Discovery and Characterization of the Streptozotocin Biosynthetic Pathway

Discovery and Characterization of the Streptozotocin Biosynthetic Pathway

A dissertation presented

by

Tai Lun Ng

to

The Department of Chemistry and Chemical Biology

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

January 2020

Dissertation advisor: Professor Emily P. Balskus Tai Lun Ng

Discovery and Characterization of the Streptozotocin Biosynthetic Pathway

Abstract

Natural products have historically provided a rich source of medicines, and natural product isolation and characterization continue to drive drug discovery and design. As the availability of genomic DNA sequences in public databases continues to increase due to decreasing costs, we now have access to an abundance of gene clusters that encode uncharacterized natural product biosynthetic pathways. Therefore, by understanding the biosynthetic processes that generate natural products, we will advance our ability to mine this abundance of available microbial sequencing data and may facilitate the discovery of additional naturally-derived bioactive compounds. Detailed *in vitro* characterization of enzymes that catalyze the chemical transformations within these biosynthetic pathways will enable applications in genome mining, biocatalysis, and design of organometallic catalysts. Here we describe the discovery and characterization of novel enzymes that install a reactive *N*-nitroso functional group during streptozotocin biosynthesis. These efforts have provided our first insights into enzymatic *N*-nitrosation in biological systems. With the elucidation of the genetic and biochemical basis for *N*-nitroso biosynthesis, our work now enables investigations of the biosynthesis of additional *N*-nitroso and N–N bond containing natural products and targeted genome mining efforts to identify additional *N*-nitroso compounds.

In **Chapter 2** of this dissertation, we present the identification and validation of the putative streptozotocin biosynthetic gene cluster. This gene cluster was identified using a resistance-guided genome mining strategy. Experimental work linked this gene cluster to the production of streptozotocin. *in vivo* genetic analysis revealed four essential biosynthetic genes. *In vivo* feeding

experiments and *in vitro* biochemical characterizations also outlined the early stages of streptozotocin biosynthesis.

Chapter 3 of this work details the biochemical characterization of the *N*-nitrosating metalloenzyme SznF. Biochemical characterization revealed a sequential oxygenation reaction on a modified L-arginine substrate. Reaction intermediates were accessed via chemical synthesis, and stable isotope tracer experiments provided insights into the mechanism of *N*-nitroso formation. We also pursued structural characterization of SznF and used site-directed mutagenesis experiments to reveal two distinct metallocofactor sites in this enzyme that have distinct catalytic capabilities. Future efforts are focused on further structural and spectroscopic characterization to elucidate the mechanism of SznF-mediated *N*-oxidation and N–N bond formation.

Finally, in **Chapter 4,** we will describe the roles of the ATP-grasp enzymes encoded in the streptozotocin biosynthetic gene cluster and assess the distribution of SznF-homologs in microbial genomes. We have reconstituted the *in vitro* activity of SznH and K and established they are L-amino acid ligases. We have concluded that these two enzymes further modify an *N*-nitrosourea-containing amino acid to generate a tripeptide product. These results suggested that additional enzyme(s) may be required for completing streptozotocin biosynthesis. Finally, phylogenetic analysis of SznF and SznF-neighborhood revealed a widely-distributed enzyme class in multiple genus of bacteria in different environments. Future efforts are focused on identifying additional enzymes required to complete streptozotocin biosynthesis and biochemically characterizing SznF-homologs identified in our bioinformatic analyses.

Acknowledgements

I acknowledge my colleagues and mentors who have assisted me during my graduate studies. I would like to first thank Prof. Emily P. Balskus for her guidance, support, and mentorship in the last five years. I would like to acknowledge Dr. Matt R. Wilson, Dr. Abraham J. Waldman, Dr. Peng Wang for their help during the first few years during my graduate studies. All the former and current colleagues in the Balskus Lab have assisted me in the technical aspects of the project in some ways.

I would like to thank our collaborators Dr. Andrew J. Mitchell, Dr. Roman Rohac, and Professor Amie K. Boal for their help with the structural characterization of SznF. They were also helpful for discussions regarding the mechanism of SznF presented in Chapter 3. I would like to thank our collaborators Molly McBride, Dr. Lauren J. Rajakovich, Dr. Debangsu Sil, Dr. Bo Zhang, Professor J. Martin Bollinger, and Professor Carsten Krebs for their helpful discussions about the mechanism of SznF and for performing experimental work on elucidating the mechanism of the SznF CADD domain.

I am also grateful to my committee members Prof. Suzanne Walker and Prof. David R. Liu for their advice. I would like to thank Dr. Nathaniel Glasser, Matt Volpe, Beverly Fu, Dr. Lauren Rajakovich, and Dr. Matt Wilson for proofreading documents, including this thesis. I would also like to thank Dr. Grace E. Kenney, Dr. Monica E. McCallum, Dr. Jennifer Wang, Christine R. Zheng, and Kelvin Wu for their work on our projects not described in this thesis.

Table of Content

List of abbreviations

ºC degree	Celsius
COSY	correlation spectroscopy d deuterium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDC•HCl	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide•hydrochloride
eV	electron volt
g	gram
h	hour
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	coupling constant
L	liter
LC-MS	liquid chromatography-mass spectrometry
M	molar
min	minute
mol	mole
MS/MS	tandem mass spectrometry
MWCO	molecular weight cut-off

NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NRRL	northern regional research laboratory
OD	optical density
ppm	parts per million
psi	pounds per square inch
PCR	polymerase chain reaction
RNA	ribonucleic acid
rpm	revolutions per minute
s	second
SZN	streptozotocin
UV	ultraviolet
V	volt
Vis	visible
×g	times gravity

Chapter 1. Natural product biosynthesis and the assembly of N–N bond-containing functional groups in living systems

Part of this Chapter is adapted from previous publication written with Dr. Abraham J. Waldman and Dr. Peng Wang.[1]

1.1 Introduction to the study of natural product biosynthesis

1.1.1 Natural products and their roles in human health

Natural products are small molecules produced by all three kingdoms of life that are not essential for the producing organism's survival.[2] While natural products are often referred to as 'secondary metabolites' in the literature, these compounds can confer important advantages to the host, such as chemical defense and intra- and interspecies signaling.[3-4] Historically, a majority of natural products isolated and characterized to date have been plant- and microbially-derived, particularly from the *Streptomyces* genus of bacteria. Recent studies have shown that marine actinomyces (e.g. *Salinospora*) are also prolific producers of natural products, rivaling related soil-dwelling microbes.[5] In addition to microbes, plants also synthesize a diverse set of bioactive molecules since their sessile nature necessitates the development of chemical defenses.[6] Current research has also investigated natural products produced by microbes from underexplored environments, such as hot-springs, the human microbiome, and plant- and sponge-symbionts with an eye towards discovering structurally unusual bioactive molecules.[7-11]

Natural products are structurally diverse (**Figure 1.1**). Classes of natural products include, but are not limited to, ribosomally-synthesized post-translationally modified peptides (RiPPs), non-ribosomal peptides, polyketides, alkaloids, steroids, terpenes, carbohydrates, and modified amino acids. These different scaffolds can be functionalized with an array of substituents, including methyl-, hydroxyl-, glycosyl-, and/or halogen groups. These

modifications can alter the physical properties (e.g. solubility) and bioactivities of the molecules.[12]

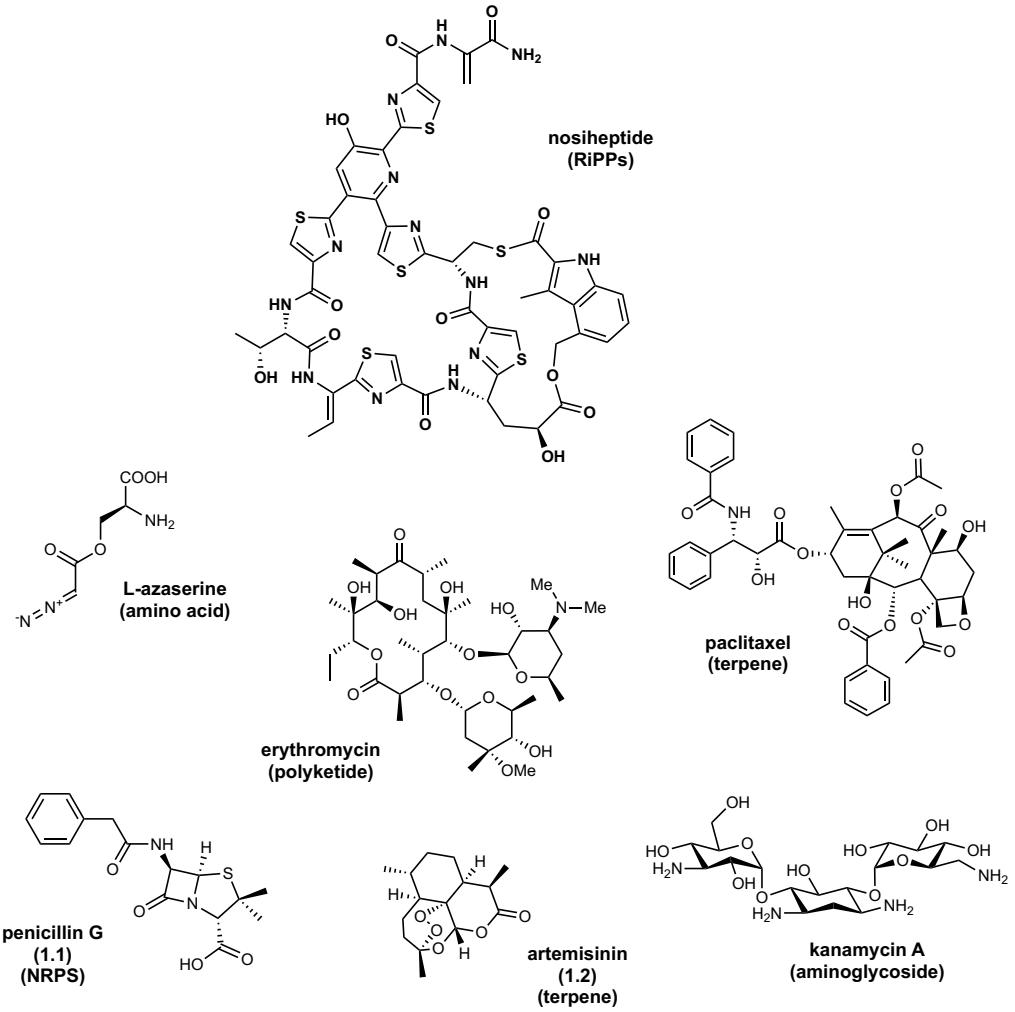


Figure 1.1: Natural products are structurally diverse. Some major classes of natural products include modified amino acids, nonribosomal peptides, terpenes, and aminoglycosides. Some common modifications to the core scaffolds include hydroxyl-, methyl-, and/or halogen groups.

Interest in natural product research has been accelerated by drug discovery programs led by the government, academia, and industry. Historically, natural products have played pivotal roles in drug development.[13] For example, the discovery of artemisinin from the sweet wormwood *Artemisia annua* (awarded 2015 Nobel Prize in Physiology or Medicine) has saved millions of lives from malaria (**1.1**, **Figure 1.1**).[14] The discovery of the antibiotic penicillin from *Penicillium chrysogenum* also marked the start of the 'Golden Age' of natural product discovery from

microorganisms (**1.2**, **Figure 1.1**).[15] Since then, more than 23,000 natural products produced by plants and microbes, in particular actinobacteria, have been characterized. While many isolated natural products are bioactive, they may not possess the ideal properties to be used directly as therapeutics. Therefore, much effort has also been invested in developing semi-synthetic natural product analogues.[16-17] As of 2014, roughly 83% of all small molecule drugs on the market are naturally derived, derivatives, or semi-synthetic analogues.[18] Altogether, naturally-derived drugs have contributed to treating cancer, bacterial and fungal infections, and autoimmune and cardiovascular diseases.

