Determining the Potential of Lanthipeptides in Drug Discovery: A Two-Pronged Approach

Hassan Abdirizak Issak

Department of Chemistry

McGill University

Montreal, Quebec, Canada

H3A 0B8

August 2019

A thesis submitted to McGill University in partial fulfillment of the Requirements of the degree of Master of Science

Abstract:

In light of the antibiotic resistance crisis, a push for the discovery and development of new antibiotics is more important than ever, and the natural world provides a treasure trove of potential treatments. Lanthipeptides represent a class of ribosomally synthesized post-translationally modified peptide natural products (RiPPs) that typically display antimicrobial activity. Their popularity has been on the rise in recent years due to their manipulability and unique modes of action that make it difficult for bacteria to develop resistance. They are characterized by the presence of thioether rings (termed lanthionine rings) that are formed by the intramolecular nucleophilic attack of cysteine residues on nearby dehydrated serine (dehydroalanine) or threonine (dehydrobutyrine) residues. The lanthionine rings are usually critical for antibiotic activity and, in the class II lanthipeptides, are installed in the lanthipeptide precursor (LanA) by a bi-domain protein called a class II lanthipeptide synthetase (LanM). In this study, a two-pronged approach, centered on lanthipeptides, was employed to aid in the discovery and development of new antibiotics. In the first approach, genome mining was used to discover a novel two-component lanthipeptide biosynthetic gene cluster (herein termed the *sael* cluster) in the actinomycete *Micromonospora saelicesensis*. Using *in vivo* co-expression experiments in *E. coli*, we demonstrate that each LanM enzyme in the cluster (SaelM1 and SaelM2) specifically modifies one of the two LanA peptides (SaelA1 and SaelA2, respectively). Moreover, we show that co-expression of the SaelM/SaelA paired with the transport enzyme, SaelT, results in the export of post-translationally modified peptides from the cell. Preliminary mass spectrometry-based structural studies suggest that these peptides, which we term saelicesin α (Saelα) and saelicesin β (Saelβ), are highly cyclized and likely contain novel lanthipeptide thioether ring topologies. In the second approach, the structural properties of the LanM, HalM2, in relation to its function was selected as a means of providing information for future attempts at protein engineering. Studies have suggested a role for enzyme conformation sampling in the installation of post-translational modifications in RiPPs, however detailed studies are not yet available. The development and use of a hydrogen-deuterium exchange mass spectrometry (HDX-MS) based approach to probe the biophysical properties of HalM2 is reported herein. Novel precursor peptide binding regions, as well as apparent long range structural communication triggered by ligand binding within HalM2 were identified. It is hoped that the insights reported will significantly aid any attempt at the rational engineering of LanMs, and highlight the utility of HDX-MS as a tool in structural biology.

Resume:

La crise actuelle de la résistance aux antibiotiques justifie plus que jamais la nécessité de développer de nouveaux antibiotiques. Les lanthipeptides sont une classe de produits naturels peptidiques provenant de modifications post-traductionnelles ribosomiques (RiPPs) et comportent typiquement une activité antimicrobienne. Depuis les dernières années, leur popularité est grandissante dû à leur versatilité et à leur mode d'action unique leur permettant de ralentir la résistance microbienne aux antibiotiques. Ces produits naturels possèdent un thioether cyclique (appelé lanthionine cyclique) qui se forme lors de l'attaque nucléophile intramoléculaire d'un résidu de la cystéine sur une serine ou une thréonine déshydratée. Les lanthionines cycliques sont habituellement critiques à l'activité antibiotique et, dans le cas des lanthipeptides de classe II, sont installés sur le précurseur au lanthipeptide (LanA) par des lanthipeptide synthase de classe II (LanM). Dans cette étude, une approche à deux volets a été employée pour contribuer au développement de nouveaux antibiotiques. Premièrement, l'analyse de données génomiques a permis de découvrir un nouveau cluster de gènes responsables de la biosynthèse de lanthipeptides provenant des actinomycètes Micromonospora saelicesensis. En co-exprimant in vivo dans E. coli, nous avons démontré que chacune des enzymes LanM du cluster (SaelM1, SaelM2) modifie spécifiquement un des deux peptide LanA (SaelA1 et SaelA2, respectivement). De plus, nous avons montré que la co-expression de SaelM/SaelA en présence de l'enzyme de transport, SaelT, permet l'exportation des RiPPs hors de la cellule. Des résultats préliminaires ont suggéré que ces peptides, appelés saelicesin α (Saelα) et saelicesin β (Saelβ), contiennent un haut niveau de cyclisation et probablement de nouvelles topologies de thioether cyclique. Dans le second volet, les particularités structurelles de LanM, HalM2, en relation avec leur fonction, ont été investiguées afin de déterminer de potentielles mutations génétiques de l'enzyme. Des études ont suggéré le rôle que jouerait les conformations qu'adopteraient l'enzyme dans l'installation de modifications post-traductionnelles des RiPPs. Le développement d'une technique d'échange de deutérium-hydrogène couplée à la spectroscopie de masse (HDX-MS) est décrite dans ce travail. Un nouveau site de liaison pour le précurseur peptidique ainsi qu'un mode de communication allostérique de HalM2 ont été identifiés. Nous espérons que cette étude contribuera à l'ingénierie rationnelle des LanMs tout en mettant l'emphase sur l'utilité de la technique HDX-MS comme outil d'analyse de la biologie structurelle.

Acknowledgements:

I would firstly like to thank my research supervisor Dr. Christopher Thibodeaux, who was willing to take a chance on me and take me under his tutelage. He showed me the wider world of biochemistry and it would not be an exaggeration to say I look to him as a model for what a true researcher should be. There were times were I lost sight of my goals, times were my motivation would wane, but he made sure I kept my head in the game and helped mold me into a researcher. I would also like to thank him for introducing me to seafood gumbo. I am shocked that it took me this long to discover something so delicious. It is for all these things, and much more, that I am truly grateful for. Next, I would like to thank a man whom I consider my brother and one of my closest friends, Kevin Uggowitzer, who stood by my side through thick and thin, through all those late nights and early mornings, day after day until we saw our work through. I would like to thank him for introducing me to hockey and helping find my inner Canadian. I would like to thank him for consoling me every time the Maple Leafs lost (a regular occurrence), and celebrating with me every time the Bruins lost too. I am going to miss the fun we had working together, and I wish you the best of luck in your PhD and all your future endeavors. I would like to thank Yeganeh Habibi for helping me process and analyze all my HDX data, and for feeding me delicious Persian food on those really long and busy days. I would also like to thank Sally Hamry for continuing my work with the saeleicesins project. I would like to recognize the many undergraduate students who helped build our lab bit by bit. Firstly, I would like to thank Lucía Ruiz Haddad for her help with characterizing the saelicesins, and for teaching me everything I know about astrology and Mexican culture. I would also like to thank Sarah Barnier for helping purify the saelicesins. Hopefully, spending hours filtering media has taught her the value of patience. I would also like to thank Corum Holtz, William Patterson, Ademola Esomojumi, Judy Huang, Lauren Tyler, Nicholas Grinberg, Sara Ghandoor, Zhiyang Liu and Marianne Dufresne, who all played their own part in making our lab what it is today.

I would like to thank NSERC, FRQNT, CFI and McGill University for funding our lab and research. I would like to thank the Auclair lab for sharing equipment, reagents and lab space while our lab was still being set up. I would also like to thank the Lennox, Damha and Tsantrizos labs for sharing their equipment. And finally, I would like to thank Julie Ducharme for the French translation of the abstract.

Table of Contents:

List of Figures:

List of Tables*:*

List of Equations:

Abbreviations:

°	Degree
m	Micro
Ala (A)	Alanine
AMP	Adenosine 5'-monophosphate
AMP-PNP	β,γ-Imidoadenosine 5′-triphosphate
ADP	Adenosine 5'-diphosphate
ATP	Adenosine 5'-triphosphate
Arg (R)	Argenine
Asn (Q)	Asparagine
Asp (D)	Aspartate
BLAST	Basic local alignment search tool
C	Celsius
CFI	Canadian Foundation for Innovation
CID	Collision induced dissociation
Cu	Copper
Cys (C)	Cysteine
Da	Daltons
Dha	Dehydroalanine
Dhb	Dehydrobutyrine
Dhx	Dehydrated residue
DNA	Deoxyribonucleic acid
ESI	Electrospray ionization
EtOH	Ethanol
FA	Formic acid
Fe	Iron
FP	Forward primer
FPLC	Fast protein liquid chromatography
GA	Glycine-Alanine
GG	Glycine-Glycine
Glu (E)	Glutamate
Gln (Q)	Glutamine
Gly (G)	Glycine
HDX	Hydrogen-deuterium exchange
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His (H)	Histidine
HPLC	High pressure liquid chromatography
HSEE	Hypothetical structure evaluation and enumeration
Ile (I)	Isoleucine
IMAC	Immobilized metal affinity chromatography
k	Kilo
K	Potassium
KA	Kinase activation
L	Litre
Lan	Lanthionine
LanA	Precursor peptide
LanB	Class I lanthipeptide dehydratase
LanC	Class I lanthipeptide cyclase
LanKC	Class III lanthipeptide synthetase
LanL	Class IV lanthipeptide synthetase
LanM	Class II lanthipeptide synthetase
LAP	Linear azole containing peptide
LB	Luria broth

Leu (L)	Leucine
Lys (K)	Lysine
MALDI	Matrix assisted laser desorption ionization
MeLan	Methyllanthionine
Met (M)	Methionine
Mg	Magnesium
mL	Millilitre
mRNA	Messenger RNA
MRSA	Methicillin resistant *Staphylococcus aureus*
MS	Mass spectrometry
MTB	*Mycobacterium tuberculosis*
MurNAC	*N*-Acetylmuramic acid
MW	Molecular weight
NAG	*N*-Acetylglucosamine
NEM	*N*-ethylmaleimide
Ni	Nickel
Ni-NTA	Nickel-nitrilotriacetic acid
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
NSERC	Natural Sciences and Engineering Research Council of Canada
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
Phe (F)	Phenylalanine
Pro (P)	Proline
RiPPs	Ribosomally synthesized post-translationally modified peptides
RNA	Ribonucleic acid
RO5	Rule of five
RP	Reverse primer
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser (S)	Serine
SPE	Solid phase extraction
TB	Terrific broth or Tuberculosis
TCEP	Tris(2-carboxyethyl) phosphine
TFA	Trifluoroacetic acid
Thr (T)	Threonine
TIC	Total ion chromatogram
TOF	Time of flight
tRNA	Transfer RNA
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
V	Volts
Val (V)	Valine
WT	Wild type
XIC	Extracted ion chromatogram
Zn	Zinc

Contribution of Authors

This thesis is written as two chapters. All work done in chapter 2 was conducted solely by the author. For chapter 3, some data processing was carried out by Yeganeh Habibi.

Chapter 1: Background

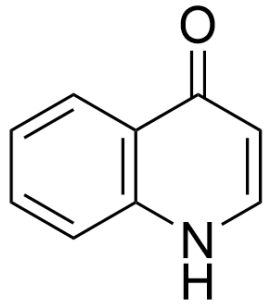
1.1.History of Antibiotics Discovery:

The discovery of penicillin, and thus the concept of antibiotics, in 1928 by Alexander Fleming ushered in a new era in medical research. The idea of medication to help stave off multiple diseases was particularly enticing during World War II, where allied troops were exposed to various pathogens in foreign lands. This prompted a push towards mass production of the drug for on-field treatments, which was achieved in 1943. The Golden Age of antibiotics, as it is sometimes referred to, led to the discovery of several drugs- some estimates point towards half of all antibiotics- between 1950 and 1960. Some had come to believe that the threat of bacterial infection had been rendered moot. Dr. Robert Petersdorf, a renowned expert on infectious diseases, in 1978 stated that "Even with my great personal loyalty to Infectious Disease, I cannot conceive of the need for 309 more [graduating trainees in] infectious disease...unless they spend their time culturing each other". He repeated that sentiment in 1985, commenting that "the millennium where fellows in infectious disease will culture one another is almost here"[1].

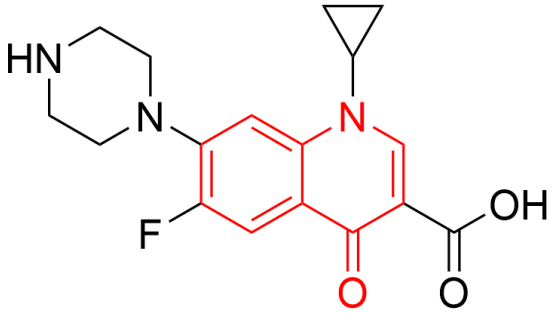
However, many scientists warned of the danger of antimicrobial resistance. In fact, Fleming himself noticed the development of resistant bacteria in early experiments, stating that "the microbes are educated to resist penicillin and a host of penicillin-fast organisms is bred out…"[2]. Methicillin, an antibiotic in the same class as penicillin, was marketed as a treatment for infection by bacteria of the *Staphylococcus* and *Streptococcus* genera, such as *Staphylococcus aureus* and *Streptococcus pneumoniae*. Within a mere two years, resistance to methicillin had become widespread, rendering the drug ineffective[3]. Vancomycin, first isolated in 1953, was used as a treatment for various infections caused by Gram-positive bacteria, such as methicillin resistant *Staphylococcus aureus* (MRSA) and *Enterococci*[4]. Unfortunately, not only was the drug associated with several undesirable side effects, including hearing loss and potentially death, vancomycin resistant bacteria (eg: vancomycin resistant *Enterococcus*) emerged in 1987[4]. Just six years later, a strain of vancomycin dependent *Enterococcus* (ie: vancomycin is required for growth) was isolated from a patient receiving long term vancomycin therapy. In one extreme case, a nosocomial outbreak of multidrug resistant, vancomycin dependent strains of *Enterococcus faecalis* was recorded in a bone marrow transplant ward at The Johns Hopkins Hospital in 1998.

1.2. Antimicrobial targets and resistance:

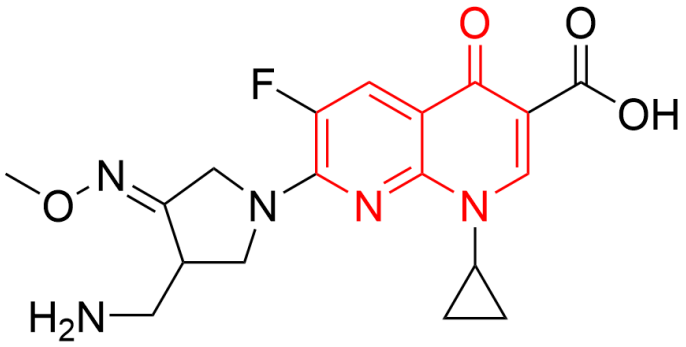
1.2.1. Inhibition of DNA replication: Targeting DNA replication mechanisms typically induces bacteriostasis, wherein cell growth and replication are inhibited. A prominent example of drugs that target DNA replication machinery are the quinolones (Fig.1.1). Quinolone antibiotics are characterized by having 4-quinolone related structures as the core of the compound. They are broad spectrum antibiotics, used to treat various diseases such as tuberculosis and infections caused by Gram-positive and Gram-negative bacteria.



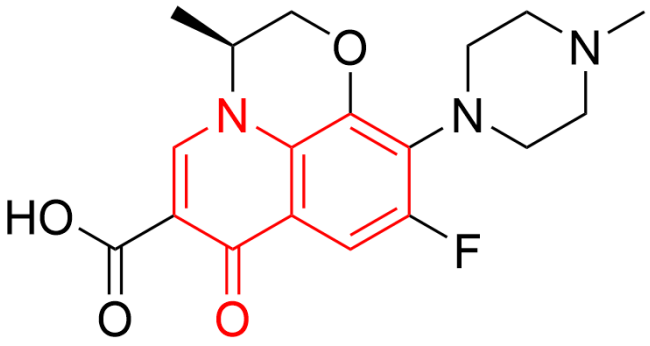
4-quinolone



Ciprofloxacin



Gemifloxacin

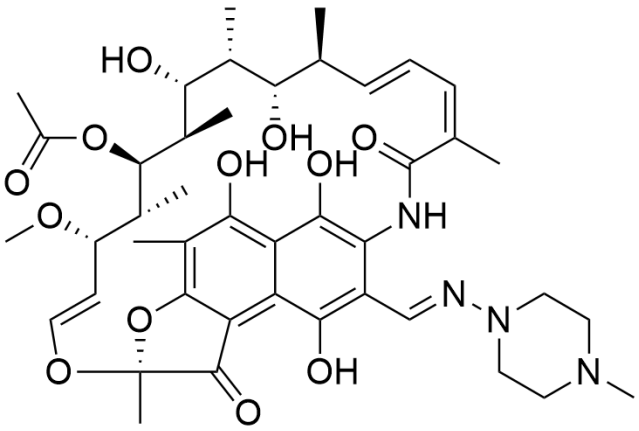


Levofloxacin

Figure 1.1. The structures of select quinolone antibiotics, with the 4-quinolone motif highlighted.

Quinolones function by binding to DNA topoisomerases (II and IV) during DNA replication, trapping them in a complex with DNA at replication forks. This leads to breaks in DNA strands that are consequently lethal to the cell[5]. The simplest and most prominent mechanism is mutation of genes encoding for either topoisomerase II or IV, leading to significantly reduced affinity of quinolones to their target. This can occur during the cell's SOS response, which is a part of the DNA repair system that causes genes associated with DNA repair to be induced and wherein mutant topoisomerases may emerge.

1.2.2. **Inhibition of RNA synthesis**: Several metabolic pathways can be blocked altogether by inhibiting RNA synthesis in bacteria, often leading to cell death. Rifamycins represent a class of RNA synthesis inhibiting antibiotics characterized by the presence of an aromatic ring with an aliphatic chain within the structure (Fig. 1.2). They were first isolated by culturing *Amycolatopsis mediterranei* in 1957 and were found to be effective at treating infections caused by difficult-to-target bacteria such Mycobacteria[6]. Rifamycins function by binding to the β-subunit of RNA polymerase, thus inhibiting transcription, and further downstream, translation and protein/peptide synthesis[7]. Mutations to the β-subunit of RNA polymerase, which occur readily, introduce resistance. As such, rifamycins are usually prescribed as part of combinatorial treatments.



Rifampicin

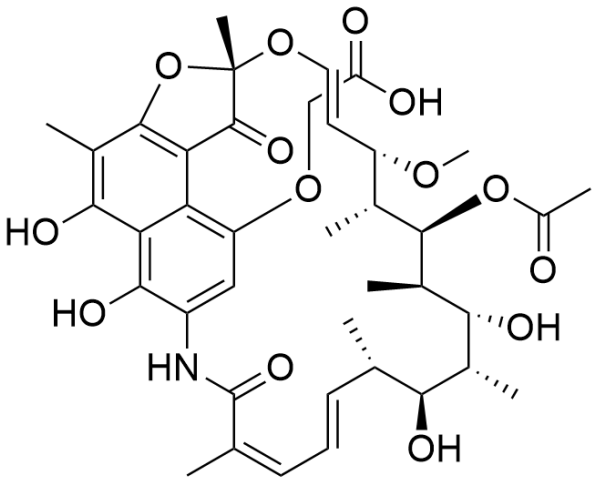


Figure 1.2. Structure of Rifampicin and Rifamycin B

1.2.3. **Inhibition of cell wall synthesis**: Bacterial cell walls are vital in maintaining the cell's integrity under various conditions. As such, disrupting cell wall synthesis can severely reduce cell fitness, and can ultimately lead to cell death. To that end, some of the most prominent antibiotics discovered were found, or designed, to target pathways in cell wall synthesis. Two classes of note for cell wall synthesis inhibitors are the β-lactams and the glycopeptides.

Peptidoglycan is a carbohydrate and amino acid based polymer that constitutes the cell wall of bacteria (Fig. 1.3). It consists of alternating units of *N*-acetylmuramic acid (NAM) and *N*-acetylglucosamine (NAG). Short-chain peptides attached to NAM units allow crosslinking between the glycan strands, forming the mesh that comprises the cell wall[8].

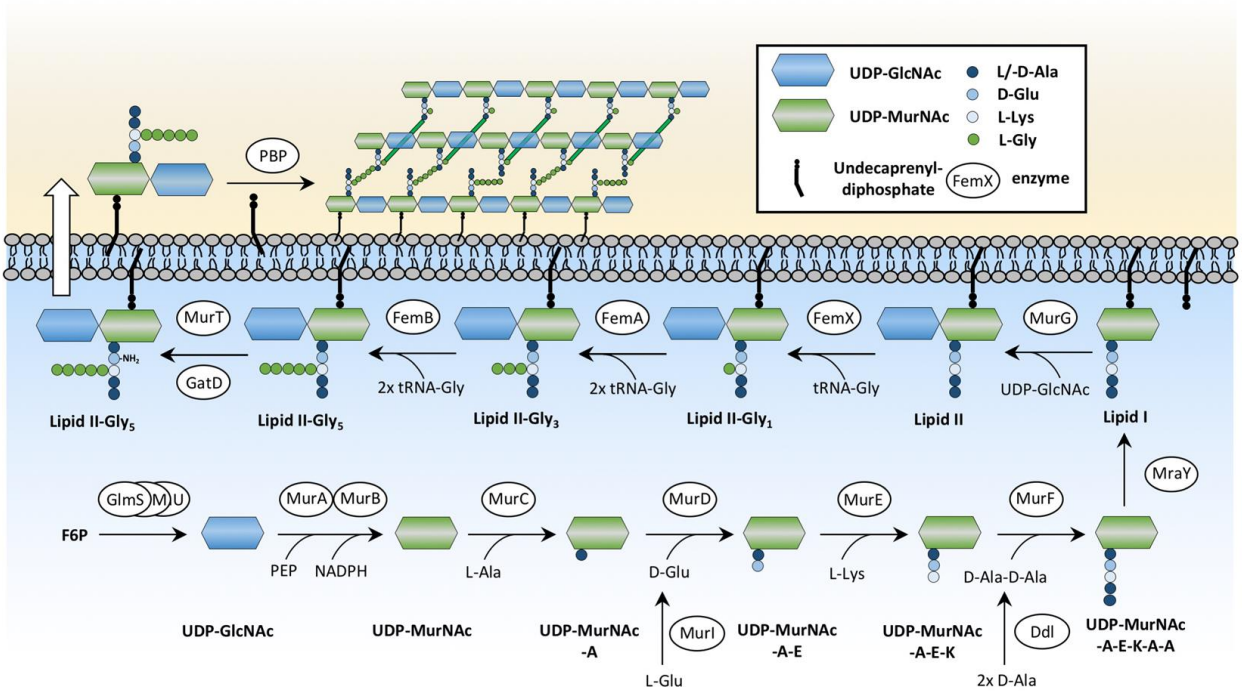


Figure 1.3. Structure of the peptidoglycan layer, highlighting the steps involved in its synthesis. Adapted with permission. [9] *Licensed under CC BY 4.0.*

In the cell wall synthesis pathway, there are two targets of note: D-Alanyl-D-Alanine subunits (D-Ala-D-Ala) and the Penicillin Binding Proteins (PBP). The D-Ala-D-Ala subunit of the peptide chains attached to the NAM units of that glycan strands are involved in the formation of peptide crosslinks mentioned earlier. In brief, the terminal D-Ala residue is eliminated via a PBP, forming a complex between the PBP and the remaining D-Ala residue. PBP then facilitates crosslinking from the amine of nearby lysine residues on the D-Ala residue[8].

β-Lactam antibiotics are broad spectrum antibiotics characterized by a β-lactam ring core structure (Fig. 1.4). The most famous drug within this class is arguably penicillin, the compound whose discovery spurred the hunt for new antibiotics. The mechanism of action for β-lactams is simple: They act as analogues for the D-Ala-D-Ala subunits and so bind to PBPs, inhibiting the formation of the peptide crosslinks[8].

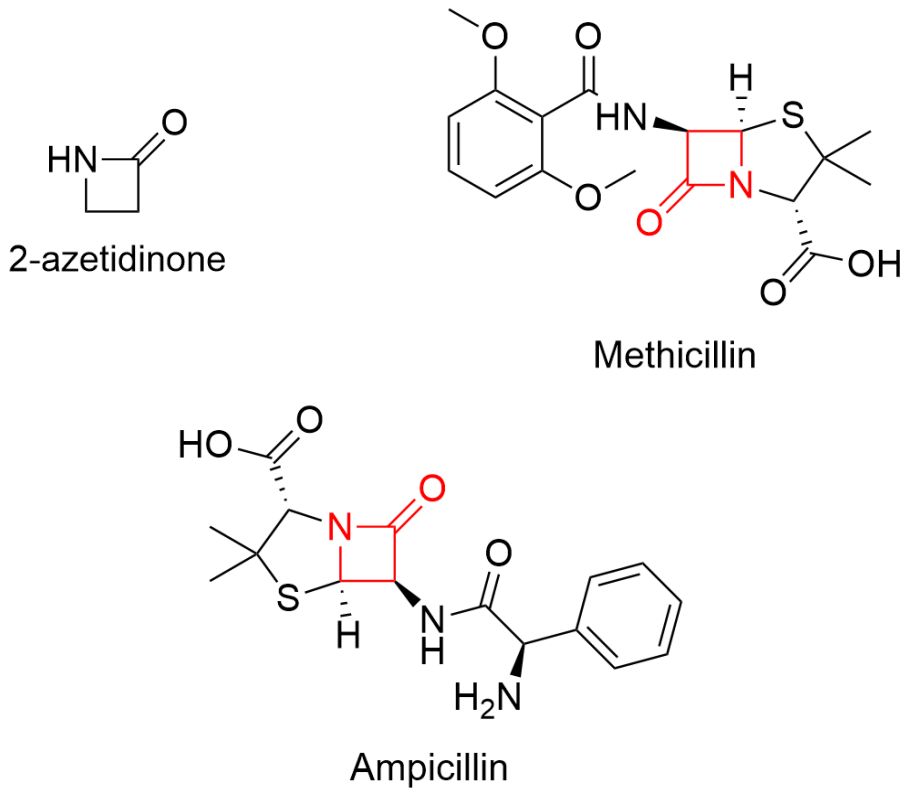
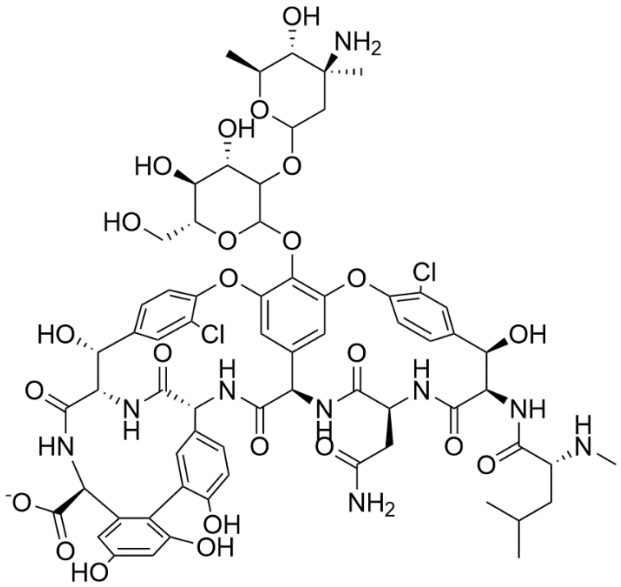


Figure 1.4 Structures of (i) the simplest possible B-lactam, (ii) methicillin and (iii) ampicillin, with the b-lactam ring highlighted

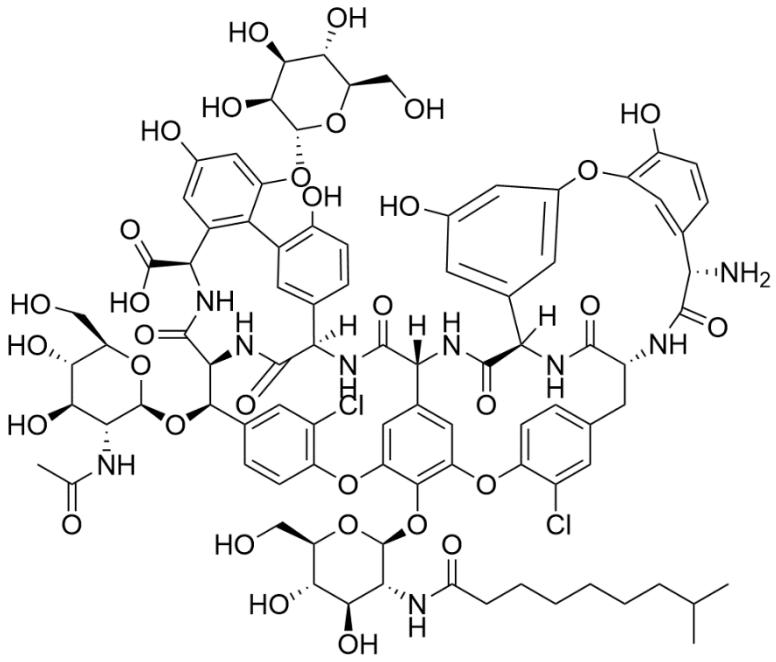
There are three main mechanisms through which bacteria develop resistance to β-Lactams: (i) Reduced affinity of PBPs for β-Lactams, (ii) Expression of β-Lactamases and (iii) Reduced access to PBPs (Gram-negative bacteria only). For case (i), mutations may arise within during treatment that may alter active site binding affinity of certain β-lactams in PBPs, thus eliminating the threat of competitive inhibition. In case (ii), β-lactamases, which are enzymes that hydrolyse the β-lactam rings, are secreted by the bacteria into their environment. This is a common mechanism of resistance in Gram-negative bacteria, however there is some evidence of Gram-positive bacteria producing β-lactamases. Finally, for case (iii), the outer membrane of Gram-negative bacteria become less permeable to β-lactam antibiotics, and efflux across the outer membrane excretes β-lactams that may have crossed into the periplasm.

Glycopeptides, similar to β-lactams, are broad spectrum antibiotics. They are peptides that have carbohydrates covalently linked to the side chains of amino acid residues. Prominent members of this class of antibiotics include vancomycin and teicoplanin (Fig. 1.5). Within the cell wall synthesis pathway, glycopeptides inhibit synthesis by binding to the D-Ala-D-Ala subunit of

the short chain peptides, preventing binding and modification by PBPs and other proteins in the pathway. Glycopeptide resistance is developed when the terminal D-Ala is mutated to a D-Lactate, thus preventing glycopeptides from binding and preventing crosslink formation.



Vancomycin



Teicoplanin

Figure 1.5. The structures of vancomycin and teicoplanin.

1.3. Platforms for drug discovery:

Over the years, bacterial resistance has become a growing concern, and a stall in the discovery of new and effective antibiotics to keep up with resistance rates is ever more crucial. To that end, several approaches to drug discovery have been developed and implemented. Herein, a few more prominent platforms for drug discovery are discussed.

1.3.1. **Species-specific antibiotic identification:** In a species-specific platform, a single strain of (or strains of closely related) bacteria are used to identify novel antibiotics. A disease of particular interest in the early days of antibiotic discovery, and one that remains a concern even today, was tuberculosis (TB). TB is a primarily pulmonary, airborne disease that causes over a million deaths annually (1.5 million deaths in 2014, 1.3 million deaths in 2016), according to a report by the World Health Organization[10]. TB infection is caused by *Mycobacterium tuberculosis* (MTB), a slow dividing, and aerobic bacillus. What makes MTB, and mycobacteria as a whole, particularly interesting is the incorporation of high levels of mycolic acid (a lipid) in its cell wall (Fig. 1.6). This characteristic reduces the penetration of hydrophilic compounds, consequently making it a difficult target for antibiotic treatment[11].

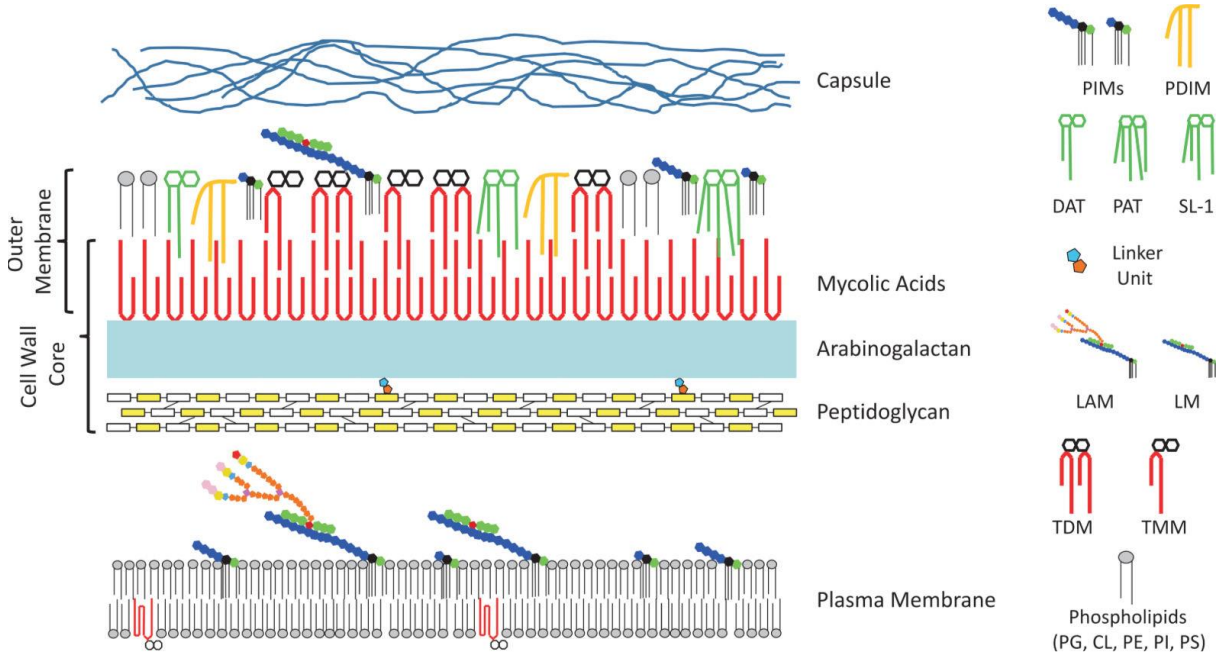


Figure 1.6 The components of mycobacterial cell membrane and cell walls, with an emphasis on the mycolic acid layer. Adapted with permission. [12]

Various compounds, both synthetic and natural, were tested against MTB for activity. Indeed, this led to the discovery of several antibiotics, most notable of which are the compounds isoniazid, pyrazinamide, ethionamide and ethambutol. These four together are used as a standard treatment for TB to this day[11]. A benefit of this approach is that the rate of discovery for active compounds is greater than other methods. This is because developing drugs for a specific target rather than a broad range of targets is simpler, allowing for more successful implementation of high throughput screening. Another benefit of this approach is that, due to the specific activity of these compounds, they are unlikely to be harmful to humans, neither targeting host cells nor symbiotic bacteria[13]. A major limitation, however, is that the discovery of narrow spectrum drugs, depending on the prevalence and threat of a target pathogen (ie: the demand for treatment), may lead to very modest returns on investment. As such, the species-specific platform for drug discovery has seen relatively limited application in medical research.

