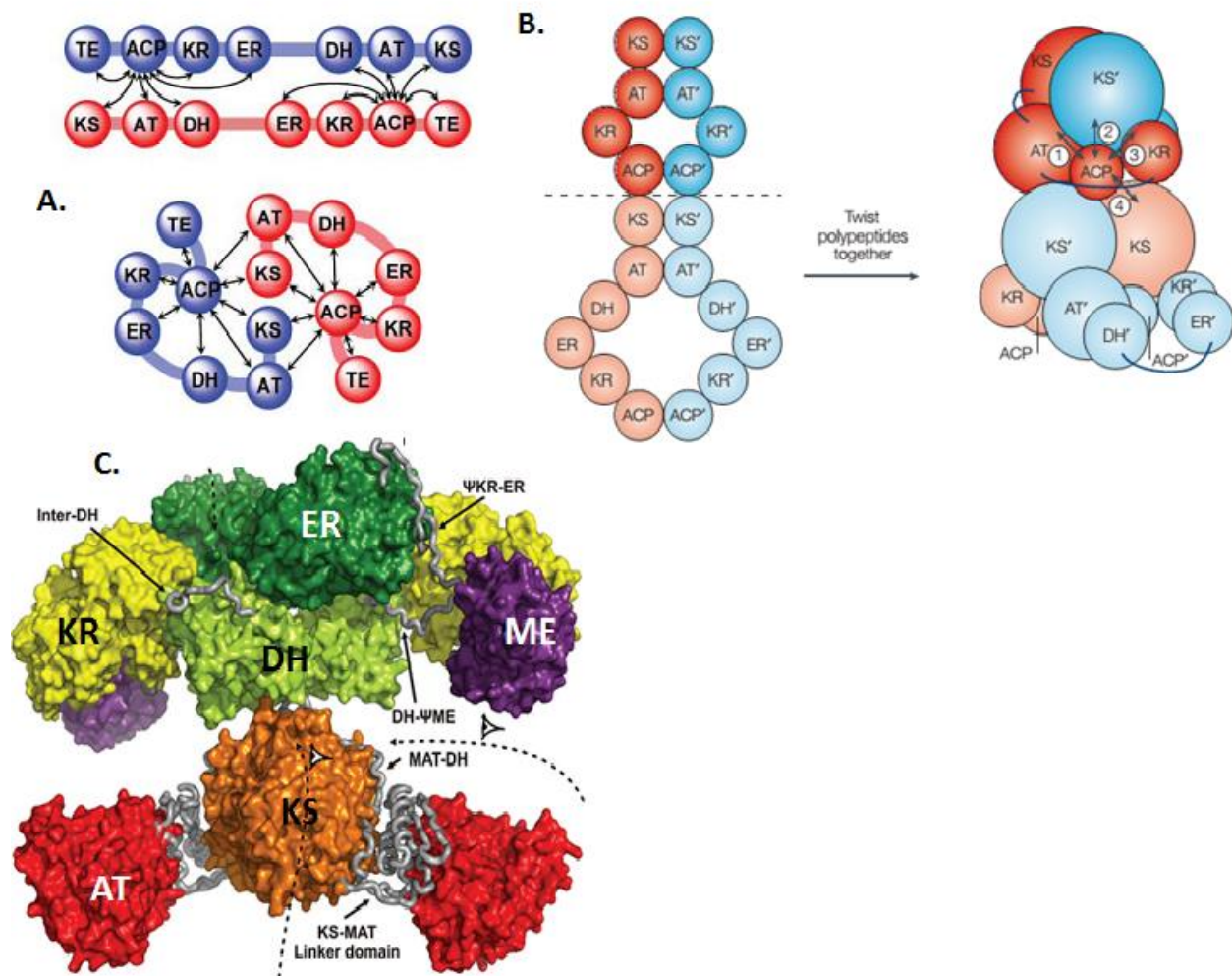
Figure 3.2. Past proposed models of PKSs. **A.** The head-to-tail model and the head-to-head model. **B.** The double helix model. Adapted from Smith and Tsai[44] with permission of The Royal Society of Chemistry. **C.** The mFAS model. Figure adapted from Maier et al[47]. Reprinted with permission from AAAS.

In 2008 the crystal structure of a mFAS was solved with 3.2 Å resolution and was found to resemble an 'X' shape[47,48]. Under the assumption that a fatty acid synthase and PKS had similar structures due to their homology, the model structure for a PKS changed yet again. The first structure of an entire PKS module, pikromycin module 5 (PikAIII), was obtained by our collaborators via high resolution cryo-electron microscopy (cryo-EM). Here, we show Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) to determine and confirm covalent protein states for each structural form found. In order to achieve these results,

improvements were needed to a previous mass spectrometry assay for characterizing PKS active site peptides.

3.2 Experimental

3.2.1 Protein preparation

Proteins were prepared by J. R. Whicher in the following manner: All expression plasmids, except ACP_4–PikAIII(C209A/ΔACP_5) (see section 3.3.3), were expressed in *Escherichia coli* Bap1 cells[49] to produce holo-ACP. ACP_4–PikAIII(C209A/ΔACP_5) was expressed in BL21 (DE3) to produce apo-ACP. Transformed bacteria were cultured at 37 °C to an $OD_{600\,nm} = 1$ in 0.5 L of TB media with 50 µg/ml kanamycin. After incubation at 20 °C for 1 h, cells were induced with 200 µM isopropyl-β-D-thiogalactoside (IPTG) and allowed to express for approximately 18 h.

Cell pellets were re-suspended in 300 mM NaCl, 10% glycerol with either 50 mM HEPES pH 8 (buffer A pH 8; ACP_4–PikAIII(C209A/ΔACP_5)) or 50 mM HEPES pH 7.4 (buffer A pH 7.4; all other constructs) containing 0.1 mg/ml lysozyme, 0.05 mg/ml DNase, 2 mM $MgCl_2$ and 20 mM imidazole. Cells were lysed by sonication, centrifuged, and the supernatant was loaded onto a 5 mL HisTrap column (GE Healthcare). A gradient of 15–300 mM imidazole in buffer A over 10 column volumes was used to elute the proteins.

Proteins were dialyzed into 50 mM HEPES pH 7.4, 100 mM NaCl before substrate loading. The holo-PikAIII and PikAIII(ΔACP) samples were not incubated with substrate. For the MM–PikAIII sample, 1 µM holo-PikAIII was incubated with 500 µM MM–CoA for 30 min at room temperature.

To prepare ACP_4–PikAIII(C209A/ΔACP_5), in which the ACP_4 was loaded with pentaketide, 6 µM apo-ACP_4–PikAIII(C209A/ΔACP_5) from the first gel filtration column was incubated with 100 µM pentaketide–CoA, 10 µM SVP (a non-specific phosphopantetheinyl transferase)[50], 10 µM $MgCl_2$ for 2 h at 30 °C in buffer A pH 8, and re-purified with a second HiPrep 16/60 Sephacryl S300 HR column equilibrated with 50 mM HEPES pH 6.8, 300 mM NaCl, 10% glycerol (buffer A pH 6.8). The peak fraction was collected from the second gel filtration and dialyzed into 50 mM HEPES pH 6.8, 100 mM NaCl.

3.2.2 Mass spectrometry sample preparation

Samples were prepared for MS analysis by Narayan, A. R. H. according to the following procedure: Twenty-five μl of 2 μM holo-PikAIII, 25 μl 2 μM MM–PikAIII, 25 μl 2 μM pentaketide–ACP$_4$–PikAIII(C209A/ΔACP$_5$), and 25 μl 2 μM holo-ACP$_4$–PikAIII(C209A/ΔACP$_5$) were diluted with 20 μl of 250 mM ammonium bicarbonate pH 8.0. Trypsin in 50 mM acetic acid was added in an enzyme: substrate ratio of 1:10. Proteolysis was allowed to proceed for 15 min at 37 °C followed by addition of formic acid (pH 4). Samples were stored at -20 °C until analysis.

3.2.3 LC/FT-ICR MS analysis

Forty-five μl of sample were injected onto a Synergi Hydro C18 hydrophilically end-capped 1 × 150 mm column with 4 μm particles (Phenomenex). A gradient was generated on an Agilent 1100 HPLC. This gradient was as follows (with isocratic elution between 40 and 50 min): 0 (98,2), 20 (70,30), 40 (50,50), 50 (50,50), 55 (30,70), 70 (2,98). Values are provided as time (%A, %B) over a total run time of 90 min. Flow was at 50 μl/min and was diverted for the first 5 min of the run. HPLC solvent A was 0.1% formic acid (ThermoFisher Scientific) in HPLC-grade water (ThermoFisher Scientific), and solvent B was 0.1% formic acid in acetonitrile (ThermoFisher Scientific). The LC was coupled to a quadrupole FTICR-MS (SolariX with 7T magnet, Bruker Daltonics). Data were gathered from *m/z* 200–2,000 in positive ion mode. Electrospray was conducted at 4,500 V with four scans per spectrum and a 256k transient. External ion accumulation in a hexapole was 0.2 s and there was 1 ICR fill before excitation and detection. External calibration used HP-mix (Agilent). PikAIII peptide products were detected over three samples in separate runs.

3.2.4 Cryo-EM preparation and analysis

For details regarding the cryo-electron microscopy preparation and analysis, please see the methods sections of the published papers from where this chapter is derived[51,52].

3.3 Results/Discussion

3.3.1 Cryo-EM details and LC/FT-ICR MS improvements

Single particle cryo-EM was applied to determine the structure of the entire PikAIII module. Cryo-EM determined the three dimensional structure of different physiological states

while FT-ICR MS determined and confirmed covalent protein states. Traditional electron microscopy suffers from poor signal to noise ratio, due to bond damage and formation of free radicals[53]. Cryo-EM reduces damage caused to the sample by six-fold, due to the low temperatures and the glass-like ice layer that forms on the sample[53]. Because the cryogenic temperatures freeze the sample into place, cryo-EM is a promising alternative technique to X-ray crystallography, which is unable to resolve proteins that have flexible areas. Single particle cryo-EM collects two-dimensional images of the sample at different orientations which are combined into a three-dimensional reconstruction of the sample's structure[53]. If an atomic model of the sample is available, it may be fitted into the density map, which greatly expands the structural information[53]. The resolution of the PikAIII samples ranged from 7.3-11.6 Å, which is very high for cryo-EM. FT-ICR MS is known for its ultra-high resolution and high mass accuracy, an ideal instrument for confirming each protein state. PikAIII was the analyte of choice, because it can be produced readily in a pure form and our collaborators have expertise in the synthesis of its substrate intermediate[54].

Previous research in the Håkansson lab has involved liquid chromatography (LC)/FT-ICR MS of natural product biosynthetic pathways[42], however improvements to the previous LC method were necessary to confirm each PikAIII active site. Table 3.1 summarizes these improvements. Previously, all cysteines were reduced and alkylated prior to digestion. In the current protocol, the cysteines are left unmodified, because the ketosynthase domain's active site peptide binds to its substrate via a cysteine. The trypsin digest time was reduced to minimize chances of substrates hydrolyzing from their active sites. An LC column more suitable for peptides was employed, with smaller particles and diameter, and with C18 particles rather than C4 particles. Accordingly, the LC column's size resulted in the selection of a lower flow rate. With the slower flow, a longer LC gradient was chosen. Additionally, a linear gradient resulted in co-elution of many peptides, so a step-wise gradient was developed.

Protocol Differences	Previous Protocol	Current Protocol
Cysteine chemistry	Reduced and alkylated	Unmodified
Trypsin digest time	15-45 minutes	15 minutes
Column	C4 2 x 250 mm 300 μm particles	Hydrophilically endcapped C18 1 x 50 mm 4 μm particles
Flow Rate	200 μl/min	50 μl/min

LC gradient	Linear, 40 minutes	Stepwise, 90 minutes

Table 3.1. A summary of the implemented sample preparation and LC improvements optimized for maximum active site peptide determination.

3.3.2 Holo vs. apo PikAIII structure determination and LC/FT-ICR MS confirmation

The overall structure of PikAIII, as determined with cryo-EM, is arch-like, with the KS at the top, followed by the AT domains on either side, with the KR domains at the bottom. The overall structure was unexpected, as the expected PikAIII structure was more linear, analogous to the mFAS structure in Figure 3.2C. Instead, the AT domains are rotated 120 degrees, relative to the mFAS crystal structure and there is a single active site chamber in PikAIII, although there are two in the mFAS structure. The cryo-EM data also revealed the existence of two separate conformers with approximately equal amounts of each; one in which the ACP is near the KS and AT domains, and another in which the ACP is at the bottom of the arch, near the KR domains. One hypothesis explaining the two conformers is that one has an apo ACP and one has a holo ACP, i.e., without and with the Ppant modification, respectively. However, cryo-EM data alone cannot confirm or deny this hypothesis. LC/FT-ICR MS (Fig. 3.4) determined that the relative amounts of apo and holo peptide were about 3% and 97% respectively, refuting the hypothesis. Alternatively, the two different conformers provide evidence that the ACP moves around the overall architecture.

Holo-ACP conformer I   Holo-ACP conformer II
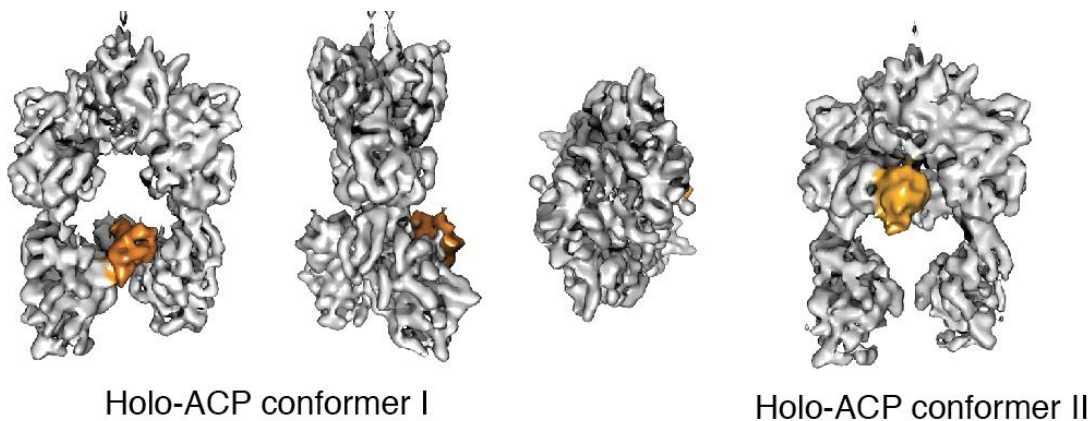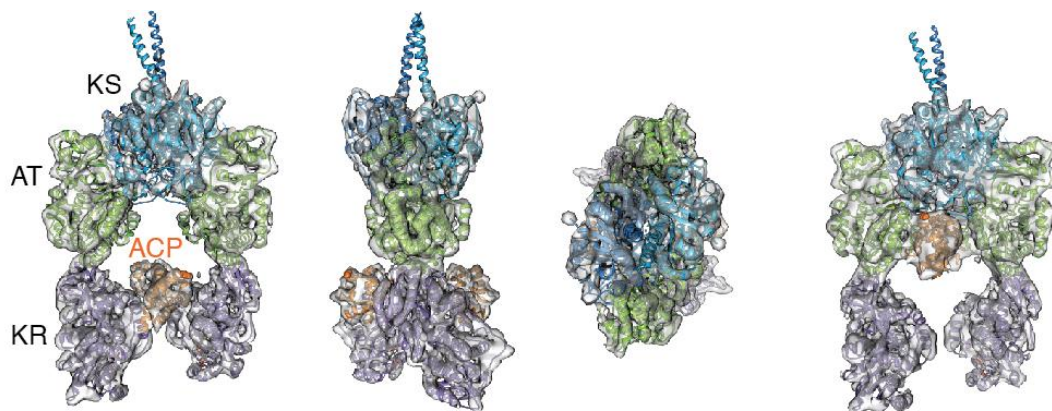
Figure 3.3. Top: Cryo-EM images of holo PikAIII. The ACPs are shown in orange. Bottom: Reconstructed crystal structures of holo PikAIII using DEBS module 5 KS, module 5 AT, module 1 KR, and an NMR structure of DEBS module 2 ACP. There are two completely different conformers, with the left one appearing in about 57% of particle projections, whereas the conformer to the right appeared in about 43% of particle projections. Data generated by S. Dutta.
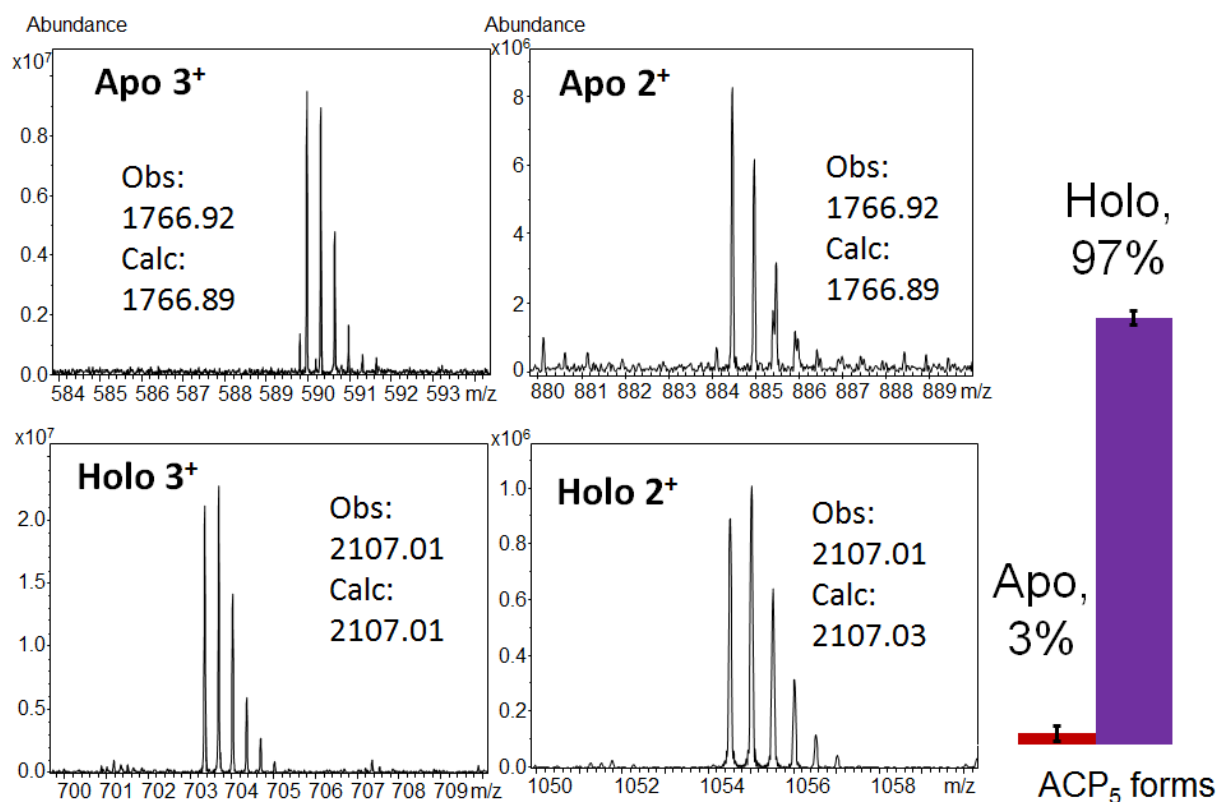
Figure 3.4. Partial mass spectra of ACP_5 active site peptides from LC/FT-ICR MS of trypsin digested PikAIII. Both apo and holo ACP_5 peptides were observed in two different charge states. Relative quantitation was performed by integrating the peak areas of all isotopes with a signal to noise ratio of three or higher.

3.3.3 PikAIIIΔACP$_5$ and ACP$_4$-PikAIII(C209A/ΔACP$_5$) structure determination and LC/FT-ICR MS confirmation

To examine how an upstream ACP interacts with its downstream module, a PikAIII construct lacking its ACP (PikAIIIΔACP$_5$) was first expressed. The overall structure was comparable to the holo PikAIII conformers shown above; however, the KR domains were rotated by 165°, indicating that ACP$_5$ has a strong influence on the structure and orientation of the module. Another construct was then expressed in which ACP$_4$ from the preceding PikAII (module 4) was fused to the N-terminus of PikAIIIΔACP$_5$ via a flexible linker. The KS active site cysteine was also mutated to an alanine, to prevent immediate transfer of the PikAII-generated pentaketide intermediate, in vitro loaded to ACP$_4$, to the KS. For this construct, the ACP$_4$ was located outside of the arch, at the top, bound to the KS, as shown in Figure 3.5. LC/FT-ICR MS (Figure 3.6) confirmed that the mutated active site KS peptide was unmodified and that the ACP$_4$ active site was loaded with pentaketide. The location of ACP$_4$ suggests that the pentaketide intermediate strongly drives the docking of ACP$_4$ at the KS active site entrance.
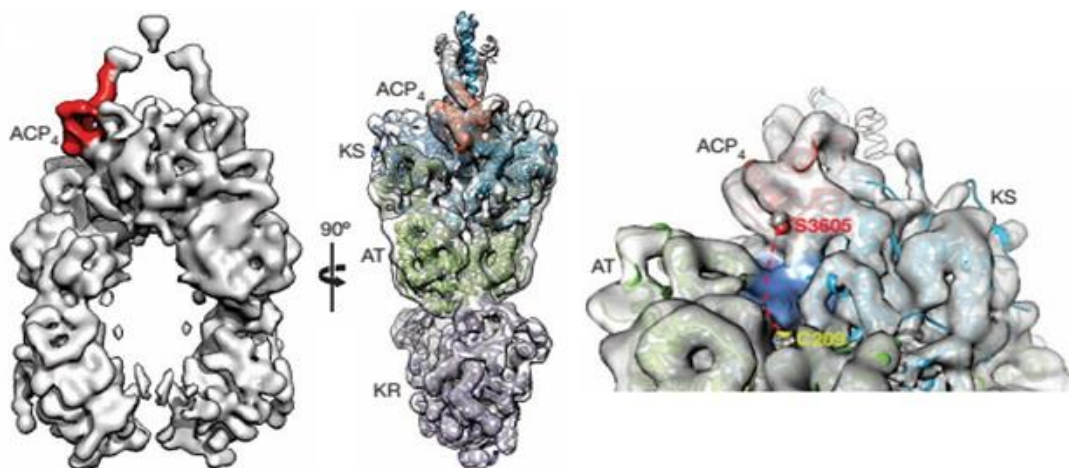
Figure 3.5. Cryo-EM structure of ACP$_4$-PikAIII(C209A/ΔACP$_5$). Note that the KR domains are rotated inward 165° compared with holo PikAIII (Fig. 3.3). ACP$_4$ (displayed in red) is outside of the arch, bound to the KS. The figure to the right shows the close proximity of the KS and ACP$_4$ active sites, about 28 Å apart. Data generated by S. Dutta.
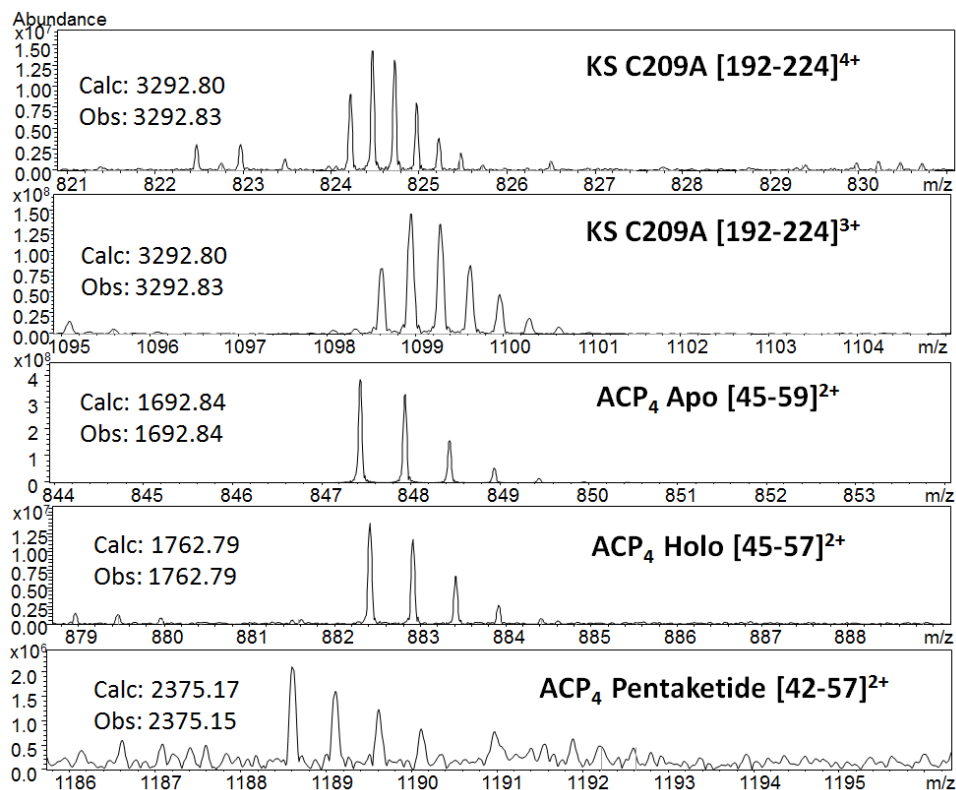
Figure 3.6. Partial mass spectra of ACP$_4$-PikAIII(C209A/ΔACP$_5$) active-site peptides from LC/FT-ICR MS of trypsin digested protein. Active site peptides include KS C209A and ACP$_4$ in apo, holo, and pentaketide- loaded states.