Research efforts to discover new natural products and characterize their biosyntheses by living organisms remain active. Sections 1.1.2 and 1.1.3 will describe how recent advances in genome sequencing and analytical techniques have accelerated natural product research. Ultimately, these approaches can not only streamline the discovery of new natural products, but also uncover new enzymatic chemistry that could be useful in biocatalysis and catalyst design. Part of this dissertation will focus on using these genome mining techniques to uncover the biosynthesis of an anticancer natural product streptozotocin and discover a novel *N*-nitrosating enzyme.

1.1.2 Natural product discovery in the post-genomic era

In microorganisms, the set of genes that encode for the biosynthetic enzymes that build a particular natural product are often colocalized within the producing organism's genome in a unit called a biosynthetic gene cluster (BGC). In addition to biosynthetic enzymes, transporter or self-resistance genes are often encoded within these clusters, which enables the host to produce enzymes that could sequester the toxic natural product. To understand the biosynthesis of a natural product, genomic libraries can be constructed and expressed heterologously in a suitable host to

identify the genes responsible for its biosynthesis. Knocking out genes in the native producer has also been used extensively to identify essential genes require for biosynthesis.

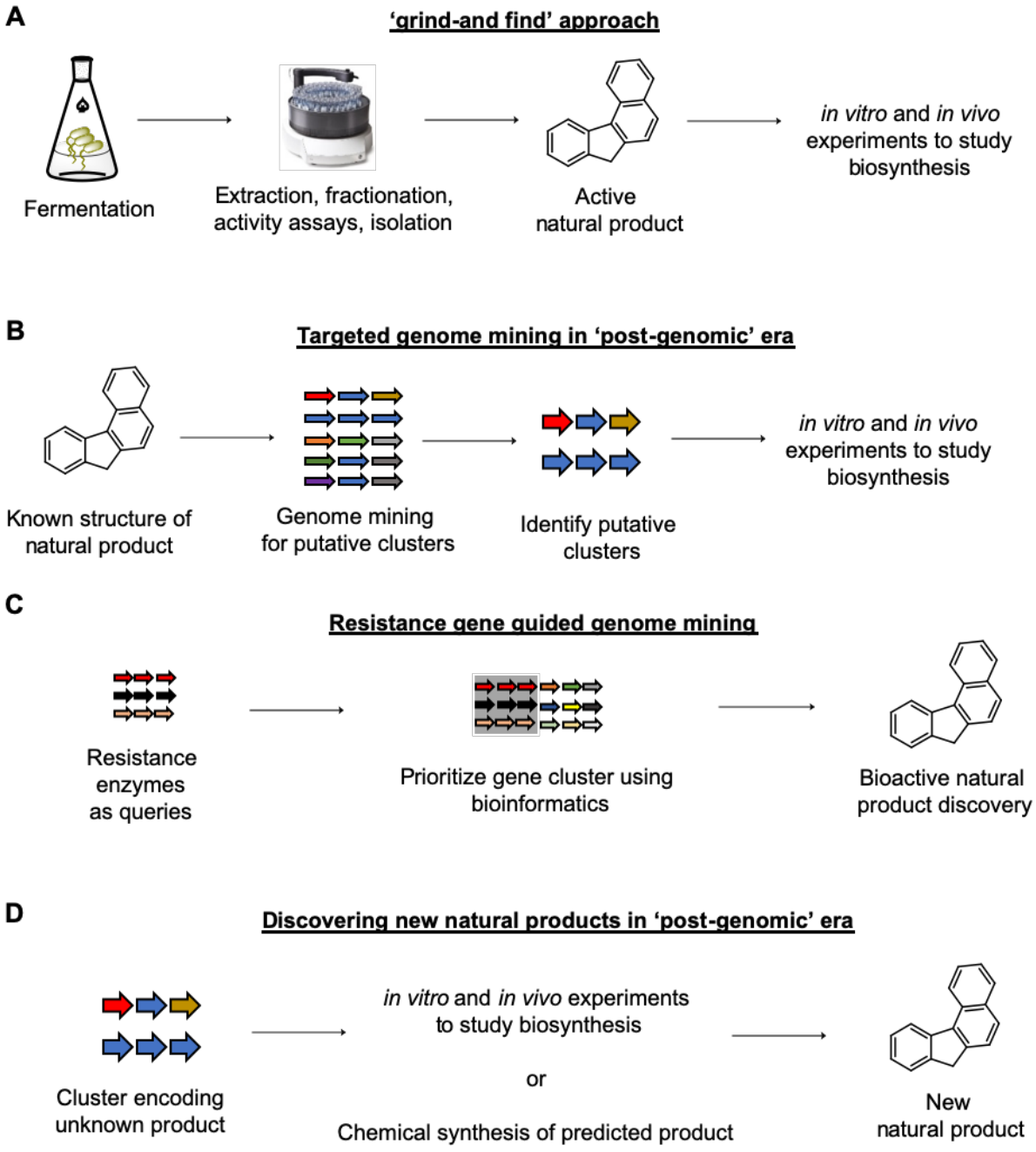


Figure 1.2: Natural product discovery before and after genome sequences became widely available. **A**) Traditional natural product discovery and characterization process. **B**) Target-oriented genome mining workflow. **C**) Resistance genes can be used as queries to discover gene clusters encoding known natural products or new bioactive natural products. **D**) Discovery of cryptic natural products in the post genomic era.

With the advent of increasing amount of publicly available microbial genomes and metagenomic sequences, natural product scientists now face many exciting opportunities and challenges.[19-20] The number of microbial genome sequences far outpaces the biochemical characterization of the corresponding gene products.[21] These new genomic sequences also revealed many putative biosynthetic gene clusters, such as gene clusters containing nonribosomal peptide synthetase, terpene cyclase, or polyketide synthase genes. However, the products of these gene clusters remain unknown because either the analytical methods are not sensitive enough to detect them, or the transcription of the biosynthetic genes is repressed. In the latter case, these gene clusters are referred to as 'cryptic' or 'silent' gene clusters. Therefore, the number of potential secondary metabolites produced by organisms are vastly under-sampled.

This ease of sequencing whole genomes has lowered the barrier of connecting natural products to their biosynthetic pathways (**Figure 1.2B**). In a typical genome mining process, one will inspect the natural product and identify chemical functionalities or structural features constructed by characterized biosynthetic pathways. The enzymes known to form those functional groups or molecular features are then used to search the genome of the producing organism for related genes. Then, manual inspection of the genomic context, such as the presence of additional predicted biosynthetic genes, leads to identification of a putative gene cluster.

In addition to biosynthetic genes, resistance genes have also been used in genome mining strategies (**Figure 1.2C**). As mentioned above, resistance genes in a biosynthetic gene cluster can encode enzymes that sequester a toxic natural product. One common resistance strategy is for the organism to express a resistant copy of the natural product's target. For example, a natural product that inhibits fatty acid synthase might result in a redundant copy of a mutated fatty acid synthase in the gene cluster to confer resistance to the native organism.[22] We could leverage this knowledge

to prioritize putative gene clusters that could encode natural products targeting fatty acid synthase. This approach has the advantage of linking a cryptic natural product to its bioactivity.[22] In another example, Tang *et al*. has employed this strategy to discovery a fungal secondary metabolite with herbicidal activity by searching for gene clusters that contain a redundant copy of dihydroxyacid dehydratase, a frequent target of herbicides.[23] Numerous computational tools like anti-SMASH ClusterScout (IMG/JGI) have been created to help prediction of natural product gene clusters.[24-25] More recent tools like Antibiotic Resistant Target Seeker (ARTS) have incorporated resistance models in the genome mining workflow, allowing users to prioritize clusters with desired bioactivities. Once a putative biosynthetic cluster is identified, one would use *in vivo* and *in vitro* techniques to verify the connection between the gene cluster and metabolite production. For example, *in vivo* gene knockouts and heterologous expression can lead to identification of essential genes for biosynthesis. *In vitro* assays of biosynthetic enzymes can confirm the biochemical functions in an isolated matrix. The discovery of the streptozotocin cluster discussed in this thesis will utilize both the biosynthetic enzymes and resistance genes as genome mining targets.

In addition to the abundance of genomic sequences, better gene synthesis and improved molecular biology techniques, and more sensitive analytical methods (e.g. high-resolution mass spectrometry (HR-MS), imaging mass spectrometry, high-throughput ultra-performance liquid chromatography (UPLC), and nuclear magnetic resonance (NMR) spectrometry) have increased the throughput of natural product discovery. Traditionally, new natural products have been discovered via the 'grind-and-find' approach (**Figure 1.2A**).[26-28] Typically, a single species of bacteria is isolated and cultured in a suitable fermentation medium. Crude or fractionated fermentation extracts are then screened for interesting bioactivity such as anti-proliferative or antibiotic activity. Further purification and analytical characterization lead to isolation and

identification of the bioactive molecule(s). While this approach has been remarkably successful, this process is labor intensive and low throughput. This method is also limited to microorganisms that can be cultured, prone to rediscovery of known natural products, and bias discovery towards non-cryptic secondary metabolites.

Methods to capture gene clusters from genomic and environmental DNA as well as a suite of heterologous hosts with well-characterized promoters have been explored to expedite natural product discovery, particularly in the context of cryptic gene clusters (**Figure 1.2D**).[29-32] The aforementioned analytical methods allow for lower detection limit of analytes, which can improve detection of natural products produced at low abundance. Methods that incorporate the use of stable isotope building blocks or molecular networking methods can be used to bias searches of metabolomics data for certain types of secondary metabolites.[33-34] Scientists have also synthesized predicted natural products for testing bioactivity, obviating the need for isolation (**Figure 1.2D**).[35-36] This strategy is amenable to pathways in which the substrates of the enzymes can be predicted bioinformatically, such as nonribosomal peptide synthases. Finally, advances in transcriptomics have enabled the discovery of biosynthetic genes that are scattered across a host genome. For natural products in which production is induced by an external stimulus, the biosynthetic genes would be transcribed, and the difference in gene expression can be detected with methods like RNA-sequencing. This approach is greatly advancing our understanding of natural product biosynthesis in plants and marine diatoms.[37-38]

1.1.3 Motivation for discovering and characterizing new enzymatic chemistry from natural product biosynthetic pathways

Studying the biosynthesis of natural products can reveal new enzymes for biocatalysis and metabolic engineering applications.[39-42] Due to their ability to perform reactions with high degrees of chemo-, enantio-, and regioselectivity under ambient conditions in aqueous solvents, enzymes can be powerful tools for chemical synthesis. Enzymes can be also engineered to accept new substrates or expressed in novel metabolic pathways to produce commodity chemicals. Utilization of enzymes in biotechnology industries has traditionally been limited to a just a few classes, including transaminases, lipases, esterases, and dehydrogenases. By studying the biosynthesis of unusual functional groups present in natural products, one may uncover novel enzymes that catalyze new reactions useful in industry.

Discovering new enzymes and elucidating their mechanisms also can provide a roadmap for discovering additional natural products with similar functionalities. Using the new enzyme(s) as a query, genome mining can reveal similar biosynthetic chemistry in biosynthetic gene clusters encoded in other organisms. Detecting homologs of the enzyme(s) of interest in unusual genomic neighborhoods may aid prioritization of new gene clusters for further investigation.

Finally, discovering and dissecting enzyme mechanisms can inspire the design of organo- and metal-based catalysts. For example, studying how first- and second-sphere active site residues in enzymes control substrate orientation and hydrogen bonding interactions may provide new strategies for catalyst design.[43-44] Studies of non-heme and diiron metalloenzymes have inspired the development of small molecule mimics that can catalyze C–H functionalization reactions.[45] This area of research could benefit from uncovering and studying new enzymes from biosynthetic pathways.