1.3.2. **Penetrability**: A more common approach to drug discovery is to design compounds based on their ability to penetrate the cell membrane of their targets. In essence, the ability of a library of compounds to penetrate cell membranes is ranked, and the characteristics of the most penetrating and least penetrating compounds are studied. This empirical approach eventually led to the development of Lipinski's Rule of Five (RO5)[14], which describes ideal characteristics that make a compound 'druglike'. These rules are: (i) no more than 5 hydrogen bond donors, (ii) no more than 10 hydrogen bond acceptors, (iii) a molecular weight of no more than 500 Daltons and (iv) octanol-water partition coefficient (LogP) of less than 5. In theory, a compound that adheres to RO5 should have ideal pharmacokinetics, allowing it to carry out its function. A prominent class of RO5 adhering antibiotics is the fluorquinolones, which includes the commonly administered, broad spectrum drug ciprofloxacin.

In situ screening is effective as well. Administering potential drugs in infected test hosts (such as mice)[13]. A noteworthy example is that of prontosil, first synthesized by Josef Klarer and Fritz Mietzsch, which exhibits broad activity against cocci bacteria, not including *Enterobacteria*. Gerhard Domagk demonstrated that prontosil was effective at reducing the mortality rate of rats infected with various *Streptococcus strains*[13]. However, had an *in vitro* test been carried out, the activity of prontosil might have been missed all together. This is because prontosil is in fact a prodrug, which is metabolized in the intestine of rats to form the active drug (containing a

sulphonamide moiety). An *in situ* screen also filters out compounds that have undesirable traits, such as acute or chronic toxicity and poor bioavailability. However, this method is not ideal for high throughput screening, as testing against millions of infected organisms (primarily murine) is neither practical nor feasible.

1.3.3. **Natural Products as Antibiotics**: An option that has proven to be the most popular approach to drug discovery over the years is to simply turn to the natural world as a source of, or an inspiration for, antibiotics. In fact, Harvey (2007) notes that "natural products have been the single most productive source of leads for the development of drugs"[15]. Between 1994 and 2007, almost half of approved drugs were based on natural products. Under the umbrella of natural products research, several approaches have been taken to discover potential drugs. However, two rapidly emerging approaches are a (i) genomics-based method and a (ii) mutasynthetic approach[15]. The genomics approach involves sampling bacterial, fungal or plant DNA, and scanning it for genes that encode for enzymes that construct potential antibiotics. This approach has grown in popularity of recent years as genome libraries have grown, and is often referred to as genome mining, wherein software is used to scan several bacterial genomes in a relatively short period. In fact, it has already led to the discovery of several families of antibiotics, such as the turbomycins.[15] The benefit of genome mining is the relative ease in which several genomes can be scanned for various potential antibiotics. Moreover, a plethora of organisms that are yet to have their genomes sequenced means that the natural world provides a veritable treasure trove of potential antibiotics that is yet to be tapped. However, growing the host species can prove challenging at times, and finding the appropriate growth conditions that lead to the expression of genes of interest compounds the difficulty. One alternative is to use heterologous expression, wherein DNA from one organism is placed in a new host that is easier to handle, typically *E. coli*. This makes expression considerably more tractable, usually leading to the synthesis of the desired natural product[15].

The mutasynthetic approach is markedly different in that it relies entirely on systems that are already known to produce antibiotics. The main objective of this approach is to take known antibiotics and engineer them to introduce more desirable traits, such as broad-spectrum activity and greater stability. This engineering can take on different forms, from directly mutating primary sequences of peptide antibiotics, such as with the many variants of mutacin, or to mutate the

different components of the synthetic pathway, such as modifying enzymes, to introduce new features. This approach requires significant knowledge of the system *a priori*, however, somewhat limiting the scope of its applicability[15].

Given the diversity of the natural world, a wide range of antibiotic compounds of varying structure and mechanism of action exist. But of the many different families of antibiotics, peptide bacteriocins[16] (antibiotic peptides produced by bacteria) distinguish themselves for two main reasons. Firstly, they often have mechanisms of action that are distinct from most therapeutic agents. Morevover, Cotter et al. (2013) explain that, "given their proteinaceous nature, [bacteriocins] are amenable to gene-based peptide engineering", allowing for improved activity[16]. This leads to the second most appealing characteristic of this class of antibiotics: tractability. Several noteworthy examples of engineered peptides can be found in literature, such as Geng's et al. (2018) study on mutacin 1140, wherein they identified several analogues that displayed greater stability and potency via mutagenesis[17].

Within the overall family of peptide natural products exists two main branches: (i) nonribosomal peptides (NRPs) and (ii) ribosomally synthesized and post-translationally modified peptides (RiPPs). As the name would imply, NRPs are synthesized independent of messenger RNA and ribosomes, instead being synthesized by proteins known as nonribosomal peptide synthetases (NRPSs). NRPs constitute some of the most well-known antibiotics in drug discovery, the foremost of which include the likes of vancomycin and daptomycin. Their activity is not limited to bacteria, however. Actinomycin D, for instance is used as a chemotherapy agent, and ciclosporin is used as an immunosuppressant[18].

RiPPs, unlike NRPs, are ribosomally synthesized peptides that are modified to some extent by enzymes post-translation. While there are no RiPP-based drugs available on the market, the peptides have been drawing considerable attention for various reasons. While NRPs have been proven to be effective antibiotics, resistance to this class of drugs is a significant concern. RiPPs provide a viable alternative, as they often have unique modes of action that most common multi-drug resistant bacteria have not evolved to inhibit. Another benefit of this class that distinguishes it from NRPs is their potential for engineering[19, 20].

A typical RiPP gene cluster encodes for a series of components in a given RiPP synthetic pathway (Fig. 1.7). The peptide synthesized by the ribosome, lacking any modifications, is referred

to as the precursor peptide. These peptides are often divided into two main regions, termed the leader peptide on the *N*-terminus, which is involved in substrate recognition for modifying enzymes, followed by the core peptide on the *C*-terminus, which is where the modifications are installed. A given RiPP cluster can encode for multiple modifying enzymes, depending on the system in question. However, what is common among RiPPs is that the leader peptide is cleaved as the final product is transported out of the cell. This can occur via a single protein that is capable of both proteolytic activity and transport, or via discrete proteins (ie: a protease and a transporter)[19].

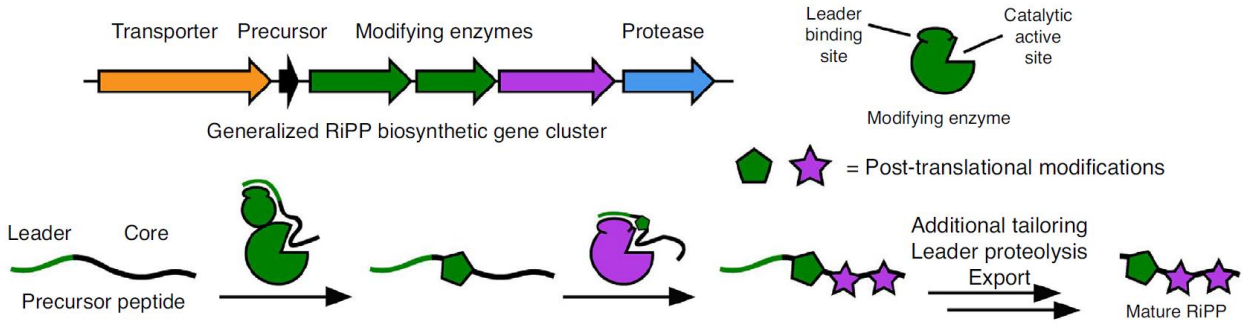


Figure 1.7. The biosynthetic pathway of a RiPPs. Adapted with permission. [21]

RiPPs are convenient to work with in that genes encoding for a RiPP system are often found in small, discrete clusters that (i) provide some cursory information on the identity of the mature RiPP and (ii) are easily manipulated, often by simply using a heterologous host, such as *E. coli*, for ease of expression and for future study[21]. In fact, it is this high degree of manipulability that makes such a promising source of antibiotics. With regard to this, there are various points along the synthetic pathway of a particular RiPP that can be altered to vary the nature of the mature RiPP. For instance, the primary sequence of the precursor peptide can be mutated. This has been shown in several instances to introduce more desirable traits, such as increased solubility and stability[22]. Moreover, since the leader peptide is involved in substrate recognition for modifying enzymes, mutating it would allow non-native modifying enzymes to process a particular RiPP, leading to the synthesis of a new product that may have more potent activity. There are a considerable number of examples in literature illustrating this, such as Burkhart's et al. (2017) study wherein custom leader peptides were designed to allow various modifying enzymes from different pathways to process the core peptide of a RiPP precursor[23]. This approach is made possible due to the fact that modifying enzymes tend to show significant permissiveness with regard to the core peptide[23].

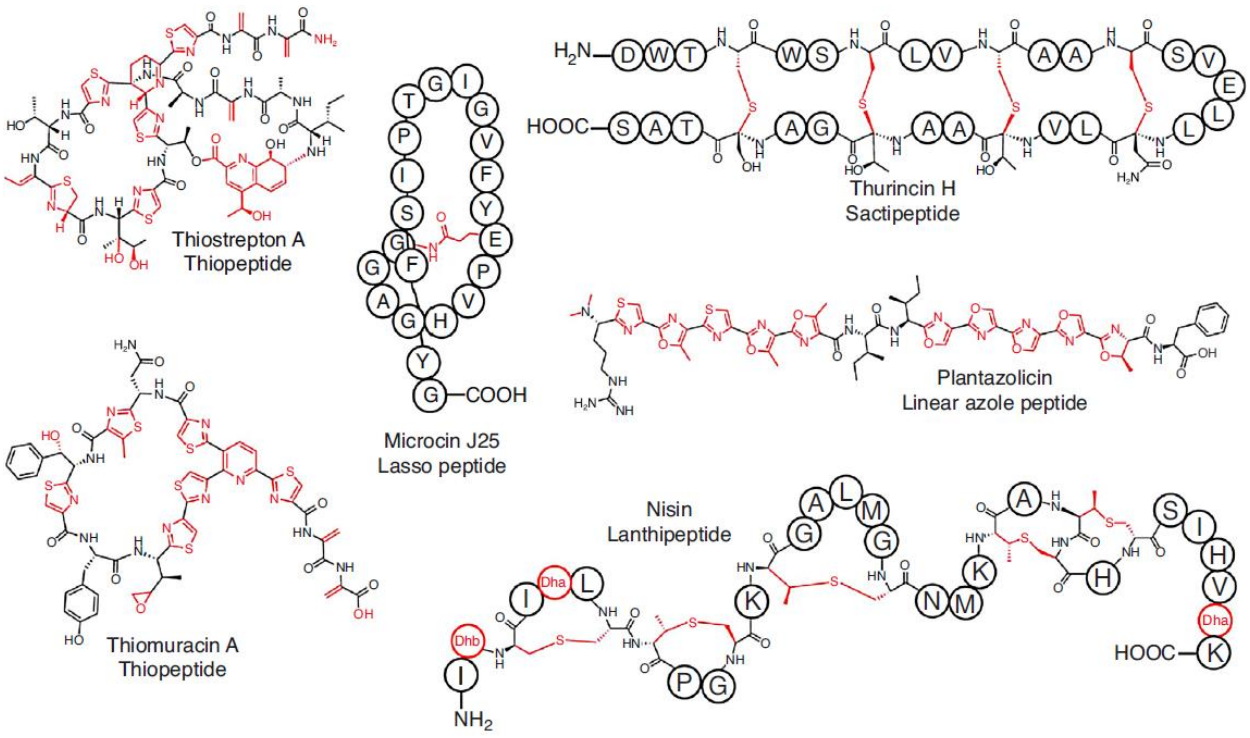


Figure 1.8. Various RiPPs of different classes. Adapted with permission. [21]

Continuing with the theme of diversity in nature, there exist several sub-classes of RiPPs, identified by the modifications introduced to the core peptide. One such class of RiPPs are the lasso peptides, characterized by their unique topology. Lasso peptides consist of an *N*-terminal macrocycle that is formed by linking the *N*-terminal residue of the core peptide to a nearby glutamate or aspartate residue. This is then followed by the *C*-terminal tail that threads through the centre of the macrocycle, and is held in place by disulfide bridges or by steric forces. A consequence of the structure of lasso peptides is that they are rather robust, being resistant to unfolding and difficult for proteases to process[19, 21]. Microcin J25 (Fig. 1.8), which acts as an RNA polymerase inhibitor, represents one of the better known lasso peptides[24]. Linear azole containing peptides (LAP) are another class of RiPPs, the most prominent of which is plantazolicin. Plantazolicin (Fig. 1.8) is a narrow spectrum antibiotic, only exhibiting activity against *Bacillus anthracis*, the pathogen that causes anthrax, and related strains. It is believed to be a DNA gyrase inhibitor[21]. A third class of RiPPs are the thiopeptides, characterized by the presence of azole or azoline heterocycles, dehydroalanine or dehydrobutyrine residues and, in particular, six member *N*-heterocycles. Most thiopeptides act as inhibitors of protein synthesis, such as thiomuracin A (Fig 1.8)[21]. Another class of RiPPs are the sactipeptides, characterized by sulfur to α-carbon

thioether linkages, called sactionine bonds. While there are few characterized examples, those that have been studied show activity against various human pathogens, including *Clostridium difficile* and *Bacillus cereus*[21].

However, the largest, and undoubtedly the most well-studied class of RiPPs are the lanthipeptides. They are characterized by the presence of eponymous lanthionine rings, which are thioether linkages formed between a cysteine residue and the β-carbon of a nearby dehydroalanine or dehydrobutyrine residues (derived from serine and threonine residues respectively). The most intriguing feature of this class of RiPPs is that they often exhibit antibiotic activity. Nisin (Fig. 1.8) is without a doubt the most well characterized example of a lanthipeptide antibiotic, more commonly known as lantibiotics, Discovered in 1928 by Rogers and Whittier, the peptide is produced by a few strains of *Lactococcus lactis* subsp. *lactis* and has long been used as a food preservative, given its broad spectrum activity[25].

Lanthipeptides are divided into several classes depending on the enzymes involved in installing the characteristic modification. However, there are still many features in common between them, primarily related to the modifications themselves. Lanthipeptide precursor peptides, known as LanAs, follow a general pathway towards maturation, beginning with the dehydration of Ser/Thr residues. This step is critical as the dehydrated residues (Dha/Dhb) act as Michael acceptors in a nucleophilic attack by nearby cysteine thiol moieties to produce the characteristic lanthionine (from Dha residues) and methyllanthionine (from Dhb residues)[26].

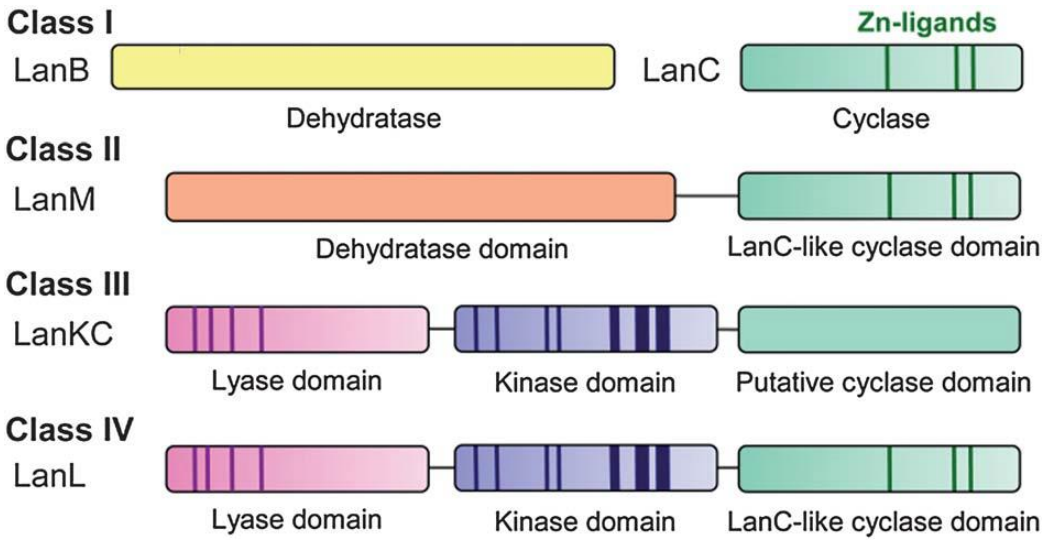


Figure 1.9. The various classes of lanthipeptide synthetases. Adapted with permission. [19]

Lanthipeptide systems are often classified based on the modifying enzymes present in the pathway. There are four main classes of lanthipeptides known to date, termed class I-IV (Fig. 1.9). Class I lanthipeptide systems actually consist of two discrete enzymes, one that dehydrates Ser/Thr residues and another that catalyzes the cyclization step called referred to as LanB and LanC respectively[19, 26]. LanBs recruit glutamyl-tRNA to first glutamylate a Ser/Thr residue, which is then followed by a subsequent elimination step. Of note is the fact that this is the only class of lanthipeptide synthetases to use glutamylation in the dehydration process. LanCs are Zn^{2+} dependent metalloenzymes that catalyze the cyclization step by coordinating cysteine thiols with the Zn^{2+} moiety, lowering the pKa of the γ-S, and increasing its nucleophilicity and propensity towards attacking Dha/Dhb residues[26]. The presence of two enzymes involved in ring installation is unique to class I lanthipeptides, as this process is handled by a single multifunctional protein in all other classes. Moreover, glutamate dependent dehydration is also unique to class I, whereas classes II-IV require ATP.

Class II lanthipeptide systems rely on a single multifunctional protein for ring installation, termed a LanM. This protein carries out its function via the activity of two domains: (i) An *N*-terminal dehydratase domain and (ii) a *C*-terminal cyclase domain[26]. The dehydratase domain of LanMs are known, through studies on LctM activity, to carry out their function via an ATP dependent phosphorylation of the hydroxyl moiety of Ser/Thr residues, followed by the elimination of the phosphate group to produce Dha/Dhb residues. Studies on the cyclase domain

of several LanMs show that they have significant sequence homology to LanCs from class I lanthipeptide systems, and are in fact also Zn^{2+} dependent[26]. The cyclase domain of CylM, a LanM produced by strains of *Enterococcus faecalis*, was shown to have significant structural homology to LanCs such as NisC, which produces nisin, as well[26, 27]. However, CylM remains the only LanM with a resolved crystal structure, making it difficult to determine the structural homology of other LanM cyclase domains to LanCs.

Class III and class IV lanthipeptide systems rely on tri-domain proteins, LanKC and LanL respectively, for ring installation. They each have a kinase domain which phosphorylates Ser/Thr residues, a lyase domain that facilitates elimination of phosphate to produce Dha/Dhb residues, and a cyclase domain that facilitates ring formation. The distinguishing feature between these two classes is that LanKCs are Zn^{2+} independent, meaning cyclizations occur via a different mechanism from LanCs and LanMs. Recently, Hegemann et al. were able to identify 5 residues within the kinase domain of SgbL, a LanL,that were critical for dehydration of Ser/thr residues (Lys248, Glu270, Asp351, Asn356, and Asp36). These residues were also found to be completely conserved in known LanKCs and LanL.[28] Moreover, the kinase domain in LanLs has previously been shown to bind the leader peptide, a ~100 residue long portion of the kinase domain of SgbL (residues 208-318) was identified as the binding site of the leader peptide.[29] Beyond this, LanKCs and LanLs remain relatively unstudied[26].

1.4.Thesis Statement:

The purpose of this thesis is to aid in the search and development of new antibiotics. This will be done via two-pronged approach. Firstly, genome mining will be employed to search for uncharacterized gene clusters that possibly encode for new lanthipeptide systems. In this study we used the LanM enzyme, HalM2, as a reference to locate a putative two-component lanthipeptide gene cluster in the actinomycete *Micromonospora saelicesencis*. The peptides were then characterized and their antibiotic activity tested. The second approach relates to the engineering of LanMs to develop novel lantibiotics. However, an important step in this process is to determine all functionally relevant residues, including binding sites, and to identify how the conformational dynamics of the system in question relates to activity. Our approach to this will be to employ the technique of hydrogen-deuterium exchange- mass spectrometry (HDX-MS) in order to probe the features of HalM2, involved in the biosynthesis of the two-component lantibiotic haloduracin. By

undertaking this multi-pronged approach, it is hoped that the information found here can help inform future experiments on the various aspects of lanthipeptide biosynthesis and engineering.

Chapter 2: Characterization of a Novel Two-Component Lanthipeptide System in *Micromonospora saelicesensis*

2.1. Introduction

Among lantibiotics, a common feature is that the peptides in question are capable of exerting their activity individually. Primary examples include the lantibiotics nisin and mersacidin. Although not as common, two-component systems have been reported wherein the peptides act synergistically to carry out their function. The first identified two-component lanthipeptide system were the cytolysins CylL and CylS (Fig. 2.1). Isolated from *Enterococcus faecium* in 1996, the peptides were identified as being lantibiotics wherein rings are installed on both precursor peptides by a single cognate LanM named CylM. Of particular interest is that the peptides show lytic activity against eukaryotic cells (erythrocytes), showing that two component lanthipeptide systems may have targets other than bacteria[30].

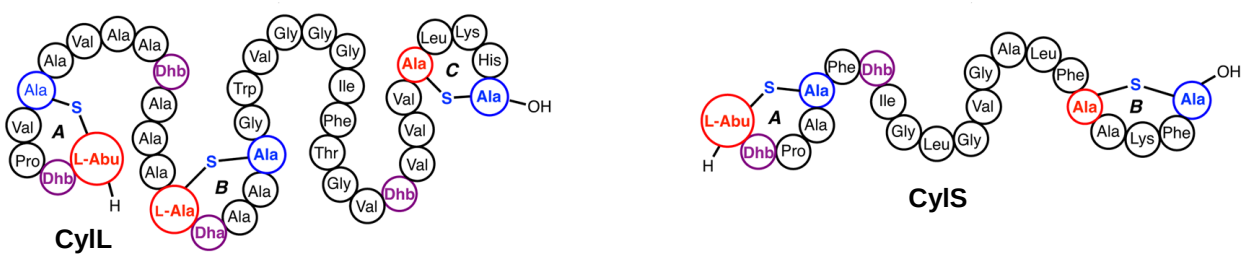


Figure 2.1. The amino acid sequences of the cytolysins CylL and CylS, with the thioether lnkages displayed. Adapted from Repka et al. Adapted with permission. [26]

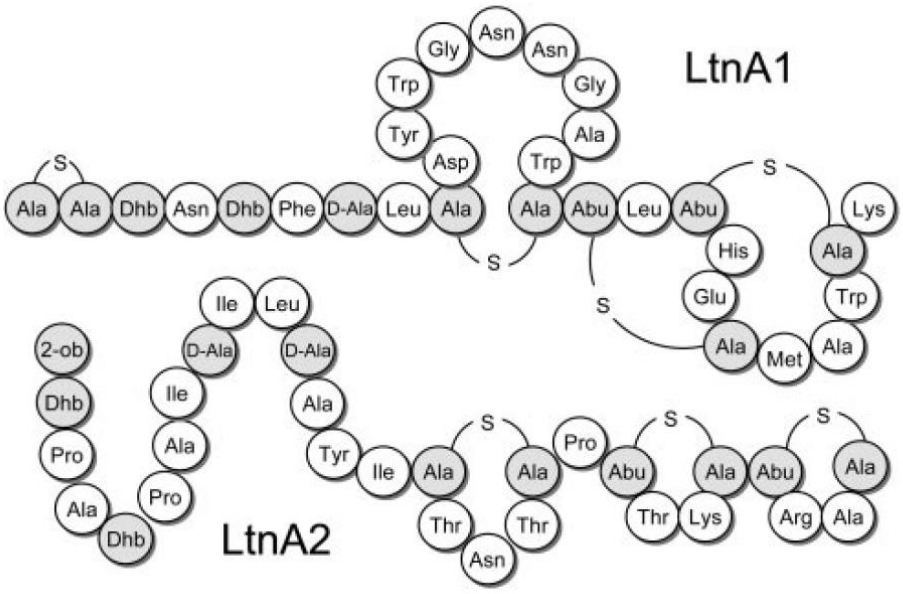


Figure 2.2 The amino acid sequences of LctA1 and LctA2, which constitute the lantibiotic lacticin 3147. Adapted with permission. [31]

Within the two-component lanthipeptide family, lacticin 3147 (Fig. 2.2) represents the most well studied member. First isolated from *Lactococcus lactis* subsp. *lactis* DPC3147 in 1996, McAuliffe et al. were able to show that lacticin 3147 is a broad-spectrum antibiotic and requires both components to exert any activity. McAuliffe et al. were also able to identify the mechanism of action, which involved formation of pores in the bacterial cell membrane that leads to a drop in intracellular K^+ levels. This was linked to an observed decrease in phosphate levels, and a consequent decrease in ATP production, leading to bacteriostasis and eventually cell death. The mechanism of action here is not dissimilar to that of nisin, which was shown to form pores that lead to rapid efflux of cellular contents. Similar to nisin, activity with lacticin 3147 is seen with concentrations in the nanomolar range. Further studies of lacticin 3147 showed that the α component is involved in binding lipid II and that β component binding to the α:lipid II complex then leads to pore formation[32].

Given the wealth of bacterial genomes available in public databases, genome mining has become a reasonable approach for the discovery of new lantibiotics. Genome mining has emerged as a powerful tool in the hunt for new antibiotics. In brief, it relies on the principle that biosynthetic machineries for natural products are often highly conserved within in each class of compound[33]. That is, certain features of each component of a particular gene cluster, such as portions of the

amino acid sequence of an encoded protein/peptide, are shared between related proteins/peptides. For instance, in the case of LanCs, the Zn ligands are expected to be present in a particular region of the protein with roughly defined spacing within the amino acid sequence. Features such as these are found *in silico* through sequence alignments, wherein several known proteins/peptides of a particular function are aligned to find residues that are conserved between them, and roughly where in the amino acid sequence these residues appear. This general process is at the core of most genome mining tools, such as antiSMASH (antibiotics and Secondary Metabolite Analysis Shell) and the NCBI basic local alignment search (BLAST) tool. AntiSmash[34, 35] searches sequenced bacterial genomes for clusters of biosynthetic genes that are known to be involved in the synthesis of specific types of natural products. The BLAST tool can be used to easily find homologues of a well-studied protein, permitting construction of detailed sequence alignments that enable more accurate functional prediction.

In this regard, the two component lanthipeptide, haloduracin (Fig. 2.3), from *Bacillus halodurans*, stands out as the very first example of an antimicrobial natural product that was identified through a genome mining approach rather than through a classical cell extract screening based methodology[36]. McClerren et al. set out to find analogues of the lantibiotic mersacidin when they discovered the *halA1* gene, encoding for the precursor peptide HalA1. The peptide was seen to have 34% sequence homology with the mersacidin precursor peptide. Of particular note was the presence of another peptide (HalA2) as well as two LanMs (HalM1 and HalM2) within the same cluster. Indeed, through heterologous expression of the peptides and activity tests, McClerren et al. were able to show that Halα and Halβ (the post-translationally modified lanthipeptides made from HalA1 and HalA2, respectively) act synergistically to exert activity. What should also be noted is that all components of the cluster were successfully heterologously expressed in *E. coli*, eliminating the need to culture the native producer.

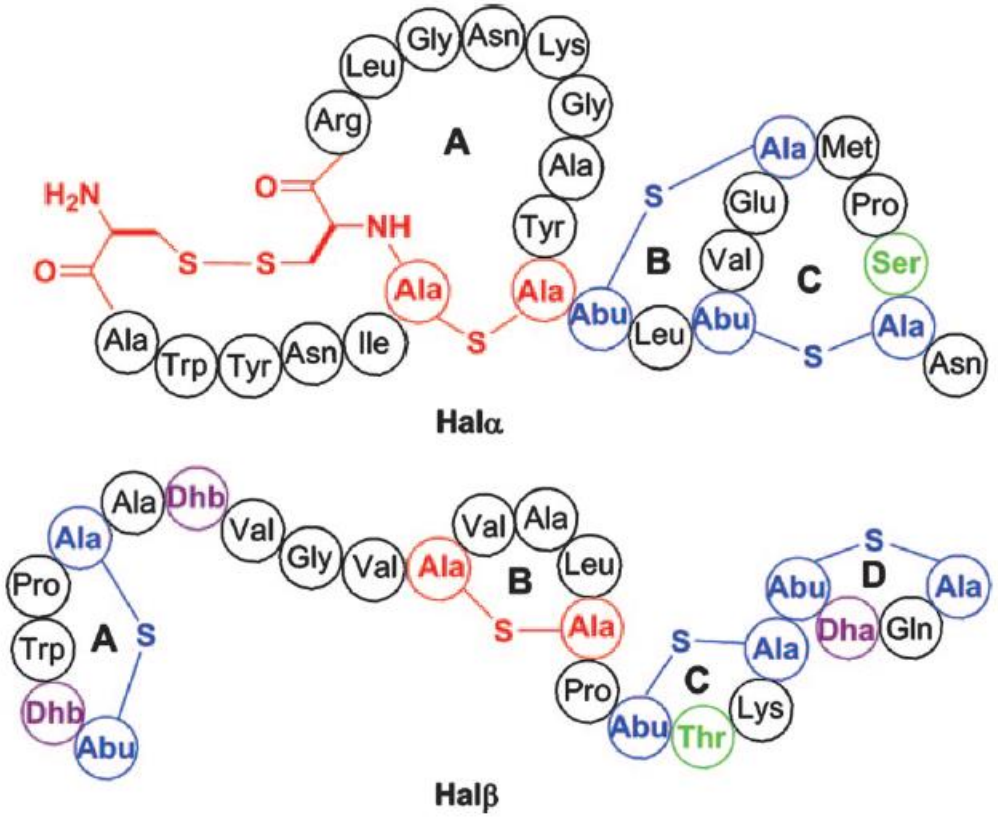


Figure 2.3. The amino acid sequence of the haloduracins, Halα and Halβ. Adapted with permission. [36]

Another example of a two component lanthipeptide system found through genome mining are the lichenicidins, first reported in 2009[37]. In this work, Begley et al. opted to look for LanM homologues, whose amino acid sequences are more highly conserved than lanthipeptide precursors. LanM enzymes homologous to CylM were discovered in the genome of *Bacillus licheniformis*, a known producer of lanthipeptides. This search revealed a gene cluster consisting of two LanM enzymes (BliM1 and BliM2) and two putative lanthipeptide precursors (BliA1 and BliA2), both of which showed significant homology to HalA1 and HalA2, suggesting a two-component system with similar structure/function. Isolation of the modified peptides Bliα and Bliβ and subsequent activity assays showed that, while they were able to exert some activity individually, their activity was significantly enhanced when assayed together. Like the haloduracin cluster, the lichenicidin cluster was also successfully heterologously expressed in *E coli*.

In a more recent example, Zhao et al. discovered a cluster in *Ruminococcus flavefaciens* that encoded for two LanMs (FlvM1 and FlvM2) but, intriguingly, a total of 12 lanthipeptides[38]. Multi-substrate tolerance has been previously reported for certain LanMs, such ProcM from *Prochlorcoccus marinus* MIT9313, which has 30 substrates[39]. However, the Flv system stands out in particular due to the presence of multiple two-component antibiotics, derived from different combinations of the encoded peptides. Of the 12 peptides, 4 had significant sequence homology with HalA1 and the remaining 8 with HalA2. As such, the peptides were designated α or β peptides based on their similarity to HalA1 or HalA2, respectively. Antimicrobial activity assays showed that one of the α peptides (Flvα.a) showed synergistic activity in the presence of 2 of the β peptides (Flvβ.b and Flvβ.g). Such diversity is unprecedented for two component lantibiotics.

Given the fact that all reported two-component lanthipeptides are antibiotics, and given the simplicity of heterologous expression in relation to culturing the native producers of lanthipeptides, we were inspired to use genome mining as tool in the discovery of novel two component lantibiotics. In this study, genome mining was used to identify a new two-component lanthipeptide system in the actinomycete, *Micromonospora saelicesencis*. We report here the heterologous expression of this gene cluster in *E. coli*, along with purification and structural analysis of the modified lanthipeptide products and preliminary biological activity assays against a panel of pathogenic human bacteria.

2.2. Results and discussion

A common feature among two-component lanthipeptides is that, in all reported cases, the peptides are modified by LanMs. And, as mentioned earlier, these proteins have well-conserved domains, one of which catalyzes dehydration and one of which catalyzes cyclization. To that end, we opted to search for homologues of HalM2 using the basic local alignment search tool (BLAST). A hypothetical protein (SaelM2) sharing 31% sequence identity with HalM2 was identified in the genome of *Micromonospora saelicesencis*, an actinomycete than grows on the root nodules of *Lupinus angustifolius*[40]. To date, there have been no characterized examples of two-component lanthipeptides from actinomycetes – the most prolific producer of known antibiotics. As such, the strain warranted closer inspection. Closer inspection of the genomic locus surrounding SaelM2 (Fig. 2.4) revealed a second class II lanthipeptide synthetase (SaelM1) as well as two precursor

peptides (SaelA1 and SaelA2), a hypothetical ABC transporter (SaelT), and a protein of unknown function (SaelX). Transport proteins homologous to SaelT are present in many lanthipeptide gene clusters, where they function to both proteolyze the leader peptide and to export the post-translationally modified products into the extracellular milieu[26].