3.3.4 MM-loaded PikAIII structure determination and LC/FT-ICR MS confirmation

The next step was to explore how ACP$_5$ interacts with the KS active site when the extender unit, methylmalonyl is introduced. For this purpose, holo PikAIII was incubated with 500-fold excess MM. LC/FT-ICR MS unambiguously confirmed that both the AT and ACP active sites were loaded with MM (Figure 3.7). Cryo-EM data showed that the overall architecture of MM-loaded PikAIII is similar to holo PikAIII, however, ACP$_5$ was only observed in one location in the MM-loaded protein, below the KS and next to the AT domains (Figure 3.8). This new location provides evidence that ACP$_5$'s location is based upon its attached substrate. Although the KS is not involved in MM loading, the KS and ACP$_5$ active sites are only 25 Å away from each other following MM loading, indicating the KS is ready to play its role. This structure revealed a second active site entrance at the bottom of the KS. Compared with the crystal structure of the DEBS KS-AT didomain, the loop surrounding the active site entrance has very low sequence conservation. Furthermore, the mFAS has a single active site entrance, so a second entrance

75

was unexpected. The second active site entrance allows for shuttling of substrates to and from the KS and may also allow for domain or module 'skipping', as seen in the Pik and DEBS systems[55,56].
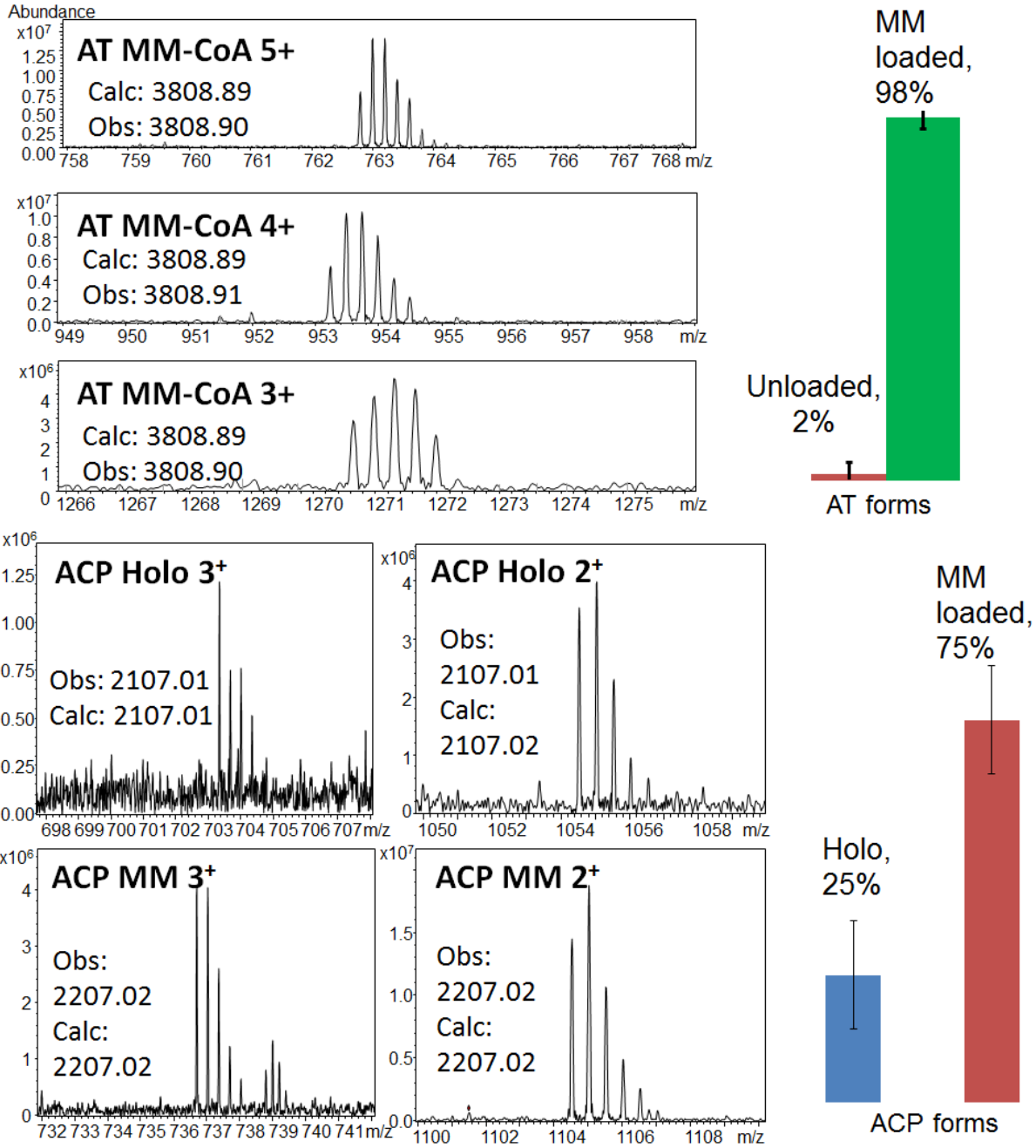
Figure 3.7. Partial mass spectra of AT and ACP active-site peptides from LC/FT-ICR MS of trypsin digested MM-loaded PikAIII. The AT domain is fully loaded with MM, whereas the ACP active site peptide was found both in holo and MM-loaded states.
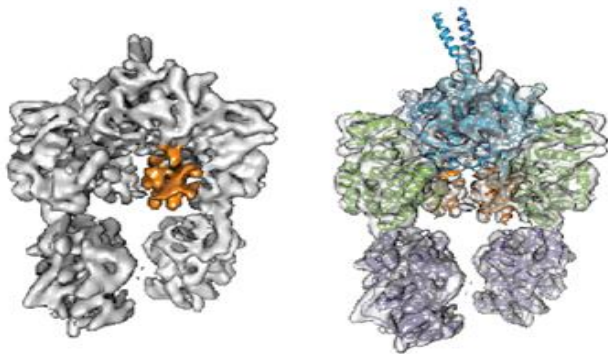
Figure 3.8. Cryo-EM and reconstructed X-ray crystallography/NMR images of MM-loaded PikAIII. Here, a single conformer exists, with the MM-loaded ACP between the AT and KS domains. Data generated by S. Dutta.

3.3.5 Pentaketide-loaded PikAIII structure determination and LC/FT-ICR MS confirmation

In the assembly line-like pikromycin biosynthetic process, there are two potential first steps: loading of the MM extender unit, as described above, or loading of the upstream intermediate, in this case, a pentaketide. When thiophenol-pentaketide was incubated with holo PikAIII, the overall structure showed multiple conformational changes as compared to holo PikAIII (Figure 3.9). The AT domains lean in toward the KS domains, partially blocking the KS active site. This adjustment may be necessary to protect the labile pentaketide until the MM extender unit is added. A more dramatic structural change is the end-to-end flip of the KR domains. As a result of this flip, the KR active sites become closer to the AT domain active sites. Mutagenesis showed that the proximity of the active sites is important to catalytic function. These data additionally demonstrate that the position and orientation of the KR domains are dependent on the other domains and reveal evidence of interdomain cross-talk. The mass spectrometry data confirmed that the KS domain was loaded with pentaketide (Figure. 3.10). Previous LC/FT-ICR MS work carried out with the PikA system was unable to detect any substrate bound to the pentaketide[54], therefore, these mass spectrometry data were particularly exciting and insightful[42]. Successful detection of pentaketide-loaded KS was very likely due to the LC/FT-ICR MS protocol changes and improvements.

77
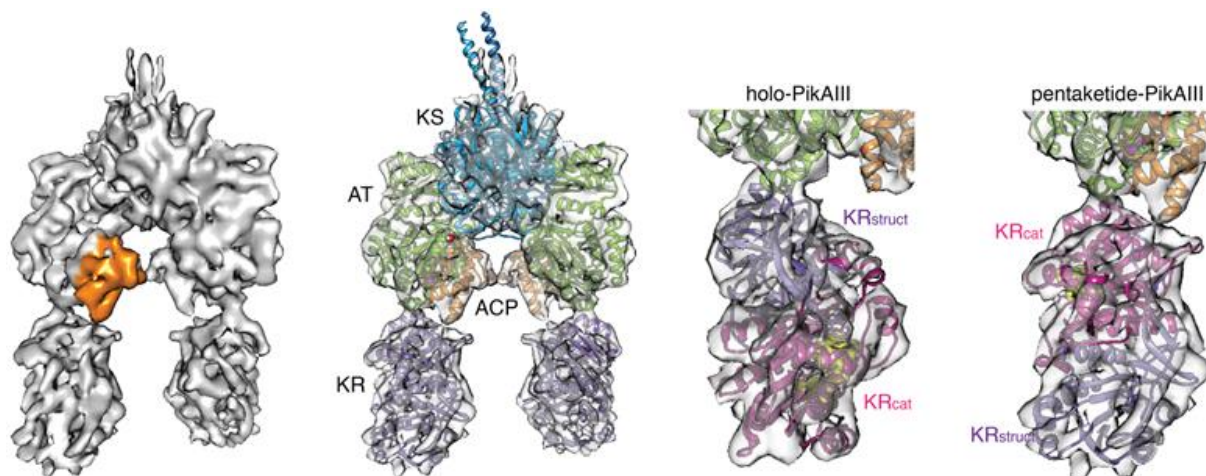
Figure 3.9. Cryo-EM and X-ray crystallography/NMR reconstructions of pentaketide-loaded holo PikAIII. The ACP shifts down and exists as a single conformer, compared with holo PikAIII. The most dramatic change is displayed in the two right-most panels, an end-to-end flip of the KR domain. Data generated by S. Dutta.
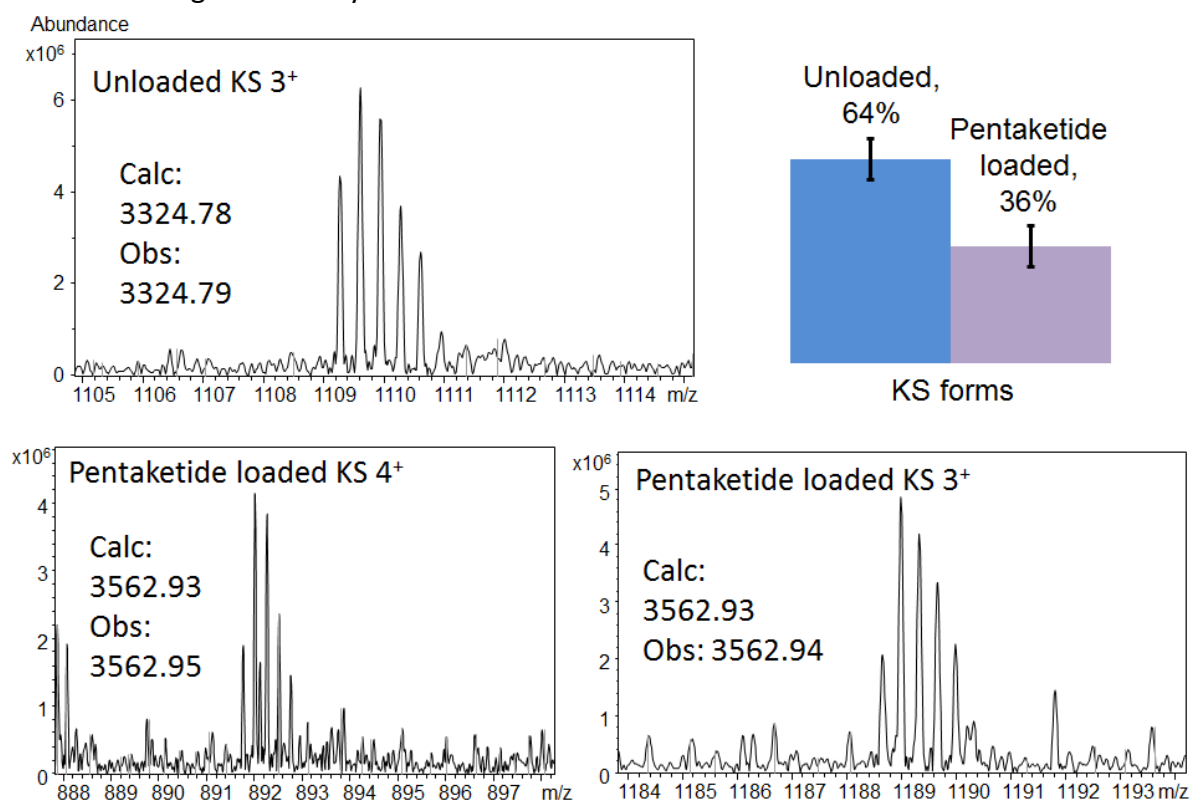
Figure 3.10. Partial mass spectra of KS active-site peptides from LC/FT-ICR MS of trypsin digested pentaketide-loaded PikAIII. Although only partial loading of the KS active site was observed, for the first time, there is mass spectral evidence of pentaketide loading on the KS.

3.3.6 MM- and pentaketide-loaded PikAIII structure determination and LC/FT-ICR MS confirmation

The condensation of MM and the pentaketide intermediate from PikAII forms a β-ketohexaketide, the first intermediate in PikAIII catalysis. The overall structure of MM and pentaketide-loaded PikAIII (Figure 3.11) is very similar to the architecture of the pentaketide loaded structure (Fig. 3.9). The AT domains are shifted even further toward the KS, blocking the side entrance to the KS active site. This shift likely prevents the immediate transfer of the pentaketide from ACP_4 to the KS active site. Although the ACP is in the same overall position, its active site points toward the NADPH binding site in the KR domain, which is involved in the subsequent catalytic step (discussed below). LC/FT-ICR MS was able to determine each domain's active site peptides and bound substrates/products (Figure 3.12). The KS was partially loaded with pentaketide, the AT was fully loaded with MM, and the ACP showed evidence of both MM and β-ketohexaketide product bound to the Ppant arm. Although each active site peptide was of low abundance, rendering tandem mass spectrometry a significant challenge, the nearly perfect mass accuracy confidently confirmed the identity of each peptide, substrate, and product.
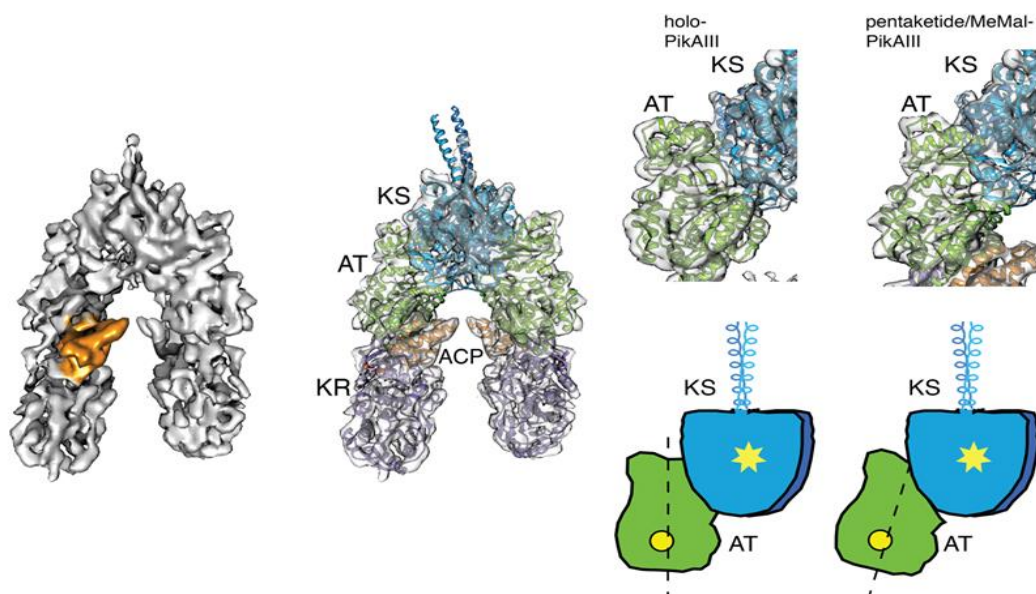
Figure 3.11. Cryo-EM and X-ray crystallography/NMR reconstructions of MM- and pentaketide-loaded PikAIII, yielding β-ketohexaketide product on the ACP. The overall architecture is similar to pentaketide-loaded PikAIII, except for the AT domains shifting toward the KS domain more dramatically, as demonstrated in the cartoon. Data generated by S. Dutta and J. Whicher.
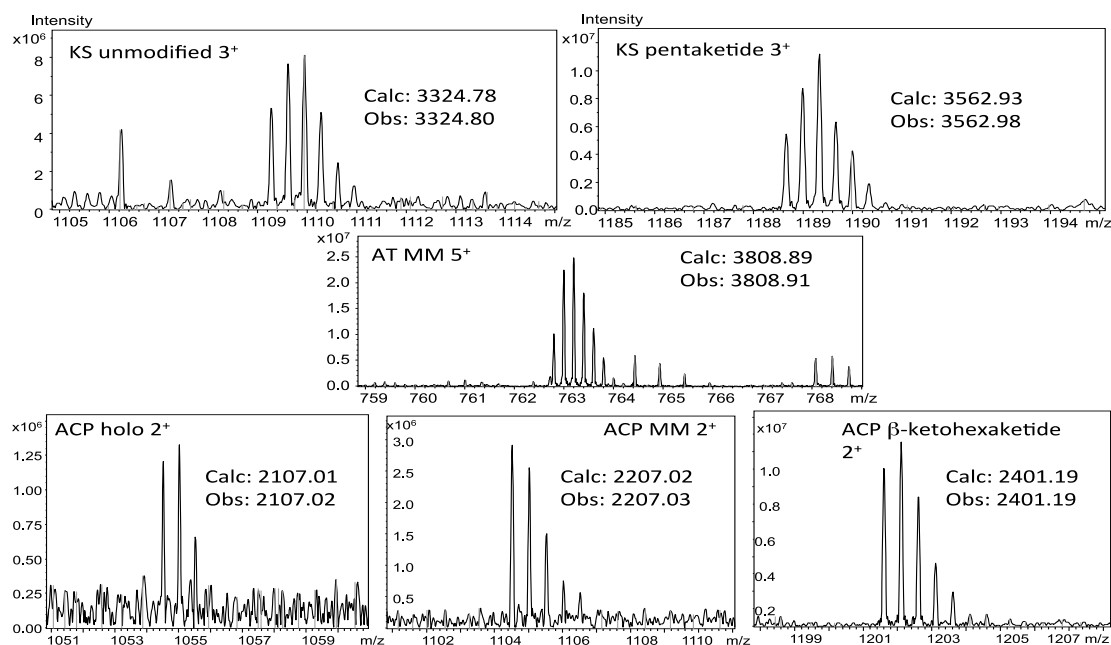
Figure 3.12. Partial mass spectra of active-site PikAIII peptides from LC/FT-ICR MS of trypsin digested MM- and pentaketide-loaded PikAIII. Pentaketide was detected on the KS, MM was detected on the AT, and both β-ketohexaketide product and MM were detected on the holo ACP.

3.3.7 MM- and pentaketide-loaded PikAIII structure determination and LC/FT-ICR MS confirmation in the presence of NADPH

The final PikAIII catalytic step is β-ketohexaketide reduction by the KR domain, resulting in a mass shift of 2 Da to yield β-hydroxyhexaketide. This reaction requires NADPH to power the KR. An initial experiment with 4 mM NADPH did not result in LC/FT-ICR MS detection of β-hydroxyhexaketide. However, both 16 mM of NADPH and 4 mM of NADPH buffered in HEPES yielded the expected reduction. High resolution mass spectrometry was vital for this part of the experiment, because the detected peptides had a 2+ charge, resulting in an m/z change of only 1. Moreover, there was a mixture of β-ketohexaketide, and β–hydroxyhexaketide, resulting in a broad isotopic pattern (Figure 3.13). Based on the isotope pattern, it appears that two ketones were reduced, entirely possible due to the excess NADPH added.

80
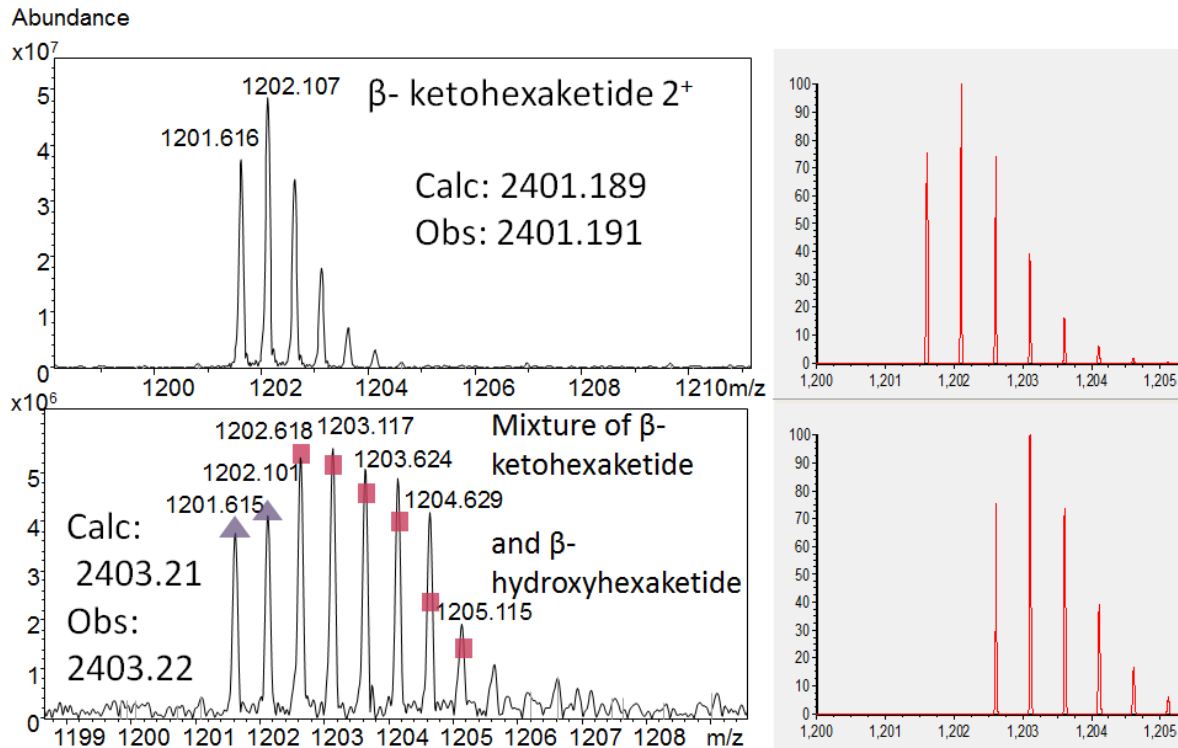
Figure 3.13. Partial mass spectra of the ACP_5 active-site peptide from LC/FT-ICR MS of trypsin digested MM- and pentaketide-loaded PikAIII in the absence (top) and presence (bottom) of NADPH. Calculated isotope patterns are shown on the right for β-ketohexaketide and β-hydroxyhexaketide, respectively, bound to the active site peptide. The bottom mass spectrum shows evidence of a mixture of these two hexaketides. The purple triangles indicate the expected abundancies of the two first isotopic peaks of the β-ketohexaketide and match well with the experimental data. The pink squares indicate the expected isotopic abundancies for a 1:1 mixture of β-ketohexaketide and β-hydroxyhexaketide. The poorer fit at higher m/z ratios indicate that a second reduction may have occurred due to excess NADPH.

In the presence of NADPH, the overall structure of MM- and pentaketide-loaded PikAIII is similar to the structure in the absence of NADPH, with the AT domains leaning in toward the KS domain and the KR domain's active site directed toward the AT. However, three different conformers were observed, each with the ACP located outside of the catalytic chamber, under the KR domains (Figure 3.14). Although there are three different positions, the ACP has little flexibility under these conditions. The ACP active site is directed away from PikAIII and is prepared to transfer the substrate to the next module, PikAIV.
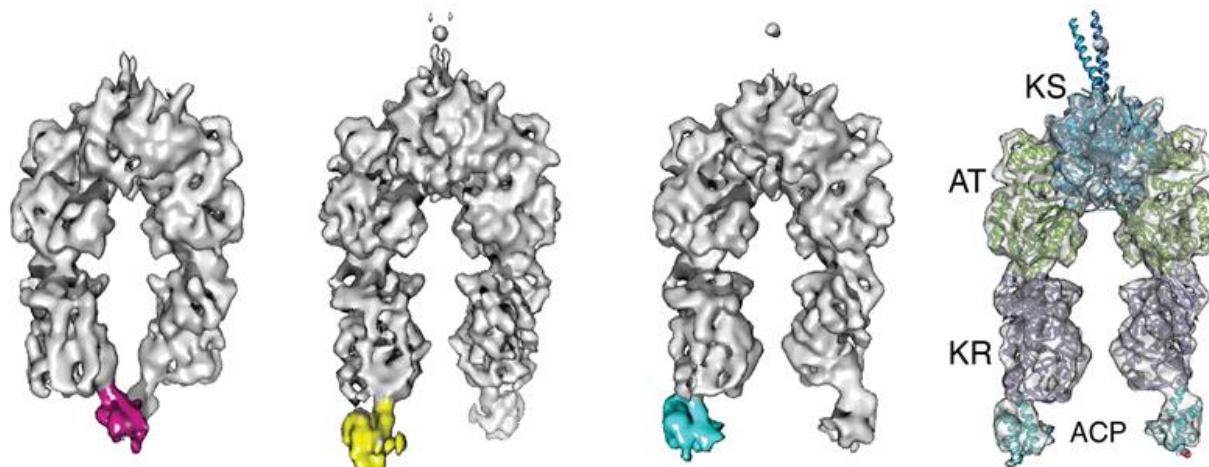
Figure 3.14. Cryo-EM and X-ray crystallography/NMR reconstructions of β-hydroxyhexaketide-bound PikAIII. Although there are three different conformers, the ACP (in fuchsia, yellow, and cyan) has limited flexibility. The ACP is positioned to transfer its substrate to the next module. Data generated by S. Dutta.