1.2 Biosynthesis of natural products containing an N–N bond

N–N bond-containing natural products represent a small, but structurally diverse class of metabolites.[46-47] There are approximately 200 members of this natural product family, ranging from the simple di-amino acid azoxy compound valanimycin to the complex diazobenzofluorene polyketide lomaiviticin (**1.18**, **Figure 1.8**). This group of natural products encompasses a wide range of functional groups, including azoxy, diazo, hydrazido, hydrazino, heterocyclic, and *N*-nitroso motifs. Many of these functional groups are believed to impart biological activity to the natural product, such as the diazo group in the kinamycins and the *N*-nitroso group in streptozotocin (discussed below).

N–N bond formation is a chemically challenging transformation as both atoms being linked in many nitrogen-containing metabolites are usually nucleophilic.[48] Therefore, enzymatic N–N bond construction requires reversal of this reactivity for one of the amine reaction partners. At the beginning of the thesis, no dedicated enzyme had been discovered to catalyze N–N bond formation in natural product assembly. However, recent investigations of natural product biosynthetic pathways have uncovered that a range of different enzymes promote this general transformation in the context of building different functional groups. The following section will discuss our current understanding of the enzymes and chemical logic involved in the construction of various N–N bond containing natural products and is organized by the type of functional group. Only natural products in which a biosynthetic gene cluster has been identified and experimentally verified will be discussed.

1.2.1 Diazo

Diazo-containing natural products are a small subset of N–N bond containing natural products that exhibit a range of structural diversity.[1,49] This group of secondary metabolites has

generated interest among chemists given their potent bioactivities, interesting mechanisms of actions, and unknown biosyntheses. Among the first diazo-containing metabolites discovered were modified α-amino acids, such as L-azaserine.[50] The bioactivities of diazo-containing compounds are thought to arise from the reactivity of the diazo functional group. For example, the mechanism of action of the diazo-containing natural product lomaiviticin A (**1.18**, **Figure 1.8**) has been elucidated, and accumulation of DNA double strand breaks requires the presence of its two diazo groups.[51]

The study of the biosynthesis of cremeomycin provided some insight into the biosynthesis of diazo groups. Cremeomycin (**1.5**, **Figure 1.3**) is an *o*-diazoquinone-containing natural product that was isolated in 1967 from the soil-dwelling bacterium *Streptomyces cremeus*.[52] In 2006, a report described 3-amino-4-hydroxybenzoic acid (3,4-AHBA; **1.6**) could be a possible biosynthetic intermediate.[53] Recently, heterologous expression of the cremeomycin (*cre*) biosynthetic gene cluster, feeding studies with stable isotope-labeled intermediates, and *in vitro* and *in vivo* characterization of biosynthetic enzymes confirmed the identity of the gene cluster and elucidated the order of the events in this pathway.[54-55]

Biosynthesis of cremeomycin begins with the production of **1.6** by the cyclase, CreH, and aldolase, CreI, from the primary metabolites L-aspartate semialdehyde (L-ASA, **1.7**) and dihydroxyacetone phosphate (DHAP, **1.8**). Synthesis of 3,4-AHBA is followed by C2 hydroxylation by the flavin-dependent monooxygenase CreL to provide 3-amino-2,4-dihydroxybenzoic acid (3,2,4-ADBA; **1.9**), which is then *O*-methylated at the C4 phenol to afford 3,2,4-AHMBA (**1.10**). Using a heterologous expression system in *Streptomyces albus,* systematic disruption of the remaining genes within the *cre* gene cluster revealed several enzymes that are essential for diazo group installation. While *creA*, *creB*, *creO*, and *creM* initially appeared to not

be involved in diazotization, Δ*creD* and Δ*creE* mutants both failed to produce cremeomycin and accumulated **1.10**.

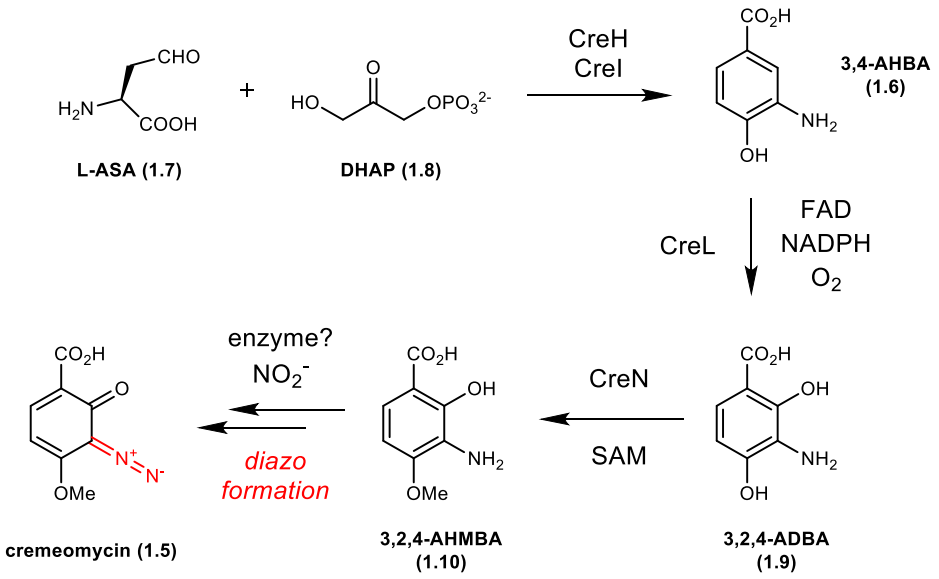


Figure 1.3: Cremeomycin biosynthesis involves a late-stage diazotization event.

These gene products were characterized *in vitro* to elucidate their roles in diazo biosynthesis. CreD is predicted to belong in the fumarase/aspartase enzyme superfamily, which catalyzes the elimination of β-heteroatoms, such as nitrogen-containing amines, from substrates with an aspartate backbone. CreE is annotated as a predicted flavin-dependent monooxygenase, a family of enzymes that could perform *N*-hydroxylation reactions to form hydroxamates in siderophore biosynthesis.[56] This enzyme was postulated to oxidize the amino group of either **1.10** or an amino acid. *In vitro* biochemical characterization of CreE demonstrated that it oxidizes L-aspartate to nitrosuccinate **1.11** (**Figure 1.4**). CreE consumed three equivalents of NADPH for every molecule of **1.11** produced, indicating that it catalyzes the full six-electron oxidation. This represents the first flavin-dependent monooxygenase capable of catalyzing the complete six-electron oxidation required to convert a primary amine into a nitro group.

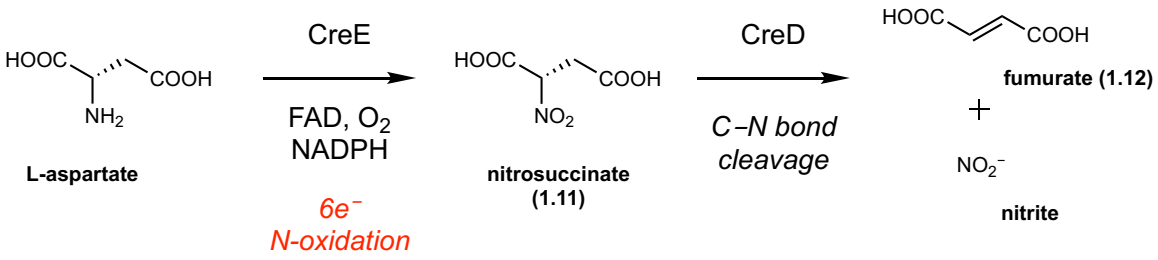


Figure 1.4: CreD and CreE catalyzes nitrite formation from L-aspartate.

The connection between CreE's reactivity and diazo group formation became clear when it was hypothesized that nitrite might be a key biosynthetic intermediate. Previous stable isotope feeding experiments in a published study of azamerone biosynthesis had implicated the involvement of nitrite in diazo biosynthesis.[57] Therefore, nitrite was hypothesized to be the source of the distal nitrogen atom of the cremeomycin diazo group. To generate nitrite, CreD was proposed to catalyze the elimination of nitrite from **1.11**. Indeed, CreD and CreE generated nitrite and fumarate (**1.12**) from L-aspartate *in vitro*. Nitrite production required the presence of both CreD and CreE, suggesting that **1.11** is first generated by CreE before CreD catalyzes the elimination reaction. The ability of CreD and CreE to generate nitrite *in vitro* strongly suggests **1.11** is an intermediate in diazo group biosynthesis. However, CreD and CreE are not capable of generating cremeomycin from 3,2,4-AHMBA and nitrite *in vitro* suggesting that they do not catalyze diazotization *in vivo*. While the crystal structure of CreD has been reported[58], the structural basis for the tight coupling between CreD and CreE activities remains to be elucidated. This control may not be surprising given the nitrite's reactivity. Queries of publicly available sequence databases have revealed that genes resembling *creD* and *creE* are widespread in actinobacterial genomes and are often clustered in close proximity to natural product biosynthetic gene clusters.[55] These findings suggest that this previously unknown pathway for nitrite generation

may be important for the biosynthesis of a diverse set of N–N bond containing natural products, many of which have yet to be discovered.

The last challenge in deciphering cremeomycin biosynthesis was identifying the enzyme(s) that catalyze N–N bond formation between nitrite and **1.10** and subsequent diazo group formation. Nitrite was demonstrated to diazotize **1.10** to form cremeomycin under mildly acidic conditions (pH 4.0), but this reaction was abolished at neutral pH.[55] The buffered, pH neutral microbial fermentation medium used in these biosynthetic studies suggested that cremeomycin production may not derive from a non-enzymatic diazotization.

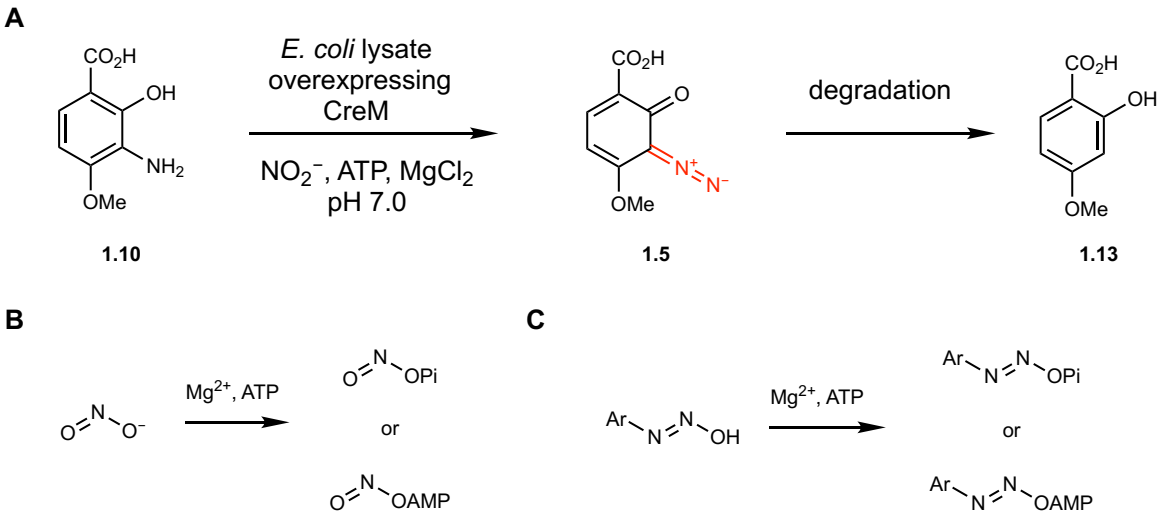


Figure 1.5: CreM's role in cremeomycin biosynthesis. **A**) CreM catalyzes diazotization of an arylamine *in vivo*. **B**, **C**) Possible diazotization mechanisms using ATP as a cofactor.