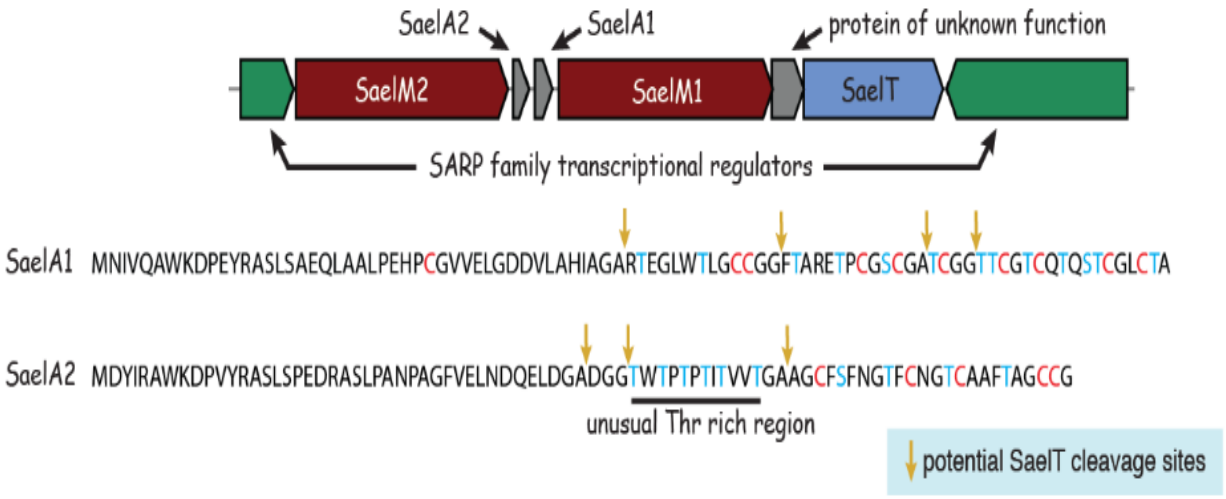


Figure 2.4 The Sael cluster, showing the sequences of the precursor peptides. Residues marked in blue are potentially dehydratable. Those marked in red are thought be involved in thioether ring formation.

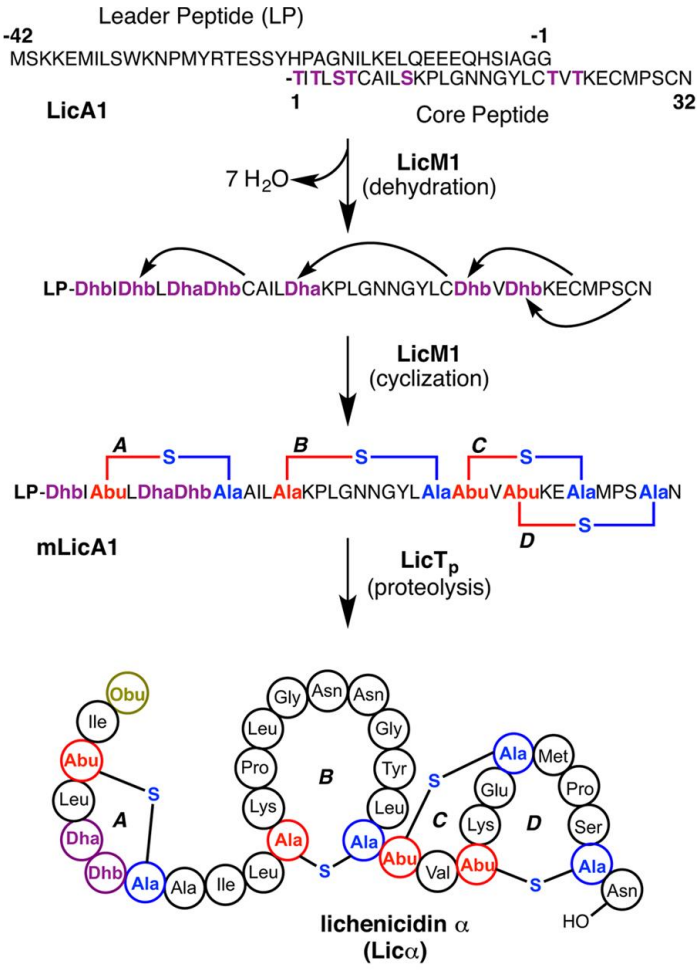


Figure 2.5 An illustration of LanM activity, using LicM1 and its final product lichenicidin α as an example. Adapted with permission. [26]

Interestingly, a sequence alignment of the two precursor peptides in the *M. saelicesensis* cluster with other known two-component lanthipeptides show that the distribution of Cys/Ser/Thr residues in both SaelA1 and SaelA2 is unique, suggesting that the thioether topology is likely unique (Fig. 2.6). This highlights one of the most beneficial aspects of RiPP natural products; namely, the sequence of the precursor peptide provides immediate evidence that the cluster is likely to encode for production of a novel molecule. More specifically, both peptides are rich in Ser/Thr residues, and SaelA1 contains 10 cysteine residues, meaning it could potentially form 10 Lan/MeLan rings, which would make it the most highly cyclized lanthipeptide characterized to date. Moreover,

SaelA2, contains an unusual Thr rich region, and both peptides contain putative SaelT cleavage motifs within the leader peptide.

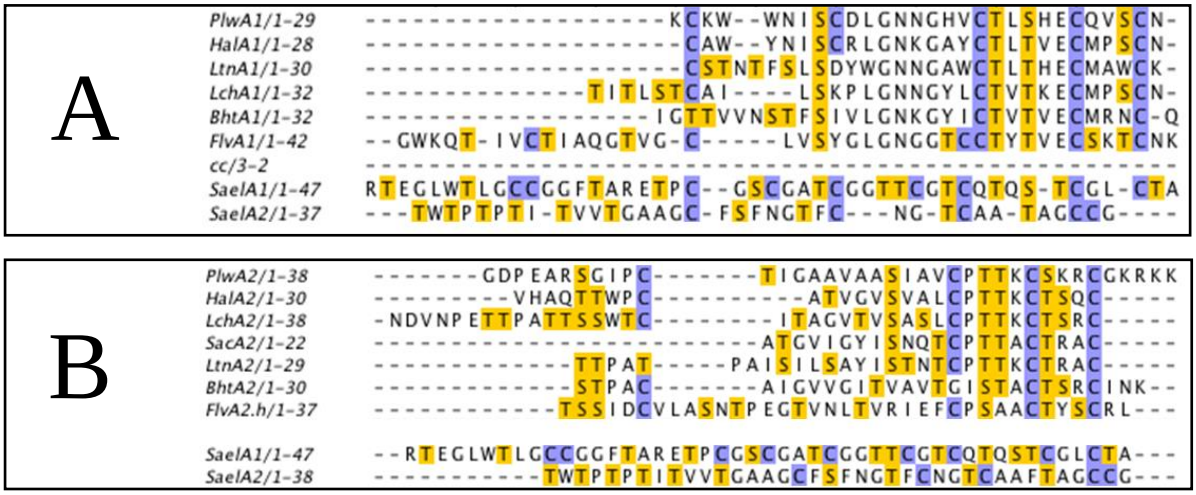


Figure 2.6 A sequence alignment of various two-component lantibiotics with the putative core peptides of SaelA1 and SaelA2, highlighting the lack of sequence homology for the SaelA peptides. (A) is an alignment of the SaelAs with α peptides, while (B) is an alignment of SaelAs with β peptides

Using Gibson assembly[41], the *saelA1* and *saelA2* genes were cloned into the *pET-DUET* and *pRSF-DUET* plasmid vectors, respectively, for heterologous expression in *E. coli*. The genes were cloned in such a way as to introduce *N*-terminal His_6-affinity tags, allowing the facile purification of the heterologously expressed peptides using immobilized metal affinity chromatography (IMAC). After purification, the His_6-tagged peptides were desalted using C4 solid phase extraction (SPE) and were analyzed using matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Indeed, analysis of the mass spectra indicated the presence of unmodified SaelA1 (Fig. 2.7) and SaelA2 (Fig. 2.8). Confident that heterologous expression would work for the *M. saelicesensis* two-component lanthipeptide system, we set out to assign each peptide's cognate LanM. This was done through *in vivo* co-expression experiments, wherein either SaelM1 or SaelM2 were cloned into the same vector as SaelA1 or SaelA2. This led to the creation of 4 clones: (i) SaelA1+SaelM1/pET-DUET, (ii) SaelA1+SaelM2/pET-DUET, (iii) SaelA2+SaelM1/pRSF-DUET and (iv) SaelA2+SaelM2/pRSF-DUET. Based on the sequence homology of SaelM1 with HalM1 (~32%) and SaelM2 with HalM2 (~31%), it was predicted that SaelM1would modify SaelA1 and that SaelM2 would modify SaelA2. Following co-expression in *E. coli* and purification of the His_6-tagged peptides, this was corroborated by MALDI-TOF-MS,

which clearly showed dehydration of SaelA1 exclusively in the presence of SaelM1 (Fig. 2.7), and dehydration of SaelA2 exclusively in the presence of SaelM2 (Fig. 2.8). As has been generally observed for other two-component lanthipeptide systems, the co-expressions suggest that the SaelM1 and SaelM2 are highly selective for their cognate substrates.

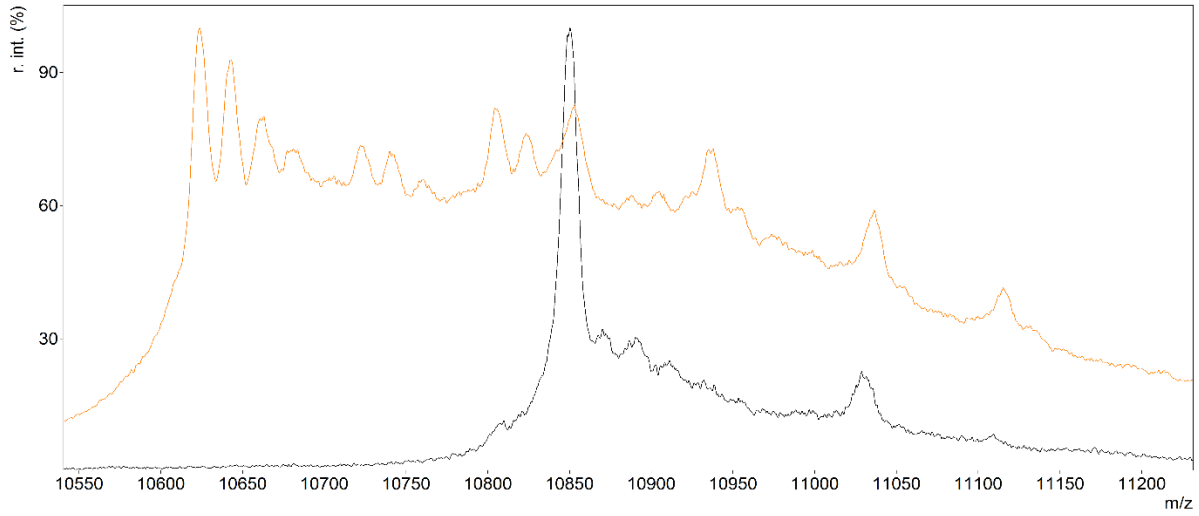


Figure 2.7 The MALDI-TOF MS spectra of SaelA1 coexpressed with SaelM1 (gold) and SaelM2 (black). Unmodified SaelA1 is expected to have an m/z of 10,856 in the +1 charge state. The fully modified SaelA1 has an m/z of 10,622 in the +1 charge state.

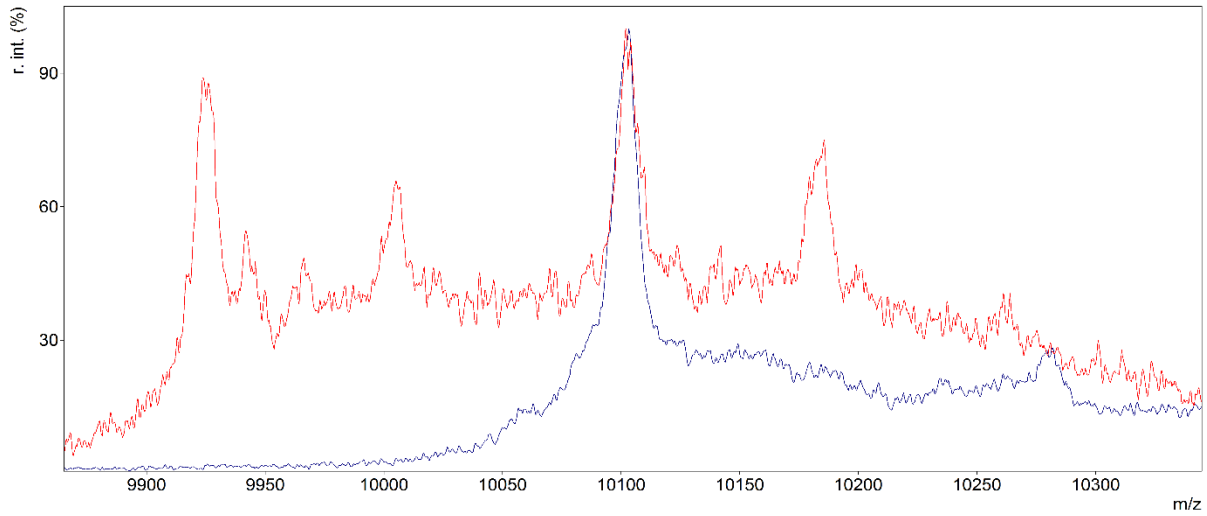
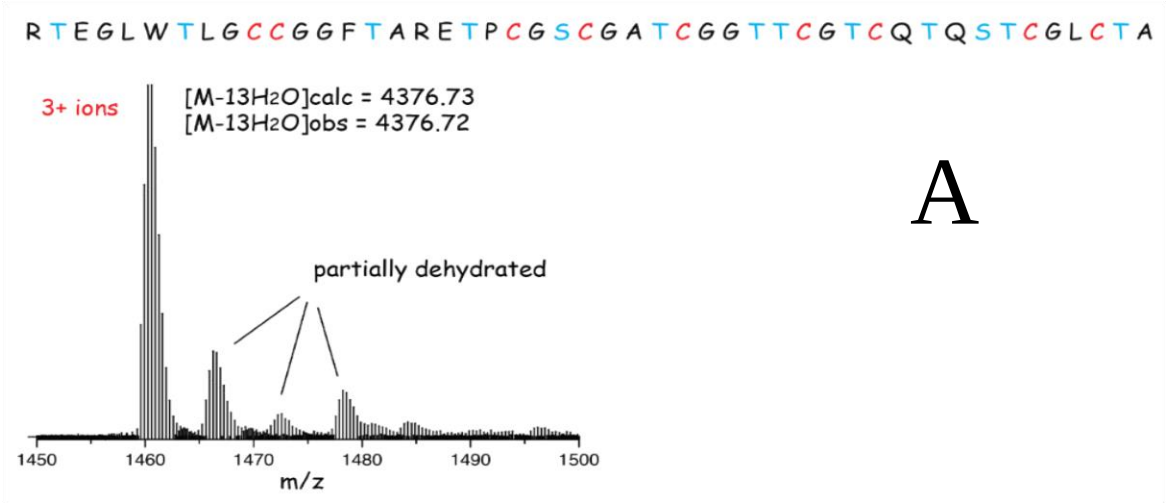


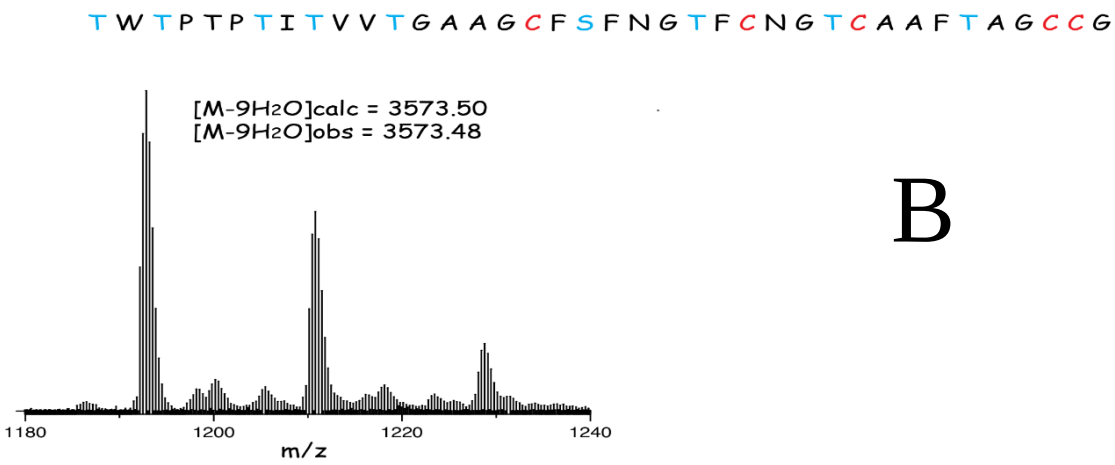
Figure 2.8. The MALDI-TOF MS spectra of SaelA2 coexpressed with SaelM1 (blue) and SaelM2 (red). Unmodified SaelA2 is expected to have an m/z of 10,100 in the +1 charge state. The fully modified SaelA2 has an m/z of 9920 in the +1 charge state.

Again using *in vivo* co-expression experiments, SaelA1+SaeM1 and SaelA2+SaelM2 were expressed with the multifunctional transport-protease enzyme, SaelT. These enzymes typically cleave modified lanthipeptide precursors *C*-terminal to GA or GG dipeptide motifs residues[26],

several of which are present in the leader peptides of both SaelA1 and SaelA2 (Figure 2.4). Since SaelT (if active) was also expected to export the modified, proteolyzed SaelA1 and SaelA2 peptides into the culture media, *E. coli* cells co-expressing the *sael* genes were first removed from co-expression cultures using centrifugation and the remaining supernatant was extracted with C8-SPE and further purified by reverse phase high performance liquid chromatography (HPLC). The purified samples were then analyzed by liquid chromatography electrospray ionization mass spectrometry (LC-ESI-MS). These data clearly showed masses corresponding to cleavage after GA or GG motifs in the SaleA1 and SaelA2 leader peptides, respectively (Fig. 2.9 A, B). SaelA1 was cleaved after Ala-45 to give a 13-fold dehydrated peptide (termed Saelα) with an observed m/z of 4376.72 (predicted m/z: 4376.73), whereas modified SaelA2 was cleaved after Gly-42 to give a 9-fold dehydrated peptide (Saelβ) with an observed m/z of 3573.48 (predicted m/z: 3573.50). The nearly complete dehydration of each Sael core peptide (13/13 possible sites in SaelA1 and 9/10 possible sites in SaelA2) indicates that, when co-expressed in *E. coli*, the SaelM1 and SaelM2-mediated dehydration reactions are extremely efficient and/or that the transport enzyme, SaelT, preferentially process and exports the fully modified peptides. Along these lines, we could detect small quantities of partially dehydrated peptides in the cell culture extract, but no unmodified peptides were detected. We also attempted to assign a function to SaelX, which does not share any detectable sequence homology with proteins of known function in the NCBI database. For these experiments, SaelX was co-expressed along with SaelT and either SaelA1+SaelM1 or SaelA2+SaelM2. The His_6-tagged peptides were again purified from these co-expression cultures and analyzed by MADLI-TOF-MS, but the mass spectra did not indicate any activity from SaelX.



Saelα



Saelβ

Figure 2.9 The ESI-MS spectra of Selα (A) and Selβ (B) extracted from the culture media in co-expression experiments with the transport/protease enzyme SelT. Residues marked in blue are dehydration sites, while ring-forming cysteine residues are marked in red.

Next, we determined the extent of thioether (Lan/MeLan) ring formation by alkylating the extracted Sel*α* and Sel*β* peptides with *N*-ethylmaleimide (NEM) and analyzing by LC-ESI-MS. This alkylation step is necessary because the LanM-mediated cyclization reaction involves no change in mass. Using the alkylation conditions described in the Methods section, NEM selectively alkylates the thiols of free cysteine residues, whereas the thioether sulfur atom of Lan/MeLan residues are unreactive. For Saelα, which contains 9 Cys residues, there was a small amount of a 7-fold cyclized species, but the major product was unreactive towards NEM, suggesting that the Saelα peptide is 9-fold cyclized (Fig. 2.10 A & B). While this experiment

cannot rule out the possibility that some of the Cys thiols in the fully modified (and presumably structured) Saelα may be inaccessible to reaction with NEM, this data strongly suggests that Sael*α* contains nine thioether rings and, is thus, the most heavily cyclized lanthipeptide characterized to date. When the NEM alkylation experiment was repeated with Sael*β*, the LC-ESI-MS results indicated that none of five cysteine residues in the Sael*β* were available for alkylation, suggesting the presence of 5 thioether rings in Sael*β* (Fig. 2.10 C & D).

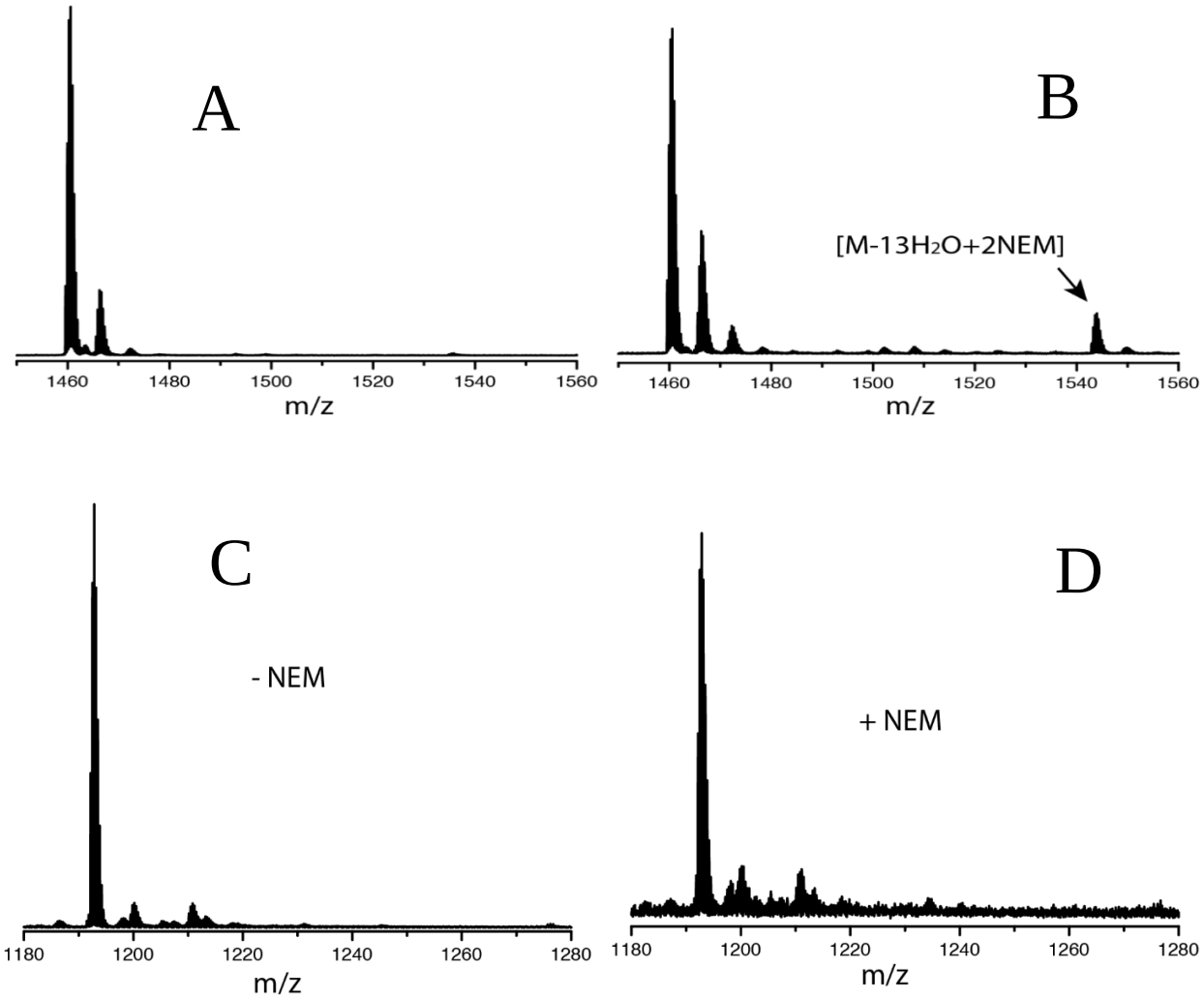


Figure 2.10 The mass spectra of Saelα (A & B) and Saelβ (C & D) in the 3+ charge state. B & D mass spectra of the peptides following NEM alkylation. As can be seen, Saelβ is unreactive towards NEM, indicating that all thiol moieties are engaged in cyclization. A small quantity of 7-fold cyclized Saelα was detectable as the two-fold NEM-alkylated product, [M-13H_2O+2NEM], but the majority of Saelα was not NEM-alkylated, suggesting the presence of 9 thioether rings in this peptide.

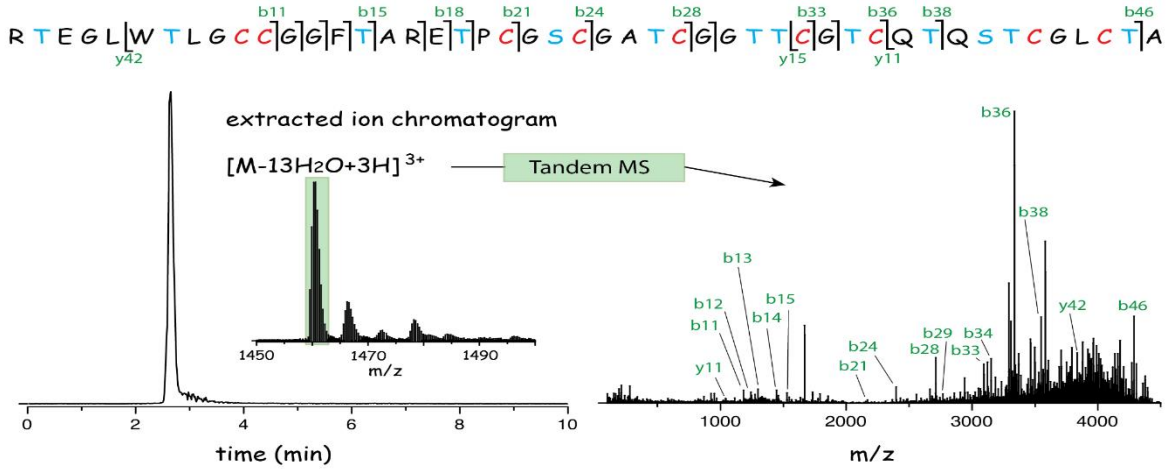


Figure 2.11 The MSMS analysis of Saelα, showing the $[M\text{-}13H_2O+3H]^{3+}$ *precursor ion selected for fragmentation by CID, and the resulting fragment ions. The fragment ions are mapped onto the Saelα peptide sequence in the inset. The collision energy used was 40-50 eV.*

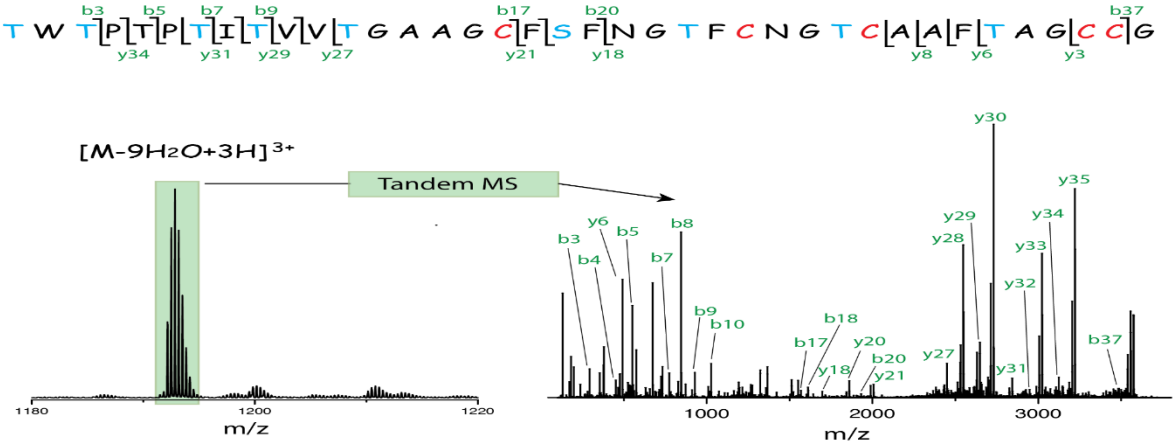


Figure 2.12. The MSMS analysis of Saelβ, showing the $[M\text{-}9H_2O+3H]^{3+}$ *precursor ion selected for fragmentation by CID, and the resulting fragment ions. The fragment ions are mapped onto the Saelβ peptide sequence in the inset. The collision energy used was 20-30 eV.*

Next, in an attempt to map out the regiochemistry of the post-translational modifications present Saelα and Saelβ, we employed tandem mass spectrometry with collision induced dissociation (CID, Figure 2.11 and 2.12). In this technique, the peptide ion of interest is first isolated in the gas phase using a quadrupole mass analyzer. The isolated ion is then accelerated through a collision cell filled with Ar buffer gas. Upon undergoing multiple collisions with the buffer gas, the internal energy of the ion is increased until the excess energy (stored in vibrational modes of the peptide ion) is released though the cleavage of covalent bonds in the peptide. The

m/z values of the resulting fragment ions are then recorded and the location of any post-translational modifications can be determined by looking for the associated mass shifts in the peptide fragment ions. For lanthipeptides, the presence of thioether rings often leads to decreased fragmentation in regions of the peptide containing the thioether rings. This is because in order to see fragments, two covalent bonds would need to be cleaved in the gas phase, which is a low probability event. In line with this property, CID fragmentation of Saelα under conditions that are typically sufficient for extensive lanthipeptide fragmentation resulted in very few fragment ions (Fig. 2.13). Increasing the collision energy did result in the formation of more fragment ions, but the total peptide coverage was still relatively poor. These results are consistent with a highly cyclized structure expected for a lanthipeptide containing potentially 9 thioether rings. Saelβ fragmented more readily (Fig. 2.12), which is to be expected, as the peptide has fewer possible thioether rings. In order to determine which of the 10 potential dehydration sites in SaelA2 was skipped by SaelM2, we analyzed the Saelβ fragmentation data with hypothetical structure enumeration and evaluation (HSEE). Reported by Zhang et al., HSEE allows users to predict the structure of modified peptides using their tandem MS profile, assuming the peptide sequence and the type and number of modifications are known[42]. In brief, the script creates a set of hypothetical structures based on the possible modifications and modification sites. Each of these structures has an assigned set of predicted MSMS fragments, and the script will analyse MSMS data and assign as many fragments as possible. The hypothetical structure with the most assigned MSMS fragments is identified as the most likely hypothetical structure. Applying this script to the Saelβ MSMS data led to two hypothetical structures with equivalent scores (lacking dehydration at either Thr7 or Thr 9) (Fig. 2.14).

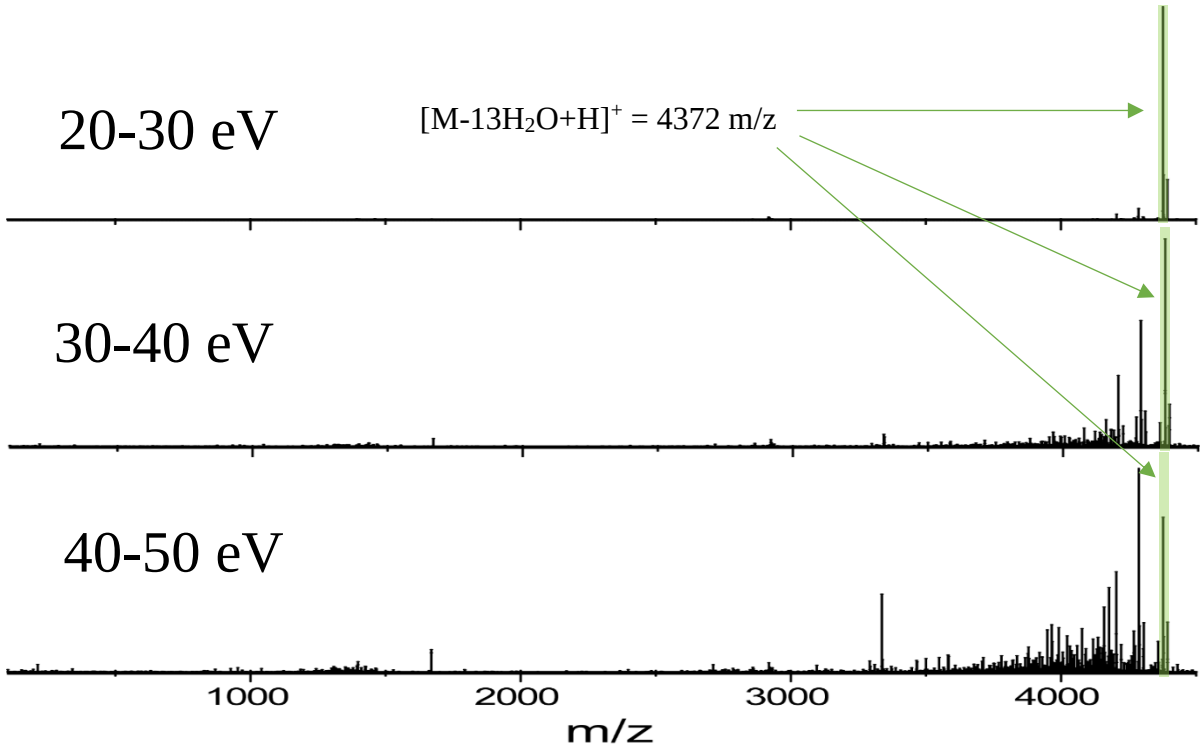


Figure 2.13 A comparison of the extent of fragmentation at different collision energies for Saelα. Saelβ showed considerable fragmentation at 20-30 eV.