3.4 Conclusions

This work presents the first structure of an entire type I PKS module and captures the movement of an ACP, for which the location depends on the bound substrate on the Ppant arm (Figure 3.15). The cryo-EM structures show the overall structure of PikAIII during each step of the assembly line-like process whereas LC/FT-ICR MS determines and confirms each domain's active site occupancy. The MS data were consistently in agreement with the cryo-EM data and provide powerful evidence to support the unexpected cryo-EM structures. The ACP does not appear to engage in protein-protein interactions with other domains unless the appropriate substrate is loaded.
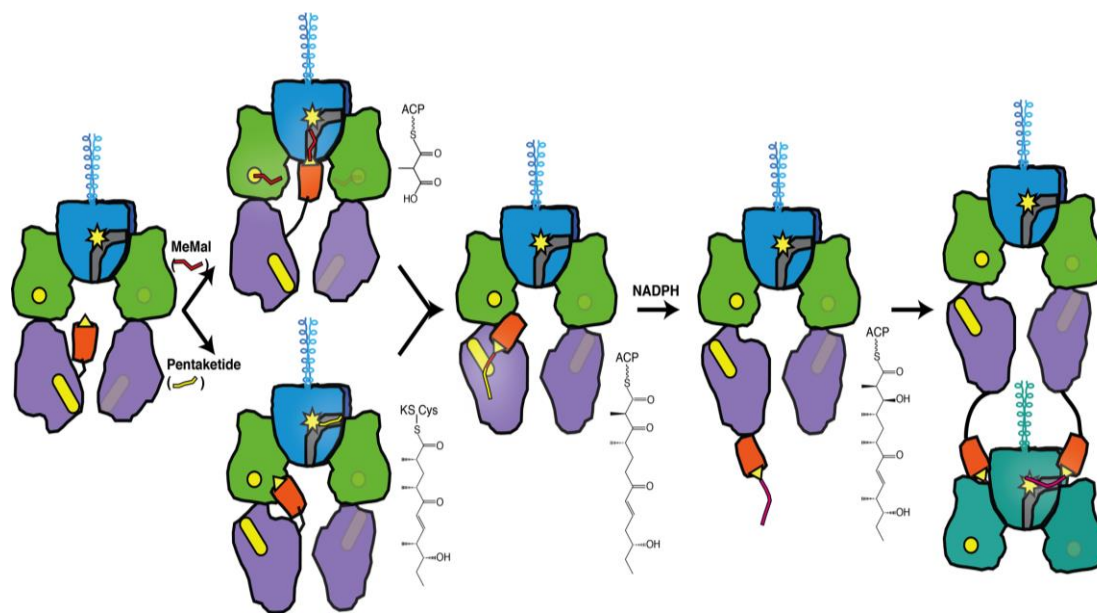
Figure 3.15. An overall summary of the ACP (in orange) movement depending on the substrate bound to it via Ppant. The figure also demonstrates the KR end-to-end flip. Figure produced by J. Whicher.

Moreover, there is evidence of interdomain cross-talk in the PikAIII module. For example, the end-to-end flip of the KR domain positions the KR active site near other active sites and the AT movements that block the KS active site entrance only occur when a pentaketide or hexaketide intermediate is present. These changes provide insight into understanding the kinetics of the module and appear to position each domain in anticipation of the next step in the process. These findings provide the first plausible mechanism, not only for ACP transfer, but also a detailed look into how type I PKSs function and produce natural products. This discovery will provide improvements in developing more effective combinatorial biosynthesis approaches and will allow for improved and more insightful synthesis using unnatural substrates.

3.5 References

(1)   Li, J. W.; Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **2009**, *325*, 161.

(2)   Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311.

(3)   Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **1990**, *348*, 176.

(4)   Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **1991**, *252*, 675.

(5)   Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat. Prod. Rep.* **2009**, *26*, 90.

(6)   Staunton, J.; Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **2001**, *18*, 380.

(7)   Brockmann, H.; Henkel, W. *Naturwissenschaften* **1950**, *37*, 138.

(8)   Brockmann, H.; Henkel, W. Pikromycin, ein bitter schmeckendes Antibioticum aus Actinomyceten (Antibiotica aus Actinomyceten, VI. Mitteil. *Chem. Ber.* **1951**, *84*, 284.

(9)   Oh, H.-S.; Kang, H.-Y. Total Synthesis of Pikromycin. *J. Organic Chem.* **2011**, *77*, 1125.

(10)   Maezawa, I.; Hori, T.; Kinumaki, A.; Suzuki, M. Biological Conversion of Narbonolide to Picromycin. *J. Antibiot.* **1973**, *26*, 771.

(11)   Kittendorf, J. D.; Sherman, D. H. The methymycin/pikromycin pathway: A model for metabolic diversity in natural product biosynthesis. *Bioorg. Med. Chem.* **2009**, *17*, 2137.

(12)   Kittendorf, J. D.; Sherman, D. H. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **2006**, *17*, 597.

(13)   Xue, Y.; Zhao, L.; Liu, H.; Sherman, D. H. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: Architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12111.

(14)   Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106*, 3468.

(15)   Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7*, 285.

(16)   Mercer, A. C.; Burkart, M. D. The ubiquitous carrier protein-a window to metabolitebiosynthesis. *Nat. Prod. Rep.* **2007**, *24*, 750.

(17)   Giraldes, J. W.; Akey, D. L.; Kittendorf, J. D.; Sherman, D. H.; Smith, J. L.; Fecik, R. A. Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat. Chem. Biol.* **2006**, *2*, 531.

(18)   Akey, D. L.; Kittendorf, J. D.; Giraldes, J. W.; Fecik, R. A.; Sherman, D. H.; Smith, J. L. Structural basis for macrolactonization by the pikromycin thioesterase. *Nat. Chem. Biol.* **2006**, *2*, 537.

(19)   Tsai, S.-C.; Lu, H.; Cane, D. E.; Khosla, C.; Stroud, R. M. Insights into Channel Architecture and Substrate Specificity from Crystal Structures of Two Macrocycle-Forming Thioesterases of Modular Polyketide Synthases†,‡. *Biochemistry* **2002**, *41*, 12598.

(20)   Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol.* **2009**, *4*, 41.

(21)   Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. The Structure of Docking Domains in Modular Polyketide Synthases. *Chem. Biol.* **2003**, *10*, 723.

(22)   Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **2007**, *16*, 2093.

(23)   Zheng, J.; Keatinge-Clay, A. T. Structural and Functional Analysis of C2-Type Ketoreductases from Modular Polyketide Synthases. *J. Mol. Biol.* **2011**, *410*, 105.

(24)   Buchholz, T. J., Kittendorf, JD., Sherman, David H. In *Wiley Encyclopedia of Chemical Biology*; Begley, T., Ed.; John Wiley and Sons: 2007.

(25)   Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H. Synthesis and Biochemical Analysis of Complex Chain-Elongation Intermediates for Interrogation of Molecular Specificity in the Erythromycin and Pikromycin Polyketide Synthases. *J. Am. Chem. Soc.* **2009**, *131*, 15784.

(26)   Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R. Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations†. *Biochemistry* **2001**, *40*, 15464.

(27)   Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortés, J.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489.

(28)   Aldrich, C. C.; Beck, B. J.; Fecik, R. A.; Sherman, D. H. Biochemical Investigation of Pikromycin Biosynthesis Employing Native Penta- and Hexaketide Chain Elongation Intermediates. *J. Am. Chem. Soc.* **2005**, *127*, 8441.

(29)   Aldrich, C. C.; Venkatraman, L.; Sherman, D. H.; Fecik, R. A. Chemoenzymatic Synthesis of the Polyketide Macrolactone 10-Deoxymethynolide. *J. Am. Chem. Soc.* **2005**, *127*, 8910.

(30)   Hans, M.; Hornung, A.; Dziarnowski, A.; Cane, D. E.; Khosla, C. Mechanistic Analysis of Acyl Transferase Domain Exchange in Polyketide Synthase Modules. *J. Am. Chem. Soc.* **2003**, *125*, 5366.

(31)   Yin, Y.; Lu, H.; Khosla, C.; Cane, D. E. Expression and Kinetic Analysis of the Substrate Specificity of Modules 5 and 6 of the Picromycin/Methymycin Polyketide Synthase. *J. Am. Chem. Soc.* **2003**, *125*, 5671.

(32)   Menzella, H. G.; Carney, J. R.; Santi, D. V. Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol.* **2007**, *14*, 143.

(33)   Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotech.* **2005**, *23*, 1171.

(34)   Sherman, D. H. The Lego-ization of polyketide biosynthesis. *Nat. Biotech.* **2005**, *23*, 1083.

(35)   Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **2002**, *19*, 542.

(36)   Khosla, C.; Zawada, R. J. X. Generation of polyketide libraries via combinatorial biosynthesis. *Trends. Biotechnol.* **1996**, *14*, 335.

(37)   Petkovic, H.; Lill, R. E.; Sheridan, R. M.; Wilkinson, B.; McCormick, E. L.; McArthur, H. A. I.; Staunton, J.; Leadlay, P. F.; Kendrew, S. G. A Novel Erythromycin, 6-Desmethyl Erythromycin D, Made by Substituting an Acyltransferase Domain of the Erythromycin Polyketide Synthase. *J. Antibiot.* **2003**, *56*, 543.

(38)   Kim, B. S.; Sherman, D. H.; Reynolds, K. A. An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. *Protein Eng. Des. Sel.* **2004**, *17*, 277.

(39)   Kakule, T. B.; Lin, Z.; Schmidt, E. W. Combinatorialization of Fungal Polyketide Synthase–Peptide Synthetase Hybrid Proteins. *J. Am. Chem. Soc.* **2014**.

(40)   Walsh, C. T. Combinatorial Biosynthesis of Antibiotics: Challenges and Opportunities. *ChemBioChem* **2002**, *3*, 124.

(41)   Weissman, K. J.; Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Micro.* **2005**, *3*, 925.

(42)   Bonnett, S. A.; Rath, C. M.; Shareef, A.-R.; Joels, J. R.; Chemler, J. A.; Hakansson, K.; Reynolds, K.; Sherman, D. H. Acyl-CoA Subunit Selectivity in the Pikromycin Polyketide Synthase PikAIV: Steady State Kinetics and Active-Site Occupancy Analysis by FTICR-MS. *Chem. Biol.* **2011**, *18*, 1075.

(43)   Jenner, M.; Afonso, J. P.; Bailey, H. R.; Frank, S.; Kampa, A.; Piel, J.; Oldham, N. J. Acyl-Chain Elongation Drives Ketosynthase Substrate Selectivity in trans-Acyltransferase Polyketide Synthases. *Angew. Chem. Int. Ed. Engl.* **2014**.

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

(45)   Sherman, D. H.; Smith, J. L. Clearing the Skies over Modular Polyketide Synthases. *ACS Chem. Biol.* **2006**, *1*, 505.

(46)   Staunton, J.; Caffrey, P.; Aparicio, J. F.; Roberts, G. A.; Bethell, S. S.; Leadlay, P. F. Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Mol. Biol.* **1996**, *3*, 188.

(47)   Maier, T.; Leibundgut, M.; Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **2008**, *321*, 1315.

(48)   Jenni, S.; Leibundgut, M.; Maier, T.; Ban, N. Architecture of a Fungal Fatty Acid Synthase at 5 Å Resolution. *Science* **2006**, *311*, 1263.

(49)   Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. *Science* **2001**, *291*, 1790.

(50)   Sánchez, C.; Du, L.; Edwards, D. J.; Toney, M. D.; Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from Streptomyces verticillus ATCC15003, the producer of the hybrid peptide–polyketide antitumor drug bleomycin. *Chem. Biol.* **2001**, *8*, 725.

(51)   Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. Structure of a modular polyketide synthase. *Nature* **2014**, *510*, 512.

(52)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014**, *510*, 560.

(53)   Milne, J. L. S.; Borgnia, M. J.; Bartesaghi, A.; Tran, E. E. H.; Earl, L. A.; Schauder, D. M.; Lengyel, J.; Pierson, J.; Patwardhan, A.; Subramaniam, S. Cryo-electron microscopy – a primer for the non-microscopist. *FEBS J.* **2013**, *280*, 28.

(54)   Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H. Biocatalytic synthesis of pikromycin, methymycin, neomethymycin, novamethymycin, and ketomethymycin. *J. Am. Chem. Soc.* **2013**, *135*, 11232.

(55)   Beck, B. J.; Yoon, Y. J.; Reynolds, K. A.; Sherman, D. H. The Hidden Steps of Domain Skipping. *Chem. Biol.* **2002**, *9*, 575.

(56)   Rowe, C. J.; Böhm, I. U.; Thomas, I. P.; Wilkinson, B.; Rudd, B. A. M.; Foster, G.; Blackaby, A. P.; Sidebottom, P. J.; Roddis, Y.; Buss, A. D.; Staunton, J.; Leadlay, P. F. Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem. Biol.* **2001**, *8*, 475.

Chapter 4

Elucidation of the Curacin and Jamaicamide Biosynthetic Pathways Via Phosphopantetheine Ejection

4.1 Introduction

A large number of antibiotics, anti-cancer agents, immunosuppressants as well as other drugs originate from nature[1-5]. Many of these natural products are synthesized by protein complexes called polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) or hybrids of the two[6-11]. Type I PKS biosynthetic pathways consist of multiple protein modules, each containing multiple domains, and act as an assembly line operating sequentially. Type I PKS modules minimally consists of three domains: a ketosynthase (KS), an acyl transferase (AT), and an acyl carrier protein (ACP)[6]. The KS domain is responsible for catalyzing carbon-carbon bond formations via a Claisen condensation, elongating the natural product intermediate by a two or three carbon chain[6,12]. The AT chooses an extender unit and transfers it to the ACP[6]. The ACP tethers and transfers natural product intermediates via a thiol residue on a phosphopantetheine (Ppant) post-translational modification[6,13,14]. Other domains add structural diversity to the natural product and include ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs), which generate β-hydroxyl, enoyl, or fully saturated products, respectively[15]. Mixing and matching these domains results in a large number of pathways that create very diverse natural products.

Moorea producens (formerly *Lyngbya majuscula*[16]) is a marine blue-green algae, a cyanobacterium which contains multiple biosynthetic pathways, generate interesting products,

such as the barbamides, jamaicamides, and curacin A[17,18]. Most cyanobacterial natural products, including those that originate from *Moorea producens*, are PK-NRP hybrids[19]. The barmamides have molluscicidal activity and have unusual chemical groups including a trichloromethyl group, a methyl enol ether, a thiazole ring, and *E*-alkene formation[19-21]. The jamaicamides, aptly named as they were first isolated in Hector's Bay, Jamaica, block sodium channels, are toxic to fish, and have unique functional groups such as an alkynyl bromide, a vinyl chloride, and a pyrrolinone ring[17,22]. Curacin A is an anti-cancer agent which inhibits tumor cell growth and contains interesting chemical features, such as a cyclopropane, a thiazoline, a *cis*-double bond, and a terminal alkene[23,24]. Both the Cur and Jam biosynthetic gene clusters, responsible for producing Curacin A, and the jamaicamides, respectively, have high sequence identity in their early segments and include a 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase-like (HCS) gene cassette encoding enzymes to create a β-branch in the polyketide intermediate[25,26]. The sections with high sequence identity in the Cur and Jam gene clusters encode for an Fe^{2+}/α-ketoglutarate (α-KG)-dependent halogenase (Hal) and three ACPs in the CurA and JamE modules[17,24,27]. The halogenase in the Jam pathway is unsurprising because there is a vinyl chloride in the final natural product and halogenation is well characterized in biosynthetic pathways[28,29]. However, Curacin A lacks a halogen, so the halogenase's function was less certain although it had been speculated to aid in cyclopropane formation[30]. Additional domains that exhibit high sequence identity are an HMG-ACP synthase in CurD and JamH, a dehydratase (ECH_1) in CurE and JamI, a decarboxylase (ECH_2) domain in CurF and JamJ, and an enoyl reductase (ER) domain in CurF and JamJ[17,24,25,31]. Generally, sequence identity over fifty percent results in the same structure and therefore function, however, the biosynthetic products from the Cur and Jam ER domains were found to be quite different: a cyclopropane for the Cur ER and a vinyl chloride for the Jam ER[31-33]. Previous collaborations between the Hakansson lab and the Smith, Gerwick, and Sherman labs have characterized these enzymes and their reaction order structurally and biochemically, revealing insight into the catalytic mechanism of each pathway[25,31,34].
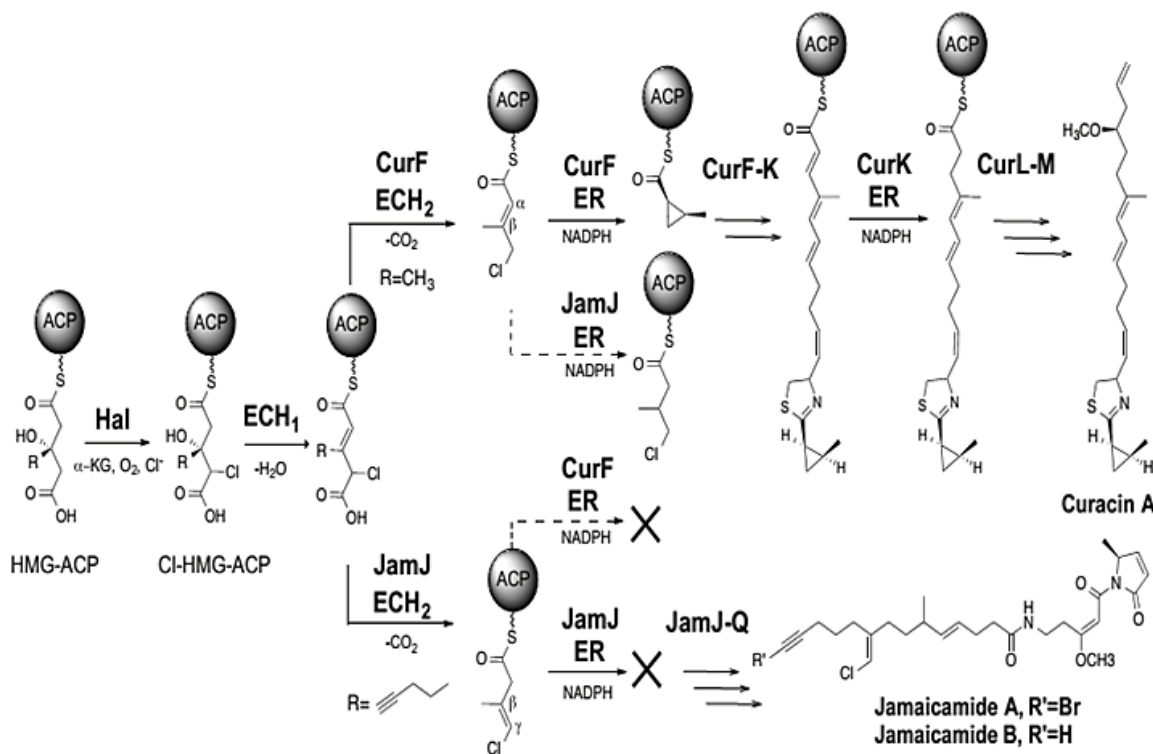
Figure 4.1. Divergence in the curacin and jamaicamaide biosynthetic pathways. The pathways diverge at the ECH_2 step, in which the CurF ECH_2 creates an α-β double bond, whereas, the Jam ECH_2 forms a β-γ double bond. JamJ ER is unreactive toward the JamJ ECH_2 product, yet reacts with the CurF ECH_2 product. Figure generated by Khare, D.

For example, it was found that, in the curacin pathway, CurA Hal chlorinates the HMG-ACP, followed by dehydration and decarboxylation by ECH_1 and ECH_2, respectively[35]. The jamaicamide pathway is identical through the dehydration step, but diverges at the ECH_2 step. Here, CurF ECH_2 forms an α,β-enoyl-γ-chloro-ACP whereas the 59% identical JamJ ECH_2 forms a β,γ-enoyl-γ-chloro-ACP[31,36]. The next domain, the ER, produces stark differences. Addition of CurF ER results in the formation of a cyclopropane, involving hydride addition from NADPH and chloride elimination, whereas the 65% identical JamJ ER has no function: it does not react with the β,γ-enoyl-γ-chloro-ACP product from the JamJ ECH_2 (Figure 4.1). However, JamJ ER reacts with the α,β-enoyl-γ-chloro-ACP product of CurF ECH_2, resulting in the reduction of an α-β double bond[31] (Figure 4.2). These two pathways are an example of how combinatorial biosynthesis, even with a single enzyme, may lead to different functional groups.

Due to the major functional differences in the ER domains between the Cur and Jam pathways, it is of interest to structurally investigate ER domains for potential differences that could explain the formation of different functional groups. Generally, an ER domain's function is to reduce an α-β alkene to an alkane[6]. This function is found in multiple ERs in the Cur and Jam pathways, including CurH ER, CurK ER, a second ER domain within JamJ, and JamL ER. The Cur and Jam ER domains belong to a superfamily of medium chain dehydrogenase/reductase (MDR) proteins that are well studied and include Zn^{2+}-dependent alcohol dehydrogenases and quinone oxidoreductases[37-39]. A common theory regarding the mechanism of the MDR and PKS ER reductions involves a hydride transfer from the NADPH nicotinamide to the β-carbon atom of the unsaturated substrate to form an enolate intermediate, followed by proton transfer from the enzyme to generate a saturated product[39]. The assumed mechanism for the CurF ER cyclopropanase includes similar initial steps of hydride transfer and formation of an enolate intermediate, however, chloride removal and a simultaneous nucleophilic attack of the γ-carbon on the α-carbon result in the formation of a cyclopropane[30].
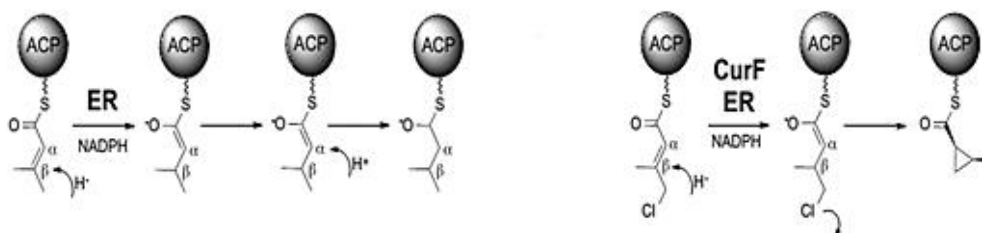
Figure 4.2. In a PKS enoyl reduction reaction the expected first step involves a hydride transfer from the cofactor NADPH to the unsaturated substrate, followed by protonation at the α-carbon atom (left). The unique CurF ER catalyzes the nucleophilic displacement of the chlorine atom in an NADPH-dependent manner to form the highly strained cyclopropane ring[31] (right). Figure generated by Khare, D.

Although some mechanistic details regarding PKS ERs have been deduced, the overall molecular and structural bases for enoyl reduction are unknown. Structural and biochemical investigations to determine a potential proton donor for reduction have been unsuccessful. For example, site-directed mutagenesis based on a homology model for the erythromycin ER identified conserved amino acids for controlling ER stereochemistry, but was unable to identify a proton donor[40,41]. Crystal structures of other PKS ERs, including a KR-ER didomain from

spinosyn (Spn) and a trans-acting ER from lovastatin, did not reveal a proton donor either[42,43]. To understand the structural and molecular basis of the unusual CurF ER cyclopropanase, a combination of X-ray crystallography and a phosphopantetheine ejection assay via FT-ICR mass spectrometry was employed.

4.2 Experimental

4.2.1 Cloning, expression and protein purification

Proteins were expressed by D. Khare as follows: CurF ER (amino acids 275-611, here numbered 1-337) and the first ER domain within JamJ (amino acids 260-594, here numbered 1-335) were amplified by PCR from plasmids encoding CurF ER and JamJ ER[31]. The amplified products were inserted into pMCSG7 (JamJ ER) and pMocr (JamJ ER and CurF ER) vectors and expressed in *E. coli* BL21 (DE3) cells[44,45]. Cells were grown to an OD_{600} of 1.0 at 37°C in TB supplemented with 4% glycerol, incubated at 20°C for 1 hr, induced with 0.4 mM IPTG, and grown for an additional 20 hr. The cell pellet from a 0.5-L culture was resuspended in 40 mL Buffer A (100 mM Tris pH 7.8, 500 mM NaCl, 5 mM imidazole and 10% glycerol) and lysed by ultrasonication. All purification steps were carried out at 4°C. The soluble fraction was loaded to a HisTrap Ni^{2+} NTA column (GE Healthcare), washed with Buffer A and eluted with a linear gradient of 5-500 mM imidazole. The His_6-Mocr tag was cleaved by TEV protease and the tag-free ER purified on a Ni^{2+} NTA column. Further purification was done on a Superdex 200 size exclusion column equilibrated with Buffer B (10 mM Tris pH 7.8, 50 mM NaCl and 10% glycerol). Selenomethionyl labeled JamJ ER was purified similarly to the wild type protein, except that the pMocr::JamJ ER plasmid was transformed into the *E. coli met⁻* auxotroph strain B834 (DE3), cells were grown in minimal media (Molecular Dimensions), and 1 mM DTT was added in all subsequent purification steps. All excised PKS ER domains eluted with an apparent molecular weight of 40 kDa, corresponding to a monomer. A JamJ-CurF ER chimera was generated by using overlap PCR to swap 12 amino acids of JamJ ER with the corresponding 14 amino acids from CurF ER. Mutants of JamJ and CurF ER in the pMocr vector were generated by site-directed mutagenesis and purified as described above.

4.2.2 Crystallization, data collection, structure determination and refinement

Details of these steps are provided by D. Khare in the manuscript associated with this chapter[46].

4.2.3 ACP linked substrate and ER activity assay

The following experiments were performed by D. Khare: The ACP-linked HMG substrate was generated and ER was assayed as described previously[31]. Briefly, 100 μM of ACP-linked substrate was incubated with 10 μM enzymes (CurA halogenase + CurE ECH_1 + CurF ECH_2) and the corresponding cofactors at 25°C for 25 min each. The reactions were initiated by exposing the reaction mixture to O_2. The γ-Cl-3-methylcrotonyl-ACP sample was further treated with CurF ER (100 nM) or JamJ ER (100 nM) for 5 min at 25°C. The reactions were quenched by addition of 10% formic acid (CH_2O_2) and the ACP-loaded sample was separated from the reaction mix by reverse phase HPLC using a Source 15RPC ST4.6/100 column (GE Healthcare). The proteins were eluted with a linear gradient from 30% to 70% of CH_3CN (0.05% CF_3COOH and 0.05% CH_2O_2)/H_2O (0.05% CF_3COOH and 0.05% CH_2O_2).

4.2.4 ACP sample analysis by FT-ICR mass spectrometry

HPLC fractions were diluted at a 40:100 ratio in 50/50 HPLC-grade H_2O (ThermoFisher Scientific) and CH_3CN (ThermoFisher Scientific) with 0.1% CH_2O_2 (ThermoFisher Scientific). Samples were directly infused into the electrospray source of an actively shielded 7T SolariX quadrupole FTICR-MS (Bruker Daltonics). Data were gathered from m/z 200-2000 in positive ion mode. Electrospray was conducted at 4,500V with 128 scans per spectrum and the transient set to 1M data points. External ion accumulation in a hexapole was 0.5 s with one ICR fill prior to excitation and detection. External calibration utilized HP-mix (Agilent) mixed with ubiquitin (Sigma Aldrich). Collision-induced dissociation (CID) was performed in the collision cell hexapole with ultra-high purity argon (Cryogenic Gases) at 10 V and all charge states present were fragmented. Quantification of the Ppant arm and attached products was carried out as previously described[35,47].

