9.11 Applications of Modern Mass Spectrometry Techniques in Natural Products Chemistry

Roland D. Kersten, Michael J. Meehan, and Pieter C. Dorrestein, University of California, San Diego, La Jolla, CA, USA

9.11.1 Introduction

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

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

The emphasis of this review is placed on two structural classes of natural products: polyketides and nonribosomal peptides (NRPs). The MS of these biosynthetic pathways is most advanced and will be covered in detail. In the following sections we will describe the current methods and applications used to study the biosynthetic pathways of natural products and provide a glimpse into upcoming techniques. In addition, a brief introduction to experimental design using high-end MS to study the biosynthesis of other natural metabolites, such as ribosomally encoded pathways and cofactors, is described.

9.11.1.1 Introduction to NRPS and PKS Biosynthesis

Many important therapeutics, in use in clinics today, are biosynthesized by the nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) paradigm. For example, many of the antibiotics (penicillin, cephalosporin, vancomycin, erythromycin, etc.), immunosuppressors (cyclosporine, rapamycin), antiviral agents (luzopeptin A), antitumor agents (bleomycin), and toxins (thaxtomin) are NRPS and PKS derived.[20–22] **Figure 1** displays a small selection of natural products that are NRPS and PKS derived and illustrates the diversity of molecular structures generated by these biosynthetic paradigms.

NRPS biosynthesis differs substantially from ribosomally encoded peptides. In the NRPS biosynthetic strategy, the substrates and intermediates are covalently linked to an active site serine on the thiolation (T) domain of an NRPS via the phosphopantetheinyl arm.[23] This posttranslational event is accomplished by a phosphopantetheinyl transferase (PPTase), which primes the active site serine with a portion of coenzyme A (CoA) to generate the holo form of the T domain (**Figure 2(b)**).[24] In traditional NRPS biosynthesis, amino

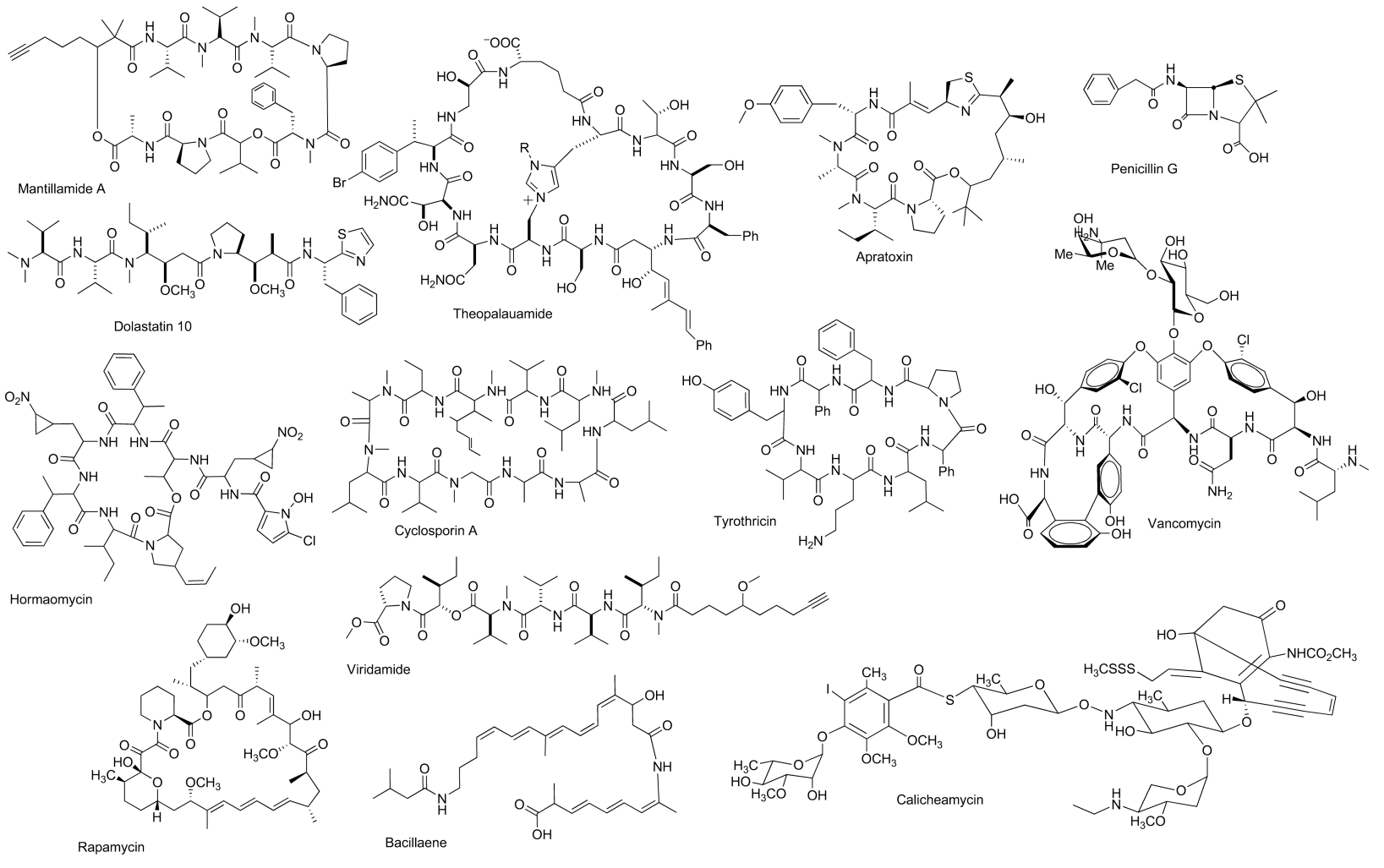


Figure 1 Some examples showing the diversity of NRPS- and PKS-derived natural products.

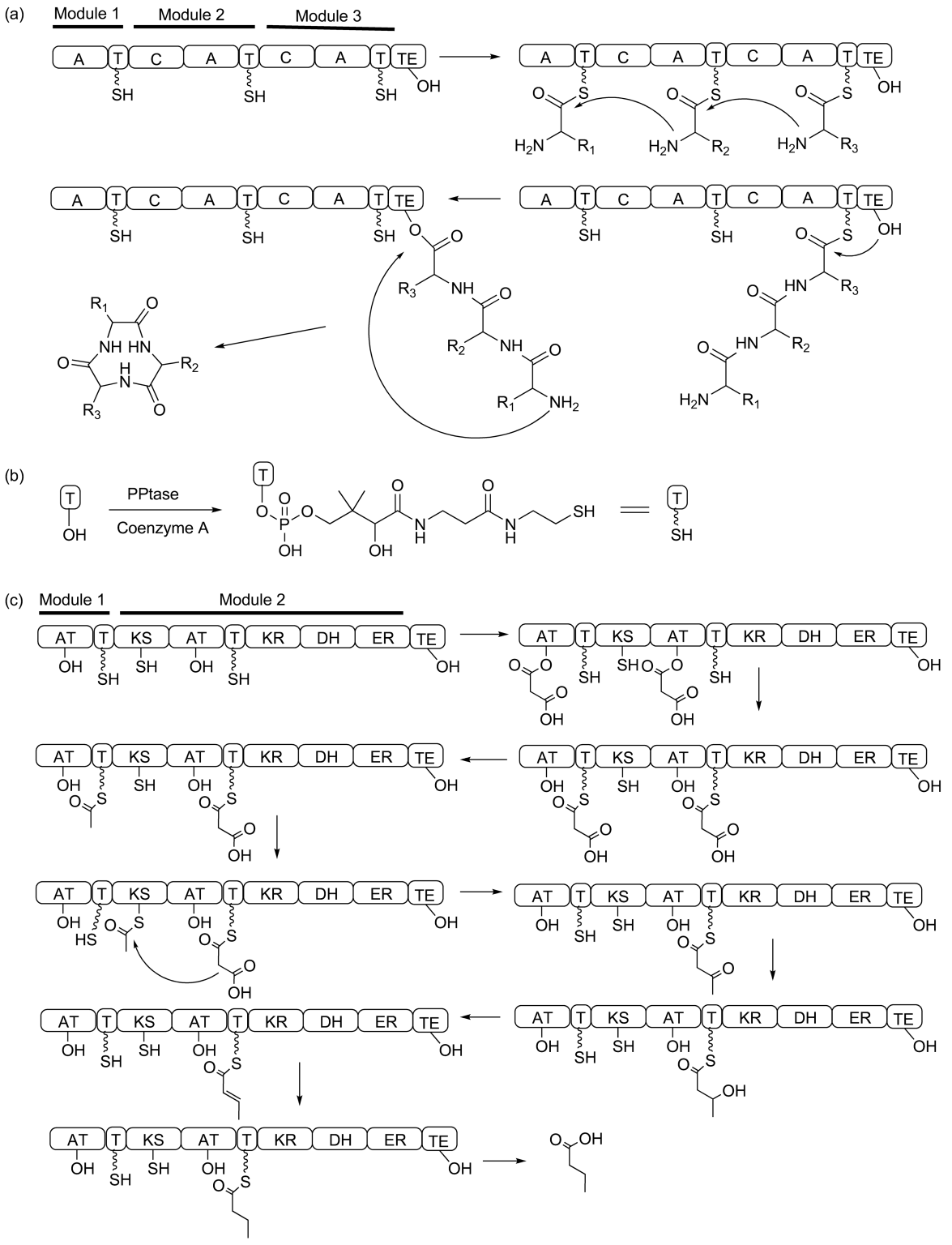


Figure 2 Overview of NRPS and PKS biosynthesis. (a) NRPS biosynthesis. (b) Priming of apo T domain with coenzyme A by a phosphopantetheinyl transferase. (c) PKS biosynthesis.

acids are activated by adenosine triphosphate (ATP). Activation is achieved by the formation of an adenylated carboxylate of an amino acid that is encoded by the first adenylation (A) domain (**Figure 2(a)**).[25] The activated amino acid then undergoes nucleophilic displacement by the thiol terminus of the phosphopantetheine (PPant) arm tethered to the first T domain. During the elongation step, the resulting aminoacyl-enzyme is then condensed with a second acyl–enzyme species on a downstream carrier domain to form a linear peptide, in a reaction catalyzed by a condensation (C) domain. Further modifications often take place while the amino acid or the growing peptide is still attached to a carrier domain (e.g., oxidation, chlorination, or other types of modifications).[26] Finally, the mature peptide is released from the last carrier domain via cyclization or removed by a termination domain.[27] The most common off-loading domain is a thioesterase (TE),[28] although other termination domains exist. While there are currently 22 known ribosomally encoded amino acids, there are hundreds of nonribosomally encoded amino acids expanding the array of structures (and therefore pharmacologically active compounds) that can be generated via the NRPS paradigm.[29]

Just like NRPS biosynthesis, polyketide biosynthesis often takes place on multidomain megasynthases.[23] PKSs also carry their substrates and intermediates on T domains, also referred to as acyl carrier proteins (ACPs). The substrates in PKS biosynthesis are not amino acids or free carboxylic acids as it is observed with NRPS systems, but rather they are CoA-activated carboxylic acids. The most common substrate for polyketide biosynthesis is malonyl-CoA. In order to divert a small subfraction of the cellular pool of malonyl-CoA, PKSs activate the malonyl by a nucleophilic attack of the active site serine from the acyltransferase (AT) domain forming a covalent oxo-ester linkage (**Figure 2(c)**). Subsequently, a transacylation takes place onto the T domain linking the malonyl to the thiol of the phosphopantetheinyl functionality via a thioester bond. While the malonyl is still attached to the T domain, it is decarboxylated by the ketosynthase (KS) domain before a second transacylation takes place to the active site cysteine of the KS domain generating an acetyl-*S*-KS intermediate, something that to date is very hard to capture directly by MS. At this point, a second malonyl group is loaded onto the upstream T domain, a reaction again catalyzed by an AT domain. As the KS domain decarboxylates the second malonyl, a Claisen condensation takes place with the acetyl group on the KS domain forming acetoacetate, the first unit in an elongating polyketide. Like NRPS biosynthetic pathways, PKS biosynthesis can incorporate a variety of tailoring reactions.[23] Tailoring reactions such as chlorination, oxidation, and cyclization may be observed.[30–32] By far the most common tailoring steps in PKS biosynthesis are ketoreduction, dehydration, and enoyl-reductions. Ketoreduction is catalyzed by the ketoreductase (KR) domain and uses the cofactor nicotinamide adenine dinucleotide (phosphate) hydride (NAD(P)H) to convert the β-carbonyl to an alcohol. This same alcohol can be dehydrated by a dehydratase (DH) domain. The DH domain catalyzes the deprotonation of the α-proton and protonates the leaving hydroxyl group in an α,β-unsaturated thioester. This thioester can then be further reduced to a completely saturated bond by the enoylreductase (ER). This reaction utilizes the cofactor NAD(P)H as required in the ketoreduction.

While there are many natural products that are either NRPS or PKS derived, there are multidomain megasynthases that contain both NRPS and PKS biosynthetic features. Examples of such biosynthetic motifs are epothilone, jamaicamide, and the endiyne C-1027.[33–35] At NRPS–PKS interfaces in these systems, the condensation of malonyl takes place with an amino acid that was activated by an A domain. Alternatively, an amino acid loaded on a T domain condenses with an elongating polyketide. Since the substrates and intermediates in NRPS, PKS, or hybrid NRPS/PKS biosynthesis introduce mass changes on T domains, they are ideal candidates for investigation by MS.

9.11.1.2 New Tools in the Characterization of Multidomain and Phosphopantetheinylated Proteins

For many years, it was very difficult to study multidomain, posttranslationally, or transiently modified proteins. These studies required several years of efforts by MS experts. But as instrumentation improved it has enabled additional investigators to carry out this type of research. Many proteins, in particular NRPS and PKS, are of substantial size. Some of the largest single open reading frames are responsible for the generation of mega dalton polypeptides. An example of this are the 2.5 MDa NRPSs of the syringomycin pathway and the megasynthase of the biosynthetic pathway of an 18-mer peptaibol.[36,37] In the majority of

cases, it is not possible to look at intact proteins of this size, so they must be digested before the active sites can be analyzed in terms of tethered biosynthetic substrates and intermediates. However, when these proteins are digested, the complexity increases and the active sites have to be found within a haystack of data. Many modern proteomic programs such as InSpecT, Sequest, OMSSA, Spectumill, Mascott, and MassLynx fail to find these active site peptides or domains of posttranslationally modified megasynthases from complete or limited digests.[38,39] There are three main reasons for this: (1) Despite the surge in the development of proteomic platforms for the analysis of digested proteins, all proteomic programs have a difficult time identifying ions based on the fragmentation data of parent ion that have charges that are greater than 4+. Most of these programs, by default, ignore 4+ ions or larger as this would increase the number of false positives. We estimate that approximately 30% of all the data in a liquid chromatography–mass spectrometry (LC–MS) of a complete tryptic digest would belong to these higher charges. In addition, because of the poorer nature of 1+ ion fragmentation, these too are often not identified. (2) While many of these programs can identify nonlabile modifications, they have to be 'trained' for labile modifications should it be possible to find them at all. For example, most of these programs have a specific scoring function for the analysis of phosphopeptides and none of these programs have a scoring function for labile modifications such as the phosphopantetheinylation found on fatty acid synthase (FAS), NRPS, and PKS multidomain proteins. (3) For a given LC–MS/MS or MS^2 run for a digest, we typically obtain 10 000–20 000 MS/MS or MS^2 spectra, yet on a good run only 2000–3000 MS/MS or MS^2 spectra are annotated. This begs the question what the remaining spectra are and clearly indicate that we are not yet identifying everything we can from the data. Thus there is an enormous opportunity for the development of programs that can capture the remaining uncharacterized spectra, including 4+ or greater charge ions or ions with different labile modifications.

Even though automated programs to facilitate the analysis of these large multidomain proteins involved in the biosynthesis of natural products do not exist, it has become possible to investigate such systems by MS. After a short protease digestion, most of the active sites are often 5–20 kDa. It is not yet reasonable to map the active sites by MS/MS or MS^2 data as it is done in proteomic investigations.[40] However, with the emergence of high-resolution instruments that have routine mass accuracies within 10 ppm, it became possible to find the active sites by the intact mass of the peptides alone giving rise to a manageable number of false positive matches. Any false match is eliminated by MS/MS or MS^2. When one performs MS/MSor MS^2, many fragment ions are detected that should match up to expected fragments for the active sites. The false positive active sites matches are eliminated on the basis that their fragments will not match up to the sequence of the active site, while all of the true positive candidate fragments match quite well. In 1999, Kelleher and coworkers were the first to recognize the usefulness of Fourier transform–ion cyclotron resonance–mass spectrometry (FT–ICR–MS) resolution and mass accuracy for the investigations of multidomain NRPS proteins.[41] At that time, FT–ICR–MS instruments were the only instruments with this degree of mass accuracy and the best of these instruments were custom built and not readily available to the general public. As instrumentation and the development of efficient strategies to analyze multidomain NRPS/PKS proteins are advancing, more and more laboratories can carry out these types of investigations, including laboratories that only have access to low-resolution instruments.[42–46] In the following sections, we describe the current and new MS approaches to analyze some of these biosynthetic pathways since a 2006 review on this topic.[47]

There are several key advances that have been in use since 2006 for the investigations of NRPS and PKS proteins: (1) The use of larger-sized magnets (e.g., 12 Tesla) enables active site mapping on an LC timescale.[48,49] (2) It was discovered that the labile posttranslational modification of T domains can be ejected during thermal activation methods and that this ejection can be used to 'observe' substrates and intermediates on the active site thiol of a phosphopantetheinyl functionality. This assay, called the phosphopantetheinyl ejection assay (PEA), has found many uses and has been adopted by several other laboratories already.[48–58] (3) The PEA can also be applied to low-resolution instruments making this assay accessible to many other researchers who work with phosphopantetheinylated proteins.[43,44] (4) Finally, there are now instances describing other methods for analyzing the phosphopantetheinylated proteins, such as the use of a phosphatase to remove the phopshopantetheinyl functionality so that it may be characterized by MS. All of these new MS capabilities are highlighted in the next sections.

9.11.1.3 A Brief Introduction to FT–ICR–MS

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

$$f_{\mathrm{cyc}} = \frac{zB}{2\pi m} \tag{1}$$

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

9.11.1.4 Interpretation of FT–ICR–MS

Due to the fact that most people investigating natural product biosynthesis do not routinely use FT–ICR–MS in their research, the interpretation of a high-resolution mass spectrum of a protein domain is highlighted in this section. The broadband spectrum of a freestanding thiolation protein Pks4 from the bikaverin biosynthetic pathway is shown in **Figure 4(a)** as a mixture of its apo and holo form.[62] This protein has a mass of 14 394 Da. The mass range for FT–ICR–MS analysis of proteins involved in the biosynthesis of natural products must fall between 200 and 2000 m/z. In order to enable the visualization of these ions within this standard m/z window, ions of Pks4, or other protein domains that are larger than 2000 Da, need to be multiply charged (z), which is experimentally accomplished by application of electrospray ionization (ESI). In the case of Pks4, we see that the same protein has multiple charge states ranging from 15 charges on the left side of the spectrum to charge 11 on

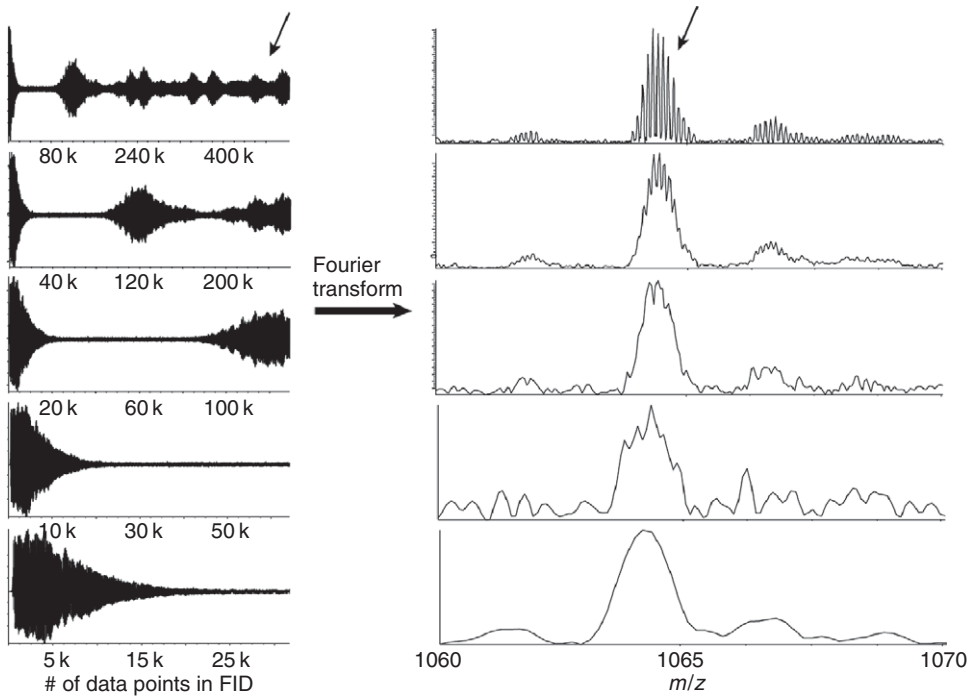


Figure 3 Correlation between the time domain duration and mass resolution. The more data points the FID (left) consists of, the higher the resolution of the mass signals (right) obtained by Fourier transform. The arrow on the left points to the beads in the time domain signal that accounts for the fine information (i.e., isotopes) shown by the arrow on the right.

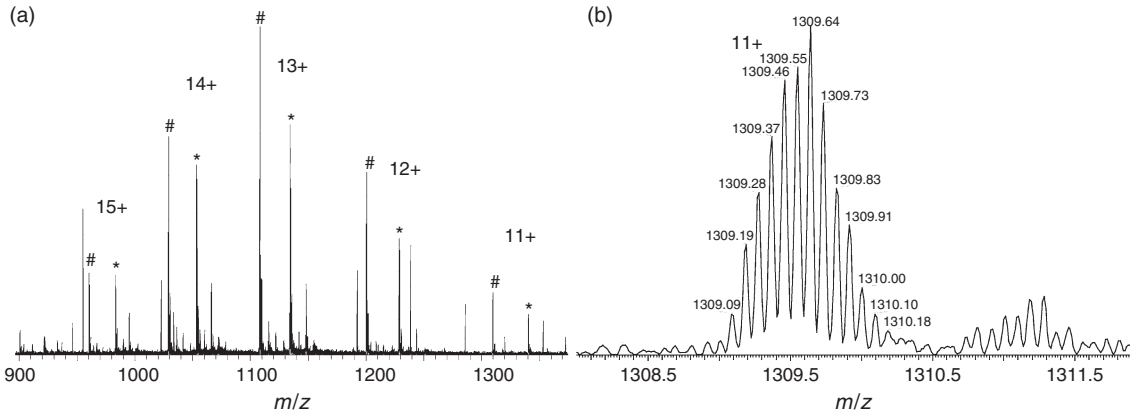


Figure 4 ESI–FT–ICR mass spectrum of freestanding T domain Pks4 from bikaverin biosynthetic pathway. (a) Multiple charge states of apo (#) and holo (*) Pks4. (b) Single charge state (11+) of Pks4 shows isotopic resolution obtained by FT–ICR–MS analysis.

the right side of the spectrum. The mass is calculated by taking the observed m/z and multiplying this value with the observed charge. Details on how to calculate the mass of such a protein or protein domain by manual means are reviewed by Dorrestein and Kelleher.[47] Pks4 is a relatively small protein and proteins of a size <20 kDa are usually well suited for FT–ICR–MS analysis. Larger proteins pose more challenges to MS characterization. For instance, if a protein of 100 kDa is analyzed, it is not uncommon to find between 100 and 150 charges on the protein and that the charge profile has >100 different charge states as opposed to the 5 observed for Pks4. **Figure 4(b)** shows an enlarged picture of a single charge state. When one zooms in on a charge state observed in **Figure 4(a)**, an isotopic profile of the ion becomes visible. This type of characteristic

isotopic profile exists because larger proteins, or protein domains, have many carbons, nitrogens, and oxygens. The natural abundance of isotopes other than the monoisotopic masses (e.g., ^{12}C, ^{14}N, and ^{16}O) is the reason why an isotopic profile is observed for a protein. The main forms of isotopes that contribute to the isotopic profile of a protein are the natural ^{13}C and ^{15}N that are present at ~1 and 0.36%, respectively. FT–ICR–MS has the resolving power that enables the visualization of the isotopes of proteins up to 110 kDa in mass.[63] It should be noted however that, in practice, the larger the protein, the more difficult it is to get isotopic resolution and that the theoretical maximum isotopic resolution of large multidomain proteins with higher field magnets (>12T) has not yet been achieved experimentally.[59]

9.11.1.5 LC–FT–ICR–MS Analysis of NRPS and PKS Proteins

Before 2006, all NRPS and PKS proteins were investigated via direct infusion of a protein or protein domain that was purified offline by HPLC, or by other C_{18}, C_8, or C_4 forms of peptide/protein purification such as Ziptips or traps. The best results were obtained using nanospray infusion, such as an Advion nanospray robot or similar nanospray devices. The advantage of a nanospray over direct microspray infusion via a syringe, a commonly used infusion method, is twofold. First, nanospray creates finer droplets than traditional forms of electrospray, making the desolvation of the droplets emitted from the spray needle or nozzle easier, and many more ions enter into the gas phase improving the detection. Second, with just 5–10 μl of a purified sample, one can analyze this sample for over an hour, sometimes up to 3 h. With a syringe infusion approach, the infusion rate is usually 2 μl min^{-1}. This is a significant limitation in situations where one has limited sample or a low concentration sample. A single FT–ICR–MS scan may take 8–120 s depending on the selected settings, and typically, 20–200 scans are required to attain a good spectrum. Therefore, nanospray becomes a critical component in the analytical platform for the biosynthetic investigations of NRPS and PKS systems. The reason why so many scans are required is illustrated in **Figure 5**. **Figure 5** shows that even when the sample ionizes well, a single scan is not sufficient for accurate data. As shown for Pks4, it takes ~10 scans to generate an accurate isotopic profile for the protein. Many more scans may be required if the sample's concentration is low, does not ionize well, or contains competing ions. Depending on the conditions, acquisition of data may take minutes to hours, and is unlike NMR where some experiments run overnight. While dependent on many factors, in practice, for every two-fold signal-to-noise improvement needed in order to collect the data, the data acquisition time needs to be increased four-fold. However, as the magnet size of an FT–ICR instrument increases, the sensitivity, and therefore the scan rate, increases compared to when the data are obtained at the same resolution on an instrument with a smaller magnet.

NRPS/PKS active site mapping on the LC timescale is possible by online LC–FT–ICR–MS. While this approach is still difficult to perform with a 7T magnet, the most common magnet size at this time, it becomes routine with a 12T magnet. The challenges with online active site mapping are the inherent signal limitation of the electrospray and the long accumulation times of FT–ICR–MS instruments. The first online LC–FT–ICR–MS analysis of a thiolation active site was described by the Marahiel group in 2006.[42] In their LC trace, they were able to observe a 1097.995 Da ion that corresponded to the posttranslationally phosphopantetheinylated form of the T domain of tyrocidine synthase B.[42] In addition, they were able to observe the amino acid Phe and dipeptide Phe–Phe loaded onto the active site thiol of the phosphopantetheinyl group. If the digestion had resulted in a 15 kDa protein domain containing the active site, these data would have been most likely much more difficult to obtain on an LC timescale. Since then, the Kelleher laboratory has built a 12T FT–ICR–MS instrument enabling routine online LC–FT–ICR–MS analysis and detection of NRPS and PKS active sites.[58,64] Hopefully, instruments with such high sensitivities and with this type of resolving power will become available to others who are tackling these types of systems for biosynthetic interrogation. Yet, as our understanding of the gas-phase fragmentation behavior of active site-tethered substrates and intermediates advances, it may not even be necessary to have high-resolution instruments for these types of experiments. One such advance is the PEA for the characterization of phopshopantetheinylated proteins.[50]

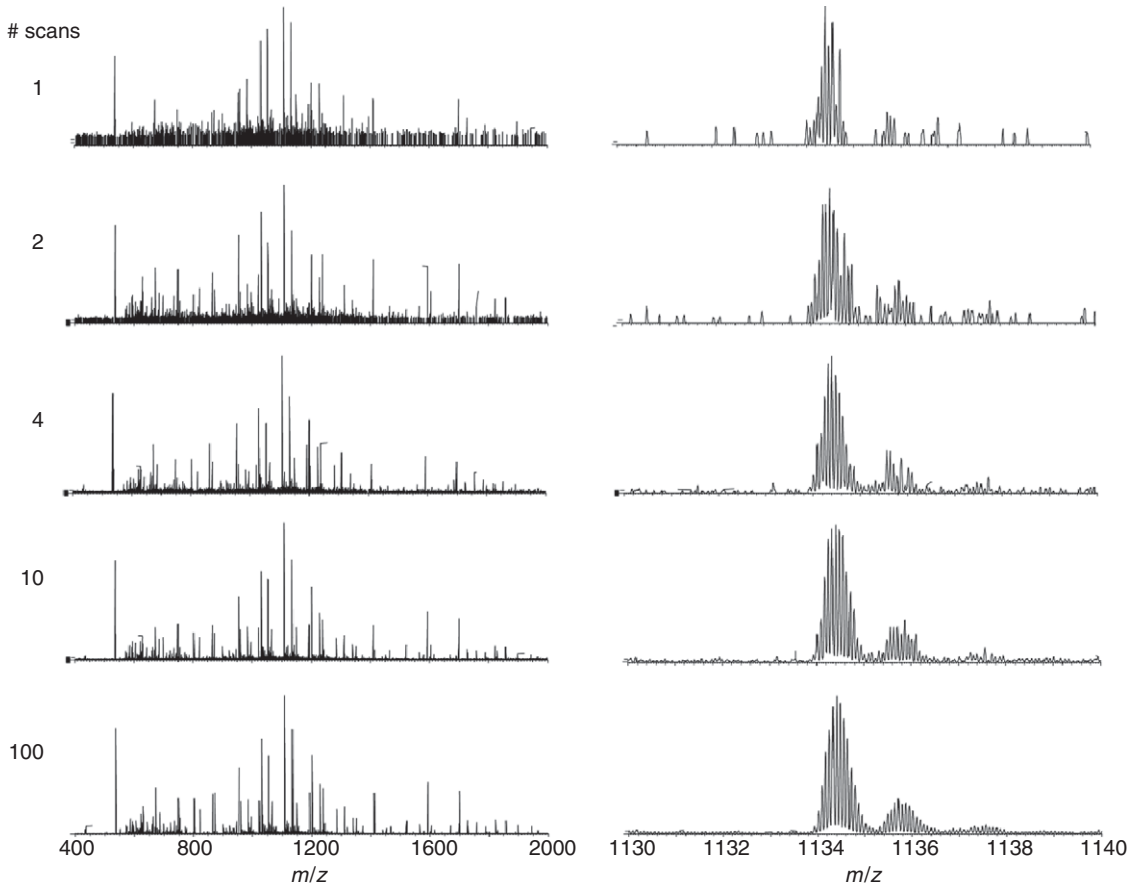


Figure 5 Correlation of FT–ICR–MS scan number and mass signal quality. The more MS scans are acquired (left), the higher the signal-to-noise ratio in the corresponding mass spectrum (right).

9.11.1.6 The Phosphopantetheinyl Ejection Assay

T domains of many biosynthetic pathways are phosphopantetheinylated. The phosphopantetheinyl posttranslational modification of serine is in many ways a novel phosphopeptide. It has been well recognized that the C–O connection of the phosphodiester bond in phosphopeptides is preferentially broken when they are subjected to thermal fragmentation methods (described in Section 9.11.1.7).[65–69] This C–O bond is energetically the most labile connection in such a peptide. Two main mechanisms are proposed for the ejection of a neutral loss phosphate (98 Da) from a phosphopeptide and are shown in **Figure 6**.[2,68] The first mechanism results in the end products that would be expected for a McLafferty-type rearrangement, the dehydroalanine and the uncharged phosphate. In this mechanism the phosphate deprotonates the α-proton on the serine, ejecting the phosphate as a neutral ion. It is not yet known if the rearrangement is a McLafferty-type homolytic rearrangement or a heterolytic cleavage as drawn in **Figure 6**. The second mechanism for neutral ion loss observed in phosphopeptides was proposed by Hunt and coworkers[2] in 2004. In this mechanism depicted in **Figure 6(b)**, a five-membered oxazoline is formed, a reaction that is promoted by the formation of protonated phosphate in the gas phase leading to the cleavage of the C–O connection and release of phosphoric acid as a neutral ion. The support for the understanding that both neutral loss mechanisms are operational has been provided by isotope labeling studies, although these studies did show that the oxazoline mechanism was favored.[68] A phosphopantetheinyl serine modification has a related C–O phosphoester connection but it also has a second C–O connection bearing the active site thiol of the T domain. These C–O linkages on the phosphodiester are the atom connections that most readily fragment in a fashion similar to phosphopeptide

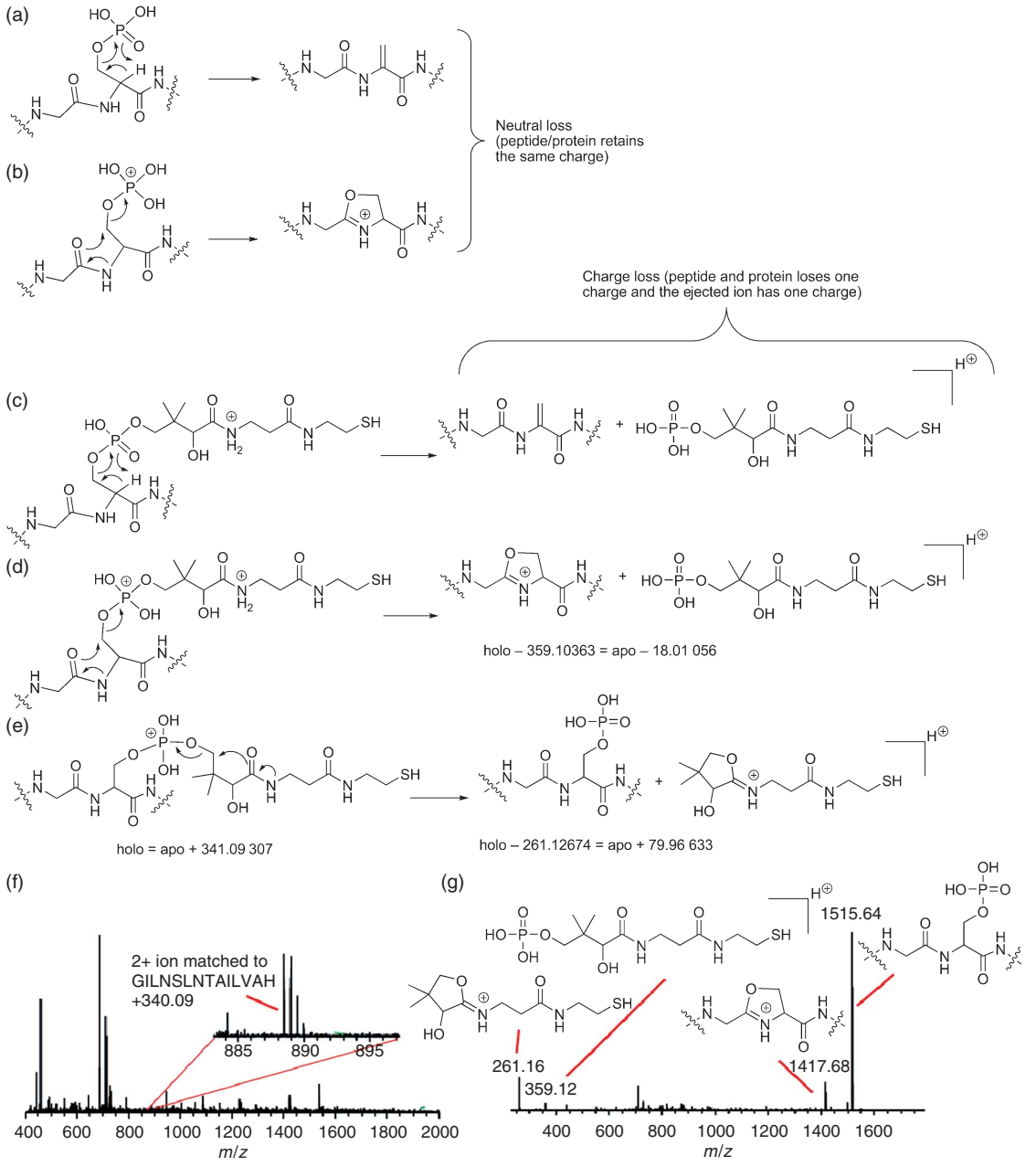


Figure 6 The phosphopantetheinyl ejection assay (PEA). (a) McLafferty-type ejection mechanism. (b) Oxazoline ejection mechanism. McLafferty-type ejection (c) and oxazoline ejection (d) on phosphopantetheinylated protein yield charge-reduced protein minus water and phospho-PPant ejection ion. (e) Iminolactone PPant ejection on phosphopantetheinylated protein yields charge-reduced phospho-apo protein and PPant ion. PEA on CouN5 fragment in broadband FT mass spectrum (f), MS^2 spectrum (g) shows PPant ion (261.16 *m/z*) and phospho-PPant ion (359.12 *m/z*).

neutral losses observed with phosphopeptides. In addition, these C–O linkages will be preferentially fragmented over the fragmentation of amide linkages normally expected of peptides. While the phosphoester ejection from a phosphopeptide results in a neutral loss, the ejection of the phosphodiester from a phosphopantetheinylated peptide or protein results in a loss of a charge on the peptide, resulting in the formation of the charge-reduced apo protein minus water in the case of the McLafferty-type ejection (**Figure 6(c)**), the oxazoline ejection (**Figure 6(d)**), and the thiazoline mechanism, or the charge-reduced apo protein plus phosphate when

an iminolactone is ejected (**Figure 6(e)**).[50] At this time the iminolactone phosphopantetheinyl ejection appears to be the ion that is the most abundant and therefore the most useful in the investigation of substrates and intermediates tethered to the active site thiol of the PPant. **Figures 6(f)** and **6(g)** show examples of phosphopantetheinyl ejection of a peptide obtained via digestion of the T protein CouN5 from the coumermycin biosynthetic pathway. The intact ion corresponding to GILNSLNTAILVAH was subjected to collisionally-induced dissociation (CID) and clearly showed all the different ejected PPant ions (observed 261.16 m/z and 359.12 m/z). This figure shows that these ions are the most prevalent ones generated in this spectrum. While they are the most abundant ions in ~70% of the phosphopantetheinylated proteins or domains investigated to date, these ejected ions are not always this abundant. Nonetheless, the PEA is a welcome addition to the arsenal of tools used to interrogate NRPS and PKS proteins.