One candidate enzyme for diazotization is CreM, a predicted fatty acid-CoA ligase from the ANL (**A**cyl-CoA synthetases, adenylation domains of **N**RPS, **L**uciferase enzymes) superfamily.[59] This type of enzyme typically uses ATP to activate the carboxylate groups of fatty acid substrates in order catalyze thioesterification with coenzyme A. Based on this biochemical logic, CreM could use ATP to active nitrite to form an electrophilic species (**Figure 1.5B**). When

the enzyme was overexpressed in *E. coli* and the cell lysate was incubated with **1.10** at neutral pH, LC-MS analysis showed the formation of cremeomycin (**1.5**) (**Figure 1.5A**). An analogous reaction using lysates of *E. coli* overexpressing an empty vector did not result in diazotization. One potential mechanism for this reaction involves the use of ATP to first activate nitrite for nucleophilic attack by **1.10** (**Figure 1.5B**). The resulting *N*-hydroxydiazenyl intermediate could be activated by another equivalent of ATP and CreM again for elimination of the hydroxyl group (**Figure 1.5C**). To further confirm the role of ATP in this reaction, an enzyme variant with the glutamic acid residue predicted to coordinate Mg^{2+} for ATP binding was mutated to alanine. This variant cannot catalyze diazotization, supporting ATP's role in this reaction. Further *in vitro* studies, such as examining the substrate scope of CreM and structurally characterizing this enzyme, will inform our understanding of diazo installation. The discovery of CreM may prompt future work to engineer this enzyme to accept other aromatic amines for diazotization.

Kinamycin and Lomaiviticin

The kinamycins (A, C, D, and F **Figure 1.6**) are a family of anti-proliferative antibiotics first isolated in 1970 and produced by several bacterial species including *Streptomyces murayamaensis* ATCC 21414.[60-62] The bioactivity of these compounds is directly linked to the reactive diazo group, with studies demonstrating a range of potential mechanisms for its activation and cytotoxicity.[67-69] The kinamycins can cleave DNA *in vitro* under biomimetic conditions with mild reductants. Both reductive and nucleophilic activation of the diazo group have been proposed to give rise to various reactive intermediates, including a vinyl radical, *o*-quinone methide, acylfulvene, and a covalent adduct. In addition, the quinone moiety may undergo redox cycling to produce reactive oxygen species.

Kinamycin biosynthesis has been studied primarily through feeding studies and the isolation of intermediates from mutant strains. The identification of the full kinamycin biosynthetic gene cluster enabled a more complete understanding of genetic and biochemical basis for kinamycin biosynthesis and laid the groundwork for understanding diazo group installation. Despite the original misassignment of their structures, much of the information gathered from early kinamycin studies remains relevant.[70]

Stable isotope feeding experiments using labeled acetate have confirmed that polyketide synthases are likely involved in the biosyntheses of the kinamycins (**Figure 1.6**).[71-73] The previously known benz[a]anthraquinone polyketide dehydrorabelomycin (**1.14**) was subsequently isolated from *S. murayamensis* ATCC 21414 and was confirmed to be a biosynthetic intermediate by feeding a deuterated analog.[74] Over the next decade, additional metabolites, such as prekinamycin (**1.17**), kinobscurinone (**1.15**) and stealthin C (**1.16**), were isolated.[75-77] Deuterated **1.15** and **1.16** were fed and incorporated at low levels into kinamycin D, suggesting they could be intermediates in kinamycin production. Based on the low levels of incorporation of deuterium into kinamycin D, the authors proposed that diazo group installation could occur via a stepwise process involving N–N bond formation on the polyketide scaffold rather than through the incorporation of an intermediate containing an intact N–N bond. However, recent isotope feeding studies and *in vitro* characterization of biosynthetic genes from the kinamycin pathway have challenged this hypothesis and provided evidence of the latter proposal (See Section 1.2.2).

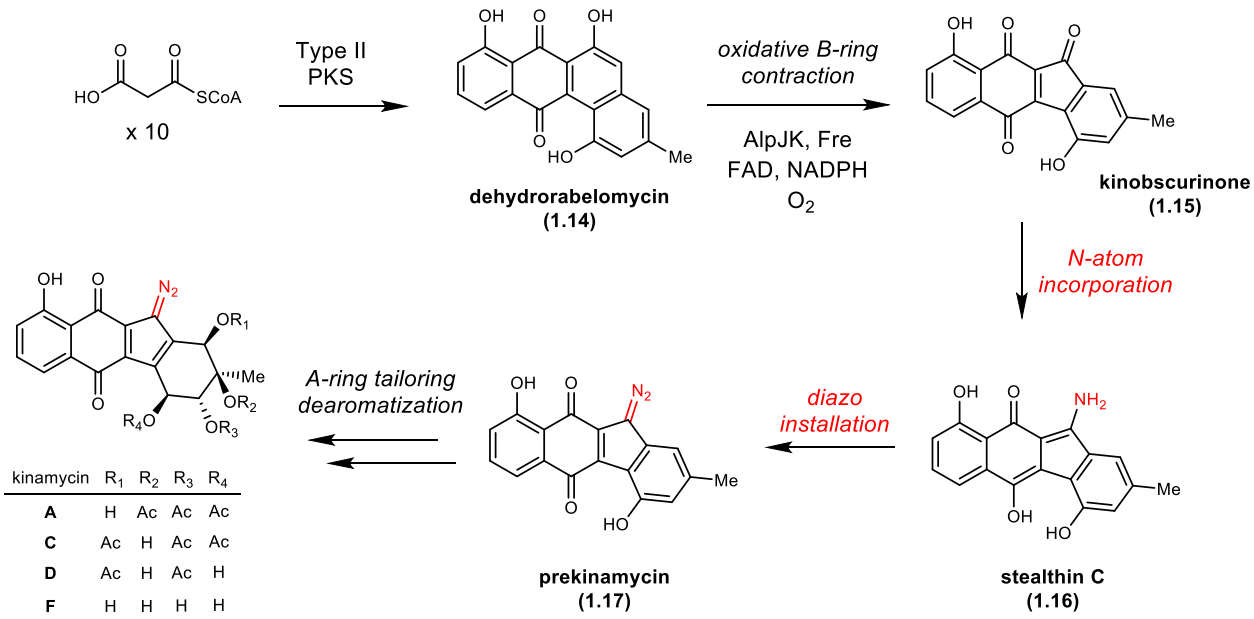


Figure 1.6: The biosynthesis of the benzo[b]fluorene scaffold of the kinamycins.

Early efforts to identify the kinamycin (*kin*) biosynthetic gene cluster led to the discovery of a cosmid from a genomic library of *S. murayamaensis* ATCC 21414 that produced **1.14**, **1.15**, and **1.16** upon expression in *S. lividans* ZX7.[78] The lack of kinamycin production was attributed to the incomplete capture of the biosynthetic gene cluster in a single cosmid. Recently, the full *kin* gene cluster (initially identified as the *alp* gene cluster) was identified in the genome of *Streptomyces ambofaciens* ATCC 23877, a strain previously not known to produce kinamycins. However, a mutant in which the transcriptional repressor (AlpW) was disrupted was able to produce several members of the kinamycin family, including kinamycin C and D**.**

The enzymatic chemistry involved in constructing the benzo[b]fluorene scaffold of the kinamycins has been elucidated (**Figure 1.6**).[79] A combination of genetics and *in vitro* and *in vivo* enzyme characterization identified a pair of flavin-dependent oxidases (AlpJ and AlpK) that appear to be responsible for B-ring cleavage of **1.14**, ring contraction, and hydroxylation to yield the hydroquinone form of the intermediate **1.15**. Surprisingly, if L-cysteine was present in the

enzyme reaction, production of stealthin C was observed (**1.16, Figure 1.7**). Stable isotope tracing experiments demonstrated that stealthin C is formed *via* a Smiles-type rearrangement of a cysteine adduct intermediate.

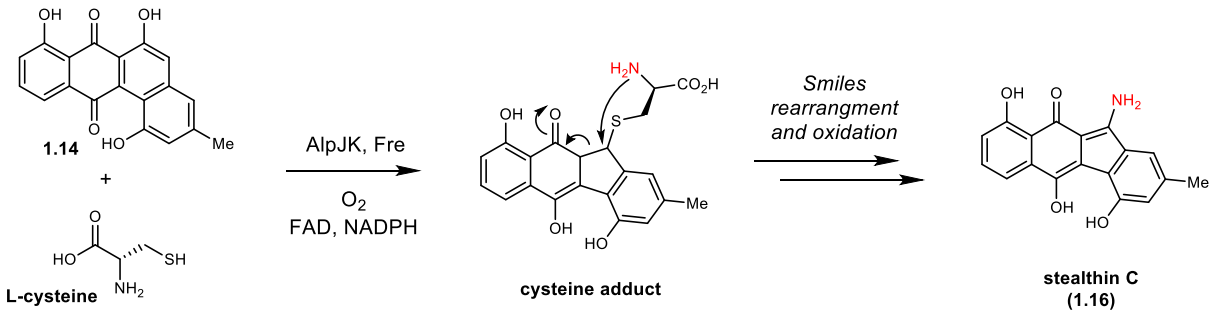


Figure 1.7: Stealthin C is formed *via* an S–N-typed Smiles rearrangement.

A variety of thiols and amine adducts were also observed when they were included in the AlpJK-catalyzed ring-contraction reaction. For example, adding *N*-acetylcytsteine to the assay mixture resulted in formation of seongomycin, a metabolite previously reported from heterologous expression of the *kin* cluster in *Streptomyces lividans* ZX7.[80] This observation suggested a mechanism in which a highly electrophilic benzofluorene intermediate is attacked nonenzymatically by a sufficiently nucleophilic amine or thiol. This result further questioned the intermediacy of stealthin C in diazo biosynthesis.[81]

Identification of the putative diazo-forming enzymes encoded in the *kin* gene cluster relied upon having access to biosynthetic gene clusters that produce the lomaiviticins, a related family of diazobenzofluorenes. A discussion of these enzymes will be included below. Additionally, this section will compare the *kin*/*lom*-type putative diazo biosynthetic machinery with that from cremeomycin assembly.

The lomaiviticins are C_2-symmetric diazobenzofluorene glycosides isolated in 2001 from the marine bacterium *Micromonospora lomaivitiensis* (later classified as *Salinospora*) using a DNA-damage assay to guide isolation.[82] Extensive studies into the mechanism of lomiaiviticin A (**1.18**, **Figure 1.8**) have provided both biochemical and structural evidence that this metabolite intercalates DNA and induces double strand breaks. The two diazobenzofluroene moieties disrupts base pairing by inserting into the DNA duplex and positions the reactive diazo functional groups in close proximity to each DNA strand.[83] These groups are then activated nucleophilically or reductively to generate vinyl radicals that abstract hydrogen atoms from the DNA backbone.