4.3 Results and Discussion

4.3.1 Enoyl Reductase Concentration Optimization and Utilization of Isotopic Fine Structure

The concentration of the ER incubated with the ACP had a large effect on the generated natural product intermediate. 100 nM of JamJ ER wildtype completely reduced the vinyl chloride, whereas 1 nM or 10 nM partially reduced the vinyl chloride (Figure 4.3). Due to the mass difference of 2 Da before and after the vinyl chloride reduction, ultra-high resolution mass spectrometry was imperative to determine the concentration effects. Differences of 1 or 2 Da are easily detectable for both the intact protein and the fragmentation products after CID. 10 nM and 100 nM were selected as suitable ER concentrations to determine the effects of the point mutations.
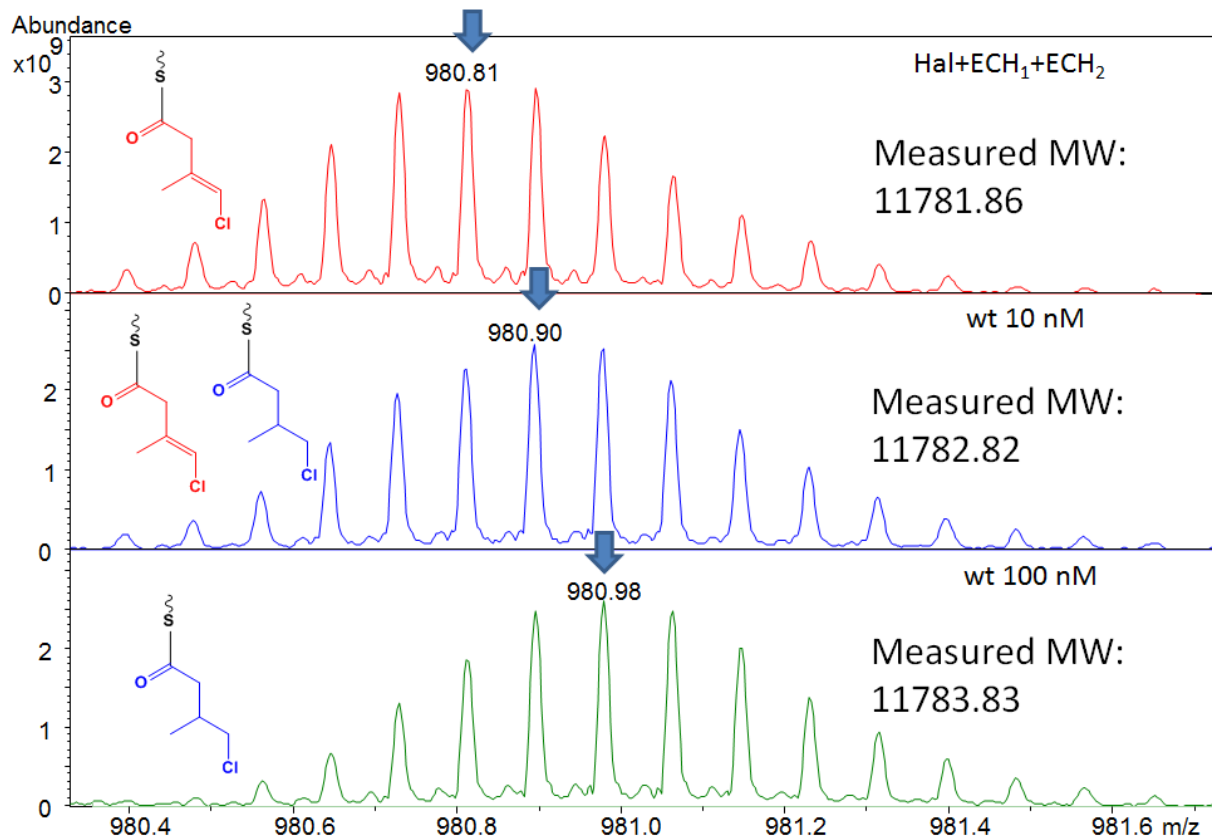
Figure 4.3. The concentration of JamJ ER has a significant effect on the ACP-bound reaction product. 0 nM, 10 nM, and 100 nM has no effect, partially reduces a vinyl chloride, and completely reduces a vinyl chloride, as shown by the measured molecular weight of the intact ACP.

A similar concentration dependence was observed for the CurF ER (Figure 4.4). The concentrations of CurF ER ranged from 1 fM to 1000 nM. Suitable concentrations ranged from 0.1 nM to 100 nM for determining CurF ER effects on the cyclopropane formation. Although the cyclopropane formation has a larger mass shift than the reduction of the vinyl chloride, ultra-high resolution mass spectrometry remains an ideal and precise method to determine the effects of the enzyme's concentration.

Ultra-high resolution mass spectrometry was also necessary for relative quantitation, conducted by measuring and normalizing peak heights. For the JamJ ER, resulting in a mass difference of 2 Da, the third isotopic peak of the halogenated, dehydrated, and decarboxylated product is less than half a milliDalton from the monoisotopic peak of the reduced product. Also, due to the chlorine atom in both substrate and product, their isotopic patterns are atypical, with the third isotopic peak being more abundant than the second isotopic peak due to the presence of ^{37}Cl (Figure 4.5).
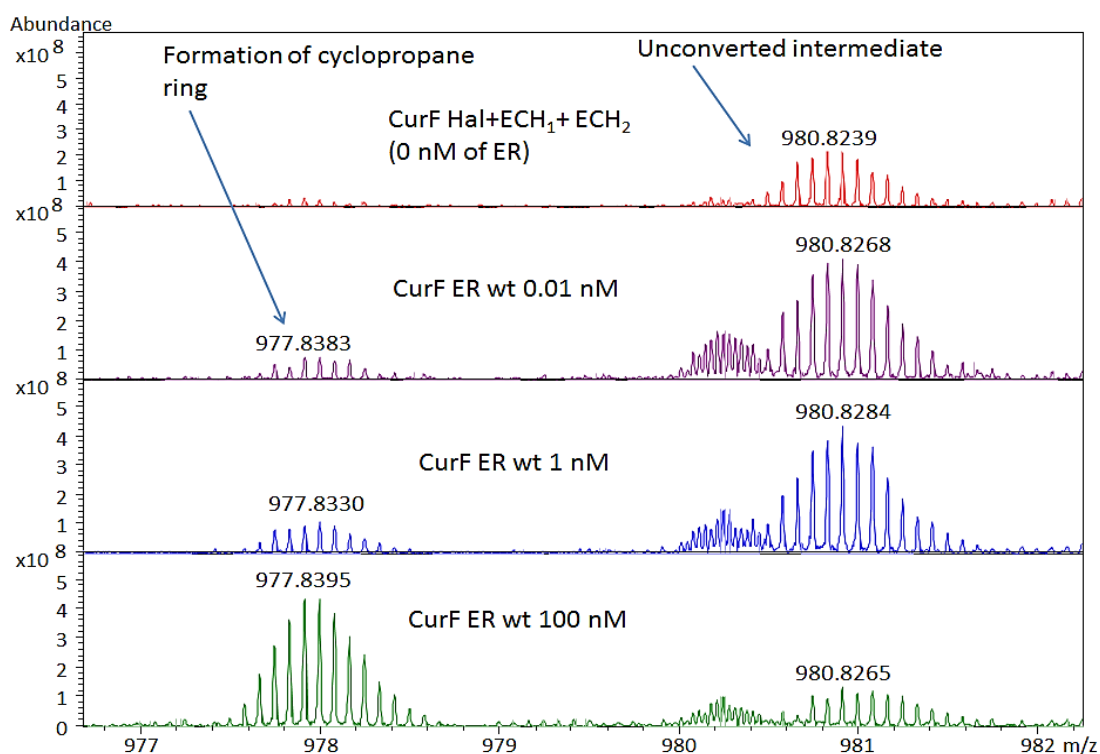
Figure 4.4. Cyclopropane formation efficiency is dependent on the concentration of the CurF ER enzyme. 0.01 nM and 1 nM partially convert the natural product intermediate, whereas 100 nM mostly converts the intermediate.
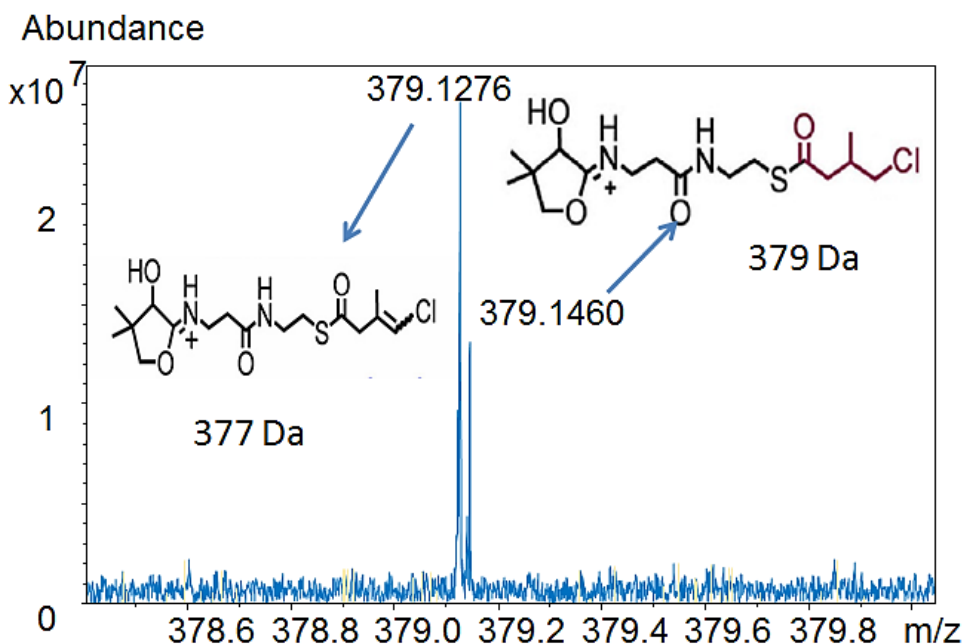
Figure 4.5. Resolved isotopic fine structure for a mixture of non-reduced substrate and reduced product. The left peak corresponds to the ^{37}Cl isotope of the non-reduced substrate and the right peak corresponds to the monoisotopic peak of the reduced product.

4.3.2 JamJ ER crystal structure and mechanism determination with FT-ICR mass spectrometry

The crystal structure for JamJ ER was solved. X-ray crystallography revealed the structure of JamJ ER with its bound cofactor at 2.25 Å and in the apo form at 1.8 Å. Its structure has the typical fold of the MDR enzyme superfamily to which it belongs[48,49]. MDR superfamily members predominantly exist as dimers or tetramers. However, the oligomeric state of PKS ER enzymes is unique as the JamJ ER is a monomer[42,43]. Like other PKS ERs, the NADPH cofactor binds to JamJ ER in a narrow cleft between the catalytic and nucleotide-binding domains through hydrogen bonds and van der Waals interactions[42,43].
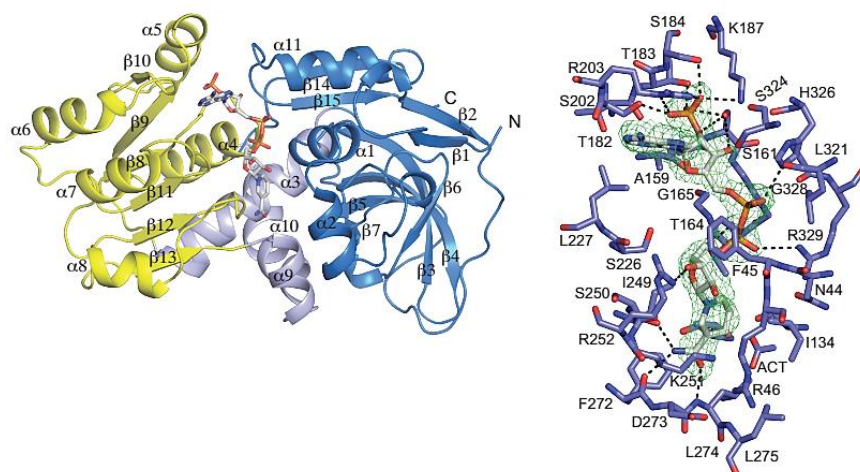
Figure 4.6. Ribbon diagram of a JamJ ER monomer (left). The putative catalytic domain is colored in blue and the NADPH binding domain is in yellow. The two domains are connected by helices α3 and α9- α10 (light purple). Cofactor NADPH is colored in red, blue, and orange. Ball and stick model (right) with a stereoview of the representative electron density around the cofactor NADPH shown from a F_o-F_c map (light green mesh) at 2.25 Å resolution. Hydrogen bonds between the cofactor NADPH and JamJ ER are illustrated as black dashed lines. Data generated by Khare, D.

The nicotinamide group from NADPH is at an open end of the active site cleft, defining the substrate entrance. D275 is at the entrance to the substrate channel, and in both structures is within reach of a conserved basic side chain (CurF R253, JamJ K251). K251 has been proposed as the proton donor in JamJ ER reduction, however mutagenesis coupled with FT-ICR mass spectrometry refutes this proposal, as shown in Figure 4.7. When this residue is substituted for an alanine or an arginine, reduction activity is retained.
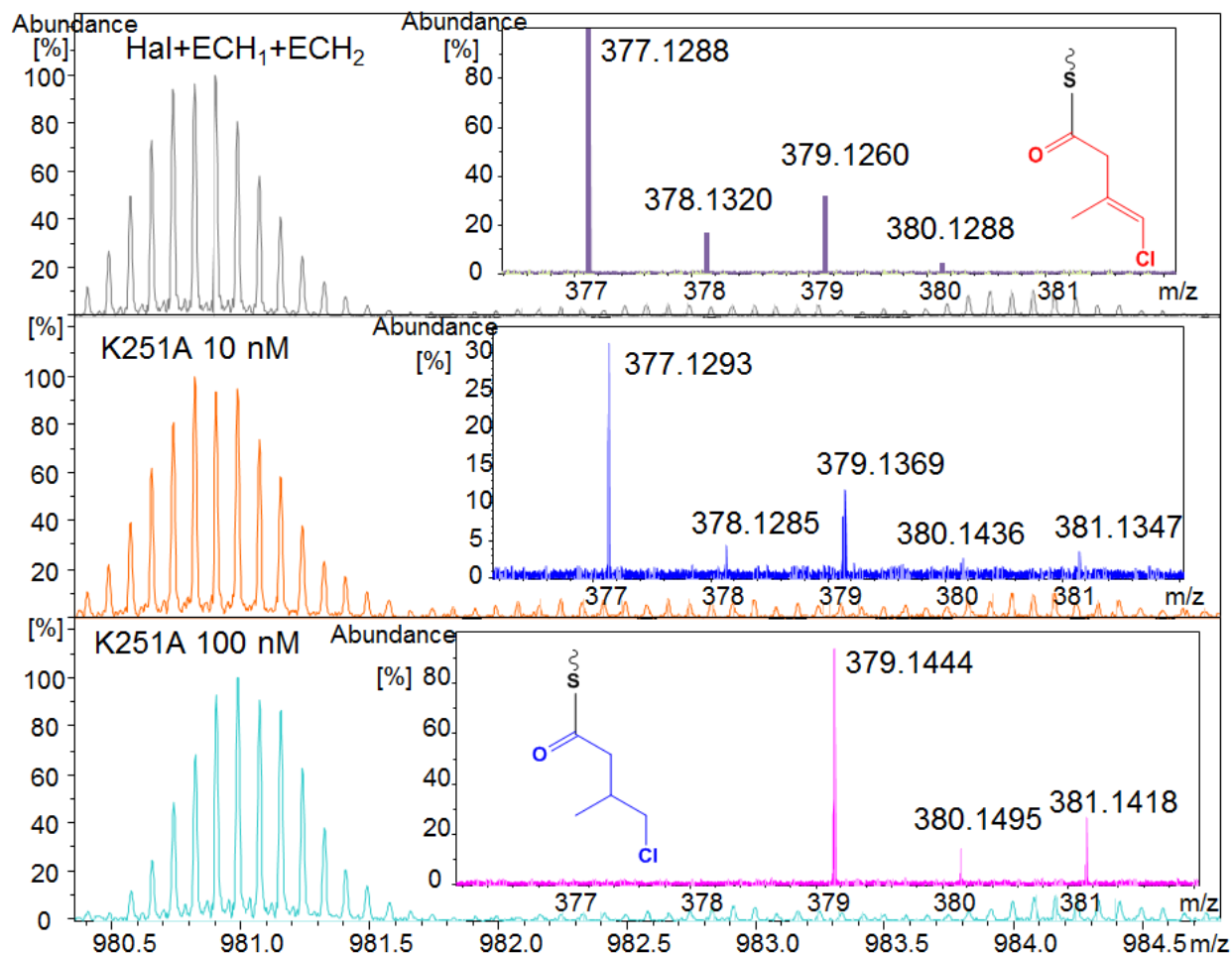
Figure 4.7. Intact 12+ charge state and Ppant ejection products for JamJ HMG-ACP following incubation with halogenase (Hal), dehydratase (ECH$_1$), and decarboxylase (ECH$_2$) (top) as well as additional incubation with K251A JamJ ER at 10 nM (middle) and 100 nM (bottom). No effect is seen from this mutation as reduction still occurs.

The phosphopantetheine assay determined and quantified the amount of reduction generated by the JamJ ER. High resolution ($\geq$ 70,000) was necessary to detect and identify reduced product: the action of JamJ ER results in the addition of two hydrogens, a mass difference of 2 Da for the intact protein. If a mixture of reduced and non-reduced protein is present, the apparent mass difference is less than 2 Da. High resolution confirms the tandem mass spectrometry (MS/MS) ejected substrate and product, respectively: isotopic fine structure is observed due to the presence of chlorine in both the substrate and product. As observed by FT-ICR mass spectrometry, the K251A mutation has no effect (Fig. 4.7). A PKS ER active site

should bind its substrate near the NADPH hydride donor and near a proton donor. It should also provide stabilizing hydrogen bonds to the enolate intermediate that forms following hydride transfer. Although the function is known, the identity of the amino acids that fulfill this role in PKS ERs and other members of the MDR superfamily are wholly unknown, including the proton donor in JamJ ER. Although D273 mutated to an alanine inhibited reduction (Figure 4.8), it did not eliminate it completely. No single amino acid was found to be essential for vinyl chloride reduction; therefore, it is possible that the proton donor is a water molecule[40,43].
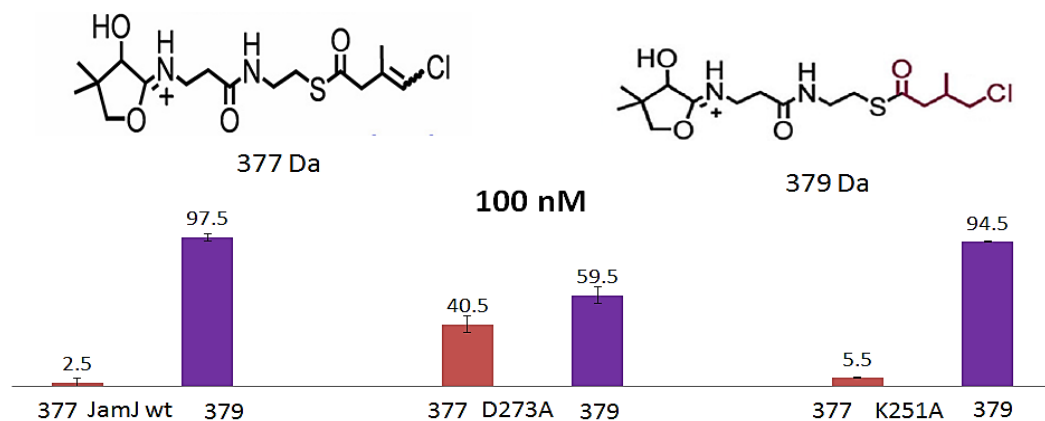
Figure 4.8. A summary of percent reduction for the three forms of JamJ ER investigated. The left bars for each form (maroon) represent the non-reduced substrate (m/z 377) and the right bars (purple) represent the reduced product intermediate at m/z 379.

4.3.3. CurF ER crystal structure and mechanism determination with FT-ICR mass spectrometry

The crystal structure of CurF ER was also solved, revealing the structure of CurF ER with its bound cofactor at 0.96 Å; overall its structure is very comparable to the JamJ ER domains, with a Cα core rmsd of 0.7 Å for JamJ ER-NADPH and 0.8 Å for JamJ apo-ER. Overall, the NADPH cofactor binds in a similar way to CurF ER. In addition, the NADPH nicotinamide position is stabilized by hydrogen bonds between the amide and the ER backbone at D275, the carbonyl of F274, and NH of M276.
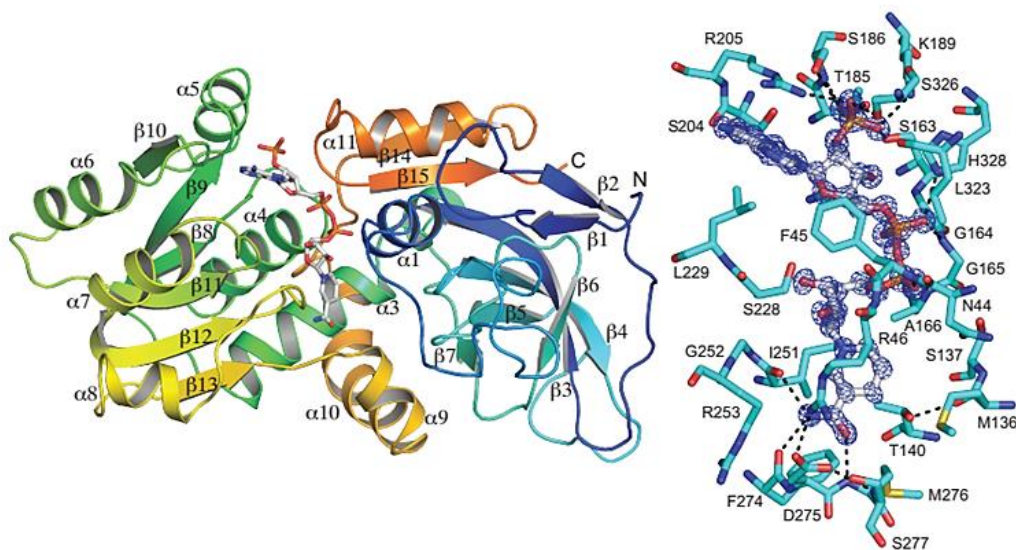
Figure 4.9. Ribbon diagram of CurF ER monomer in rainbow colors (left). Cofactor NADPH is colored in red, blue, and orange. Ball and stick model (right) with stereoview of the environment around the cofactor NADPH at 0.96 Å resolution in blue. Hydrogen bonds between the cofactor NADPH and CurF ER are indicated as black dashed lines. Data generated by Khare, D.

The proximity of the conserved and oppositely charged side chains of D275 and R253 may be critical to an ER conformation that exists only when the substrates are bound; both amino acids have been proposed as the proton donor in ER reduction[40,42]. The phosphopantetheine assay determined and quantified the amounts of natural product intermediates generated by the CurF ER. These sites as well as other amino acids in or near the active site were mutated to determine a proton donor. Interestingly, substitution of CurF ER R253 with the more typical lysine residue reduced cyclopropanation the most (Figure 4.10). The D275 mutation reduced cyclopropanation moderately. When Y58 and Y108 were mutated to alanine residues, cyclopropanation was also reduced. CurF R46 is another conserved basic residue near the substrate entrance that interacts with D275, but had a modest reduction in cyclopropanation when mutated to an alanine. The S277 is further away from the active site and mutated to alanine as a control. Predictably, this mutation had no effect on the cyclopropanation activity. Like JamJ ER, no single amino acid in the CurF ER was essential for cyclopropane formation, providing further evidence for a universal proton donor, such as a water molecule.

100
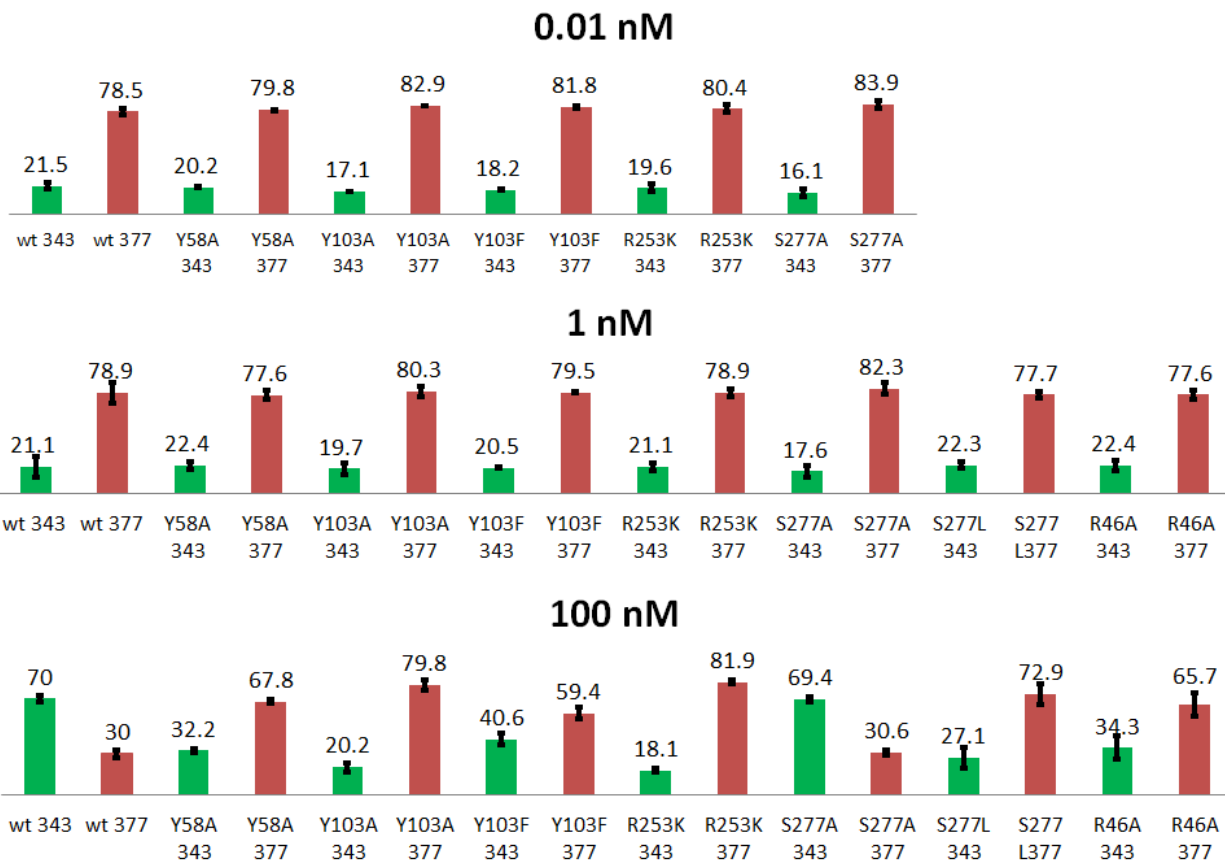
Figure 4.10. A summary of cyclopropane formation efficiency for the CurF ER mutants investigated. The left bars for each form in green represent the cyclopropane product (m/z 343) and the right bars in blue represent the non-reduced substrate at m/z 377.