PEA can also be carried out to monitor time courses as shown in **Figure 7**. In this reaction, the condensation of malonyl-*S*-PigH with pyrrolyl-*S*-PigG catalyzed by PigJ was directly monitored by the PEA of both of the two T domains of PigH in a time-dependent manner and shows that changes can be monitored using PEA. This implies that kinetic information may be obtained from monitoring the ejected ion only. However, caution should be taken to prevent the overinterpretation of the kinetics from the observed ejected ions. While changes in a time-dependent manner can be observed using PEA, there are three concerns that should be considered if one wants to obtain true kinetics using this method. First, the ejected ions may have different ionization efficiencies. Altering the ratios of two forms of a phosphopantetheinylated protein and then analyzing this by PEA may not respond in a linear behavior. Some of the discrepancy of what the ejected ions report is evidenced in **Figure 7**, as the intact spectrum on the left of **Figure 7** does not correlate with the PEA relative ratios. In

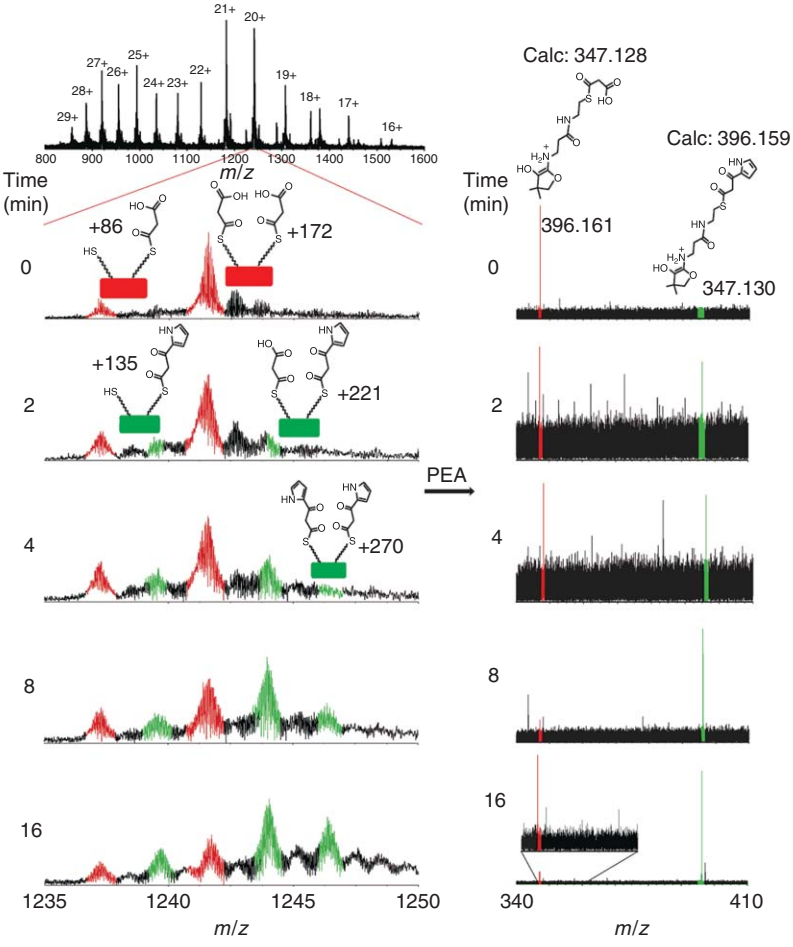


Figure 7 Monitoring time courses by PEA. The condensation of malonyl-*S*-PigH with pyrrolyl-*S*-PigG catalyzed by PigJ is monitored in a time course by broadband FTMS (left) and PEA (right) on the two active sites of PigH.

addition to the differences in ionization efficiencies, the substrate is often partially eliminated from the ejected ion as well. Therefore, the strength of the thioester of the substrate or biosynthetic intermediate will affect the ratios of the ejected ions observed. Because PEA is usually performed on a single charge state of the protein or peptide, the ratio of the ejected ion will vary as well, depending on the ionization efficiency of the parent ion. That change in the ratios between ions of different charge states is directly evident from the Pks4 data in **Figure 4(a)**. Looking at the apo versus holo ratio of each charge state of Pks4, the ratio of apo (annotated with # in **Figure 4(a)**) versus the phosphopantetheinylated form (annotated with * in **Figure 4(a)**) of the protein changes from 0.95 to 0.8. This same phenomenon is observed with different substrate-loaded or biosynthetic intermediate-loaded forms of T domains. While the ratios of different charge states will affect the ejected ions, the manner in which the ions are accelerated by CID also affect the PEA ratios, especially when a small parent ion isolation window is applied. Therefore, it is recommended that the isolation window width for CID, when multiple species are analyzed, is greater than the isotopic profile width of any of the ions that one is interested in fragmenting. Otherwise, different fragmentation energies are applied to the different parent ions and differential ejection can be observed. Despite these inherent caveats, PEA has been used to provide some kinetic information[55,69] and follows the generally accepted accuracy of 10–20% that can be obtained with the investigations of NRPS and PKS protein domains.[70]

9.11.1.7 How Is the PEA Accomplished?

PEA is accomplished using thermal activation methods. There are a large number of thermal activation methods that could be used for this. Source fragmentation, blackbody infrared radiation dissociation (BIRD), sustained off-resonance irradiation–collisionally activated dissociation (SORI–CAD), infrared radiation multiphoton dissociation (IRMPD), and CID are examples of such thermal MS/MS or MS^2 methods.[71–76] Currently, source fragmentation, CID, and IRMPD are the only methods that have been utilized for the PEA, and because of this, these are the only three that are covered in this section of this chapter. Before we begin to look at how and when one should use these methods, it is important to describe the common instrumentation configurations used in NRPS and PKS studies. Since most of PEAs have been performed on FT–ICR–MS instruments, the two main commercially FT–ICR–MS configurations are described in this section. In the first configuration, after the sample is introduced, the ions pass through a heated capillary inlet to a quadrupole where ions can be isolated and then passed on to a linear ion trap (linear IT). The ions can then be detected in the ion trap but at low resolution. Alternatively, the ions can be passed to the ICR cell and the ions can be detected with high resolution (**Figure 8**). Thermo Finnigan hybrid instruments are typically configured in this fashion. In this instrument configuration, thermal fragmentation of ions can take place by exciting the ions at the source by increasing the voltage and colliding the ions with air. On the other hand, CID can be accomplished via excitation of the ions in the ion trap resulting in helium gas collisions. Finally, the instrument can also be equipped with an optional IRMPD and the ions will be fragmented directly in the cell of the instrument. Other commercial FT–ICR instruments from Varian (formerly IonSpec) and Bruker often have a different configuration. Both of these instruments have a sample inlet, usually with a heated capillary, an optional ion funnel (not depicted) to capture as many ions as possible, followed by an isolation quadrupole and an accumulation octupole before passing the ions on to the cell of the ICR instrument. Thermal activation can be accomplished (1) at the source via collisions with air, (2) in the accumulation octupole via collisions with helium or other inert

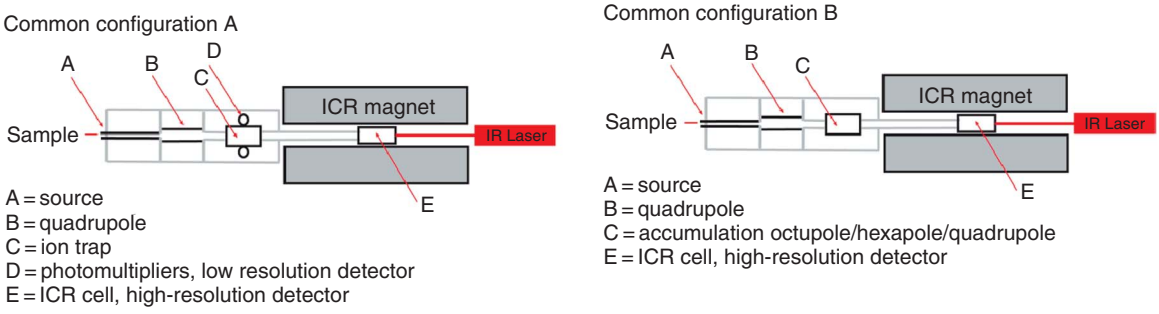


Figure 8 The two main FT–ICR–MS configurations that have been utilized in the investigations of NRPS and PKS systems: LTQ–FT–ICR–MS configuration (left) and accumulation multipole FT–ICR–MS configuration (right).

gas, and (3) inside the cell using a pulsed laser (IRMPD). Below we describe the types of experiments that can be performed by each setup.

9.11.1.7.1 The LTQ–FT–ICR–MS configuration for PEA (common configuration A)

A hybrid instrument such as the hybrid linear trap quadrupole–Fourier transform–ion cyclotron resonance–mass spectrometry (LTQ–FT–ICR–MS) configuration (**Figure 8**) allows one to perform some very interesting experiments with respect to NRPS and PKS multidomain proteins. First, we can fragment at the source of the instrument, resulting in phosphopantetheinyl ejection for any ion that enters the instrument. The advantage of this approach is that we are not just isolating a single charge state for fragmentation but all the charge states from a protein, thereby increasing the signal intensity of the ejected ion. The disadvantage of this approach is that one is likely to get many more signals in the low m/z region, requiring additional confirmation of the signals observed. This confirmation can be accomplished using substrates with stable isotopes or an additional round of fragmentation in a data-dependent manner as done for proteomics experiments. This is of particular importance when the ejection of the PPant is performed on low-resolution instruments as described in Section 9.11.1.8. An additional disadvantage of using source fragmentation for PEA is that the ejected ion cannot be correlated with the parent ion. This is important when one wants to map active sites. While source fragmentation is one of the approaches to obtain PPant ejection, the most common approach is CID. In CID, ions of interest are accelerated and collide with gas ions (helium being the most commonly used). During this collision, most of the kinetic energy is subsequently converted to thermal energy and if the phosphopantetheinylated ion undergoing the collisions has enough vibrational energy amassed, which likely requires multiple collisions it will dissociate into two ions. CID can be accomplished in an inert gas-filled quadrupole or in an ion trap. In the case of a Thermo Finnigan instrument (configuration A), CID is accomplished in the linear ion-trap portion of the instrument. Once dissociated, the fragment ions can be observed using a photomultiplier or the ions can be passed on to the ICR cell for high-resolution analysis. Unfortunately, ion traps suffer from a major but well-documented limitation that is defined as the 1/3rd rule (**Figure 9**).[77,78] When the activation q, that is, the energy that is responsible for accelerating all of the ions to be fragmented, is raised, low m/z product ions start to lose stable trajectories causing them to be ejected from the trap.[77,78] For example, when an activation q is set to 0.25 and the parent ion is isolated at 1200 m/z, one cannot observe fragment ions below 400 m/z. Some improvements in terms of the detectability of the low m/z scanning range can be gained by lowering the activation q, but this change necessitates an increase in the time for activation from 30 to 100 ms in order to obtain significant fragmentation. On the author's LTQ–FT–ICR–MS instrument, when q is set to 0.2, the 1/3rd rule becomes a 1/4th rule, thereby increasing the size of the ions that can be fragmented, and ultimately detected, for the PEA assay. A partial solution for this limitation of ion traps is a related software-controlled mechanism called pulsed-Q dissociation (PQD). Although this is not yet available for commercial LTQ–FT–ICR–MS instruments, it is available for two-dimensional linear trap quadrupoles (LTQs), three-dimensional linear quadrupole (LCQs), and hybrid linear trap quadrupole-orbitraps (LTQ–ORBIs).[79,80] Most ion traps in existence today are not equipped with a PQD software upgrade so that its utility is limited at this time. While it is possible to equip an LTQ–FT–ICR–MS instrument with IRMPD, it has not yet been applied toward PEA on such instruments.

9.11.1.7.2 The accumulation multipole setup for PEA (common configuration B)

The other common setup for FT–ICR–MS instrumentations is configuration B (**Figure 8**), which differs from the ion trap configuration by having an accumulation multipole instead of a linear ion trap. Such an instrument is also capable of PEA by conducting fragmentation at the source of the instrument. The combination of the quadrupole-accumulation multipole setup is very useful in PEA. The advantage is that one can select out ions of interest in the quadrupole and then accumulate ions in the accumulation octupole as shown in **Figure 10(b)**. The ions can then be excited for CID inside the accumulation multipole or they can be passed on to the cell of the instrument. Inside the cell of the instrument, the isolated ions can be subjected to PEA using infrared radiation (IRMPD). While both CID and IRMPD are thermal activation methods and the resulting fragmentation are quite similar it is not yet investigated if one results in improved PPant ejection over the other. The advantage of CID or IRMPD on an accumulated ion signal is that the intensity of the ejected ion is much more intense. The disadvantage is that it can take many seconds, sometimes up to 60 s, to collect a single scan on an 8.4T instrument. Therefore this configuration is typically not amenable to the LC timescale.

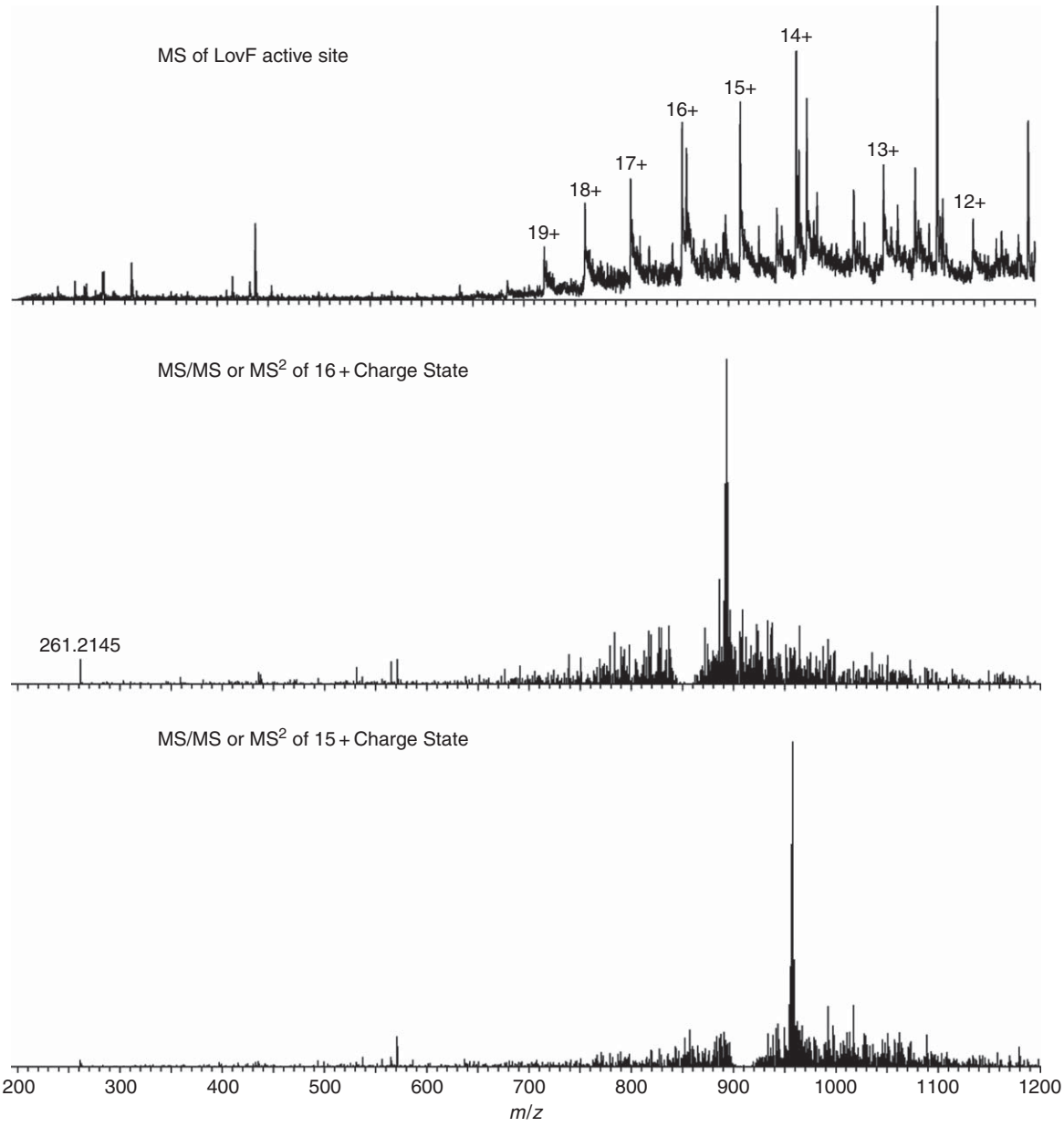


Figure 9 Illustration of the 1/3rd rule. The top panel shows multiple charge states of the phosphopantetheinylated active site containing peptide from PKS LovB. The 16+ ion at 854.636 *m/z* and the 15+ ion at 911.636 *m/z* of the peptide were subjected to CID. The middle panel shows that PPant ejection (261 *m/z*) is clearly detected in the MS/MS or MS^2 spectrum resulting from CID of the 16+ ion. The bottom panel shows that PPant ejection resulting from fragmentation of the 15+ ion cannot be detected due to limitations defined by the 1/3rd rule.

9.11.1.8 PEA on Non-ICR Instruments: Low-Resolution Phosphopantetheinyl Ejection

PEA is an important discovery in the characterization of phosphopantetheinylated proteins. It allows one to analyze nonradioactively labeled substrates thereby making many more substrates accessible to biosynthetic studies of these types of proteins. In addition, this approach often eliminates the need for preparing synthetic standards for comparison of hydrolyzed thioester intermediates, which may be challenging and time consuming to accomplish. Moreover, PEA immediately shows that the substrate or intermediate is connected to the PPant posttranslational modification and not elsewhere on the protein. This is not possible to detect with radioactivity or any other

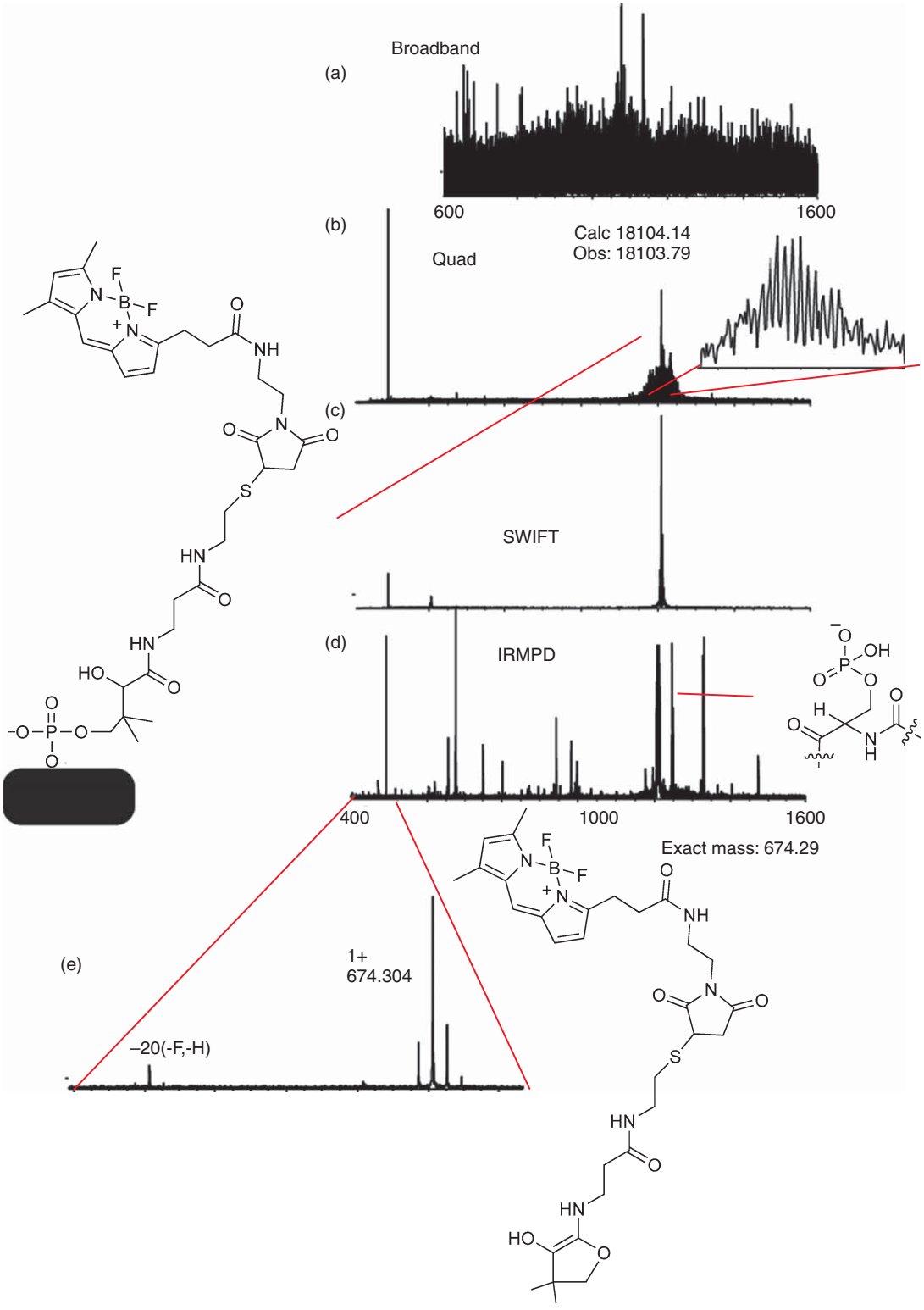


Figure 10 The accumulation multipole setup for PEA. Ions of interest are selected by the quadrupole (a), accumulated in the accumulation octupole (b), isolated by SWIFT (c), and subjected to PEA by IRMPD (d,e) in the analyzer cell.

conventional assay used to study these types of systems. An additional advantage that PEA has over radioactive assays is that it provides a mass signature and therefore unexpected intermediates can be observed. Even though PEA is such a useful tool in the investigation of substrates and biosynthetic intermediates, it has not been widely used by other investigators because it requires costly FT–ICR–MS instrumentation and highly skilled researchers to use them. Because most MS instruments are able to perform source fragmentation, CID, or both, this method should be readily be applicable to low-resolution instruments. Since properly calibrated low-resolution instruments have mass accuracies well within 0.5 Da, they can be used to differentiate molecular species within 1 Da for the 1+ ejected PPant ions, making such an instrument a very useful tool for the characterization of phosphopantetheinylated proteins. As PEA can be applied to low-resolution instruments, it is an assay that is accessible to most researchers studying phosphopantetheinylated proteins. The challenge of this assay is that even though the ejected ion in the low m/z range is often the most abundant one, there are other ions in this region of the spectrum. Due to these other ions and to noisier low-resolution detectors, which are not based on image current measurement, it can be more difficult to confirm the candidate ejected ion than anticipated. This confirmation can be carried out using labeling studies or directly by PPant fragmentation, a new MS^3 (additional fragmentation of ions generated by MS/MS or MS^2) method, if an instrument has the ability to perform MS^3. When this MS^3 method was originally reported, detection of 12 diagnostic ions were reported for the second round of MS/MS or MS^2 on the ejected ion, but there are many more fragments with lower abundance, including an important and abundant ejected ion fragment that reports on the substrates loaded onto the thiol of the phosphopantetheinyl functionality (**Figure 11**, Dorrestein, unpublished observations).[81] Thus far, the confirmation of the ejected ions has only been accomplished on an ion trap instrument in two different modes. In the first mode, the phosphopantetheinylated protein is observed by MS, the phosphopantetheinylated ions are isolated in the ion trap and subsequently subjected to CID (MS/MS or MS^2). Then the ejected ion is isolated and fragmented again using CID (MS^3). This will result in the diagnostic fragment pattern for the phosphopantetheinylated peptide ejection ions. The second mode relies on the initial use of source fragmentation followed by CID on the ejected ions in the low m/z region (MS^2). In this case, the method can be performed in a data-dependent manner, as it is done for proteomic experiments, to find the characteristic phosphopantetheinyl 'MS^3' signature from any of the ions that enter the instrument. The main disadvantage of the source fragmentation

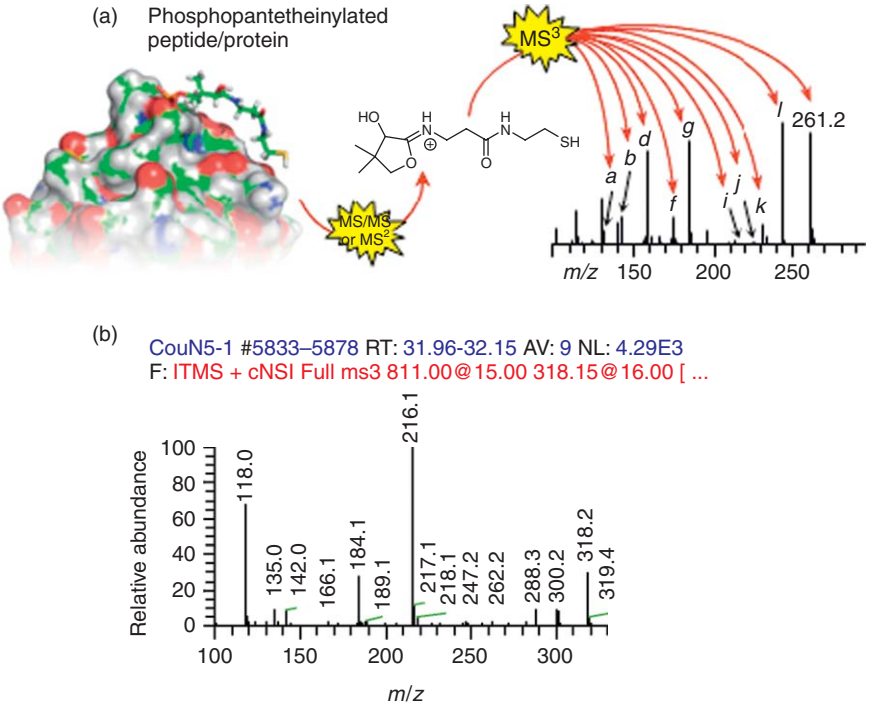


Figure 11 The PPant fragmentation method. (a) A phosphopantetheinylated protein is first subjected to the PEA and, subsequently, the PPant ejected ion is fragmented again to yield diagnostic MS/MS or MS^2 peaks (b).

method, as mentioned earlier, is that all the information of the precursor ion is lost, but it should still enable the analysis of phosphopantetheinylated proteins on nonion trap, low-resolution instruments that have only one stage of MS capabilities by additional confirmation of the PEA ions.

9.11.1.9 Low-Resolution Capillary LC–MS on Ion Traps

To date, FT–ICR–MS instruments are the main instruments used for PEA but they are not the best instruments to analyze NRPS/PKS on an LC timescale because, currently, their scan rates are very long. The scan rates of most low-resolution instruments are shorter and, therefore, much better suited for interfacing with LC. In proteomics, it has become standard to use small-bore columns (30–100 $\mu mol\,l^{-1}$ inner diameter) with nanoflow (200–500 $nl\,min^{-1}$) gradients, but this has not yet found much use in the investigations of biosynthetic pathways. More recently, with the ability to perform PEA, these LC–MS approaches have emerged as useful tools in the investigation of phosphopantetheinylated peptides and proteins. The advantage of LC with a 100 $\mu mol\,l^{-1}$ column as opposed to traditional 1, 2.1, or 4.6 mm HPLC columns is that it uses much less material. In the case of a 4.6 mm column, injection of >100 μg of sample is not uncommon to attain good signals, while in the case of a 100 $\mu mol\,l^{-1}$ diameter column, one typically uses 0.1–1 μg of material. This approach has recently been used to observe the phosphopantetheinylation of a carrier domain from the hemolytic toxin pathway from *Streptococcus agalactiae* and its corresponding phosphopantetheinyl ejection.[81]

9.11.1.10 Mass Spectrometry of Intact NRPS and PKS Multidomain Proteins

There have only been a few reports on the MS of intact multidomain NRPS and PKS proteins.[82] The NRPS proteins GrsA and NikP1 have been analyzed by FT–ICR–MS but not with isotopic resolution. In addition, the 6-deoxyerythronolide B synthase (DEBS) PKS didomain has been investigated by MS.[82] In most cases, a mass shift can be observed upon phosphopantetheinylation or substrate loading. The main challenge with intact NRPS and PKS protein analysis is to obtain sufficient quantities of very pure proteins. A small amount of contamination by a small peptide or small domain will compete for the signal very efficiently, making it nearly impossible to observe the larger ion. While the examples of GrsA, NikP1, and DEBS illustrate that it is possible to interrogate these larger proteins, it should even be possible to analyze entire NRPS and PKS multiprotein complexes such as the ones observed for bacillaene.[83] These NRPS/PKS multiprotein complexes are similar in size to intact viral particles or intact ribosomes that have been investigated by MS.[84,85] The MS of intact protein and multiprotein complexes is an area that remains wide open for exploration.

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

In many respects, the phosphopantetheinyl functionality is similar to a phosphoserine and treatments that work with phosphorylated proteinaceous materials will often work with phosphopantetheinylated materials as well. For instance, phosphatases are responsible for the removal of phosphates but were recently shown to remove the phosphopantetheinyl functionality as well.[86] While this assay has not yet been utilized on the biosynthesis of natural products, this assay has been applied to the investigation of the phosphopantetheinylated protein 10-formyltetrahydrofolate dehydrogenase, which is involved in the formation of formyltetrahydrofolate.[86] Donato *et al.* used the phosphatase assay to confirm the presence of the phosphopantetheinylation. The characterized phosphopantetheinyl modification suggests that there may be many other phosphopantetheinylated proteins that have not been identified. This assumption has recently been confirmed using a phage display approach.[87] In the study by Donato *et al.* a common matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) instrument was used for analysis. One caution about their interpretation of the spectra should be noted and it is a direct result that the experiments were performed by a novice and not an MS expert. In their spectra, these researchers observed a mass of 360.08 Da and in their text they report that the mass of the ion should be 358.33 Da, which is not correct. We recalculated the mass of the hydrolyzed ion and anticipate that the mass of this hydrolyzed species should be 359.104 Da in the positive mode while in the negative mode it is 357.089 Da. In addition to the monomer unit, the authors saw a dimer with a mass of 550.60 Da where both phosphates had been removed. According to our calculations, this dimer should have a mass of 555.252 Da for a

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

9.11.1.12 The Development of Recognition Software for PEA on an LC Timescale

While there are many proteomics programs designed to find peptides or neutral ion losses such as phosphoric acid from phosphopeptides, there are no programs or algorithms developed that can identify phosphopantetheinylated peptides. The main challenges in the annotation of phosphopantetheinylated peptides by such programs is the observation that the ejected ions are often very abundant and, therefore, limiting to the intensity and number of normal fragment ions typically encountered with peptides. In the case of phosphopeptides, there were enough examples available that software could be trained to recognize such peptides.[89] In the case of phosphopantetheinylated proteins, there are a limited number of such data training sets with which to train new software. In addition, phosphopantetheinylated peptides have two phosphoester linkages while phosphopeptides only have one, making it more likely that this ion is ejected instead of amide cleavage. Finally, the thiol of the phosphopantetheinyl functionality is modified with substrates and intermediates making it even harder to identify and find phosphopantetheinylated peptides. Such modifications to phosphopantetheinylated peptides will need to be taken into account when conducting searches for active sites. Current efforts are underway to overcome some of these limitations. Once solutions to this problem are obtained, it will become possible to study the biosynthesis of the therapeutics that are biosynthesized on phosphopantetheinylated proteins at their native levels using proteomic approaches.