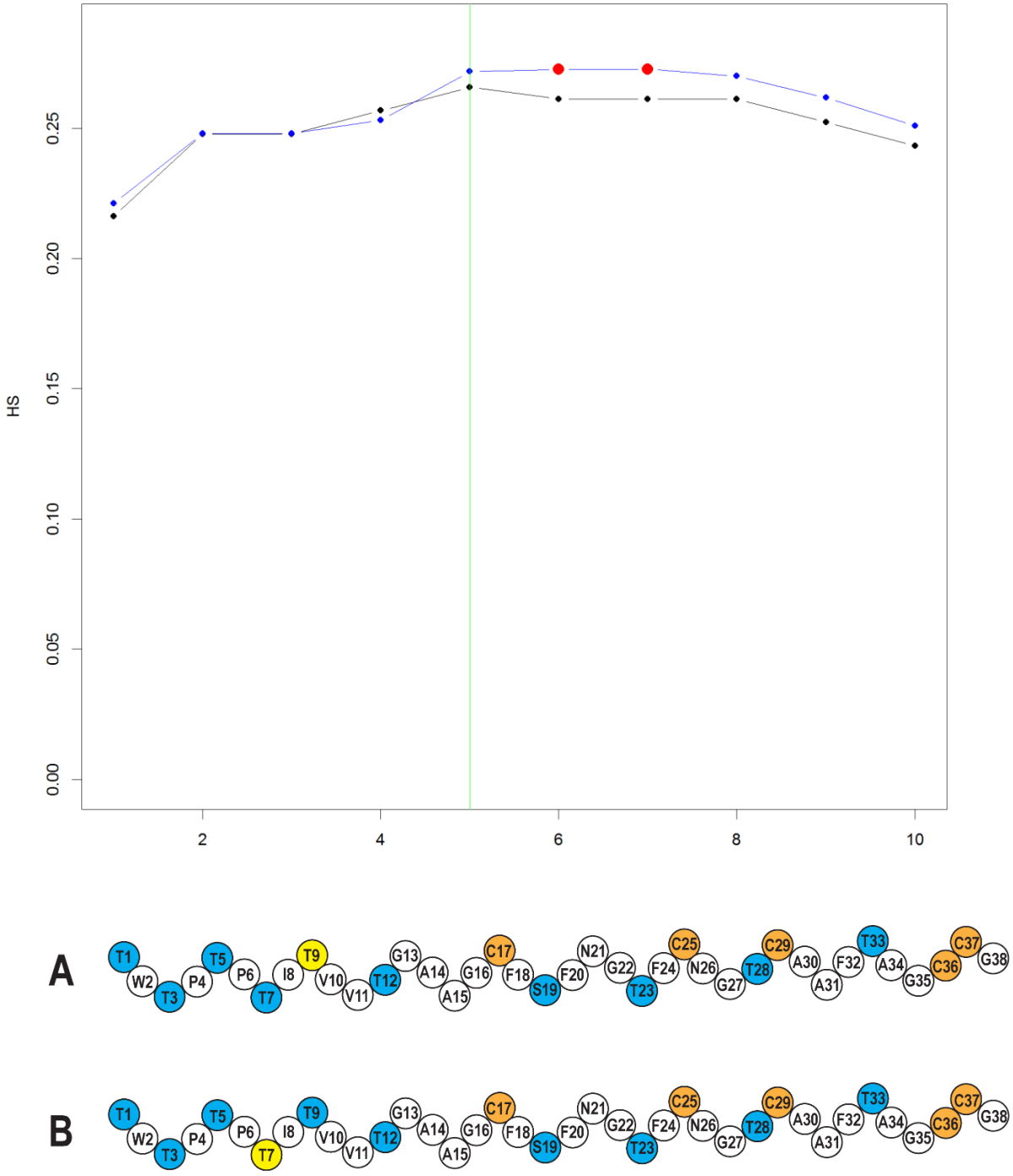


Figure 2.14. The results of HSEE analysis for Saelβ. The x-axis represents the hypothetical structure and the y-axis represents the score. Structures 6 (A) and 7 (B), with skipped dehydrations at Thr7 and Thr9, respectively, showed the highest score after analysis. Blue represents dehydrated residues, yellow represents the unmodified Thr residue and cysteine residues are marked in orange.

In order to test the biological activity of Saelα and Saelβ, we grew 12 L cultures of *E. coli* expressing SaelT, SaelX, and either SaelA1+SaelM1 or SaelA2+SaelM2 and extracted the mature lanthipeptides using a methanol extraction protocol reported previously[43], followed by semi-preparative and analytical RP-HPLC. Typical yields were 0.5-1 mg/L for Saelα and 0.25-0.5 mg/L for Saelβ. With a reliable method of purification in hand, we next tested for antibiotic activity of Saelα and Saelβ. Initially, the two peptides were assayed against *Micrococcus luteus*, a Gram-positive bacterium that is commonly used an indicator strain for antimicrobial compounds. Liquid culture assays were first carried out wherein the Saelα and Saelβ peptides were added independently or in combination at a concentration of 10 μM. The antimicrobial activity was determined by measuring the optical density of the cultures (i.e. the absorption at 600 nm) which is an indication of cell density and hence, cell growth. During the course of these studies, no significant differences in the OD_{600} was seen between the negative control (lacking Sael peptides) and any of the samples with peptides. Given the relative insolubility of the peptides in water (which typically require approximately 30% acetonitrile to be soluble in water), the lack of activity may have been due to the peptides precipitating or forming inactive micro-aggregates in solution when added to culture. To that end, we opted to use an agar plate assay instead. We employed a disk diffusion assay; 1 mL of *M. luteus* liquid culture was added to an agar plate and left to dry. To this, sterile discs loaded with 10 μM Saelα, Saelβ or both were placed on to the plate and left to incubate for 16 hours. No zone of inhibition was seen on the plate, suggesting again that the peptides were ineffective as antibiotics against *M. luteus*.

Under the assumption that the peptides were ineffective against *M. luteus*, we moved on to testing them against several other strains of bacteria, both gram-positive and gram-negative, in collaboration with the Karine Auclair group. These strains were as follows: (i) *Acinetobacter baumanii* ATCC 19606, (ii) *Klebsiella pneumoniae* ATCC13883, (iii) *Pseudomonas aeruginosa* ATCC 27853, (iv) *Escherichia coli* ATCC 25922, (v) *Salmonella typhimirium* ATCC 14028, (vi) *Bacillus subtilis* ATCC 6051, (vii) *Enterococcus faecium* ATCC 19434 and (viii) methicillin resistant *Staphylococcus aureus* ATCC 43300. Unfortunately, as with *M. luteus*, disk diffusion assays employing 10 μM or 100 μM peptide did not result in any observable zones of inhibition, indicating that Saelα and Saelβ are not functioning as antimicrobial agents under the conditions tested.

We then hypothesized the peptides may be involved in metal transport. While there are no reports of lanthipeptides being involved in metal transport, the highly cyclized nature of the products led us to believe that they could potentially chelate metals. There are several examples of peptides that bind metal, such as the methanobactins, which are also cyclized peptides, and the fuscachelins[44, 45]. To that end, we tested the ability of Saelα and Saelβ to bind Zn(II), Cu(II) and Ni(II), both independently and in an equimolar mixture. The divalent forms of the metals were selected because literature shows metal binding peptides and proteins most often interact with these, rather than other charge states. Moreover, studies with monovalent metals proved challenging due to oxidation of the metals, potentially confounding the data. Oxidation was especially problematic for iron, wherein Fe(II) was rapidly oxidized to Fe(III), precipitating in solution. Consequently, these metals were not included in this experiment.

We opted to use UV-Vis spectroscopy, with a CARY 5000 UV-Vis-near IR spectrophotometer, in order to detect and characterize these interactions, scanning wavelengths between 240 and 800 nm. Solutions of Saelα and Saelβ were prepared in water with 10% (v/v) DMSO, independently and in combination, and were mixed 1:1 and 1:10 molar ratio with metal solution. The final concentration of each peptide was 100 μM, and each metal was either prepared at 100 μM or 1 mM. The metal solutions prepared were from the chloride salts $ZnCl_2$, $NiCl_2$, and $CuCl_2$. Controls were as follows: (i) 100 μM Saelα, (ii) 100 μM Saelβ, (iii) 100 μM Saelα and Saelβ, (iv) 100 μM of each metal individually (10% DMSO) and (v) 1 mM (10% DMSO) of each metal individually. After mixing, the solutions were left to incubate for 10 min at room temperature before being analyzed via the spectrophotometer. The spectra collected for solutions of metal alone were subtracted from the spectra of metal with peptide(s). Binding of metal to the peptides should lead to a shift in the spectra, and so this approach would highlight regions of differences, clearly indicating if a binding event had occurred. However, the data show that there was no clear shift in the spectra for each metal in the presence of each peptide, individually or in combination. Thus, it can be presumed that the peptides do not bind the metals tested.

2.3.Conclusion:

In this study, a two-component lanthipeptide system from the actinomycete *Micromonospora saelicesencis* was discovered through genome mining. The components of the

predicted cluster were heterologously expressed to yield the peptides Saelα and Saelβ. The peptide were characterized using mass spectrometry based techniques. It was seen that Saelα, a 47-mer peptide, had all Ser/Thr residues dehydrated, leading to a total of 13 dehydrations. Additionally, Saelα was found to contain a total of 9 thioether rings, leading a highly cyclized structure. These results were corroborated by MSMS analysis, wherein significant energy was required to fragment the peptide, indicative of a highly cyclized and interlocked structure. To our knowledge, this is the most highly cyclized lanthipeptide reported. Analysis of Sael*β* indicated a lesser degree of modification, with 9 dehydrations and 5 thioether rings. However, unlike Saelα, which had all Ser/Thr residues dehydrated, either Thr7 or Thr9 in Saelβ eludes dehydration. Antimicrobial activity studies were carried out for both Saelα and Saelβ, but no discernable activity was detected against a panel of bacteria.

One possible explanation for the lack of activity is that these peptides diffuse slowly through the agar plate as a consequence of their insolubility in water. However, it also possible that these peptides serve some other function. For instance, the cytolysin system, while showing activity against bacteria, functions as a virulence factor for *Enterococcus faecalis* by lysing mammalian cells, such as erythrocytes and immune cells[30]. And similar to the Sael peptides, the cytolysins share no significant sequence and structural homology with other two-component lanthipeptides. However, the cytolysins represent the only reported case of a lanthipeptide displaying lytic activity against mammalian cells, and they are produced by a known animal pathogen, rather than by a soil bacterium (*M. saelicesencis*) that functions as a plant symbiont.

To that end, another possible role for the Sael peptides is to act as metal binders. This may take on one of two forms: (i) They may be metal dependent antibiotics, such as the bacitracins[46], or (ii) they may act as metal transporters, such as the siderophores[45]. The bacitracins are a class of cyclic peptides that chelate metal ions (typically divalent ions such as Zn^{2+}) in order to exert antibiotic activity. Much like two-component lanthipeptides, the bacitracins target Gram-positive bacteria by inhibiting cell wall synthesis. However, instead of targeting lipid II, as most other two-component lanthipeptides are thought to, the bacitracins bind to and inhibit the C55-isoprenyl pyrophosphate (IPP), which is a required component of the cell wall biosynthesis pathway. A defining feature of the bacitracins is the presence of a macrocycle whose purpose is to chelate a

divalent metal ion[46]. As such, the highly cyclized Sael peptides have the potential to chelate metal ions, and may need to do so in order to exert activity.

Moreover, while there are no reported cases of lanthipeptides acting as metal chelators, given the extent of cyclization in both peptides, it would not be unreasonable for them to act as one. There is already significant precedence for metal transporting peptides in literature. For instance, the siderophores are iron binding peptides that act as shuttles for iron across bacterial membranes. The methanobactins, another example of metal binding peptides, are known to bind copper, and are believed to recruit copper from the extracellular matrix for use in intracellular processes[44]. Moreover, the host species, *Micromonospora saelicesencis,* is found growing on the root nodules of blue lupine (*Lupinus angustifolius*)[40], meaning it may be involved in symbiotic metal transport. To that end, the ability for the peptides Saelα and Saelβ to bind Zn^{2+}, Ni^{2+} and Cu^{2+} was tested. The tests show no evidence of metal binding.

A final explanation may be that some component of the biosynthetic pathway for Saelα and Saelβ may have been excluded from our attempts at heterologous expression. This would not be surprising, as several examples of lanthipeptide systems wherein proteases and other tailoring enzymes that are endogenous to the native producer but not clustered with the genes for the other components are required for activity. There may also be an enzyme we have simply missed in our analysis of the genome that may be involved in modification. We attempted to characterize a gene of unknown function (*saelX*) encoded in the saelicesencin cluster, but we could not detect any change in the masses of Saelα or Saelβ when produced in the presence of SaelX. However, since we did not perform any assays in the native *M. saelicesensis* producer, we cannot exclude the possibility that SaelX requires for activity some cellular component not present in the *E. coli* heterologous host.

2.4.**Future Work**:

Given the proximity of *M. saelicesencis* to blue lupine, it is possible that it exerts some form of antifungal activity. In fact, the pinensins are one recently characterized example of lanthipeptides acting as antifungal agents[20]. Moreover, actinobacteria have been noted to be involved in upregulating genes involved in plant defense pathways[47]. To that end, the Thibodeaux

lab will be collaborating with Martha Trujillo at the University of Salamanca to determine if the peptides play either role.

2.5.Experimental Methods:

2.5.1. **Genome mining**: Analyses were carried out using the National Center for Biotechnology Information (NCBI) database. The protein basic local alignment search tool (BLASTp) was employed using the sequence of HalM2 as the query. NCBI database and antiSMASH were used to further analyze the ORF wherein SaelM2 was detected to look for other ORFs within the cluster.
2.5.2. **Cloning**: Unless otherwise indicated, all cloning reagents were purchased from New England Biolabs (NEB) and all oligonucleotide primers were purchased from Integrated DNA Technologies. *Micromonospora saelicesensis* genomic DNA was purchased from DSMZ. Genes from the putative saelicesin biosynthetic gene cluster were amplified from *M. saelicesensis* genomic DNA using the polymerase chain reaction (PCR). A typical PCR reaction contained 50 ng genomic DNA, 0.5 µM of each oligonucleotide primer (Table 2.1), 1 mM deoxynucleotide-5′-triphosphates (dNTPs), 5% DMSO, 1x Phusion Polymerase buffer (either HF or GC buffer), and 1 unit of Phusion Polymerase. The oligoneuclotide primer sequences used for each amplification are given in Table 1. Typical PCR cycling conditions were as follows: 3 min incubation at 98 °C, followed by 35 cycles composed of a 15 s denaturation step at 98 °C, a 20 s annealing step at a temperature ranging from 65 – 72 °C depending on the gene being amplified, and an elongation step at 72 °C for 1.5 min. After the last cycle, a final elongation was carried out for 5 min at 72 °C. The sizes of the amplicons were verified by 0.5% agarose gel electrophoresis and the PCR products were purified using the NucleoSpin Gel and PCR cleanup kit from Macherey Nagel following the manufacturer's instructions. The pDUET vectors were digested with either BamHI (when cloning the *saelA1* or *saelA2* genes with *N*-terminal His_6-tags) or with NdeI (when cloning *saelM1*, *saelM2*, *saelT* or *saelX*). After overnight restriction enzyme digestion at 37 C, the digested vectors were purified using the NucleoSpin Gel and PCR cleanup kit from Macherey Nagel following the manufacturer's instructions. After purification of all required DNA fragments, the *E. coli* expression constructs for the *sael* genes were generated using the Gibson assembly[41] using a 1:1 molar ratio of insert to vector. In brief, Gibson assembly is a tool used in combining multiple DNA fragments into a single construct. In this study, Gibson assembly was employed to insert a gene of interest into a vector, which imparts a specific antibiotic resistance,

such that bacterial colonies can be selectively screened for the plasmid. The process begins with the PCR amplification of a gene of interest using primers that introduce a small number of bases to either end that overlap with a particular region of the desired vector. The vector is then cleaved such that the region of overlap is exposed. A 5' exonuclease is introduced that digests in the 5' to 3' direction, creating what are referred to as 'sticky ends'. These are simply the portion of the vector and insert that are complementary. After annealing, a polymerase introduces the missing bases to close gaps. Finally, a DNA ligase seals any nicks in the DNA. For Gibson assembly, a 5 uL solution containing equimolar amounts of each fragment (100 ng of larger fragment) was prepared, and to this, 15 uL of Gibson mix was added. The mixtures were then incubated at 50 ºC for 1 hour.

Table 2.1. Primers used in this study, with corresponding sequences and T_m values.

Primer #	Primer name	Sequence (5'-3')	Tm (°C)
1	SaelA1_f	GAACATATGAACATCGTGCAAGCCTGGAAGGACCCCGAG	75
2	SaelA1_r	GGAACTCGAGTCAGGCGGTGCACAGACCGCAG	77
3	SaelA2_f	GGACATATGGACTACATCAGGGCGTGGAAGGACCCGGTCTACCG	79
4	SaelA2_r	GAAACTCGAGTCAGCCGCAGCAGCCGGCGGTG	79
5	SaelM1_f	GGACATATGACGCAGTCGCACCTGCCCGGCCAC	79
6	SaelM1_r	GAAGGAATTCAGAGCTGCTGCTCAGCCACGCGGGTCACC	79
7	SaelM2_f	GAACATATGGACGACGGCCTTGCCCCGGTGCAG	77
8	SaelM2_r	GGAACTCGAGCTATGGCAGATCGATGGCACCGCGATGGTC	78
9	SaelX_f	GTATATTAGTTAAGTATAAGAAGGAGATATACATATGACCCGCGTGGCTGAG	71
10	SaelX_r	GGCCGGCCGATATCCAATTGAGATCTGCCATTCATCGCAGGACCCCCTC	81
11	SaelT_f	CACCATCATCACCACAGCCAGGATCCGATGACGATGAGCTGGCTGCG	81
12	SaelT_r	TCGACCTGCAGGCGCGCCGAGCTCGAATTCGCTACAGCGGCCGGGCC	88
13	Duet_colony_FP	GAGCGGATAACAATTCCCC	56
14	Duet_colony_RP	GCTAGTTATTGCTCAGCGG	56

The *saelA1*, *saelA2*, and *saelX* genes were first cloned into the BamH1 sites of *pET-DUET*, *pRSF-DUET*, and *pACYC-DUET* to make the SaelA1/pET-DUET, SaelA2/pRSF-DUET, and SaelX/pACYC-DUET constructs. These constructs were then used as templates for cloning of the *saelM1*, *saelM2*, and *saelT* genes into the Nde1 site. All constructs generated by Gibson assembly were transformed into chemically competent *E. coli* DH5α cells using the following procedure. 10 uL of mixture was transformed into chemically competent E. Coli DH5α. The cells were incubated on ice for 30 minutes before the addition of 1 mL of LB broth. The cells were then incubated at 37 ^{o}C for 1 hour before being streaked on an antibiotic infused LB agar plate (50 ug/mL kanamycin for *pET28b*, 100 ug/mL ampicillin for *pET15B*, 25 μg/mL chloramphenicol for *pACYC-DUET*). The plates were then allowed to dry and then incubate at 37 ^{o}C for 16-20 hours.

Several colonies were selected for colony PCR. Colony PCR is a process used to screen cells for a desired plasmid by attempting to PCR amplify a portion of said plasmid. Cells from a single clonal population are introduced into a PCR tube, and a desired PCR mix is added. All the components of the mix were added following the manufacturer's instructions for OneTaq polymerase (0.5 μM primer, 1 mM dNTPs, 1X Taq standard buffer). The standard colony PCR protocol was as follows: 3 min incubation at 98 °C, followed by 35 cycles composed of a 15 s denaturation step at 98 °C, a 30 s annealing step at 55 °C, and an elongation step at 72 °C for 1 min per kb. After the last cycle, a final elongation was carried out for 5 min at 72 °C. The size of the amplicons were determined using a 0.5% agarose gel. Positive hits from the colony PCR analysis were selected from the agar plate and were used to inoculate 5 mL of growth media and incubated at 37 ^{o}C overnight. Plasmid DNA was recovered from the cells using the Nucleospin Plasmid kit from Macherey Nagel, following the manufacturer's instructions. All plasmid constructs were sequenced by Genome Quebec using the appropriate primers. Upon confirmation of the plasmid sequence, the plasmids were then transformed into chemically competent *E. coli* BL21 (DE3) cells for co-expression studies. Cell lines were stored at -80 ^{o}C until further use.

2.5.3. **Expression and purification**: Single colonies from BL21 (DE3) transformations were used to inoculate overnight starter cultures of LB broth containing the necessary antibiotics and were incubated at 37 ^{o}C. Starter cultures were diluted 100x into fresh LB along with appropriate antibiotic. These cultures were grown at 37 ^{o}C until an OD_{600} of 0.8 was reached, at which point expression of the *sael* genes was induced with a final concentration of 250 μM isopropyl *β*-D-1-

thiogalactopyranoside (IPTG). Cultures were allowed to grow at 37 ^{o}C for an additional 3-4 hours, at which point, cells were pelleted by centrifuging at 8,000 x *g* for 10 minutes at 4 ^{o}C. The cell pellets were stored at -80 ^{o}C until further use.

The His_6-tagged SaelA1 and SaelA2 peptides were purified using immobilized metal affinity chromatography (IMAC). The cell pellets were resuspended in wash buffer (6M guanidinium hydrochloride, 20 mM NaH_2PO_4 pH 7.5 at 25 ^{o}C, 300 mM NaCl, and 30 mM imidazole) and were sonicated for lysis. The lysate was then centrifuged at 16,000 x *g* for 45 minutes to pellet cell debris. The supernatant was then loaded onto a HisTrap FF (5 mL) column from GE Healthcare that had been pre-equilibrated with 5 column volumes (CVs) of wash buffer. After loading the cell lysate, the column was washed with 10 column volumes (CVs) of wash buffer after loading. The peptides were then eluted using 5 CVs of elution buffer (4M guanidinium hydrochloride, 20 mM NaH_2PO_4 pH 7.5 at 25 ^{o}C, 300 mM NaCl, and 30 mM imidazole). The eluate was desalted using a C8 SPE column (30 mL) from Waters using the following procedure. The SPE column was loaded with sample, then washed with 5 CVs of 0.1% TFA solution. The peptide was then eluted using a solution of 80% acetonitrile and 0.1% TFA. The samples were then lyophilized and resuspended in a desired volume of 30% acetonitrile, 0.1% TFA solution. Afterwards, the crude peptide was further purified by RP-HPLC using a C8 semiprep column (SUPELCO). Solvent A was 0.1% TFA solution and solvent B was acetonitrile with 0.1% TFA. The gradient chosen was 30-100% B over 40 minutes. Fractions were screened via MALDI-TOF-MS, and the peptides were lyophilized and stored at -80 ^{o}C until further use.

Given that the *N*-terminal His_6-tag is cleaved from both SaelA1 and SaelA2 upon expression with their cognate LanMs and SaelT, IMAC could not be used to purify the post-translationally modified peptides. As such, we adapted a methanol extraction procedure described by Hegemann et al. (2018). After co-expression and cell pelleting, the cell pellets (~5 g per litre of culture) were resuspended in a minimal volume of water (10 mL water per litre of culture). The suspension was then diluted with methanol to give a final concentration of 80% methanol. This was then incubated at 4 ^{o}C while shaking overnight. The suspension was then centrifuged at 10,000xg for 10 minutes at 4 ^{o}C, and the methanol was removed by rotary evaporation prior to freezing the reaming aqueous phase and lyophilisation. The crude peptide was then resuspended in a solution 50% aacetonitrile with 0.1% TFA and purified by RP-HPLC using a Waters

Symmetry C4 column (4.6x250mm, 5 μm). Solvent A was 0.1% TFA and solvent B was acetonitrile with 0.1% TFA. The gradient chosen was 30-100% B over 40 minutes. The fractions collected were analyzed via MALDI-TOF-MS.

2.5.4. **Matrix assisted laser desorption/ionization mass spectrometry**: All samples for MALDI analysis were prepared by mixing 1 μL of peptide-containing sample 1:1 with a saturated solution of MALDI matrix (α-Cyano-4-hydroxycinnamic acid, HCCA) dissolved in 80% acetonitrile with 0.1% TFA. The peptide and matrix were mixed directly on the stainless steel MALDI target and allowed to air dry prior to analysis. All MALDI analyses were performed on a Bruker Microflex instrument using the following conditions: 20 kV grid voltage, 200 ns delay, 30-50% power setting. Data were recorded in linear positive mode.

2.5.5. **Electrospray ionization mass spectrometry**: All ESI-MS and tandem MS analyses were carried out using a Waters Synapt G2-Si system coupled to an Acquity UPLC chromatography system. Unless otherwise noted, MS data was collected with the following instrument settings: polarity = positive, resolution mode, capillary voltage = 3 kV, cone voltage = 40 V, cone gas flow rate = 50 L/hr, source offset = 80 V, source temperature = 100 ^{o}C, desolvation temperature = 250 ^{o}C. Samples of HPLC-purified Saelα and Saelβ were delivered to the ESI source using a Waters ACQUITY BEH C8 column (1.7μm, 2.1mm x 100 mm). The column was eluted using a linear gradient of solvent A (99.9% water with 0.1% formic acid) and solvent B (99.9% acetonitrile with 0.1% formic acid), where the percentage of solvent B was raised from 3 – 100% over 10 min. The instrument was calibrated using a solution of 1% (w/v) NaI in a solution of 50%-50% water-isopropanol. [Glu1]-fibrinopeptide (GluFib) was used as an external calibrant.

Both Saelα and Saelβ were analyzed using tandem MS in order to further elucidate their respective structures. The LC and source settings were identical to those described above. The ions of interest (4371 m/z for Saelα and 3555 m/z for Saelβ) were transmitted to the collision cell via the quadrupole analyzer. Argon was used as the collision gas, and the collision voltages used were 20-30, 30-40, 45-50 V for Saelα and 20-30 V for Saelβ over an interval of 1 s. The spectra were record on the range of 50-2000 m/z, in continuum mode, with glu-1-fibrinopeptide used as the external standard. The spectra were deconvoluted using MaxEnt3, a function in the MassLynx software

The hypothetical structure enumeration and evaluation (HSEE) method, developed by Zhang et al., was used to determine the possible structures of Saelβ[42]. Using a combination of tandem MS studies and NEM derivatization of Saelβ, the structure of the peptide can be readily determined. In summary, the dehydration of Ser/Thr residues leads to a mass change of -18 Da and NEM alkylation of free cysteine residues leads to a mass change of +125 Da. Since the sites of modification (Cys, Ser and Thr residues) are known, a series of MSMS fragment ions can be predicted. The HSEE method creates a series of MSMS predictions based on constitutional isomers of the lanthipeptide, and then compares the generated fragmentation data to the experimental data and creates a ranking for each structure. The best match is that which gives the highest HSEE score.

2.5.6. **Antimicrobial assays**: Disk diffusion assays were carried out by first plating liquid culture of bacteria (prepared from revived glycerol stocks) onto agar plates (1.5% agar). For *M. luteus*, the medium used for both liquid and solid culturing was ATCC Medium 3. For all other strains, nutrient broth was used. Sterile disks purchased from Fisher were placed onto the agar plates. 10 μL of varying concentrations of Sael*α* or Sael*β* or both (dissolved in 50% acetonitrile) were pipetted onto the disks and allowed to dry before being incubated overnight. For *M. luteus*, the incubation was carried out at 30 ^{o}C whereas all other strains were incubated at 37 ^{o}C. Control disks were spotted with 10 uL of 50% acetonitrile. Liquid culture assays were carried out for *M. luteus* only. Revived glycerol stocks were used to inoculate 10 mL started cultures of *M. luteus*. 200 μL of this culture was diluted to 1.8 mL with fresh ATCC Medium 3. To this, 100 μL of 200 μM of Saelα and Saelβ stock were added respectively, bringing the final concentration of each to 10 μM. For single peptide assays, 100 μL of 50% acetonitrile was supplemented to bring the final volume of culture to 2 mL. The control assay was simply inoculated with 200 μL of 50% acetonitrile.

2.5.7. **Metal binding assays**: Stock solutions of Saelα, Saelβ, $ZnCl_2$, $NiCl_2$ and $CuCl_2$ were prepared in water with 10% DMSO. Assays contained 100 μM of peptide in all cases. They also contained either 100 μM or 1 mM metal. The following mixtures were prepared: (i) Saelα + Zn^{2+}, (ii) Saelα + Ni^{2+}, (iii) Saelα + Cu^{2+}, (iv) Saelβ + Zn^{2+}, (v) Saelβ + Ni^{2+}, (vi) Saelβ + Cu^{2+}, (vii) Saelα + Saelβ + Zn^{2+}, (viii) Saelα + Saelβ + Ni^{2+}, (ix) Saelα + Saelβ + Cu^{2+}. All experiments were carried out in triplicate. After mixing, samples were left to incubate on the benchtop for 10

min, before analysis using a CARY 5000 UV-Vis-near IR spectrophotometer, scanning between 240 and 800 nm.

Chapter 3: Probing The Conformational Dynamics And Structural Properties Of The Class II Lanthipeptide Synthetase HalM2 Using Hydrogen-Deuterium Exchange Mass Spectrometry

3.1. Introduction

While relying on the natural world to provide antibiotics for use is a valid option, as exemplified by genome mining, manipulating natural synthetic machinery can be equally as effective[48]. And while this can take on many forms, protein engineering serves as a promising tool by allowing for the modification of antibiotic biosynthetic pathways to produce novel antibiotics. Given that a biosynthetic pathway can be easily modified via cloning, rational mutation of enzymes allows for the biosynthesis of desired compounds. And with the advancement of cloning techniques, and the facile nature of *E. coli* culturing, production of considerable quantities of novel antibiotics is more feasible than ever.[48] Naturally, a thorough understanding of a protein's structure and conformational dynamics and the role they play in facilitating function is a perquisite for any rational attempts at protein engineering. There are already a significant number of studies in the literature which indicate that the evolution of specific conformational dynamics is critical to function. For instance, the increased flexibility of an enzymatic active site has been associated with increased promiscuity[49]. Moreover, more global changes to protein structure can introduce features, such as the introduction of new folds, that alter dynamics so significantly that enzymes can develop altogether new functions.[49]

As mentioned above, lantibiotics present a potentially promising class of antibiotics for rational manipulation. Being able to modify them to improve their chances of becoming full scale drugs presents a challenge, however, in part due to the complexity of the multistep reaction sequences needed for lanthipeptide maturation and the iterative activity of the modifying enzymes. To that end, studying the conformational dynamics of lanthipeptide synthetases and their intermolecular interactions with the precursor peptide substrate may provide valuable insights into how these interesting enzymes carry out their multiple functions – in time and space – so as to produce final products with specific structures. Indeed, kinetic studies performed previously by Thibodeaux et. al. have provided numerous lines of evidence that conformational changes within the LanM:LanA Michaelis complex likely gate the steady state rates of the dehydration and cyclization reactions needed for lanthipeptide maturation.[50] Thus, conformational changes appear to be the most critical factor that governs the biosynthetic sequence in LanM systems.

Class II lanthipeptide synthetases (LanMs) are bi-domain proteins that install thioether rings on their substrates through an ATP dependent dehydration of Ser/Thr residues followed by an intramolecular cyclization event (Fig. 3.1). The *N*-terminal dehydratase domain uses ATP to first phosphorylate Ser/Thr residues, which is then followed by an elimination of the phosphate group to produce Dha/Dhb residues. The *C*-terminal cyclase domain, uses a tightly bound Zn^{2+} ion to facilitate the nucleophilic attack of the thiol moiety of cysteine residues on nearby Dha/Dhb, residues thus forming lanthionine or methyllanthionine rings[26].

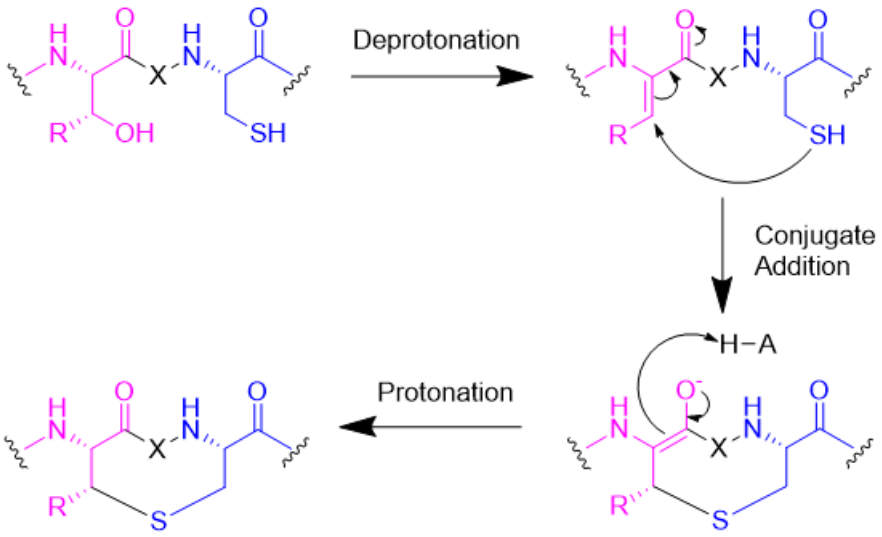


Figure 3.1. The mechanism of ring installation in lanthipeptide systems. X represents any number of amino acids.

While the chemical mechanism of ring formation is well known, the structural dynamics involved in facilitating the reaction and the observed regioselectivity of the post-translational modifications remain largely unknown. With each LanM-catalyzed reaction, the structure of the maturing peptide intermediate changes. As the structure of the intermediate evolves, the intermolecular interactions between the intermediate and the synthetase must also change. The dehydration of Ser/Thr residues alters the local stereoelectronics of the peptide backbone, and the installation of thioether rings rigidifies the peptide backbone. Both of these modifications could have significant effects on the binding energy of the maturing intermediate in the dehydratase and cyclase active sites of the LanM. How does the LanM:LanA Michaelis complex sample the various conformational states in a manner that allows these enzyme systems to install modification in a defined, orderly manner, such that biosynthetic fidelity is maintained?

The dearth of high-resolution structures for LanMs (and for RiPP biosynthetic enzymes in general) has presented a significant barrier to assessing the importance of protein-protein interactions and conformational changes in these systems. To date, CylM remains the only LanM for which a crystal structure has been solved, and there are not yet any crystal structures with a LanM bound to its cognate LanA precursor peptide. Thus, very little is known regarding the peptide binding mode. Nevertheless, the CylM structure resolved by Dong et al.[27], was still highly informative (Fig. 3.2). While the dehydratase domain of LanMs shows no significant amino acid sequence homology with any other class of protein, the CylM structure clearly shows that this domain is structurally homologous to lipid kinase enzymes – a result that is in line with ATP-dependence of the LanM-catalyzed dehydration reaction. Interestingly, the kinase activation loop of CylM is ordered in the crystal structure, whereas it is normally *disordered* in lipid kinases. This observation could explain the previously observed low level background rate of ATP hydrolysis catalyzed by LanMs. As expected based on sequence homology to other proteins, the cyclase domain shared structural homology with enzymes that catalyze Zn-dependent thiol alkylation reactions such as human protein farnesyl transferase and the class I lanthipeptide cyclases, LanC. Importantly, the active sites of the dehydratase and cyclase domains are separated by nearly 50 Å, clearly necessitating some sort of conformational sampling within the LanM:LanA complex in order to achieve complete modification of the substrate.