In many ways, JamJ ER and CurF ER are comparable with other ERs in this family; they both have similar topology to the MDR superfamily and do not have a single amino acid essential for catalytic activity, whether it performs a reduction or forms a cyclopropane. Therefore, the proton donor for both ERs is likely to be a water molecule as hypothesized by others[40,43]. Notably, the CurF ER structure also reveals that the size of the substrate binding pocket correlates with the size of the substrate acyl group. The binding pocket extends into the hydrophobic core of the substrate-binding domain, indicating that the pocket size could be important for substrate selectivity. Overall, the active site and substrate cleft have conserved amino acids, except for a substrate loop which lines the substrate cleft, facing the nicotinamide ring and may be used for substrate recognition and selectivity. In JamJ and CurF, it is a longer

loop and also differently structured from similar ERs such as Spn ER. The JamJ and CurF ER

sequences vary the most amongst this loop region as well, as seen in Table 4.1.

Table 4.1. Sequences of various enoyl reductases, highlighting the varying loop regions. Table generated by Khare, D.

4.3.4 JamJ-CurF ER chimera formation and catalytic activity determination with FT-ICR mass spectrometry

Although the JamJ and CurF ER structures have a high degree of similarity, with 65% identity, they differ mostly in the substrate loop area and generate products with very different properties and functions. Two questions beg to be answered: are the differences in the substrate loop responsible for the differences in function for the two ERs and can the JamJ ER be mutated and manipulated into a cyclopropanase, similar to the CurF ER? Due to the difference in both amino acids and length in the loop, a chimera ER was designed with CurF amino acids 54-68 dubbed as the "cyclopropanase loop." Overall, as observed in the CurF ER crystal structure, the cyclopropanase loop is very ordered, despite the lack of alpha helices and beta sheets, with at least 15 hydrogen bonds for support. Another reason for selecting this group of amino acids was R62, which is only found in the CurF ER and is important for the highly structured loop. From the CurF ER crystallography data, it was determined that R62 forms four hydrogen bonds from its sidechain to uncharged carbonyl oxygen atoms and a fifth hydrogen bond with a water molecule in close proximity to the chlorine atom in the CurF substrate. This residue is also within 9 Å of the nicotinamide group and is ideally positioned to stabilize a chloride leaving group near the bound water. The function of the cyclopropanase loop was

tested by transplanting the loop into JamJ ER K251R, creating an active and stable enzyme. Additional substitutions were performed near the loop, however, these changes led to an unstable protein. With the natural CurF ER substrate, 3-chloromethyl-crotonyl-ACP, both the reduced product and cyclopropane product were detected with FT-ICR MS, although the cyclopropane was only about 10% of the total product (Figure 4.11). This chimera protein demonstrates the ability to form a cyclopropane and provides evidence that the cyclopropanase loop determines the unique activity of the curacin biosynthetic pathway. Another outcome to note is that although the substrate loop is a large contributing factor to forming the cyclopropane in the CurF ER, it is not a large contributing factor to reductase activity in the JamJ ER, as the chimera demonstrates both activities.
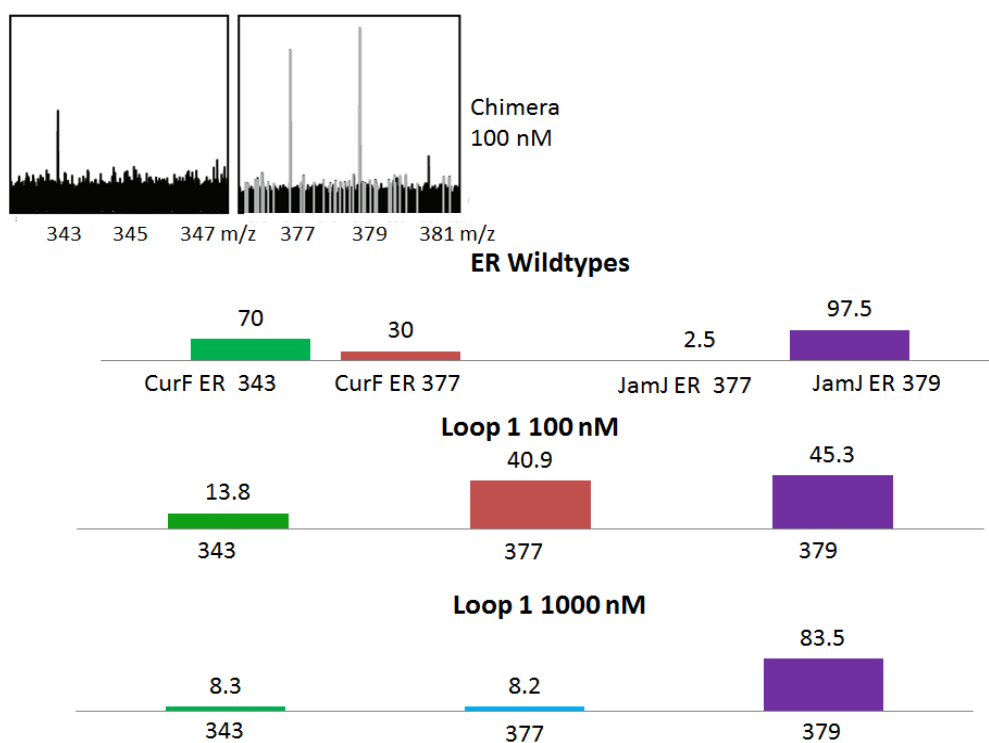
Figure 4.11. Partial mass spectra from Ppant ejection of the ACP-bound products following incubation of γ-Cl-3-methylcrotonyl-ACP with the JamJ-CurF ER chimera (top). The observed products constitute a mixture of the products following incubation with JamJ ER and Cur F ER, respectively. ACP-bound products were ejected by CID and detected and quantified by FT-ICR MS. The cyclopropane product intermediate is at m/z 343, the non-reduced substrate intermediate is at m/z 377, and the reduced product intermediate is at m/z 379. Bar graphs (bottom) display activity of the wildtype and chimera ERs as normalized percent formation of

the cyclopropane (in green), unreacted (in maroon and light blue), and the reduced product (in purple) at different ER concentrations.

4.4 Conclusions

Overall, this study sought to determine the structural and mechanistic differences between two enoyl reductase domains with high identity but dramatically disparate properties. JamJ ER applied to an HMG-ACP, also incubated with halogenase, dehydratase, and decarboxylase, results in the expected reduction of an α-β alkene, whereas CurF ER results in the formation of a cyclopropane in the natural product intermediate covalently bound to the ACP. Although the two biosynthetic pathways have high identity and therefore theoretically similar functions, the jamaicamide pathway generates products known as the jamaicamides, toxic to fish and blocking sodium channels, while the curacin pathway produces Curacin A, an anti-cancer agent. The crystal structures for the CurF ER and JamJ ER were solved. Based on the crystal structures, point mutations were made near the active site in each ER to determine which amino acids were essential for catalytic activity. Neither ER had a single amino acid responsible for catalytic activity. However, D273 in Jam and R253, Y58, and Y103 in Cur, contributed more to activity, as reduction or cyclopropane formation decreased when these amino acids were mutated to alanine residues. Crystallography showed major differences between JamJ ER and CurF ER in the substrate-binding region, which led to the design of a 'chimeric' ER enzyme, which generates both reduced product and cyclopropane. The amino acids from the CurF ER substrate loop were transplanted into the JamJ ER K251R enzyme to create this chimeric protein. This novel enzyme demonstrated that the substrate loop is important, though not fully responsible for ER action, as reduction still occurred with the substitution. The insights into the structure and function of the CurF ER and the JamJ ER, gained by high resolution mass spectrometry and X-ray crystallography, provide important details into the inner workings of the enzymes, including the potential for transforming a toxic product into an anti-cancer drug.

4.5 References

(1)   Li, J. W.; Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **2009**, *325*, 161.

(2)   Newman, D. J.; Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311.

(3)   Gerwick, W. H.; Fenner, A. M. Drug Discovery from Marine Microbes. *Microb. Ecol.* **2013**, *65*, 800.

(4)   Clardy, J.; Fischbach, M. A.; Walsh, C. T. New antibiotics from bacterial natural products. *Nat. Biotech.* **2006**, *24*, 1541.

(5)   Gerwick, W. H.; Moore, B. S. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.* **2012**, *19*, 85.

(6)   Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106*, 3468.

(7)   Jones, A. C.; Gu, L.; Sorrels, C. M.; Sherman, D. H.; Gerwick, W. H. New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Curr. Opin. Chem. Biol.* **2009**, *13*, 216.

(8)   Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 4688.

(9)   Kakule, T. B.; Lin, Z.; Schmidt, E. W. Combinatorialization of Fungal Polyketide Synthase–Peptide Synthetase Hybrid Proteins. *J. Am. Chem. Soc.* **2014**.

(10)   Wipf, P.; Reeves, J. T.; Balachandran, R.; Day, B. W. Synthesis and Biological Evaluation of Structurally Highly Modified Analogues of the Antimitotic Natural Product Curacin A. *J. Med. Chem.* **2002**, *45*, 1901.

(11)   Schwecke, T.; Aparicio, J. F.; Molnár, I.; König, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortés, J.; Lester, J. B. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7839.

(12)   Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **2003**, *7*, 285.

(13)   Staunton, J.; Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **2001**, *18*, 380.

(14)   Mercer, A. C.; Burkart, M. D. The ubiquitous carrier protein-a window to metabolitebiosynthesis. *Nat. Prod. Rep.* **2007**, *24*, 750.

(15)   Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortés, J.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489.

(16)   Engene, N.; Rottacker, E. C.; Kastovsky, J.; Byrum, T.; Choi, H.; Ellisman, M. H.; Komarek, J.; Gerwick, W. H. Moorea producens gen. nov., sp. nov. and Moorea bouillonii comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1171.

(17)   Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. Structure and biosynthesis of the jamaicamides, new mixed polyketide-

peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chem. Biol.* **2004**, *11*, 817.

(18)   Burja, A. M.; Banaigs, B.; Abou-Mansour, E.; Grant Burgess, J.; Wright, P. C. Marine cyanobacteria—a prolific source of natural products. *Tetrahedron* **2001**, *57*, 9347.

(19)   Chang, Z.; Flatt, P.; Gerwick, W. H.; Nguyen, V.-A.; Willis, C. L.; Sherman, D. H. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* **2002**, *296*, 235.

(20)   Orjala, J.; Gerwick, W. H. Barbamide, a Chlorinated Metabolite with Molluscicidal Activity from the Caribbean Cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* **1996**, *59*, 427.

(21)   Flatt, P. M.; O'Connell, S. J.; McPhail, K. L.; Zeller, G.; Willis, C. L.; Sherman, D. H.; Gerwick, W. H. Characterization of the Initial Enzymatic Steps of Barbamide Biosynthesis. *J. Nat. Prod.* **2006**, *69*, 938.

(22)   Jones, A. C.; Gerwick, L.; Gonzalez, D.; Dorrestein, P.; Gerwick, W. H. Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium Lyngbya majuscula and identification of possible regulatory proteins. *BMC Microbiol.* **2009**, *9*, 247.

(23)   Verdier-Pinard, P.; Sitachitta, N.; Rossi, J. V.; Sackett, D. L.; Gerwick, W. H.; Hamel, E. Biosynthesis of Radiolabeled Curacin A and Its Rapid and Apparently Irreversible Binding to the Colchicine Site of Tubulin. *Arch. Biochem. Biophys.* **1999**, *370*, 51.

(24)   Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P. M.; Jia, J.; Sherman, D. H.; Gerwick, W. H. Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2004**, *67*, 1356.

(25)   Gu, L.; Jia, J.; Liu, H.; Håkansson, K.; Gerwick, W. H.; Sherman, D. H. Metabolic Coupling of Dehydration and Decarboxylation in the Curacin A Pathway: Functional Identification of a Mechanistically Diverse Enzyme Pair. *J. Am. Chem. Soc.* **2006**, *128*, 9014.

(26)   Calderone, C. T. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* **2008**, *25*, 845.

(27)   Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'Connor, S. E.; Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **2005**, *436*, 1191.

(28)   Vaillancourt, F. H.; Yin, J.; Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII α-ketoglutarate- and O2-dependent halogenase. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10111.

(29)   Galonić, D. P.; Vaillancourt, F. H.; Walsh, C. T. Halogenation of Unactivated Carbon Centers in Natural Product Biosynthesis: Trichlorination of Leucine during Barbamide Biosynthesis. *J. Am. Chem. Soc.* **2006**, *128*, 3900.

(30)   Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; Garneau-Tsodikova, S.; Walsh, C. T. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. *Chem. Rev.* **2006**, *106*, 3364.

(31)   Gu, L.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Hakansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459*, 731.

(32)   Kaczanowski, S.; Zielenkiewicz, P. Why similar protein sequences encode similar three-dimensional structures? *Theor. Chem. Acc.* **2010**, *125*, 643.

(33)   Kosloff, M.; Kolodny, R. Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins* **2008**, *71*, 891.

(34)   Akey, D. L.; Gehret, J. J.; Khare, D.; Smith, J. L. Insights from the Sea: Structural Biology of Marine Polyketide Synthases. *Nat. Prod. Rep.* **2012**, *29*, 1038.

(35)   Khare, D.; Wang, B.; Gu, L.; Razelun, J.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L. Conformational switch triggered by α-ketoglutarate

in a halogenase of curacin A biosynthesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14099.

(36)   Geders, T. W.; Gu, L.; Mowers, J. C.; Liu, H.; Gerwick, W. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L. Crystal structure of the ECH2 catalytic domain of CurF from Lyngbya majuscula. Insights into a decarboxylase involved in polyketide chain beta-branching. *J. Biol. Chem.* **2007**, *282*, 35954.

(37)   Persson, B.; Hedlund, J.; Jörnvall, H. Medium- and short-chain dehydrogenase/reductase gene and protein families: The MDR superfamily. *Cell Mol. Life Sci.* **2008**, *65*, 3879.

(38)   Persson, B.; Zigler, J. S.; Jörnvall, H. A Super-Family of Medium-Chain Dehydrogenases/Reductases (MDR). *Eur. J. Biochem.* **1994**, *226*, 15.

(39)   Nordling, E.; Jörnvall, H.; Persson, B. Medium-chain dehydrogenases/reductases (MDR). *Eur. J. Biochem.* **2002**, *269*, 4267.

(40)   Kwan, D. H.; Leadlay, P. F. Mutagenesis of a Modular Polyketide Synthase Enoylreductase Domain Reveals Insights into Catalysis and Stereospecificity. *ACS Chem. Biol.* **2010**, *5*, 829.

(41)   Kwan, D. H.; Sun, Y.; Schulz, F.; Hong, H.; Popovic, B.; Sim-Stark, J. C. C.; Haydock, S. F.; Leadlay, P. F. Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **2008**, *15*, 1231.

(42)   Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **2012**, *8*, 615.

(43)   Ames, B. D.; Nguyen, C.; Bruegger, J.; Smith, P.; Xu, W.; Ma, S.; Wong, E.; Wong, S.; Xie, X.; Li, J. W. H.; Vederas, J. C.; Tang, Y.; Tsai, S.-C. Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11144.

(44)   DelProposto, J.; Majmudar, C. Y.; Smith, J. L.; Brown, W. C. Mocr: A novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Express. Purif.* **2009**, *63*, 40.

(45)   Stols, L.; Gu, M.; Dieckman, L.; Raffen, R.; Collart, F. R.; Donnelly, M. I. A New Vector for High-Throughput, Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site. *Protein Express. Purif.* **2002**, *25*, 8.

(46)   Khare, D.; Hale, W. A.; Gu, L.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L. Crystal Structure of Enoyl Reductase Domain from curacin and jamaicamide Pathways: From Cyclopropanation to Enoyl Reduction. *In Manuscript.* **2015**.

(47)   Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry* **2006**, *45*, 12756.

(48)   Hedlund, J.; Jornvall, H.; Persson, B. Subdivision of the MDR superfamily of medium-chain dehydrogenases/reductases through iterative hidden Markov model refinement. *BMC Bioinformatics* **2010**, *11*, 534.

(49)   Knoll, M.; Pleiss, J. The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence–structure–function relationship. *Protein Sci.* **2008**, *17*, 1689.

Chapter 5

Phosphopantetheine IR Absorptivity for Data Independent Selective Identification of Natural Product Biosynthetic Pathways

5.1 Introduction

Bottom-up or shotgun proteomics has transformed into a widespread practice, in which identifying and quantifying proteins and peptides has become routine[1-3]. Most proteomics approaches involve a data-dependent liquid chromatography tandem mass spectrometry (LC/MS/MS) procedure, in which peptides from proteolytic digestion of the protein sample are selected for fragmentation in order of abundance as they elute from the LC column(s). Modern mass spectrometers typically acquire one precursor ion mass spectrum followed by 5-10 peptide MS/MS spectra. In order to avoid fragmenting the same peptides multiple times, dynamic exclusion is employed in which an m/z ratio chosen for fragmentation populates an exclusion list for a set period of time. For complex samples such as whole cell proteomes, these procedures may include additional upstream clean-up or enrichment steps prior to LC/MS/MS[4-6]. Although shotgun proteomics has demonstrated its success, this method cannot completely map a proteome; rather, it can only identify the most abundant peptides in a complex mixture. One study demonstrated that approximately 100,000 peptides elute in a traditional bottom-up LC/MS/MS run, but only about 16% of these peptides are detected and analyzed[7]. Evolving instrumentation improvements, including sensitivity increases, are not yet powerful enough to sequence each peptide in a complex mixture[8,9]. Fourier transform (FT) mass spectrometers are

particularly vulnerable to this shortcoming of data-dependent analysis, as their trademark ultra-high resolution is directly proportional to the length of the time domain signal, thus, in order to maintain performance, they have lower throughput compared to other mass spectrometers[10,11]. One strategy for circumventing issues associated with data dependent acquisition (DDA) is to increase the dynamic range of the sample, primarily accomplished with fractionation or enrichment[8,12]. Examples of such strategies include subcellular fractionation, affinity enrichment, 2D gel electrophoresis, isoelectric focusing and cation or anion exchange methods such as MudPIT[13-17].

Many labs have recognized the shortcomings of data-dependent analysis and have developed alternative approaches, including elimination of MS/MS, or data independent LC/MS/MS methods (Table 5.1). Data independent acquisition (DIA) procedures produce product ion spectra without first selecting a precursor ion, i.e., all co-eluting peptides are fragmented together[12]. The accurate mass and corresponding retention time (AMRT) method was developed for quantitative label-free proteomics[18] and is similar to the accurate mass and time (AMT) tag concept that compares accurate masses and their corresponding LC elution times to a custom database[19]. This latest iteration of the AMT tag approach is an improvement of a previous AMT tag strategy that only utilized accurate mass. In this former implementation, "AMT" was an abbreviation for "accurate mass tag"[20]. The improved AMT tags are unique to mass spectral peaks 88% of the time and can identify thousands of peaks in a single LC/MS run[19]. Although the AMT tag is best suited to an FT mass spectrometer to ensure high mass accuracy, it is also amenable to time-of-flight instruments[19]. The main disadvantage of the AMT tag approach is the requirement for a custom database.

The precursor acquisition independent from ion count (PAcIFIC) DIA method increases dynamic range of complex samples by repeatedly analyzing a sample via multiple LC/MS/MS runs at specific but different m/z ranges, simultaneously fragmenting all precursor ions within the set range, typically a few Daltons[21]. Product ions can be observed even when a precursor ion was not detected. With this setup, PAcIFIC has been shown to identify over 2,000 proteins and up to 13,000 unique peptides from a whole cell lysate. Although the numbers of identifications are unprecedented, the data collection time is about five days[21]. The time

consuming data acquisition in PAcIFIC is an obvious flaw that was later addressed by utilizing higher throughput instruments to expedite the process, shortening the acquisition time to two or three days[11]. Another PAcIFIC flaw was the lack of quantitative data, which was addressed by including isobaric tagging[11]. Although the data collection duration is still very long, PAcIFIC can detect proteins across the entire measurable dynamic range.

Another LC/MS/MS DIA approach is the MS[E] method, introduced by Waters for their ultrahigh pressure LC (UPLC)/Synapt ion mobility-quadrupole time-of-flight (TOF) instruments. In this technique, the collision cell rapidly switches between low and high collision energies, producing alternating precursor and fragment ion spectra, both with high mass accuracy[22]. The duty cycle is very high, ensuring that both MS and MS/MS data are obtained for each narrow peak in the UPLC chromatogram. There is no quadrupole isolation; however low collision energy precursor ion spectra are matched with their associated high collision energy product ion spectra through sophisticated computer algorithms based on accurate retention time. Gas-phase ion mobility separation can be employed to further separate co-eluting peptides. This method has successfully been applied to phosphopeptides, which are generally dynamic and of low abundance[23]. MS[E] is an accurate method and has been utilized for comparing the reproducibility of two enrichment workflows for low abundance proteins[24]. The main disadvantage of MS[E] is that it is only available on a single commercial platform. The "all ion Fragmentation" (AIF) technique, designed for Orbitrap instruments, is similar to MS[E]. In this method, an MS scan is first analyzed in the orbitrap, followed by analysis of all product ions from high-energy collisional dissociation (HCD)[25]. With AIF, 48 proteins were detected in a HeLa cell lysate; thus the dynamic range is much lower than other DIA methods.

Another commercially available DIA technique is AB Sciex's SWATH acquisition on triple TOF mass spectrometers. SWATH is also applicable to other platforms such as the Orbitrap and was first designed by the Aebersold lab. SWATH fragments all ions in the 400-1200 m/z range by repeatedly acquiring 32, 25 Da wide windows during a single LC injection with a 1 Da overlap[26]. Thus, SWATH is similar to PAcIFIC, but has larger isolation windows. Unlike PAcIFIC, the data analysis portion involves a targeted data extraction strategy, in which product ions are compared to a spectral library containing information on product ion abundance and

chromatographic retention times[26]. Compared with other DIA strategies, SWATH-MS has a relatively high duty cycle and high dynamic range, though the former attribute is not as high as for MS^E and the latter is not as high as in PAcIFIC[26]. The data analysis strategy has recently been improved for SWATH-MS through automation[27]. SWATH is a modification of a previous method originating in the Yates lab. This method fragments ions in windows of 10 or 20 Da with no overlapping m/z values[28]. The Yates method utilizes MudPIT and metabolic labeling for increased chromatographic resolution and quantification, respectively. Although the Yates method is very similar to SWATH, it suffers from a slightly lower throughput and a lower dynamic range due to the difference in instrumentation and the comparative lack of data mining. Additionally, the Yates method identified a much lower number of proteins and peptides than similar methods with wide isolation windows although all methods analyzed whole cell lysates.

One of DIA's main advantages is also a disadvantage: the rich and complex data sets it produces renders data analysis challenging and time consuming. The OpenSWATH program adds a huge advantage, as it is applicable for multiple DIA data sets across different vendors and greatly simplifies data analysis. XDIA is an alternative computational strategy, which combines two MS/MS techniques, electron transfer dissociation and collision-induced dissociation (CID), to deconvolute complex spectra[29]. DeMux is another program designed to deconvolute complex data sets resulting from DIA, however, it is only applicable to data from ion trap mass spectrometers[30]. Upon designing DeMux, the authors found that data independent MS/MS provides more reproducible results, a necessary aspect for proteomics experiments. DIA-Umpire is a new program that enables untargeted peptide identification from DIA-derived data, using spectral libraries[31]. MS/MS data are matched against these spectral libraries. Quantitation can be performed with either precursor ions or fragment ions intensities. Additionally, DIA-Umpire is open-source and is applicable to numerous DIA strategies.

Method	Description	Max # proteins/peptides (original paper)
PAcIFIC	Fragments all peptides in 2.5 Da	2320 proteins, 13,049 peptides

	windows, overlap of 1.5 m/z in windows	
MS[E]	Alternates low collision energy scan (MS) and high collision energy (MS/MS) scan	N/A metabolites - not whole cell lysate
All Ion Fragmentation	MS scan in orbitrap followed by MS/MS scan in HCD cell	45 proteins, 348 peptides
SWATH-MS	Fragments all peptides in 25 Da windows, 1 m/z overlap in windows	original paper cites 4 log dynamic range, SWATH website cites over 3000 proteins[32]
DIA- Yates lab	Fragments all peptides in 10 or 20 Da windows, no overlap of m/z in windows	599 proteins, 1933 peptides

Table 5.1. A summary of various DIA approaches.

Data independent approaches to complex biological samples increase the prospects for detection of relevant peptides or proteins of low abundance. We are particularly interested in applying DIA techniques to cell extracts from natural product biosynthetic organisms. Such organisms express large polyketide synthases (PKSs) and/or non-ribosomal peptide synthetases (NRPSs), responsible for synthesizing bioactive natural products. These proteins include acyl carrier protein (ACP) and peptidyl carrier protein (PCP) domains, containing a phosphopantetheine (Ppant) post-translational modification (PTM). Via this PTM, natural product intermediates are covalently tethered to the biosynthetic protein modules. Without this PTM, the proteins are inactive, or in apo states, whereas, with Ppant attached, the proteins are considered active, or in holo states[33,34]. Because these large enzymes can easily be megadaltons in combined size,[33,34] it is highly challenging to detect active site peptides in a bottom-up approach. In addition, many natural product producers have unsequenced genomes and/or have symbiont origin, presenting additional challenges. Although an untargeted conventional DDA approach is possible, for example, our research group demonstrated detection of tryptic peptides from the ET-743 (a molecule approved as an anti-cancer drug in Europe) symbiont biosynthetic pathway in a metaproteomic sample, this method is akin to pulling a needle from a haystack[35]. Therefore, a targeted approach is desired, particularly for identifying and interrogating active site peptides. One example of a targeted approach to discovering novel natural products and their associated biosynthetic pathways is "proteomic investigation of secondary metabolism" (PrISM)[36], based on the gas-phase lability of

the Ppant modification[37]. This DDA workflow, which can be combined with digestion of only high molecular weight gel bands for additional specificity, selectively identifies ACP/PCP active site peptides, based on two characteristic PPant ejection product ions; m/z 261.1267 and 359.1036, resulting from bond cleavage on either side of the Ppant phosphate group. However, this approach only searches for 'holo' active site peptides and will miss any Ppantylated peptide with an attached substrate. Also, although a DIA implementation has been proposed[38], PrISM typically uses DDA and thus comes with all of the associated disadvantages. Another example of a targeted approach for biosynthetic pathway discover is orthogonal active site identification system (OASIS), which utilizes chemical probes to target either thioesterases, the protein domains found in the terminal PKS/NRPS module, or an ACP/PCP[39]. Although OASIS is a targeted approach, it has not yet been tested for complex samples and has only been implemented on low resolution mass spectrometers. Additionally, OASIS involves DDA and conventional protein databases, which bias the data towards genomes of sequenced organisms. A true unbiased targeted approach in this area is lacking.