9.11.2 Applications of Mass Spectrometry on NRPS Systems

NRPs are important bioactive and medicinally applied natural products, including compounds such as cephalosporins, penicillins, and vancomycin. These natural products are biosynthesized by NRPSs, one of the two thiotemplate biosynthetic machineries found in fungal and microbial secondary metabolism. As mentioned before, two factors made NRPS characterization difficult until the late 1990s: (1) their large size (up to 700 kDa) and (2) their complex multimodular structure comprising a diversity of catalytic and carrier domains. Consequently, the characterization of NRPS intermediates and substrates from enzymatic assays was a laborious task exemplified by the dissection of the gramicidin biosynthetic system by Stein *et al.*[90] Herein, multiple steps of radioactive labeling of carrier protein active sites, digests, HPLC separations, and low-resolution mass spectrometric analysis were required to identify all biosynthetic substrates from milligram quantities of the corresponding NRPS proteins GS1 and GS2. In 1999, Shaw-Reid *et al.*[41] applied electrospray ionization Fourier transform mass spectrometry (ESI–FTMS) to characterize enterobactin NRPS enzymology, which simplified and accelerated the identification process of NRPS intermediates significantly. In addition, they reduced the required protein amount to microgram quantities. Since then, MS has held its initial promise as a key method to investigate NRP biosynthesis by the improvement of instrumentation of large molecule MS in combination with advanced molecular biology techniques and the development of innovative MS-based assays to study NRPS – such as the PEA[50] and the substrate screening assay.[91] Many methods and investigations of NRPS biosynthetic pathways and their tailoring reactions are summarized in the comprehensive 2006 review by Dorrestein and Kelleher.[47] This section will highlight the recent applications of the modern MS approaches applied to NRPS proteins since this 2006 review and the progress that has been made since then.

In the 2006 review, the basic experimental procedures to detect intermediates on NRPS carrier proteins are explained and the mechanistic insights for the application of those thiotemplate systems gained by MS are described. This section connects to this review by summarizing recent MS-based studies of nonribosomal biosynthetic machineries from 2006 until January 2009 and emphasizes the application of the new MS methods described in the previous section of this review. As described above and in the review by Dorrestein and Kelleher, the major advantages of MS as an investigative tool of NRPS systems are: (1) *In vitro* substrate identification of NRPS domains by mapping of T domain active sites. The direct detection of T domain-tethered substrates and intermediates allows the investigation of substrate specificity of NRPS catalytic domains and tailoring enzymes. (2) Parallel detection of NRPS T domains and rapid quench methods allow the investigation of relative T domain occupancy by intermediates, the investigation of intermediate flux and pseudokinetic interrogation of NRPS downstream processing. (3) Substrate screening methods allow verification of predicted substrates and characterization of orphan NRPS gene clusters by fast substrate identification from a complex substrate pool.

Since most current papers that are published on NRPS systems utilize some form of MS, we have divided the application of MS methods to understand NRPS pathways into five different categories: (1) investigation of substrate specificity in NRPS systems, (2) characterization of new NRPS enzymology and of deviations from colinearity, (3) characterization of tailoring reactions, (4) characterization of multistage assembly line action, and (5) time courses.

In the following sections, it is explained briefly what makes NRPS biochemistry accessible to mass spectrometric investigation and a new active site mapping approach for NRPS and PKS systems, the online LC–MS–PEA assay. Subsequently, each of the above categories of MS applications in NRPS biosynthesis research is presented based on recent research.

9.11.2.1 MS Accessibility of NRPS Systems

NRPS enzymology has been outlined in several excellent reviews.[23,92] There are three main features of NRPS that make these biosynthetic systems amenable for MS characterization. The first feature is that all substrates and intermediates are covalently tethered to carrier domains. The covalent tethering of intermediates allows their detection by isolation of the carrier protein active sites. Therefore, NRP chain elongation and tailoring reactions along the biosynthetic pathway can be characterized by detection of corresponding mass changes of active site-bound intermediates. Second, NRPs are usually biosynthesized along a thiotemplate with multiple carrier domains. The timing of biosynthetic events on an NRPS assembly line can be dissected by parallel mass spectrometric detection of those multiple carrier proteins and their covalently bound intermediates. Third, substrates and intermediates are covalently bound on the carrier proteins as thioesters. The weakness of the thioester C–O bond enables the cleavage of intermediates from carrier proteins as PPant species and, therefore, enables the PEA (see Section 9.11.1.6).

9.11.2.2 Active Site Screening and PPant Ejection Assay in NRPS Investigation

One of the most important steps in NRPS characterization by MS is active site mapping, that is, isolation and confirmation of a carrier protein active site in the mass spectrometer. Currently, there are two experimental types of techniques to map active sites by MS – offline-LC active site mapping and online-LC active site mapping. The general offline-LC active site mapping approach and recent techniques to accelerate active site detection are reviewed by Dorrestein and Kelleher.[47] Since 2006, online LC–MS and the PEA have been applied to speed up active site mapping even more. The LC–MS–PEA active site mapping approach has been used to map NRPS active sites by the Kelleher group[51,93] and the Marahiel group.[42,52] It is generally pursued as follows and as shown in **Figure 12**.

The first step is the priming of the NRPS active site and a subsequent limited tryptic digest of the protein. The digested sample is loaded on a reverse-phase liquid chromatography (RPLC) C18 column, which is directly connected to the inlet of an FT mass spectrometer. During online LC separation, the eluent is analyzed by MS and MS^2 on an LC timescale. In the mass spectrometer the eluent is first analyzed by broadband Fourier transform mass spectrometry (FTMS). Then, peaks in the resulting broadband FT mass spectrum are

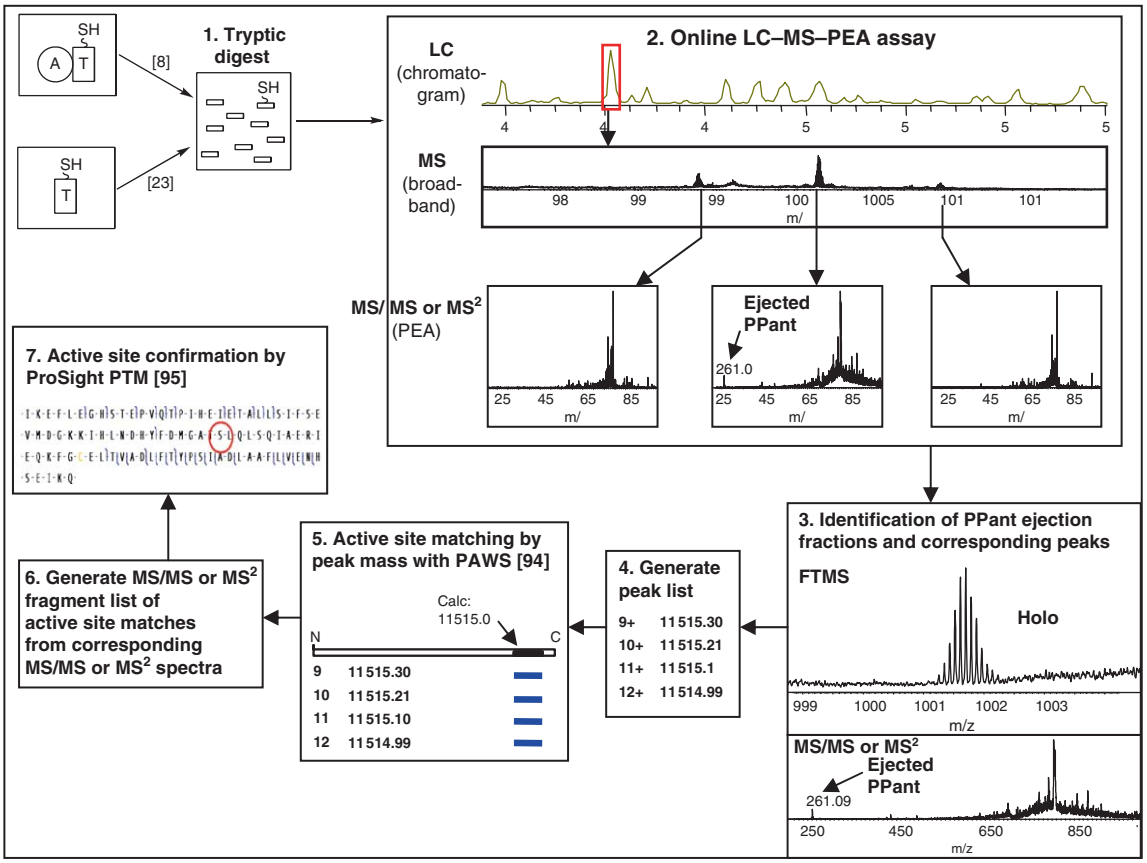


Figure 12 Active site mapping by online LC–FT–ICR–MS–PEA assay.

fragmented selectively and successively and MS^2 data for each mass peak are collected for a programmed number of scans. This MS^2 analysis of eluting mass peaks is called data-dependent analysis because the eluent MS^2 analysis is only carried out on occurring mass signals. The data-dependent fragmentation of eluting peaks enables the PEA simultaneously. The streamline collection of MS and MS^2 data during the LC–MS–PEA assay shortens the time of MS measurement for active site mapping to ~1 h per experiment. The data analysis for mapping the active site within the LC chromatogram first utilizes PPant ejection data to characterize elution fractions that comprise active site fragments. The occurrence of the ejected PPant species 261.127 m/z and 359.104 m/z indicates that the fragmented peak within an elution fraction was phosphopantetheinylated and, therefore, might be an eluting active site fragment. The corresponding mass of the putative active site fragment can be determined as described by Dorrestein and Kelleher[47] from the FTMS broadband spectrum and mapped in the NRPS amino acid sequence by PAWS.[94] If an active site fragment could be identified, the MS^2 peaks of the identified active site fragment are analyzed with ProSight PTM[95] in order to confirm the mapped sequence by detected b- and y-ions.

Ultimately, the LC–MS–PEA active site mapping approach has advantages and disadvantages in comparison to recent offline-LC active site mapping approaches. Major advantages are the acceleration and simplification by automatization of active site mapping and the substantial decrease of required protein quantities. A disadvantage is the limited accumulation time of ions to obtain high-quality MS data and, thus, the limited number of peaks that can be selected and fragmented during data-dependent MS^2 analysis. In addition, the required instrumentation is also more expensive as the collection of the FT–ICR–MS data on an LC timescale from a complex protein sample demands stronger magnets for faster accumulation time. Therefore, the LC–MS–PEA assay for active site mapping is currently ideally accomplished on a 12T FT–ICR–MS instrument[93] but it was already conducted on a 7T FT–ICR–MS instrument.[51]

Offline-LC active site mapping might still be ideal for active site mapping of complex NRPS systems because active site fragments can appear as low intensity peaks relative to the other coeluting ions in the FT broadband mass spectrum and, consequently, then might not trigger data-dependent fragmentation and analysis on the LC timescale in the online-LC approach. In offline-LC active site mapping experiments, collected fractions can be directly infused into the FT–ICR–MS and data can be collected for an extended period of time in order to increase the quality of weak peptide signals. Offline-LC approaches have gained more reliability from the PEA that is an additional step to confirm a primed active site fragment by ejected PPant species. Current examples of applied offline-LC active site mapping are investigations of NRPS systems of the prodigiosin biosynthesis,[96] the microcystin biosynthesis,[97] and the vibrobactin biosynthesis.[98]

9.11.2.3 Investigation of Substrate Specificity in NRPS Systems

One of the main applications of MS in NRPS research is the investigation of NRPS substrate specificity. Herein, MS is used for two purposes: substrate identification and determination of substrate tolerance of different NRPS domains. Both approaches are referred to as substrate specificity assays in this section. MS-based substrate identification is the *in vitro* and *in vivo* characterization of the native substrates utilized by a catalytic NRPS domain or tailoring enzyme during biosynthesis of an NRP natural product. MS-based determination of substrate tolerance is the *in vitro* characterization of alternative substrates that a catalytic NRPS domain or tailoring enzyme uses besides its native substrates.

Prior to MS-based substrate specificity assays, certain NRPS substrate specificities can be predicted by bioinformatics. Adenylation domain substrates can be predicted based on their '10 letter code'[99,100] by substrate prediction tools such as the NRPS predictor.[101] Methyltransferases can be predicted in their substrates and methylation sites by bioinformatic analysis too.[102] In addition, substrates of catalytic NRPS domains and tailoring enzymes can be predicted by the structure of the known NRP natural product. Either way, predicted substrates of NRPS domains need to be experimentally verified. A traditional technique to determine substrate specificity of an A domain is the adeonsine triphosphate–pyrophosphate (ATP–PP_i) exchange assay. The ATP–PP_i exchange assay characterizes substrates indirectly by observing the radioactive pyrophosphate incorporation into ATP from a reverse reaction with pyrophosphate and the acyl–adenylate of the substrate.[103] Because the PP_i exchange measures the back exchange of pyrophosphate into ATP, the determined substrate can deviate from the true substrate as it may be only the kinetically most competent substrate of the reverse adenylation reaction. In contrast to this assay, MS has become a more reliable tool to identify NRPS substrates because it determines the true substrate specificity by detection of the complete adenylation reaction product, that is, the substrate tethered on a T domain.

Before selected publications are presented, a general guideline of MS-based substrate specificity assays for specific NRPS domains is given. This guideline emphasizes three experimental aspects of an MS-based NRPS substrate specificity assay. First, which MS instrumentation should be applied? Second, which substrates have to be considered for a catalytic NRPS domain or tailoring enzyme? Third, which substrate pool can be utilized to identify a native or alternative substrate?

The applied MS instrumentation for a substrate specificity assay depends on the size of the carrier domain construct whose active site is monitored. If multidomain constructs or T domain constructs >20 kDa are investigated, active site mapping is required and, therefore, ESI–FTMS instruments have to be applied because of their established active site mapping capabilities described above. If the monitored T domain constructs are freestanding and <20 kDa, no active site mapping is required and low-resolution MS instruments can be utilized for characterization of loaded substrates too.

Each type of catalytic NRPS domain and tailoring enzyme has characteristic substrates. Herein, a differentiation can be made between biosynthetic substrates and carrier protein substrates. A biosynthetic substrate is a building block or intermediate of the NRP natural product. For instance, an amino acid can be the biosynthetic substrate of an A domain or an NRP intermediate can be the biosynthetic substrate of a tailoring enzyme. A carrier protein substrate is the T domain recognized by a catalytic NRPS domain or tailoring enzyme for substrate loading or turnover of tethered biosynthetic substrates. In the following, biosynthetic substrates and carrier protein substrates of A domains, C domains, tailoring enzymes, and aminoacyl transferases (**Figure 13**) are summarized.

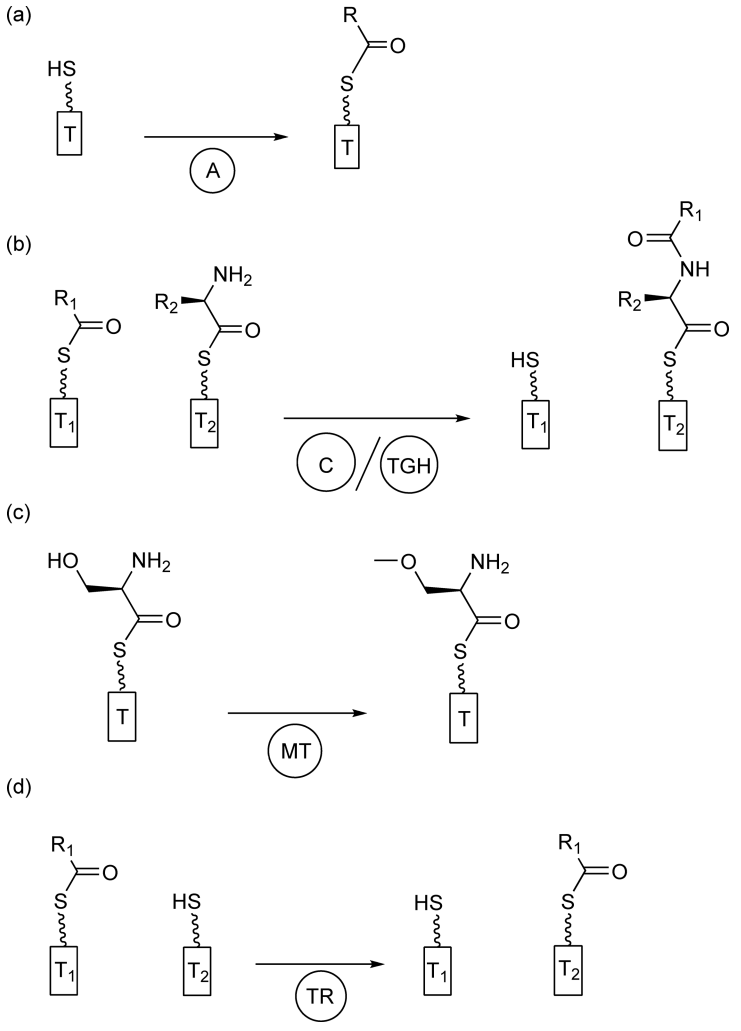


Figure 13 General substrate specificity assays. (a) Adenylation domain. (b) Condensation/TGH domain. (c) Tailoring enzyme, for example, *O*-methyltransferase. (d) Aminoacyl transferase.

Adenylation domains utilize free amino acids,[97] aryl acids,[104] or fatty acids,[51] biosynthetic substrates, and one T domain for tethering their substrates on the thiotemplate. C domains have an upstream nucleophile and a downstream electrophile as biosynthetic substrates, and two T domain substrates: one upstream and one downstream T domain.[93,105] Transferases, such as the aminoacyl transferase CmaE in the crotonine biosynthetic pathway,[106] are similar to C domains in that they use two T domain substrates but they have only one biosynthetic substrate tethered to the downstream T domain. Tailoring enzymes can have two forms of biosynthetic substrates: T domain-bound substrates[52,55] or non-T domain-bound substrates. Non-T domain-bound substrates can be free carbon acids,[46,107] CoA-activated species, or the analogue of the natural product lacking the assayed chemical modification.

MS-based substrate specificity assays that are applied to NRPS domains are mainly aimed to characterize native and alternative biosynthetic substrates. **Table 1** presents biosynthetic substrate pools for substrate screening of catalytic NRPS domains and tailoring enzymes. Biosynthetic substrate pools are specified for the two purposes of assaying NRPS substrate specificity as mentioned above: (a) substrate identification and (b) determination of substrate tolerance. The biosynthetic substrate pools for each specific domain and for each experimental purpose are explained as follows.

Table 1 Biosynthetic substrate specificity assays for different NRPS domains and tailoring enzymes: Substrate pools for two experimental purposes

Domain	*Purpose of substrate specificity assay*	*Biosynthetic substrate pool*	*References*
Adenylation domain	Substrate identification	(1) Multiple biosynthetic substrates, e.g., algal lysate	91 – *in vitro* screening 97 – *in vivo* screening
		(2) Predicted native biosynthetic substrate	104, 107, 108 – *in vitro* screening 97 – *in vivo* screening
	Substrate tolerance	(1) Multiple biosynthetic substrates excluding native substrate	
		(2) Single biosynthetic substrate except native substrate	107 – *in vitro* screening 97 – *in vivo* screening
Condensation domain/ TGH domain	Substrate identification	Predicted native nucleophile and electrophile tethered to upstream and downstream T domain, respectively	105
	Substrate tolerance	Nonnative nucleophile and electrophile tethered to upstream and downstream T domain, respectively	93
Tailoring enzyme	Substrate identification	(1) Predicted native substrate tethered to T domain	52, 55, 105
		(2) Predicted native substrate	52, 105, 109
		(3) NRP natural product analogue lacking predicted chemical modification	52, 105
	Substrate tolerance	(1) Nonnative substrate tethered to T domain	52, 55
		(2) Nonnative substrate	109
		(3) NRP natural product derivatives	
Transferase	Substrate identification	Predicted native substrate tethered to downstream T domain	106
	Substrate tolerance	Nonnative substrate tethered to downstream T domain	106

Adenylation domains are the only domain type that can be studied by substrate screening with multiple biosynthetic substrates in one assay. All other NRPS domains are screened for biosynthetic substrates in a 'one-assay-one-substrate' approach. If the investigated biosynthetic substrate is T domain bound, it has to be tethered to a T domain substrate by a promiscuous PPTase, for example, Sfp, before the enzymatic reaction. If it is non-T domain bound, it is incubated with the enzyme without prior loading on a carrier protein. Biosynthetic substrate identification of A domains can be done either with a multiple substrate pool, for example, algal lysate, or with the predicted native biosynthetic substrate only. As noted above, substrate specificity assays are usually conducted *in vitro*, but recently, *in vivo* substrate specificity assays for an A domain were done with McyG adenylation–thiolation (A–T) didomain by coexpression with PPTase Svp in *Escherichia coli* and the application of varying growth media.[97] Biosynthetic substrate tolerance of A domains can be investigated *in vitro* by incubation with a multiple substrate pool lacking the native substrate or single substrates except the native substrate.

Biosynthetic substrate identification of C domains is pursued by tethering the predicted native nucleophile and electrophile to the corresponding native upstream and downstream T domains, respectively, and detection of the condensation product on the upstream active site after the enzymatic reaction. Biosynthetic substrate tolerance of C domains is characterized by the same approach as for substrate identification, except that nucleophiles and electrophiles different than from the biosynthetic pathway are screened. Electrophile biosynthetic substrate tolerance is screened with alternative electrophiles on the downstream T domain in the presence of the native nucleophile on the upstream T domain. Nucleophile biosynthetic substrate tolerance is screened with alternative nucleophiles on the upstream T domain in presence of the native electrophile on the downstream T domain.

Aminoacyl transferases can be characterized in their native biosynthetic substrate by tethering the predicted biosynthetic substrate on the downstream T domain and MS detection of substrate transfer to the upstream T domain upon incubation with the transferase. Biosynthetic substrate tolerance of transferases is tested by loading biosynthetic substrates differing from the native substrate on the downstream T domain and by the same approach as for substrate identification.

MS-based substrate identification assays of tailoring enzymes depend on whether the biosynthetic substrate is T domain bound or not. If the biosynthetic substrate is T domain bound, the predicted native substrate is loaded on the T domain and the mass change upon the tailoring reaction is detected by ESI–FTMS. If the biosynthetic substrate is non-T domain bound, conversion of the predicted native substrate by tailoring reaction can be detected by low-resolution MS. Biosynthetic substrate tolerance assays for tailoring enzymes are conducted in a 'one-assay-one-substrate' approach like substrate identification assays but with alternative biosynthetic substrates.

MS-based assays that are aimed to characterize the specificity of catalytic NRPS domains and tailoring enzymes for carrier protein substrates can be done on high-resolution mass spectrometers or, for small substrate T domains (<20 kDa),[105] on low-resolution mass spectrometers. For investigation of T domain substrate tolerance, the native T domain substrates of a catalytic NRPS domain or tailoring enzyme are exchanged by different T domains, for example, from different NRPS systems. In addition, the tolerance of T domain order can be tested for C domains and aminoacyl transferases by reverse-ordered tethering of native biosynthetic substrates to the native T domains and MS detection of the reaction product on the assayed upstream active site.[93]

Most of the latest publications on NRPS substrate specificity are focused on A domain specificity because their substrate screening is straightforward in terms of biosynthetic substrate form (free amino acids/fatty acids/aryl acids) and T domain substrates (one T domain). Four studies focus on substrate specificity of NRPS loading modules of microcystin biosynthesis,[97] mycosubtilin biosynthesis,[51] daptomycin biosynthesis,[108] and leinamycin biosynthesis.[108] The A domains of microcystin, mycosubtilin, and daptomycin biosynthesis initiation showed fatty acid specificity. The initial domain from leinamycin biosynthesis has D-amino acid specificity. Another paper presents the elucidation of aryl acid-specific AsbC adenylation enzyme from petrobactin biosynthesis.[104]

The first example in which MS was utilized to determine the specificity of an A domain from an NRPS gene cluster was published in a joint effort by the Moore and Kelleher laboratories. Hicks *et al.*[97] investigated the substrate specificity of the loading module McyG of microcystin synthetase. Microcystin is a cyclic NRPS–PKS hybrid toxin derived from various cyanobacteria genera. The initiation module McyG comprises an A–T didomain (**Figure 14(b)**), which was predicted based on the '10 letter code' to activate and load

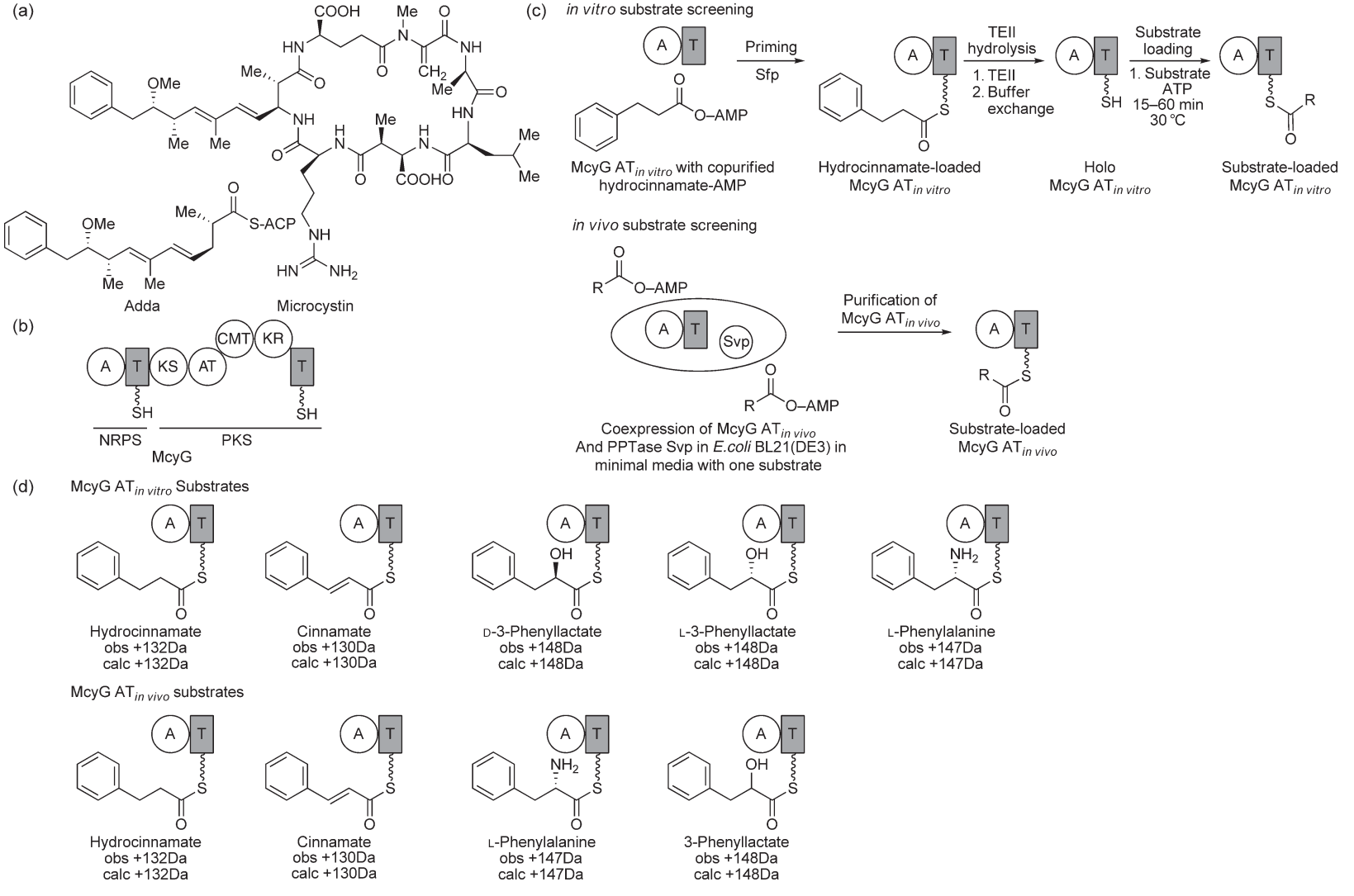


Figure 14 *In vivo* and *in vitro* substrate screening of loading module of microcystin biosynthesis. (a) Microcystin and Adda. (b) Loading protein McyG of microcystin synthetase. (c) *In vitro* and *in vivo* substrate screening assays of McyG AT. (d) Characterized substrates of McyG $AT_{in\ vivo}$ and McyG $AT_{in\ vitro}$ by ESI–FTMS (observed and calculated mass shifts from holo McyG AT active site).

phenylacetate as a starter unit for subsequent polyketide extension and formation of the aromatic β-amino acid (2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid (Adda) within the microcystin structure (**Figure 14(a)**). FT–ICR–MS was utilized to determine the substrate specificity of this domain. Two McyG AT constructs were used for substrate identification – holo $AT_{in\ vivo}$ and holo $AT_{in\ vitro}$ The holo $AT_{in\ vivo}$ construct was generated by heterologous coexpression of McyG AT with Svp PPTase and allowed *in vivo* substrate screening and, subsequently, the first detection of *in vivo* McyG intermediates (**Figure 14(c)**). *In vivo* and *in vitro* phosphopantetheinylation of McyG A–T was confirmed by MALDI–TOF MS, which allowed detection of intact McyG AT (>78 kDa). Nonetheless, no holo McyG $AT_{in\ vivo}$ and low holo $AT_{in\ vitro}$ could be detected. Furthermore a posttranslational modification with a higher mass was detected on both McyG species and was proposed as a copurified substrate bound to the McyG AT active site. This putative substrate was characterized by electrospray ionization–Fourier transform–ion cyclotron resonance–mass spectrometry (ESI–FT–ICR–MS) on the mapped T active site to have a mass that corresponded to hydrocinnamate. Subsequent hydrolysis and small-molecule MS by gas chromatography/electron impact–mass spectrometry (GC/EI–MS) confirmed this hypothesis. This is a surprising finding because it is not clear how hydrocinnamate could be used in the microcystin pathway. Because *in vivo* feeding studies could not confirm the expected phenylacetate substrate, specificity of McyG AT, Hicks *et al.* also tested alternative substrates of McyG by ATP–PP_i exchange assay. Surprisingly, the A–T didomain accepted a wide range of phenylpropanoids including hydrocinnamate and was in agreement with these original findings by MS. To follow up on these observations, the McyG AT substrate specificity was further investigated by MS methods. *In vitro* substrate screening was done by hydrolysis of hydrocinnamate to free the thiol of the phosphopantetheinyl arm of the purified AT by a type II thioesterase. Once the free thiol was obtained, the sample was buffer exchanged so that the released dihydrocinnamate is removed from the reaction mixture to allow subsequent mass spectrometric activity screens with other phenylpropanoids (**Figure 14(c)**). This confirmed the substrate tolerance of McyG loading module by loading of five phenylpropanoids with different efficiencies (**Figure 14(d)**). In addition, *in vivo* substrate screening of holo McyG $AT_{in\ vivo}$ was carried out. This was done by *E. coli* growth in complex medium leading to the observed preference of hydrocinnamate loading. However, supplementation of complex growth medium with excess cinnamate led to preference of cinnamate loading on McyG AT and growth in defined media led to alternative substrate loading of 3-phenyllactate and 3-phenylalanine (**Figure 14(d)**). In summary, this study proved substrate tolerance of McyG for phenylpropanoids but not the predicted phenylacetate. This study, as well as an additional recent study that identified that the substrate phenylalanine of GrsA copurifed with a heterologously expressed A domain, indicate that this may not be an uncommon phenomenon in the investigations of NRPS proteins *in vitro*. In such a case, MS will be critical to understand what is loaded onto T domains. The characterization of *in vivo* intermediates by MS as introduced by Hicks *et al.* is the first example that demonstrates the *in vivo* reconstitution of NRPS systems in complement to MS-based *in vitro* reconstitution approaches. Therefore, it is the first example that connects *in vivo* substrate loading, albeit the protein is overproduced.

In a paper by Hansen *et al.*[51] both biochemical assays and FT–ICR–MS were utilized for characterization of the substrate specificity for another NRPS initiation module: the loading module of mycosubtilin biosynthesis. Mycosubtilin, a potent antifungal natural product of the iturin class of cyclic lipopeptides, is a β-amino fatty acid-containing octapeptide (**Figure 15(a)**). The initial protein of mycosubtilin biosynthesis is MycA, which is predicted to comprise an acyl ligase (AL) and a T domain as a loading module (**Figure 15(b)**). This didomain is called MycA10 and is a loading module with fatty acid specificity. Although being an AL, MycA10 is considered in this NRPS section because of its similar chemistry to NRPS A domains and its upstream position to an NRPS assembly line. Hansen *et al.* showed by *in vitro* assays with radiolabeled decanoic acid that the AL activates its substrate by adenylation. Substrate tolerance of the AL for fatty acids with a 10–16 carbon-comprising chain was detected by a radiolabeled chase experiment. The loading of a fatty acid on MycA was further characterized by the online LC–MS–PEA assay. AL-T_1 didomain was incubated with ATP and decanoic acid for 1.25 h and subsequently limited digested by trypsin. The digestion mixture was separated by reverse-phase high-performance liquid chromatography (RP–HPLC) and T domain-tethered decanoic acid was detected by LC–MS and data-dependent MS/MS or MS^2 of intact peptides for T domain active site mapping and by PEA (**Figure 15(c)**). This was accomplished using a 4.6 mm diameter C18 column on a 7 tesla FT–ICR–MS

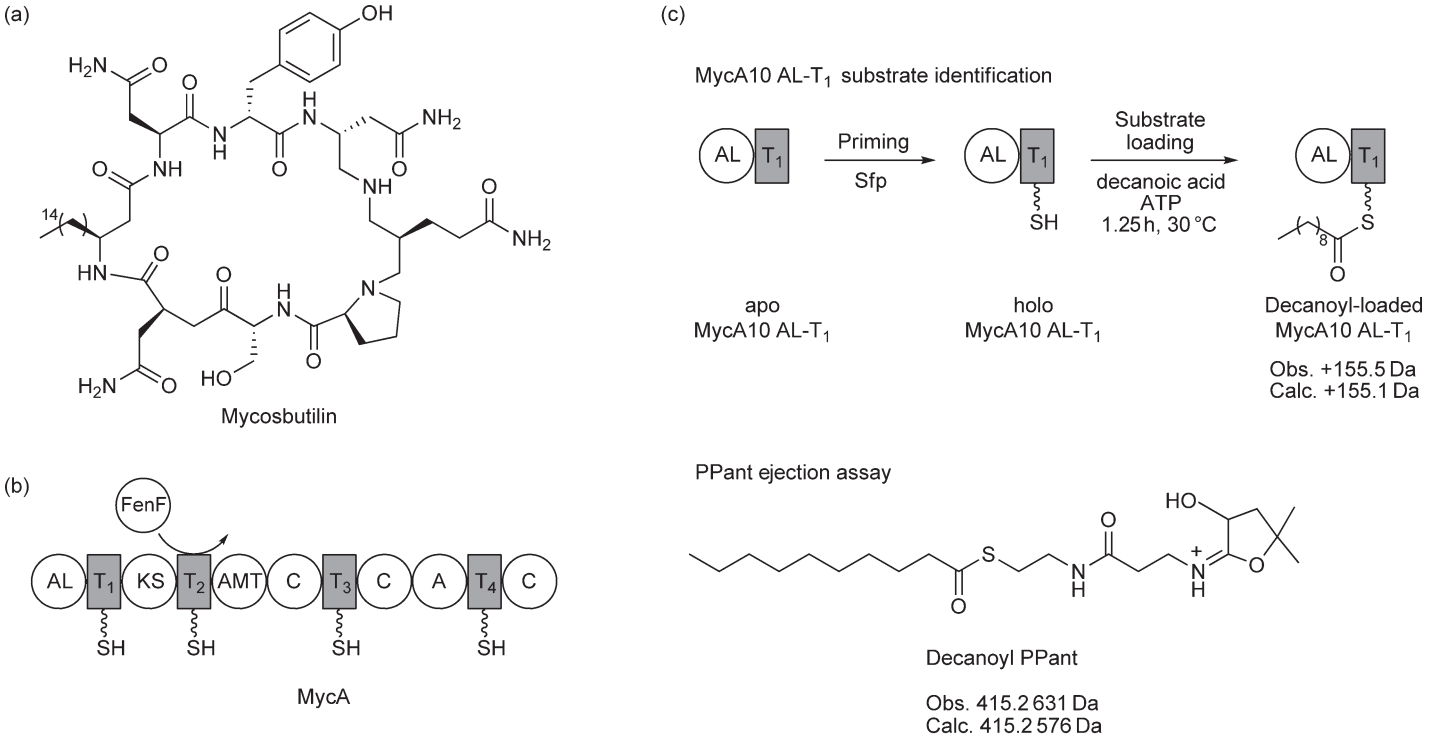


Figure 15 Substrate identification of McyA10 AL-T_1 loading module of mycosubtilin biosynthesis. (a) Mycosubtilin. (b) Loading protein MycA of mycosubtilin biosynthesis. (c) MycA10 AL-T_1 substrate identification by ESI–FTMS (observed and calculated mass shift from T_1 active site) and PEA.

instrument. This study characterized the loading module of mycosubtilin biosynthesis, MycA10, as an A domain by biochemical assays and confirmed a predicted fatty acid substrate specificity by ESI–FT–ICR–MS.