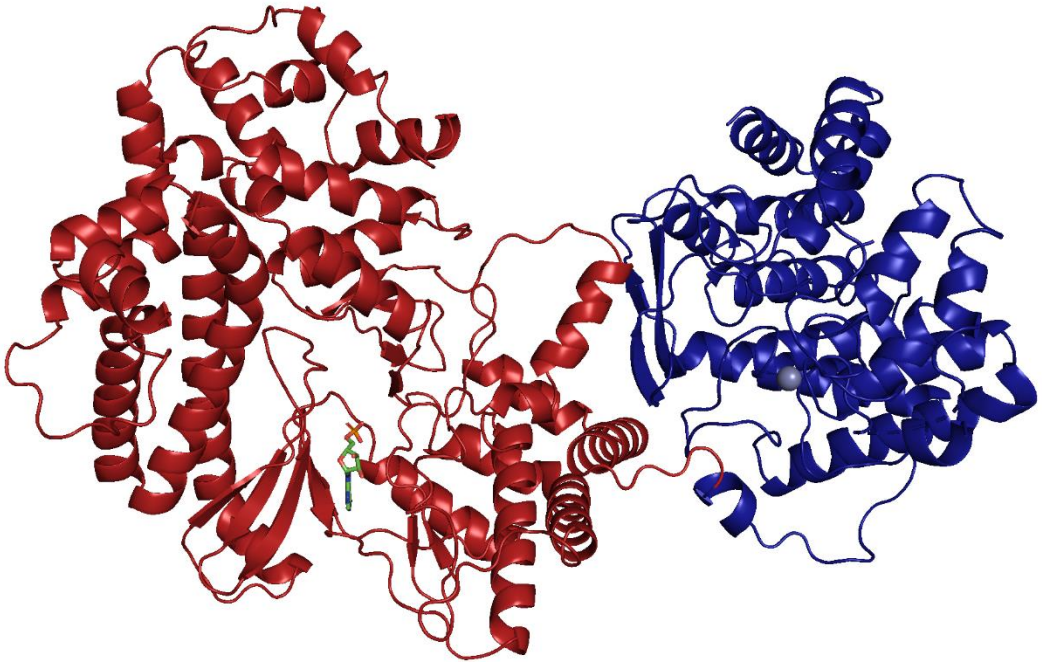


Figure 3.2. The crystal structure of CylM, as reported by Dong et al. The dehydratase domain is marked in red and the cyclase domain in blue.[27]

The crystal structure of CylM also supported prior mechanistic studies of LctM, involved in the synthesis of the lantibiotic lacticin 481, where the residues involved in dehydration and cyclization reactions were identified via multiple amino acid sequence alignments and their roles were assessed by site directed mutagenesis and biochemical assays.[51] In these studies, residues Asp242 and Asp259 (Asp347 and Asp364 in CylM respectively) were identified as being important for phosphorylation, while residues Arg399 and Thr405 (Arg506 and Thr512 in CylM respectively) were important for elimination of the phosphate (Fig. 3.3). Moreover, LanM cyclase domains typically coordinate Zn^{2+} via a His and two Cys residues, which are highly conserved. In LctM these residues were identified as Cys781, Cys836 and His837 via mutagenesis.[26, 52]

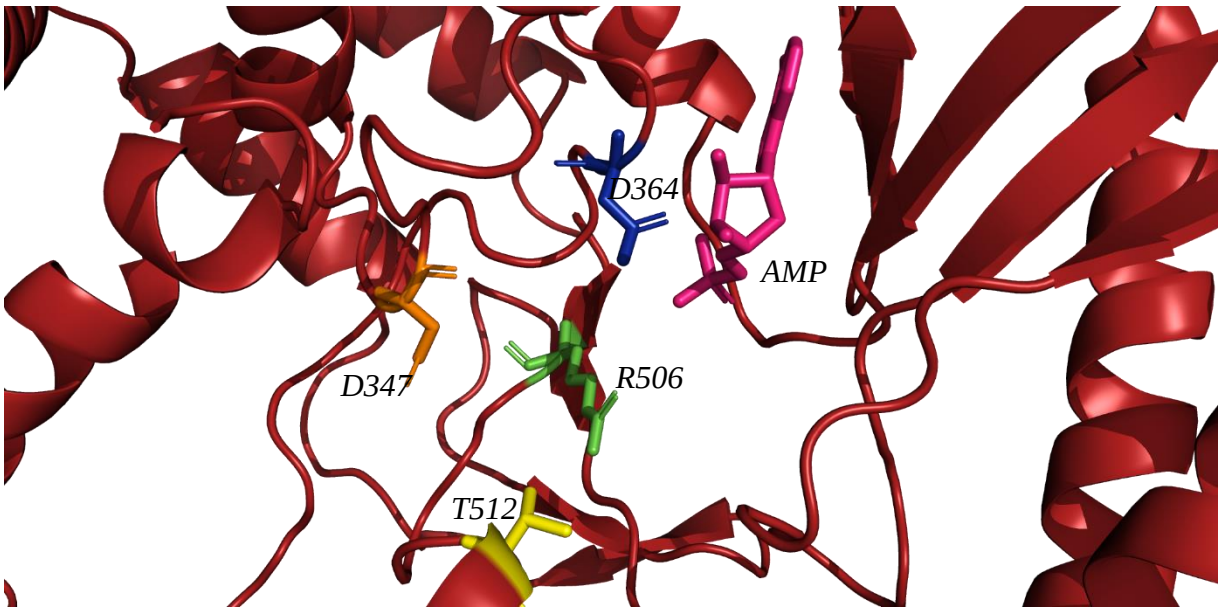


Figure 3.3. A close up of the dehydratase domain active site in CylM. Residues Asp347 (orange) and Asp364 (blue) are believed to be involved in phosphorylation of Ser/Thr residues. Residues Arg506 (green) and Thr512 (yellow) are believed to be involved in phosphate elimination. AMP is highlighted in pink.

In addition to LctM, HalM2 represents one of the best-studied LanM enzymes.[50] The protein is involved in the synthesis of Halβ, one of the two peptides of the two-component lantibiotic haloduracin, produced by *Bacillus halodurans*. The HalM2 substrate peptide (HalA2), the precursor to Halβ, is modified to include a total of 7 dehydrations and 4 cyclizations. Despite the increased complexity of this LanM system relative to LctM, the HalA2 precursor peptide is considerably more soluble in aqueous solution than LctA, making the HalM2/HalA2 system a much more tractable model for LanM mechanistic studies. Kinetic studies on HalM2 have shown that the multiple modifications are installed in an orderly fashion with an overall *N*- to *C*-terminal directionality. Namely, residues in the *N*-terminal region of the HalA2 core peptide are dehydrated/cyclized prior to residues in the *C*-terminal region (Fig. 3.4). Such directionality in enzymes with iterative activity is usually associated with motor proteins – enzymes that harness the energy of ATP hydrolysis to drive directional motion and conformational changes. While LanMs indeed consume ATP during the dehydration reaction, LctM has been shown to install the correct set of thioether rings into chemically synthesized, dehydrated substrates in the absence of ATP.[52] Thus, motor protein function does not appear to be absolutely required for LanM activity.

However, whether or not these systems exploit the energy generated by phosphoryl transfer to enhance the kinetic coupling of dehydration and cyclization events has not been tested.

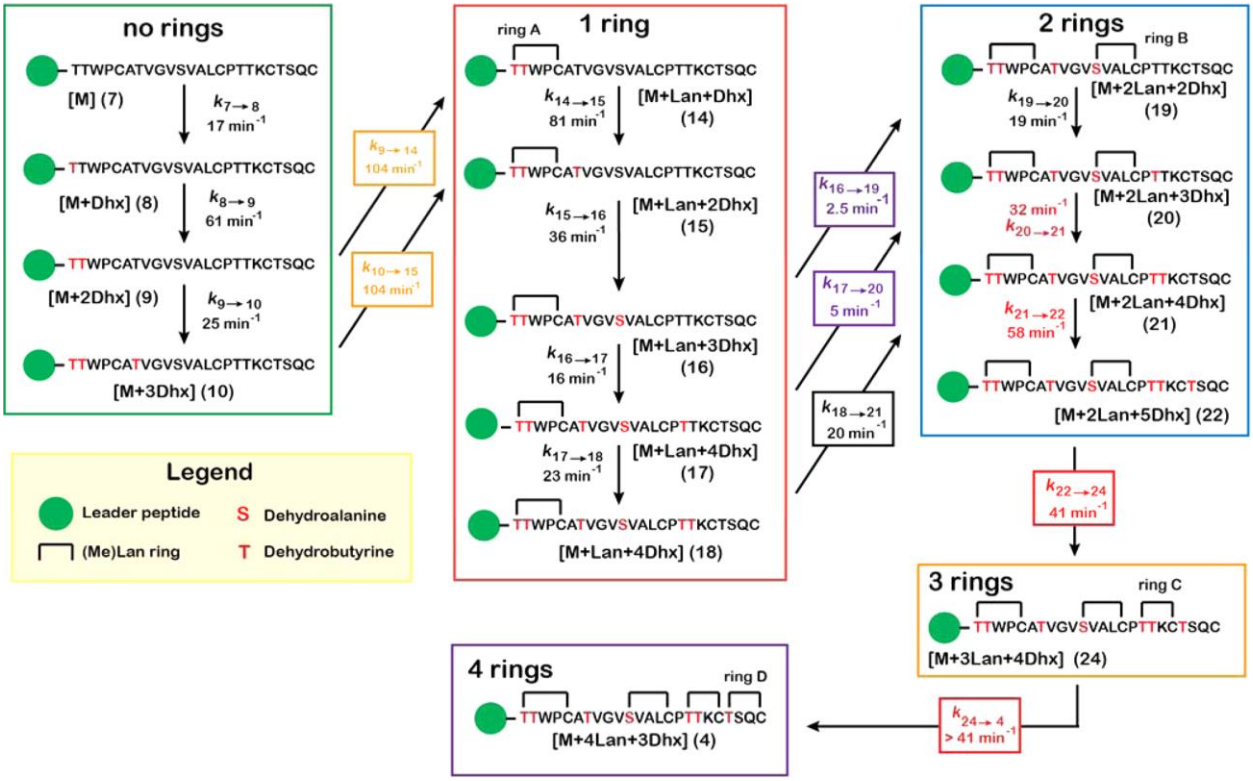


Figure 3.4. The kinetic model of HalA2 maturation proposed by Thibodeaux et al. Adapted with permission. [50]

Another interesting observation derived from HalM2 kinetic studies is that the overall reaction sequence contains both processive and distributive segments. In a processive kinetic mechanism, an enzyme with iterative activity catalyzes multiple rounds of a chemical reaction without dissociating from its substrate. The canonical example of a processive enzyme is DNA polymerase, which catalyzes many thousands of nucleotide incorporation events within a single binding event to the template DNA. Like DNA polymerases, processive enzymes are typically motor proteins that convert the chemical energy of phosphodiester bonds into protein conformational changes that drive some sort of directional motion of the catalyst. In contrast, in a fully distributive kinetic mechanism, the enzyme dissociates from the substrate after each round of catalysis[50]. HalM2 seems to exhibit both patterns during the maturation of HalA2. Namely the first two dehydrations (of Thr1 and Thr2) and the first cyclization to form ring A occur at rates faster than the HalA2 dissociation rate as measured by single molecule fluorescence (Fig. 3.4).[50]

Thus, to some extent, these initial reactions will be coupled, allowing HalM2 to install these modifications in a processive manner within a single HalM2/HalA2 binding event. Similarly, after the formation of thioether ring B (the kinetic bottle neck in the overall pathway), the final dehydration of Thr21 and the final two cyclizations also occur with fast rates that could allow their installation to be processive. In contrast, certain intermediates (most notably the singly-cyclized intermediates that accumulate prior to the rate limiting formation of ring B) accumulate to concentrations that exceed the enzyme concentration used in the assay – a clear indication of distributive kinetic behavior.

Thibodeaux et al. also proposed that conformational changes are likely a major factor in determining LanM kinetic parameters.[50] Namely, the rates of ring installation between HalM2 and ProcM (a LanM with 30 natural substrates) differ significantly, with HalM2 processing its substrate to completion in 20 minutes under the conditions used, whereas ProcM processes its substrate peptide on the scale of hours. Moreover, in HalM2, the installation of the second ring (ring B), a Lan ring, is the slowest step in the pathway. This is curious because the three other rings installed in HalA2 are MeLan rings. Given the added steric hindrance introduced by the additional methyl group in Dhb, these rings should be installed at lower rates, as has been noted in previous model studies[26]. This would suggest something other than nucleophilic attack of the Cys thiol on the Dha/Dhb residue (i.e. a conformational change) is the rate determining step during the cyclization reaction. Finally, mutagenesis of Cys11, which is involved in the formation of the ring B, adversely affects the ability of HalM2 to form the remaining rings.[53] A possible explanation for this is that the formation of the rings brings the residues that are to be modified closer to the leader peptide for processing, and also reduces the dynamic nature of the substrate, allowing for better docking in the active site of HalM2. All of this provides considerable evidence for the importance of conformational dynamics within the HalM2-HalA2 complex.

Given the mounting evidence that conformational changes play a critical role in the functional properties of LanM systems, and the relative dearth of structural studies on these enzymes, additional methods capable of characterizing conformational dynamics are needed. To that end, we describe in this thesis the development of a hydrogen-deuterium exchange mass spectrometry (HDX-MS) platform as a versatile tool to probe local protein structural dynamics.[54] The typical workflow for a bottom-up HDX experiment is shown in Figure 3.5. In this workflow,

the protein is first diluted into buffer prepared in D_2O, which triggers solvent H/D exchange of NH, OH, and SH bonds in the protein. The exchange rate of the amide protons in the peptide backbone is a reflection of the solvent accessibility and the hydrogen bonding to that amide. Thus, segments of the protein that are highly structured as α-helices, β-sheets, or other structural elements tend to exchange more slowly than backbone amides from unstructured regions of the protein where the amide N-H bond is readily accessible to the solvent. As such, HDX can provide information on how protein structure is changing under different conditions.

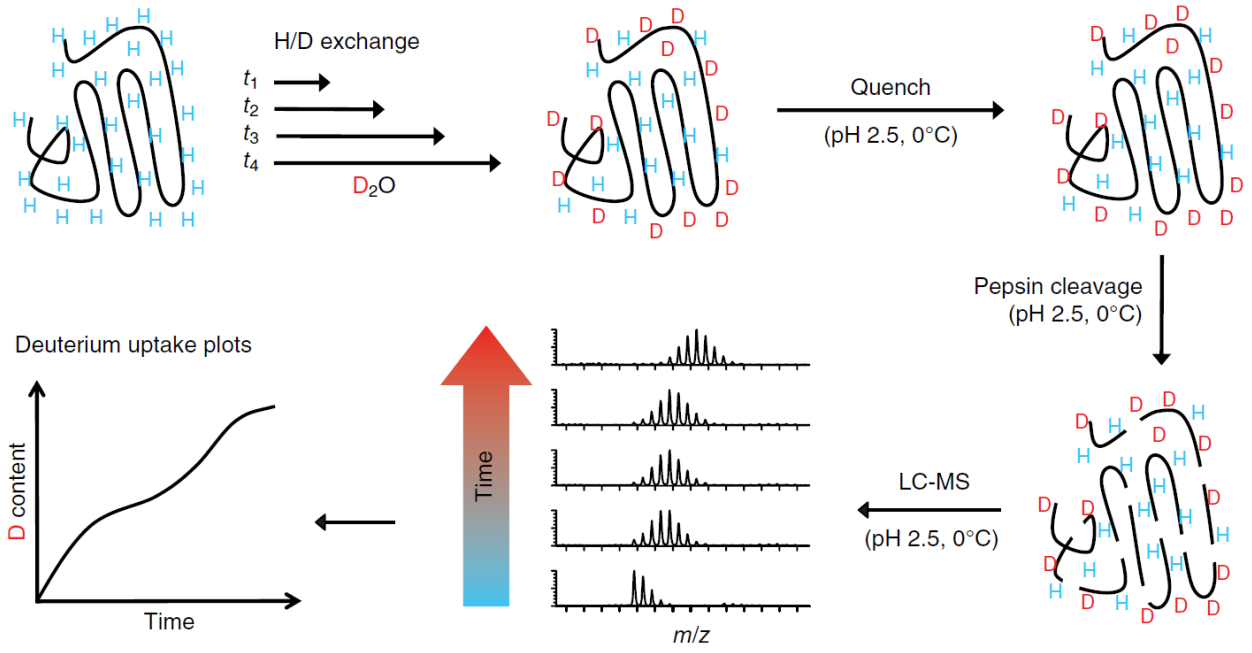


Figure 3.5. A standard HDX-MS workflow. Adapted with permission.[55] Our modified workflow involves freezing samples with liquid nitrogen for storage before thawing and digesting with pepsin. Moreover, digestion occurs at 15 °C.

In a typical HDX-MS experiment (Fig. 3.5), the protein is allowed to exchange in solvent D_2O for variable lengths of time ranging from seconds to several hours. The exchange reaction is then quenched by acidifying the reaction mixture to a value near pH ~ 2.5 (where the rate of amide solvent exchange is at its minimum) and by flash freezing the sample in liquid nitrogen. These treatments are necessary to reduce back exchange (and thus, retain deuterium label) during the subsequent steps in the analysis, all of which are carried out in protiated solvents. In the next phase of the pipeline, the deuterated protein samples are proteolytically digested into peptides. This "bottom-up" approach ultimately enables localized measurements of HD exchange. Given that the solution must remain acidic in order to reduce back exchange, the proteolysis step is

limited to the use of acid stable proteases, most commonly pepsin. This enzyme has very relaxed specificity for peptide bond cleavage and tends to generate multiple overlapping peptides that provide significant coverage of the protein sequence.

After pepsin digestion (which in our workflow is carried out on a UPLC column), the complex mixture of deuterated peptides is then separated on a C18 reverse phase UPLC column (operated at 0.4 °C to minimize back exchange) and injected into the source of an electrospray ionization (ESI) mass spectrometer. Despite the C18 separation (which separates peptides on the basis of polarity) and the high resolution of our time-of-flight mass spectrometer (which separates peptides based on their mass to charge ratios), these samples still contain thousands of signals and are incredibly complex. In order to more confidently identify as many peptides as possible, we also employ a gas phase ion mobility separation, wherein peptide ions are passed at low potential into a chamber filled with a neutral carrier gas (nitrogen). Under the gentle electric field, larger ions undergo more collisions with the carrier gas and hence, move through the ion mobility cell more slowly. In contrast, more compact ions move through the chamber more quickly. As such, ion mobility can separate isobaric species having the same mass and chromatographic retention time on the basis of their shapes, thus introducing a new layer of separation into the workflow and dramatically increasing our ability to identify peptides for HDX-MS analysis. In the final stage of the pipeline, signals for the peptic peptides derived from the enzyme are identified, extracted from the mass spectra, and the time-dependent mass increase in each peptide (due to deuterium exchange) is calculated.

As such, the bottom-up HDX-MS workflow provides a vast amount of information on local structural dynamics in proteins in a manner that does not require labeling of the protein with fluorophores or other reporter groups that might alter the native structural properties of the protein. In addition, the HDX reaction is carried out in solution at physiological pH and ionic strength, under conditions where the enzymatic activity can be directly measured. This ensures that the HDX-MS measurement is capturing equilibrium conformational dynamics that are relevant to function. As we will illustrate below, the platform is most powerful as a tool to compare HD exchange differences between different enzymatic states (*e.g.* between the free enzyme and enzyme:substrate complex), allowing the relative perturbations between different conditions to be rapidly identified. Using this highly tractable platform, we report the use of HDX-MS to probe

the conformational dynamics of HalM2 upon binding to AMP-PNP (a non-hydrolyzable ATP analogue, Fig. 3.6) and to the HalA2 precursor peptide substrate. From these initial studies, we have identified for the first time a potential binding site for HalA2 as well as a number of previously overlooked HalM2 structural elements that undergoing significant conformational changes upon binding AMP-PNP and/or HalA2. These conformational changes could be reporting on catalytically competent Michaelis complexes and providing the first glimpse of functionally relevant conformational changes in LanM systems.

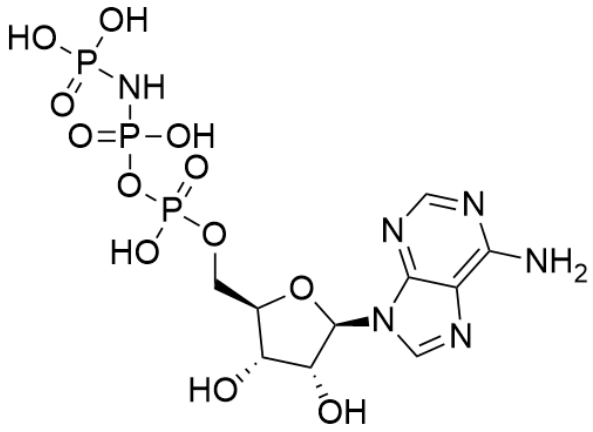
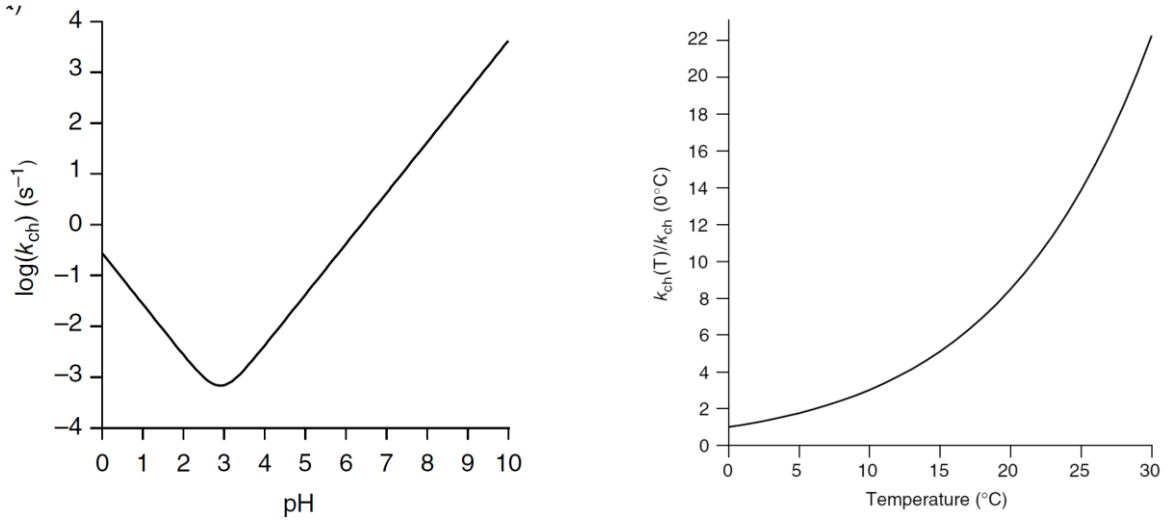


Figure 3.6. The structure of AMP-PNP.

3.2. Results and Discussion:

3.2.1. **Assay development and sample preparation considerations**: A typical HDX experiment requires significant fine tuning in order to minimize back exchange and to enhance the reproducibility and information content of the data. Steps in the process that need to be optimized include the exchange reaction, the quenching step, the pepsin digestion of the protein, the C18 chromatography step and the MS analysis. With regards to the exchange reaction, a major consideration is the buffer to be used, as ionic strength is a factor in exchange. Fortunately, most commonly used buffers are HDX compatible.[55] For this study, the buffer used was adapted from Thibodeaux et al., wherein the activity of HalM2 was studied using kinetics and MS[50]. To maximize the deuterium content in the exchange reaction, all buffer components and HalA2 were deuterated by repeated cycles of lyophilization and re-dissolving in 100% D_2O. Moreover, concentrated stocks (~ 200 μM) of HalM2 in protiated storage buffer were prepared by centrifugal filtration such that a minimal quantity of protons would be added to the exchange reaction upon

dilution of HalM2 into the mixture. All in all, this allowed us to achieve a deuterium content of approximately 98% in our exchange reactions. As mentioned earlier, the exchange of backbone amide protons is minimized at pH 2-3 (Fig. 3.7 A). The rate of back exchange increases an order of magnitude for every 1 unit pH change in either direction of the pH 2.5 minimum. Thus, the final pH of the quenched solutions had to be as close to pH 2.5 as possible and the assays had to be performed in a careful, standardized manner so as to prevent systematic differences in back exchange between different samples, which would adversely affect our ability to interpret the exchange data across all our data sets. To that end, in designing the workflow, a mock HDX reaction containing all buffer components was prepared, and the quench buffer (100 mM K_2PO_4, pH 2.0) was titrated in triplicate. The exact volume of quench buffer required to adjust the pH of the final solution to 2.5 was then recorded, and that volume was then used to quench all future assays.



A B

Figure 3.7. (A) A plot of the rate of deuterium exchange for peptide backbone amide moieties as a function of pH.. (B) The rate of deuterium exchange as a function of temperature. Adapted with permission. [55]

The amide exchange rate is also very sensitive to temperature (Fig. 3.7 B). In order to minimize back exchange, the mixture must be cooled considerably and the low temperature must be maintained during the on-column pepsin digestion and subsequent C18 chromatography[56]. In all of our assays, quenched reaction aliquots were immediately flash frozen in liquid nitrogen, then stored at -80 °C until pepsin digestion and LC-MS analysis. On the day of analysis, the samples were thawed over a precisely-timed 2 min interval, prior to injection in the LC-MS system. Errors

in the timing and temperature of this step will ultimately reduce the precision of the HDX measurement. As the pepsin protease activity is also affected by the temperature and digestion time, a balance between proteolytic activity and back exchange must be met. Allowing long periods of pepsin digestion will considerably reduce resolution by producing peptide fragments that are too small to be confidently assigned. On the other hand, large peptic fragments produced by poor digestion will reduce the spatial resolution of the measurement by limiting our ability to localize deuterium uptake to specific regions of HalM2. In addition, the amount of time that each sample is digested must be the same, as differences will lead to different digestion patterns and poor reproducibility. To address these issues, a UPLC column with immobilized pepsin was used so as to consistently digest all HDX samples. As shown below, we ultimately found that a 3 min pepsin digestion at 15 °C using a 0.1% aqueous formic acid solvent (pH 2.5) provided sufficient coverage of the HalM2 sequence, while maintaining enough of the deuterium label to allow reproducible HDX quantitation for most peptides. Following pepsin digestion, the proteolyzed mixture was separated with a 10 min gradient on a BEH C18 UPLC column at 0.4 °C with the following gradient: Gradient of 3-100% acetonitrile with 0.1% formic acid.

3.2.2. **Collection of mass spectrometry data***:* Following elution from the BEH C18 UPLC column, peptides were introduced into the ESI source of a Waters Synapt G2-Si mass spectrometer (Fig. 3.8) and were ionized in positive ion mode using typical source temperature and desolvation settings (see Methods). The instrument was operated without quadrupole selection, so as to allow as many ions as possible to reach the ion mobility cell and the time-of-flight (TOF) detector. The total ion chromatogram for a typical digested HalM2 sample is shown in figure 3.9. Each time point in this chromatogram is a mass spectrum that shows all of the peptide ions within the detection window (m/z = 100-2000) that eluted from the C18 column at that time. The spectrum from a single 1 s MS scan at 6 min is shown in Figure 3.10, illustrating the numerous species detected. The sum of the intensities of all of these ions is used to generate the y-axis scale in the total ion chromatogram (Fig. 3.9). Each time point contains a different mass spectrum, with a different set of peptide ion peaks derived from the digestion of HalM2 (Fig. 3.10).

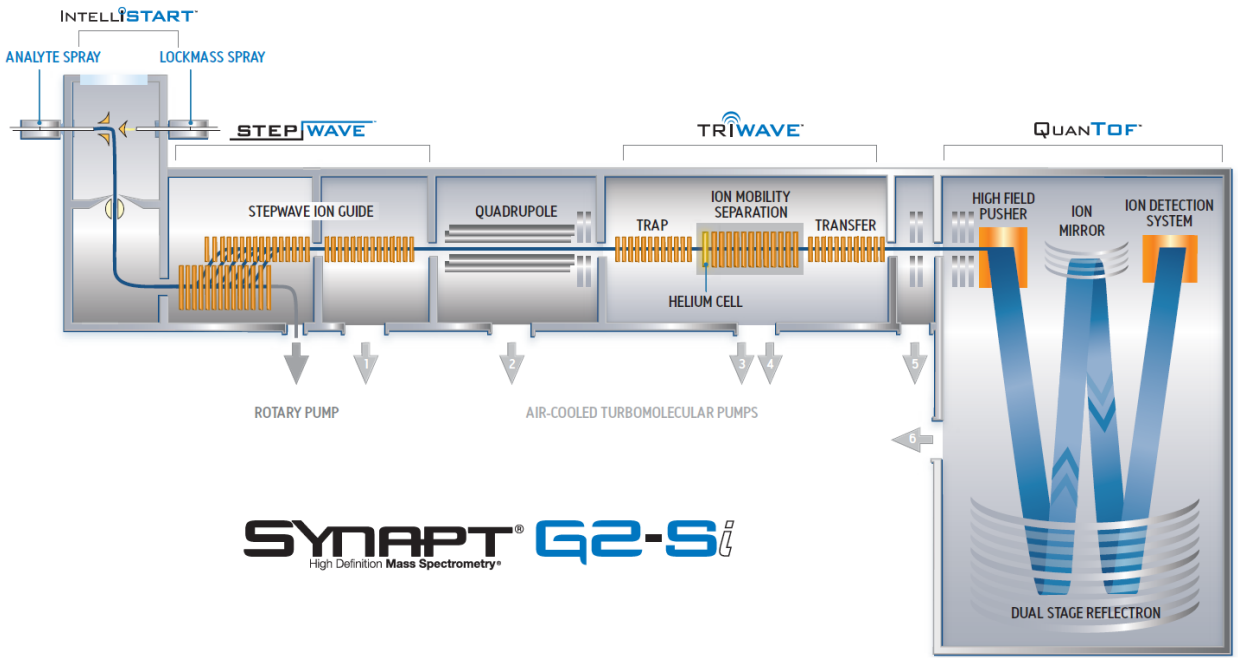


Figure 3.8. Schematic representation of the Waters Synapt G2-Si mass spectrometer.

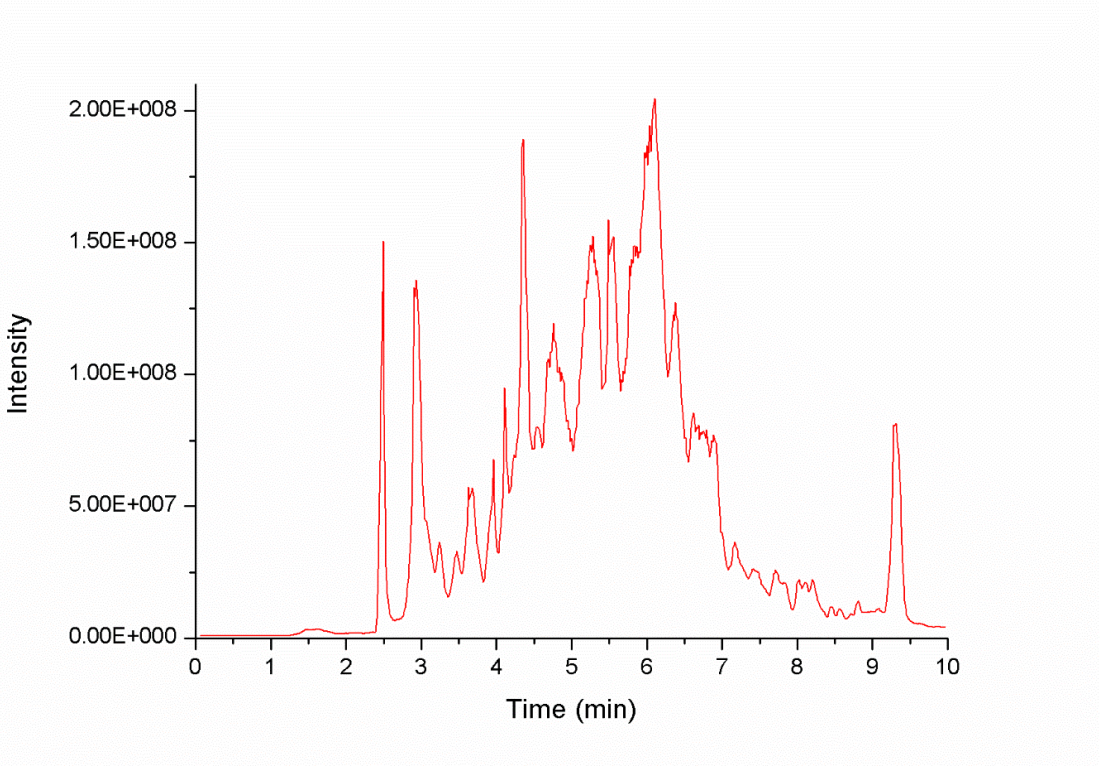


Figure 3.9. A representative total ion chromatogram (TIC) from a typical HDX sample. Most of the peptides elute off the BEH C18 UPLC column between 3-9 minutes.

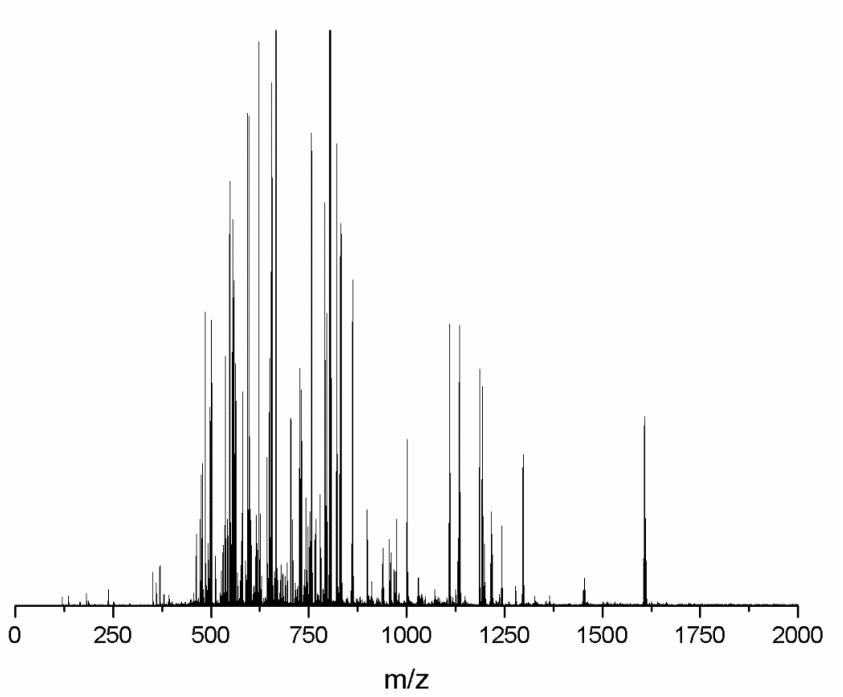


Figure 3.10. An example of a mass spectrum extracted from a single time point the TIC, 6 minutes into the elution gradient. The majority of these signals represent HalM2-derived peptic peptides – illustrating the vast amount of signals that are detected by this method.