Lomaiviticin biosynthesis was hypothesized to follow closely with the biosynthetic pathway leading to kinamycin production.[84-85] Highly similar lomaiviticin (*lom*) biosynthetic gene clusters containing roughly 60 genes were identified in *Salinispora tropica* CNB-440, *Salinispora pacifica* DPJ-0019 and *Salinispora pacifica* DPJ-0016 using a combination of bioactivity-guided fractionation, genomic library screening, and genome mining. Comparison with the *kin* gene clusters from *S. murayamaensis* and *S. ambofaciens* supported the use of an analogous route for diazobenzofluorene construction and identified putative enzymes involved in dimerization and propionate starter unit generation. The biochemistry of starter unit generation and A-ring oxidation has been investigated *in vitro*.[86-87]

This comparative analysis also revealed strong candidates for the diazo-forming enzymes. Comparison of the complete *lom* and *kin* gene clusters revealed a conserved, co-localized set of genes encoding homologs of several C–N bond forming and bond cleaving enzymes, as well as enzymes that are predicted to perform redox chemistry. These genes are absent from the partial *kin* gene cluster that was identified in *S. murayamaensis*.[78] This observation is consistent with the inability of this partial gene cluster to generate diazo-containing compounds when heterologously

expressed in *S. lividans*. The general biochemical functions encoded within the predicted *kin*/*lom*-type diazo gene cassettes are consistent with the types of transformations needed to construct a diazo group from two separate nitrogen-containing metabolites (**Figure 1.8**). This includes predicted C–N bond forming enzymes glutamine synthetase (Lom32/KinL) and an *N*-acetyltransferase (Lom35/KinN) and homologs of C–N bond cleaving enzymes [an amidase (Lom33/KinK) and an adenylosuccinate lyase (Lom34/KinM)]. Additionally, a hypothetical protein (Lom29/KinJ) encoded adjacent to a predicted ferredoxin (Lom30/KinI) may be involved in redox chemistry.

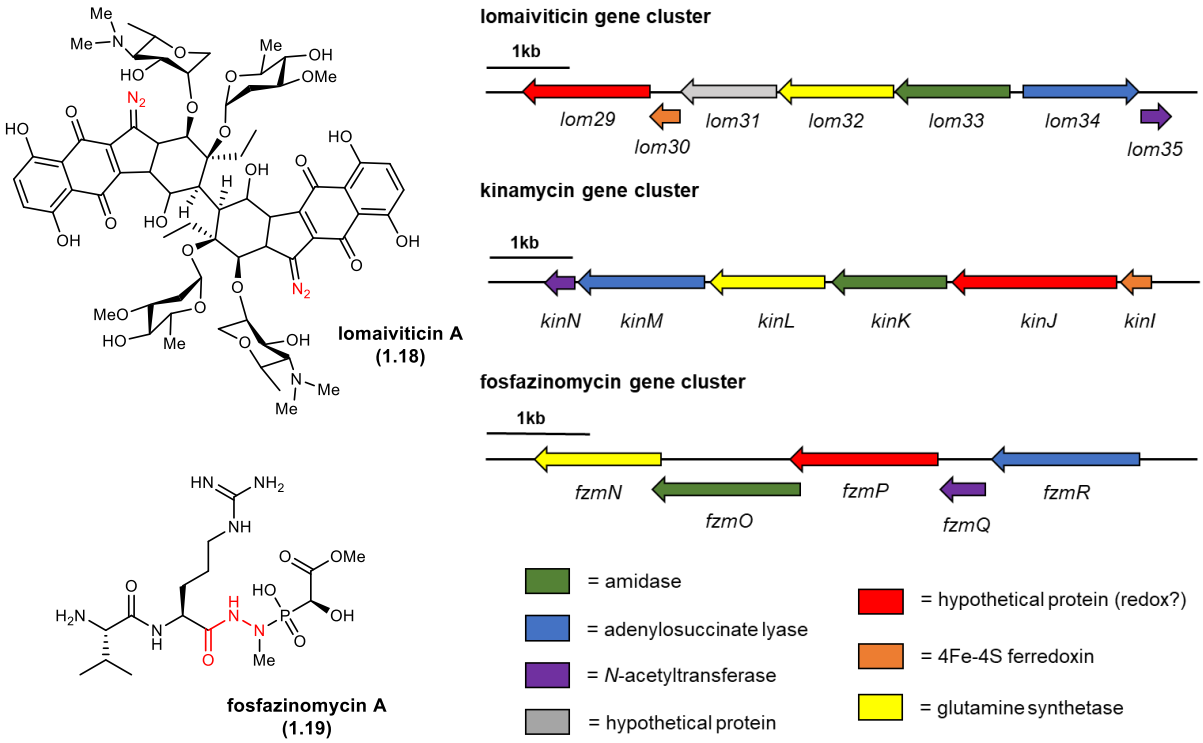


Figure 1.8: **A**) Structures of lomaiviticin A and fosfazinomycin A. **B**) Comparison of the diazo cassette found in kinamycin and lomaiviticin gene clusters and the hydrazide-forming genes from fosfazinomycin biosynthesis.

All but one (Lom30/KinI) of the *kin*/*lom*-type diazo-forming enzymes are encoded within the gene cluster that produces the hydrazide-containing natural product fosfazinomycin A (**1.19**),

further supporting their role in N–N bond formation (**Figure 1.8**).[88] The relationship between the *lom/kin*-type diazo-forming gene cassette and the *fzm*-type hydrazide-forming gene cassette will be discussed below in the section covering fosfazinomycin biosynthesis (Section 1.2.2). In short, *in vivo* feeding experiments with ^{15}N-nitrite and *in vitro* biochemical assays of the 'diazo' gene cassette in the kinamycin producer uncovered that L-glutamylhydrazide is a common intermediate in both fosfazinomycin and kinamycin biosynthesis (discussed in Section 1.2.2, **Figure 1.11**) These results provided insight into the sequence of biosynthetic events and the origin of the diazo group. However, *in vitro* reconstitution of diazo group formation remains elusive.

Queries of publicly available genome sequences revealed similar *lom/kin*-type gene cassettes in 57 gene clusters encoding putative secondary metabolites.[85] The wide distribution of this putative diazo biosynthetic machinery, as well as that from the cremeomycin pathway suggests that diazo-containing natural products may be far more numerous than previously expected. Targeted genome mining approaches may now be used to identify new metabolites that contain diazo groups, expanding this family of exceptionally bioactive natural products.

1.2.2 Hydrazine, hydrazide, hydrazone

Hydrazides are well represented among N–N bond containing natural products and can be found in both linear and cyclic scaffolds.[46-47] This family of natural products has been discovered in fungi, both terrestrial and marine bacteria, and plants since the 1960s. These natural products exhibit antidepressant, anti-inflammatory, antiproliferative, antibacterial, and antifungal bioactivities.[46] In synthetic organic chemistry, hydrazides serve as important reagents and intermediates, as well as final target scaffolds.[89] Hydrazides are most commonly accessed by acylation of hydrazine and hydrazine derivatives. Hydrazide natural products may be

biosynthesized using related logic in which an activated carboxylic acid derivative undergoes nucleophilic attack by hydrazine or a hydrazine derivative. Currently, our understanding of the genes and enzymes involved in hydrazide biosynthesis come from studying the fosfazinomycin, piperazic acid-containing nonribosomal peptides, and a recently uncovered natural product s56-p1.

Fosfazinomycin and kinamycin

Fosfazinomycin A (**1.19**) were isolated in 1983 from the culture filtrate of *Streptomyces lavendofoliae* No. 630 during a screening program for antifungal antibiotics (**Figure 1.9**).[90-92] This compound was also isolated from *Streptomyces* sp. WM6372, and a fosfazinomycin (*fzm*) biosynthetic gene cluster was identified using genome mining and a novel enzymatic isotope labeling approach called SILPE (stable isotope labeling of phosphonates in extracts).[88] Surprisingly, the *fzm* gene cluster contains putative diazo-forming gene cassettes from both the cremeomycin and kinamycin/lomaiviticin pathway. *In vitro* characterization of enzymes from both of these gene cassettes has led to proposals for their roles in hydrazide biosynthesis (**Figure 1.8**).

In vitro characterization of the *fzm* biosynthetic enzymes has clarified the order of events in this pathway, confirmed the proposed activities of several enzymes, and elucidated key biosynthetic intermediates in hydrazide installation (**Figure 1.9**).[93-94] Based on this work, it was hypothesized that fosfazinomycin is assembled via a convergent process involving the incorporation of phosphonate and amino acid building blocks. In the phosphonate side of the pathway, the enzymes FzmBCDG, which consist of an *O*-methyltranferase, phosphoenolpyruvate (PEP) phosphomutase, phosphonopyruvate (PnPy) decarboxylase, and α-ketoglutarate dioxygenase, respectively, convert PEP (**1.20**) to the intermediate (*S*)-2-hydroxy-2-phosphono-

acetate (**1.21**). This phosphonate intermediate **1.21** is then coupled with methylated arginine hydrazide (**1.22**) to form **1.23**, which can be further elaborated to **1.19** by ligation of L-valine. While the enzymatic chemistry involved in constructing **1.21** is well-precedented, the biosynthetic route to the hydrazide coupling partner **1.22** has not been completely elucidated.

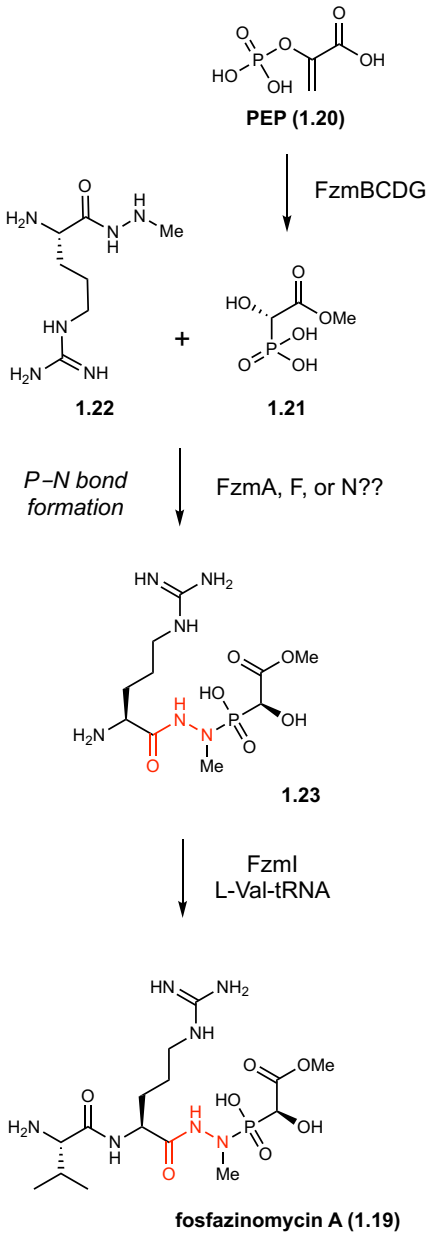


Figure 1.9: Biosynthesis of fosfazinomycin A is a convergent process.

An early biosynthetic hypothesis for hydrazide construction was proposed based on the contents of the *fzm* gene cluster and the *in vitro* activities of the *fzm* biosynthetic enzymes.[94] Arginine hydrazide (**1.23, Figure 1.10**) was proposed to be a precursor to the late-stage intermediate **1.22** as it is selectively methylated by FzmH to generate **1.22**. The synthesis of **1.23** is thought to involve nitrite production as the *fzm* gene cluster encodes homologs of CreD and CreE (FzmL and FzmM, respectively) from cremeomycin biosynthesis (Section 1.2.1). Consistent with this proposal, these enzymes FzmM and FzmL convert L-aspartate to nitrite and fumarate *in vitro*. Based on the hypothesis that L-hydrazinosuccinate (**1.25**) is a biosynthetic intermediate, nitrite could diazotize a second equivalent of L-aspartate, forging the N–N bond that appears in the final hydrazide natural product. This diazo acid intermediate (**1.24**) may then be reduced via an unknown process to generate **1.25**.