Another example of a substrate identification of an A domain is a study by Wittmann *et al.*[109] about the lipidation of daptomycin. Daptomycin is a clinically important semisynthetic derivative of the A21978 branched cyclic lipopeptide isolated from *Streptomyces roseosporus*. It comprises a 13 amino acid peptide core coupled to a decanoic acid moiety (**Figure 16(a)**). Wittmann *et al.* characterized the putative adenylating enzyme DptE as a fatty acid adenosine monophosphate (AMP) ligase that activates the natural substrate decanoic acid by adenylation and, subsequently, tethers it to the freestanding T domain DptF (**Figure 16(a)**). Thiolation activity of DptF was assayed by Bodipy labeling[110] and ESI–FTMS detection of holo DptF. Lipidation activity of DptE was characterized by detection of decanoic acid adenylation and loading on DptF (**Figure 16(b)**). Substrate tolerance of the adenylation enzyme DptE for alternative biosynthetic substrates and T domain substrates was investigated by ATP–PP_i exchange assay. Herein, Wittmann *et al.* showed that DptE can load various fatty acid substrates on DptF but not on other T domains.

Tang *et al.*[108] characterized the substrate specificity of the NRPS initiation module of leinamycin biosynthesis by ATP–PP_i exchange assay and offline-HPLC electrospray ionization mass spectrometry (ESI–MS). Leinamycin is a hybrid NRP–polyketide natural product isolated from *Streptomyces atroolivaceus* S-140. It comprises a D-alanine and shows potent antitumor activity. The NRPS initiation domain of leinamycin biosynthesis consists of a freestanding A domain, LnmQ, and a freestanding T domain, LnmP (**Figure 17(a)**), which enabled detection of apo-, holo-, and substrate-loaded T domain by low-resolution ESI–MS. The MS-based substrate screening assays were conducted by testing one substrate per assay. D-Alanine and glycine loading on LnmP was detected by ESI–MS and ATP–PP_i exchange (**Figure 17(b)**). Hereby, D-alanine was identified as the native A domain substrate because it resulted in a higher ATP–PP_i exchange activity by LnmQ, whereas glycine was characterized as an alternative substrate of the investigated A domain. LnmQ is the first known A domain with D-amino acid stereospecificity.

Finally, the characterization of adenylation enzyme AsbC within the petrobactin biosynthetic pathway by Pfleger *et al.*[104] is an example of an adenylation enzyme with aryl acid specificity. Although petrobactin is an NRPS-independent siderophore from *Bacillus anthracis* (**Figure 18(a)**), it is covered in this NRPS section too, because of the NRPS homologue enzymology of adenylation enzyme AbsC and freestanding T domain AsbD (**Figure 18(b)**). The native substrate and substrate tolerance were determined by single substrate screens and LC–ion trap MS. Therefore, the detection of AsbC specificity for aryl acid 3,4-dihydroxybenzoic acid (3,4-DHBA) is another example for substrate identification by low-resolution MS in combination with the ATP–PP_i exchange assay. AsbC substrate tolerance for multiple substituted benzoic acid derivatives could be detected by the same approach (**Figure 18(c)**).

Substrate specificity of a transglutaminase homologue (TGH) domain, which is a new type of NRPS C domain was characterized by MS application within two studies from the Walsh group.[93,105] First, Fortin *et al.* identified the substrates of TGH domain AdmF from the andrimid biosynthetic pathway by electrospray ionization–quadrupole–time-of-flight–mass spectrometry (ESI–Q–TOF MS). Then, Magarvey *et al.* investigated AdmF substrate tolerance by ESI–FTMS. Andrimid is a hybrid NRP–polyketide antibiotic isolated from various bacteria that shows nanomolar inhibition of the bacterial acetyl-CoA carboxylase. The six-module hybrid NRPS–PKS assembly line is of interest because of six interfaces between NRPS and PKS enzymology and a new type of amide bond forming C domain. At the first NRPS–PKS interface, an octatrienoyl group tethered on the T domain AdmA and (*S*)-β-Phe tethered on the T domain AdmI are condensed by TGH AdmF to give the intermediate octatrienoyl-β-Phe-*S*-AdmI (**Figure 19(a)**).

Fortin *et al.*[105] determined the native AdmF substrates by condensation product formation on the upstream T domain of AdmF, which is AdmI. The predicted electrophile was confirmed as an octatrienoyl moiety tethered to the downstream T domain AdmA. The predicted nucleophile was confirmed as a β-phenylalanine bound to the upstream T domain AdmI. The AdmF reaction products octatrienoyl-β-Phe-*S*-AdmI and holo AdmA were detected by ESI–Q–TOF MS (**Figure 19(a)**).

In the study by Magarvey *et al.*[93] substrate tolerance of multiple components of the NRPS–PKS system for andrimid biosynthesis was studied by the application of biochemical radiolabel assays and MS assays. Magarvey *et al.* characterized the stereoselective formation of (*S*)-β-Phe from L-Phe by aminomutase AdmH by incubation with radiolabeled substrates and HPLC analysis. The (*S*)-β-Phe specificity of the downstream, freestanding A

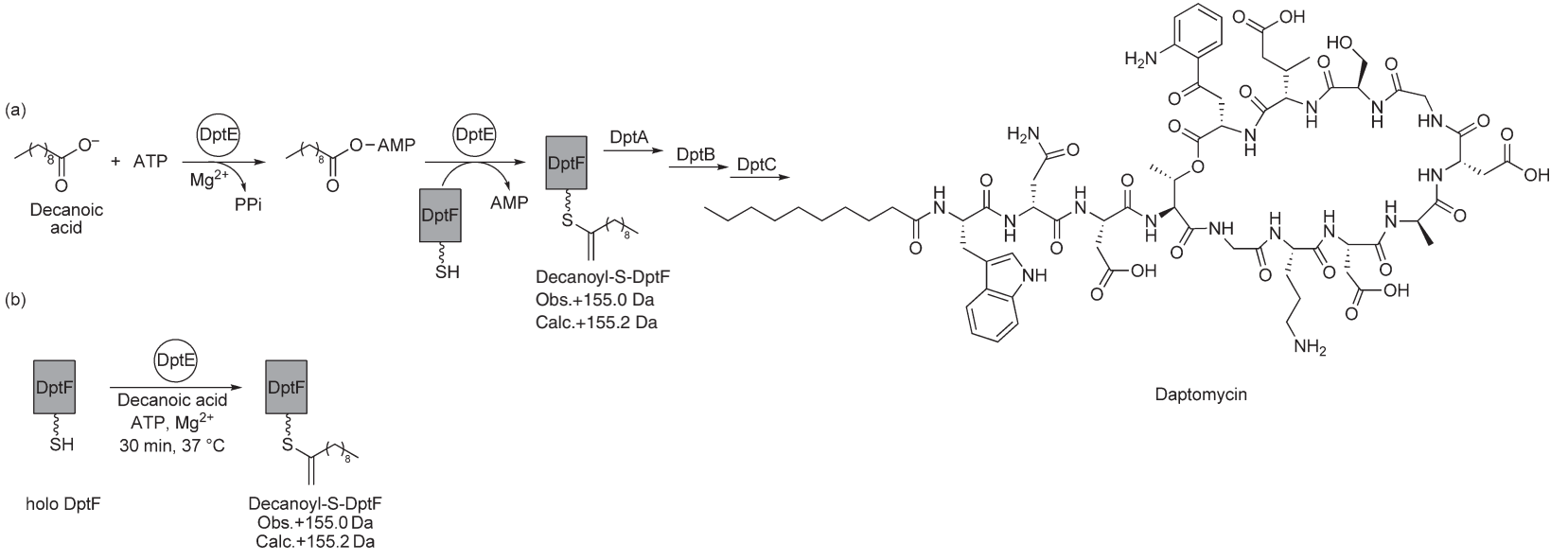


Figure 16 Lipidation of daptomycin. (a) Role of DptE and DptF in daptomycin lipidation. DptE adenylates decanoic acid and tethers it on T domain DptF for insertion into daptomycin. (b) DptE substrate identification assay (observed and calculated mass shifts from holo DptF characterized by ESI–FTMS).

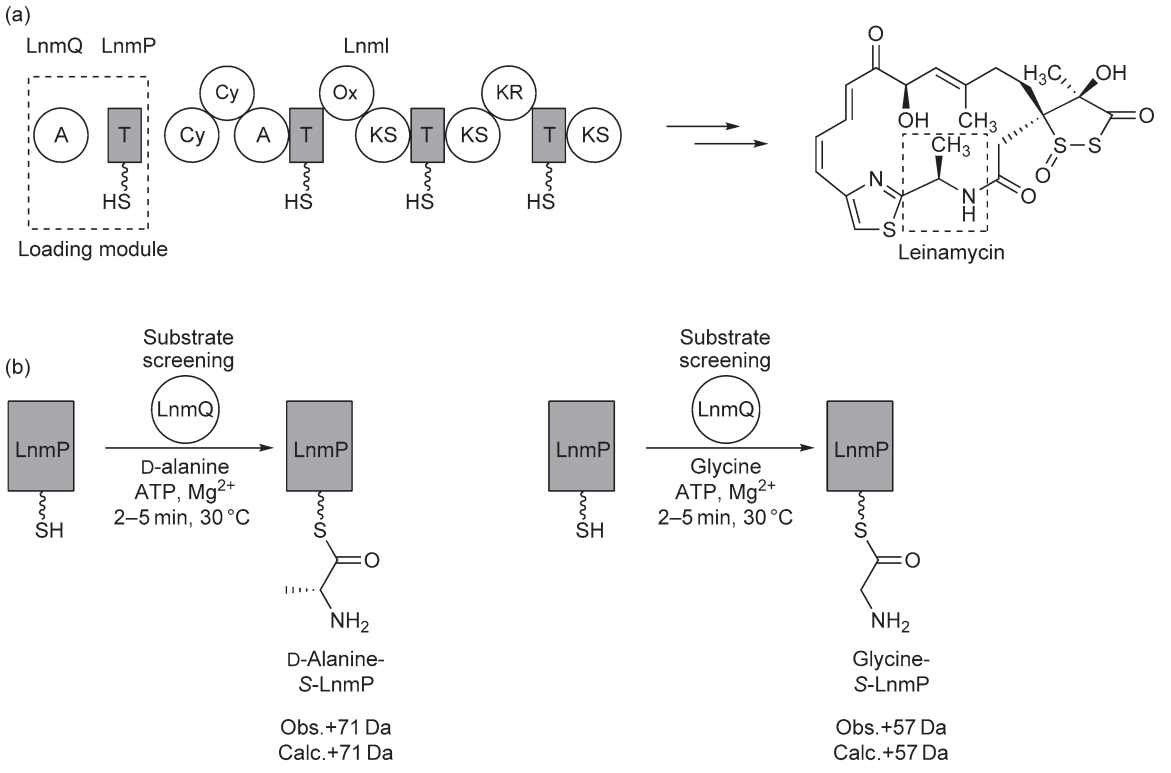


Figure 17 The loading module of leinamycin biosynthesis. (a) Loading module components (LnmQ and LnmP) of leinamycin synthetase and leinamycin structure. (b) Substrate screening assays of LnmQ. D-Alanine and glycine loading was detected by ESI–MS (observed and calculated mass shifts of holo LnmP).

domain AdmJ, which adenylates and loads (*S*)-*β*-Phe only on the T domain AdmI, was identified by ATP–PP_i exchange assay. Thus, AdmH and AdmJ are considered as specificity gatekeepers within the early stages of andrimid biosynthesis. In addition, the substrate tolerance of TGH AdmF was characterized in terms of nucleophile specificity, electrophile specificity, and T domain specificity. It was shown that AdmF was specific for the T domains AdmA and AdmI in any order (**Figure 19(b)**). AdmF was characterized by MS to accept a wide array of acyl chain donors (**Figure 19(d)**) and nucleophile amines (**Figure 19(c)**) for amide bond formation, which shows its relaxed substrate specificity. All AdmF reaction products were identified by broadband FTMS of T domain-tethered intermediates and PEA through the online LC–MS–PEA assay. Therefore, AdmF was characterized as a promiscuous enzyme.

The substrate identification and determination of substrate tolerance of tailoring enzymes by MS is exemplified by two recent studies from the Marahiel group about calcium-dependent antibiotic (CDA) biosynthesis.[46,52] CDA is another nonribosomal lipopeptide, like daptomycin and mycosubtilin, with bioactivity against multidrug-resistant pathogens. CDA comprises an 11 amino acid chain cyclized to a 10-membered ring and it contains two characteristic functionalities – a unique 2,3-epoxyhexanoyl moiety and a *β*-hydroxyasparagine residue (**Figure 20(a)**).

In the first study, Kopp *et al.*[52] investigated epoxidation enzymes within CDA biosynthesis to form the 2,3-epoxyhexanoyl moiety. Two putative oxygenases, HxcO and HcmO, were cloned, expressed, and characterized in terms of substrate specificity by ESI–FTMS. HxcO was predicted as an acyl-CoA dehydrogenase that catalyzes C2–C3 bond dehydrogenation of an alkanoic acid substrate and subsequent epoxidation of the C2–C3 double bond. HcmO was predicted as a flavin-dependent monooxygenase that epoxidizes an alk-2-enoic acid substrate. First, the substrate forms of the putative epoxidation enzymes were determined. Three forms of the HxcO-predicted native substrate hexanoic acid were tested for epoxidation by HxcO: hexanoyl-CoA, hexanoyl-CDA analogue, and T domain-bound hexanoic acid. The free substrate and natural product analogue

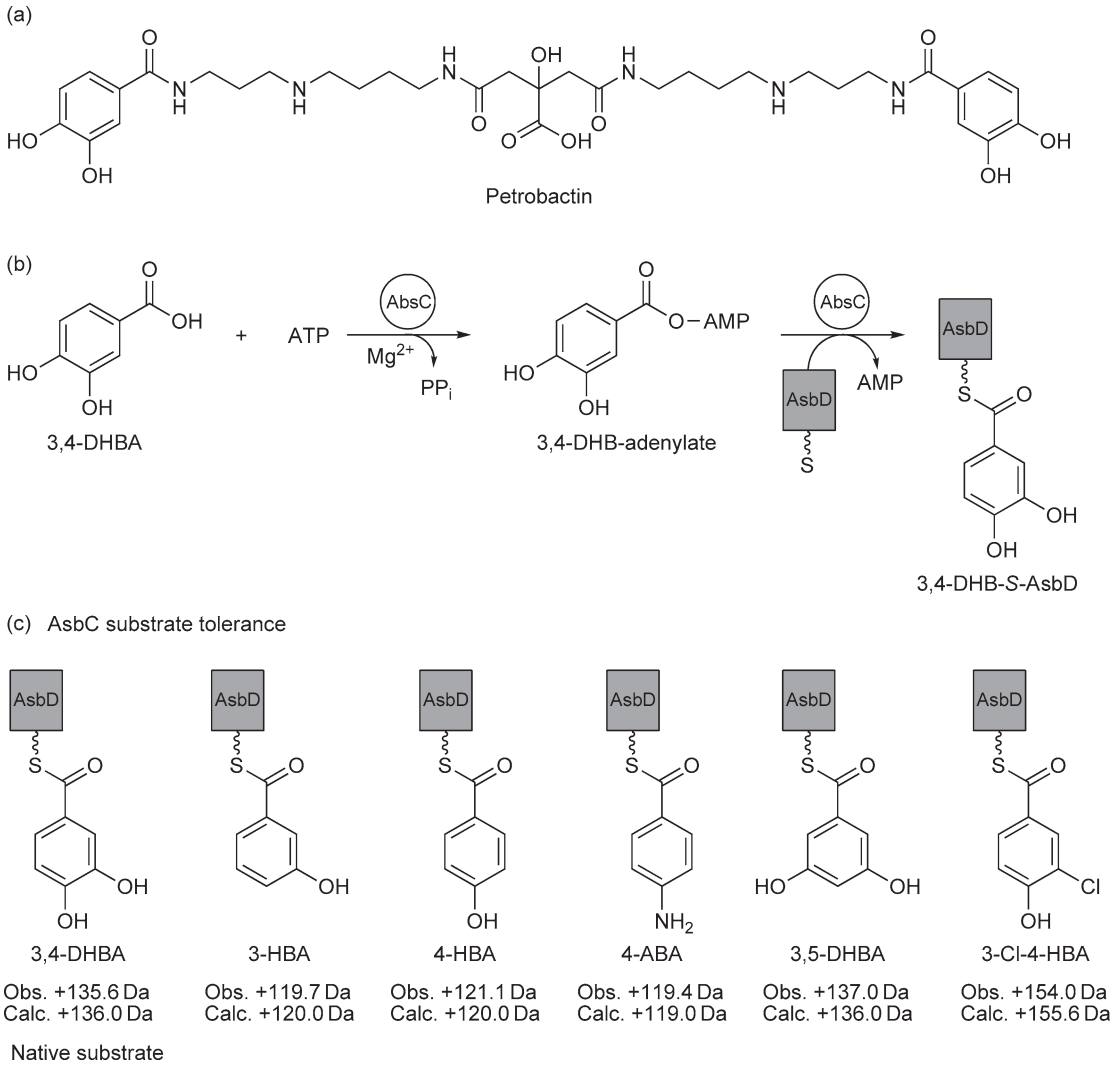


Figure 18 Adenylation enzyme AsbC from petrobactin biosynthesis has aryl acid specificity. (a) Petrobactin. (b) AsbC enzymology. AsbC adenylates native substrate 3,4-DHBA and tethers it to thiolation domain AsbD. (c) AsbC substrate tolerance characterized by LC–IT–MS (observed and calculated mass shifts of AsbD).

were not epoxidized but the T domain tethered substrate was. For HcmO, hexenoyl-CoA and T domain-bound hex-2-enoic acid were assayed as HcmO substrates and also only T domain-bound substrate epoxidation was detected. Epoxidation of the hexanoyl-*S*-T by HxcO and epoxidation of hex-2-enoyl-*S*-T by HcmO was characterized by online HPLC–ESI–FTMS and PEA (**Figure 20(b)**). In addition, substrate tolerance of HxcO and HcmO was investigated by testing epoxidation of various T domain-bound alkanoic and alkenoic acids, respectively. Herein, one alternative substrate was screened per assay and epoxidation of the substrates was characterized as before. Kopp *et al.* showed that HxcO could epoxidize various T domain-bound fatty acid substrates with different chain lengths whereas HcmO showed only epoxidation of one alternative substrate (crotonyl-*S*-T), which is similar to the HcmO natural substrate (**Figure 20(c)**). Overall, this study is a very good reference of MS-based substrate specificity assays for tailoring enzymes.

The second study on CDA tailoring enzymes is the substrate identification of the nonheme Fe^{2+}/α-ketoglutarate-dependent oxygenase AsnO, which was predicted to catalyze Cβ-hydroxylation of Asn9 side chain to yield the CDA functionality β-hydroxyasparagine. Strieker *et al.*[46] showed in this study by MS that AsnO is not hydroxylating a CDA analogue lacking the corresponding hydroxyl group or a T

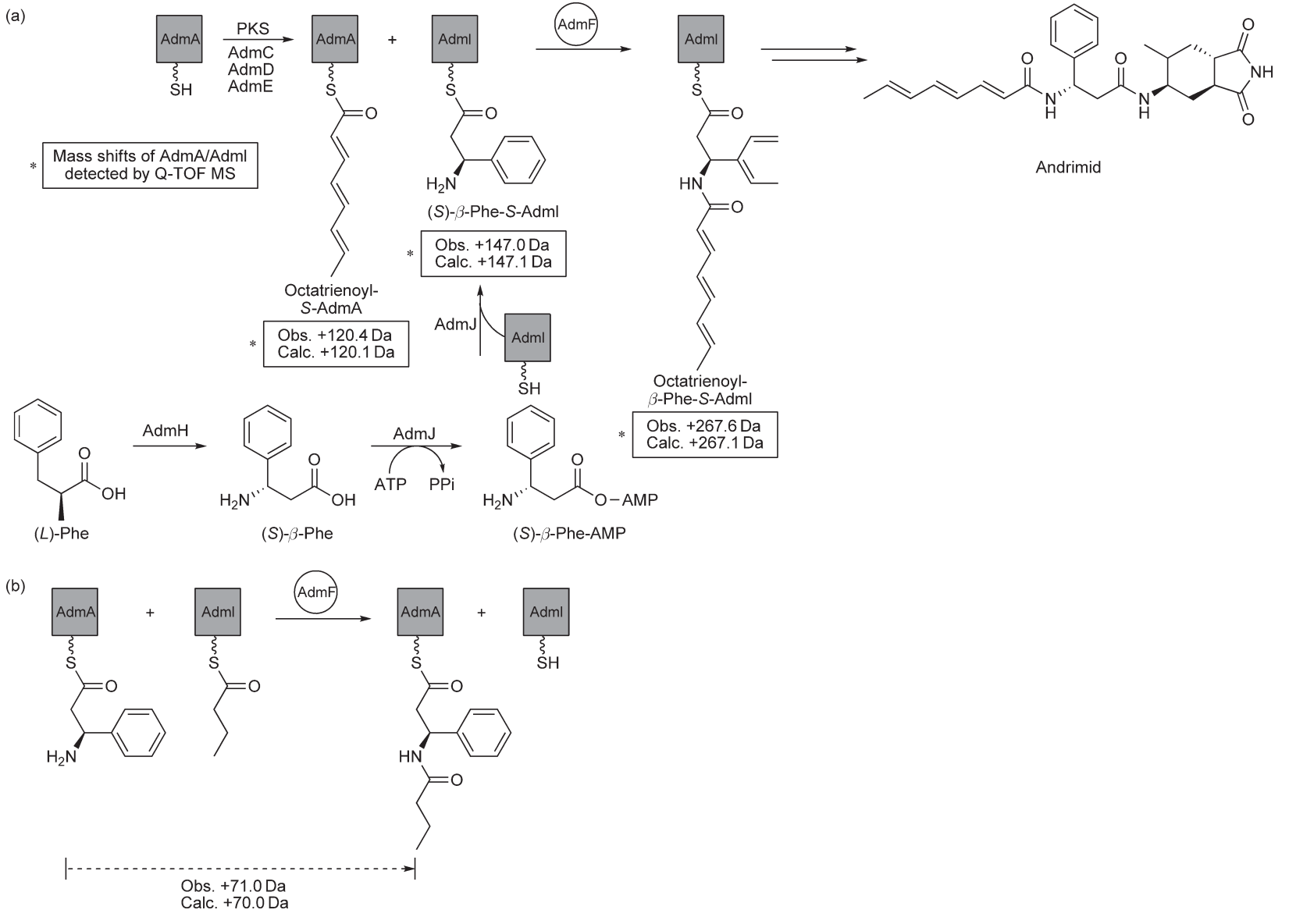
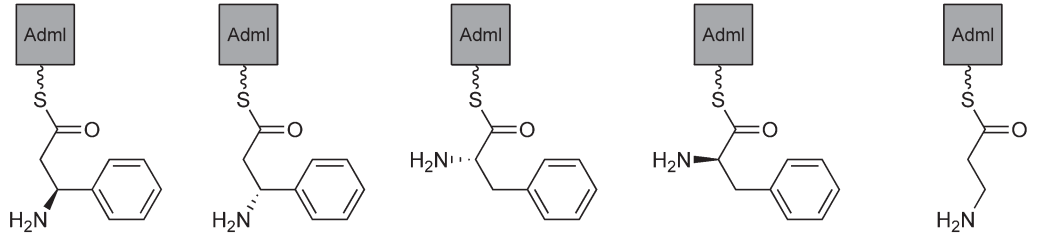
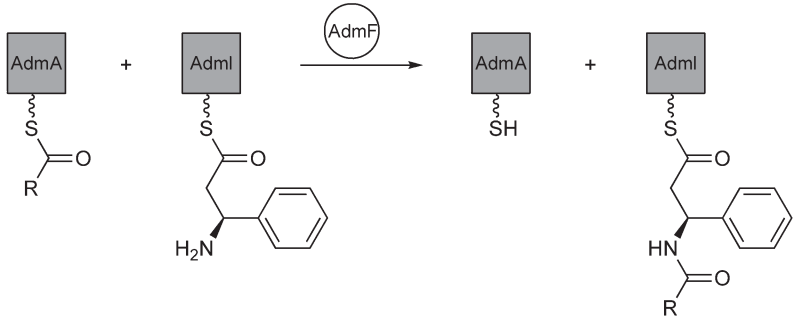


Figure 19 (Continued)

(c) AdmF nucleophile tolerance



(d) AdmF electrophile tolerance



	R	CH_3								
Mass shift of AdmI (Da)	Obs.	189.1	218.1	241.1	276.1	302.2	328.2	384.3	265.1	294.1
	Calc.	189.1	217.1	241.1	273.2	301.2	329.2	385.3	265.1	294.1

Figure 19 Investigation of substrate specificity in early stages of andrimid biosynthesis. (a) Early stages in andrimid biosynthesis and characterization of AdmF reaction. AdmF condensation product octatrienoyl-β-Phe-*S*-AdmI was detected by ESI–Q–TOF MS (observed and calculated mass shifts of AdmI/AdmA). (b) AdmF tolerance for reverse T domain order. AdmF condensation product was detected on AdmA by ESI–FTMS. (c) AdmF nucleophile tolerance. Depicted nucleophiles were characterized by ESI–FTMS to undergo AdmF condensation with butyryl-*S*-AdmA. (d) AdmF electrophile tolerance. Electrophiles depicted in the table were characterized by ESI–FTMS to undergo AdmF-catalyzed condensation with (*S*)-β-Phe.

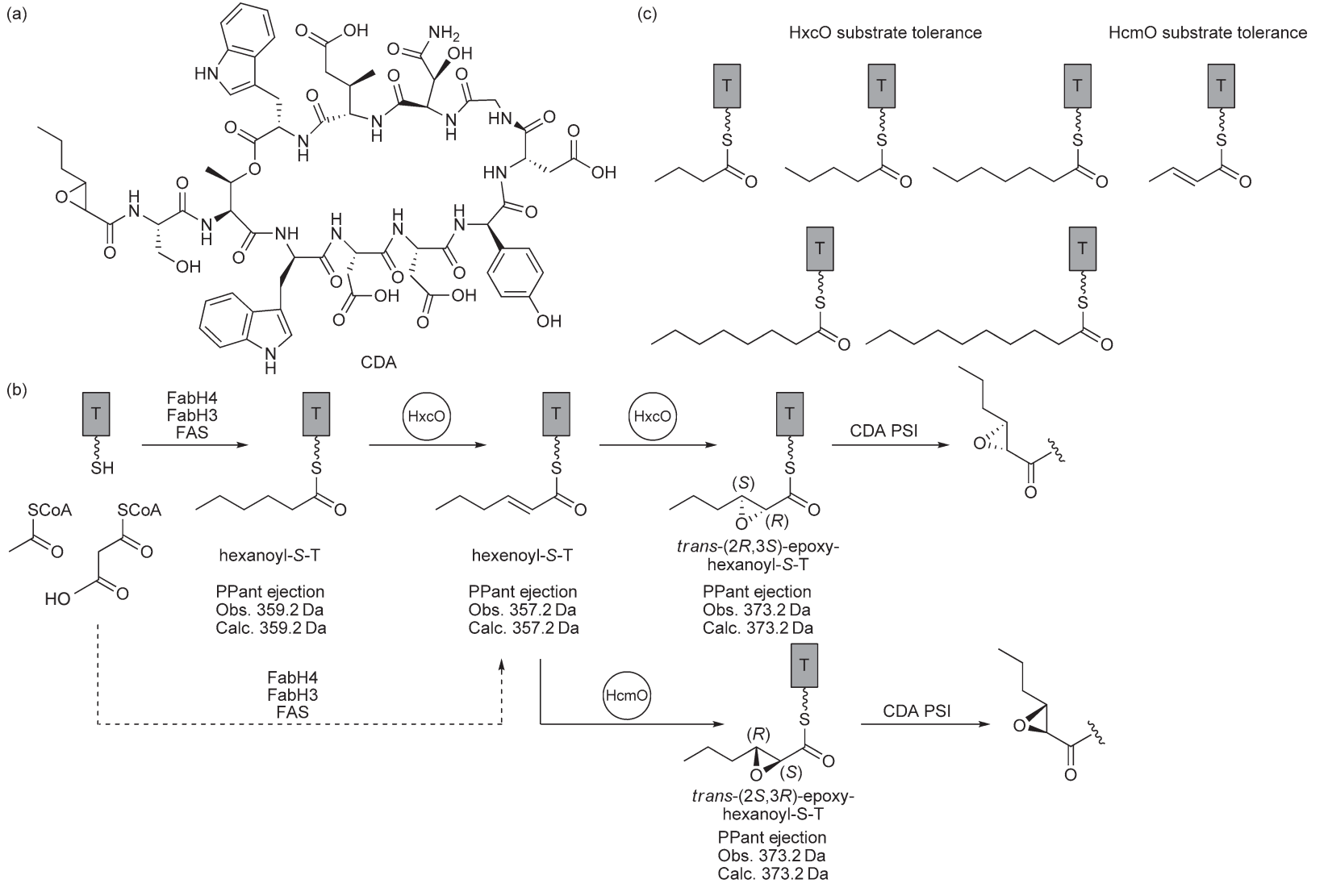


Figure 20 CDA epoxidation by tailoring enzymes HxcO/HcmO. (a) CDA. (b) Proposed mechanism of HxcO/HcmO epoxidation to form *trans*-2,3-epoxyhexanoyl moiety, reaction intermediates, and products were characterized by online LC–FTMS and PEA. The stereochemistry of HxcO/HcmO reaction products was characterized by the amide ligation assay.[52] (c) HxcO and HcmO substrate tolerance characterized By LC-IT-MS.

domain-bound L-/D-asparagine because no +16 Da mass shifts upon hydroxylation were detected after AsnO reactions with these putative biosynthetic substrates. Therefore, the most likely substrate was a free amino acid, that is, L- or D-asparagine. X-ray could identify the free amino acid L-asparagine as the native AsnO substrate.

Further recent MS applications to investigate substrate specificity were the characterization of aminoacyl-transferase CmaE in coronamic acid biosynthesis pathway[106] (**Figure 21(a)**). Purified CmaE transferred various chemically differing aminoacyl groups between various T domains, which were detected by MALDI–TOF MS, and therefore CmaE promiscuity was identified (**Figure 21(b)**).

The studies by Fortin *et al.*,[105] Strieter *et al.*,[106] Pfleger,[104] and Tang[108] showed that it is not always necessary to apply high-resolution MS for substrate identification on NRPS systems and for determination of their substrate tolerance.

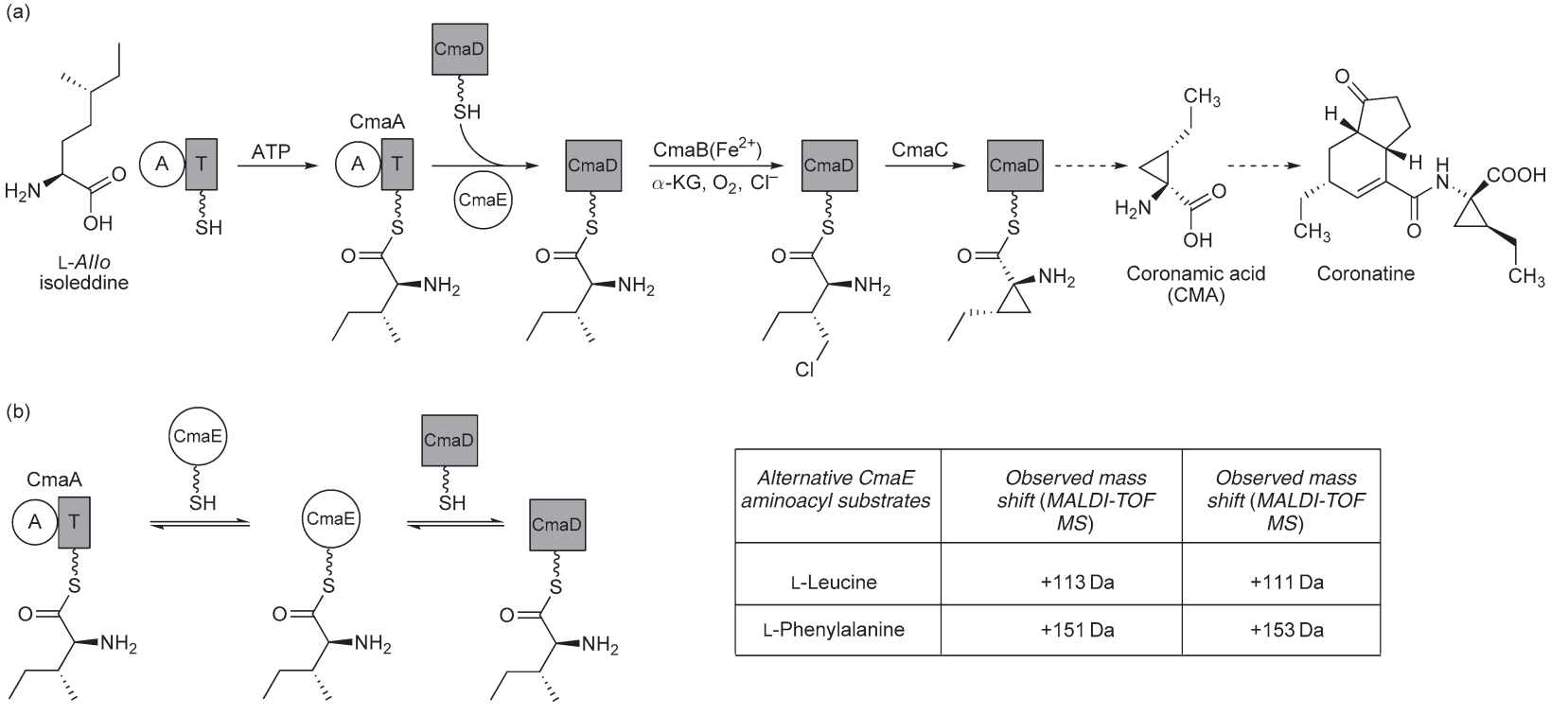
9.11.2.4 Dissection of New NRPS Enzymology and of Deviations from NRPS Colinearity

Several recent studies applied MS for mechanistic insights into new NRPS enzymology and tailoring reactions. As mentioned earlier, a new type of thiotemplate C domain, the TGH AdmF, was predicted within the andrimid biosynthesis. Before the described study by Magarvey *et al.*[93] characterizing the AdmF promiscuity, Fortin *et al.*[105] dissected amide bond formation catalyzed by AdmF applying HPLC and ESI–Q–TOF MS for detection of condensation product octatrienoyl-β-Phe tethered on AdmI T domain. In contrast to regular NRPS C domains, the TGH C domain performs covalent catalysis by an acyl–enzyme intermediate on an active site cysteine (C90) within a catalytic triad Cys–His–Asp (**Figure 22**). This mechanism of a new type of thiotemplate C domain was dissected by radiolabeling of the acyl–AdmF intermediate and its inactivation by site-directed mutagenesis (C90A).