After exiting the quadrupole, the peptide ions pass through the trap region of the ion mobility cell and enter the IMS region, where the precursor (parent) ions were separated by travelling wave ion mobility using nitrogen buffer gas. Thus, at every time point in the chromatogram, an additional gas phase separation was performed to help resolve peptides that coincidentally had the same C18 retention times and overlapping isotope distributions in the m/z dimension. As can be seen from the representative mobiligram in Figure 3.11, the gas phase separation indeed resolves peptides that have otherwise overlapping m/z signals. Finally, after leaving the IMS region, the mobility-separated parent ions enter the transfer region of the mobility cell, where the ions were subjected to alternating high and low collision energy regimes. The energy is manipulated by altering the potential across the transfer cell. The higher the potential, the more a given ion will accelerate in the cell and the higher its kinetic energy will be. The cell is additionally filled with argon collision gas to collisionally activate the precursor ions. In the low energy regime, the precursor ions traverse the cell without fragmentation and reach the TOF detector intact. In the high energy regime, multiple collisions with the argon gas lead to fragmentation of the precursor ions into daughter (product) ions. Thus, by alternating the low and

high energy cycles, we can obtain time-correlated m/z measurements of the parent ions and their fragmentation products, which allows highly confident assignment of parent ions to specific amino acid sequences derived from HalM2.

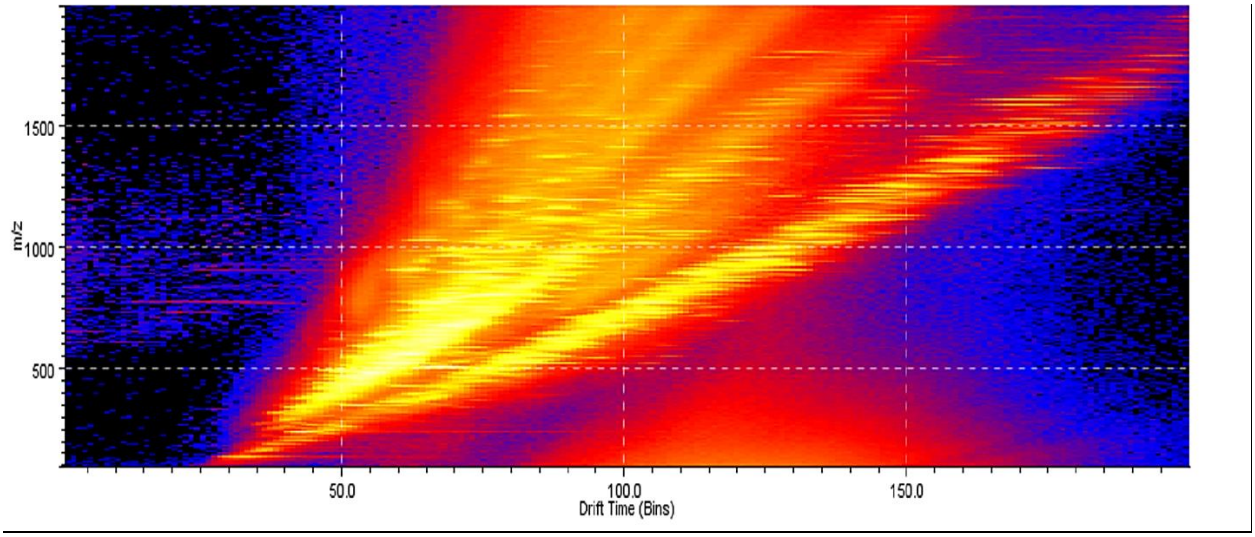


Figure 3.11. An example of a mobiligram recorded during an HDX experiment, wherein the x-axis is the drift time (the time spent in the IM chamber) and y-axis is the m/z value of the ions detected. As can be seen, most ions travel through the chamber between 50 and 150 bins (approximately 10-30 ms).

3.2.3. **Processing of MS data***:* In order to measure the deuterium content of HalM2-derived peptides, we first needed to generate a peak list of all of the peptides that were reproducibly detected by our workflow. To do this, we prepared samples containing HalM2 in triplicate in protiated buffer and subjected these reference samples to the same procedure used for the workup of the deuterium exchanged samples. The raw MS data from these protiated reference samples was then uploaded to the ProteinLynx Global Server (PLGS) software for analysis. The software scans the raw data for peaks associated with expected peptides masses. These expected masses are derived from an *in silico* digestion of the known HalM2 amino acid sequence. Putative spectral assignments are then scored by a variety of user-defined criteria such as the ppm error between the expected and observed m/z, the difference in chromatographic retention time (between parent and daughter ions), the difference in ion mobility drift time (between parent and daughter ions), and the number of matched MSMS daughter ions. After the initial assignment by PLGS, the PLGS output (which is basically a peak list with stats associated with each peak) is uploaded into DynamX 3.0 for additional thresholding. Here, we restricted the final HalM2 peptide list to include only those peptides within 5 ppm of their theoretical value, which generated at least 3 daughter

ions, and that were found in all three replicate reference samples. This resulted in a final list of 224 peptides that covered ~87% of the HalM2 sequence with 2.77 redundancy (Fig. 3.12).

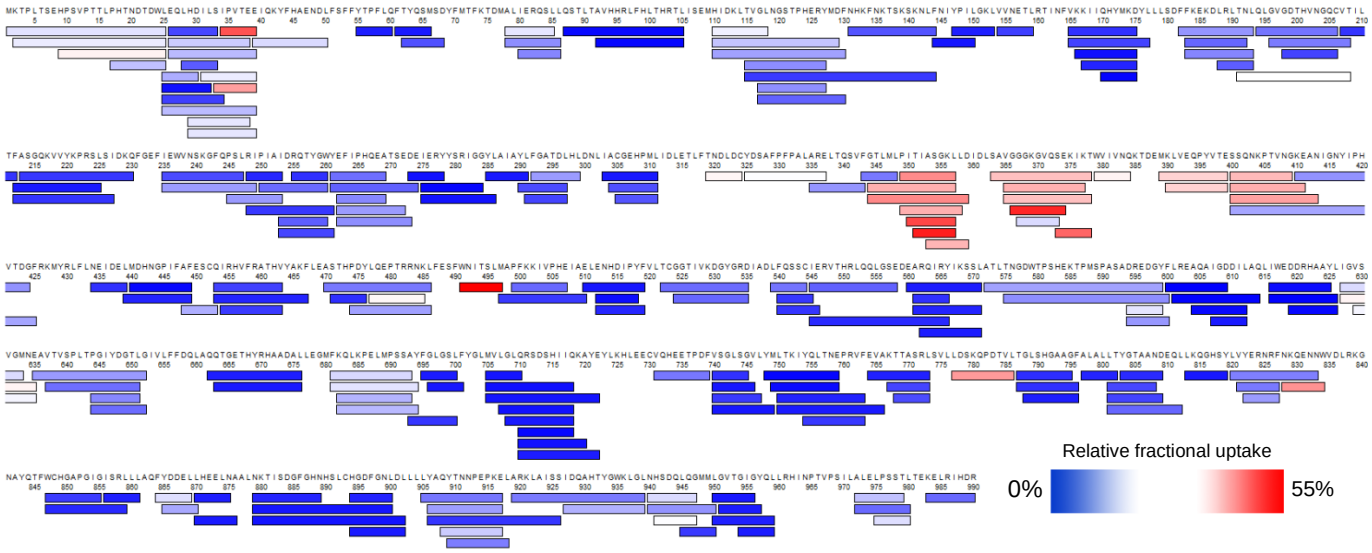


Figure 3.12. The coverage map recorded for apo HalM2. The 5 min time point was used for this illustration. Each bar represents a HalM2-derived peptide that was detected in each of three replicate samples with 5 ppm mass accuracy and that produced at least 3 fragments ions by MSMS. The color scale indicates the extent of deuterium uptake (number of exchanged amide groups normalized by the total number of exchangeable sites on the peptide). Maximal fraction uptake was 55% and sequence coverage is ~87%.

3.2.4. **Determining deuterium content**: Having settled on a final list of confidently identifiable HalM2-derived peptides, we next employed DynamX 3.0 to quantify the deuterium content of each peptide in every detectable charge state at each exchange time point (0.25, 0.5, 1, 5, 30, 60, and 240 min) for every enzymatic "state" investigated ([HalM2], [HalM2+AMP-PNP], and [HalM2+HalA2+AMP-PNP]). DynamX analyzes the isotopic distribution of peptides from the processed MS data, and uses the weighted average of the peak distribution to determine the number of deuteria that have been exchanged into each peptide. This value is then normalized by the total number of exchangeable sites on that peptide to calculate a relative fractional uptake (RFU). Unless otherwise noted, this is how we will express all HD exchange values reported in this study. Figure 3.13 shows a typical output from DynamX, where the time dependent change in the isotopic distribution of a given peptide is clearly visible in the stacked spectral plot, and the quantified time-dependent exchange values are observable for each "state" in the uptake plot (Fig. 3.16). Note the very good precision in the replicate RFU determinations at each time point (we typically find a relative standard error of less than 10% of the mean for each peptide at every time point in

all states), which enables even very slight differences in the RFUs between different states to be analyzed with confidence. It should be noted that while DynamX 3.0 performs admirably in the automation of most of this quantitation, the data still requires a significant amount of curation (usually about a week per data set) in order to most confidently extract mechanistically relevant information and to ensure that assignments and RFU values are determined appropriately. This curation is a rather straightforward process that is greatly facilitated by the interactive and dynamic workflow windows provided by DynamX 3.0. During this process, we check that each of the isotope peaks is properly quantified, that the reference and deuterated peptides have matching retention and drift times, and that all peptide charge states are assigned properly. If all of these criteria are not met, the peptide is removed from the analysis. The time-dependent deuterium uptake plots for several important peptides are shown in Figure 3.14.

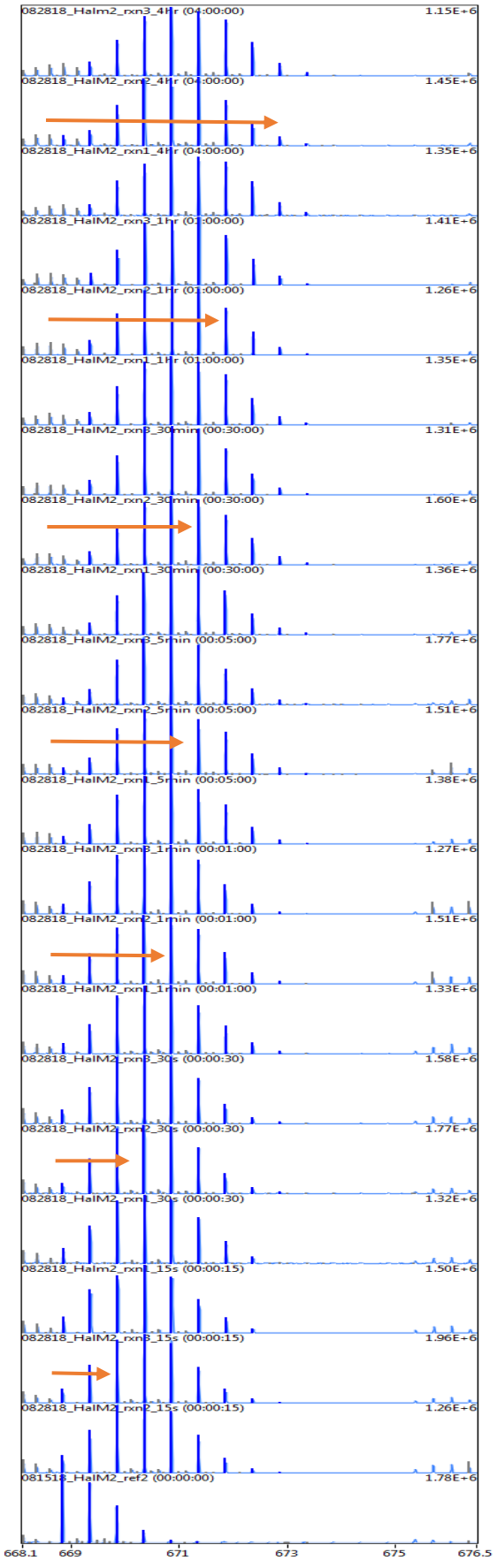
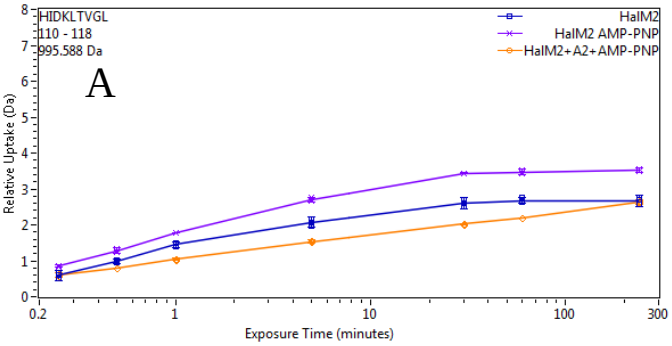
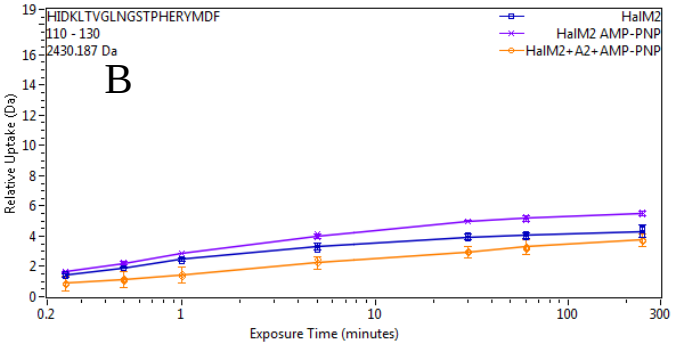

082818_HalM2_rxn3_4Hr (04:00:00) 1.15E+6
082818_HalM2_rxn2_4Hr (04:00:00) 1.45E+6
082818_HalM2_rxn1_4Hr (04:00:00) 1.35E+6
082818_HalM2_rxn3_1Hr (01:00:00) 1.41E+6
082818_HalM2_rxn2_1Hr (01:00:00) 1.26E+6
082818_HalM2_rxn1_1Hr (01:00:00) 1.35E+6
082818_HalM2_rxn3_30min (00:30:00) 1.31E+6
082818_HalM2_rxn2_30min (00:30:00) 1.60E+6
082818_HalM2_rxn1_30min (00:30:00) 1.36E+6
082818_HalM2_rxn3_5min (00:05:00) 1.77E+6
082818_HalM2_rxn2_5min (00:05:00) 1.51E+6
082818_HalM2_rxn1_5min (00:05:00) 1.38E+6
082818_HalM2_rxn3_1min (00:01:00) 1.27E+6
082818_HalM2_rxn2_1min (00:01:00) 1.51E+6
082818_HalM2_rxn1_1min (00:01:00) 1.33E+6
082818_HalM2_rxn3_30s (00:00:30) 1.58E+6
082818_HalM2_rxn2_30s (00:00:30) 1.77E+6
082818_HalM2_rxn1_30s (00:00:30) 1.32E+6
082818_HalM2_rxn1_15s (00:00:15) 1.50E+6
082818_HalM2_rxn3_15s (00:00:15) 1.96E+6
082818_HalM2_rxn2_15s (00:00:15) 1.26E+6
081518_HalM2_ref2 (00:00:00) 1.78E+6
668.1
669
671
673
675
676.5
m/z

Figure 3.13. An example of a stacked spectral plot for peptide M389-E399 in the 3+ charge state, clearly showing a shift towards higher m/z at longer exchange times (from bottom to top), indicating increased deuterium uptake.

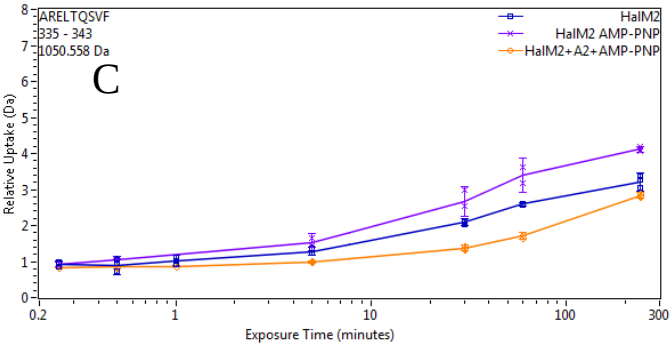
3.2.5. HalM2 is a conformationally dynamic protein:



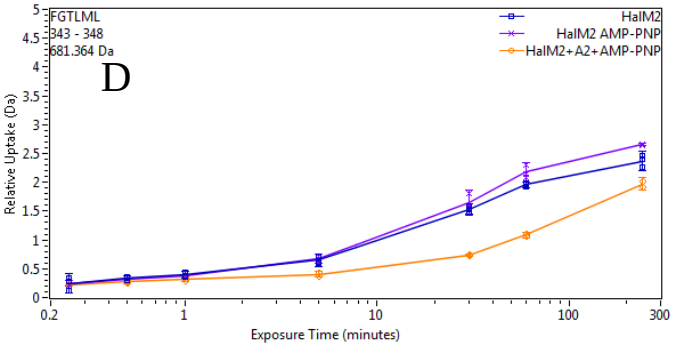
H110-L118



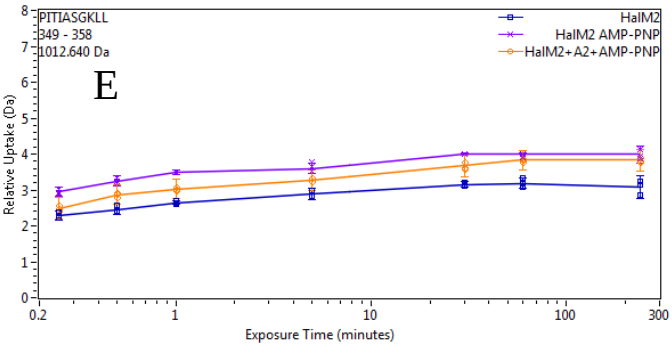
H110-F130



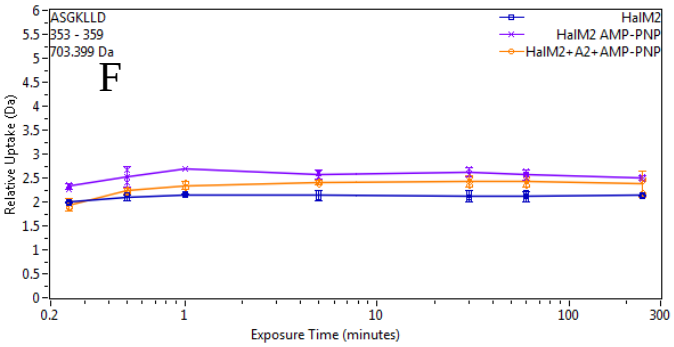
A335-F343



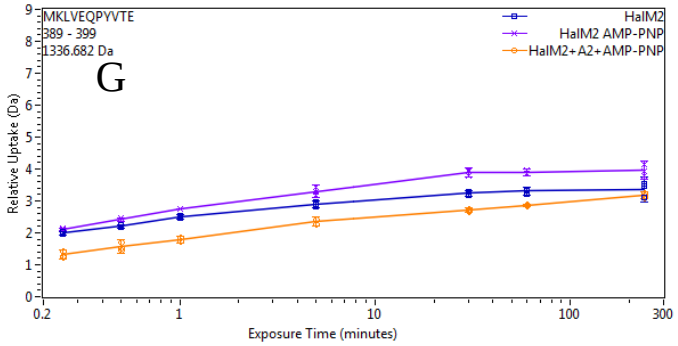
F343-L348



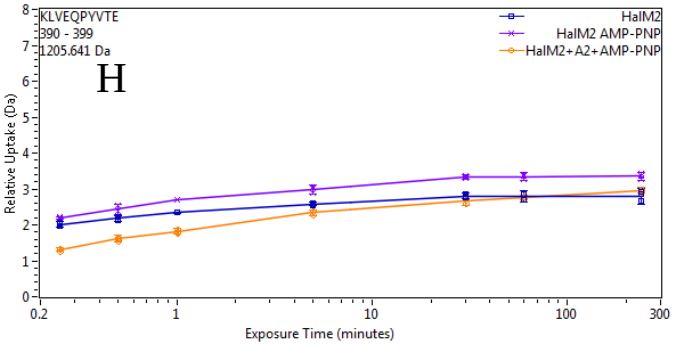
P349-L358



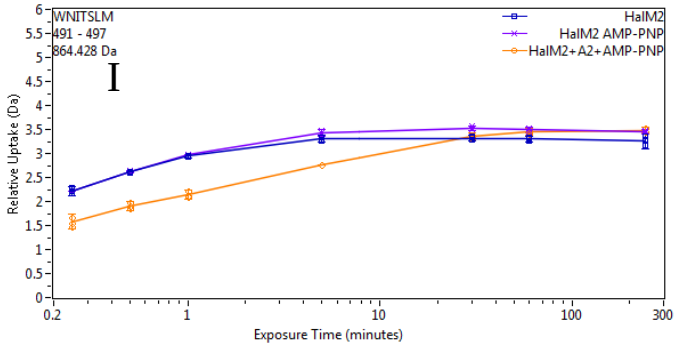
A353-D359



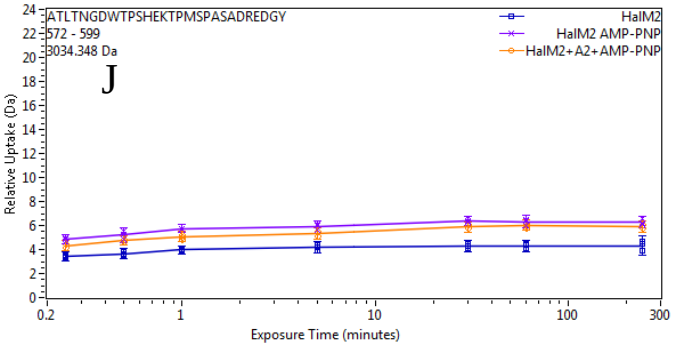
M389-E399



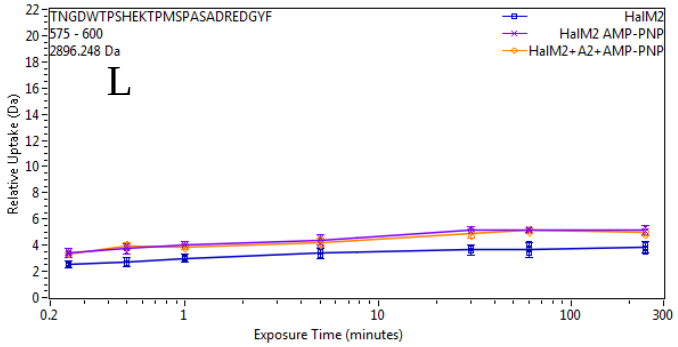
K390-E399



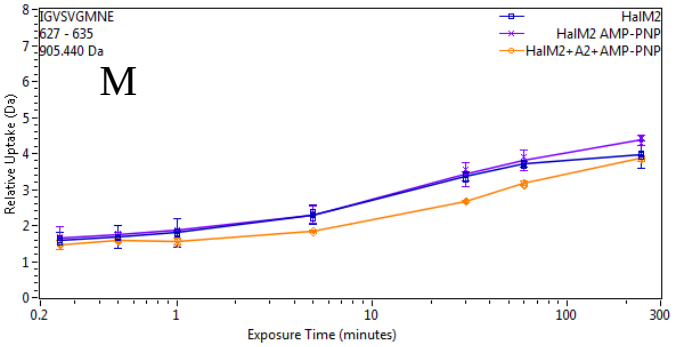
W391-M497



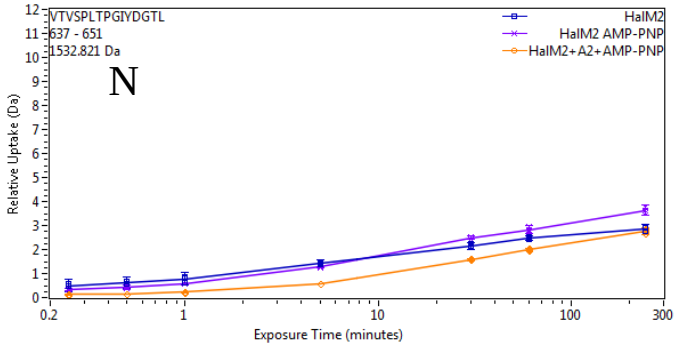
A572-Y599



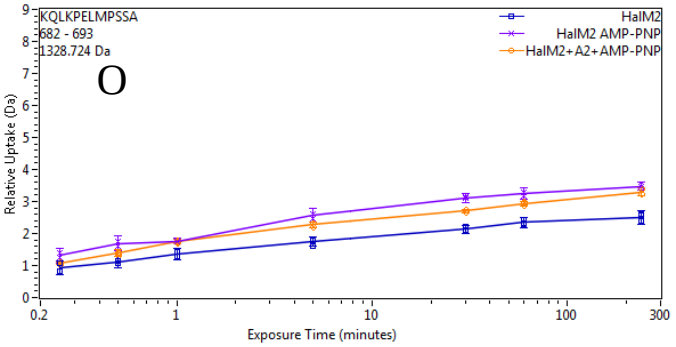
T575-F600



I627-E635



V637-L651



K682-A693

Figure 3.14. The uptake plots for peptides of interest identified in this study. Apo-HalM2 is in blue, the [HalM2+AMP-PNP] state is in purple, and [HalM2+Hal2+AMP-PNP] state is in gold.

The first step to the study involved analyzing the HDX properties of HalM2 alone, in the absence of HalA2 and the ATP analogue, AMP-PNP. What is immediately apparent is that the protein is considerably dynamic (Fig. 3.12, 3.15). Many of the most dynamic regions of the HalM2 sequence map to regions that are near in space and oriented towards the active site of either the dehydratase or cyclase domain, based on the HalM2 homology model (Fig. 3.16). This implies a potential role for these residues in substrate binding. The regions observed to exhibit significant exchange are H110-L118, P330-P410, N492-M497 and D778-L786.

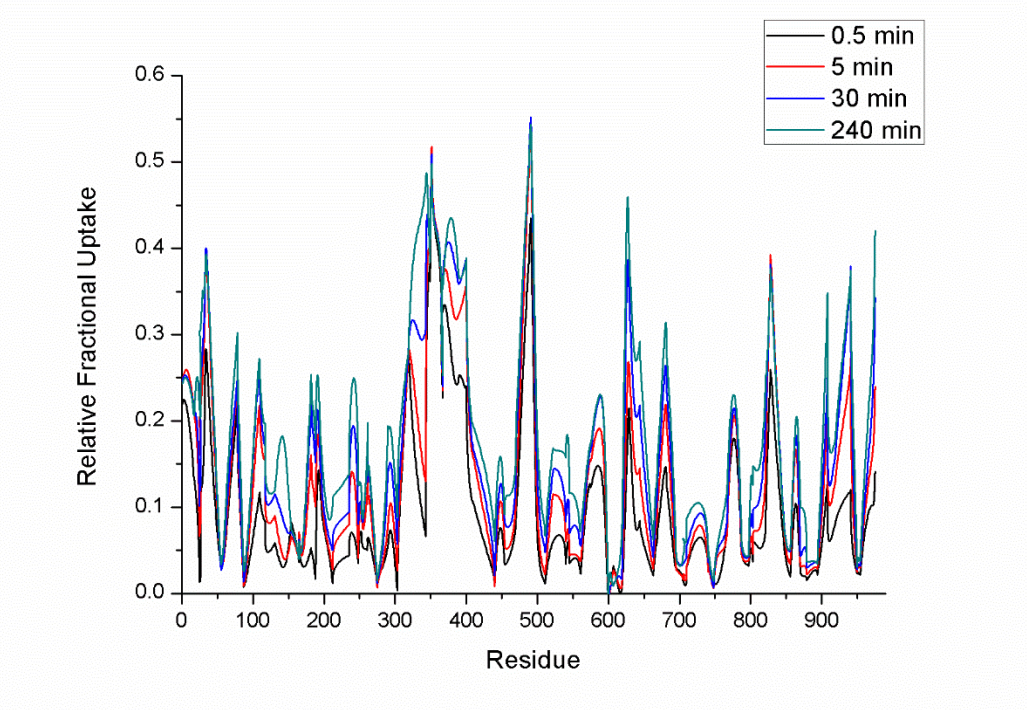


Figure 3.15. A butterfly plot displaying the relative fractional uptake of residues in apo HalM2. The residues are displayed N to C terminal along the x-axis.

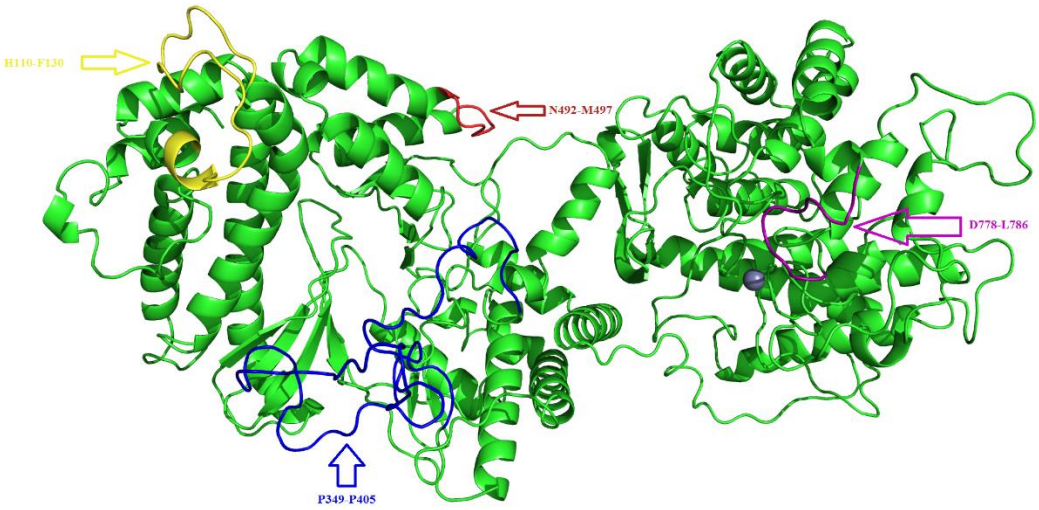


Figure 3.16. HalM2 homology model with dynamic regions of interest for apo HalM2 highlighted.

The most prominently dynamic region is found between residues G344 and K410, which we shall refer to as the "P349-P405 loop". Many of the peptides derived from this region underwent maximal exchange within the first 15 s of the exchange reaction to RFU values of ~ 50% (Fig. 3.14). PSIPRED analysis (a method used to predict secondary structural features) did not assign secondary structure to the region other than the presence of two small *β*-stands (W379-N383 and M389-Y396) that flank a moderately conserved Thr-Asp (TD) motif (Fig. 3.17). While the dynamic nature of the region suggests that it is disordered, it is possible that substrate binding may induce ordering, or that transient ordering of the region may facilitate substrate binding[57], which would explain the predicted β-sheet character. A comparison to the region in the presence of substrate is required in order to determine if this is the case (*vide infra*).

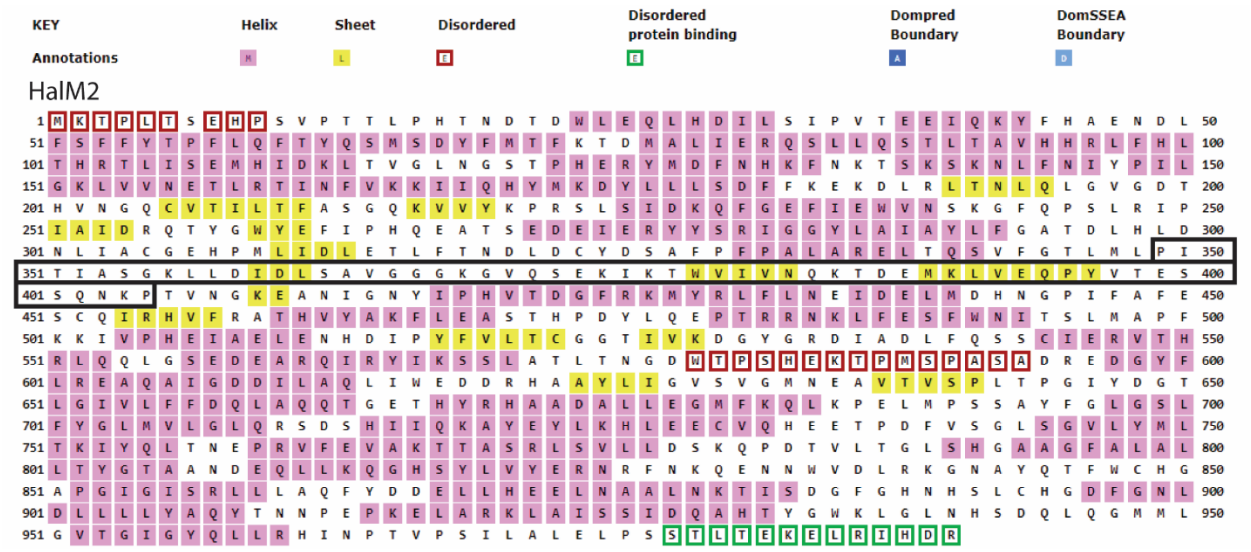


Figure 3.17. *PSIPRED analysis of HalM2, illustrating predicted secondary structural features. The P349-P405 loop is boxed in black and the three putative β-sheet regions within this loop are highlighted.*

Another curious region of significant deuterium uptake is found between residues N492 and M497, which maps to a loop between two α-helices of the dehydratase domain (Fig. 3.16, 3.17). While there are no conserved residues within the loop, PSIPRED analysis of the HalM2 sequence predicts the presence of two flanking α-helices. This is likewise observed in the HalM2 homology model (Fig. 3.16). The function of loop N492-M497 is not immediately obvious, but it is part of the kinase activation (KA) domain, a feature unique to LanMs that is critical to their activity. The loop is near in space to the established catalytic residues R455 and T461 (R506 and T512 in CylM), which facilitate phosphate elimination.[52] Moreover, the loop is oriented towards the putative HalA2 binding cleft, where it could engage in direct interactions with the HalA2 leader peptide and/or with the large, dynamic P349-P405 loop discussed above. The loop is also near in space (< 5 Å) to the kinase activation loop in the HalM2 homology model, which suggests that it may have a role in keeping the activation loop in its active conformation.