Inspiration for a method to target unsequenced genomes for discovery of novel biosynthetic pathways and natural products originates from a selective gas-phase dissociation approach for phosphopeptides via infrared multiphoton dissociation (IRMPD)[40,41]. The CO_2 laser most frequently used for IRMPD emits photons at a wavelength of 10.6 μm, at which phosphate groups strongly absorb[42]. Thus, phosphorylated peptides dissociate readily upon IR irradiation while unphosphorylated peptides do not[40,41]. Research has shown that tyrosine phosphorylation is more stable than serine or threonine phosphorylation upon CID[43], however, with IRMPD, tyrosine phosphopeptides dissociate faster than serine phosphopeptides[44]. The Brodbelt lab successfully coupled LC to IRMPD-MS in a quadrupole ion trap mass spectrometer to separate several mixtures containing unphosphorylated and phosphorylated peptides[40]. Unmodified peptides fluctuate in intensity between 6% and 29% of their original abundances, whereas phosphorylated peptides dissociate between 96% and 100%. This approach was also utilized by the same lab for design of chromogenic cross-linkers, containing an ethyl phosphate group in the middle of the cross-linker for protein structural analysis[45]. Because the Ppant modification also contains a secondary phosphate, we hypothesized that IRMPD could be used

in a data independent manner for targeted natural product biosynthetic pathway discovery. Here, this universal approach, which operates well in both positive and negative ion mode, is presented for selective detection of phosphopantetheinylated peptides, regardless of their covalent state, i.e., lacking or carrying substrate or product.

5.2 Experimental

Proof-of-principle experiments were conducted with a mixture of unphosphorylated peptides (bradykinin, angiotensin I, and occasionally substance P) and phosphorylated peptides DAM 1 (SFVLNPTNIGM-pS-KSSQGHVTK), NPF (KRS-pY-EEHIP), and the tyrosine phosphorylated peptide (H-TSTEPQ-pY-QPGENL-NH$_2$). Peptides were purchased from Sigma-Aldrich, except for the tyrosine phosphorylated peptide which was purchased from Millipore. Data independent LC/FT-ICR MS/MS was performed with a hydrophilically endcapped C18 column (Phenomenex Synergi Hydro 80 Å 4 µm 1 x 150 mm). A non-linear gradient of 2-98% acetonitrile with 0.1% formic acid over ninety minutes was generated with an Agilent 1100 HPLC coupled to a 7 Tesla Apex or SolariX Q FT-ICR mass spectrometer (Bruker Daltonics). The gradient was as follows (with isocratic elution between 40 and 50 min): 0 (98,2), 20 (70,30), 40 (50,50), 50 (50,50), 55 (30,70), 70 (2,98). Values are provided as time (%A, %B) over a total run time of 95 min. Flow was at 50 µl/min and was diverted for the first 5 min of the run. HPLC solvent A was 0.1% formic acid (ThermoFisher Scientific) in HPLC-grade water (ThermoFisher Scientific) and solvent B was 0.1% formic acid in acetonitrile (ThermoFisher Scientific). IRMPD was performed with a 25 W CO$_2$ laser (Synrad) between 40% and 60% power with a 100 to 150 millisecond irradiation time.

The CurM ACP from the Curacin biosynthetic pathway[46] is 119 amino acids long, corresponding to 13,379 Da. The protein's concentration was 50 µM and it was kept in a buffer containing 10 µM of *Streptomyces verticillus* ATCC15003 (SVP), 10 mM of MgCl$_2$ and 100 mM of Tris at pH 7.5. The protein was expressed and purified by a member of the Sherman Lab at the University of Michigan and fractions were collected between 50 and 55 minutes for the apo and holo protein, respectively. Apo and holo ACPs were digested with sequencing grade trypsin at 37 degrees Celsius for 16 hours. The LC gradient was identical to the gradient for the proof-of-

principle experiment. IRMPD was performed with a 25 W CO_2 laser at 50% power with a 100 ms irradiation time.

PikAIII from the pikromycin biosynthetic pathway[47,48] is 1,562 amino acids long, corresponding to 163,593 Da. The dimeric Pik module was expressed, purified and digested by a member of the Smith lab at the University of Michigan. PikAIII was digested with sequencing grade trypsin for 15 minutes at a 4:1 protein:trypsin ratio and quenched with formic acid. The final protein concentration was 1 µM. The LC gradient was identical to the gradient for the proof-of-principle phoshopeptide experiment. IRMPD was performed with a 25 W CO_2 laser at 40% power with a 100 millisecond irradiation time. External calibration for all experiments utilized HP-mix (Agilent) or 0.1 mg/mL sodium trifluoroacetate clusters (Sigma-Aldrich). The intelligent pulse sequence organizer (IPSO) was re-programmed to allow the d12 command to trigger IRMPD rather than electron capture dissociation, thus firing the laser in every other LC/MS/MS scan on the SolariX instrument. Quadrupole isolation was turned off in all cases to allow DIA.

5.3 Results and Discussion

5.3.1 Implementation of DIA LC/IRMPD: Proof-of-Principle and Optimization

Initially, all experiments were conducted on the older Apex-Q FT-ICR mass spectrometer. Apex does not have the IPSO program; therefore, peptides were compared from separate LC/MS runs, i.e, one run with the laser off and one with the laser on. On the Apex instrument, the laser was tuned for optimum power and irradiation time through direct injection of the peptide mixture, as shown in Table 5.2.

Laser Power	40%	40%	40%	40%	40%	40%	40%	50%	60%	60%
Irradiation Time (s)	0	0.05	0.10	0.15	0.20	0.25	0.30	0.15	0.15	0.10

Bradykinin (A)	100%	94%	95%	101%	82%	79%	49%	92%	90%	102%
Angiotensin I (B)	100%	89%	91%	90%	74%	71%	36%	79%	82%	88%
Substance P (C)	100%	90%	92%	103%	61%	59%	30%	77%	63%	94%
DAM1 (M)	100%	95%	82%	68%	27%	27%	17%	37%	29%	79%
pY peptide (N)	100%	82%	36%	26%	15%	14%	11%	24%	21%	34%

Table 5.2. Different laser powers and irradiation times were tested to find the optimal condition for selective IRMPD. Shown in the table are the relative abundance of the selected ions (A $[A + 2H]^{2+}$, B $[B + 2H]^{2+}$, C $[C + 2H]^{2+}$, M $[M + 3H]^{3+}$, and N $[N + K + H]^{2+}$) following IR irradiation. A laser power of 60% and an irradiation time of 0.15 s were chosen as the abundance of phosphorylated species (M, N) were significantly reduced upon IR irradiation whereas the unphosphorylated species (A, B, C) were not. Data collected by a summer undergraduate student under my supervision, Yilun Li.

Following optimization of the laser power and irradiation time, the peptide mix was subjected to LC/FT-ICR MS with and without laser irradation. This proof-of-principle experiment was successful in that all five peptides were detected. Peak areas from extracted ion chromatograms (EICs) with and without IR irradiation remained within 43 percent of each other for the unphosphorylated peptides. The tyrosine phosphopeptide's EIC peak area decreased by about 96 percent, whereas the DAM1 EIC peak area completely disappeared upon IR irradiation, thus demonstrating selectivity of IRMPD for phosphopeptide dissociation (Figure 5.1).
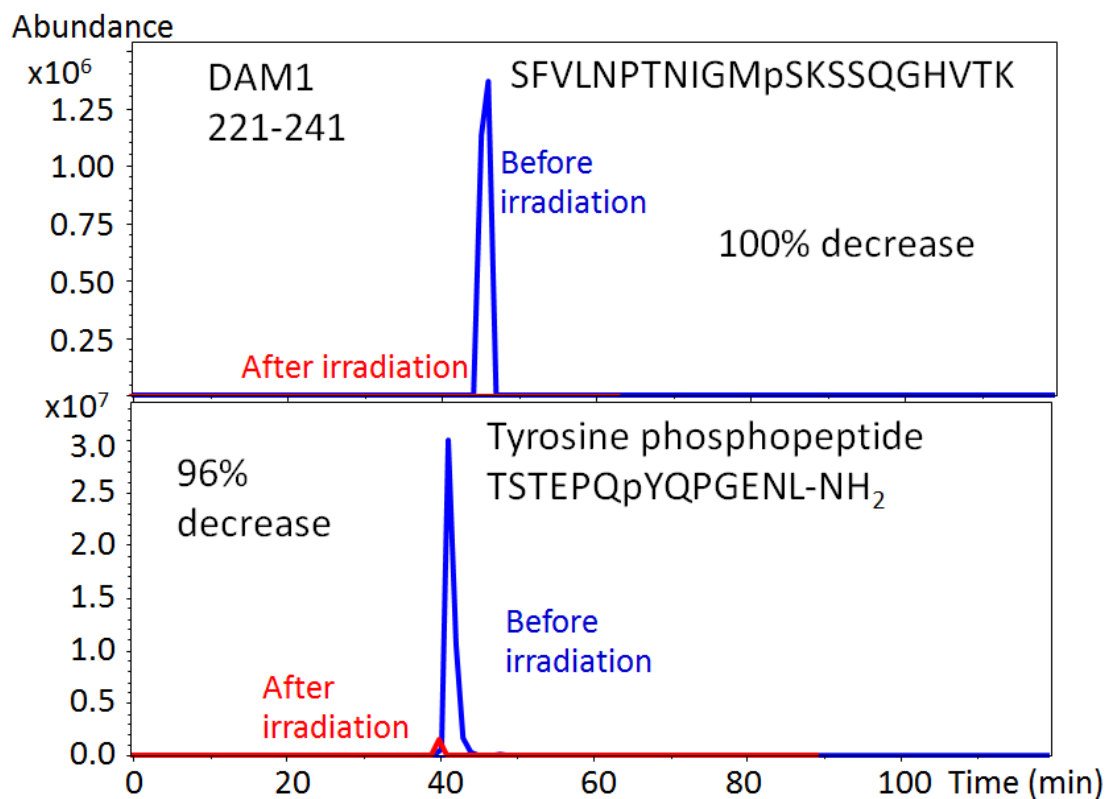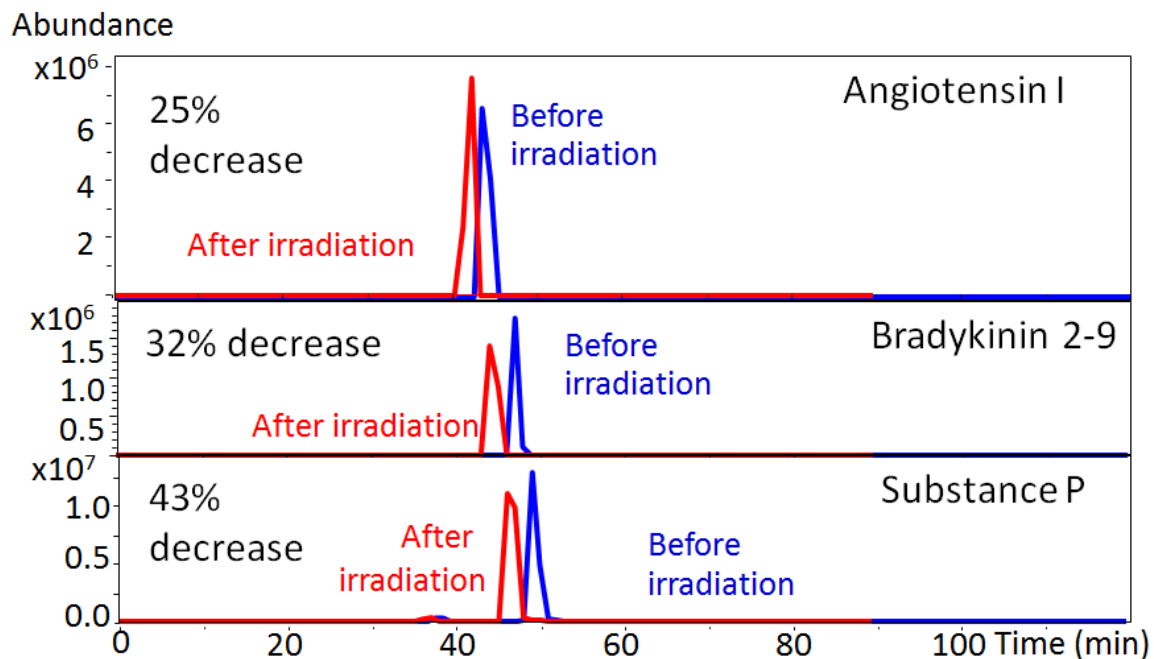
Figure 5.1. Extracted ion chromatograms (EICs) of the monoisotopic peak of the unphosphorylated and phosphorylated peptides without (blue) and with (red) IR irradiation from separate LC/MS runs on an Apex-Q FT-ICR mass spectrometer. Unmodified peptides' peak areas decreases by less than fifty percent upon irradiation, whereas phosphorylated peptides decreased greater than 95 percent.

An added advantage of this DIA method is that it is insensitive to ion polarity (Figure 5.2). The by far most common way of analyzing peptides in a mass spectrometer is in positive ion mode, in which the ion source forms positively charged analytes. However, acidic molecules, such as phosphorylated peptides and proteins, can be difficult to ionize in positive ion mode[49]. Analyzing acidic analytes in negative ion mode can increase the sensitivity and limit of detection[50,51]. Therefore, it is important to develop and operate complementary fragmentation techniques and data-independent strategies in both polarities. Operating this data independent method in both positive and negative ion modes may increase the number of identified Ppant-containing peptides as complementary ionization is known to occur between the two modes.
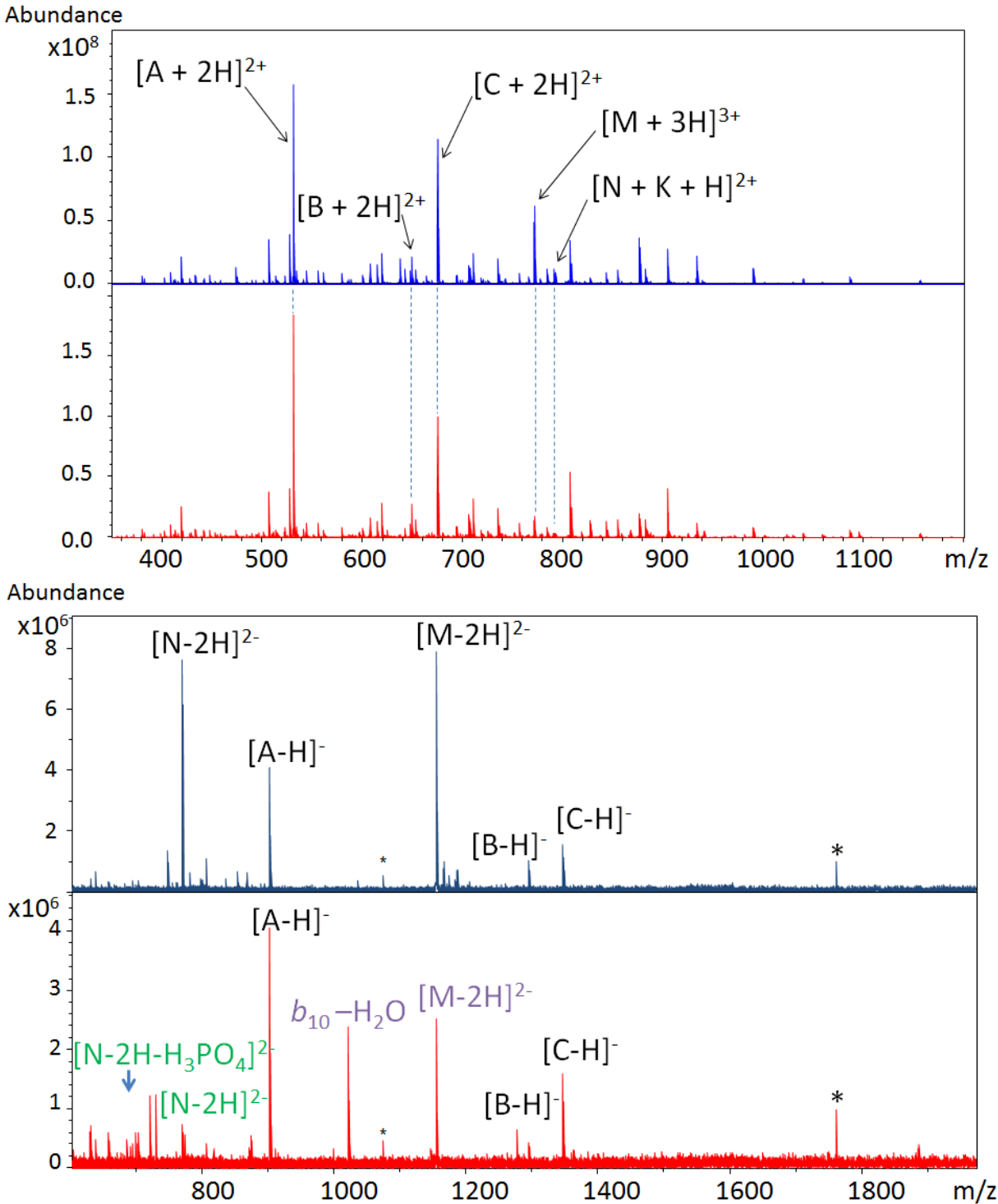
Figure 5.2. The identities of peptides A, B, C, M, and N are listed in Table 5.2. The unmodified peptides A, B, and C do not dissociate efficiently upon IR irradiation in neither positive (top), nor negative (bottom) ion mode. Phosphorylated peptides nearly completely dissociate in both ionization modes. The b_{10} - H_2O fragment ion colored in purple is from the M peptide, also

labeled in purple. The N peptide loses phosphoric acid upon irradiation; both precursor and fragment ions are color coded in green.

Although the older Apex instrument yielded satisfactory data, the LC peaks were undersampled and thus have triangular shapes, as seen in Figure 5.1. The observed slight shift in retention time with and without IR irradiation is due to the separate LC/MS runs. Furthermore, between seemingly identical LC/MS runs, there may be differences in ionization efficiency and ion transport; therefore, it would be ideal to implement our DIA approach in a single LC/MS run with the IR laser firing every other scan. This improvement would allow for twice the number of runs during a set duration. Therefore, the IR laser was moved to the newer SolariX mass spectrometer and a modified pulse program was written to allow for LC/MS with the laser firing every other scan. The SolariX instrument is more sensitive than the Apex instrument and writes data to the hard drive faster, thus allowing for more data acquisition points in a given time period. Moreover, SolariX has a more advanced pulse programming capability (IPSO) than Apex. The default MS/MS activation methods for LC/MS/MS on SolariX are CID and electron capture dissociation (ECD) with the d12 parameter controlling the electron irradiation time. The IPSO was manipulated to instead trigger the infrared laser during d12. A 27 pin connector for external triggering was connected to the laser. This SolariX implementation resulted in more reliable retention times and improved peak shapes as well as a more accurate comparison of the peak shapes prior to and after irradiation, as compared with the Apex implementation. For quantitation purposes, EICs were created for all observed isotopic peaks above a signal to noise ratio of three. On SolariX, the laser power and irradiation time were reduced to 40% and 100 ms, respectively, for optimal conditions. When more than on peptide charge state was observed, peaks areas were normalized to charge and summed. The NPF peptide dissociated by 93% and the pY peptide dissociated by 99%. The unmodified peptides, bradykinin and angiotensin I dissociated by 4% and increased by 12%, respectively, as shown in Figure 5.3.
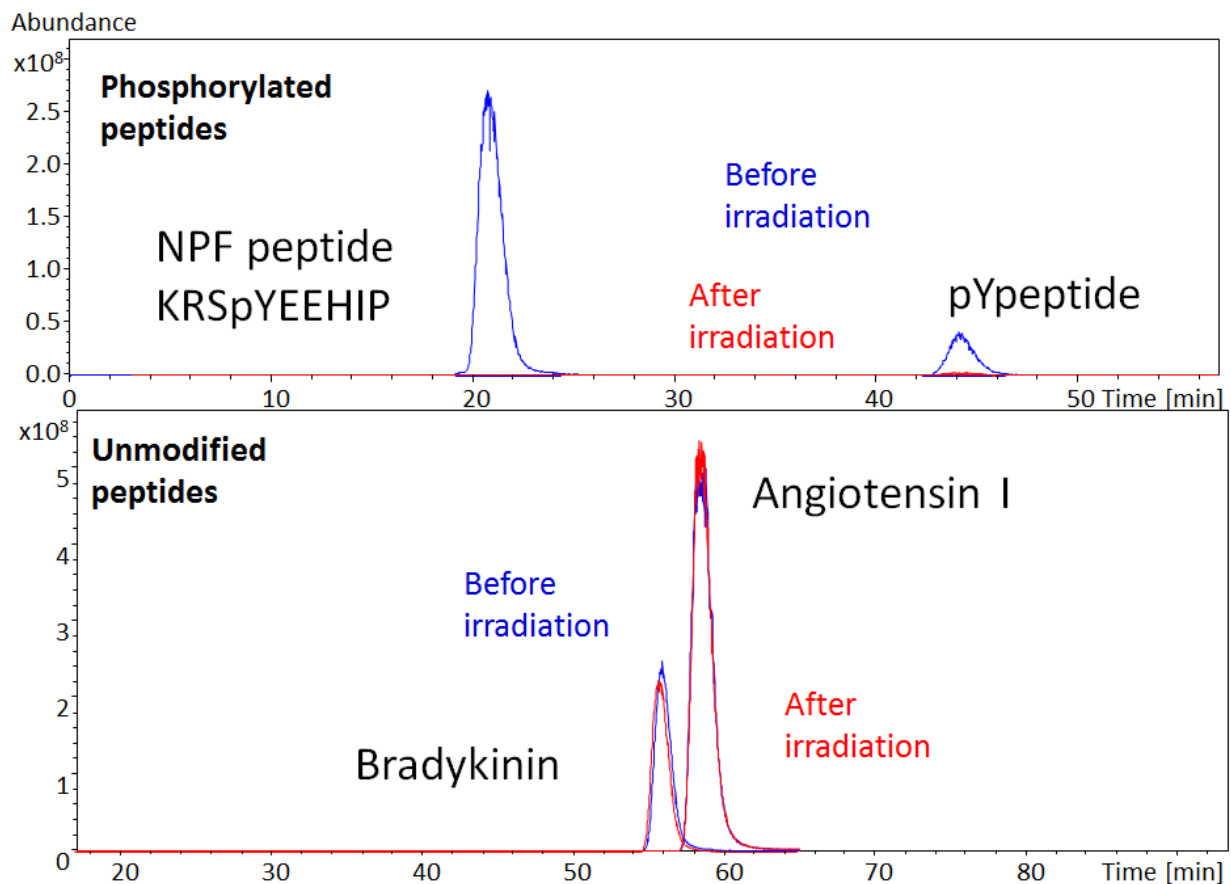
Figure 5.3. DIA LC/FT-ICR MS of a mixture of four peptides, half unmodified, half phosphorylated, on a newer SolariX Q-FT-ICR mass spectrometer with the laser firing every other scan. This experiment resulted in smoother peak shapes than the data in Fig. 5.1 with retention times completely overlapping due to the MS and IRMPD data being collected in the same run. Overall, similar results ensued, with unmodified peptides partially dissociating, whereas phosphorylated peptides dissociated nearly to completion.

5.3.2 DIA LC/IRMPD of an acyl carrier protein

After the proof-of principle experiments were successful, a single ACP was tryptically digested and subjected to data-independent LC/IRMPD to explore whether the Ppant-modified active site peptide would show analogous behavior to the phosphopeptides above and to the previously published phosphate-containing cross-linker[45]. The CurM ACP from the curacin A hybrid PKS/NRPS biosynthetic pathway was selected for this experiment, which was carried out on the older Apex instrument, due to availability and laser location at the time. Ten tryptic peptides, including one that contained Ppant, were detected in LC/MS as well as in LC/IRMPD

analysis. Upon IR irradiation, the Ppant-containing peptide decreased in peak area by 88%, as determined from the EIC (Figure 5.4). The EIC peak area difference without and with IR irradiation for unmodified peptides ranged from 3 to 29 percent. This experiment demonstrated the selectivity of the DIA IRMPD method for a more complex sample and confirmed our hypothesis that the Ppant modification would show selective dissociation.
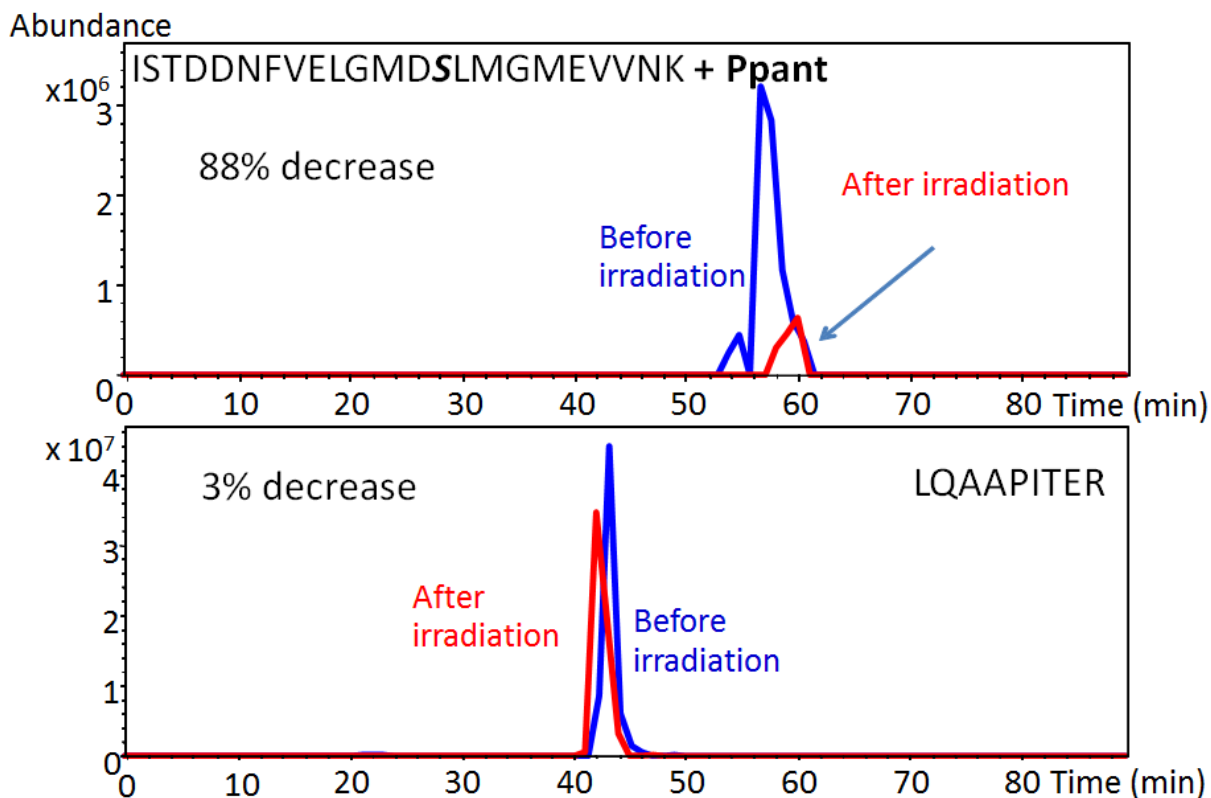
Figure 5.4. LC/FT-ICR MS without (blue) and with (red) IR irradiation. The top spectrum demonstrates the effectiveness of IRMPD for a phosphopantetheinylated peptide with a decrease in EIC peak area by 88%. An example unmodified peptide, LQAAPITER, only dissociated by 3%. In this tryptic digest, unmodified peptides dissociated 29% or less.