In another mechanistic study of NRPS catalytic domains, Stein *et al.*[42] differentiated epimerization (E) domains of NRPS assembly lines into aminoacyl E domains and peptidyl E domains by monitoring intermodular transfer. The characterized differences in intermodular transfer activity by the two types of E domains from the tyrocidine biosynthesis (one from an initiation module and one from an elongation module) were done by formation of the two dimodule constructs $TycB_{2\text{-}3}$-AT-CATE/COM_{tycA} with the initiation E domain of tyrocidine biosynthesis at its C-terminus and $TycB_{2\text{-}3}$-AT-CAT/E_{tycA} with an elongation E domain of tyrocidine biosynthesis at its C-terminus. For both the constructs, Phe–Phe dipeptide formation rate was characterized by ESI–FTMS quantification of $TyrB_{2\text{-}3}$ T_3 domain-loaded intermediates. $TyrB_{2\text{-}3}$ T_3 domain active site was mapped by preliminary fluorophore labeling, HPLC separation, and online LC–ESI–FTMS analysis of labeled fractions. For investigation of intermodular transfer of aminoacyl or dipeptidyl groups by E domains, a reporter construct $TyrB_1$-CAT/TE_{srf} was formed. $TyrB_1$-CAT could be recognized by both E domains and it monitored aminoacyl transfer by dipeptide formation (**Figure 23**(**a**,1)) and peptidyl transfer by tripeptide formation (**Figure 23**(**a**,2)). Both products were detected and quantified as offloaded molecules by HPLC–ESI–TOF MS after incubation of each of the E domain constructs with the reporter construct. The E domain of the initial module showed more activity of aminoacyl transfer to the downstream T domain and is called aminoacyl E domain. This type of E domain causes misinitiation within an NRPS assembly line by unselective formation of a D-amino acid and its subsequent intermodular transfer. The E domain of the TycB elongation module showed only selective D-Phe formation in Phe–Phe dipeptidyl intermediate and subsequent transfer of L-Phe–D-Phe to the reporter construct. This type of E domain is called peptidyl E domain, which is a gatekeeper for intermediate downstream processing in nonribosomal biosynthetic machineries (**Figure 23(b)**). This approach by Stein *et al.* demonstrates how mechanistic insights can be gained through formation of well-chosen NRPS constructs and mass spectrometric analysis.

The recent studies of Hansen *et al.*[51] and Wittmann *et al.*[109] revealed a new mechanism for lipidation of lipopeptide biosynthesis such as mycosubtilin or daptomycin biosynthesis by application of ESI–FTMS. Both papers describe that fatty acid incorporation is catalyzed by an A domain with fatty acid specificity in the loading module of the mycosubtilin NRPS (**Figure 15**) or by a preassembly line A and T domain in daptomycin biosynthesis (**Figure 16**).

In addition, Tang *et al.*[108] characterized a new mechanism for D-amino acid incorporation into NRP chains by NRPS, which was previously thought to happen only by epimerization of T domain-tethered L-amino acid



Alternative CmaE aminoacyl substrates	*Observed mass shift (MALDI-TOF MS)*	*Observed mass shift (MALDI-TOF MS)*
L-Leucine	+113 Da	+111 Da
L-Phenylalanine	+151 Da	+153 Da

Figure 21 Characterization of aminoacyl transferase CmaE in coronamic acid biosynthesis pathway. (a) Within the biosynthetic pathway, CmaE carries out substrate shuttling from the CmaA T domain to the CmaD T domain. (b) CmaE substrate tolerance was characterized by MALDI–TOF MS (observed and calculated mass shift of CmaD in the table). In addition, evidence of reversible aminoacyl transfer by CmaE was detected.

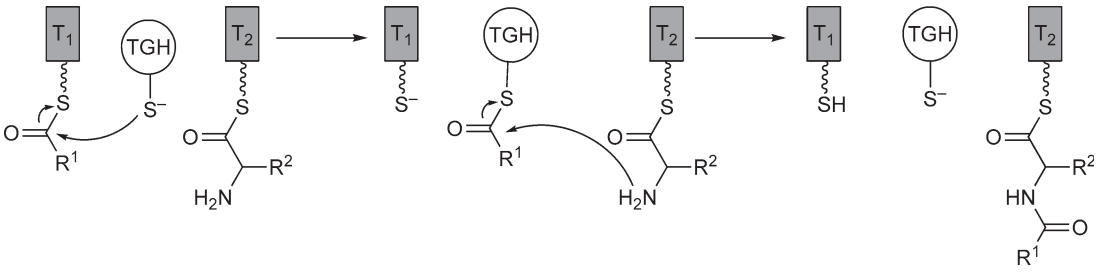


Figure 22 Condensation mechanism of transglutaminase homologue domain by covalent atalysis.[105]

substrates via $^{L}C_{D}$ domains or E domains. The initiation module of leinamycin biosynthesis comprises an A domain that has D-alanine specificity (**Figure 17**) and, therefore, it constitutes a new direct mechanism of D-amino acid incorporation.

Wenzel *et al.*[111] applied MS to investigate a deviation from NRPS colinearity in the myxochromide S biosynthetic pathway – the first reported NRPS module skipping process. Myxochromides are lipopeptides isolated from several myxobacteria. Myxochromides A are structurally similar lipohexapeptides produced by *Myxococcus xanthus* and myxochromides S are structurally similar lipopentapeptides produced by *Stigmatella aurantiaca*. Both the myxochromide types are biosynthesized by hybrid PKS/NRPS that are identical in module and domain arrangement. Because myxochromides S contain only a five amino acid chain despite a six module NRPS (**Figure 24(a)**), Wenzel *et al.* investigated the absence of an L-proline in myxochromides S that is integrated by module 4 of the correlating $MchC_A$ NRPS into myxochromides A. Adenylation activity was detected for both A4 domains from $MchC_A$ NRPS and $MchC_S$ NRPS by ATP–PP_i exchange assay with a slightly lower activity of A4 from $MchC_s$. The T domain activity was investigated by coexpression of each T4 with PPTase MtaA in *E. coli* and ESI–FTMS and MALDI–TOF MS analysis of the GST-tagged T domain constructs. Only phosphopantetheinylation of T4 from $MchC_A$ was detected by the corresponding +340 Da shift of the apo T domain peak. Although the GST-T domain constructs were ~35 kDa in size, no active site mapping was necessary, which shows that very pure NRPS constructs larger than 20 kDa can be characterized by ESI–FTMS or MALDI–TOF MS. The lacking $MchC_S$ T4 domain activity is due to a point mutation in the carrier domain active site that results in the complete deactivation of the module 4 in myxochromides S biosynthetic assembly line. Therefore, Wenzel *et al.* proposed a module skipping mechanism based on the characterized 'loss-of-function' point mutation in the T4 domain, which is the first skipping process described in a multimodular NRPS (**Figure 24(b)**).

9.11.2.5 Characterization of Tailoring Reactions

Recently, two new tailoring reactions were analyzed using MS.[55,107] Other NRP tailoring mechanisms that have been characterized by utilization of MS and protein crystallography have been reviewed elsewhere.[47,112]

Gatto *et al.*[107] characterized the mechanism of L-pipecolic acid formation by cyclodeaminase RapL from L-lysine within rapamycin biosynthesis, which is a hybrid NRP–polyketide antibiotic (**Figure 25(a)**). RapL was characterized by biochemical assays to require cofactor nicotinamide adenine dinucleotide (NAD^+) and an oxidative cyclodeamination reaction mechanism corresponding to ornithine cyclodeamination was proposed based on ESI–FTMS analysis of RapL reaction products (**Figure 25(b)**).

Another cyclization mechanism was investigated by Kelly *et al.*[55] Cyclopropane ring formation by CmaC catalysis from a γ-chloro-L-*allo*-Ile intermediate within coronamic acid assembly line was characterized. Coronamic acid is a fragment of coronatine, a hybrid NRP–polyketide phytotoxin (**Figure 21(a)**). In this study, mechanistic insights were gained by ESI–Q–FT–ICR–MS detection of T domain CmaD tethered intermediates of CmaC reaction. Isotopically labeled substrates at specific positions combined with the high mass accuracy of ESI–FT–ICR–MS allowed the dissection of CmaC-catalyzed propane ring formation by detection of γ-chloride loss (**Figure 26(a)**) and α-hydrogen exchange (**Figure 26(b)**). Therefore, a new mechanism of cyclopropane formation was proposed in which CmaC-Zn^{2+}-mediated carbanion formation is

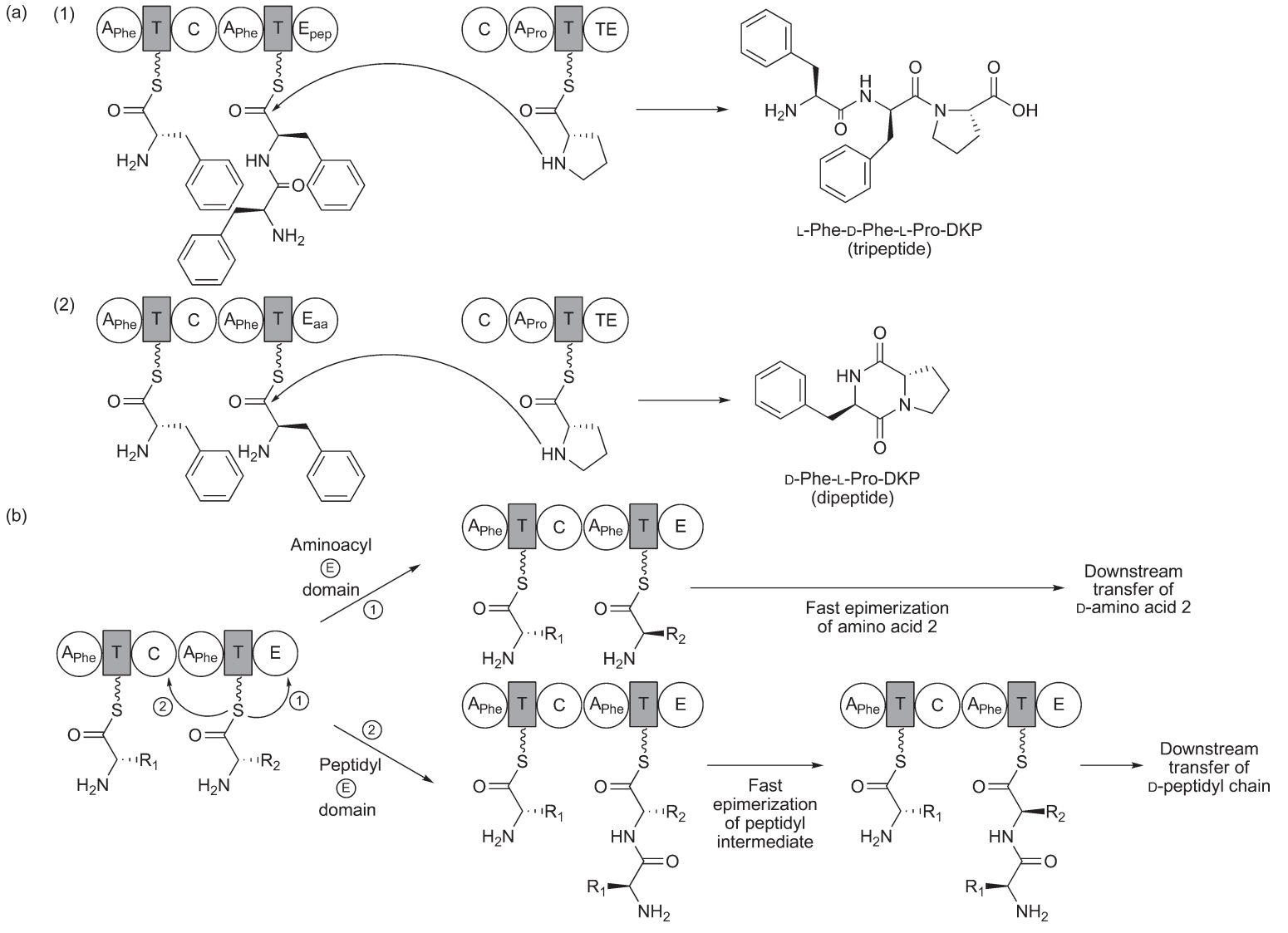


Figure 23 Differentiation of initiation and elongation epimerization domains by intermodular transfer activity. (a) Intermodular transfer activity assay. Peptidyl or aminoacyl specificity of epimerization domains is monitored by tripeptide formation or dipeptide formation, respectively. (b) Proposed mechanism of aminoacyl epimerization domain and peptidyl epimerization domain in NRPS.[42]

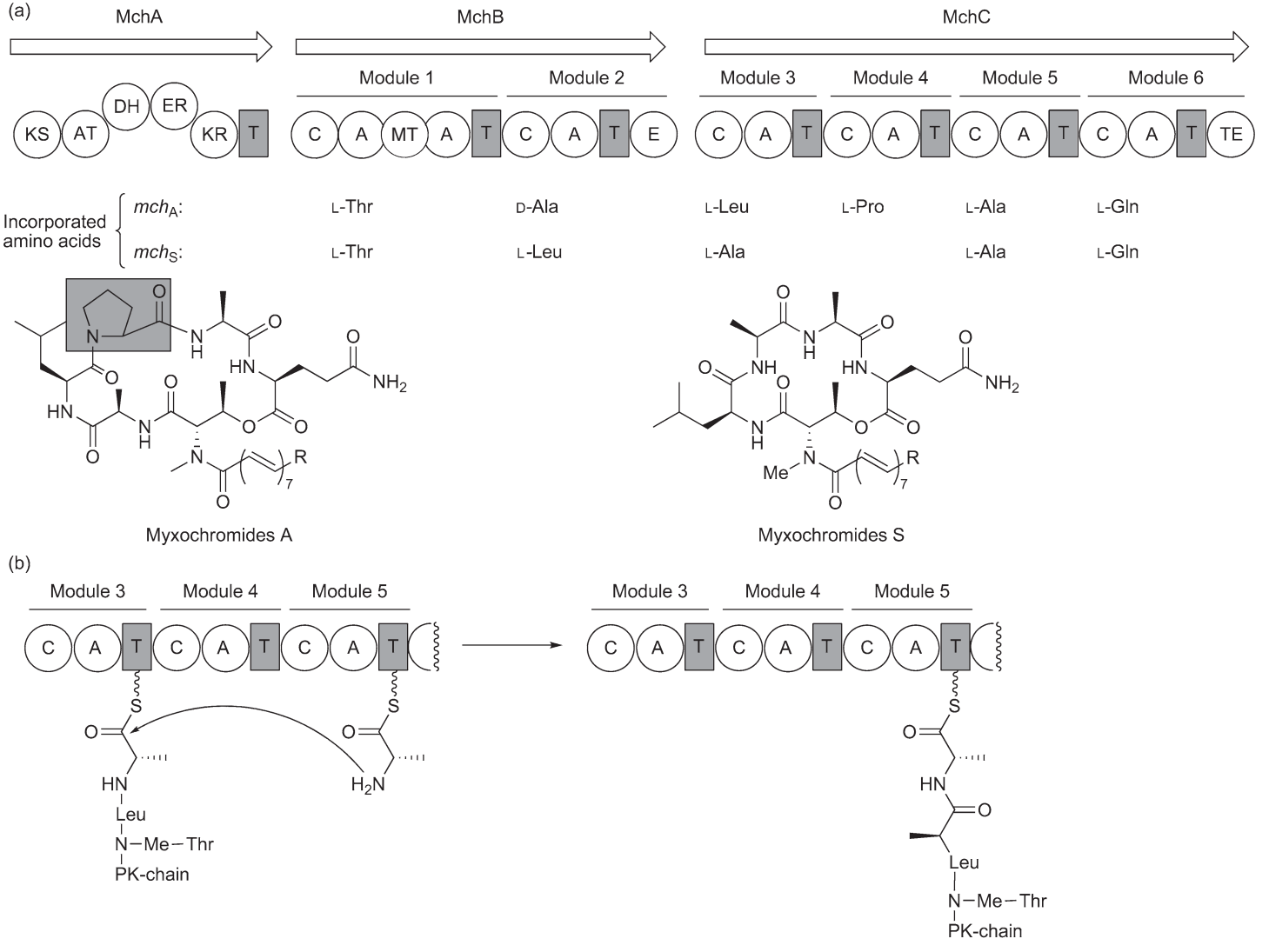


Figure 24 NRPS module skipping mechanism revealed by myxochromides A and S biosynthesis. (a) Myxochromides A and S structures and module and domain arrangement in their biosynthetic machineries. (b) Proposed NRPS-module skipping mechanism by 'loss-of-function' mutation in T domain of the skipped module.

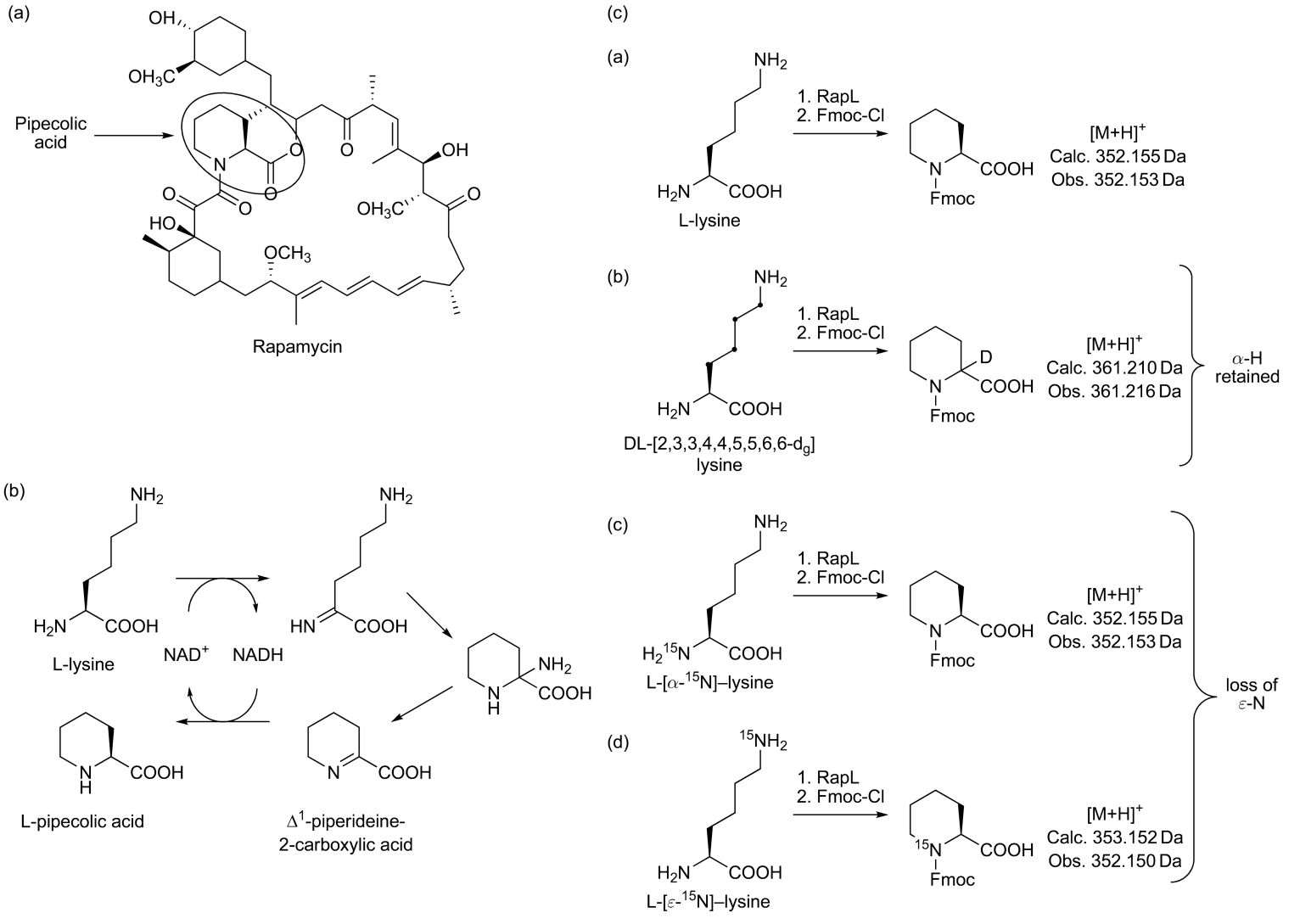


Figure 25 L-Pipecolic acid formation by cyclodeaminase RapL in rapamycin biosynthesis. (a) Rapamycin and incorporated pipecolic acid moiety. (b) Proposed oxidative cyclodeamination mechanism of pipecolic acid formation from L-lysine. (c) RapL activity assays and exact ESI–FTMS analysis of derivatized reaction products revealing mechanistic insights such as α-H retainment and loss of ε-N.

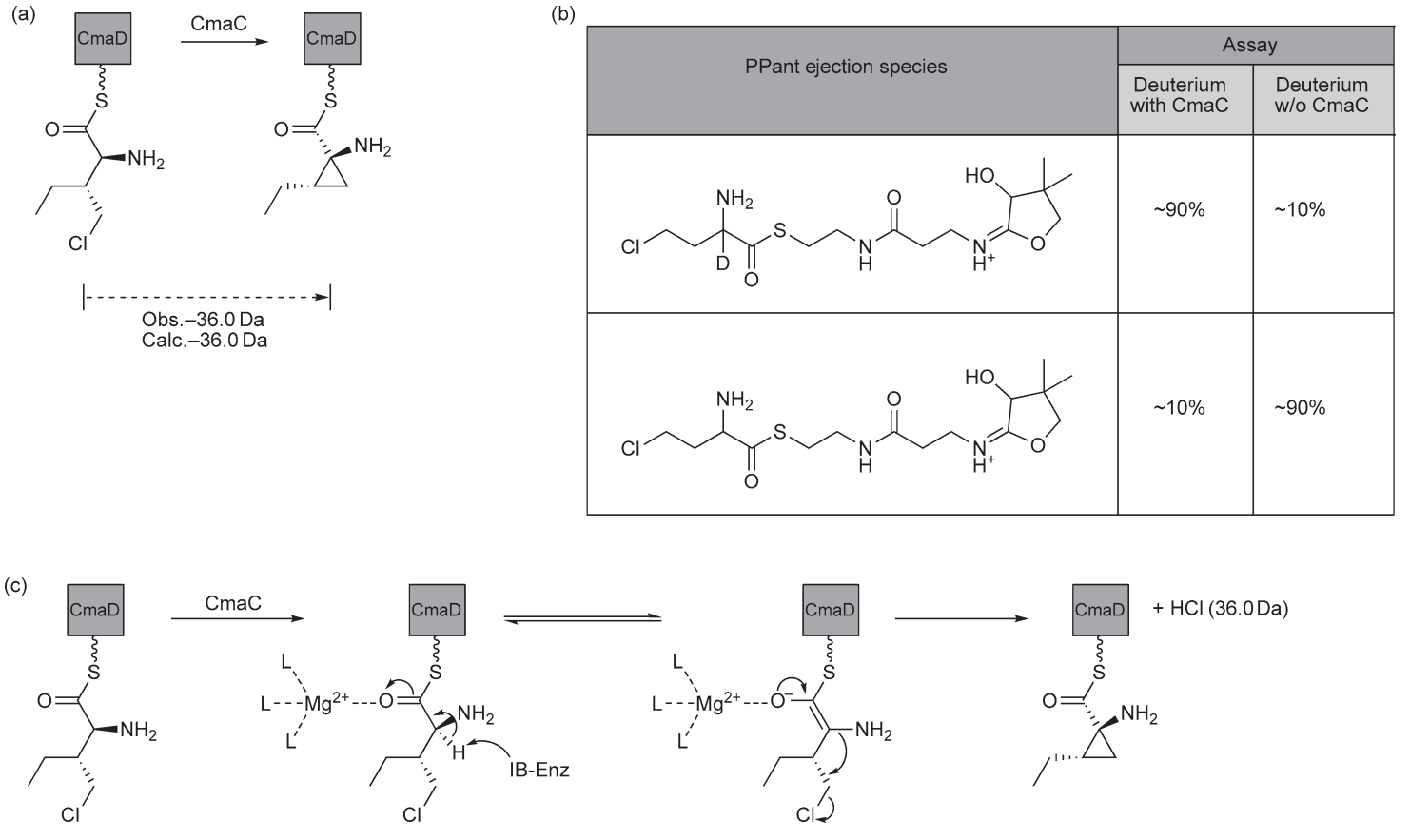


Figure 26 Cyclopropane ring formation by CmaC in coronamic acid biosynthesis. (a) Characterization of γ-Cl loss by ESI–FTMS. (b) Characterization of α-H exchange by deuterium solvent exchange and PEA. CmaC promotes deuterium incorporation at α-C position. (c) Proposed cyclopropane ring formation mechanism.[55]

followed by nucleophilic substitution of a γ-Cl-leaving group (**Figure 26(c)**). Alternative mechanisms such as azetidine formation were experimentally denied. Before CmaC-catalyzed cyclization, γ-chloro-L-*allo*-Ile is transferred from CmaA T domain to CmaD T domain by predicted aminoacyl transferase CmaE. In a subsequent study, Strieter *et al.*[106] investigated this shuttling step within the coronamic acid biosynthesis by application of MALDI–TOF MS. As described above, CmaE shuttling promiscuity was characterized by MS. In addition, the reversibility of CmaE transfer within the coronamic acid biosynthesis was also detected by MALDI–TOF MS. L-Valine, tethered on upstream CmaD T domain, was detected after several minutes incubation with CmaE on downstream holo CmaA T domain. This suggested reversible aminoacyl transfer by CmaE.

A novel epoxidation tailoring reaction is described by the work of Kopp *et al.*[52] about the formation of the 2,3-epoxyhexanoyl moiety in CDA. Online LC–ESI–FT–ICR–MS and the PEA characterized that both the putative oxygenases HxcO and HcmO of the CDA gene cluster are involved in the fatty acid epoxidation. It is shown that HxcO catalyzes the C2–C3 dehydrogenation of hexanoyl-*S*-T and subsequent hexenoyl-*S*-T epoxidation to yield 2,3-epoxyhexanoyl-*S*-T as the main product. Hexenoyl-*S*-T was detected as an HxcO reaction side product. HcmO utilizes only hexenoyl-*S*-T to give the 2,3-epoxyhexanoyl-*S*-T. Therefore, Kopp *et al.* concluded that HxcO is the main fatty acid epoxidation catalyst within CDA biosynthesis and HcmO epoxidizes the hexenoyl side product of the HxcO reaction to the epoxyhexanoyl moiety (**Figure 20(b)**). Interestingly, it was shown by a novel amide ligation assay with an amine enantiomer and subsequent chiral HPLC analysis that the HxcO and HcmO epoxides have opposite stereochemistry. This is an example for a limitation of MS for characterization of biosynthetic pathways.

9.11.2.6 Characterization of Multistage Assembly Line Action

The characterization of multistage assembly line action by high-resolution MS was developed in particular on yersiniabactin assembly line components. Five of its active sites were detected in parallel to monitor intermediate downstream processing on a biosynthetic thiotemplate.[47,113]

In a recent study by Garneau-Tsodikova *et al.*,[96] the early stages of prodigiosin biosynthesis were characterized by ESI–FTMS detection of intermediates on two active sites. Although the biosynthetic steps could only be characterized in a single turnover fashion, it is an example for dissection of multistage assembly line action by MS. Prodigiosin of the prodiginine class of natural products comprises three pyrrole rings and is derived by *Serratia marcescens*. Each of its three pyrrole rings is proposed to be formed by a different mechanism. In the first stages of prodigiosine biosynthesis, dipyrrole formation is proposed as depicted in **Figure 27(a)**. Adenylation domain PigI adenylates and loads L-proline onto freestanding PigG carrier protein. L-Prolyl-*S*-PigG is double dehydrogenated by flavoprotein desaturase PigA yielding pyrrolyl-*S*-PigG intermediate. The pyrrolyl group is transferred to the PigJ active site, a ketosynthase with a chain length factor (CLF) partner domain. The ketosynthase catalyzes decarboxylation of a malonyl moiety to a carbanion tethered on the T_1 active site (we changed the original notation from this paper as ACP_1 to T_1 to reflect the notation used in this review) of the tridomain PigH and subsequent formation of pyrrolyl-β-ketoacyl-*S*-PigH. PigH comprises a putative PLP-containing seryltransferase (SerT), which catalyzes the formation of the second pyrrole ring by serine insertion. Garneau-Tsodikova *et al.* could confirm most of these biosynthetic steps by *in vitro* reconstitution assays and ESI–FTMS detection of corresponding intermediates on PigG T domain and PigH T_1. On FTMS-mapped PigG active site, tethered L-proline and L-pyrrolyl intermediates could be detected after PigI adenylation or PigA double dehydrogenation, respectively. In addition, simultaneous substrate loading of both PigH active sites (T_1 and T_2) could be confirmed (**Figure 27(b)**). On FTMS-mapped PigH T_1 active site, malonyl and pyrrolyl-β-ketoacyl intermediates were detected upon PigL malonyl loading or PigJ Claisen condensation, respectively. Serine insertion by SerT and bipyrrole formation have not been detected by ESI–FTMS yet. The study shows the dissection of predicted NRPS–PKS assembly line action only by ESI–FTMS and manifests its potential as a stand-alone investigative tool of biosynthetic thiotemplates. Once again, an NRPS–PKS interface is investigated.