The residues H110-F130 are also seen to be rather dynamic, and form part of a structural feature known as the helical capping domain, which is also unique to LanMs and helps stabilize the KA domain. In the cyclase domain, a dynamic region in the free enzyme with no obvious function is D778-L786. Interestingly, this loop is located close in space to Zn^{2+} in the cyclase domain of both HalM2 and CylM (based on alignments), and H791 in HalM2 is known to protonate the enolate intermediate formed during cyclization (Fig. 3.18).

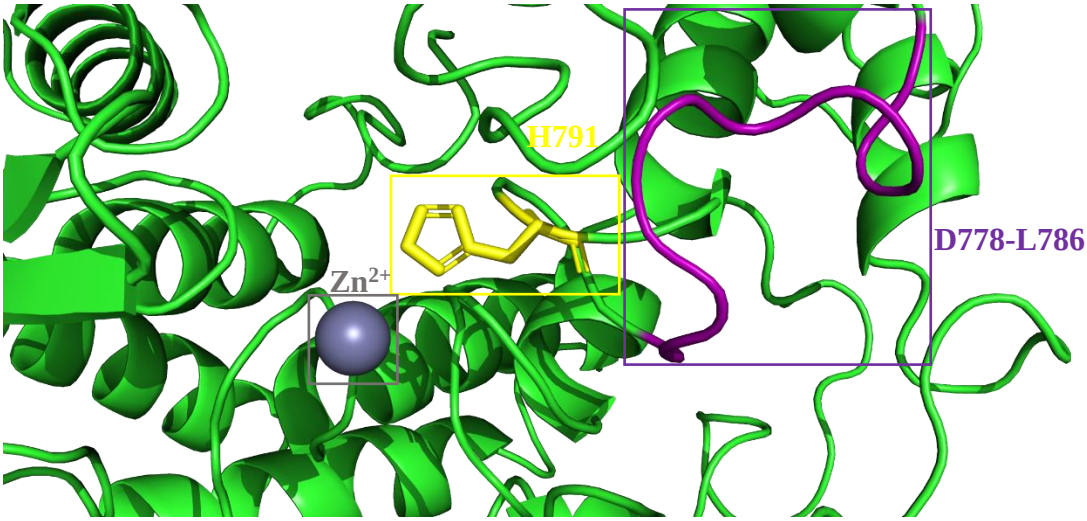


Figure 3.18. Close view of the cyclase active site of the HalM2 homology model, highlighting the catalytic residue H791 (which protonates the enolate formed during the cyclization reaction) and the dynamic D778-L786 loop. The catalytic zinc ion is displayed as a grey sphere.

3.2.6. **AMP-PNP binding primes HalM2 for peptide binding**: HDX-MS is most informative when used in comparative studies wherein the relative fractional uptake *difference* between two or more states are studied. The two "states" could be ligand bound and unbound enzyme forms, wild type and mutant enzymes, or different solvent exchange conditions. In this initial HDX study, we focus on the effects of ligand binding to HalM2. For HalM2, the two ligands of interest are HalA2 and ATP, the latter of which is needed for dehydratase activity. We hypothesize that ATP binding would perturb the local dynamics of HalM2, and that these changes could play a role in HalM2 achieving its catalytically active conformation. The role of ATP binding on conformation has been noted previously in other well studied systems[58], and it is possible that such behaviour may be seen here.

Introduction of both ATP and substrate would lead to catalytic activity (dehydration and cyclization of HalA2) within the reaction mixture. Under these conditions, the structure of the HalA2 peptide would be changing during the course of the H/D exchange. If there are differential interactions between the various partially modified HalA2 intermediates and HalM2, this mix of different conformations could lead to a range of effects on the HDX rate, thereby confounding any subsequent analyses. To sidestep this problem, the non-hydrolyzable ATP analogue β,γ-imidoadenosine 5′-triphosphate (AMP-PNP), was used in place of ATP. In short, AMP-PNP contains a non-hydrolysable N-P bond, meaning that, while it should bind to the HalM2

dehydratase active site in a similar fashion to ATP,[59] it is not competent in the phosphoryl transfer reaction needed to initiate Ser/Thr dehydration. To test for inhibitory activity of AMP-PNP, we conducted kinetic experiments under previously reported conditions[50] in the presence of (i) 5 mM ATP, (ii) 0.5 mM ATP, and (iii) 0.5 mM ATP + 5 mM AMP-PNP. A study by Thibodeaux et al., from which the procedure for this experiment was adapted, showed that HalM2 is saturated with ATP at 0.5 mM, leading to extensive conversion of HalA2 to Halβ under these conditions within 10 min.[50] Thus, the addition of a 10-fold molar excess of AMP-PNP should lead to a striking decrease in activity over this time frame if AMP-PNP is a competitive inhibitor with a binding affinity that is similar to ATP. As can be seen from the MS data in figure 3.19, AMP-PNP indeed significantly inhibits conversion of the starting material into the various intermediates and the final product, suggesting that AMP-PNP should be a good structural mimic for ATP binding.

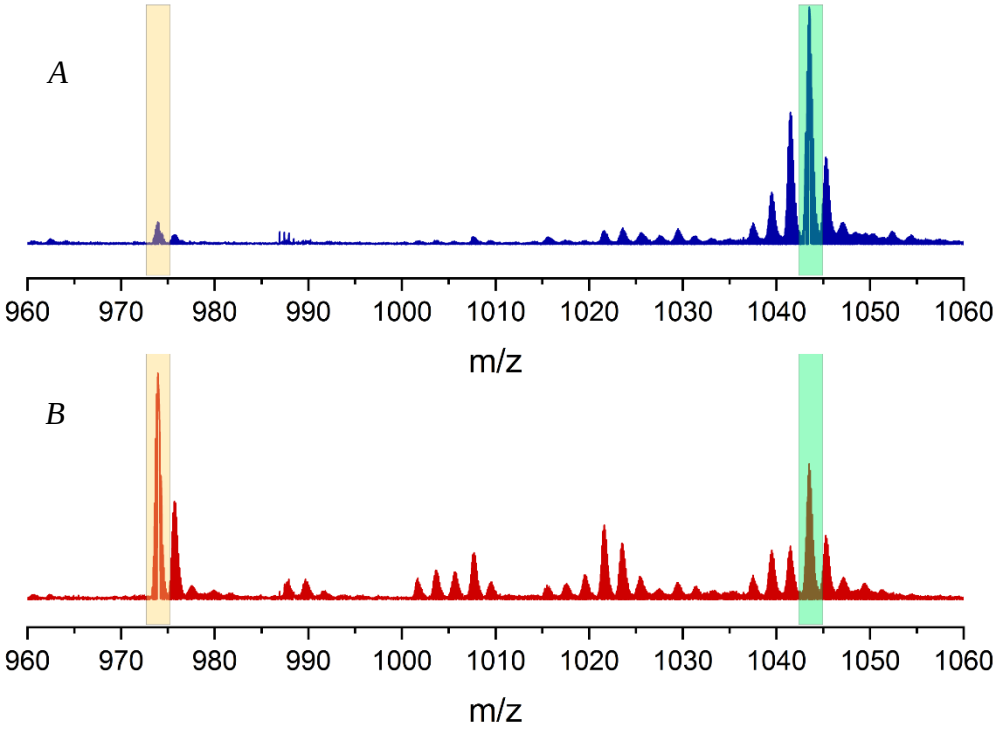


Figure 3.19. MS analysis of the competitive inhibition assay, monitoring the formation of the HalA2 intermediates in the 9+ charge state after 10 minutes of reaction time. (A) represents the assay containing 5 mM AMP-PNP with 0.5 mM ATP. (B) represents the assay with 0.5 mM ATP alone. The eneolope in green represents unmodified, fully NEM-alkylated HalA2 (m/z: 1042.9486; Obs.: 9377.4474 Da; Theoretical: 9377.4704 Da). The envelope in yellow represents fully modified HalM2 (m/z: 973.3499; Obs.: 8751.1491 Da; Theoretical: 8751.2017 Da). The other signals represent various partially modified forms of HalA2. Note the decreased abundance of final product and intermediates in the presence of AMP-PNP (panel A).

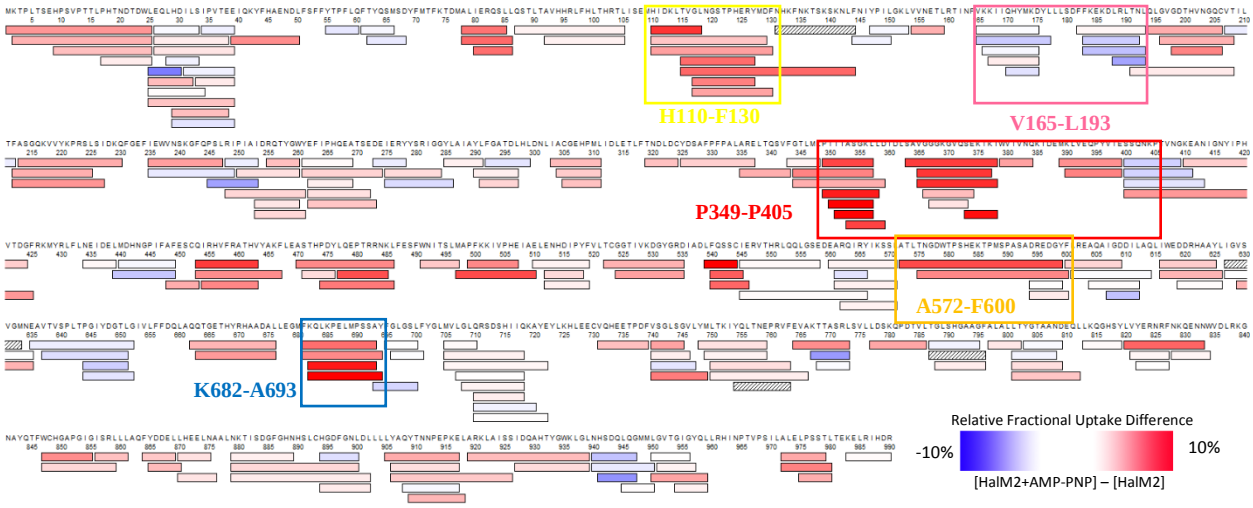


Figure 3.20. The coverage map displaying the relative fractional uptake difference between [HalM2+AMP-PNP] and apo-HalM2. Regions marked in red and blue display greater and less exchange, respectively, relative to the apo-HalM2 state.

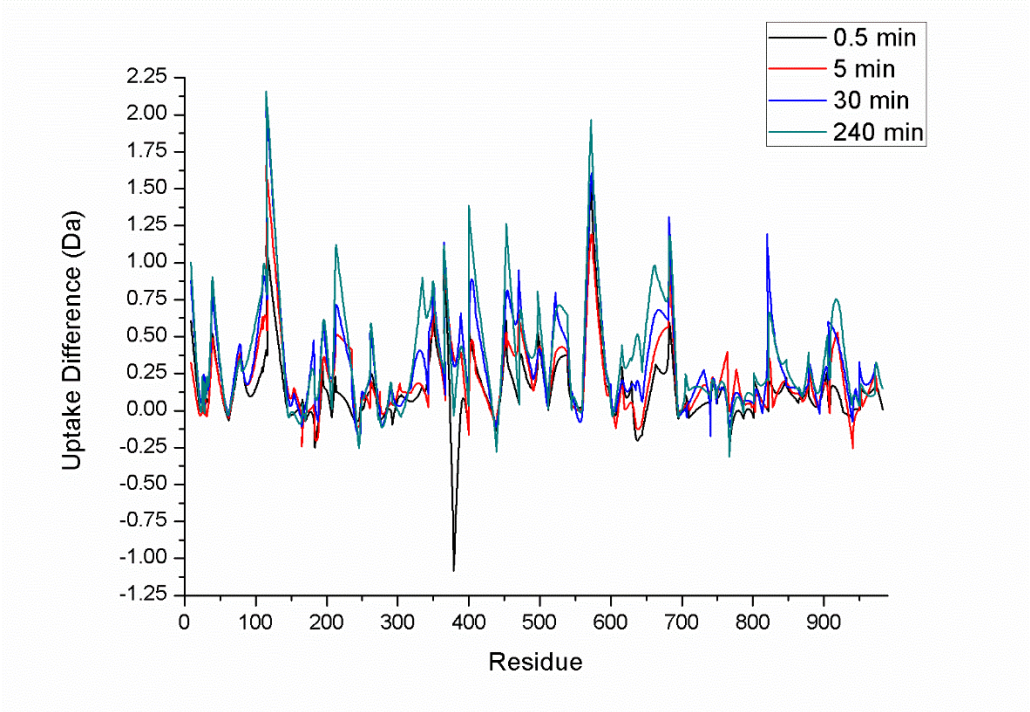


Figure 3.21. A butterfly plot showing the difference in deuterium uptake for each residue of HalM2 between [HalM2+AMP-PNP] and apo-HalM2 over time.

Comparing apo HalM2 to [HalM2+AMP-PNP], we see that overall, the protein becomes more dynamic, with the majority of the detected peptides showing greater exchange in this state

than in the absence of AMP-PNP (Fig. 3.20, 3.21). Regions that display significantly altered dynamics are as follows: (i) H110-F130, (ii) V165-L193, (iii) P349-P405 loop, (iv) A572-F600 and (v) K682-A693 (Fig. 3.20, 3.22). The region H110-F130 seems to become particularly more dynamic with the addition of AMP-PNP. As noted earlier, the residues H110-F130 are located in a cluster of α-helices referred to as the helical capping domain, which is noted for being unique to LanMs and are believed to help organize the KA domain. This suggests that the residues may play a significant role in the enzyme's activity, such as potentially being involved in binding substrate. Thus, the addition of AMP-PNP may prime the region to bind HalA2 by introducing structural flexibility. The residues V165-L193 appear more protected in the presence of AMP-PNP (Fig. 3.20), and based on the alignment to CylM, appears to be near the ATP binding site of HalM2, thus explaining the reduced exchange. The P349-P405 loop again appears to be highly dynamic, more so than in the case of apo-HalM2. It is possible that AMP-PNP binding leads to an increase in dynamics in order to facilitate binding.

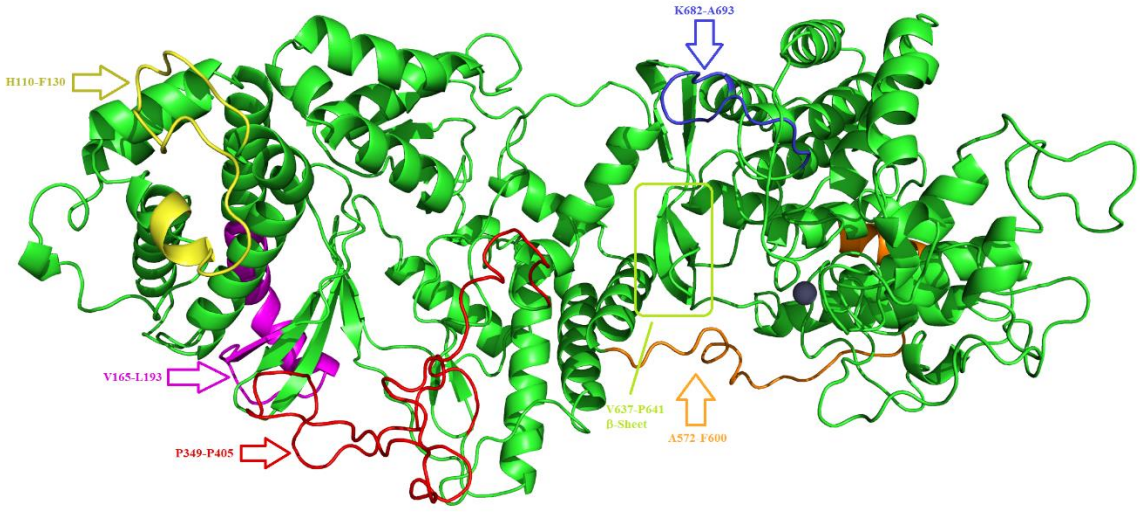


Figure 3.22. HalM2 homology model with regions of interest for [HalM2+AMP-PNP] highlighted.

In the cyclase domain, A572-F600 appears to be considerably more dynamic in the presence of AMP-PNP (Fig. 3.20). The homology model suggests that it is a loop region that wraps around an α-helix in the cyclase domain (Fig. 3.22). Moreover, PSIPRED analysis shows that the residues W579-A593 are expected to be disordered, explaining the high levels of exchange recorded in the study (Fig. 3.17). No function has yet been assigned to this region, but the increase

in dynamics marks it as a region of interest. Another region of interest is K682-A693. With AMP-PNP present, the region becomes noticeably more dynamic (Fig. 3.20), There is no obvious function for this region, but based on the HalM2 homology model (Fig. 3.22), this loop is close in space to the cyclase active site. Thus, it may play a role in cyclization, but further study is required to confirm this. Thus, in summary, HDX-MS identified previously overlooked regions likely associated with enzyme activity, and provided evidence for AMP-PNP binding inducing a more catalytically active conformation.

3.2.7. HalA2 binding significantly perturbs exchange patterns, leading to global exchange reduction:

Upon the addition of both AMP-PNP and HalA2 to HalM2, the exchange dynamics changed markedly from apo HalM2 and [HalM2+AMP-PNP], with various regions across the protein displaying increased protection, mostly in the dehydratase domain, and we hypothesize that many of these regions form direct interactions with HalA2 (Fig. 3.23). Regions of interesting exchange patterns are (i) H110-F130, (ii) P349-D359, (iii) M389-E399, (iv) N492-M497, (v) A572-L600 and (vi) I627-G652 (Fig. 3.24). As noted earlier, H110-F130 is more dynamic with AMP-PNP, and we postulated that AMP-PNP binding primes this region to bind HalA2. Moreover, we noted that the region represents a relatively well conserved loop that appears to be unique to LanMs, lending credence to the claim that this loop plays an important role in HalM2 activity. Comparing the state [HalM2+HalA2+AMP-PNP] to apo HalM2, there is a clear reduction in exchange for the residues H110-F130, providing evidence for some type of interaction between HalA2 and the region in question (Fig. 3.23 A), and a comparison to [HalM2+AMP-PNP] only highlights the observed protection even more (Fig. 3.23 B). To ascertain whether or not this is the case, we propose three separate mutations spanning this region for future studies: Replacing H110-L115 with a GS4 linker (4 repeating glycine-serine units), T116-T123 with a GS4 linker and H124-K133 with a GS5 linker. These mutants are expected to display significantly reduced binding affinity for HalA2.

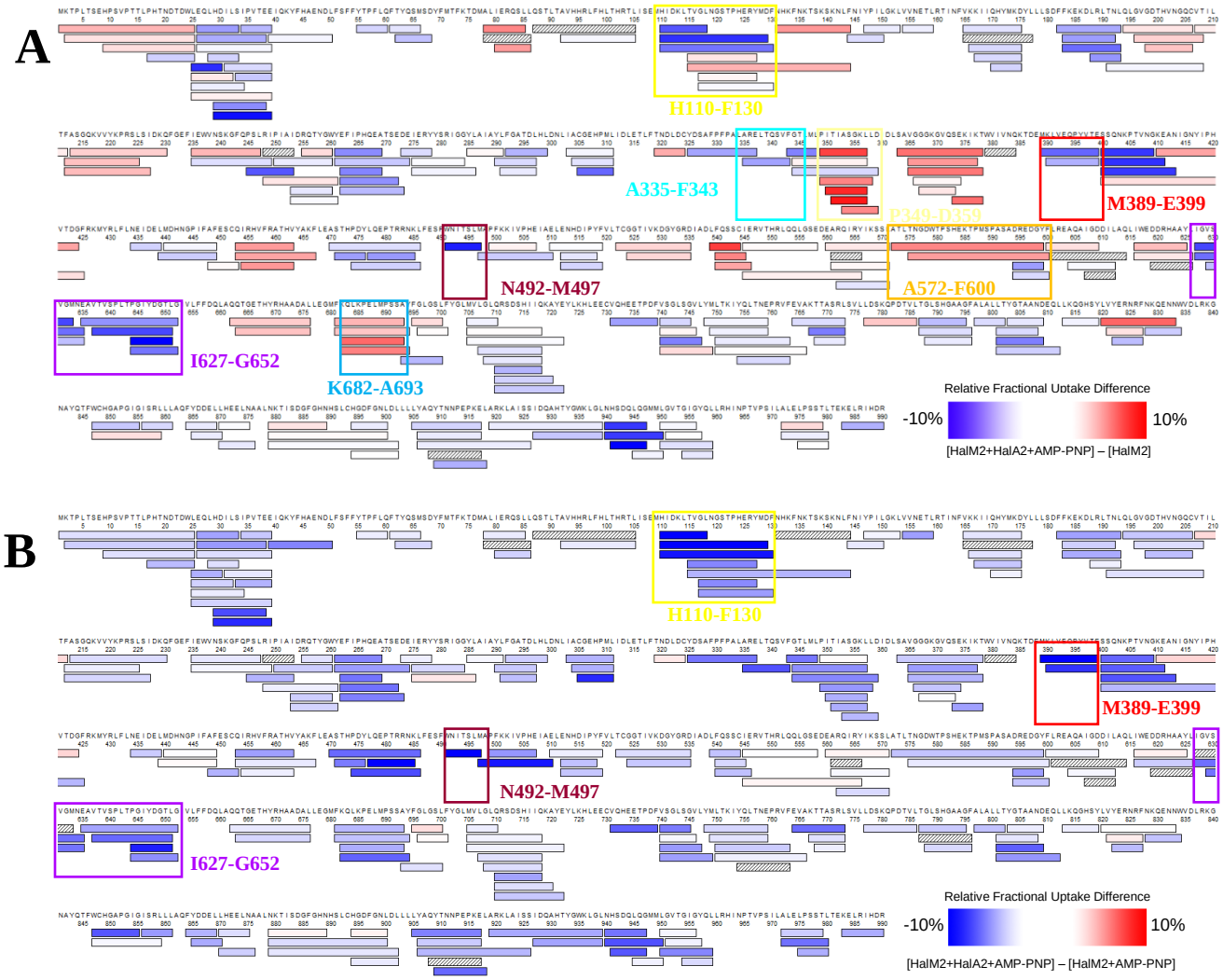


Figure 3.23. Coverage map displaying the relative uptake difference between [HalM2+HalA2+AMP-PNP] and apo-HalM2 (A) and [HalM2+HalA2+AMP-PNP] and [HalM2+AMP-PNP] (B). Regions marked in red display greater exchange in [HalM2+HalA2+AMP-PNP].

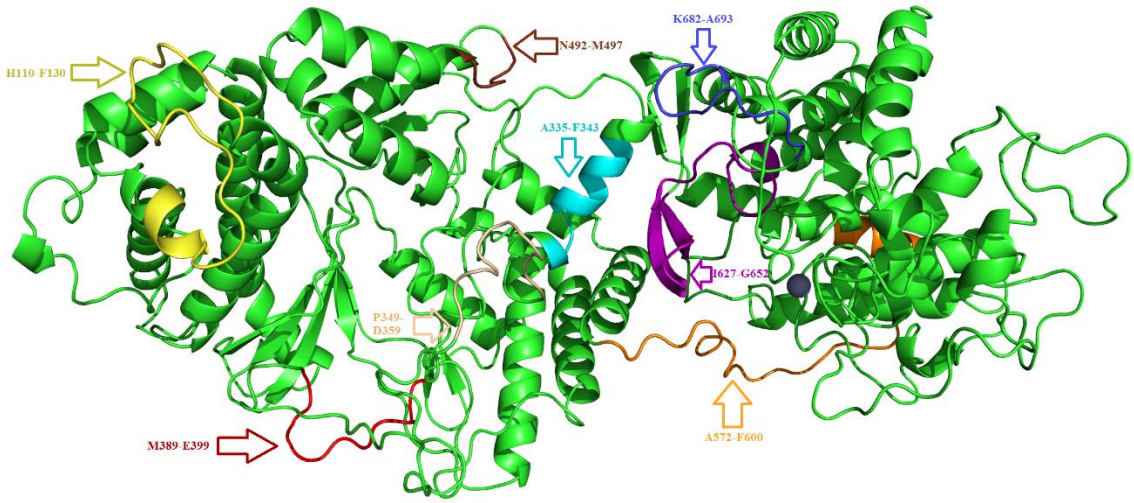


Figure 3.24. The HalM2 homology model with the regions of interest for [HalM2+HalA2+AMP-PNP] highlighted.

Adjacent to the aforementioned P349-P405 loop is a predicted α-helix spanning the residues A335-F343 (cyan, Fig. 3.24). Despite not showing any particular response to AMP-PNP binding, we find that the region displays significantly reduced exchange in the presence of HalA2, suggesting some role in substrate binding (Fig. 3.23). The proximity of the A335-F343 α-helix to the active site of the dehydratase domain, and in particular to the P349-P405 loop seems to strongly suggest that the helix forms some type of interaction with substrate, possibly acting as a major component of the dehydratase domain binding pocket. Additionally, the fact that the region already exhibits defined secondary structural elements provides greater evidence that the reduced exchange is a consequence of direct HalA2 interactions.

In the presence of both HalA2 and AMP-PNP, significant perturbation in exchange is also seen throughout the P349-P405 loop. Within the loop, the greatest change in exchange behaviour is seen for residues P349-D359, which represents the *N*-terminal portion of the loop, and M389-E399 toward the C-terminal end (Fig. 3.24). The increased exchange exhibited by residues P349-D359 indicate increased flexibility, perhaps allowing the large, disordered P349-P405 loop to sample various conformations, which could be of relevance to substrate binding and catalysis. Of greater interest, the M389-E399 region is predicted to form a β-sheet according to PSIPRED predictions (Fig. 3.17), suggesting that the region may become ordered in the presence of substrate in order to facilitate interaction. This ordering is consistent with the reduced HDX observed in this

region upon HalA2 binding. Consequently, we surmise that the A335-F343 α-helix and the residues M389-E399 act as points of contact with the substrate HalA2, perhaps with the assistance of the H110-F130 region, while P349-D359 likely facilitate conformational changes within the HalM2:HalA2 complex that are necessary for modification. For initial mutagenesis studies, we propose replacing residues F331-Q340 with a GS5 linker, S341-L348 with a GS4 linker and M389-Y396 with a GS4 linker. If the above-stated hypothesis about their role in binding HalA2 is true, then all of these mutations should lead to reduced binding affinity, and a consequent drop in activity.

A curious find, residues A572-F600 contain a loop that appears to stretch across the cyclase domain (Fig. 3.24). It appears to be more dynamic in the presence of ligand, be it AMP-PNP alone or HalA2 and AMP-PNP together (Fig. 3.20, 3.23). PSIPRED predicts the presence of a disordered loop between residues W579 and A593. Proline residues often flank disordered loops, as they introduce 'kinks' into a peptide chain[60]. Thus, residues P581 and P590 are believed to mark the N and C-terminal portions of the loop respectively. Moreover, acts as a bridge between the dehydratase and cyclase domains. While the role of this region is not immediately clear, one possibility is that it simply acts as a bridge between the dehydratase and cyclase domains, allowing long-range structural communication between the two. This is likely the case given that the residues are distant from the active face of the protein. Regardless, the significant perturbation of the exchange pattern upon the addition of HalA2 marks it as a region of interest. As such, the residues P581-P590 were selected to be replaced with a GS5 linker. While the exact effect of this mutation is difficult to predict, it is very likely to impact activity and may very well affect cyclization. It may also reduce the stability of the [HalM2:HalA2] complex and lead to the release of various intermediates, some with improper ring installation.

Further along the cyclase domain, residues I627-G652 are apparently more protected with addition of both AMP-PNP and HalA2 (Fig. 3.23). PSIPRED predicts the presence of β-sheets (V637-P641) which, as noted in the case of β-sheet forming residues I666-L690 in CylM, are known to act as RiPP recognition elements and are critical for the protein to engage its target. The increased protection and the likelihood of β-sheet formation in between residues I627 and G652 point towards this region being an RRE. Given the tendency of RREs to interact with leader peptides, it is possible that the region interacts with the leader peptide region exclusively rather

than the core peptide. As such, its role would be limited to primarily substrate recognition as opposed to any unique chemistry. It should also be noted that residues K682-A693 display increased exchange in the presence of peptide (Fig. 3.23). However, the significance of this relative to the function of the protein is difficult to surmise. Similar to residues P349-D359, this region may simply introduce the flexibility needed to sample active conformations within the HalM2:HalA2 complex. To that end, we selected residues E635-P644 to be replaced with a GS5 linker. Such a mutation is likely to impact substrate recognition and reduce activity.

There exists the possibility that observed effects on exchange as a consequence of ligand binding are simply artifacts. To determine whether or not this is the case, a final control experiment was performed by titrating varying amounts of HalA2 in order to look for a concentration dependent response in the HDX signals. In short, HalM2 was exposed to a steadily increasing concentration of HalA2 (1, 2, 5, 15 and 25 uM), and exchange was monitored for the peptide M389-E399. As can be seen from figure 36, the data show a clear trend with increasing HalA2 concentration (Fig. 3.25). By fitting the data to the equation 3.1 listed in the experimental section, we were able to derive a K_d value of ~3.49 μM, which is in reasonable agreement with the K_d of ~1.8 μM reported by Thibodeaux et al. [50] This indicates that our current workflow allows for accurate and reproducible detection of exchange perturbations.

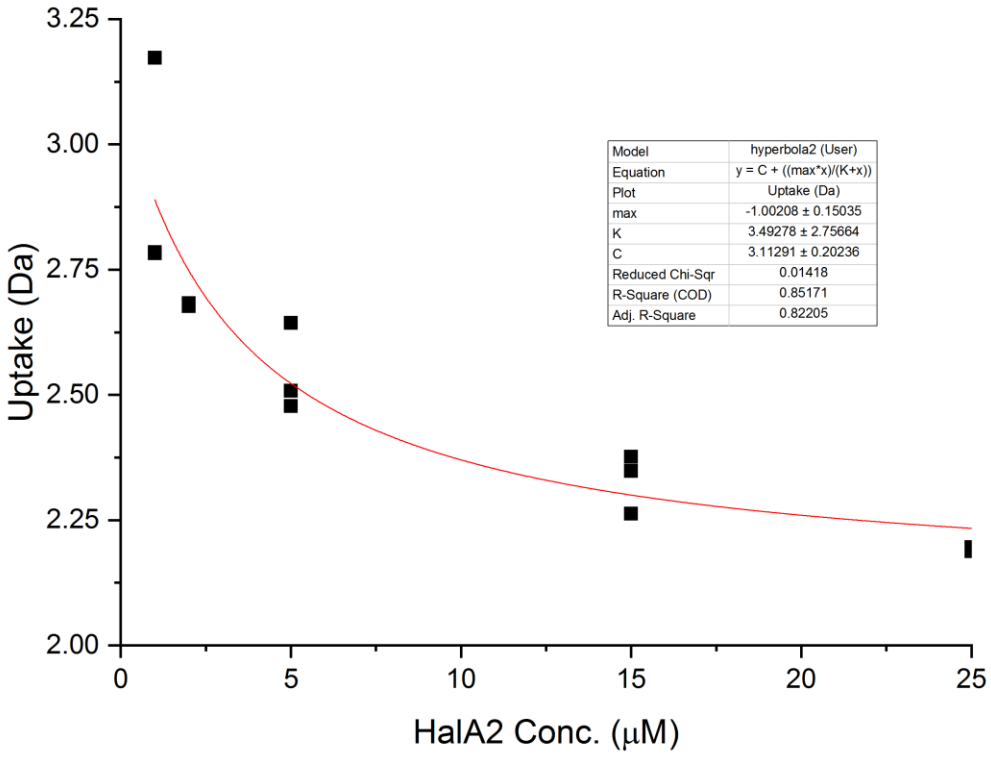


Figure 3.25. The exchange pattern recorded for residues M389-E399 with increasing HalA2 concentration. There is a clear trend towards greater protection in the presence of substrate. The data was fitted to equation 3.1 to derive K_d.

3.3 Conclusion:

RiPPs represent a promising means of developing new antibiotics, due to their unique modes of action, and in particular, their potential for engineering. Lanthipeptides are especially interesting as they often display antibiotic activity. To engineer lanthipeptide biosynthetic systems, a thorough understanding of the mechanism behind ring installation by various enzymes is required. LanMs are bifunctional enzymes that are capable of both dehydratase and cyclase activity, and among the many factors that contribute to the formation of the [LanM:LanA] Michaelis complex during lanthipeptide maturation, conformational sampling is believed to play a significant role.[50] To shed light on the matter, we developed a method to probe the conformational dynamics of HalM2 using hydrogen-deuterium exchange mass spectrometry, and were able to provide the first insights into changes in HalM2 dynamics upon ligand binding. Putting together all the data presented above, we propose a model wherein HalM2 conformational

dynamics are heavily dictated by the binding of ligands, and where the majority of the binding interactions occur in the dehydratase domain. More specifically, AMP-PNP binding leads to an increase in exchange throughout the protein, with the exception of residues V165-L193, which is near the ATP binding pocket. Thus, it is very likely that AMP-PNP helps to prime the protein by increasing the accessibility of residues that are required to interact with HalA2. The addition of HalA2 leads to significant protection in various regions of the protein, mainly in the dehydratase domain, however the β-sheet forming residues I627-G652, which is believed to be a RiPP recognition element, shows a drastic reduction in exchange, providing evidence for its role in binding HalM2. The residues A572-F600 are unique in that it is the only region that shows significant perturbation upon ligand binding that is not located prominently on the active face of the enzyme. The increase in dynamics, coupled with its distance from the active sites provides evidence for its role in facilitating long range structural communication in HalM2 rather being involved in peptide binding. To validate these findings, we proposed a series of mutations for use in future studies. Thus, HDX-MS has proven its utility in helping to identify previously overlooked regions of functional significance in HalM2, and by extension, LanMs as a whole. As such, we believe that this approach will aid future studies on lanthipeptide and other RiPP biosynthetic enzymes.