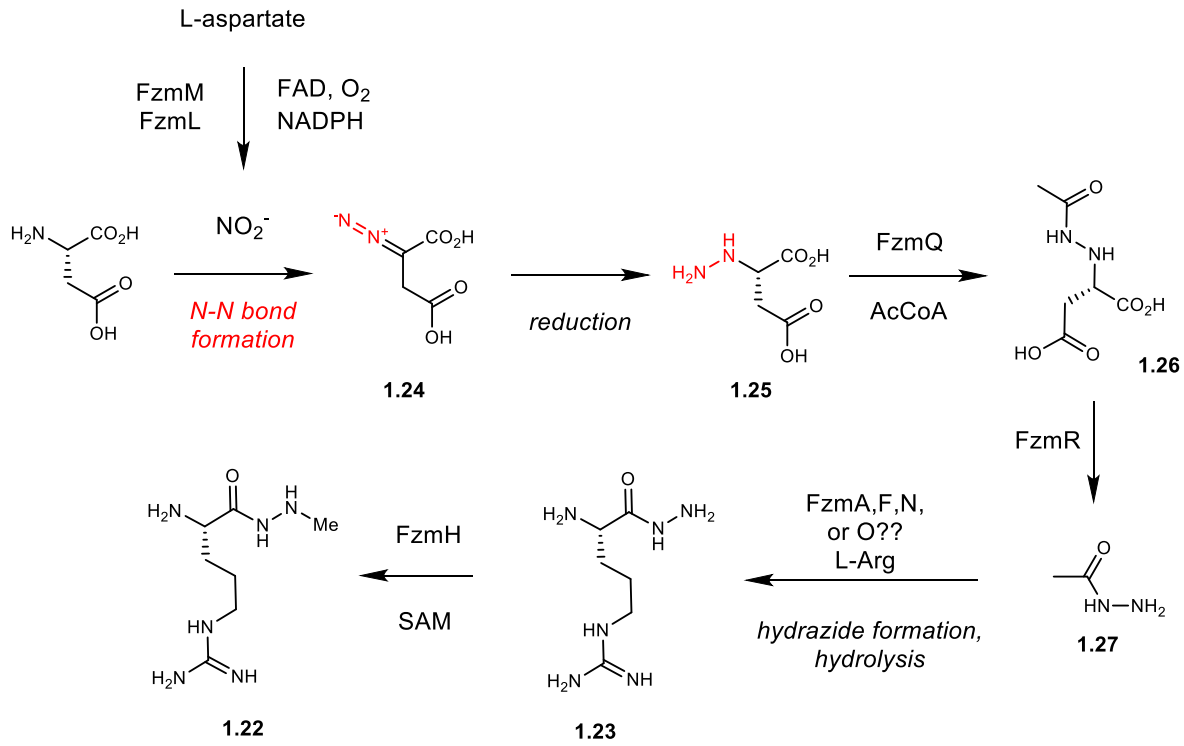


Figure 1.10: L-hydrazinosuccinate is a biosynthetic intermediate in the biosynthesis of L-argininylhydrazide, a precursor to fosfazinomycin A.

As noted earlier, in addition to homologs of CreD and CreE the *fzm* gene cluster contains most of the putative diazo-forming gene cassette from the kinamycin and lomaiviticin pathways, including the glutamine synthetase (*fzmN/kinL*), the amidase (*fzmO/kinK*), the hypothetical protein (*fzmP/kinJ*), the *N*-acetyltransferase (*fzmQ/kinN*), and the adenylosuccinate lyase (*fzmR/kinM*). Recent feeding studies in *Streptomyces sp.* XY332 and *Streptomyces murayamaensis* ATCC 21414 and *in vitro* characterization of the hydrazide/diazo-gene cassette linkage revealed a convergent biosynthetic pathway in kinamycin and fosfazinomycin.[95]

In vitro characterization of two biosynthetic enzymes, FzmQ/KinN and FzmR/KinM, has confirmed that these enzymes participate in converting **1.25** to *N*-acetylhydrazine (**1.27, Figure 1.11**). The *N*-acetyltransferase FzmQ/KinN was shown to selectively acylate the terminal nitrogen atom of L-hydrazinosuccinate with acetyl-CoA, providing **1.26**. The adenylosuccinate lyase homolog FzmR/KinM catalyzed the elimination of **1.27** from **1.26** forming fumarate as a co-product. Feeding studies with *N*-acetylhydrazine and L-hydrazinosuccinate confirm that both of the nitrogen atoms in the hydrazine group are incorporated into the hydrazide and diazo linkages respectively. For example, feeding $^{15}N_2$-acetylhydrazine to *Streptomyces murayamaensis* ATCC 21414 resulted in ~50% enrichment of both diazo nitrogen atoms.

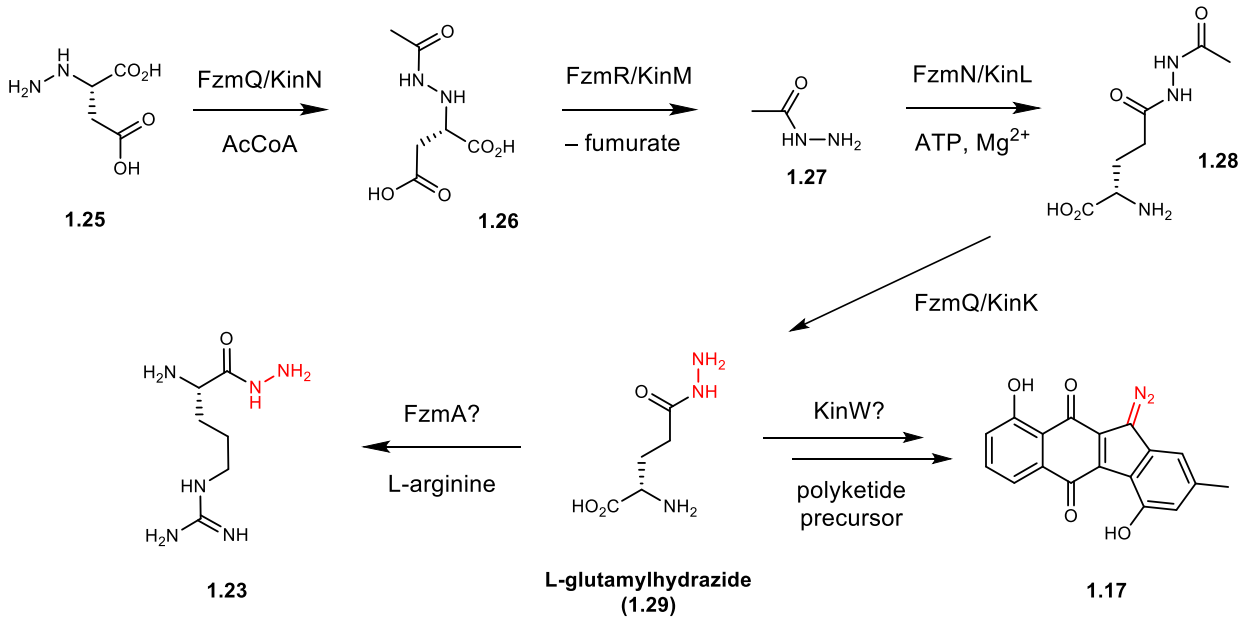


Figure 1.11: L-glutamylhydrazide is a common biosynthetic intermediate in the biosynthesis of kinamycin and fosfazinomycin.

After *N*-acetylhydrazine is formed, this intermediate is proposed to be ligated to L-glutamic acid by FzmN/KinL using ATP and subsequently deacetylated to form L-glutamytlhydrazide (**1.29**) by the amidase FzmQ/KinK. These steps were reconstituted *in vitro* by purified FzmN and FzmQ. Analogously, *E. coli* overexpressing KinL was able to form *N*-acetylglutamylhydrazide in the presence of added *N*-acetylhydrazine, and feeding studies also demonstrated both hydrazide nitrogen atoms of glutamylhydrazide are incorporated into the diazo group. Furthermore, feeding ^{15}N-nitrite to *S. murayamaensis* resulted exclusively in labeling of proximal diazo nitrogen (~90%), consistent with the biosynthetic hypothesis outlined. The lack of scrambling also suggested delivery of hydrazine from glutamyl hydrazide to the polyketide scaffold is enzymatic. L-glutamylhydrazide is proposed to the last common intermediate in the biosynthetic pathways of kinamycin and fosfazinomycin.

The transfer of the hydrazine unit to arginine to form L-argininylhydrazine **1.29** in fosfazinomycin biosynthesis, or the polyketide scaffold in kinamycin biosynthesis, has not been reconstituted *in vitro*. In fosfazinomycin, this enzyme is proposed to be the asparagine synthase FzmA. On the other hand, KinW, a predicted glutamine synthetase, is proposed to cleave and transfer the hydrazine unit from L-glutamylhydrazide to the polyketide scaffold. The generation of **1.28** rather than direct use of hydrazine may have evolved to avoid accumulation of this toxic and reactive species. Furthermore, the N–N bond formation between nitrite and L-aspartic acid to form L-hydrazinoaspartate has not yet been reported. Nonetheless, this study demonstrated that the N–N bond is made in an independent pathway that forms a branch of a convergent route to structurally complex natural products.

s56-p1

The biosynthesis of s56-p1 (**1.30**, **Figure 1.12A**), a hydrazone-containing natural product, has been recently reported.[96] This natural product was isolated from fermentation cultures of *Streptomyces sp*. SoC090715LN-17. A putative biosynthetic gene cluster was identified using a genome mining strategy that targeted homologs of amino group carrier proteins that produced the nonproteinogenic amino acid core scaffold (2*S,*6*R*)-diamino-(5*R*,7)-dihydroxy-heptanoic acid (DADH). Subsequent heterologous expression of a partial gene cluster in *Streptomyces lividans* TK23 produced a hydrazine intermediate **1.32** (**Figure 1.12B**). Using gene knockouts, the authors have identified two genes, *spb38* and *spb40*, that are essential for the production of **1.32**. Overexpressing these two genes in *E. coli* also resulted in generation of **1.32**.[96] Spb38 is annotated as a flavin-dependent enzyme. Subsequent *in vitro* assays and labeling studies demonstrated that

Spb38 is a flavin-dependent *N*-monooxygenase that oxygenates L-lysine to form *N*-hydroxy-L-lysine (**1.31**) using FAD, NADH, and O_2 as cofactors.

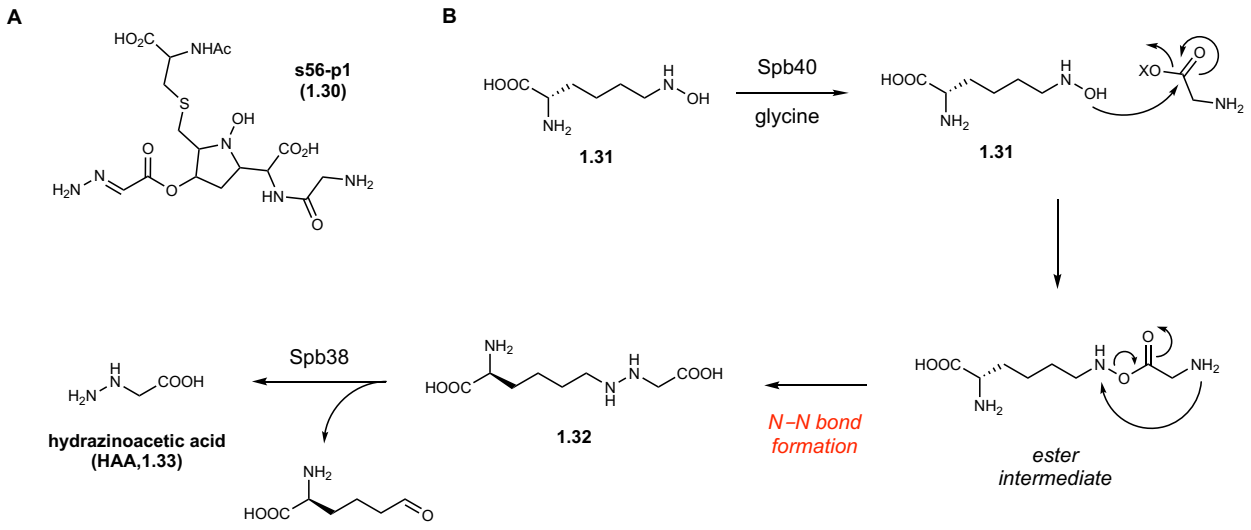


Figure 1.12: **A**) The structure of sp56-p1. **B**) The N–N bond forming reaction catalyzed by Spb40 ad Spb38.