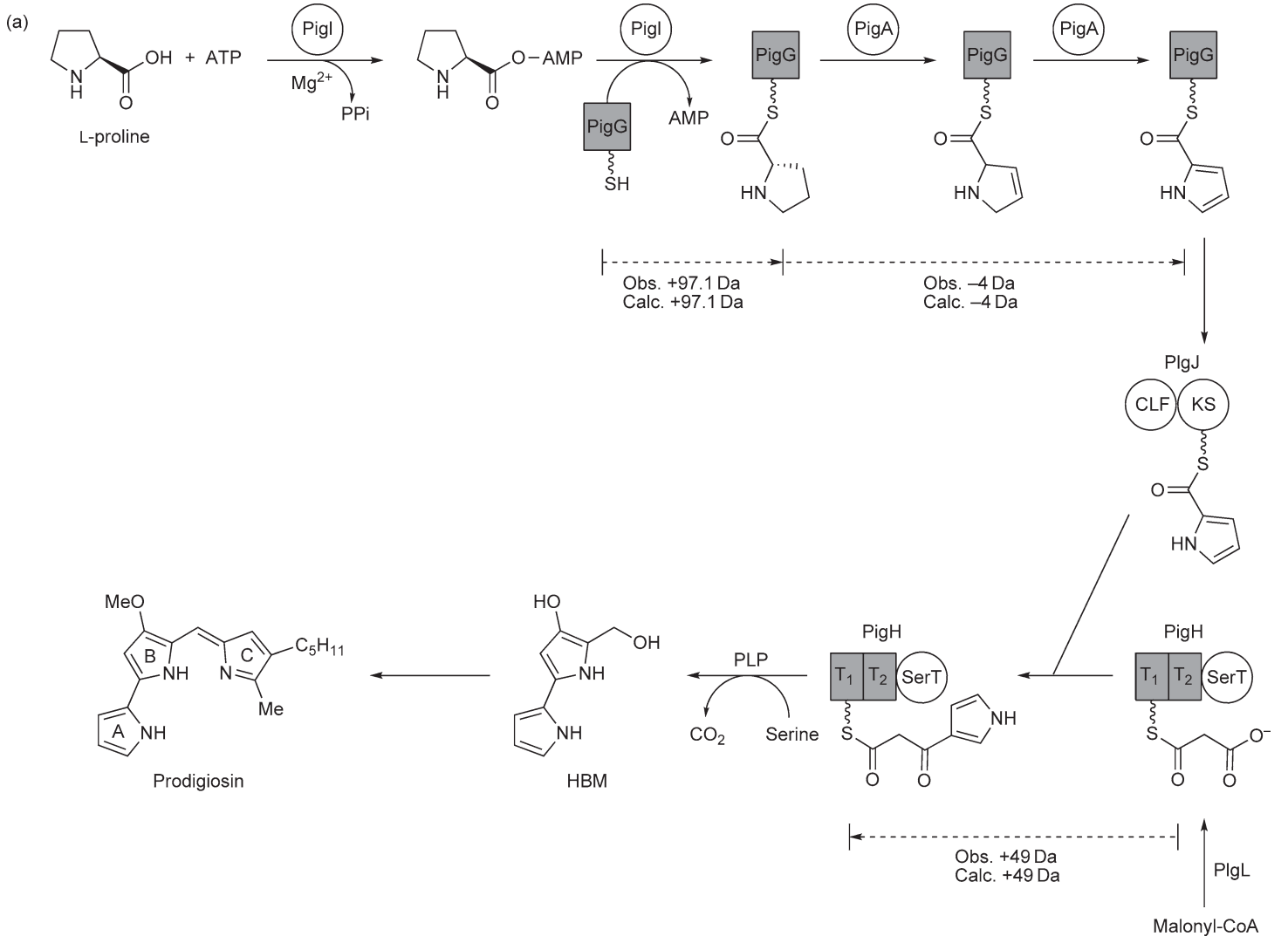


Figure 27 (Continued)

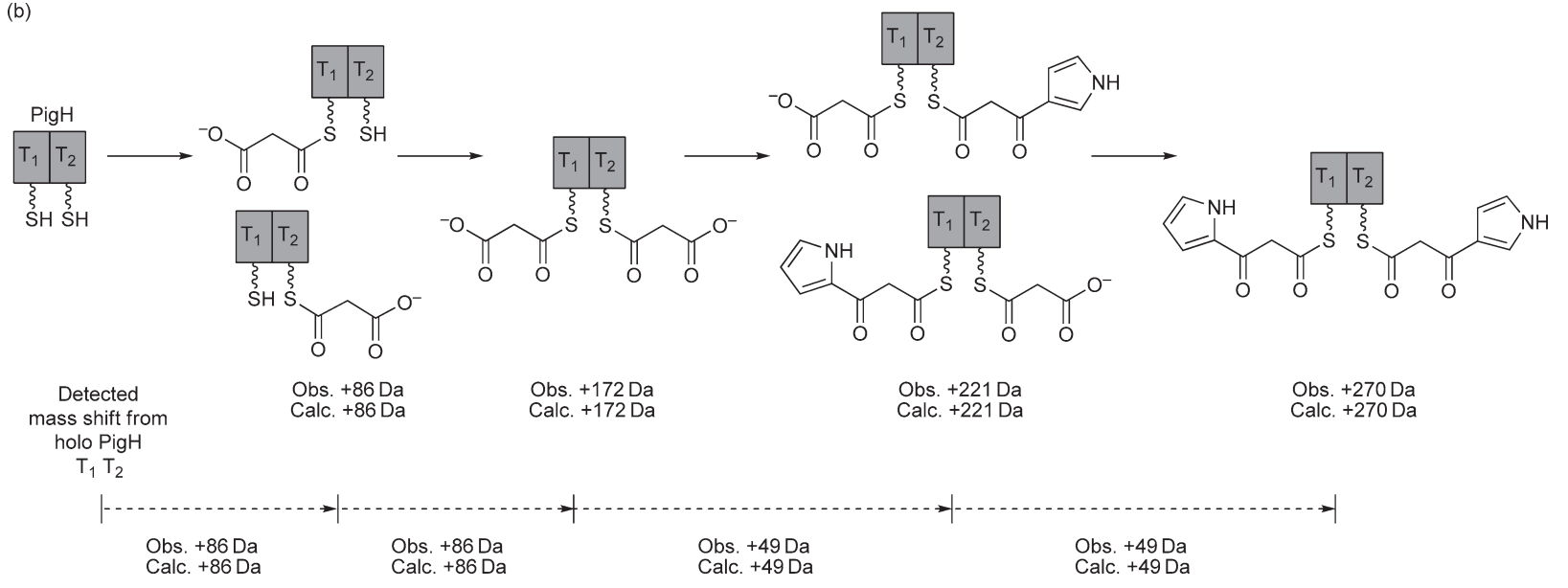
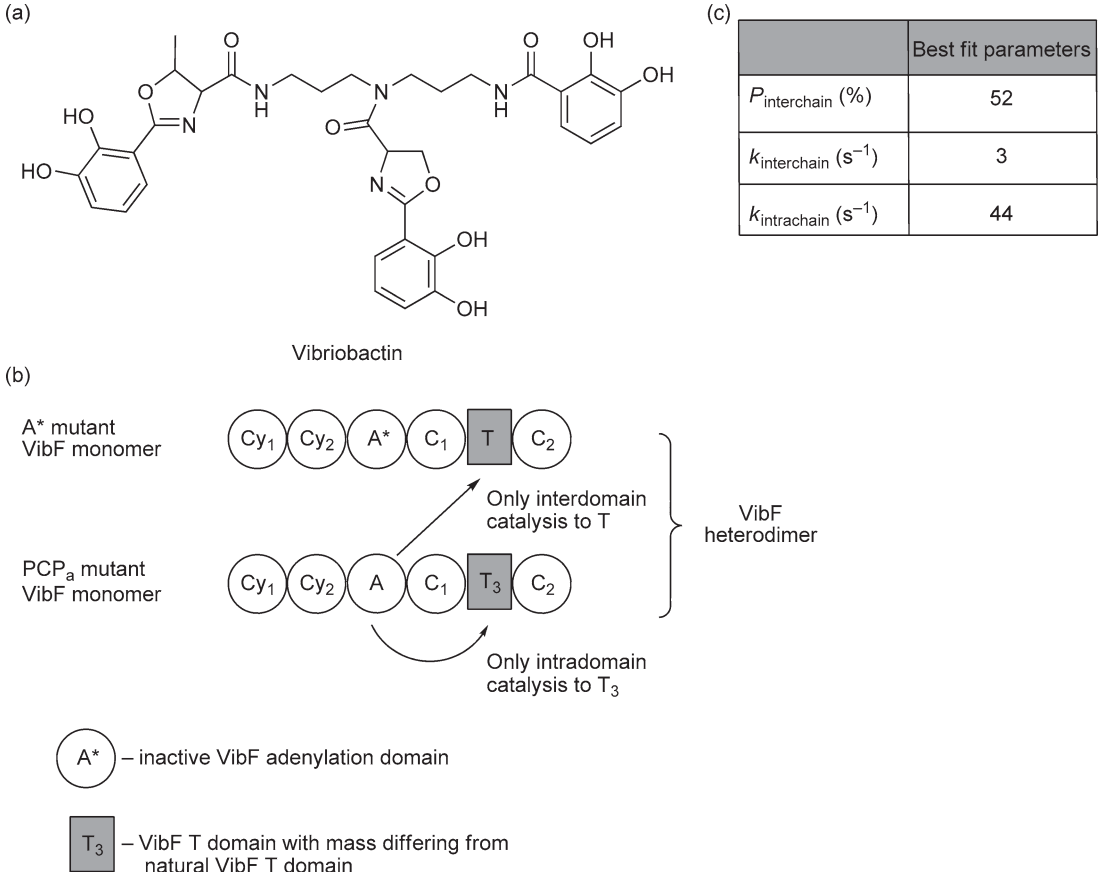


Figure 27 Monitoring multiple active sites in prodigiosin biosynthesis. (a) Early stages of prodigiosin biosynthesis dissected by *in vitro* reconstitution and ESI–FT–ICR–MS (observed and calculated mass shifts of PigG T domain and PigH T_1 domain active sites). (b) Characterization of substrate loading on both PigH active sites T_1 and T_2 by ESI–FT–ICR–MS (observed and calculated mass shifts of PigH T_1T_2).

9.11.2.7 Time Courses

The investigation of time courses of NRPS assembly line processing by MS was realized by semiquantitative approaches until 2006, as highlighted by Dorrestein and Kelleher.[47] The problem of obtaining true-kinetic time courses for intermediate flux on thiotemplates is the limited complexity of peptide mixtures,which can be analyzed by ESI–FTMS and different ionization behavior of the same active site species loaded with different intermediates causing deviations from true intermediate quantities. So far kinetic time course experiments on NRPS systems have been considered as pseudokinetic. An example is a study by Hicks *et al.*[98] in which interchain and intrachain acylation in dimeric VibF of vibrobactin NRP biosynthesis was detected and semiquantitative rates for both processes in presteady state were calculated. By detection of time-dependent holo T domain decrease during L-threonine loading by VibF A domain, the relative occupancy of T domains with L-threonyl intermediate was indirectly measured because of unsuccessful L-threonyl-*S*-T domain detection by ESI–FTMS. Interchain and intrachain acylation in the homodimeric VibF was investigated by formation a of a heterodimeric construct with an inactive A domain on one chain and two T domains with differing masses (**Figure 28(b)**). Acylation rates were determined by numeric modeling of time-dependent occupancy curves of the interchain or intrachain T domains. The intrachain acylation rate was significantly faster than the interchain rate and an equal flux of intermediates was detected for both alternative pathways (**Figure 28(c)**). This was the first approach to gain pseudokinetic data of competitive pathways in a dimeric NRPS system.



(c)

	Best fit parameters
$P_{interchain}$ (%)	52
$k_{interchain}$ (s^{-1})	3
$k_{intrachain}$ (s^{-1})	44

Figure 28 Investigation of interchain and intrachain acylation in dimeric VibF. (a) Vibrobactin. (b) VibF heterodimer for interchain and intrachain kinetic assay and ESI–FTMS analysis. (c) Obtained best-fit kinetic parameters for VibF intrachain and interchain acylation, $P_{interchain}$ – probability of interchain flux, $k_{interchain}$ – interchain acylation rate, $k_{intrachain}$ – intrachain acylation rate.

This section highlighted recent applications of modern MS to dissect biosynthesis on NRPS systems and NRPS–PKS interfaces by *in vitro* reconstitution. MS is increasingly applied for *in vitro* characterization of gatekeeping and promiscuity of NRPS components and a first *in vivo* NRP intermediate was detected by MS. In addition, MS is a valuable method to dissect single reaction mechanisms, for example, of tailoring reactions or to dissect multistage NRPS assembly line action. The investigation of NRPS time courses has been accomplished to date only in a pseudokinetic fashion. As mentioned above, several laboratories gained excess to FTMS instrumentation since its first application to an NRPS system in 1999 by Kelleher and coworkers.[41] This trend is shown here by the presented NRPS studies. Future diversification and a wider application of FTMS methods in an NRPS context can be expected. Additionally, several of the presented studies applied low-resolution MS, in particular MALDI–TOF MS. These studies do not include NRPS active site mapping, which is only possible with high-resolution MS.

9.11.2.8 Orphan Gene Cluster Characterization by Mass Spectrometry

As outlined recently,[47] the ability of ESI–FTMS to characterize NRPS substrates, intermediates, and tailoring reactions could be applied for elucidation of natural product chemistry from orphan NRPS gene clusters. Orphan gene clusters are gene clusters with unknown natural products. Various strategies have been developed to identify their secondary metabolites. The majority of these so-called genome mining approaches, which are reviewed elsewhere,[114] are aimed to isolate the unknown natural product. In contrast to these discovery strategies, the *in vitro* reconstitution approach is aimed to dissect the structure of the unknown natural product by characterization of the recombinant orphan biosynthetic enzymes. Herein, MS would complement recent *in vitro* reconstitution tools and could provide the main information in the structure elucidation process by reliable identification of substrates and chemical modifications. MS-based *in vitro* reconstitution of an orphan NRPS gene cluster could be pursued as follows and depicted as in **Figure 29**.

First, the functions of all orphan gene products are predicted by BLAST or other protein analysis tools. NRPS and tailoring enzymes are differentiated. In addition, the substrates of the A domains can be predicted from their '10 letter codes'[99,100] by bioinformatic tools such as NRPS Predictor[101] although this is not a prerequisite for the *in vitro* reconstitution approach but a routine in NRPS characterization.

Second, all A domains are expressed heterologously and screened for their native substrates by MS-based substrate screening.[91] The substrate screening assay depends on the A domain construct: If an AT didomain or larger construct is screened, active site mapping or PEA-based 'top down' analyses are required to identify the substrates. If the A domain loads onto a freestanding T domain, substrates can be characterized directly by the T domain mass shift without prior active site mapping. The T domain substrate should be the pathway-specific T domain corresponding to the A domain, biosynthetic substrates can be screened in a defined way, for example, a specific amino acid mixture, or in a undefined way, for example, algal lysate. Substrate identification from an A domain of an orphan gene cluster has been achieved on two A–T didomains, PksN and PksJ, from the *pksX* pathway from *Bacillus subtilis* by Dorrestein *et al.*[91] Problems in this step can be a difficult and laborious protein expression and A domain inactivity upon heterologous expression.

Third, based on the identified A domain substrates, NRP intermediates and a complete NRP scaffold are synthesized by solid-phase peptide synthesis.[115] These small peptides will serve as substrates to elucidate tailoring enzymes.

Fourth, the tailoring enzymes are characterized in their substrates, their chemical modifications on the NRP-scaffold, and their reaction mechanisms. The appropriate substrate of each tailoring enzyme has to be identified. The substrate can be a T domain-tethered substrate or intermediate, CoA-activated acid, or an analogue of the complete natural product. Subsequently, the chemical modification is dissected by *in vitro* tailoring reaction of the corresponding substrate or intermediate and FTMS detection of its mass shift. Additionally, chemical modification can be localized within the NRP by FT–MS^2-based structure elucidation. Herein, the tailoring reaction substrate and product are fragmented in the mass spectrometer separately and the NRP fragment with the chemical modification is identified by the corresponding mass shift.

Based on this MS dissection of the native substrates and chemical modifications, a related or even identical structure of the unknown natural product can be synthesized and tested in terms of bioactivity and physiological target. The result from the *in vitro* reconstitution should be confirmed by another genome mining

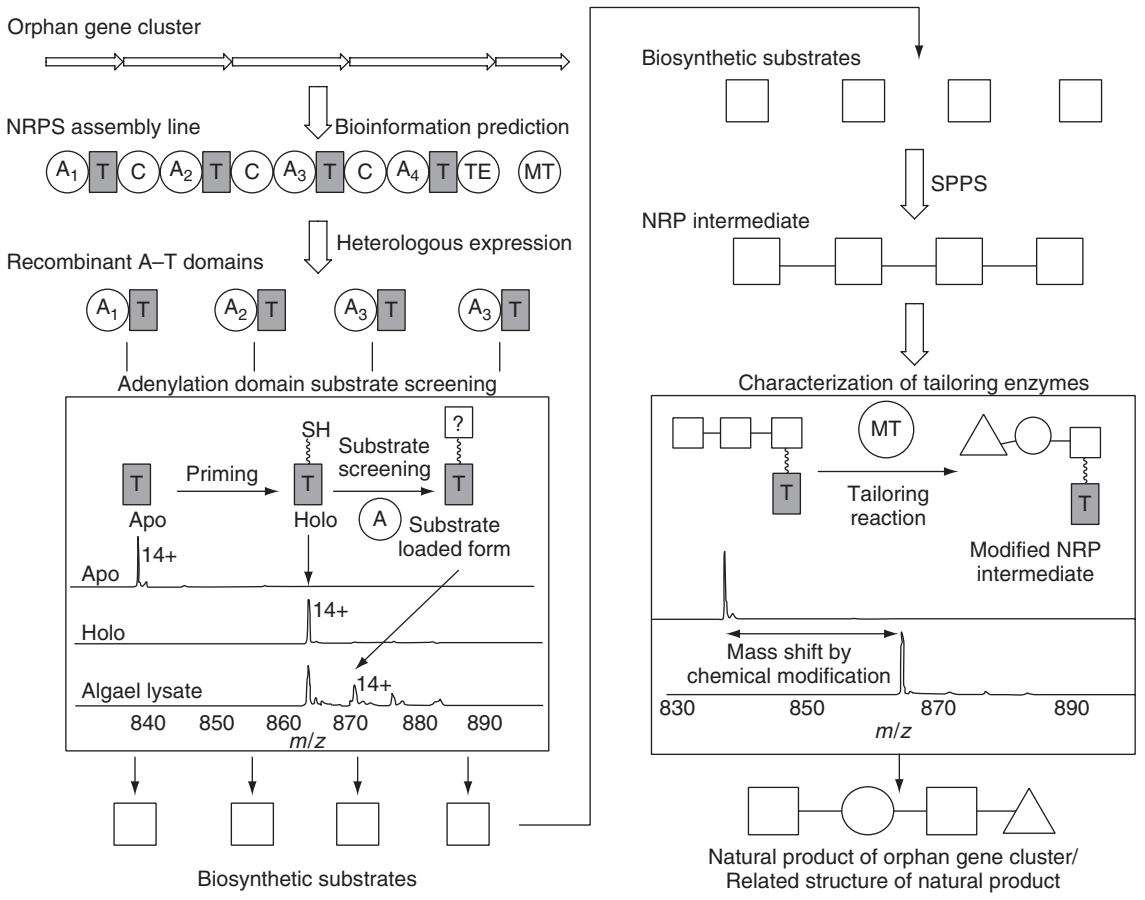


Figure 29 MS-based *in vitro* reconstitution of orphan NRPS gene clusters. Substrates and chemical modifications of unknown NRP natural products can be dissected by FT–ICR–MS methods such as substrate screening and PEA. Based on these biosynthetic informations a structure related to the actual natural product can be drawn.

approach aimed to discover the natural product, for example, genomeisotopic approach based on the identified substrates. The disadvantage of the MS-based *in vitro* reconstitution of orphan gene clusters is on the one hand the laborious task of heterologous protein expression and on the other hand the requirement of high-resolution MS, which is recommended because of its high mass accuracy and MS^n features. The advantage of the approach is that any NRPS gene cluster – silent or expressed – can be identified in its natural product chemistry. Advances in bioinformatic prediction tools, gene synthesis, and heterologous protein expression should lay a more rigid foundation for the role of MS as a major genome mining tool.

9.11.3 Applications of Mass Spectrometry on PKS Systems

Polyketides represent a source of numerous pharmacologically and commercially useful compounds.[116] The appeal of polyketides is that they are a structurally diverse class of compounds, yet these complex molecules can be synthesized from much smaller, simpler acyl-CoAs. Polyketides are synthesized by large multidomain megasynthase PKSs.[23,117] These megaenzymes efficiently carry out the addition of the acyl-CoAs to form elongated intermediates that undergo a variety of different enzymatic tailoring steps. PKSs are classified as type I, type II, or type III. Type I PKSs are a single protein consisting of a linear arrangement of the various catalytic and carrier domains. Type II PKSs consist of the various domains that exist as individual proteins that interact with each other. Type III PKSs function without the use of a T domain. In addition, polyketides can be

classified as either being modular or iterative. Modular PKSs have multiple domains that function in an assembly-line fashion in which the substrate is bound to a carrier protein of the first domain, undergoes modifications, and then is transferred to a carrier protein within the next domain. The growing intermediate is passed from one catalytic domain to the next and undergoes elongation and additional modifications at each domain until a full-length intermediate is released. Iterative PKSs possess only a single module consisting of a few catalytic and carrier domains that are reused over and over during the elongation of the intermediate. The intermediate undergoes cycles of addition and modification. Type I PKSs and type II PKSs have been the subject of recent investigations and the fact that they possess the phosphopantetheinyl functionality on their T domains make them suitable targets for studies involving MS. Using high-resolution FTMS, the intermediates of PKS biosynthetic pathways can be detected while still covalently bound to active site of the T domain via PPant. Confirmation of the exact mass of PPant-bound intermediates can be confirmed by subsequent PEA.

9.11.3.1 Bacillaene Biosynthesis: *Bacillus subtilis* HMG-CoA/Trans-Enoylreductase and α-/β-Ketoreductase

9.11.3.1.1 Orphan gene cluster pksX

The *pksX* gene cluster of the *Bacillus subtilus* encodes a hybrid PKS–NRPS that produces a previously unknown secondary metabolite. This orphan gene cluster was found to share many similarities with gene clusters involved in the biosynthesis of curacin, jamacamide, pederin, as well as others. Of particular interest were several biosynthetic tailoring enzymes expressed such as a trans-acting AT (PksC). Also of interest were several freestanding proteins, such as a T domain (AcpK), a ketosynthase (PksF), a 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase (PksG), two enoyl-CoA hydratases (PksH and PksI), and a larger protein containing multiple T domains (PksL, **Figure 30(a)**). In addition to various biochemical techniques, high-resolution MS was applied to designate functional roles to these various proteins encoded by the *pksX* gene cluster. The genes for PksC, AcpK, PksF, PksG, PksH, PksI, and the region of the gene encoding the pair of tandem T domains PksL were heterologously expressed in *E. coli* and then purified. The AcpK protein and the tandem T domains (PksL-T_2) were phosphopantetheinylated *in vivo* by coexpressed Sfp.[57]

Substrate loading onto AcpK and subsequent alterations to the PPant-tethered intermediate that were hypothesized to be carried out by the array of proteins encoded by *pksX* were determined by FTMS. PksC, when incubated with holo AcpK and malonyl-CoA, resulted in a +86 Da shift of the AcpK protein, corresponding to an 8.6 m/z shift of the 10+ of AcpK ion. This mass difference was consistent with the formation of malonyl-*S*-AcpK. This observation, coupled with detection of a +86 Da mass shift of PksC itself when incubated with malonyl-CoA, confirmed the malonyl-AT function of PksC. The function of PksF was determined by incubating the protein with malonyl-*S*-AcpK. The resulting 44 Da loss in mass, representing decarboxylation of malonyl-*S*-AcpK to acetoacetyl-*S*-AcpK, was detected with FTMS and helped verify the function of PksF as that of a ketosynthase. To probe the possible function of PksG as that of an HMG-CoA synthase (HCS), acetoacetyl-*S*-PksL-T2 was incubated with acetyl-*S*-AcpK and PksG. The addition of PksG facilitated the formation HMG-*S*-PksL-T_2, at one or both tandem T domain active sites, observed as a +60 Da mass shift by FTMS. IRMPD resulted in a PPant ejection ion with a mass of 503.152 Da, 60.022 Da larger than the mass of acetoacetyl (Acac)-loaded PPant (**Figure 30(b)**). The functional characterization of PksH and PksI activity was again facilitated by high-resolution MS and confirmed by PPant ejection analysis. HMG-*S*-PksL-T_2 was incubated first with PksH alone, then PksI alone. On its own, PksH did not yield any change in the PPant-tethered intermediate. Incubation of HMG-*S*-PksL-T_2 with PksI, on the other hand, resulted in a loss of 18 Da representing dehydration of the intermediate. Incubation of HMG-*S*-PksL-T_2 with both PksH and PksI resulted in a mass shift of 62 Da. This corresponds with the dehydration (−18 Da) and subsequent decarboxylation (−44 Da) of the HMG intermediate. These findings were further validated by observation of mass shifts in the PPant ejection ions generated by IRMPD of the various PksL-T_2-bound intermediates (**Figure 30(c)**). These studies highlighted the utility of FTMS and the PEA as means to elucidate the functions of the various products of an orphan PKS gene cluster such as *pksX* by *in vitro* reconstitution.

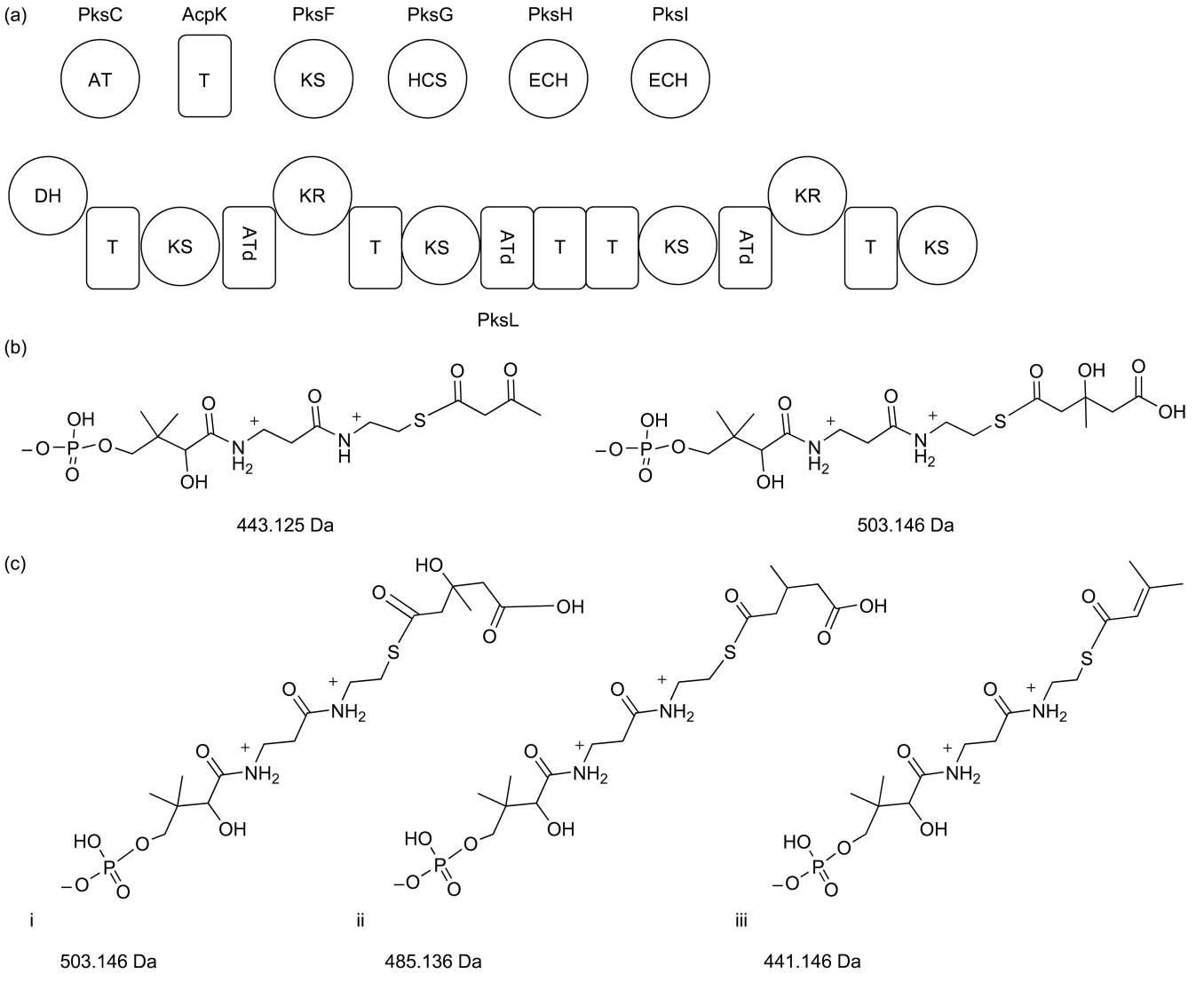


Figure 30 Bacillaene biosynthesis. (a) Annotation of the function of multiple proteins encoded by the *pksX* gene. (b) Determination of PksG (HCS) function. Acetoacetyl (Acac)-*S*-AcpK phosphopantetheine ejection ion was observed to have a mass of 443.125 Da. Incubation of Acac-*S*-AcpK with PksG results in an increase of 60.02 Da, consistent with formation of HMG-*S*-AcpK and confirmation of HCS activity by PksG. (c) Functional determination of PksH and PksI. HMG-*S*-PksL-T_2 was incubated with PksI, and PksI and PksH. Phosphopantetheinyl ejection ions were generated using IRMPD. (1) The ejection ion from HMG-*S*-PksL-T_2 has a mass of 503.146 Da. (2) Incubation with PksI results in a loss of 18 Da consistent with dehydration. (3) Incubation with PksI and PksH results in a loss of 62 Da, corresponding to dehydration followed by decarboxylation. Incubation of HMG-*S*-PksL-T_2 did not result in any mass changes.

9.11.3.1.2 Trans-enoylreductase and α- and β-ketone reduction

More recently, the ultimate biosynthetic product of the *pksX* gene cluster was identified and its structure was elucidated.[118] This product was identified as dihydrobacillaene, which is later converted to bacillaene (**Figure 31(a)**) by PksS.[119] Dihydrobacillaene is produced by numerous enzymes and tailoring domains discussed previously, as well as an additional four megasynthase complexes. The first of these megasynthases, PksJ, contains two NRPS modules followed by two PKS modules. Recent investigation into the dihydrobacillaene biosynthetic pathway involved the determination of the origin of the α-hydroxyacyl N-cap.[58] While the first two modules of PksJ seemed capable of accepting α-hydroxyisocaproate (α-HIC) directly, before being transferred to the third PksJ module, researchers found that PksJ preferentially loads α-ketoisocaproate (α-KIC) first, then transfers the intermediate to the third module. Module 3 of PksJ possesses a pair of tandem T domains, which receive the α-KIC-containing intermediate. The presence of a KR domain and the tandem T domains within the third PksJ module led to the hypothesis that the α-KIC-containing intermediate is reduced to α-HIC within the third module by the single KR domain acting on both the α-keto and β-keto groups of the α-KIC intermediate (**Figure 31(b)**).

To test the functions of the KR domain, investigators heterologously overexpressed the region of PksJ containing the KR domain and the tandem T domains, PksJ(KR-T_3-T_4). To assess the typical function of the PksJ-KR, as a β-ketoreductase, apo PksJ(KR-T_3-T_4) was incubated with Acac-CoA and Sfp in order to generate Acac-*S*-PksJ(KR-T_3-T_4). Reduction was carried out using either nicotinamide adenine dinucleotide hydride (NADH) or nicotinamide adenine dinucleotide phosphate hydride (NADPH), followed by ArgC digestion. The T domain active site fragments incubated with NAD(P)H were compared to controls. A small shift in mass of the active site was detected in samples treated with NADH/NADPH, and the mass shift of +2.0 166 Da was verified by carrying out source-induced dissociation for PPant ejection (**Figure 31(c)**). This mass shift is consistent with β-ketoreduction of Acac-*S*-PksJ(KR-T_3-T_4) to β-hydroxybutyrl-*S*-PksJ(T_3T_4), thus confirming the function of the PksJ module 3-KR as a β-ketoreductase. While a β-ketoreduction represents the rule for KR function, α-ketoreduction represents the exception. In order to test the ability of PksJ module 3-KR to reduce the distant α-ketone of α-KIC, researchers utilized α-ketoisocaproyl-γ-aminobutyrate (α-KIC-GABA) as a model substrate. Reduction of α-KIC-GABA-*S*-PksJ(KR-T_3-T_4) occurred in an NAD(P)H-dependent manner and was detected by protease digestion followed by FTMS and PPant ejection analysis (**Figure 31(c)**).

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

9.11.3.2 Curacin A Biosynthesis: ECH1 and ECH2

The application of high-resolution MS has allowed characterization of enzymes involved in the biosynthesis of curacin A, a mixed polyketide–NRP produced by *Lyngbya majuscule*, which possesses cytotoxic properties.[120] Researchers were intrigued by the unusual structure of this molecule (**Figure 32(a)**) and during their investigations of the enzymes involved in curacin A biosynthesis, an HCS-like gene cassette was identified, similar to that of the *pksX* gene cluster in *B. subtilis*, as well as others. The HCS-like cassette, involved in curacin A biosynthesis, encodes five separate enzymes that include a T domain, a ketosynthase, an HCS, and two separate enoyl-CoA synthases.[121] This set of enzymes was hypothesized to be responsible for the formation of

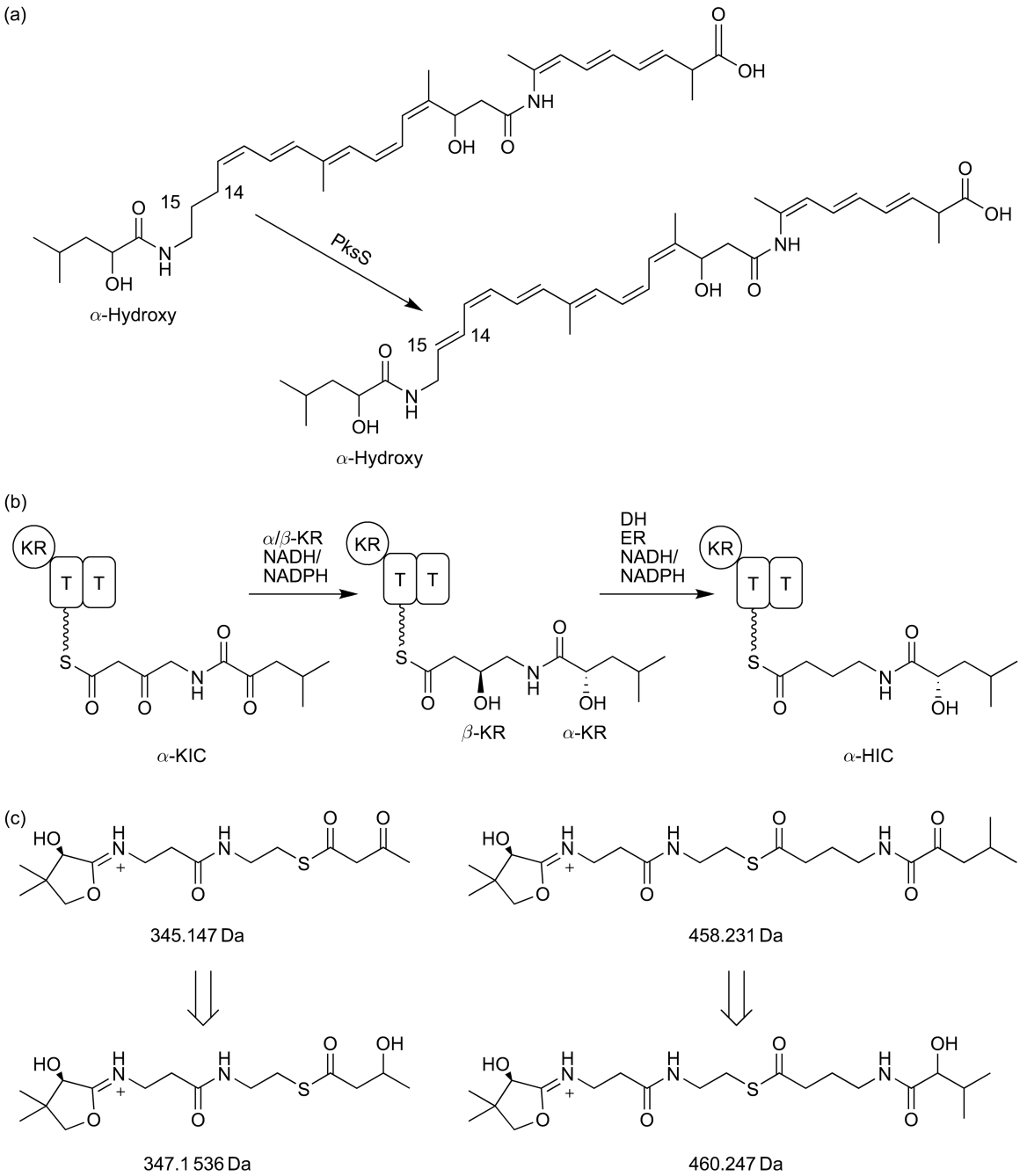


Figure 31 Bacillaene biosynthesis. (a) Structure of dihydrobacillaene and bacillaene. PksJ oxidizes the C14′–C15′ bond after dihydrobacillaene has been synthesized. Also, note the α-hydroxyacyl N-cap. This particular N-capping has been reported very rarely. (b) α- and β-Ketoreduction of α-KIC to α-HIC. The KR domain of the first PKS module in PksJ is capable of reducing both the α-KIC amide and the β-ketone in an NAD(P)H-dependent fashion. The order in which these two reductions occur is unknown. Ultimately, keto-reduction is followed by dehydration and enoyl reduction. (c) Theoretical structure of PPant ejection ions used to analyze PksJ ketoreduction. Right: PPant ejection ion resulting from IRMPD of Acac-*S*-PksJ(T_3-T_4) incubated with PksJ. Mass shift of $+2.017$ Da corresponds with reduction of the β-ketone. Left: PPant ejection ion resulting from IRMPD of α-KIC-GABA-*S*-PksJ(T_3-T_4) incubated with PksJ. Shift of $+2.015$ Da is observed in PPant ejection ions.