5.3.3 DIA LC/IRMPD of a polyketide synthase module

The selectivity of the established data-independent LC/IRMPD method was further validated for the fifth module from the pikromycin biosynthetic pathway, PikAIII. This module was selected due to familiarity with the pikromycin biosynthetic pathway. PikAIII consists of four domains, a ketosynthase, an acyl transferase, an ACP, and a ketoreductase. The application to an entire PKS module increases sample complexity from a single ACP (MW

13,379) to greater than 150 kDa. All PikAIII LC/IRMPD replicates were carried out on the SolariX mass spectrometer, ensuring higher quality chromatography data and higher sensitivity. For comparison, active site peptides from the ketosynthase and acyl transferase domains were selected, as well as other abundant unmodified peptides.
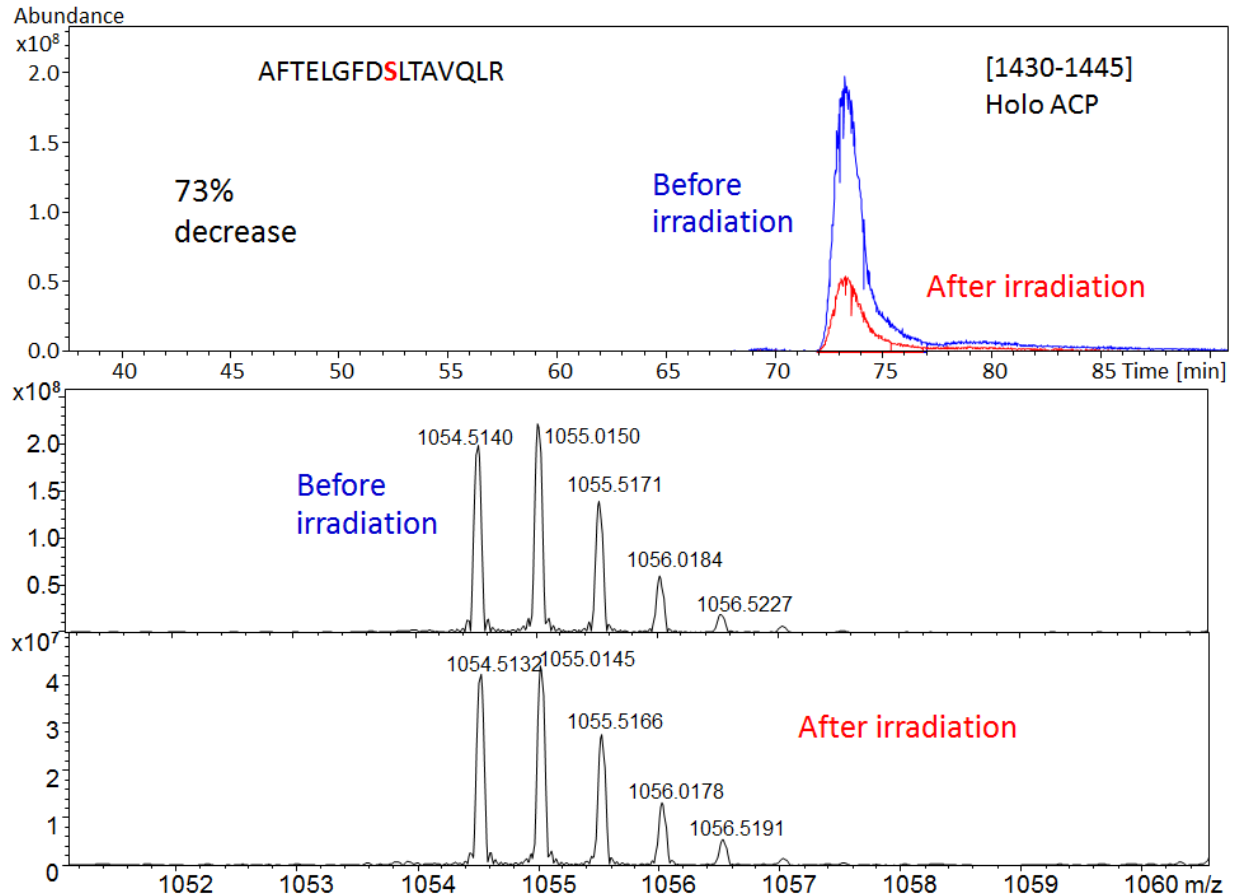
Figure 5.5. Extracted ion chromatograms of the active site holo ACP peptide without and with IR irradiation (top). FT-ICR mass spectra (bottom) verify the loss in signal abundance upon IR irradiation. The holo ACP active site peptide (containing Ppant) dissociates by 73%.
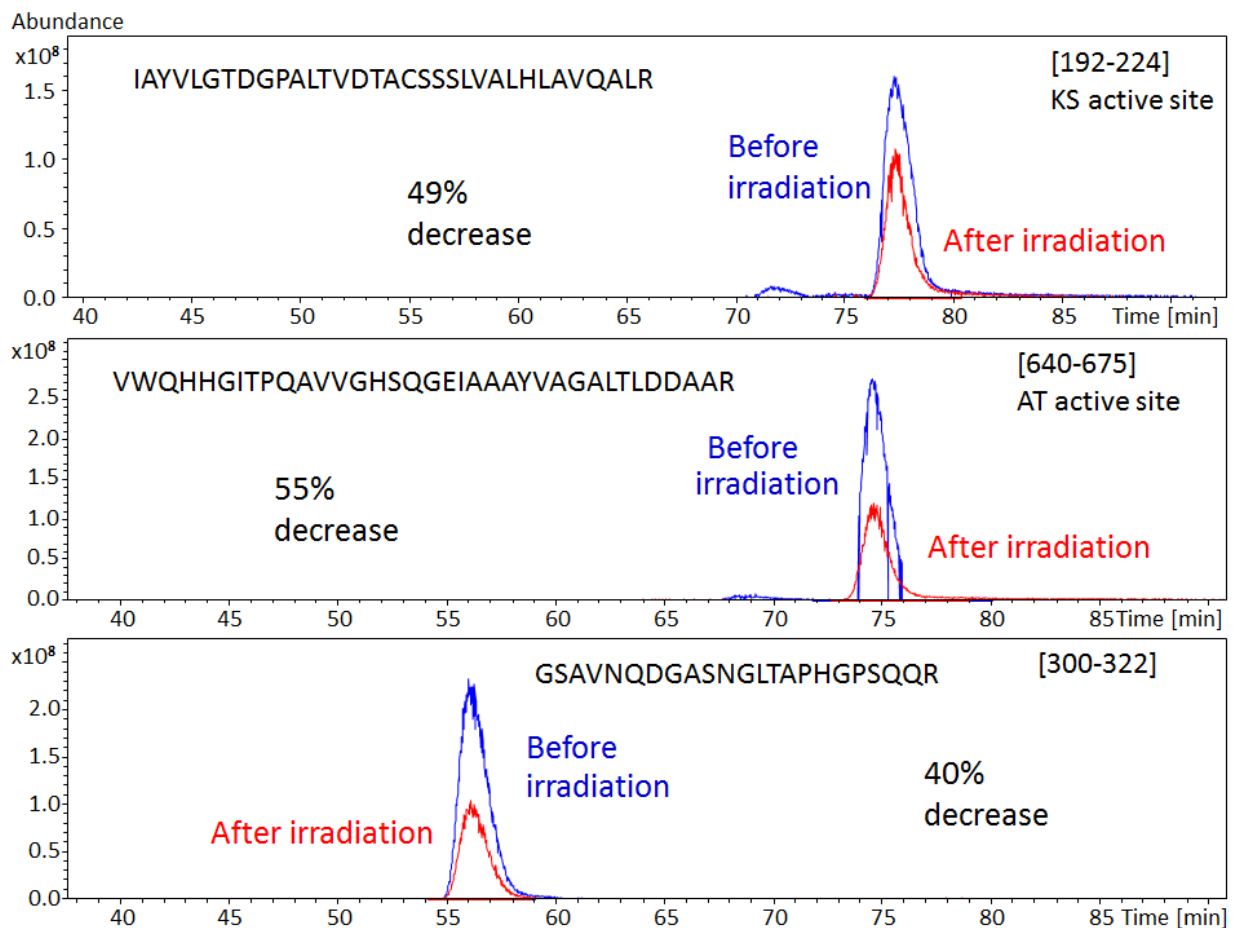
Figure 5.6. EICs of selected unmodified peptides from a PikAIII tryptic digest without (blue) and with (red) IR irradiation. Peak areas decrease less than that of holo ACP (Fig. 5.5). When multiple charge states were present, peaks were normalized to charge and summed. The acyl transferase and ketosynthase active site peptides are shown as well as a PikAIII non-active site peptide that was abundant in each replicate.

The IR laser was tuned and optimized prior to each LC/IRMPD run, data were collected with 55% power and 200 ms irradiation time, with 50% power and 150 ms irradiation time, and ultimately with 40% power and 100 ms irradiation, chosen as the optimal condition for PikAIII. Overall, the experiment was successful in that the holo ACP dissociated more readily than unmodified peptides, 73% (Figure 5.5). For relative quantitation, EICs for all isotopic peaks above a signal to noise ratio of three were created. Three selected unmodified peptides dissociated by 49%, 55%, and 40%, respectively (Figure 5.6). Thus, for the PikAIII digest, there was a smaller difference in dissociation percentage between unmodified and Ppantylated peptides than the previous experiments. When the laser power and irradiation time were

increased, the PPantylated peptide dissociated up to 93%, however, the unmodified peptides dissociated at a higher percent as well, particularly the acyl transferase active site peptide. Experiments to further optimize the laser power and/or irradiation time may improve selectivity between unmodified and Ppantylated peptides and should be carried out in the future.

5.4 Conclusions

A novel data independent LC/IRMPD approach has been developed to selectively detect phosphopantetheinylated peptides from PKS and NRPS biosynthetic pathways. Targeted and unbiased proteomic approaches are lacking in the natural product biosynthesis field and this method fills this gap. IRMPD selectively detects and dissociates phosphopantetheinylated peptides, allowing for eventual discovery of novel natural product biosynthetic pathways. This method was optimized and established in proof-of-principle experiments with a mixture of unmodified and phosphorylated peptides. Subsequently, the complexity of the sample was increased to a single ACP domain and further to a single module from a PKS biosynthetic pathway. Future work for further development and improvement of this method includes additional optimization for the PKS module and further increase in complexity to the overexpressed pikromycin biosynthetic pathway in an *E. coli* lysate and eventually lysates of *Streptomyces venezuelae*, the native pikromycin producer. This method will eventually be applied to organisms with unsequenced genomes, such as *Karenia brevis*, a marine dinoflagellate responsible for producing multiple natural products, including a potential respiratory therapeutic and a toxin responsible for the Florida red tides.

5.5 References

(1)   Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **2001**, *101*, 269.

(2)   Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198.

(3)   Bensimon, A.; Heck, A. J.; Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **2012**, *81*, 379.

(4)   Johnson, H.; Del Rosario, A. M.; Bryson, B. D.; Schroeder, M. A.; Sarkaria, J. N.; White, F. M. Molecular Characterization of EGFR and EGFRvIII Signaling Networks in Human Glioblastoma Tumor Xenografts. *Mol. Cell Proteomics* **2012**, *11*, 1724.

(5)   Hendriks, I. A.; D'Souza, R. C. J.; Yang, B.; Verlaan-de Vries, M.; Mann, M.; Vertegaal, A. C. O. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **2014**, *21*, 927.

(6)   Froehlich, J. W.; Vaezzadeh, A. R.; Kirchner, M.; Briscoe, A. C.; Hofmann, O.; Hide, W.; Steen, H.; Lee, R. S. An in-depth comparison of the male pediatric and adult urinary proteomes. *Biochim. Biophys. Acta* **2014**, *1844*, 1044.

(7)   Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 1785.

(8)   Chapman, J. D.; Goodlett, D. R.; Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **2014**, *33*, 452.

(9)   Law, K. P.; Lim, Y. P. Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Rev. Proteomics* **2013**, *10*, 551.

(10)   Amster, I. J. Fourier Transform Mass Spectrometry. *J. Mass Spectrom.* **1996**, *31*, 1325.

(11)   Panchaud, A.; Jung, S.; Shaffer, S. A.; Aitchison, J. D.; Goodlett, D. R. Faster, Quantitative, and Accurate Precursor Acquisition Independent From Ion Count. *Anal. Chem.* **2011**, *83*, 2250.

(12)   Canterbury, J.; Merrihew, G.; MacCoss, M.; Goodlett, D.; Shaffer, S. Comparison of Data Acquisition Strategies on Quadrupole Ion Trap Instrumentation for Shotgun Proteomics. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 2048.

(13)   Borner, G. H. H.; Hein, M. Y.; Hirst, J.; Edgar, J. R.; Mann, M.; Robinson, M. S. Fractionation profiling: a fast and versatile approach for mapping vesicle proteomes and protein–protein interactions. *Mol. Biol. Cell* **2014**, *25*, 3178.

(14)   Keilhauer, E. C.; Hein, M. Y.; Mann, M. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Mol. Cell Proteomics* **2015**, *14*, 120.

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

(16)   Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. Linking genome and

proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14440.

(17)   Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Stephenson Jr., J. L.; Bundy, J. L. Immobilized pH Gradient Isoelectric Focusing as a First-Dimension Separation in Shotgun Proteomics. *J. Biomol. Tech.* **2005**, *16*, 181.

(18)   Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G.-Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. *Anal. Chem.* **2005**, *77*, 2187.

(19)   Strittmatter, E. F.; Ferguson, P. L.; Tang, K.; Smith, R. D. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 980.

(20)   Lipton, M. S.; Paša-Tolić, L.; Anderson, G. A.; Anderson, D. J.; Auberry, D. L.; Battista, J. R.; Daly, M. J.; Fredrickson, J.; Hixson, K. K.; Kostandarithes, H.; Masselon, C.; Markillie, L. M.; Moore, R. J.; Romine, M. F.; Shen, Y.; Stritmatter, E.; Tolić, N.; Udseth, H. R.; Venkateswaran, A.; Wong, K.-K.; Zhao, R.; Smith, R. D. Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11049.

(21)   Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **2009**, *81*, 6481.

(22)   Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K. UPLC/MS[E]; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1989.

(23)   Blackburn, K.; Goshe, M. B. Challenges and strategies for targeted phosphorylation site identification and quantification using mass spectrometry analysis. *Brief. Funct. Genomic. Proteomic.* **2008**, *8*, 90.

(24)   Hakimi, A.; Auluck, J.; Jones, G. D. D.; Ng, L. L.; Jones, D. J. L. Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent LC-MS. *Proteomics* **2014**, *14*, 4.

(25)   Geiger, T.; Cox, J.; Mann, M. Proteomics on an Orbitrap Benchtop Mass Spectrometer Using All-ion Fragmentation. *Mol. Cell Proteomics* **2010**, *9*, 2252.

(26)   Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell Proteomics* **2012**, *11*, O111.016717.

(27)   Rost, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinovic, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmstrom, J.; Malmstrom, L.; Aebersold, R. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotech.* **2014**, *32*, 219.

(28)   Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Meth.* **2004**, *1*, 39.

(29)   Carvalho, P. C.; Han, X.; Xu, T.; Cociorva, D.; Carvalho, M. d. G.; Barbosa, V. C.; Yates, J. R. XDIA: improving on the label-free data-independent analysis. *Bioinformatics* **2010**, *26*, 847.

(30)   Bern, M.; Finney, G.; Hoopmann, M. R.; Merrihew, G.; Toth, M. J.; MacCoss, M. J. Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 833.

(31)   Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Meth.* **2015**, *12*, 258.

(32)   Sciex, A.; AB Sciex: Framingham, MA, 2014.

(33)   Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106*, 3468.

(34)   Sieber, S. A.; Marahiel, M. A. Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chem. Rev.* **2005**, *105*, 715.

(35)   Rath, C. M.; Janto, B.; Earl, J.; Ahmed, A.; Hu, F. Z.; Hiller, L.; Dahlgren, M.; Kreft, R.; Yu, F.; Wolff, J. J.; Kweon, H. K.; Christiansen, M. A.; Håkansson, K.; Williams, R. M.; Ehrlich, G. D.; Sherman, D. H. Meta-omic Characterization of the Marine Invertebrate Microbial Consortium That Produces the Chemotherapeutic Natural Product ET-743. *ACS Chem. Biol.* **2011**, *6*, 1244.

(36)   Bumpus, S. B.; Evans, B. S.; Thomas, P. M.; Ntai, I.; Kelleher, N. L. A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* **2009**, *27*, 951.

(37)   Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry* **2006**, *45*, 12756.

(38)   Chen, Y.; Ntai, I.; Ju, K.-S.; Unger, M.; Zamdborg, L.; Robinson, S. J.; Doroghazi, J. R.; Labeda, D. P.; Metcalf, W. W.; Kelleher, N. L. A Proteomic Survey of Nonribosomal Peptide and Polyketide Biosynthesis in Actinobacteria. *J. Proteome Res.* **2012**, *11*, 85.

(39)   Meier, J. L.; Niessen, S.; Hoover, H. S.; Foley, T. L.; Cravatt, B. F.; Burkart, M. D. An Orthogonal Active Site Identification System (OASIS) for Proteomic Profiling of Natural Product Biosynthesis. *ACS Chem. Biol.* **2009**, *4*, 948.

(40)   Crowe, M. C.; Brodbelt, J. S. Differentiation of Phosphorylated and Unphosphorylated Peptides by High-Performance Liquid Chromatography-Electrospray Ionization-Infrared Multiphoton Dissociation in a Quadrupole Ion Trap. *Anal. Chem.* **2005**, *77*, 5726.

(41)   Flora, J. W.; Muddiman, D. C. Selective, Sensitive, and Rapid Phosphopeptide Identification in Enzymatic Digests Using ESI-FTICR-MS with Infrared Multiphoton Dissociation. *Anal. Chem.* **2001**, *73*, 3305.

(42)   Correia, C. F.; Balaj, P. O.; Scuderi, D.; Maitre, P.; Ohanessian, G. Vibrational Signatures of Protonated, Phosphorylated Amino Acids in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130*, 3359.

(43)   Affolter, M.; Watts, J. D.; Krebs, D. L.; Aebersold, R. Evaluation of Two-Dimensional Phosphopeptide Maps by Electrospray Ionization Mass Spectrometry of Recovered Peptides. *Anal. Biochem.* **1994**, *223*, 74.

(44)   Flora, J. W.; Muddiman, D. C. Determination of the relative energies of activation for the dissociation of aromatic versus aliphatic phosphopeptides by ESI-FTICR-MS and IRMPD. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 121.

(45)   Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. Chromogenic Cross-Linker for the Characterization of Protein Structure by Infrared Multiphoton Dissociation Mass Spectrometry. *Anal. Chem.* **2008**, *80*, 4807.

(46)   Gu, L.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Hakansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459*, 731.

(47)   Xue, Y.; Zhao, L.; Liu, H.; Sherman, D. H. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: Architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12111.

(48)   Xue, Y.; Sherman, D. H. Biosynthesis and Combinatorial Biosynthesis of Pikromycin-Related Macrolides in Streptomyces venezuelae. *Metab. Eng.* **2001**, *3*, 15.

(49)   Janek, K.; Wenschuh, H.; Bienert, M.; Krause, E. Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1593.

(50)   Wilm, M.; Neubauer, G.; Mann, M. Parent Ion Scans of Unseparated Peptide Mixtures. *Anal. Chem.* **1996**, *68*, 527.

(51)   Carr, S. A.; Huddleston, M. J.; Annan, R. S. Selective Detection and Sequencing of Phosphopeptides at the Femtomole Level by Mass Spectrometry. *Anal. Biochem.* **1996**, *239*, 180.

Chapter 6

Conclusions and Future Directions

6.1 Dissertation Summary

Although each chapter in this dissertation tells a distinct story, each involves Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometric quantitation of dynamic peptides or proteins that change abundance or modifications readily. Chapter 2 explored the phosphorylation kinetics of Sic1, a protein in budding yeast. Phosphorylation of six of the nine sites triggers a switch-like behavior, resulting in degradation of the protein, which regulates the transition from the G1 to the S phase of the cell cycle[1,2]. A novel bottom up quantitative method was developed to determine the kinetics of each phosphorylation site. Internal standard peptides with Val to Leu or Leu to Val substitutions were tested for similar ionization efficiencies and served as internal standard peptides for the quantitative method. If a Val or Leu was not present in the Sic1 peptide of interest, a proline was labeled with ^{13}C and ^{15}N. Additionally, all spectra were collected in negative ion mode and the first three isotopes were quantified to ensure accuracy, particularly for low signal to noise ratio peptides. To establish the technique, a mutant protein was quantified, Sic1 2p, in which every phosphorylation site except T45 and S76 were mutated to an alanine. The two phosphorylation sites had different

phosphorylation rates, with S76 phosphorylating faster than T45. Phosphorylation rates for each known Sic1 phosphorylation site in the wildtype protein were determined with bottom-up FT-ICR mass spectrometry (MS). For the 33-50 peptide with two proline-directed phosphorylation sites, evidence for a third phosphorylation event was found, although the site of additional phosphorylation proved difficult to determine with tandem mass spectrometry (MS/MS) due to its relatively low abundance. Evidence for a total of ten phosphorylation events was also seen in previous top down experiments in our laboratory[3]. Each of the nine known sites on the wildtype protein has a different phosphorylation rate, with T45, S76, T5, and T33 phosphorylating faster than T2, S69, S80, T173 and S191. These findings agree with previously published data stating that the four aforementioned sites contribute more significantly to Sic1 degradation[1]. Future work for this project includes comparing the kinetic data with mathematical models and repeating the experiments *in vivo*.

Chapter 3 confirmed, identified, and quantified substrate- and product-bound active site peptides in the fifth module of the pikromycin biosynthetic pathway, the polyketide synthase (PKS) PikAIII. Changes in the small molecule occupancy of the different active sites in this PKS caused dramatic PKS structural changes, shown with cryo-electron microscopy (cryo-EM). The combination of cryo-EM and ultra-high resolution mass spectrometry provided greater insight into how these multienzyme complexes synthesize bioactive natural products. Ultra-high resolution mass spectrometry was necessary for these experiments, particularly to detect a mixture of one active site peptide with two different products bound, differing by only two hydrogens, i.e. 2 Da. Relative quantitation was performed by adding the peak areas of all detectable isotopes with a signal to noise ratio greater than three for all active site peptides. Additionally, the sample preparation and liquid chromatography (LC) steps were optimized by eliminating the reduction and alkylation of cysteine residues, choosing an LC column more appropriate for peptides, and gradient optimization, allowing for the detection of the ketosynthase active site peptide with bound substrate.

Chapter 4 investigated the effect of two enoyl reductase (ER) domains on a natural product intermediate in two pathways from the same organism, *Moorea producens*. The jamaicamide (Jam) pathway produces Jamaicamide A, a small molecule toxic to fish with sodium channel

blocking properties, whereas the curacin (Cur) pathway produces Curacin A, an anti-cancer agent[4]. The JamJ ER reduces a vinyl chloride, an ER's typical function, whereas the CurF ER produces a cyclopropane. Collision-induced dissociation (CID) was applied to the intact acyl carrier proteins (ACPs) excised from the JamJ and CurF PKSs to identify natural product intermediates generated by the JamJ and CurF ERs. High resolution ($\geq$ 70,000) was necessary to detect and identify natural product intermediates: the action of JamJ ER results in a mass difference of 2 Da for the intact protein. High resolution also confirms the MS/MS ejected intermediate product: isotopic fine structure is observed due to chlorine in the natural product intermediate. Point mutations near the ER active site were carried out to determine which amino acid acts as a proton donor for performing the ER's function. Both the JamJ and CurF ER mutants show a concentration dependence: 1 or 10 nM of the ER partially reduced a vinyl chloride intermediate or partially formed a cyclopropane, whereas 100 nM or greater completely reduced the vinyl chloride and mostly (>80%) formed the cyclopropane in the presence of NADPH. JamJ D273 and CurF Y58, Y103, and R253 appear to play significant roles in their respective active sites, although no single site is completely responsible for catalytic activity, confirming the hypothesis that a water-mediated pathway may be the proton donor necessary for catalytic activity[5,6]. Natural product intermediates were relatively quantified to determine the effects of ER mutations. Additionally, the JamJ R251K mutant was further altered to incorporate a structural loop from Cur, resulting in a novel cyclopropanase. The cyclopropanase activity was confirmed and quantified with mass spectrometry.