Spb40 is an enzyme of unknown function that is predicted to contain two-domains, a central methionyl-tRNA ligase and a C-terminal cupin domain. The authors proposed that Spb40 uses ATP to ligate the hydroxyl group of *N*-hydroxyl-L-lysine to the carboxylate of glycine to form an ester intermediate. This intermediate then rearranges intramolecularly to afford **1.32**. This proposal is supported with ^{18}O-labeling experiments in which the labeled oxygen atom of the hydroxylamine is transferred to the carboxylate oxygen after N–N bond formation. However, *in vitro* reconstitution of N–N bond formation has not been successful, as some unidentified cofactors might be required. Mutational analysis also preliminarily demonstrated the metal binding residues in the cupin domain are required for activity, but the identity of the metal cofactor and its role in N–N bond formation remain unknown. The authors also tested the activity of Spb39, a predicted D-amino oxidase. Incubating purified Spb39 with **1.32** *in vitro* resulted in formation of hydrazinoacetic acid **1.33** (**Figure 1.12B**), which could be ligated to the core DADH scaffold and

oxidized to the hydrazone group. Since the report of Spb38–40, homologs of these enzymes have been found in several natural product biosynthetic pathways, such as triacsin C[97]and pyrazomycin (discussed below), highlighting a potentially diverse application of this hydrazine building block.

1.2.3 Pyrazole and piperazate

Pyrazole (**1.34**) and β-pyrazol-1-ylalanine (**1.35, Figure 1.13**) are structurally and biosynthetically related metabolites isolated from seed extracts of the watermelon *Citrullus vulgaris*.[98] The latter molecule has also been detected in several cucumber varieties including *Cucumis sativus*.[99] The identification of **1.35** in 1959 marked the first discovery of a pyrazole-containing natural product.[100] The biosynthesis of these metabolites has been studied using stable isotope feeding experiments and *in vitro* assays with partially purified protein extracts. These studies support a biosynthetic hypothesis for **1.35** in which 1,3-diaminopropane (**1.36**) undergoes a six-electron oxidation to **1.34** via the intermediate 2-pyrazoline (**1.37**), followed by condensation with *O*-acetylserine.[101-102] A partially purified protein extract from *Cucumis sativus* seedlings was capable of converting **1.36** to **1.34**.[99] This activity was stimulated by the addition of FAD, but not NAD^+ or $NADP^+$. Dialysis of the protein extract against EDTA did not inhibit its activity nor did the addition of various metals ions. Taken together, these results suggest that one or more flavin-dependent enzymes are involved in the six-electron oxidative cyclization and dehydrogenation that converts **1.36** to **1.34**. The lack of a stimulatory effect by addition of NAD^+/$NADP^+$ and lack of an inhibitory effect of EDTA suggest that this oxidative process is not catalyzed by nicotinamide- or metal-dependent oxidases. A related assay with crude protein extracts revealed the final coupling reaction that affords **1.35**.[103-104] Although direct evidence is lacking, it is interesting to consider whether the biosynthetic logic for N–N bond formation in

pyrazole biosynthesis may be similar to that of other pathways which use flavin-dependent enzymes, including the azoxy compound valanimycin and the hydrazine piperazic acid.

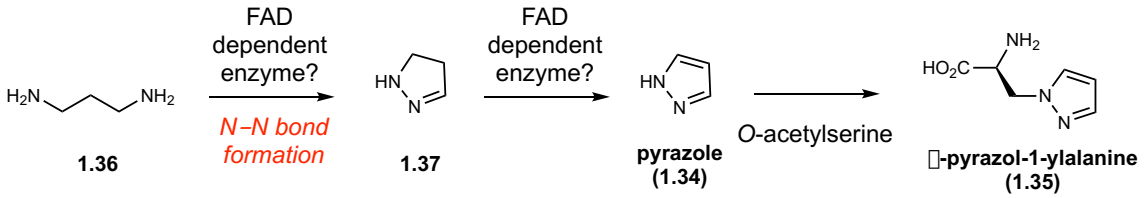


Figure 1.13: The proposed biosynthetic pathway of β-pyrazol-1-ylalanine involves a pyrazole intermediate.

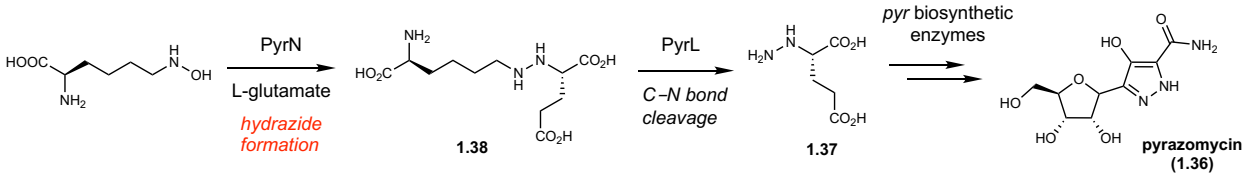


Figure 1.14: The biosynthetic pathway of the C-nucleoside pyrazomycin uses a hydrazine intermediate.

Recent insights into pyrazole biosynthesis came from studying the biosynthesis the *C*-nucleoside pyrazomycin (**1.36**, **Figure 1.14**).[105] Pyrazomycin was isolated from *Streptomyces candidus* NRRL 3601 and displays weak antibiotic activity. The biosynthetic gene cluster was identified by searching for homologs of genes that produce the nucleoside family of natural products, such as formycin.[106] The putative gene cluster also contained homologs of the aforementioned hydrazine-forming enzymes Spb38 (PyrM) and Spb40 (PyrN) (Section 1.2.2). *In vivo* biochemical characterization of PyrMN demonstrated that the homologs analogously catalyzes N–N bond formation when overexpressed in *E. coli*, but *in lieu* of glycine, L-glutamic acid is utilized. Feeding ^{15}N-labeled L-glutamic acid resulted in labeling of one of the pyrazole nitrogen atoms, further supporting this amino acid as the coupling partner to form the hydrazine intermediate **1.37**. Oxygen isotope tracing experiments resulted in similar labeling pattern to that

of Spb40-catalyzed N–N bond formation. Specifically, the molecular oxygen-derived oxygen atom of N6-hydroxy-L-lysine is not retained in the hydrazine intermediate **1.38**. Similar to Spb40, the N–N bond forming reaction has not been reconstituted *in vitro*, suggesting additional cofactor(s) may be required. For example, the authors proposed that the activation of L-glutamic acid is dependent on tRNA. Further *in vitro* activity of N–N bond and the transfer of hydrazine unit is to a pyrazole group would advance our understanding on the biosynthesis for this natural product family.

Piperazic acid (**1.39**, **Figure 1.15**) is a nonproteinogenic amino acid incorporated into numerous nonribosomal peptides, such as himastatin and the kutznerides (**1.40**).[107-108] Biochemical evidence demonstrated *N*-oxygenation in piperazic acid formation in the kutzneride pathway with L-ornithine as the starting precursor.[109] The kutznerides are a family of piperazic acid-containing antimicrobial and antifungal cyclodepsipeptides produced by *Kutzneria* sp. 744. Discovery and annotation of the kutzneride (*ktz*) biosynthetic gene cluster revealed a putative flavin-dependent lysine/ornithine *N*-monooxygenase (KtzI).[110] KtzI was characterized *in vitro* and was demonstrated to hydroxylate the N5 nitrogen of L-ornithine (**1.41**). The enzyme was specific towards its substrate, as D-ornithine, D/L-lysine, D/L-glutamat, and D/L-glutamine were not accepted as substrates.

To demonstrate *N*5-hydroxy-L-ornithine (**1.41**) was important in piperazic acid biosynthesis, feeding studies with labeled amino acids and analysis of extracts by LC-MS/MS were performed.[109] While $^{13}C_5$-Glu and $^{13}C_5$-Gln were incorporated into the kutznerides, mass shifts ranging from +1 Da to +5 Da suggested that extensive scrambling of the label had occurred, likely due to metabolism in primary metabolic pathways. Alternatively, feeding studies with $^{13}C_5$-L-ornithine and $^{13}C_5$-N5-hydroxy-L-orinithine resulted in incorporation levels as high

as 35%, with the labeling largely isolated to the piperazic acid residues. These results indicate that L-ornithine and N5-hydroxy-L-orinithine are more proximal precursors to piperazic acid than glutamate or glutamine. Label suppression feedings studies with $^{13}C_5$-L-ornithine and $^{13}C_5$-N5-hydroxy-L-orinithine further demonstrated that $^{13}C_5$-*N*5-hydroxy-L-orinithine is downstream of $^{13}C_5$-L-ornithine in piperazic acid biosynthesis. Overall, these results support that L-glutamate and L-glutamine are first converted into L-ornithine before incorporation into piperazic acid.

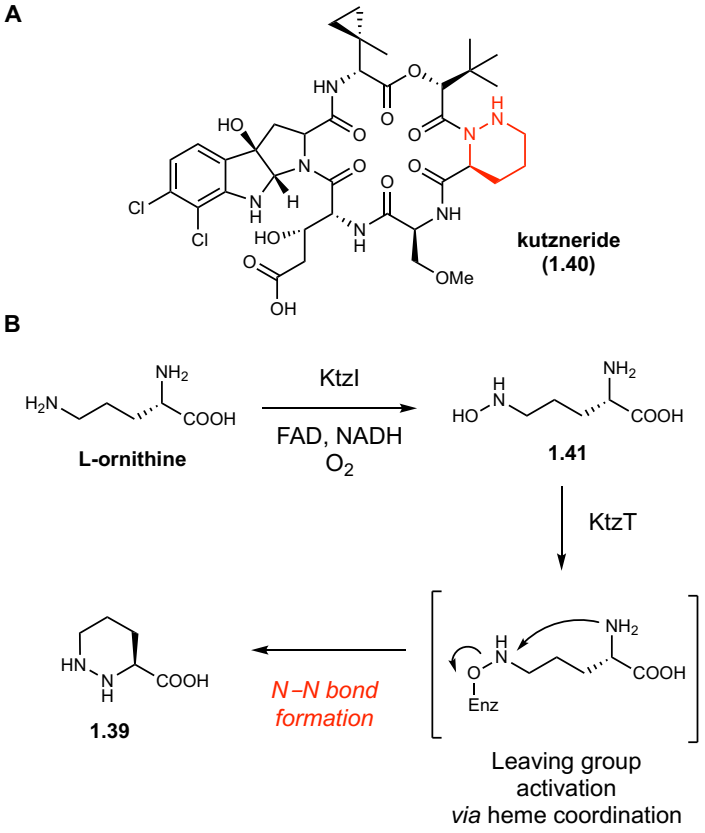


Figure 1.15: **A**) The structure of kutzneride with the piperazic acid highlighted in red. **B**) KtzT is a heme-containing metalloenzyme that catalyzes N–N bond formation.

Building on these preliminary results, N–N bond formation in piperazic acid biosynthesis has been reconstituted *in vitro* recently (**Figure 1.15B**).[111] Specifically, a heme-dependent enzyme KtzT utilizes *N*-hydroxy-L-ornithine (**1.41**) to form the heterocycle. KtzT was initially annotated

as having an FMN-binding domain; however, subsequent analysis of the heated proteins did not reveal bound flavin cofactors. Instead, spectroscopic features suggested that KtzT may bind heme. Indeed, a pyridine hemochrome assay revealed a heme-*b* cofactor. The piperazic acid formation has been shown to be redox-independent, since the reaction proceeds with high turnover using both as-purified and dithionite-reduced KtzT. Furthermore, no oxygen was consumed during the reaction, and the reaction can proceed anaerobically. D- or L-N6-hydroxyornithine is also accepted as a substrate. Finally, the authors also identified several active site residues that might be important in heme-binding and catalysis. These activities suggested that the heme iron may act as a Lewis acid that activates the hydroxylamine for an intramolecular S_N2-like displacement. This discovery may lead to developing or finding other KtzT homologs for biocatalytic applications. Further structural characterization of KtzT may shed light into the mechanism for N–N bond formation.