3.4.Future Work:

While this study provides significant information on functionally relevant regions in HalM2, the results still need to be validated in further studies. To that end, several mutations were proposed in the preceding sections. These mutants are to have their activity studied to determine if the mutations lead to any perturbations. The Thibodeaux lab has already carried out studies on the proposed mutants and determined that several of the identified regions of interest, many of which have not been previously identified in LanM studies, are in fact functionally relevant. Further work remains however to determine if these structural elements are a common feature of LanMs in general or are specific to HalM2. Thus, the Thibodeaux lab is currently utilizing the HDX method developed in this study to probe the dynamics of other LanMs, such as ProcM. It is possible that many of the interesting differences in the functional properties of LanM enzymes (such as the differences in catalytic efficiency, substrate specificity, and reaction directionality) derive from fundamental differences in the structural dynamics of these enzyme systems. Finally, the HDX-

MS platform can be utilized to study similar phenomena in other RiPP biosynthetic enzymes, as these systems are all likely to exploit conformational dynamics to some extent in order to carry out their iterative catalytic activities.

3.5.Experimental:

3.5.1. Cloning: Unless otherwise stated, all cloning reagents were purchased from NEB and all oligonucleotides were purchased from IDT. The sequences of the primers used in this study are listed in table 3.1.

Table 3.1. List of primers used in this study, with corresponding sequences and T_m.

Primer #	Primer name	Sequence (5'-3')	Tm (°C)
1	HalM2_f	GGGTATCCGCTCGAGATGAAAACTCCTCTAACAAGT	76
2	HalM2_r	TATAAACGCGGATCCTTATCTGTCATGAATTCTCAA	74
3	HalA2_f	GCGCCGCATATGGTAAATTCA	68
4	HalA2_r	AAACTCGAGTTAGCACTGGCT	62
5	pET15_colony_FP	GAGGATCGAGATCTCGATCCC	59
6	pET15_colony_RP	GGCTTGGTTATGCCGGTACTG	52
7	pET28_colony_RP	AACCCCTCAAGACCCGTTTAG	60
8	T7	TAATACGACTCACTATAGGG	49
9	T7_terminator	GCTAGTTATTGCTCAGCGG	56

A typical PCR reaction contained 50 ng genomic DNA, 0.5 µM of each oligonucleotide primer, 1 mM deoxynucleotide-5′-triphosphates (dNTPs), 5% DMSO, 1x Phusion Polymerase buffer (either HF or GC buffer), and 1 unit of Phusion Polymerase. Typical PCR cycling conditions were as follows: 30 s incubation at 98 °C, followed by 35 cycles composed of a 10 s denaturation step at 98 °C, a 20 s annealing step at a temperature ranging from 55 – 65x °C depending on the gene being amplified, and an elongation step at 72 °C for 30 s per kb being amplified. After the last cycle, a final elongation was carried out for 10 min at 72 °C. The sizes of the amplicons were verified by 0.5% agarose gel electrophoresis and the PCR products were digested with DpnI to eliminate parental DNA. To every 50 uL of PCR reaction mixture, 6 uL of 10x CutSmart buffer is added along with 3 uL of sterile water and 1 uL of 20 U/ uL DpnI. The fragments were incubated for 1 hour at 37 ^{o}C then purified using 0.5% agarose gel electrophoresis. The desired fragments

were then cut from the gel and purified using the NucleoSpin Gel and PCR cleanup kit from Macherey Nagel following the manufacturer's instructions. The concentration of each of the amplicons was determines using Thermo Scientific Nano Drop Lite. All fragments were stored at -20 ^{o}C until needed. The vectors (pET15b and pET28b) were prepared for Gibson assembly by digesting 50 uL stocks with BamHI (6 uL 10x CutSmart buffer, 3 uL sterile water and 1 uL BamHI). Fragments were then purified using gel electrophoresis and the Nucleospin Gel and PCR cleanup kit, as mentioned above. Concentrations were determined using a nanodrop and fragment were stored at -20 ^{o}C until needed.

For Gibson assembly, a 5 uL solution containing equimolar amounts of each fragment (100 ng of larger fragment) was prepared, and to this, 15 uL of Gibson mix was added. The mixtures were then incubated at 50 ^{o}C for 1 hour. Afterwards, 10 uL of mixture wass transformed into chemically competent E. Coli DH5a. The cells were incubated on ice for 30 minutes before the addition of 1 mL of LB broth. The cells were then incubated at 37 oC for 1 hour before being streaked on an antibiotic infused LB agar plate (50 ug/mL kanamycin for pET28b, 100 ug/mL ampicillin for pET15B). The plates were then allowed to dry and then incubate at 37 ^{o}C for 16-20 hours. From the plates, colonies were selected for colony PCR (as described in chapter 2) and the colonies with a positive result from gel electrophoresis were selected for overnight growth and subsequent DNA extraction (as described in chapter 2). The extracted plasmids were then submitted for sequencing to the Genome Quebec Innovation Centre with the appropriate primers (###). Once sequences were verified, the plasmids were then transformed in DH5a (for maintenance) and BL21 (for expression) according to the method described above, and the cell lines were stored at -80 ^{o}C until further use.

3.5.2. Protein and peptide expression: 120 mL of LB broth were inoculated with the desired BL21 cell line and the appropriate antibiotic, and were then incubated at 37 ^{o}C overnight. These cultures were then used to inoculate 10 L of Terrific Broth (TB) that has already been inoculated with the appropriate antibiotic. The cultures were then incubated at 37 ^{o}C until an OD_{600} of 0.8 was reached, after which expression was induced with the addition 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). For HalA2, the cultures were left to incubate at 37 ^{o}C for 3 hours after the addition of IPTG. For HalM2, the cultures were cooled to 18 ^{o}C and left to incubate for

16 hours after the addition of IPTG. After incubation, cells were harvested by centrifugation (8,000 xg, 10 mins) at 4 ^{o}C. The cell pellets were then collected and stored at -80 ^{o}C until use.

3.5.3. Purification of protein/peptide:

3.5.3.1 Protein: His_6-tagged HalM2 was purified using immobilized metal affinity chromatography (IMAC). The cell pellet was resuspended in lysis buffer (20 mM Tris, 500 mM NaCl, 10% glycerol, pH 7.5) and kept cool with ice. The suspension was then sonicated at 60% amplitude with a 2.0 second pulse on time and 8.0 seconds off. Sonication was continued for a total of 5 minutes of pulse on time, with pauses for stirring and to make sure the suspension remained cool. After sonication, the suspension was then centrifuged (15,000 x g, 45 mins) at 4 ^{o}C to remove cell debris. The supernatant after centrifugation was then collected and loaded on to a 5 mL HisTrap FF column from GE Healthcare that was pre-equilibrated with 3-5 CVs of lysis buffer at 4 ^{o}C. The loaded column was then washed with 10 CVs of wash buffer (20 mM Tris, 30 mM imidazole, 1M NaCl, pH 8.0). The column was then connected to a BIO-RAD NGC Quest 10 Fast Protein Liquid Chromatography (FPLC) system, also at 4 ^{o}C, and washed with wash buffer until the UV trace ($A_{280\ nm}$) had reached baseline. The protein was then eluted using elution buffer (20 mM Tris, 500 mM NaCl, 200 mM imidazole, pH 7.5) using a gradient of 0-100% elution buffer over 60 minutes at a flow rate of 1.5 mL/min. Protein containing fractions were pooled together and glycerol was added (10% v/v) to help maintain protein stability. The pooled fractions were then concentrated using an AMICOM Ultra-15 centrifugal filter (50 kDa MW cut off) by centrifuging at 5000 x g for 10 mins at 4 ^{o}C. The concentrated eluate was then injected onto a Biorad Enrich Size Exclusion Chromatography 650 column (SEC-650, 10 x 300, 24 mL) that was pre-equilibrated with 2 CV of gel filtration buffer (20 mM HEPES, 300 mM KCl, 10% glycerol, pH 7.5) at 4 ^{o}C with a flow rate of 1 mL/min. The fractions containing monomeric protein were collected and their concentration determined by their A_{280} reading and the extinction coefficient of HalM2 (ε_{280} = 113,020). They protein was aliquoted, frozen with liquid nitrogen and stored at -80 ^{o}C until needed.

3.5.3.2. Peptide: The His_6-tagged HalA2 was purified using immobilized metal affinity chromatography (IMAC). The cell pellet was resuspended in wash buffer (6 M guanidinium hydrochloride, 20 mM NaH_2PO_4 pH 7.5 at 25 ^{o}C, 300 mM NaCl, and 30 mM imidazole) and was

sonicated for lysis (4.4 s pulse on, 9.9 s pulse off, 15 min total pulse on time) at room temperature. The lysate was then centrifuged at 16,000 x g for 45 minutes to pellet cell debris. The supernatant was then loaded onto a HisTrap FF (5 mL) column from GE Healthcare that had been pre-equilibrated with 5 column volumes (CVs) of wash buffer. After loading the cell lysate, the column was washed with 10 column volumes (CVs) of wash buffer after loading. The peptide was then eluted using 5 CVs of elution buffer (4M guanidinium hydrochloride, 20 mM NaH2PO4 pH 7.5 at 25 oC, 300 mM NaCl, and 30 mM imidazole). The eluate was desalted using a C8 SPE column (C_8 Sep-Pak, 30 mL) from Waters using the following procedure. The SPE column was loaded with sample, then washed with 5 CVs of 0.1% TFA solution. The peptide was then eluted using a solution of 80% acetonitrile and 0.1% TFA. The samples were then lyophilized and resuspended in a minimal volume of 0.1% TFA solution. Afterwards, the crude peptide was further purified by RP-HPLC using a C8 semiprep column (SUPELCO). Solvent A was 0.1% TFA solution and solvent B was acetonitrile with 0.1% TFA. The gradient chosen was 3-100% B over 37 minutes. Fractions were screened via MALDI-TOF-MS, and the peptides were lyophilized, resuspended in sterile water. The solution was then aliquoted, flash frozen with liquid nitrogen and stored at -80 ^{o}C until further use.

3.5.4. Kinetic assays: Aliquots of HalM2 and and HalA2 were thawed on ice. Meanwhile fresh stock solutions of ATP and TCEP, 100 mM each, were prepared. Stock solutions of HEPES (0.8 M, pH 7.5), $MgCl_2$ (500 mM) and AMP-PNP (100 mM) were prepared in advance. Two separate pre-incubation solutions were prepared containing the protein and peptide respectively. This was done to allow both the peptide and the enzyme time to equilibrate in solution before initiating the reaction. The peptide solution was composed of 100 mM HalA2, 100 mM HEPES pH 7.5, 1 mM TCEP, 5 mM $MgCl_2$, and varying amounts of either ATP, AMP-PNP or both.

The protein pre-incubation solution was composed of the exact same components as the peptide solution, except for the addition of 2 uM HalM2 instead of HalA2. The pH of the solutions were adjusted to 7.5 using 1 uL of 1 M NaOH for every 100 uL of sample, and the pH was confirmed using pH paper (Micro Essential, Hydrion pH 6-11). The pre-incubation solutions were then allowed to equilibrate at on the benchtop for 30 min. Afterward, the solutions were mixed 1:1 to give a final solution containing 1 uM HalM2, 50 uM HalA2, 100 mM HEPES pH 7.5, 5 mM $MgCl_2$, 1 mM TCEP and the varying amounts of ATP. With regards to ATP and AMP-PNP, 3

conditions were tested: (i) 5 mM ATP, (ii) 0.5 mM ATP and (iii) 0.5 mM ATP + 5 mM AMP-PNP.

After mixing, the samples were allowed to incubate for 10, 30 and 60 minutes. At the desired time points, 100 uL aliquots of reaction mixture were removed and quenched using 900 uL of 100 mM sodium citrate and 1 mM EDTA, pH 3.0. The quenched aliquots were then spiked with TCEP, giving a final concentration of 10 mM TCEP, and then incubated at 37 ^{o}C for 10 minutes. This ensures the reduction of disulfide bonds that may confound MS studies of the assays. The pH of the solution is then adjusted to ~6.2 with the addition of approximately 40 uL of 5 M NaOH. To this, 11 uL of 1 M N-ethylmaleimide (NEM) dissolved in ethanol was added, giving a final concentration of 10 mM NEM. NEM is used in the derivatization of free cysteines in HalA2. This introduces a clear mass shift in MS studies, allowing one to determine the number of cysteine residues not engaged in thioether ring formation. This is also why TCEP was used earlier, so as to limit the number of disulfide linkages that may lead to false results. After the addition of NEM, the solution are then incubated at 37 ^{o}C for 10 minutes before being quenched with 11 uL of concentrated TFA, giving a final concentration 1% (v/v) TFA. The samples were then desalted using a C_4 SPE column (Chromabond, 1 mL). The column was activated using 4 mL of 80% AcCN, 0.1 % TFA solution. It was then washed with 4 mL of 0.1% TFA solution, and sample was then loaded on to the column. The column was then washed with 4 mL 0.1% TFA and eluted with 2 mL 80% AcCN, 0.1% TFA solution. The eluate was then frozen with liquid nitrogen and lyophilized.

3.5.5. MS analysis of kinetic assays: Lyophilized samples were redissolved in 100 uL of 0.1% formic acid solution (FA) in water. The peptides were separated using a Waters ACQUITY BEH C_8 column (1.7 uM, 2.1 mm x 100 mm) using water and acetonitrile with FA (Solvent A: H_2O + 0.1% FA, Solvent B: AcCN + 0.1% FA).15 uL of approximately 25 uM peptide was injected onto the column with a flow rate of 50 uL/min and a gradient of 3-100% B over 10 min. The various modified forms of HalA2 elute at 4.5-5.5 min with this gradient. The eluate was then directed into the ESI source of Waters Synapt G2-Si instrument with the following settings: polarity = positive ion mode, resolution mode, capillary voltage of 3 kV, cone voltage of 40 V, source offset of 80 V, cone gas flow rate of 50 L/h, source temperature of 100 ^{o}C, desolvation temperature of 250 ^{o}C.

Glu-1-fibrinopeptide (Glu-Fib) was used as the external standard, and spectra were collected over a range of 100-2000 m/z, 1 scan per second, in centroid mode.

3.5.6. HDX assays: Buffer components (HEPES, TCEP, $MgCl_2$, AMP-PNP) and HalA2 were deuterated by first dissolving them in D_2O (Sigma-Aldrich), lyophilizing and then dissolving them again in D_2O. This was done to ensure maximal deuteration of the components, and to minimize the introduction of protons in the assay. A standard HDX assay was a total 50 uL, with 50 mM HEPES (pD = 7.5), 5 mM $MgCl_2$, 1 mM TCEP. Where AMP-PNP was used, 5 mM were added to the assay, and where HalA2 was used, 15 uM were added. All exchange reactions were carried out at room temperature. The 1 M HEPES buffer stock was adjusted to a pH of 7.1 (pD = pH + 0.4) with concentrated hydrochloric acid. A stock solution of quench buffer (100 mM potassium phosphate) was prepared and the pH adjusted to ~2.0. Quenching needs to bring the pH of the assay to 2.5 so as to minimize back exchange before analysis. To determine how much quench buffer was needed to do so, varying amounts of quench buffer were added to aliquots of 50 mM deuterated HEPES buffer, and the pH measured using an electrode and pH paper. The addition of 75 uL of quench buffer to 50 uL of HEPES buffer was found to give optimal results.

Assays were carried out over a 4 hour period, and quench at a total of 7 time points (15s, 30s, 1 min, 5 min, 30 min, 60 min and 240 min). The assays were prepared by making a stock buffer solution equivalent in volume to 8 individual assays that contains all components, including peptide, except for protein. As soon as protein was added to this stock, the timer is started, and 50 uL aliquots of the assay are pipetted into Eppendorf tubes containing 75 uL of quench buffer at room temperature once the desired time point is reached. Once quenched, the aliquoted are immediately frozen with liquid nitrogen and stored at -80 ^{o}C until use. This experiment was repeated in triplicate.

To prepare samples for MS, they were thawed in a water bath at 37 ^{o}C for exactly 1 min before being injected into the LC/MS system. A Waters Synapt G2-Si mass spectrometer was used for the study. It comes equipped with a drift tube for ion mobility MS studies. The LC system used was the Waters Acquity M-class UPLC with HDX technology. Buffer A was water with 0.1% formic acid and buffer B was AcCN with 0.1% formic acid. The system is designed with an injection port that feeds sample, using buffer A, into a temperature controlled (15 ^{o}C) immobilized

pepsin column (Waters BEH) where the protein undergoes online digestion (3 min). The eluate is then collected and passed through an analytical C18 column (Waters C18) that is chilled to 0.4 ºC to minimize back exchange. The gradient is 3-97% B over 10 min. After separation of the peptides produced from digestion, the eluate is fed directly into the ESI source for analysis. For ion mobility, the IMS cell pressure was 3.13 mbar and had a He gas flow rate of 90 mL/min. Traveling wave height was set to 40 V and velocity was set to 650 m/s. For tandem MS analysis of the peptide, parent ion were collected over a 12 s window, with a scan interval of 1 s, and fragmented using a collision energy of 15 V for the low energy regime and 30 V for the high energy regime. Argon was used for the fragmentation. Glu-Fib was used as an external standard.

3.5.7. MS data processing and analysis: We first needed to generate a reference list of peptides that could be reproducibly detected via our workflow. This was accomplished by preparing mock exchange reactions in triplicate with protiated buffer, absent HalA2 and AMP-PNP. The samples were then put through the standard workflow described above. The MS data collected from these samples was then uploaded to the ProteinLynx Global Server (PLGS) software from Waters to generate the peptide list. The software sifts through the MS data to detect peaks associated with the masses of peptides that are predicted to be formed by an *in silico* digestion of HalM2. From there, a series of spectral assignments are ranked based on user-defined filters, such as the difference in chromatographic retention time, differences between expected and observed m/z values and observed m/z values between replicates, differences in MSMS detections and ion mobility drift time. The PLGS data is then exported to DynamX 3.0 (Waters), were further thresholds are applied. More specifically, DynamX was set to generate a peptide list wherein peptides with 5 ppm of their expected m/z, with at least 3 assigned daughter ions and that were detected in all replicates were included. This led to the generation of a peptide list that included 224 peptides, covering ~87% of the protein sequence with 2.77 redundancy.

In determining the K_d of HalM2:HalA2, we employed the following equation to fit our data:

$$y = C + \left(\frac{U_{max} * x}{K_d + x}\right)$$

Equation 3.1. The equation used to derive the Kd of HalM2:HalA2, where U is uptake.

4. Conclusion:

In this thesis, we sought to aid the discovery and development of new antibiotics through the study of lanthipeptide systems, which often exhibit antibiotic activity, using a two pronged approach. In the first study, genome mining was employed to identify a novel two-component lanthipeptide system that we termed the saelicesin system produced by the actinomycete *Micromonospora saelicesensis*. The components that constitute the identified gene cluster were heterologously expressed to yield the peptides Saelα and Saelβ. These peptides were then characterized using mass spectrometry based techniques. One finding of note was that the peptide Saelα had all of its 13 Ser/Thr residues dehydrated, and all 9 Cys residues formed thioether linkages, making this peptide, to our knowledge, the most highly cyclized lanthipeptide reported in literature. Saelβ exhibited less extensive post-translational modification, with 9 Ser/Thr dehydrations and 5 thioether rings. One of the Saelβ Thr residues appears to remains unmodified in the final product, unlike in Saelα, and we were able to narrow down the identity of the unmodified residue to either Thr7 or Thr9. The peptides were screened against a panel of bacteria, however no antibiotic activity was observed. While it is likely that the peptides may exhibit activity against a bacterial strain we have not tested, it may also have a novel function. In contrast, it is also possible that these peptide serve some other, novel function. To that end, we tested whteher the peptides could act as metal chelators, similar to another RiPP - methanobactin. However, the peptides did not exhibit affinity for the metals tested. It is also possible that there remains an unidentified component of the saelicesin gene cluster that is needed to confer antimicrobial activity of the peptides. To help answer questions around the role of the Sael system, the Thibodeaux lab is currently collaborating with the Trujillo lab in the University of Salamanca, Spain, who are experts in the biology of *Micromonospora* soil bacteria and the symbiotic interaction of these organisms with plants. They will be testing for a variety of novel lanthipeptide functions including as antifungal agents, or as compounds that regulate plant gene expression or defensive pathways.

In the second study, we sought to probe the conformational dynamics of the enzyme HalM2, a class II lanthipeptide synthetase that plays a role in the synthesis of the haloduracins, a known two-component lantibiotic. It is hoped that any insights into the function of HalM2 can help inform future attempts at rationally engineering HalM2, and by extension other LanMs, to produce novel antibiotics. Previous studies have shown that conformational sampling potentially plays an important role in the [HalM2:HalA2] Michaelis complex that forms during peptide maturation. To that end, we developed a method that employs hydrogen-deuterium exchange mass

spectrometry to help identify regions in the protein structure that may be functionally relevant. Probing the effect of ligand binding on HalM2 exchange patterns, we were able to identify several regions of interest and provide the first insights into the relation between structural dynamics and function in HalM2. Overall, we note that apo-HalM2 is rather structurally dynamic. Upon binding the ATP analogue AMP-PNP, a general relaxation in the structure is observed. The observed relaxation suggests that AMP-PNP binding primes HalM2 to bind HalA2 and leads to easier access of residues involved in binding. When HalA2 is introduced, several regions of HalM2 show considerably perturbed exchange patters, particularly in the *N*-terminal dehydratase domain. The residues I627-G652, which are predicted to have β-sheet character, constitute the only region in the cyclase domain to show significant protection in the presence of HalA2, and we propose that it acts as a RiPP recognition element. Interestingly, the residues A572-F600 are unique among those residues that are perturbed by ligand binding in that they are distal from the active face of the enzyme, suggesting that the region may be involved in long range structural communication rather than binding ligands. We proposed a series of mutations for future studies to determine if the identified regions are, in fact functionally relevant. As such, this study provides new insights into how HalM2, and other LanMs by extension, carry out their function. It is hoped that the information provided here can help inform any attempts at rationally engineering LanMs in the future.

5. References

1. Spellberg, B., Dr. William H. Stewart: Mistaken or Maligned? *Clin Infect Dis* **2008,** *47* (2), 293-294.
2. Rosenblatt-Farrell, N., The Landscape of Antibiotic Resistance.
3. Center, U. o. C. M. R. http://mrsa-research-center.bsd.uchicago.edu/timeline.html.
4. Levine, D. P., Vancomycin: A History.
5. Correia, S.; Poeta, P.; Hebraud, M.; Capelo, J. L.; Igrejas, G., Mechanisms of quinolone action and resistance: where do we stand? *J Med Microbiol* **2017,** *66* (5), 551-559.
6. Margalith, P.; Beretta, G., Rifomycin. XI. taxonomic study on streptomyces mediterranei nov. sp. *Mycopathologia et Mycologia Applicata* **1960,** *13* (4), 321-330.
7. Floss, H. G.; Yu, T.-W., RifamycinMode of Action, Resistance, and Biosynthesis. **2005,** *105* (2), 621-632.
8. Gautam, A.; Vyas, R.; Tewari, R., Peptidoglycan biosynthesis machinery: A rich source of drug targets. **2011,** *31* (4), 295-336.
9. Jarick, M.; Bertsche, U.; Stahl, M.; Schultz, D.; Methling, K.; Lalk, M.; Stigloher, C.; Steger, M.; Schlosser, A.; Ohlsen, K., The serine/threonine kinase Stk and the phosphatase Stp regulate cell wall synthesis in Staphylococcus aureus. *Scientific Reports* **2018,** *8* (1), 13693.
10. Organization, W. H. https://www.who.int/news-room/fact-sheets/detail/tuberculosis.
11. Pai, M.; Behr, M. A.; Dowdy, D.; Dheda, K.; Divangahi, M.; Boehme, C. C.; Ginsberg, A.; Swaminathan, S.; Spigelman, M.; Getahun, H.; Menzies, D.; Raviglione, M., Tuberculosis. *Nature Reviews Disease Primers* **2016,** *2*, 16076.
12. Angala, S. K.; Belardinelli, J. M.; Huc-Claustre, E.; Wheat, W. H.; Jackson, M., The cell envelope glycoconjugates of Mycobacterium tuberculosis. *Critical Reviews in Biochemistry and Molecular Biology* **2014,** *49* (5), 361-399.
13. Lewis, K., Platforms for antibiotic discovery. *Nat Rev Drug Discov* **2013,** *12* (5), 371-87.
14. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **2001,** *46* (1-3), 3-26.
15. Harvey, A. L., Natural products in drug discovery. *Drug Discov Today* **2008,** *13* (19-20), 894-901.
16. Cotter, P. D.; Ross, R. P.; Hill, C., Bacteriocins - a viable alternative to antibiotics? *Nat Rev Microbiol* **2013,** *11* (2), 95-105.
17. Geng, M.; Smith, L., Modifying the lantibiotic mutacin 1140 for increased yield, activity and stability. *Applied and Environmental Microbiology* **2018**, AEM.00830-18.
18. Süssmuth, R. D.; Mainz, A., Nonribosomal Peptide Synthesis-Principles and Prospects. *Angewandte Chemie International Edition* **2017,** *56* (14), 3770-3821.
19. Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K. D.; Fischbach, M. A.; Garavelli, J. S.; Goransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Muller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Ross, R. P.; Sahl, H. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Sussmuth, R. D.; Tagg, J. R.; Tang, G. L.;

Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A., Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **2013,** *30* (1), 108-60.
20. Hetrick, K. J.; Van Der Donk, W. A., Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Current Opinion in Chemical Biology* **2017,** *38*, 36-44.
21. Hudson, G. A.; Mitchell, D. A., RiPP antibiotics: biosynthesis and engineering potential. *Curr Opin Microbiol* **2018,** *45*, 61-69.
22. Guiotto, A.; Pozzobon, M.; Canevari, M.; Manganelli, R.; Scarin, M.; Veronese, F. M., PEGylation of the antimicrobial peptide nisin A: problems and perspectives. *Il Farmaco* **2003,** *58* (1), 45-50.
23. Burkhart, B. J.; Kakkar, N.; Hudson, G. A.; Van Der Donk, W. A.; Mitchell, D. A., Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products. *ACS Central Science* **2017,** *3* (6), 629-638.
24. Wilson, K.-A.; Kalkum, M.; Ottesen, J.; Yuzenkova, J.; Chait, B. T.; Landick, R.; Muir, T.; Severinov, K.; Darst, S. A., Structure of Microcin J25, a Peptide Inhibitor of Bacterial RNA Polymerase, is a Lassoed Tail. *Journal of the American Chemical Society* **2003,** *125* (41), 12475-12483.
25. de Arauz, L. J.; Jozala, A. F.; Mazzola, P. G.; Vessoni Penna, T. C., Nisin biotechnological production and application: a review. *Trends in Food Science & Technology* **2009,** *20* (3-4), 146-154.
26. Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A., Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem Rev* **2017,** *117* (8), 5457-5520.
27. Dong, S. H.; Tang, W.; Lukk, T.; Yu, Y.; Nair, S. K.; van der Donk, W. A., The enterococcal cytolysin synthetase has an unanticipated lipid kinase fold. *Elife* **2015,** *4*.
28. Hegemann, J. D.; Shi, L.; Gross, M. L.; van der Donk, W. A., Mechanistic Studies of the Kinase Domains of Class IV Lanthipeptide Synthetases. *ACS Chemical Biology* **2019,** *14* (7), 1583-1592.
29. Hegemann, J. D.; van der Donk, W. A., Investigation of Substrate Recognition and Biosynthesis in Class IV Lanthipeptide Systems. *Journal of the American Chemical Society* **2018,** *140* (17), 5743-5754.
30. Booth, M. C.; Bogie, C. P.; Sahl, H.-G.; Siezen, R. J.; Hatter, K. L.; Gilmore, M. S., Structural analysis and proteolytic activation of Enterococcus faecalis cytolysin, a novel lantibiotic. **1996,** *21* (6), 1175-1184.
31. Cotter, P. D.; Connor, P. M.; Draper, L. A.; Lawton, E. M.; Deegan, L. H.; Hill, C.; Ross, R. P., Posttranslational conversion of l-serines to d-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (51), 18584.
32. McAuliffe, O.; Ryan, M. P.; Ross, R. P.; Hill, C.; Breeuwer, P.; Abee, T., Lacticin 3147, a broad-spectrum bacteriocin which selectively dissipates the membrane potential. *Applied and environmental microbiology* **1998,** *64* (2), 439-445.
33. Ziemert, N.; Alanjary, M.; Weber, T., The evolution of genome mining in microbes - a review. *Nat Prod Rep* **2016,** *33* (8), 988-1005.

34. Medema, M. H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **2011,** *39* (suppl), W339-W346.
35. Blin, K.; Kim, H. U.; Medema, M. H.; Weber, T., Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Briefings in Bioinformatics* **2017**.
36. McClerren, A. L.; Cooper, L. E.; Quan, C.; Thomas, P. M.; Kelleher, N. L.; van der Donk, W. A., Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc Natl Acad Sci U S A* **2006,** *103* (46), 17243-8.
37. Begley, M.; Cotter, P. D.; Hill, C.; Ross, R. P., Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Applied and environmental microbiology* **2009,** *75* (17), 5451-5460.
38. Zhao, X.; van der Donk, W. A., Structural Characterization and Bioactivity Analysis of the Two-Component Lantibiotic Flv System from a Ruminant Bacterium. *Cell chemical biology* **2016,** *23* (2), 246-256.
39. Li, B.; Sher, D.; Kelly, L.; Shi, Y.; Huang, K.; Knerr, P. J.; Joewono, I.; Rusch, D.; Chisholm, S. W.; van der Donk, W. A., Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proceedings of the National Academy of Sciences* **2010,** *107* (23), 10430.
40. Trujillo, M. E.; Kroppenstedt, R. M.; Fernandez-Molinero, C.; Schumann, P.; Martinez-Molina, E., Micromonospora lupini sp. nov. and Micromonospora saelicesensis sp. nov., isolated from root nodules of Lupinus angustifolius. **2007,** *57* (12), 2799-2804.
41. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **2009,** *6* (5), 343-345.
42. Zhang, Q.; Ortega, M.; Shi, Y.; Wang, H.; Melby, J. O.; Tang, W.; Mitchell, D. A.; van der Donk, W. A., Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS. *Proc Natl Acad Sci U S A* **2014,** *111* (33), 12031-6.
43. Hegemann, J. D.; van der Donk, W. A., Investigation of Substrate Recognition and Biosynthesis in Class IV Lanthipeptide Systems. *J Am Chem Soc* **2018,** *140* (17), 5743-5754.
44. McCabe, J. W.; Vangala, R.; Angel, L. A., Binding Selectivity of Methanobactin from Methylosinus trichosporium OB3b for Copper(I), Silver(I), Zinc(II), Nickel(II), Cobalt(II), Manganese(II), Lead(II), and Iron(II). *J Am Soc Mass Spectrom* **2017,** *28* (12), 2588-2601.
45. Dimise, E. J.; Widboom, P. F.; Bruner, S. D., Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile <em>Thermobifida fusca</em&gt. *Proceedings of the National Academy of Sciences* **2008,** *105* (40), 15311.
46. Ming, L.-J.; Epperson, J. D., Metal binding and structure–activity relationship of the metalloantibiotic peptide bacitracin. *Journal of Inorganic Biochemistry* **2002,** *91* (1), 46-58.
47. V. M. Conn, A. R. W. a. C. M. M. F., Endophytic Actinobacteria Induce Defense Pathways in Arabidopsis thaliana. *Molecular Plant-Microbe Interactions* **2007,** *21* (2), 208-218.
48. Khosla, C.; Keasling, J. D., Metabolic engineering for drug discovery and development. *Nature Reviews Drug Discovery* **2003,** *2* (12), 1019-1025.
49. Tokuriki, N.; Tawfik, D. S., Protein Dynamism and Evolvability. *Science* **2009,** *324* (5924), 203-207.

50. Thibodeaux, C. J.; Ha, T.; van der Donk, W. A., A Price To Pay for Relaxed Substrate Specificity: A Comparative Kinetic Analysis of the Class II Lanthipeptide Synthetases ProcM and HalM2. *Journal of the American Chemical Society* **2014,** *136* (50), 17513-17529.
51. Uguen, P.; Le Pennec, J. P.; Dufour, A., Lantibiotic Biosynthesis: Interactions between Prelacticin 481 and Its Putative Modification Enzyme, LctM. *Journal of Bacteriology* **2000,** *182* (18), 5262-5266.
52. You, Y. O.; van der Donk, W. A., Mechanistic investigations of the dehydration reaction of lacticin 481 synthetase using site-directed mutagenesis. *Biochemistry* **2007,** *46* (20), 5991-6000.
53. Cooper, L. E.; McClerren, A. L.; Chary, A.; van der Donk, W. A., Structure-activity relationship studies of the two-component lantibiotic haloduracin. *Chem Biol* **2008,** *15* (10), 1035-1045.
54. Zhang, Z.; Smith, D. L., Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. **1993,** *2* (4), 522-531.
55. Jensen, P. F.; Rand, K. D., Hydrogen Exchange. In Hydrogen Exchange Mass Spectrometry of Proteins. Weis, D. D., Ed. Wiley: 2016; pp. 1-17. https://doi.org/10.1002/9781118703748.ch1 (accessed 03/16).
56. Wang, L. C., Krishnamurthy, S. and Anand, G. S., Hydrogen exchange mass spectrometry of proteins : fundamentals, methods, and applications. Weis, D. D., Ed. Wiley: Chichester, West Sussex, 2016.
57. Vilar, M.; Saurí, A.; Marcos, J. F.; Mingarro, I.; Pérez-Payá, E., Transient Structural Ordering of the RNA-Binding Domain of Carnation Mottle Virus p7 Movement Protein Modulates Nucleic Acid Binding. **2005,** *6* (8), 1391-1396.
58. Iizuka, R.; Yoshida, T.; Shomura, Y.; Miki, K.; Maruyama, T.; Odaka, M.; Yohda, M., ATP Binding Is Critical for the Conformational Change from an Open to Closed State in Archaeal Group II Chaperonin. **2003,** *278* (45), 44959-44965.
59. Krasteva, M.; Barth, A., Structures of the Ca2+–ATPase complexes with ATP, AMPPCP and AMPPNP. An FTIR study. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2007,** *1767* (1), 114-123.
60. Theillet, F.-X.; Kalmar, L.; Tompa, P.; Han, K.-H.; Selenko, P.; Dunker, A. K.; Daughdrill, G. W.; Uversky, V. N., The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins* **2013,** *1* (1), e24360.