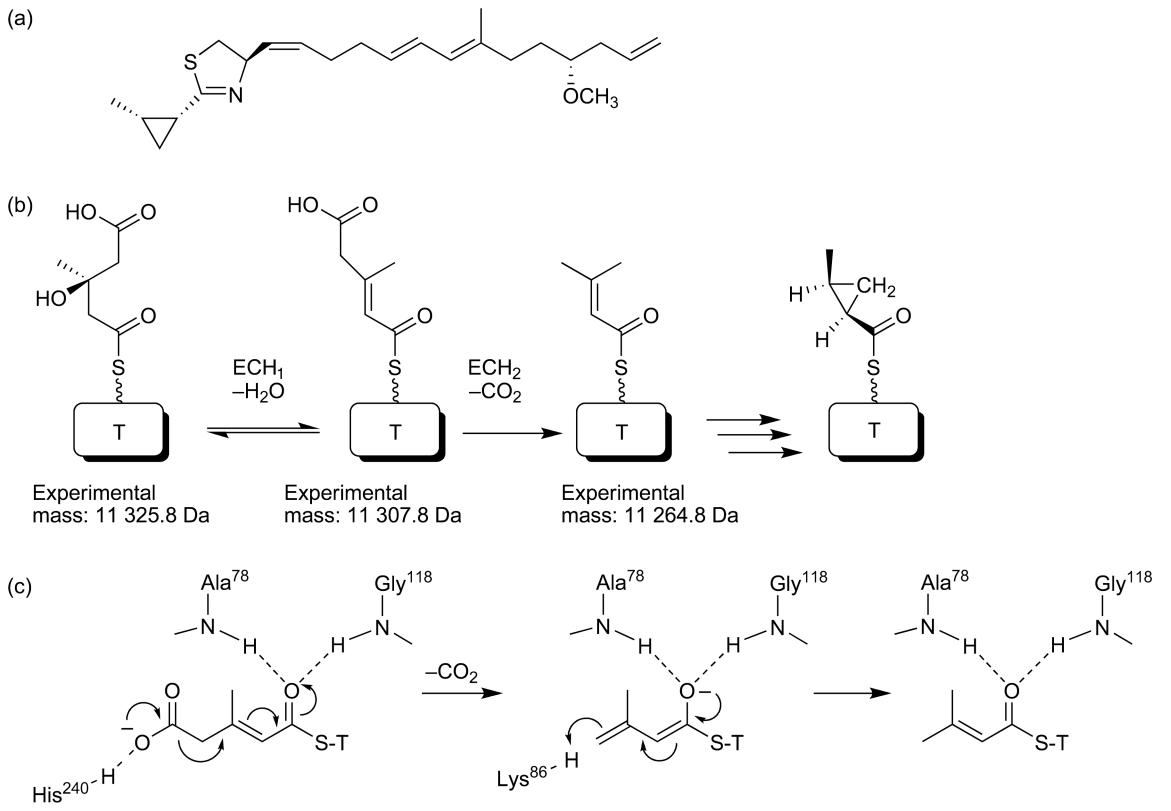


Figure 32 Curacin biosynthesis. (a) Curacin A structure. (b) T domain bound intermediates involved in cyclopropyl ring formation. 3-Hydroxy-3-methylglutaryl (HMG)-*S*-T undergoes dehydration catalyzed by ECH1 to produce 3-methylglutaconyl-*S*-T. ECH2 catalyzes the subsequent decarboxylation to yield 3-methylcrotonyl-*S*-T. (c) Proposed mechanism of decarboxylation of 3-methylglutaconyl-*S*-T by ECH2. The His240 residue of ECH2 acts to position the substrate and prime its decarboxylation. Lys86 donates a proton to the enolate anion. Point mutations of these two residues substantially diminished the production of 3-methylcrotonyl-*S*-T, as detected by FT–ICR–MS.

the cyclopropanyl ring of curacin A. In particular, the functions of the two enoyl-CoA synthases ECH1 (CurE) and ECH2 (CurF) were determined using high-resolution MS.

As members of the functionally diverse crotonase superfamily, ECH1 and ECH2 were expected to have different roles in the formation of the cyclopropanyl ring precursor of curacin A.[122] To probe the function of ECH1 and ECH2, the T domain (CurB) was first incubated with (*S*)-HMG-CoA, in order to covalently attach HMG to the PPant arm of the T domain active site. Next, the individual activities of each enzyme were determined by incubating ECH1, ECH2, or both enzymes with HMG-*S*-T. ESI–FTMS was used to analyze the products of these reactions by detecting the mass differences observed in the various reactions compared to HMG-*S*-T alone. The 12+ ion was used to detect these mass differences. The most abundant mass of HMG-*S*-T was determined to be 11325.8 Da. Incubation with ECH2 alone did not result in any new products. However, incubation of HMG-*S*-T with ECH1 resulted in the detection of a new product with a mass of 11307.8 Da, and incubation of HMG-*S*-T with ECH1 and ECH2 yielded two products with masses of 11 307.8 and 11 264.8 Da. These differences observed by FTMS corresponded to losses of 18 and 62 Da. This provided evidence that ECH1 functions to dehydrate HMG-*S*-T (–18 Da) to form 3-methylglutaconyl-*S*-T. The 3-methylglutaconyl-*S*-T undergoes subsequent decarboxylation, catalyzed by ECH2, to form 3-methylcrotonyl-*S*-T (**Figure 32(b)**). After identifying the exact functions of ECH1 and ECH2, researchers determined the crystal structure of the N-terminal domain ECH2 (CurF).[123] Structural alignments of CurF ECH2 with other members of the crotonase superfamily revealed several key features of the enzyme active site. Crystallization of ECH2 complexed with product analogues was not successful, so computational modeling was used to

identify three polar side chains within the active site chamber that possessed potential catalytic function: Tyr82, Lys86, and His240. The previously established ECH1/ECH2 enzymatic assay was carried out using ECH2 mutants containing Y82F, K86A, K86Q, H240A, and H240Q. The wild-type and mutant ECH2 enzymes were incubated with ECH1 and (*S*)-HMG-*S*-T (CurA-S-T(II)), then the different incubated reaction mixtures were run on a C4 column and eluted with acetonitrile. The samples were analyzed by ESI–FTMS after being redissolved in an electrospray solution (55% acetonitrile:45% water, with 0.05% formic acid and 0.05% TFA). FTMS was used to detect the presence of 12+ charged (*S*)-HMG-*S*-T, 3-methylglutaconyl-*S*-T, and 3-methylcrotonyl-*S*-T as a means of confirming the effect of the site-directed mutagenesis of the various residues. Using this ECH1/ECH2 assay and FTMS, it was determined that substitution of Tyr82 resulted in only minimal reduction in the production of 3-methylcrotonyl-*S*-ACP from HMG-*S*-T, while substitutions of Lys86 and His240 resulted in drastic decreases in product formation. The identification of essential active site residues allowed researchers to propose a mechanism of action for ECH2 (CurF, **Figure 32(c)**). The work carried out with the curacin A biosynthetic pathway highlights the importance of FT–ICR–MS in both the determination of intermediates in NRPS/PKS pathways as well as evaluation of PKS domain functions.

9.11.3.3 Enediyne Biosynthesis: SgcE

MS has proven to be essential in investigations of the biosynthetic pathways of C-1027, an enediyne antitumor antibiotic. C-1027 is an extremely cytotoxic compound, isolated from *Streptomyces globisporus*, consisting of a binding protein (CagA) and a reactive nine-membered enediyne core containing a pair of conjugated acetylenic groups (**Figure 33(a)**). The enediyne core is the key to the cytotoxicity of C-1027 because when it is released from the protein complex it can form a transient biradical species that can induce single-strand and double-strand breaks in DNA molecules.[124] Characterization of the biosynthetic gene cluster revealed the presence of a type I iterative PKS (SgcE) that catalyzes the formation of the nine-membered enediyene core from acyl-CoAs. A unique feature of this particular PKS is the presence of an integrated C-terminal PPTase that covalently attaches 4′-phosphopantetheine to the active site serine of the T domain. Typically, PPTases are freestanding components in PKS biosynthetic pathways. *In vivo* experiments involving the inactivation of proposed SgcE active sites were carried out, resulting in loss of C-1027 production by SgcE. In order to confirm the function of both the T domain and the PPTase domain, high-resolution MS was employed.[54] SgcE was digested briefly with trypsin, HPLC purified, and the fragments were analyzed by FTMS. A 3+ charge peptide fragment containing the T domain active site was determined to contain a mass shift of 340.1 Da, corresponding to the addition of PPant. By using CID to fragment this ion, the PPant modification was localized to the serine residue at position 974. The same 3+ T domain active site peptide was subjected to IRMPD and the characteristic PPant ejection ion was observed, along with the dehydroalanine form and the phosphoserine form of the peptide. The PEA was used to verify the role of the various active sites in the phosphopantetheinylation of the T domain of SgcE. Mutational inactivation of the active site of the T domain and the inactivation either of the two key residues of the PPTase domain resulted in loss of phosphopantetheinylation of the T domain, which was confirmed by the absence of the 3+ T domain active site peptide fragment with the 340.1 Da mass shift. This provided clear evidence that the C-terminal domain does in fact serve the role of a PPTase.

The information gathered during the analysis of the domain functions of SgcE can be used in a hypothesis-driven approach for mapping higher charge states of the active site peptides and to identify substrate- and intermediate-loaded forms of the T domain active site. During the investigations into the function of the PPTase domain of SgcE, the tryptic peptide containing the phosphpantetheinylated active site of the T domain was identified and found to have a mass of 4134.0 Da. Recent work was carried out in the Dorrestein laboratory to detect substrate loading onto the SgcE T domain active site. Armed with the knowledge that the active site containing peptide has a mass of 4134 Da and a 3+ charge, calculations were carried out to predetermine the expected m/z values at which malonyl, acetoacyl, and acetyl loaded forms of the active site peptide would be found. Preliminary substrate loading assays were performed using SgcE. After a 1 h incubation with malonyl-CoA, SgcE was digested for 10 min using trypsin, and then analyzed by capillary–LC–FTMS. MS data were collected using FTMS at a resolution of 50k, and the top 10 most abundant peptides in each scan were fragmented by CID in a data-dependent fashion. This method was ideal because the mass spectra could quickly be searched for the hypothetical substrate-loaded peptides and MS/MS or MS^2 data were already collected to

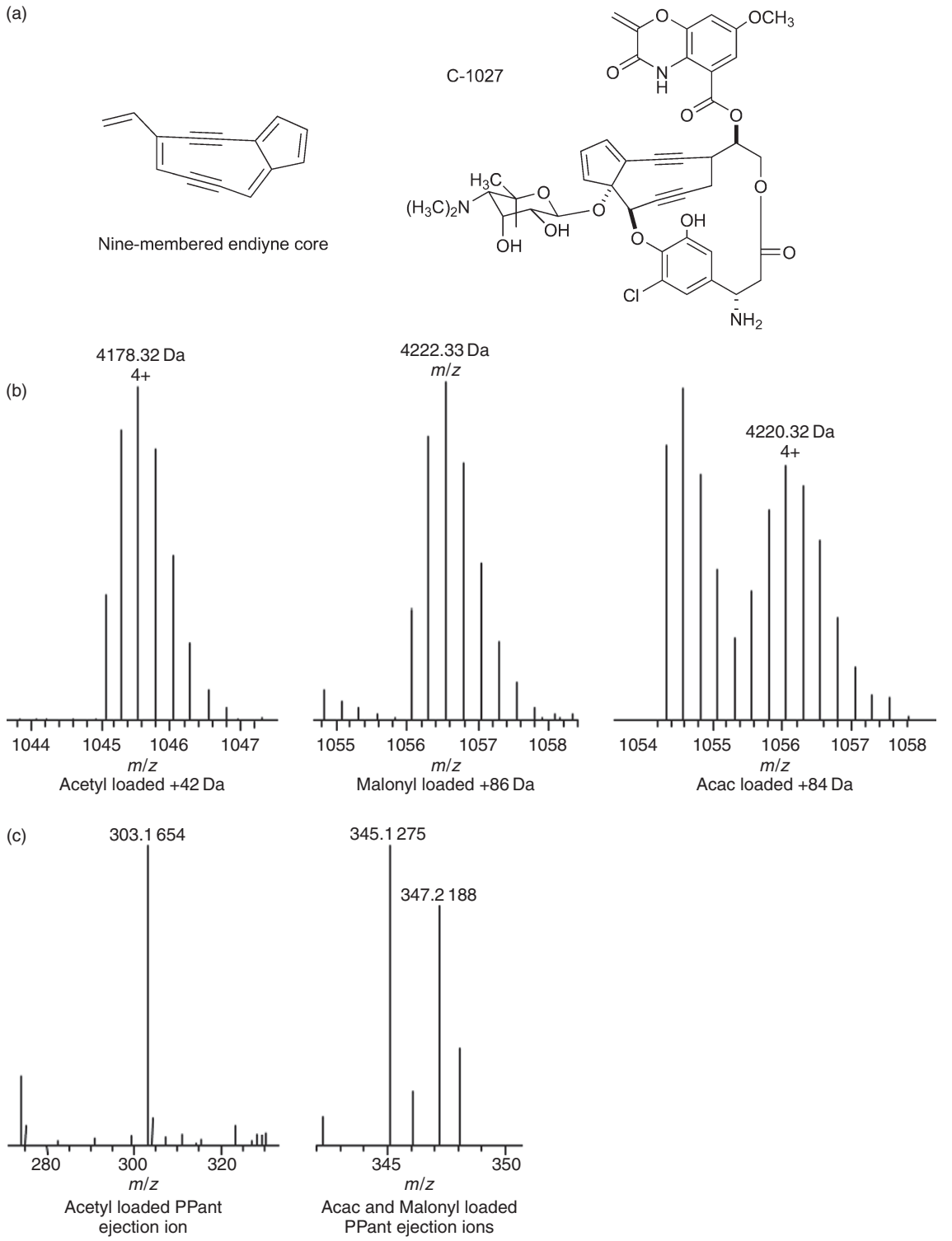


Figure 33 SgcE analysis. (a) The structure of the nine-membered endiyne core (left) produced by SgcE and the complete structure of the C-1027 molecule (right). (b) Active site containing tryptic peptide signals with detected mass shifts corresponding to the loaded substrate/intermediate: acetyl (left), malonyl (middle), and acetoacetyl (left). (c) Phosphopantetheinyl ejection ions. Ejection ion confirms the mass of the bound intermediate. The malonyl intermediate undergoes decarboxylation resulting in the formation of acetyl-(*S*)-T. Malonyl and acectoacetyl-ACP active site fragments coelute during LC–MS. The result is that the *m/z* of the two species are close enough that they are both fragmented and their resulting ejection ions can be detected in the same MS/MS or MS^2 spectrum.

confirm the mass of the loaded substrate by identifying the PPant ejection ion. Using this method, the 4+ phosphopantetheinylated T domain active site peptides were detected having mass shifts of +42 Da (acetyl), +86 Da (malonyl), and +84 Da (acetoacetyl, **Figure 33(b)**). The active site-containing peptide loaded with malonyl has only a 2 Da difference in mass from the active site peptide loaded with acetoacetate, and they partially coelute. While FTMS is capable of resolving the signals of the malonyl and acetoacyl loaded peptides, the PPant ejection ions were clearly detected by MS/MS or MS^2 as shown in **Figure 33(c)**.

Given that LC–FTMS followed by data-dependent fragmentation is capable of detecting the loaded PPant ejection ions, this method can be quite useful for detecting PPant ejection signatures in other PKS systems as well. However, in systems in which phophopantetheinylated active site containing peptides are unknown, or during proteome-level screening, subsequent data analysis of LC–MS data in search of PPant ejection ions can be extremely challenging due to the overwhelming amount of MS/MS or MS^2 data and the difficulty that the preexisting MS/MS or MS^2 analysis programs have with multiple charged precursor peptides. Currently, collaborative work is being carried between the Dorrestein laboratory and the Bafna research group in the computer science department of UCSD to develop a program capable of rapidly scanning LC–FT–MS/MS or MS^2 data for the presence of PPant ejection ions (loaded or not), the corresponding dehydroalanine (−18 Da, $z-1$) form and the phosphoserine (+80, $z-1$) form of the active site peptide in the MS/MS or MS^2 data. This software has already successfully identified PPant ejection signatures in data collected from a number of PKS and NRPS systems. As development of this software progresses, searching LC–FTMS data for PPant ejection will be dramatically increased, and this possesses the potential for *de novo* detection of PPant ejection in highly complex samples that are being analyzed by LC–MS in proteomics screening experiments.

9.11.3.4 Deconstructive Analysis of PksA

Recent efforts have been made to better understand eukaryotic iterative polyketide synthases (IPKSs) and how these biosynthetic pathways ensure the formation of specific products. As opposed to much larger modular PKSs that carry out the biosynthesis and tailoring of their products in an assembly-line fashion, iterative PKSs contain many fewer catalytic domains and the same domains are used multiple times prior to yielding the final product. It has been unclear what exactly determines the number of catalytic cycles the growing intermediate of an IPKS undergoes before it is released. To clarify the 'global division of labor' between the domains of an IPKS, researchers studied PksA by dissecting out the individual domains and reassembling them *in vitro*.[48] By expressing the various domains of the PksA as smaller units and recombining them in different combinations, the hope was to paint a better picture of the function of each domain in the biosynthesis of aflatoxin B_1. Various mono-, di-, and tridomains were expressed. The various domains of PksA were then mixed back with the PksA starter-CoA:acyl carrier protein acyltransferase (SAT)-KS-malonyl-CoA:acyl carrier protein acyltransferase (MAT) domain, and the different products of the reactions between the substrate, SAT-KS-MAT, and the other domains were analyzed. In theory, the full complement of the 6 PksA domains (**Figure 34(a)**) would be required in order to efficiently produce norsolorinic acid, an isolable precursor of aflatoxin B_1, from a starter hexanoyl-CoA and seven malonyl-CoAs.

The products formed by the various combinations of the deconstructed PksA domains, incubated with substrate, were monitored by HPLC. It was found that the combination of SAT-KS-MAT with the putative product template (PT) domain and the T domain yielded significant quantities of a product, which was not norsolorinic acid. Addition of the thioesterase/Claisen-like cyclase (TE/CLC) to the mixture resulted in substantial formation of norsolorinic acid. These findings revealed that product formation was not occurring to an appreciable extent in the absence of the PT domain. Therefore, it was hypothesized that the PT domain acts as a cyclase/aromatase and catalyzes the formation of the first two rings of a putative intermediate. This putative intermediate can then undergo spontaneously C–O ring closure in the absence of the TE/CLC domain to form the shunt product napthopyrone, or it can undergo C–C ring closure in the presence of the TE/CLC domain to from the norsolorinic acid anthrone precursor (**Figure 34(b)**).

In order to verify the proposed structures of the phosphopantetheinyl-bound intermediates, an LC–FTMS assay was employed to detect PPant ejection ions containing bound intermediate. Reactions with SAT-KS-MAT and either the T domain or a PT-T didomain were carried out with both hexanoyl-CoA and malonyl-CoA. The reactions were carried out for varying times, followed by limited trypsin digestion (15 min). The

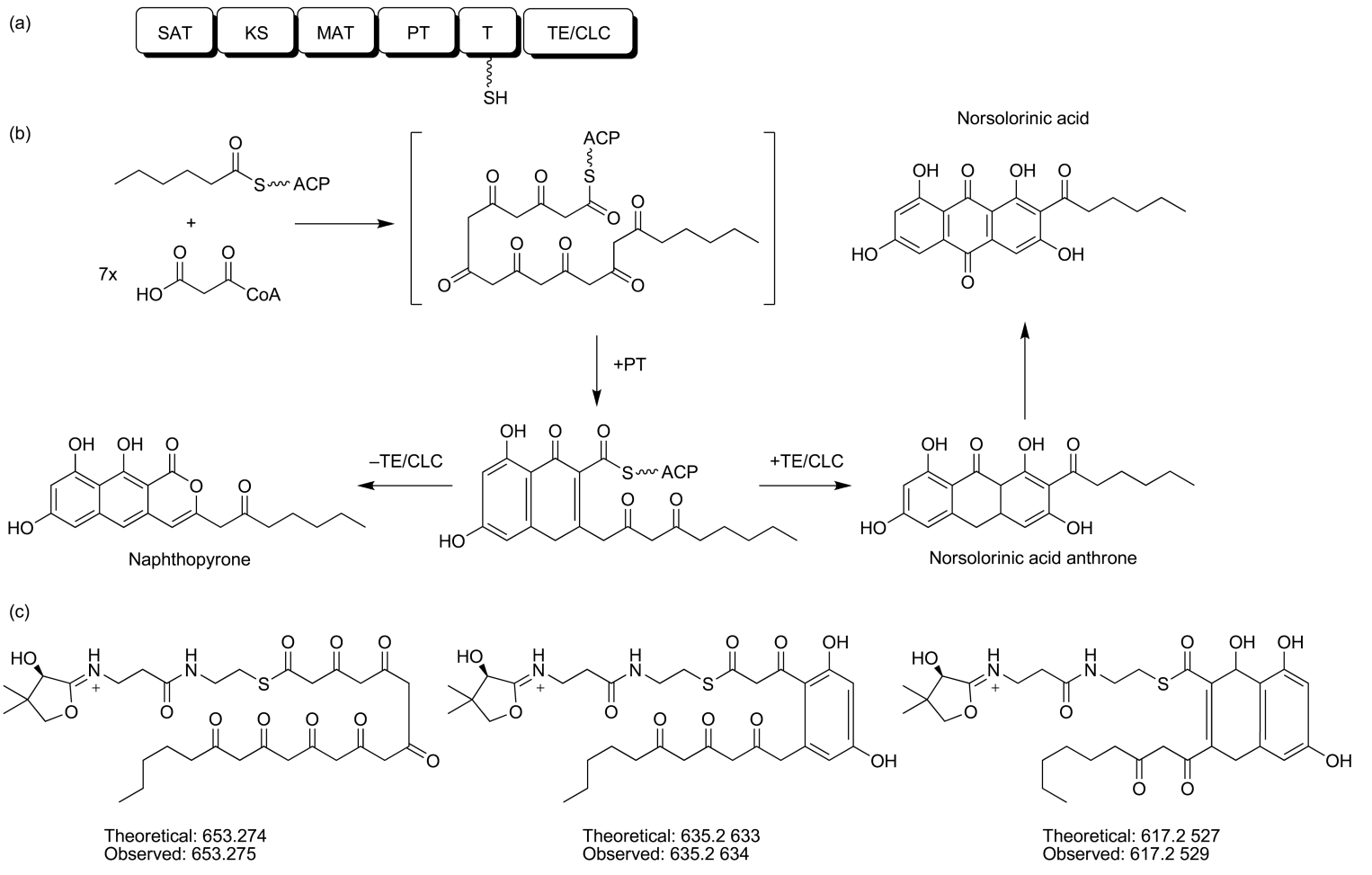


Figure 34 PksA deconstruction. (a) Enzymatic domain architecture of PksA. (b) PksA utilizes a starter hexanoyl-CoA and seven malonyl-CoAs to produce the covalently linked intermediate (brackets). The PT domain acts as an aromatase/cyclase facilitating the closure of the first two rings on the intermediate. In the absence of the TE/CLC domain the intermediate undergoes C–O cyclization to spontaneously form the naphthopyrone. In the presence of the TE/CLC domain, the intermediate undergoes C–C cyclization to from the norsolorinic acid anthrone, which autooxidizes to form norsolorinic acid. (c) Observed PPant ejection ions confirming the structures of the proposed intermediates bound to the active site of the PksA T domain. The first intermediate (left) was detected on the T domain active site after incubation of SAT-KS-MAT with T domain alone. Incubation of SAT-KS-MAT with PT-T results in the formation of the intermediates containing first a single-cyclization product (middle) followed by a double-cyclization product (right).

digested reactions were run separately on a C4 column over a 1 h water/acetonitrile gradient and injected directly into an FTMS. The phosphopantetheinylated active site peptides possessing various intermediates were observed and the structures were confirmed by MS/MS or MS^2 and PPant ejection. The fully extended intermediate generated by a single starter hexanoyl unit and seven malonyl units was detected bound to the T domain active site. In addition to this PPant-bound intermediate, the intermediate was detected as a single dehydrated compound with a signal aromatic ring or it was detected as a double dehydrated compound with two closed rings. The MS/MS or MS^2 PPant ejection ions reveal mass shifts with less than 1.5 ppm mass accuracy. These ejection ions were subjected to an additional round of fragmentation and the resulting ions were mapped to the structure of the intermediate (**Figure 34(c)**). The T domain was observed to accept both hexanoyl-CoA and malonyl-CoA, but PksA preferentially loads hexanoyl-CoA as the starter unit. The MS/MS or MS^2 PPant ejection ions reveal mass shifts with less than 1.5 ppm mass accuracy. The high mass accuracy of the FTMS allowed for the accurate determination of the mass of PPant-bound intermediates, even for relatively large intermediates such as the full extended octa-ketide intermediate of norsolorinic acid.

9.11.4 Prospective Applications of Current Natural Product MS Methods on NRPS and PKS Systems

Although more high-resolution mass spectrometrists are entering the field, there are currently just a handful of investigators that are developing novel approaches to investigate biosynthetic pathways of secondary metabolites. McLafferty has certainly been one of the early driving forces in the 1990s in this area but since that time others, including Kelleher, Hakansson, and Leary, began to use high-resolution MS creatively to investigate biosynthetic pathways. Some of these creative approaches are highlighted in this section.

9.11.4.1 Isotopically Depleted Proteins and Peptides

There are several factors that add to the complexity of a mass spectrum and that thus make identification of a specific protein difficult. For instance, it is not uncommon that a protein forms phosphate, sodium, or potassium adducts and that it exists in multiple oxidized forms. In addition, other protein or peptides may be ionized at the same time. These various forms of a specific protein and contaminants entering the MS instrument can make a single mass spectrum very complex. This complexity may make it tremendously difficult to identify the ions of interest with a high degree of certainty. For example, the biosynthesis of lacticin 481 by the protein LctM from the ribosomally encoded peptide LtcA requires multiple phosphorylations and dehydrations followed by thioether bridging (**Figure 35(a)**). It was not possible to discern the details of the mechanism observed for lacticin 481 biosynthesis when LtcA contained normal abundance isotopic levels because it was prone to oxidation during this process. Therefore LtcA was grown in isotopically ^{13}C- and ^{15}N-depleted material and the maturation of LtcA by LtcM was analyzed by FT–ICR–MS. **Figure 35(b)** shows the time course of the conversion of LtcA to lacticidin 481. Very little buildup of biosynthetic intermediates is observed, which characterizes LtcM as a processive enzyme. The abovementioned example shows how useful ^{13}C- and ^{15}N-depleted materials can be if the spectrum is too complex to reliably interpret the data. The depleted protein strategy is also useful when the protein is of reasonable size and yet one is interested in observing small mass changes such as dehydrogenation, deamidation, or transfer of an isotopically labeled substrate to the recipient protein. One such scenario has been found in the thiazole phosphate biosynthetic pathway in prokaryotes.

One of the pathways to thiazole phosphate, the active portion of thiamin utilizes a ubiquitin-like protein called ThiS, an E1-type protein called ThiF, a flavin-dependent oxidase ThiO, the thiazole synthase ThiG, a sulfur source, ATP, 1-deoxy-D-xylulose-5-phosphate, and glycine.[125] This pathway is initiated by the adenylation of ThiS by ThiF (**Figure 36**). Once the acyl–adenylate is formed, a sulfur, presumably from a cysteine desulferase-mediated reaction, displaces the adenylate to form a C-terminal ThiS-thiocarboxylate. Subsequently, it reacts with deoxy-D-xylulose-phosphate (DXP) that is transiently but covalently tethered to the thiazole synthase ThiG **4**. The sulfur from ThiS is transferred to the C3 carbon of DXP **5**. During this process, one of the hydroxyl groups from DXP is transferred to the C-terminal end of ThiS forming a stable

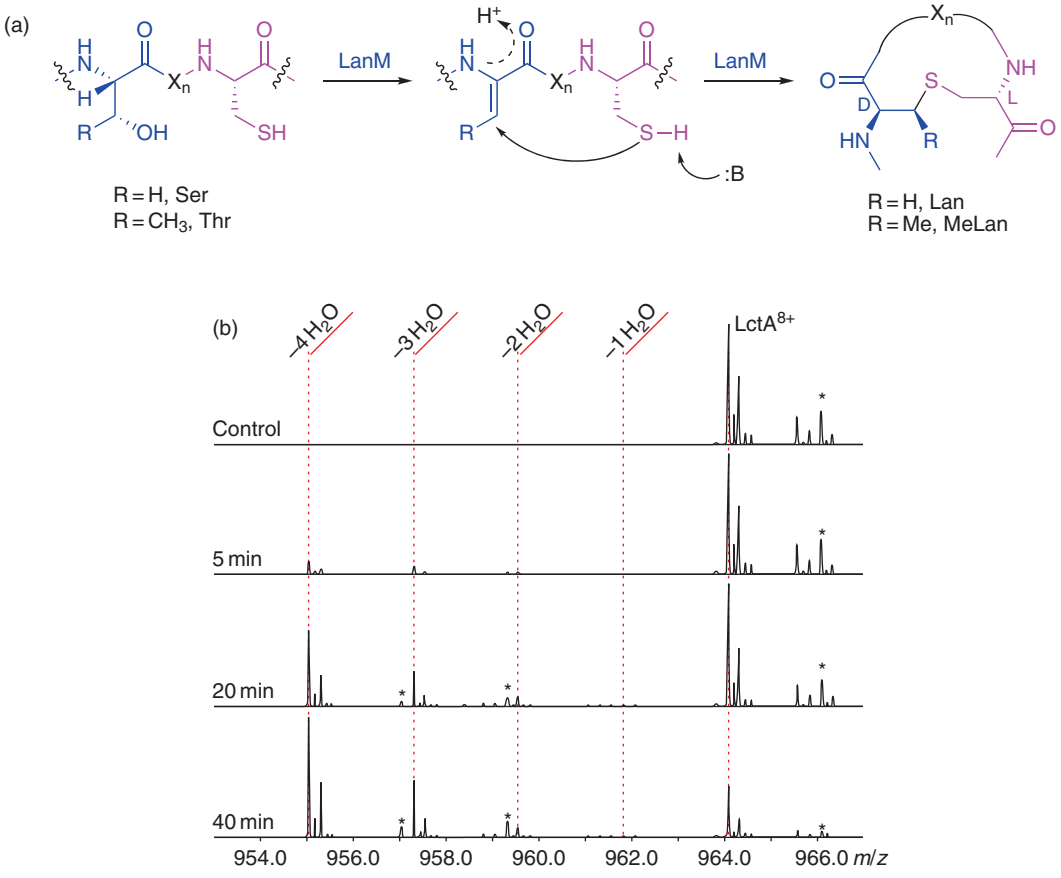


Figure 35 Conversion of LctA to the mature form of lacticidin 481 catalyzed by LtcM. (a) Thioether bridge formation in lantibiotic biosynthesis leading to dehydration (−18 Da). (b) MS time course of isotopically depleted LctA conversion to lacticidin 481 by LtcM shows that little dehydration intermediates are observed.

thioenolate intermediate on ThiG **6**. Once ThiO oxidizes glycine **15** and transfers the corresponding oxidized glycine **16**, it cyclizes and forms a stable carboxy-thiazolephosphate species **11** that is decarboxylated when it is coupled to the pyrimidine portion of thiamin.[126] This pathway is arguably one of the most complex biosynthetic pathways, involving several protein substrate interactions described in the literature to date.

^{13}C- and ^{15}N-depleted ThiS was necessary to visualize oxygen transfer shown as steps 5–8 in **Figure 36**. Oxygen transfer from DXP to the C-terminal end of ThiS was monitored by the incorporation of ^{18}O from ^{18}O-labeled DXP. Since ^{18}O-DXP was generated enzymatically, it was only partially ^{18}O-labeled. Therefore the incorporation was limited. When nondepleted material was used and the conversion of ThiS-thiocarboxylate to ThiS was monitored, a small shift was observed in the spectral profile but it could not be conclusively demonstrated that ^{18}O had indeed been incorporated. There were three reasons for this: (1) there was only partial incorporation, (2) there was still some ThiS-carboxylate sample left in the ThiS-thiocarboxylate sample at the onset of the reaction, and (3) there was no good way to deconvolute the contribution of an ^{18}O to the overall spectral profile. However when ^{13}C- and ^{15}N-depleted ThiS was generated, the shift of 2 Da was readily visualized. (**Figure 37**). A protein that is depleted shows a rather different isotopic envelope compared to the one described for Pks4 in **Figure 4**. In the depleted protein the main species observed is the monoisotopic ion. The second most abundant ion is due to the contribution of ^{18}O from naturally abundant oxygen isotopes. The +2 isotope dramatically increases in intensity when ThiS-thiocarboxylate is incubated with ^{18}O-DXP. Subsequent MS/MS or MS^2 demonstrated that this incorporation was at the C-terminal end of ThiS. This provides a good example of the utility of ^{13}C- and ^{15}N-depleted protein strategy.

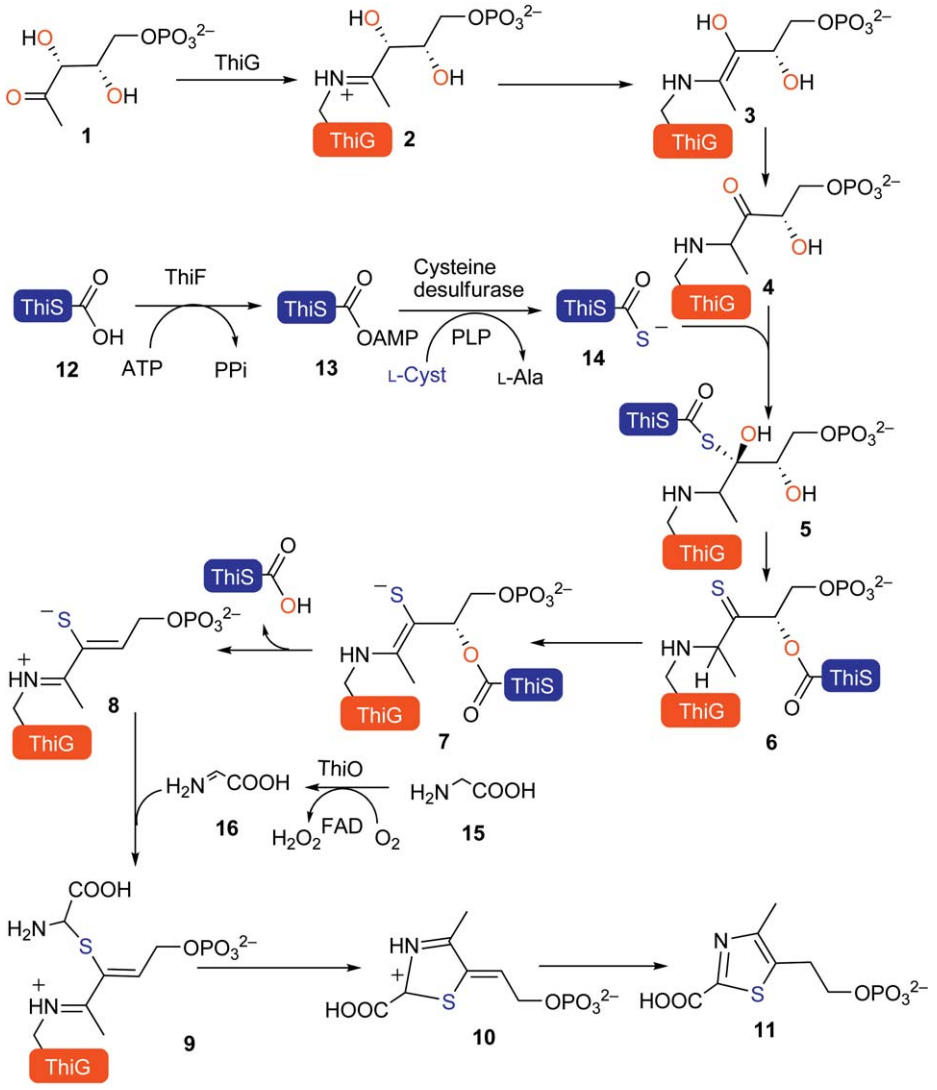


Figure 36 A thiazole phosphate biosynthetic pathway that has been studied by application of isotopically depleted proteins and mass spectrometry. The blue boxes represent ThiS.