Chapter 5 strove to develop a novel data-independent acquisition method for discovery of novel PKS and non-ribosomal peptide synthetase (NRPS) pathways in unsequenced organisms. This approach is also novel in that it is a targeted approach toward the phosphopantetheine (Ppant) moiety, which is responsible for tethering and transferring natural product intermediates. Infrared multiphoton dissociation (IRMPD) at 10.6 μM selectively dissociates peptides with both primary and secondary phosphates[7,8]. Ppant, a post-translational modification of ACPs and peptidyl carrier proteins (PCPs) in NRPSs has a secondary phosphate, acting as a chromophore for the IR photons[9]. As proof-of-principle, a mixture of phospho- and non-phosphopeptides was subjected to LC/FT-ICR MS without and with IR irradiation.

Phosphopeptides decreased in peak area by 80-99 percent, whereas unmodified peptides decreased in peak area by less than twenty nine percent. Ten peptides, including one containing Ppant, were found in LC/FT-ICR MS analysis of tryptically digested CurM ACP from the curacin A biosynthetic pathway. With IRMPD, all unmodified peptides decreased in peak area by 30 percent or less, including some peptides that increased in peak area. With IRMPD, the Ppant-containing peptide decreased in peak area by 88 percent. The complexity of the mixture was increased to an entire module of a polyketide synthase, module five from the pikromycin biosynthetic pathway. Unmodified peptide peak areas decreased by 55 percent or less, whereas the phosphopantetheinylated peptide's peak area decreased by 73 percent. We are currently exploring the applicability of this method, which can be conducted in either positive or negative ion mode, for more complex NRPS/PKS mixtures.

6.2 Prospects for Future Work

6.2.1 Polyketide synthase natural product discovery

Bolstered by the success of the combination of cryo-EM and LC/FT-ICR mass spectrometry described in Chapter 3, there are already future projects in the works for additional PKS and NRPS systems. For example, PikAIV, the module following PikAIII, is thought to be much less specific in regard to substrate size and structure than PikAIII[10,11]. Determining the structural differences that could point to the specificity differences in the two modules is an important question. Although cryo-EM data has not yet been collected, some preliminary mass spectrometry data for trapped catalytic intermediates in PikAIV have been collected. PikAIV condenses the β-hydroxyhexaketide from PikAIII with methyl-malonyl (MM) to form a heptaketide. We were able to detect PikAIV ACP-bound heptaketide at high yield in a construct for which the terminal thioesterase (which off-loads this product) was knocked out. Additional active site peptides with bound substrates were also identified and quantified with FT-ICR mass spectrometry, as summarized in Figure 6.1. Chapter 3 made several protocol changes from a previous lab member's procedures, including reducing the trypsin digestion time to fifteen minutes to avoid hydrolyzation. A 15 minute digestion produced acceptable results. However, it would be advantageous to further reduce the digestion time to increase the chance of

trapping intermediates, such as the penta- and hexaketide on the ketosynthases. Digestion times around 1 minute have been reported when coupled with ultrasonication and as low as 50 seconds for trypsin immobilized on a hybrid silica monolith[12,13].
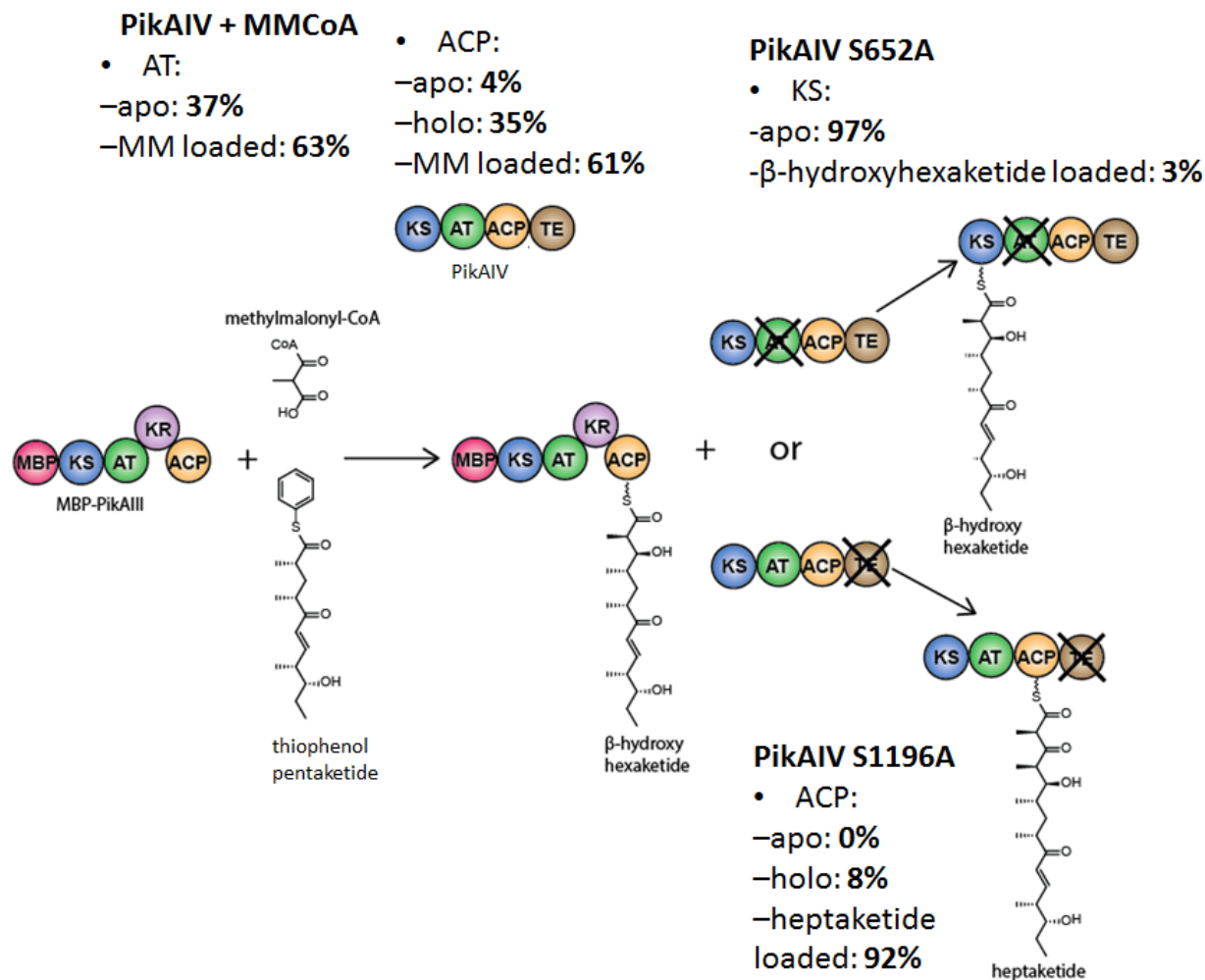

Figure 6.1. Quantitative LC/FT-ICR MS data and depiction of the formation and trapping of PikAIV catalytic intermediates. The generated heptaketide had an unexpectedly high yield. PikAIV cartoons generated by Meredith Skiba.

Bryostatin-1, a polyketide natural product (see biosynthetic scheme in Figure 6.2), is a highly potent inhibitor of protein kinase C and has been investigated as an agent against cancer, HIV, and Alzheimer's[14,15]. Additionally, bryostatin-1 could bring back memory loss subsequent to a postischemic stroke[16]. Although this molecule shows great promise, the bryostatins are produced from the bacterium, *Candidatus Endobugula sertula,* which is unculturable[17]. As a

135

result, it is a costly product, about $14 million/gram. For total synthesis, the lowest number of steps reported to date is thirty.
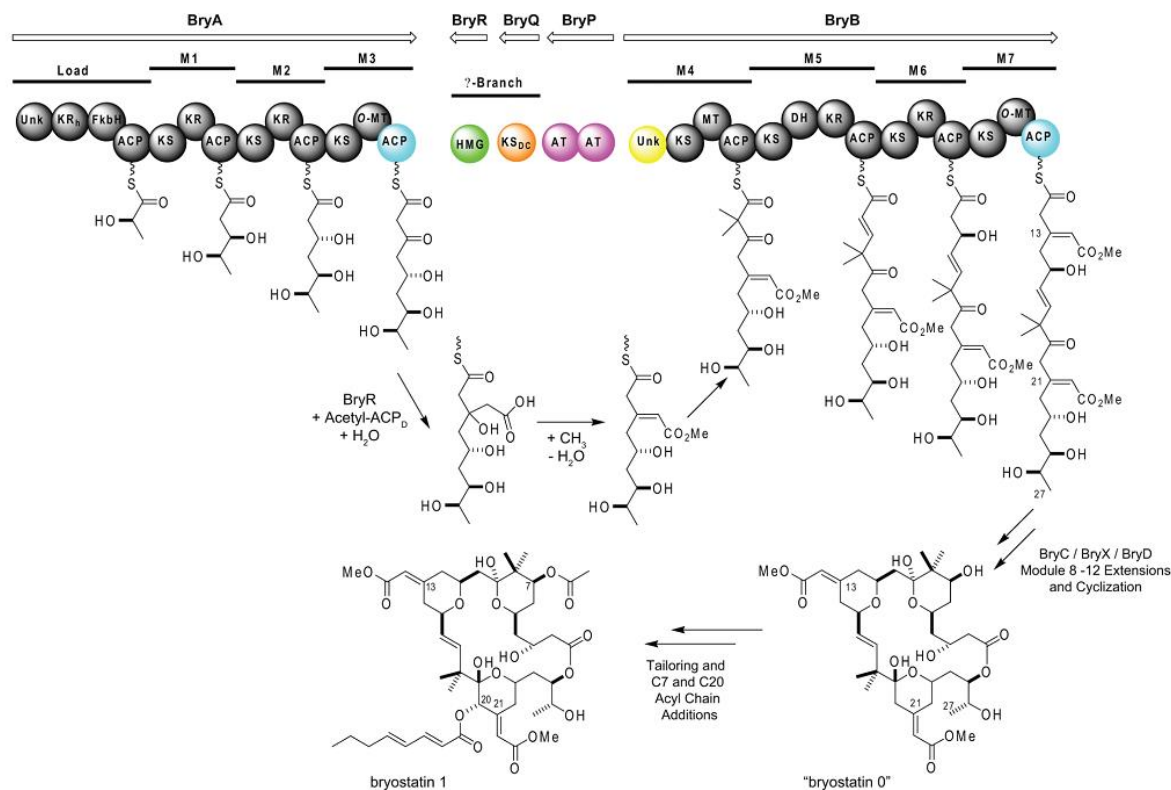

Figure 6.2. The bryostatin biosynthetic pathway. Only one product, bryostatin-1, is shown, though there are several other natural products that are produced by this pathway. Figure from Buchholz et al[18].

The bryostatins are another example of the need for chemoenzymatic synthesis. The enzymatic selectivity of BryR, a portion of the bryostatin biosynthetic pathway, has been investigated by determining binding affinities between BryR and ACPs from other biosynthetic pathways[18]. The great therapeutic potential of the bryostatin family may be further developed by feeding unnatural substrates into the bryostatin pathway to form novel natural products, or to create more efficient natural products. FT-ICR mass spectrometry would be a valuable tool for confirming and quantifying active site occupancy and kinetics of each module.

6.2.2 Determination of PikAIII dynamics via hydrogen/deuterium exchange mass spectrometry

One of the most surprising discoveries in the PikAIII structural analysis (Chapter 3) was the end-to-end flip of the ketoreductase (KR) domain upon addition of thiophenol-pentaketide (Figure 6.3)[19]. Although the reason for this end-to-end flip was speculated upon; the flip brings the KR active site closer to the ACP active site, the mechanism of the flip is unknown.
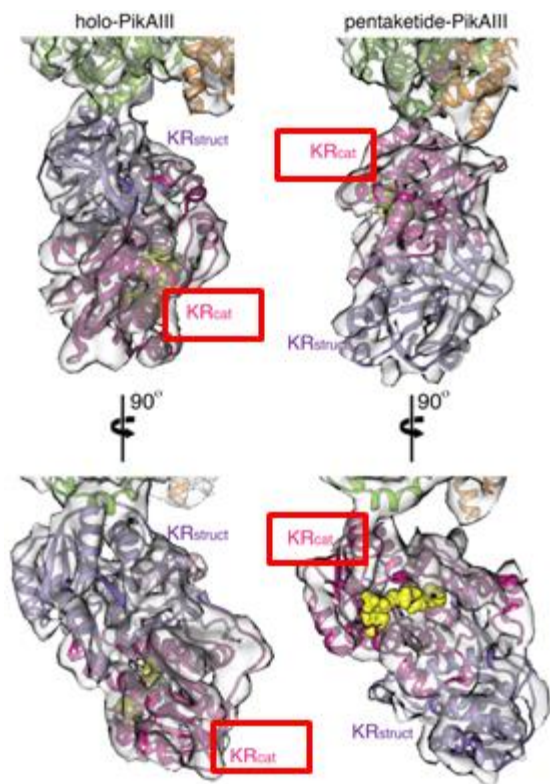

Figure 6.3.   PikAIII cryo-EM structures fitted with a KR crystal structure from the 6-deoxyerythronolide B synthase (DEBS) natural product biosynthetic pathway. An end-to-end flip occurs for the KR domain upon addition of the pentaketide substrate to PikAIII. Figure generated by Somnath Dutta.

The dynamics of the end-to-end flip can be interrogated with hydrogen deuterium exchange (HDX) mass spectrometry. Due to the large size of PikAIII, 150 kDa for the monomer, a bottom-up approach is the most appropriate for such experiments, with pepsin as the enzyme of choice, as it is active at pH 2.5, preventing deuterium back exchange. The Håkansson lab has previously worked in this area and is now returning to this field. Preliminary experiments have been carried out to optimize HDX conditions, such as ratio of protein to pepsin, pepsin digestion period, and buffer conditions for protein and pepsin. Thus far, these conditions have been optimized for a small protein, myoglobin (Figure 6.4). Future experiments will include

137

HDX mass spectrometry on larger proteins, optimizing the conditions for each and finally performing HDX mass spectrometry on PikAIII incubated with the pentaketide. HDX can help determine the mechanism of the conformational changes the KR undergoes for the end-to-end flip, as solvent exposed amino acids are revealed, leading to more structural insight into this natural product biosynthetic pathway. We have focused on a bottom-up approach, which can suffer from lower sequence coverage than its counterparts, top-down and middle-down analysis. However, top-down analysis of a 150 kDa protein is likely beyond the range of a 7T FT-ICR mass spectrometer. Therefore, development of a novel middle-down technique may increase the KR sequence coverage. Pepsin is still the ideal enzyme for HDX; therefore, other factors must be explored to increase proteolytic peptide size. Reducing the digestion time, increasing the protein to pepsin ratio, or decreasing the digestion temperature may yield larger peptides and each of these factors should be explored systematically to optimize the middle-down approach for highest sequence coverage.
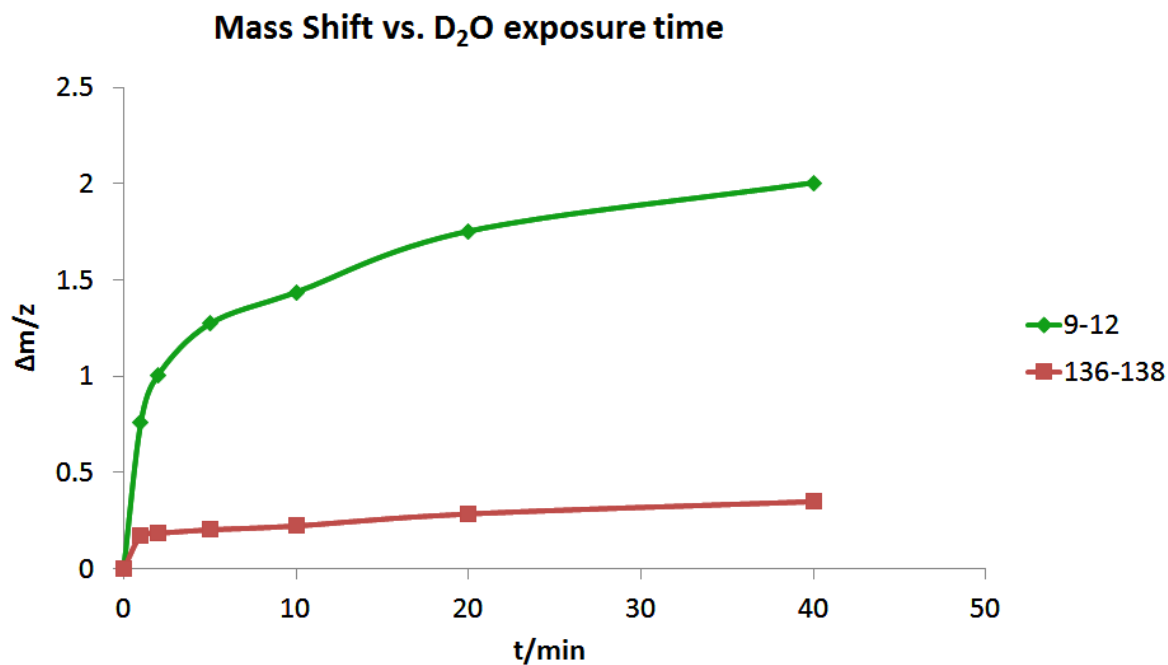
Emma Wang

Figure 6.4. Example graph of the mass increase of two peptic peptides from myoglobin upon D_2O exposure. The green trace represents a peptide more exposed to solvent than the maroon

trace, and thereby showing a larger mass shift. Data collected with Emma Wang. Figure generated by Emma Wang.

6.2.3 Data independent LC/IRMPD to identify natural product biosynthetic enzymes in a red tide toxin-producing dinoflagellate

Dinoflagellates such as *Karenia brevis* (*K. brevis*) produce secondary metabolites that are large and complex, some toxic, whereas others are potential medicines[20]. The *K. brevis* genome is rather inaccessible due to its large size, 50-100 times the size of the human genome[21]. Organisms such as these are ideal targets for a data independent LC/MS approach. A collaboration with Dr. David Sherman's lab has been set up to specifically study natural products from this organism, including brevenal, an anti-toxin, that has the potential to treat lung diseases, and ladder polyethers, brevetoxins, which are responsible for the dangerous Florida red tides (see structures in Figure 6.5)[22,23].
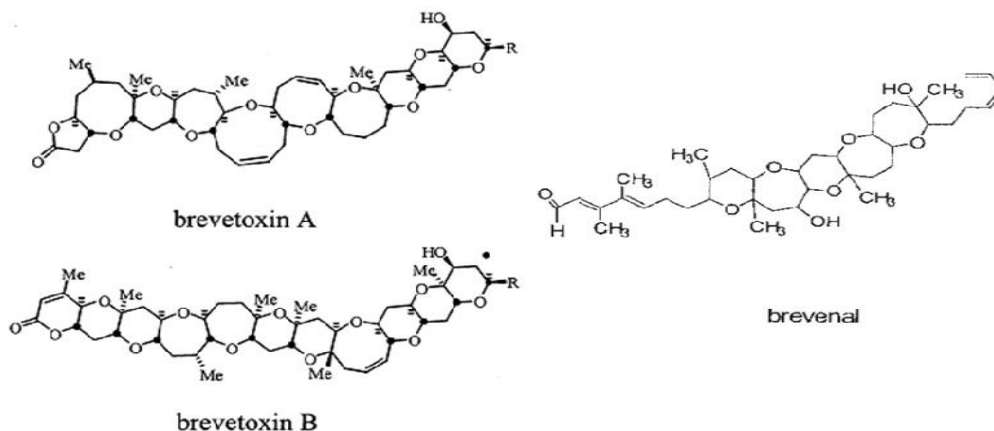

Figure 6.5. Brevetoxins are thought to cause the Florida red tides and have also been linked to airway complications in humans, whereas the antitoxin brevenal has been shown to provide symptomatic relief in diseases such as cystic fibrosis[23].

One approach to discovering the enzymes responsible for generating the brevetoxins and brevenal is to synthesize chemical probes designed to locate and isolate either holo-ACPs or the thioester bond in the ketosynthase domain. These chemical probes will also include a primary or secondary phosphate group, to allow for the data-independent LC/IRMPD method described in Chapter 5. FT-ICR MS will also be used in model systems to ensure that the probes function as desired and are targeting the chosen chemical groups Once chemical probes have been

shown to function well, they can be used to isolate PKSs from *K. brevis* through affinity enrichment, e.g., through incorporation of a biotin functionality.   Data independent LC/IRMPD will allow identification of active site biosynthetic peptides, which can be sequenced via an arsenal of tandem mass spectrometry techniques. This sequence information will subsequently be used to design PCR primers for amplifying and sequencing the corresponding genes for the first time.   This information may also further efforts in sequencing other ladder polyether-generating organisms.

6.3 References

(1)   Nash, P.; Tang, X.; Orlicky, S.; Chen, Q.; Gertler, F. B.; Mendenhall, M. D.; Sicheri, F.; Pawson, T.; Tyers, M. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **2001**, *414*, 514.

(2)   Ferrell, J. E. J. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* **1996**, *21*, 460.

(3)   Varedi, K., SM; Song, H.; Hale, W. A.; Hakansson, K.; Lin, X. Multisite Phosphorylation Undergoes Ultrasensitive Binding of Sic1 to Cdc4 in Response to Cln2-Cdc28. *Mol. Cell* **2014**, *In revision.*

(4)   Gu, L.; Wang, B.; Kulkarni, A.; Geders, T. W.; Grindberg, R. V.; Gerwick, L.; Hakansson, K.; Wipf, P.; Smith, J. L.; Gerwick, W. H.; Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification. *Nature* **2009**, *459*, 731.

(5)   Kwan, D. H.; Leadlay, P. F. Mutagenesis of a Modular Polyketide Synthase Enoylreductase Domain Reveals Insights into Catalysis and Stereospecificity. *ACS Chem. Biol.* **2010**, *5*, 829.

(6)   Ames, B. D.; Nguyen, C.; Bruegger, J.; Smith, P.; Xu, W.; Ma, S.; Wong, E.; Wong, S.; Xie, X.; Li, J. W. H.; Vederas, J. C.; Tang, Y.; Tsai, S.-C. Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11144.

(7)   Flora, J. W.; Muddiman, D. C. Selective, Sensitive, and Rapid Phosphopeptide Identification in Enzymatic Digests Using ESI-FTICR-MS with Infrared Multiphoton Dissociation. *Anal. Chem.* **2001**, *73*, 3305.

(8)   Crowe, M. C.; Brodbelt, J. S. Differentiation of Phosphorylated and Unphosphorylated Peptides by High-Performance Liquid Chromatography-Electrospray Ionization-Infrared Multiphoton Dissociation in a Quadrupole Ion Trap. *Anal. Chem.* **2005**, *77*, 5726.

(9)   Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *Biochemistry* **2006**, *45*, 12756.

(10)   Mortison, J. D.; Kittendorf, J. D.; Sherman, D. H. Synthesis and Biochemical Analysis of Complex Chain-Elongation Intermediates for Interrogation of Molecular Specificity in the Erythromycin and Pikromycin Polyketide Synthases. *J. Am. Chem. Soc.* **2009**, *131*, 15784.

(11)   Yin, Y.; Lu, H.; Khosla, C.; Cane, D. E. Expression and Kinetic Analysis of the Substrate Specificity of Modules 5 and 6 of the Picromycin/Methymycin Polyketide Synthase. *J. Am. Chem. Soc.* **2003**, *125*, 5671.

(12)   Domínguez-Vega, E.; García, M. C.; Crego, A. L.; Marina, M. L. First approach based on direct ultrasonic assisted enzymatic digestion and capillary-high performance liquid chromatography for the peptide mapping of soybean proteins. *J. Chromatogr. A* **2010**, *1217*, 6443.

(13)   Ma, J.; Hou, C.; Liang, Y.; Wang, T.; Liang, Z.; Zhang, L.; Zhang, Y. Efficient proteolysis using a regenerable metal-ion chelate immobilized enzyme reactor supported on organic–inorganic hybrid silica monolith. *Proteomics* **2011**, *11*, 991.

(14)   Nelson, T. J.; Alkon, D. L. Neuroprotective versus tumorigenic protein kinase C activators. *Trends Biochem. Sci.* **2009**, *34*, 136.

(15)   Banerjee, S.; Wang, Z.; Mohammad, M.; Sarkar, F. H.; Mohammad, R. M. Efficacy of Selected Natural Products as Therapeutic Agents against Cancer⊥. *J. Nat. Prod.* **2008**, *71*, 492.

(16)   Sun, M.-K.; Hongpaisan, J.; Alkon, D. L. Postischemic PKC activation rescues retrograde and anterograde long-term memory. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14676.

(17)   Sudek, S.; Lopanik, N. B.; Wagonner, L. E.; Hildebrand, M.; Anderson, C.; Liu, H.; Patel, A.; Sherman, D. H.; Haygood, M. G. Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from "Candidatus Endobugula sertula", the Uncultivated Microbial Symbiont of the Marine Bryozoan Bugula neritina. *J. Nat. Prod.* **2007**, *70*, 67.

(18)   Buchholz, T. J.; Rath, C. M.; Lopanik, N. B.; Gardner, N. P.; Hakansson, K.; Sherman, D. H. Polyketide B-Branching in Bryostatin Biosynthesis: Identification of Surrogate Acetyl-ACP Donors for BryR, an HMG-ACP Synthase. *Chem. Biol.* **2010**, *17*, 1092.

(19)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **2014**, *510*, 560.

(20)   Rein, K. S.; Borrone, J. Polyketides from dinoflagellates: origins, pharmacology and biosynthesis. *Comp. Biochem. Phys. B* **1999**, *124*, 117.

(21)   Hou, Y.; Lin, S. Distinct Gene Number-Genome Size Relationships for Eukaryotes and Non-Eukaryotes: Gene Content Estimation for Dinoflagellate Genomes. *PLoS One* **2009**, *4*, 1.

(22)   Fleming, L. E.; Backer, L. C.; Baden, D. G. Overview of Aerosolized Florida Red Tide Toxins: Exposures and Effects. *Environ. Health Persp.* **2005**, *113*, 618.

(23)   Abraham, W. M.; Bourdelais, A. J.; Sabater, J. R.; Ahmed, A.; Lee, T. A.; Serebriakov, I.; Baden, D. G. Airway Responses to Aerosolized Brevetoxins in an Animal Model of Asthma. *Am. J. Reprod. Im. Mic.* **2004**, *171*, 26.