9.11.4.2 Trapping Reactive Intermediates

In addition to the unusual transfer of a hydroxyl group from a substrate (DXP) to the C-terminal end on a protein, there are several reactive intermediates on the thiazole synthase pathway.[127] In order to visualize these intermediates they had to be trapped before they could be visualized by MS. Imine **2** or its Amadori rearrangement product **4** in **Figure 35** were trapped with $NaBH_4$. $NaBH_4$ reduction resulted in an irreversible linkage to the thiazole synthase ThiG. Similar type of sugar substrate–imine linkage has also been observed for other biosynthetic pathways such as the ones found in pyridoxal phosphate biosynthesis.[128–130] Trapping intermediates like this will no doubt be important in the investigations of NRPS and PKS biosynthetic pathways as well.

9.11.4.3 High-Resolution Mass Spectrometry of Noncovalent Interactions

Another promising MS approach is the investigation of noncovalent complexes. A beautiful example of this was provided by Leary in a study of 5′-adenylylsulfate (APS) reductase, a 4F–4S iron–sulfur cluster-containing protein.[131] APS reductase catalyzes the reduction of APS to sulfite. Leary and coworkers were able to observe

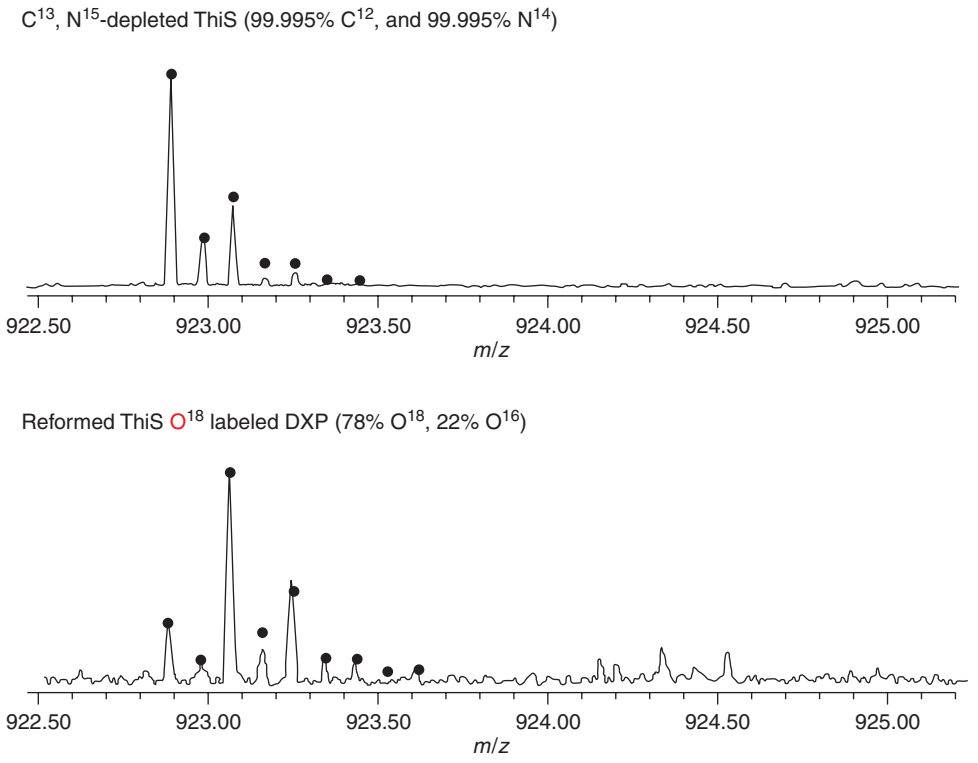


Figure 37 FT–ICR–MS signal of isotopically depleted ThiS (top) and the effect of ^{18}O incorporation (bottom) in ThiS.

the intact iron–sulfur cluster and even to show the interaction of the protein with APS, AMP, and thioredoxin, a cofactor in this biosynthetic process. This result of visualizing the intact APS reductase with its noncovalent substrates is remarkable because it was always believed that iron–sulfur clusters are unstable. Undoubtingly, the application of this approach to NRPS and PKS systems would reveal new insights into their biosynthetic pathways.

9.11.5 Up-and-Coming Advances in Mass Spectrometry Tools for the Investigation of Natural Products and Their Biosynthetic Pathways

In the last few years, mass spectrometers have undergone a revolution. Just 5 years ago, FT–ICR–MS was the only type of MS that was able to obtain sufficient mass accuracy and resolution that enabled the investigations of NRPS and PKS systems in any detail. Since then new methods have evolved, instruments have improved, and additional mass spectrometers have become available that have a mass accuracy within 10 ppm. An example is the ORBI-trap, which is the only other FT analyzer and which measures an image current too. This instrument should be capable of doing any of the experiments described in this chapter. An ORBI-trap can achieve resolutions >100 000. In addition to the ORBI-trap, the most recent Q–TOFs have a resolution of 60 000 and a mass accuracy of sub-parts per million. This mass accuracy rivals the mass accuracy that can be achieved by modern FT–ICR–MS instrumentation. As high-resolution instruments are becoming commonplace, we anticipate that many other investigators will begin using the approaches described in this chapter as well.

In addition to higher mass accuracy instruments, there are also new MS tools that are going to be very exciting in the investigations of NRPS and PKS systems or biosynthetic pathways in general. One such tool is ion mobility. Ion mobility will enable us to look at large complexes and the changes that take place on such complexes. It is also likely that the PEA can be taken to another level as it may be possible to separate the

ejected ion from the rest of the peptides in the gas phase. All these exciting developments in MS will ensure that the investigations of these important natural products will be possible with higher accuracy, with smaller samples, and at a faster throughput. Maybe, someday in the next few decades, we may be able to perform these biosynthetic studies at single cell levels.

Acknowledgments

Pieter C. Dorrestein appreciates all the funding he has received from the ACS and Moores Cancer Center, Beckman Foundation, PhRMA Foundation, V-Foundation, and NIH (1R01GM086283 and 1R01GM085128), and the Dorrestein laboratory members and all his colleagues and collaborators on the numerous projects described in this chapter. In addition, the authors thank Michael Thomas for providing the adenylation T didomain construct from **Figure 10**, Yi Tang for providing Pks4, and Ben Shen for providing SgcE. RDK appreciates his USA scholarship from the Studienstiftung des deutschen Volkes.

Abbreviations

3,4-DHBA	3,4-dihydroxybenzoic acid
A (domain)	adenylation
Acac	acetoacetyl
ACP	acyl carrier protein
AL	acyl ligase
AMP	adenosine monophosphate
APS	5′-adenylylsulfate
A–T (didomain)	adenylation–thiolation (didomain)
AT	acyltransferase (domain)
ATP	adenosine triphosphate
ATP–PPi	adeonsine triphosphate–pyrophosphate exchange
BIRD	blackbody infrared radiative dissociation
C (domain)	condensation (domain)
CD	circular dichroism
CDA	calcium-dependent antibiotic
CID	collisionally-induced dissociation
CLC	Claisen-like cyclase
CLF	chain length factor
CoA	coenzyme A
DH (domain)	dehydratase (domain)
DXP	deoxy-D-xylulose-phosphate
E (domain)	epimerization (domain)
ER (domain)	enoyl reductase (domain)
ESI	electrospray ionization
ESI–FT–ICR–MS	electrospray ionization–Fourier transform–ion cyclotron resonance–mass spectrometry
ESI–FTMS	electrospray ionization–Fourier transform mass spectrometry
ESI–MS	electrospray ionization–mass spectrometry
ESI–Q–TOF MS	electrospray ionization–quadrupole–time-of-flight–mass spectrometry
ETD	electron transfer dissociation
FAS	fatty acid synthase
FID	free induction decay
FT–ICR–MS	Fourier transform–ion cyclotron resonance–mass spectrometry

FTMS/FT–MS	Fourier transform–mass spectrometry
GC/EI–MS	gas chromatography/electron impact–mass spectrometry
HCS	HMG-CoA synthase
HIC	hydroxyl isocaproate
HMG	3-hydroxy-3-methylglutaryl
HPLC	high-performance liquid chromatography
IPKS	iterative polyketide synthase
IR	infrared
IRMPD	infrared multiphoton dissociation
KIC	ketoisocaproate
KR (domain)	ketoreductase (domain)
KS (domain)	ketosynthase (domain)
LC	liquid chromatography
LC–MS	liquid chromatography–mass spectrometry
LC–MS–PEA	liquid chromatography–mass spectrometry–phosphopantetheinyl ejection assay
LCQ	three-dimensional linear quadrupole
Linear IT	linear ion trap
LTQ	two-dimensional linear trap quadrupole
LTQ–FT–ICR–MS	hybrid linear trap quadrupole–Fourier transform–ion cyclotron resonance–mass spectrometry
LTQ–ORBI	hybrid linear trap quadrupole–orbitrap
MALDI	matrix-assisted laser desorption/ionization
MALDI–TOF	matrix-assisted laser desorption/ionization–time-of-flight
MAT (domain)	malonyl-CoA:acyl carrier protein acyltransferase
MDa	megadalton
MS	mass spectrometry
MS/MS (or MS^2)	tandem mass spectrometry
MS^3	additional fragmentation of ions generated by MS/MS or MS^2
NAD(P)H	nicotinamide adenine dinucleotide (phosphate) hydride
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide hydride
NADPH	nicotinamide adenine dinucleotide phosphate hydride
NMR	nuclear magnetic resonance
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthetase
NRPS/PKS	hybrid nonribosomal peptide synthetase/polyketide synthase
PEA	phosphopantetheinyl ejection assay
PKS	polyketide synthase
PPant	phosphopantetheine
PPTase	phosphopantetheinyl transferase
PQD	pulsed-Q dissociation
Q–TOF	quadrupole–time-of-flight
RP–HPLC	reverse-phase high-performance liquid chromatography
RPLC	reverse-phase liquid chromatography
SAT (domain)	starter-CoA:acyl carrier protein acyltransferase
SerT	seryltransferase
SORI–CAD	sustained off-resonance irradiation–collisionally activated dissociation
T (domain)	thiolation (domain)
TE (domain)	thioesterase
TE/CLC (domain)	thioesterase/Claisen-like cyclase
TGH	transglutaminase homologue

UV–vis	ultraviolet–visible spectroscopy
X-ray	X-ray diffraction techniques

Nomenclature

Asn	asparagine
Cys	cysteine
Da	dalton
His	histidine
kDa	kilo dalton
Lys	lysine
m/z	mass/charge
nl min^{-1}	nanoliters per minute
Phe	phenylalanine
ppm	parts per million
–S–	bonding sulfur of thioester
T (unit)	tesla
Tyr	tyrosine
***z* (unit)**	charge
μg	microgram
μl min^{-1}	microliter per minute
μmol l^{-1}	micromolar

References

1. D. J. Newman; G. M. Cragg, *J. Nat. Prod.* **2007**, *70*, 461–477.
2. J. E. Syka; J. J. Coon; M. J. Schroeder; J. Shabanowitz; D. F. Hunt, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
3. R. A. Zubarev; D. M. Horn; E. K. Fridriksson; N. L. Kelleher; N. A. Kruger; M. A. Lewis; B. K. Carpenter; F. W. McLafferty, *Anal. Chem.* **2000**, *72*, 563–573.
4. H. J. Cooper; K. Håkansson; A. G. Marshall, *Mass Spectrom. Rev.* **2005**, *24*, 201–222.
5. J. Pól; P. Novák; M. Volný; G. H. Kruppa; R. Kostiainen; K. Lemr; V. Havlíček, *Eur. J. Mass Spectrom.* **2008**, *14*, 391–399.
6. A. A. Shvartsburg; R. D. Smith, *Anal. Chem.* **2008**, *80*, 9689–9699.
7. N. E. Manicke; T. Kistler; D. R. Ifa; R. G. Cooks; Z. Ouyang, *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 321–325.
8. Z. Takáts; J. M. Wiseman; B. Gologan; R. G. Cooks, *Science* **2004**, *306*, 471–473.
9. N. D. Udeshi; P. D. Compton; J. Shabanowitz; D. F. Hunt; K. L. Rose, *Nat. Protoc.* **2008**, *3*, 1709–1717.
10. G. C. McAlister; W. T. Berggren; J. Griep-Raming; S. Horning; A. Makarov; D. Phanstiel; G. Stafford; D. L. Swaney; J. E. Syka; V. Zabrouskov; J. J. Coon, *J. Proteome Res.* **2008**, *7*, 3127–3136.
11. T. R. Northen; J. C. Lee; L. Hoang; J. Raymond; D. R. Hwang; S. M. Yannone; C. H. Wong; G. Siuzdak, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3678–3683.
12. T. R. Northen; O. Yanes; M. T. Northen; D. Marrinucci; W. Uritboonthai; J. Apon; S. L. Golledge; A. Nordström; G. Siuzdak, *Nature* **2007**, *449*, 1033–1036.
13. H. Han; Y. Xia; M. Yang; S. A. McLuckey, *Anal. Chem.* **2008**, *80*, 3492–3497.
14. Y. Nakata; Y. Honda; S. Ninomiya; T. Seki; T. Aoki; J. Matsuo, *J. Mass Spectrom.* **2009**, *44*, 128–136.
15. A. Brunelle; O. Laprévote, *Anal. Bioanal. Chem.* **2009**, *393*, 31–35.
16. H. Hazama; J. Aoki; H. Nagao; R. Suzuki; T. Tashima; K. Fujii; K. Masuda; K. Awazu; M. Toyoda; Y. Naito, *Appl. Surf. Sci.* **2008**, *255*, 1257–1263.
17. P. Chaurand; S. A. Schwartz; R. M. Caprioli, *Curr. Opin. Chem. Biol.* **2002**, *6*, 676–681.
18. Q. Hu; R. J. Noll; H. Li; A. Makarov; M. Hardman; R. Graham Cooks, *J. Mass Spectrom.* **2005**, *40*, 430–443.
19. J. Van der Greef; R. Van der Heijden; E. R. Verheij, *Adv. Mass Spectrom.* **2004**, *16*, 145–165.
20. A. L. Demain, *Appl. Microbiol. Biotechnol.* **1999**, *52*, 455–463.
21. J. Dubois; D. Guenard; F. Gueritte, *Expert Opin. Ther. Pat.* **2003**, *13*, 1809–1823.
22. L. Du; B. Shen, *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 215–228.

23. M. A. Fischbach; C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468–3496.
24. Z. Zhou; P. Cironi; A. J. Lin; Y. Xu; S. Hrvatin; D. E. Golan; P. A. Silver; C. T. Walsh; J. Yin, *Chem. Biol.* **2007**, *2*, 337–346.
25. U. Linne; A. Schäfer; M. T. Stubbs; M. A. Marahiel, *FEBS Lett.* **2007**, *581*, 905–910.
26. C. T. Walsh; H. Chen; T. A. Keating; B. K. Hubbard; H. C. Losey; L. Luo; C. G. Marshall; D. A. Miller; H. M. Patel, *Curr. Opin. Chem. Biol.* **2001**, *5*, 525–534.
27. F. Kopp; M. A. Marahiel, *Curr. Opin. Biotechnol.* **2007**, *18*, 513–520.
28. K. M. Hoyer; C. Mahlert; M. A. Marahiel, *Chem. Biol.* **2007**, *14*, 13–22.
29. Y. Hu; V. Phelan; I. Ntai; C. M. Farnet; E. Zazopoulos; B. O. Bachmann, *Chem. Biol.* **2007**, *14*, 691–701.
30. L. C. Blasiak; F. H. Vaillancourt; C. T. Walsh; C. L. Drennan, *Nature* **2006**, *440*, 368–371.
31. N. Peric-Concha; B. Borovicka; P. F. Long; D. Hranueli; P. G. Waterman; I. S. Hunter, *J. Biol. Chem.* **2005**, *280*, 37455–37460.
32. A. Li; J. Piel, *Chem. Biol.* **2002**, *9*, 1017–1026.
33. L. Tang; S. Shah; L. Chung; J. Carney; L. Katz; C. Khosla; B. Julien, *Science* **2000**, *287*, 640–642.
34. D. J. Edwards; B. L. Marquez; L. M. Nogle; K. McPhail; D. E. Goeger; M. A. Roberts; W. H. Gerwick, *Chem. Biol.* **2004**, *11*, 817–833.
35. W. Liu; S. D. Christenson; S. Standage; B. Shen, *Science* **2002**, *297*, 1170–1173.
36. E. Guenzi; G. Galli; I. Grgurina; D. C. Gross; G. Grandi, *J. Biol. Chem.* **1998**, *273*, 32857–32863.
37. X. Wei; F. Yang; D. C. Straney, *Can. J. Microbiol.* **2005**, *51*, 423–429.
38. S. Tanner; H. Shu; A. Frank; L. C. Wang; E. Zandi; M. Mumby; P. A. Pevzner; V. Bafna, *Anal. Chem.* **2005**, *77*, 4626–4639.
39. L. McHugh; J. W. Arthur, *PLoS Comput. Biol.* **2008**, *4*, e12.
40. F. X. Wu; P. Gagné; A. Droit; G. G. Poirier, *Bioinformatics* **2008**, *9*, S13.
41. C. A. Shaw-Reid; N. L. Kelleher; H. C. Losey; A. M. Gehring; C. Berg; C. T. Walsh, *Chem. Biol.* **1999**, *6*, 385–400.
42. D. B. Stein; U. Linne; M. Hahn; M. A. Marahiel, *ChemBioChem* **2006**, *7*, 1807–1814.
43. R. Gerber; L. Lou; L. Du, *J. Am. Chem. Soc.* **2009**, *131*, 3148–3149.
44. C. Qiao; D. J. Wilson; E. M. Bennett; C. C. Aldrich, *J. Am. Chem. Soc.* **2007**, *129*, 6350–6351.
45. L. Gu; T. W. Geders; B. Wang; W. H. Gerwick; K. Håkansson; J. L. Smith; D. H. Sherman, *Science* **2007**, *318*, 970–974.
46. M. Strieker; F. Kopp; C. Mahlert; L. O. Essen; M. A. Marahiel, *Chem. Biol.* **2007**, *2*, 187–196.
47. P. C. Dorrestein; N. L. Kelleher, *Nat. Prod. Rep.* **2006**, *23*, 893–918.
48. J. M. Crawford; P. M. Thomas; J. R. Scheerer; A. L. Vagstad; N. L. Kelleher; C. A. Townsend, *Science* **2008**, *320*, 243–246.
49. S. B. Bumpus; N. A. Magarvey; N. L. Kelleher; C. T. Walsh; C. T. Calderone, *J. Am. Chem. Soc.* **2008**, *130*, 11614–11616.
50. P. C. Dorrestein; S. B. Bumpus; C. T. Calderone; S. Garneau-Tsodikova; Z. D. Aron; P. D. Straight; R. Kolter; C. T. Walsh; N. L. Kelleher, *Biochemistry* **2006**, *45*, 12756–12766.
51. D. B. Hansen; S. B. Bumpus; Z. D. Aron; N. L. Kelleher; C. T. Walsh, *J. Am. Chem. Soc.* **2007**, *129*, 6366–6367.
52. F. Kopp; U. Linne; M. Oberthür; M. A. Marahiel, *J. Am. Chem. Soc.* **2008**, *130*, 2656–2666.
53. Y. A. Chan; M. T. Boyne, II; A. M. Podevels; A. K. Klimowicz; J. Handelsman; N. L. Kelleher; M. G. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14349–14354.
54. J. Zhang; S. G. Van Lanen; J. Ju; W. Liu; P. C. Dorrestein; W. Li; N. L. Kelleher; B. A. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1460–1465.
55. W. L. Kelly; M. T. Boyne, II; E. Yeh; D. A. Vosburg; D. P. Galonić; N. L. Kelleher; C. T. Walsh, *Biochemistry* **2007**, *46*, 359–368.
56. C. T. Calderone; D. F. Iwig; P. C. Dorrestein; N. L. Kelleher; C. T. Walsh, *Chem. Biol.* **2007**, *14*, 835–846.
57. C. T. Calderone; W. E. Kowtoniuk; N. L. Kelleher; C. T. Walsh; P. C. Dorrestein, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8977–8982.
58. C. T. Calderone; S. B. Bumpus; N. L. Kelleher; C. T. Walsh; N. A. Magarvey, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12809–12814.
59. A. G. Marshall; C. L. Hendrickson, *Annu. Rev. Anal. Chem.* **2008**, *1*, 579–599.
60. A. G. Marshall; C. L. Hendrickson; G. S. Jackson, *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
61. T. Liu; M. E. Belov; N. Jaitly; W. J. Qian; R. D. Smith, *Chem. Rev.* **2007**, *107*, 3621–3653.
62. W. Zhang; Y. Li; Y. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20683–20688.
63. N. L. Kelleher; M. W. Senko; M. M. Siegel; F. W. McLafferty, *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 380–383.
64. N. L. Kelleher, University of Illinois at Urbana-Champaign, Urbana, IL. Unpublished work, 2006.
65. P. J. Ulintz; A. K. Yocum; B. Bodenmiller; R. Aebersold; P. C. Andrews; A. I. Nesvizhskii, *J. Proteome Res.* **2009**, *8*, 887–899.
66. A. Tholey; J. Reed; W. D. Lehmann, *J. Mass Spectrom.* **1999**, *34*, 117–123.
67. M. Edelson-Averbukh; R. Pipkorn; W. D. Lehmann, *Anal. Chem.* **2006**, *78*, 1249–1256.
68. A. M. Palumbo; J. J. Tepe; G. E. Reid, *J. Proteome Res.* **2008**, *7*, 771–779.
69. S. G. Van Lanen; S. Lin; P. C. Dorrestein; N. L. Kelleher; B. Shen, *J. Biol. Chem.* **2006**, *281*, 29633–29640.
70. M. P. Barrow; W. I. Burkitt; P. J. Derrick, *Analyst* **2005**, *130*, 18–28.
71. L. Sleno; D. A. Volmer, *J. Mass Spectrom.* **2004**, *39*, 1091–1112.
72. W. D. Price; P. D. Schnier; E. R. Williams, *Anal. Chem.* **1996**, *68*, 859.
73. J. W. Gauthier; T. R. Trautman; D. B. Jacobson, *Anal. Chim. Acta.* **1991**, *246*, 211.
74. D. P. Little; J. P. Speir; M. W. Senko; P. B. O'Connor; F. W. McLafferty, *Anal. Chem.* **1994**, *66*, 2809–2815.
75. W. Li; C. L. Hendrickson; M. R. Emmett; A. G. Marshall, *Anal. Chem.* **1999**, *71*, 4397–4402.
76. M. W. Senko; J. P. Speir; F. W. McLafferty, *Anal. Chem.* **1994**, *66*, 2801–2808.
77. A. H. Payne; G. L. Glish, *Anal. Chem.* **2001**, *73*, 3542–3548.
78. A. H. Racine; A. H. Payne; P. M. Remes; G. L. Glish, *Anal. Chem.* **2006**, *78*, 4609–4614.
79. J. C. Schwartz; J. E. P Syka; S. T. Quarmby, In Proceedings of the 53rd ASMS Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, 5–9 June 2005.
80. T. Schlabach; T. Zhang; K. Miller; R. Kiyonami, The 2006 ABRF Conference, Long Beach, CA, 11–14 February 2006.
81. D. Meluzzi; W. H. Zheng; M. Hensler; V. Nizet; P. C. Dorrestein, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3107–3111.
82. L. M. Hicks; M. T. Mazur; L. M. Miller; P. C. Dorrestein; N. A. Schnarr; C. Khosla; N. L. Kelleher, *ChemBioChem* **2006**, *7*, 904–907.
83. P. D. Straight; M. A. Fischbach; C. T. Walsh; D. Z. Rudner; R. Kolter, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 305–310.

84. B. Bothner; G. Siuzdak, *ChemBioChem* **2004**, *5*, 258–260.
85. A. R. McKay; B. T. Ruotolo; L. L. Ilag; C. V. Robinson, *J. Am. Chem. Soc.* **2006**, *128*, 11433–11442.
86. H. Donato; N. I. Krupenko; Y. Tsybovsky; S. A. Krupenko, *J. Biol. Chem.* **2007**, *282*, 34159–34166.
87. J. Yin; P. D. Straight; S. Hrvatin; P. C. Dorrestein; S. B. Bumpus; C. Jao; N. L. Kelleher; R. Kolter; C. T. Walsh, *Chem. Biol.* **2007**, *14*, 303–312.
88. R. S. Brown; A. Aman, *J. Org. Chem.* **1997**, *62*, 4816–4820.
89. S. H. Payne; M. Yau; M. B. Smolka; S. Tanner; H. Zhou; V. Bafna, *J. Proteome Res.* **2008**, *7*, 3373–3381.
90. T. Stein; J. Vater; V. Kruft; A. Otto; B. Wittmann-Liebold; P. Franke; M. Panico; R. McDowell; H. R. Morris, *J. Biol. Chem.* **1996**, *271*, 15428–15435.
91. P. C. Dorrestein; J. Blackhall; P. D. Straight; M. A. Fischbach; S. Garneau-Tsodikova; D. J. Edwards; S. McLaughlin; M. Lin; W. H. Gerwick; R. Kolter; C. T. Walsh; N. L. Kelleher, *Biochemistry* **2006**, *45*, 1537–1546.
92. K. J. Weissman; R. Müller, *ChemBioChem* **2008**, *9*, 826–848.
93. N. A. Magarvey; P. D. Fortin; P. M. Thomas; N. L. Kelleher; C. T. Walsh, *Chem. Biol.* **2008**, *3*, 542–554.
94. PAWS is freeware originally developed by ProteoMetrics, LLC and now made available by Genomic Solutions (http://bioinformatics.genomicsolutions.com/).
95. http://prosightptm.scs.uiuc.edu (accessed 1 September 2006).
96. S. Garneau-Tsodikova; P. C. Dorrestein; N. L. Kelleher; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 12600–12601.
97. L. M. Hicks; M. C. Moffitt; L. L. Beer; B. S. Moore; N. L. Kelleher, *Chem. Biol.* **2006**, *1*, 93–102.
98. L. M. Hicks; C. J. Balibar; C. T. Walsh; N. L. Kelleher; N. J. Hillson, *Biophys. J.* **2006**, *91*, 2609–2619.
99. T. Stachelhaus; H. D. Mootz; M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493–505.
100. G. L. Challis; J. Ravel; C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211–224.
101. C. Rausch; T. Weber; O. Kohlbacher; W. Wohlleben; D. H. Huson, *Nucleic Acids Res.* **2005**, *33*, 5799–5808.
102. M. Z. Ansari; J. Sharma; R. S. Gokhale; D. Mohanty, *Bioinformatics* **2008**, *9*, 454.
103. P. Berg; F. H. Bergmann; E. J. Ofengand; M. Dieckmann, *J. Biol. Chem.* **1961**, *236*, 1726–1734.
104. B. F. Pfleger; J. Y. Lee; R. V. Somu; C. C. Aldrich; P. C. Hanna; D. H. Sherman, *Biochemistry* **2007**, *46*, 4147–4157.
105. P. D. Fortin; C. T. Walsh; N. A. Magarvey, *Nature* **2007**, *448*, 824–827.
106. E. R. Strieter; F. H. Vaillancourt; C. T. Walsh, *Biochemistry* **2007**, *46*, 7549–7557.
107. G. J. Gatto, Jr.; M. T. Boyne, II; N. L. Kelleher; C. T. Walsh, *J. Am. Chem. Soc.* **2006**, *128*, 3838–3847.
108. G. L. Tang; Y. Q. Cheng; B. Shen, *J. Biol. Chem.* **2007**, *282*, 20273–20282.
109. M. Wittmann; U. Linne; V. Pohlmann; M. A. Marahiel, *FEBS J.* **2008**, *275*, 5343–5354.
110. J. J. La Clair; T. L. Foley; T. R. Schegg; C. M. Regan; M. D. Burkart, *Chem. Biol.* **2004**, *11*, 195–201.
111. S. C. Wenzel; P. Meiser; T. M. Binz; T. Mahmud; R. Müller, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 2296–2301.
112. S. A. Samel; M. A. Marahiel; L. O. Essen, *Mol. Biosyst.* **2008**, *4*, 387–393.
113. S. M. McLoughlin; N. L. Kelleher, *J. Am. Chem. Soc.* **2005**, *127*, 14984–14985.
114. M. Zerikly; G. L. Challis, *ChemBioChem* **2009**, *10*, 625–633.
115. R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
116. D. Hoffmeister; N. P. Keller, *Nat. Prod. Rep.* **2007**, *24*, 393–416.
117. S. G. Van Lanen; B. Shen, *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 186–195.
118. R. A. Butcher; F. C. Schroeder; M. A. Fischbach; P. D. Straight; R. Kolter; C. T. Walsh; J. Clardy, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1506–1509.
119. J. J. Reddick; S. A. Antolak; G. M. Raner, *Biochem. Biophys. Res. Commun.* **2007**, *358*, 363–367.
120. P. Verdier-Pinard; J. Y. Lai; H. D. Yoo; J. Yu; B. Marquez; D. G. Nagle; M. Nambu; J. D. White; J. R. Falck; W. H. Gerwick; B. W. Day; E. Hamel, *Mol. Pharmacol.* **1998**, *53*, 62–76.
121. H. M. Holden; M. M. Benning; T. Haller; J. A. Gerlt, *Acc. Chem. Res.* **2001**, *34*, 145–157.
122. L. Gu; J. Jia; H. Liu; K. Håkansson; W. H. Gerwick; D. H. Sherman, *J. Am. Chem. Soc.* **2006**, *128*, 9014–9015.
123. T. W. Geders; L. Gu; J. C. Mowers; H. Liu; W. H. Gerwick; K. Håkansson; D. H. Sherman; J. L. Smith, *J. Biol. Chem.* **2007**, *282*, 35954–35963.
124. Y. Sugimoto; T. Otani; S. Oie; K. Wierzba; Y. Yamada, *J. Antibiot.* **1990**, *43*, 417–421.
125. P. C. Dorrestein; H. Zhai; F. W. McLafferty; T. P. Begley, *Chem. Biol.* **2004**, *11*, 1373–1381.
126. A. Hazra; A. Chatterjee; T. P. Begley, *J. Am. Chem. Soc.* **2009**, *131*, 3225–3229.
127. P. C. Dorrestein; H. Huili Zhai; S. V. Taylor; F. W. McLafferty; T. P. Begley, *J. Am. Chem. Soc.* **2004**, *126*, 3091–3096.
128. K. E. Burns; Y. Xiang; C. L. Kinsland; F. W. McLafferty; T. P. Begley, *J. Am. Chem. Soc.* **2005**, *127*, 3682–3683.
129. J. W. Hanes; K. E. Burns; D. G. Hilmey; A. Chatterjee; P. C. Dorrestein; T. P. Begley, *J. Am. Chem. Soc.* **2008**, *130*, 3043–3052.
130. T. Raschle; D. Arigoni; R. Brunisholz; H. Rechsteiner; N. Amrhein; T. B. Fitzpatrick, *J. Biol. Chem.* **2007**, *282*, 6098–6105.
131. H. Gao; J. Leary; K. S. Carroll; C. R. Bertozzi; H. Chen, *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 167–178.

Biographical Sketches

Roland D. Kersten was born in Berlin, Germany, in 1985. He obtained his Diplom in 2008 in biochemistry from the Free University, Berlin. The final 9 months of Diplom thesis, on the characterization of orphan gene clusters from the marine organism *Salinipora tropica*, was supervised by Dr. Dorrestein at the University of California, San Diego. Following the completion of his Diplom, Roland continued to answer natural product biosynthetic questions by modern mass spectrometry. Roland is currently working in the Dorrestein laboratory as a research associate. In the summer of 2009, he will be joining the graduate program at the Scripps Institution of Oceanography, San Diego, where he will continue to implement mass spectrometry as a genome mining tool for natural product discovery under the dual mentorship of Dr. Bradley Moore and Dr. Pieter C. Dorrestein.

Michael J. Meehan was born in Downer's Grove, Illinois. He received his bachelor's degree in biology from the University of San Francisco (USF) in 2001. During his studies at USF, Michael was primarily interested in immunology, and he also worked for Dr. Eugene V. Benton on a joint project between the USF Department of Physics and the Loma Linda University Cancer Institute, studying the interaction of high-energy proton beams with tissue-equivalent materials by atomic force microscopy. Michael entered the master's program in chemistry at UCSD in the fall of 2007 and joined the laboratory of Dr. Pieter Dorrestein in early 2008. During his time in the Dorrestein laboratory, Michael's research focused on mapping of active sites of polyketide synthases and detection of intermediates on these enzymes using high-resolution mass spectrometry. Michael will complete his master's thesis by the summer of 2009.

Dr. Pieter C. Dorrestein was born in Utrecht, the Netherlands, in 1974. He received his undergraduate chemistry training at the Northern Arizona University in 1998 under the tutelage of Dr. John MacDonald. Pieter remained in John's laboratory as a masters student for an additional year. In 1999, Pieter moved to the Department of Chemistry and Chemical Biology at Cornell University, where he received his Ph.D. in 2004. His Ph.D., on the biosynthetic investigations of thiamin, was supervised by Dr. Tadgh Begley. Following his Ph.D., Pieter, as an NIH NRSA Kirchstein fellow, combined his interests in mass spectrometry and natural products. During this time Pieter used high-resolution mass spectrometry methods to investigate the biosynthesis of therapeutic agents under the supervision and cosponsorship of Dr. Neil Kelleher (University of Illinois) and Dr. Christopher T. Walsh (Harvard Medical School). In September 2006, Pieter joined the UCSD Skaggs School of Pharmacy and Pharmaceutical Sciences and the Departments of Pharmacology, Chemistry, and Biochemistry. His current research aims to develop new mass spectrometry approaches to detect and characterize natural products as well as their biosynthesis. In addition, his laboratory aims to characterize the function of posttranslational modifications involved in disease processes. Since his arrival in UCSD, Pieter has received several honors, including the Beckman Young Investigator Award, the V-Foundation Scholar Award, the PhRMA Foundation Award, the Eli Lilly Young Investigator Award in Analytical Chemistry, and enjoys funding from the National Institute of Health.