1.2.4. Azoxy

The first natural product discovered to possess an N–N bond was the azoxy-containing compound macrozamin in 1951.[112] Azoxy compounds have attracted attention given their varied bioactivities, which include antibacterial, antifungal, anticancer and carcinogenic effects.[46-47] While azoxy-containing metabolites are rare among known natural products, these compounds have well-established roles as intermediates and reagents in synthetic organic chemistry.[113] This functional group can be furnished by the oxidation of azo compounds, a reaction route that involves constructing the N–N linkage first. Alternatively, both aliphatic and aromatic azoxides can be accessed through coupling of *C*-nitroso compounds and hydroxylamines, with N–O bond of the azoxy linkage installed prior to N–N bond formation.

The azoxymycins are a class of dimeric aromatic azoxy compounds that were isolated from *Streptomyces chattanoogensis* L10 in 2015 (**Figure 1.16**).[114] Azoxymycin A, B, and C (**1.42-1.44**) are bright yellow solids and analysis of these compounds using HPLC-diode array detection demonstrated that they exist as interconverting cis-trans isomers at room temperature. The azoxymycin (*azo*) biosynthetic gene cluster was discovered using a genome mining approach based upon the hypothesis that these compounds were polyketide-derived.[115] Involvement of this gene cluster in azoxymycin production was confirmed by gene knockout experiments. Importantly, when *azoC* was disrupted two amino aromatic polyenes (**1.45 and 1.46, Figure 1.17A**) accumulated in the mutant.[116] The protein encoded by *azoC* is a putative *para*-aminobenzoic acid (PABA) *N*-oxygenase resembling the extensively studied nitro group-forming non-heme diiron *N*-oxygenase AurF (See Chapter 1.4).[117] This annotation suggests that AzoC may be involved in oxidizing the amino group of **1.45** and **1.46** en route to azoxy group formation.

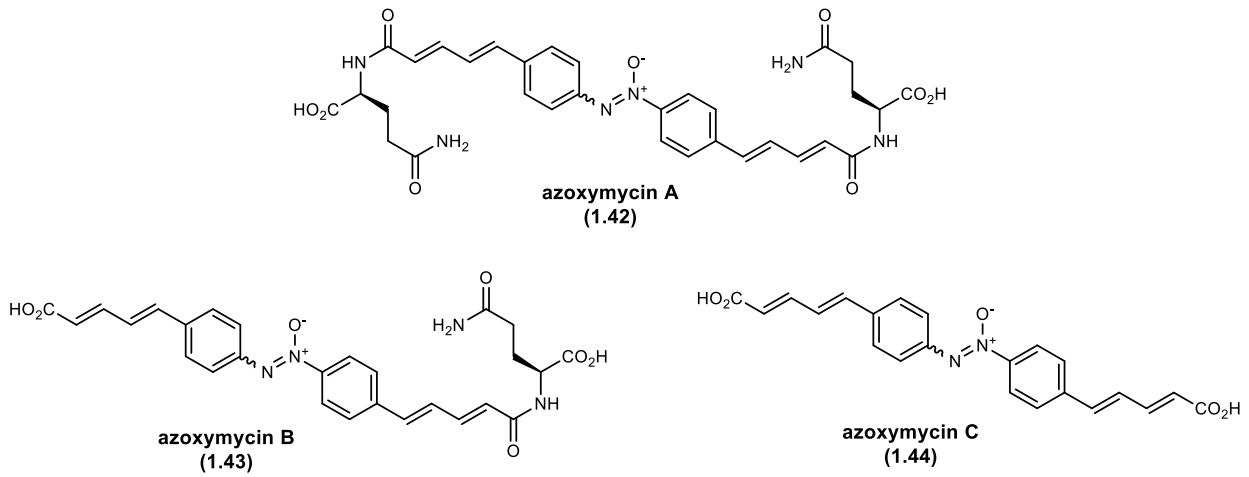


Figure 1.16: The structures of the azoxymycins**.**

Confirmation of AzoC-catalyzed *N*-oxygenation *in vitro* was reported recently.[118] Specifically, purified AzoC oxidizes both putative biosynthetic benzylamine precursors **1.45** and **1.46** to form the azoxy-containing products **1.42**, **1.43**, **1.44** in the presence of NADH, Fe^{II} and

phenazine methosulfate. Under single turnover conditions with a reduced Fe^{II}-AzoC, the authors observed conversion of the primary amine to the hydroxylamine product (**Figure 1.17B**). Using a synthetic hydroxylamine as the substrate, this intermediate can be further oxygenated to form the nitroso congener. However, AzoC cannot oxidize the nitroso group to the corresponding nitro group. Furthermore, no azoxy dimer was formed under single turnover conditions. Instead, a nonenzymatic radical coupling reaction between the hydroxylamine and nitroso intermediates promoted by cofactors NAD^+ and NADH induces dimerization (**Figure 1.17C**). This coupling reaction is inhibited by radical scavengers TEMPO and BHT. This discovery provided evidence of an enzymatic *N*-oxygenation reaction catalyzed by AzoC, and a nonenzymatic N–N bond radical coupling step that forms azoxymycin. These reactivities might provide insight into biosynthesis of other azoxy-containing natural products.[118-119]

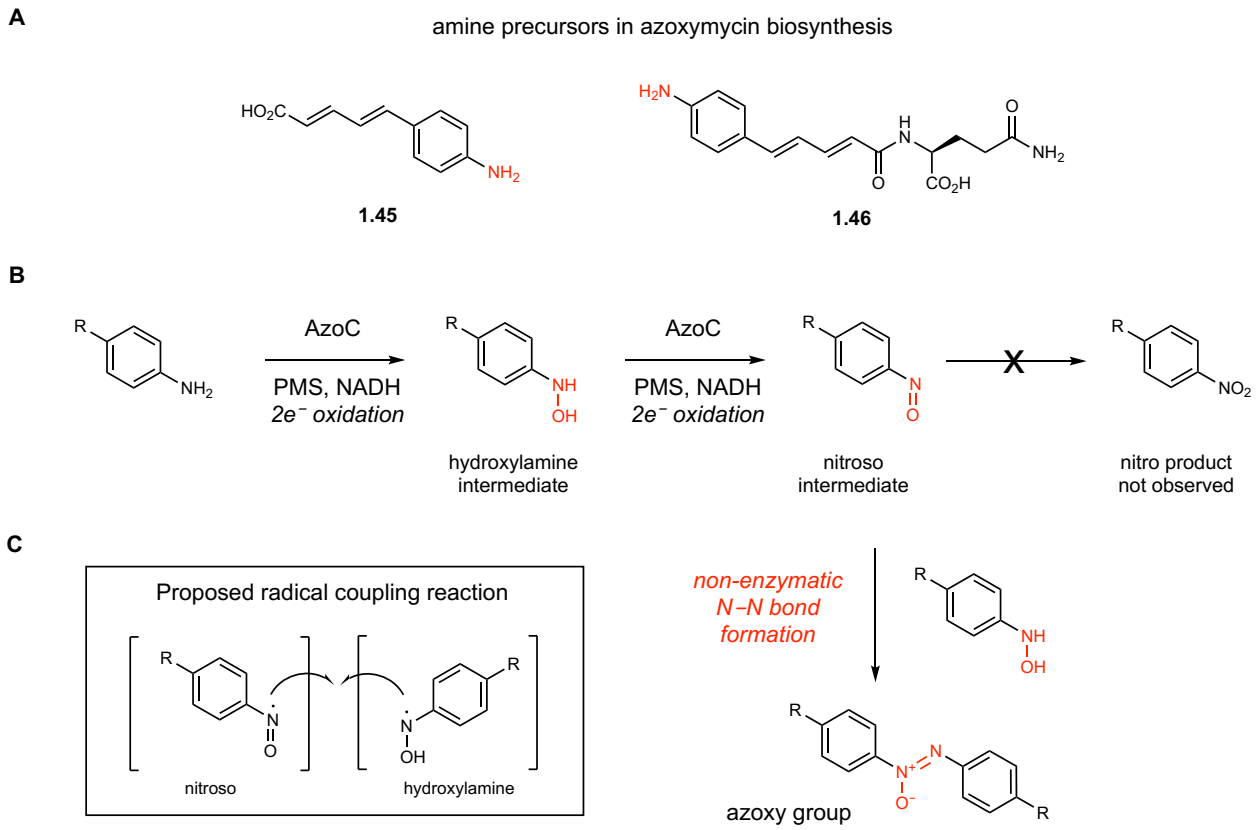


Figure 1.17: AzoC catalyzes azoxy formation *in vitro*. **A**) Structures of the amine substrates of AzoC. **B**) AzoC catalyzes two sequential oxygenations of the amines to form the nitro congeners. **C**) Azoxy formation involves radical coupling between the nitroso and hydroxylamine radical intermediates. PMS = phenazine methosulfate.

1.3 Biosynthesis of *N*-nitrosamines in primary and secondary metabolism

N-nitrosation of primary metabolites has been extensively studied due to this functional group's connection to gastric cancer and inflammation.[120-121] *N*-nitrosation of secondary amines confers deleterious bioactivities to otherwise inactive compounds, as *N*-nitrosamines and alkylnitrosoureas are precursors to diazonium alkylating agents (**Figure 1.18A**). These compounds alkylate the nucleobases of DNA, which can disrupt base pairing and cause depurination. The result of these modifications can lead to mutations, DNA strand breaks, and cell death.[121]

Nitrite is central to nitrosamine formation *in vivo,* as it can be activated under acidic conditions to nitrous acid ($pK_a = 3.2$) or reduced by metalloenzymes like nitrite reductases to afford

reactive nitrogen species (**Figure 1.18A**). Recently, *N*-nitrosation of secondary amines by nitrite-derived nitrous anhydride (N_2O_3) was also shown to be a physiologically relevant reaction.[122] At the beginning of this thesis, this process is not known to be catalyzed by dedicated enzymes. Most of our understanding of nitrite's role in *N*-nitrosation has been driven by the food industry and health-monitoring agencies, as our exposure to nitrite is mainly through diet. Nitrite is often used as a food additive in the cured meat manufacturing industry dating as early as the 1900s. Several seminal studies reported that nitrite, which can react with itself in acidic conditions to form nitric oxide (NO) and nitrogen dioxide (NO_2), can preserve the red color of meat.[123] This is due to the generated NO that can bind to hemoglobin. Since then, the meat industry has widely adopted nitrite to preserve meat under federal regulatory control in the U.S. and E.U. However, in recent years, epidemiology studies have found a link between high consumption of processed meat to incidences of increased gastric cancer risk.[124] This has prompted the World Health Organization (WHO) to classified cure meats as Group 1 carcinogens, a category in which the organization finds sufficient evidence that consumption cure meat is linked to causing cancer.

While a lot of studies have focused on the nonenzymatic formation of *N*-nitrosamines in the human body, *N*-nitrosation in secondary metabolism has not been studied at the beginning of this thesis work. In recent years, several biosynthetic gene clusters that produce *N*-nitroso-containing natural products, such as the antimetabolite L-alanosine and the siderophore gramibactin, have been reported (**Figure 1.18B**).[125-126] The enzymes that form the N–N bond in these metabolites have not been reported. The investigation of N–N bond forming reaction that results in a *N*-nitrosamine in the biosynthesis of streptozotocin (SZN) is the subject of this thesis.

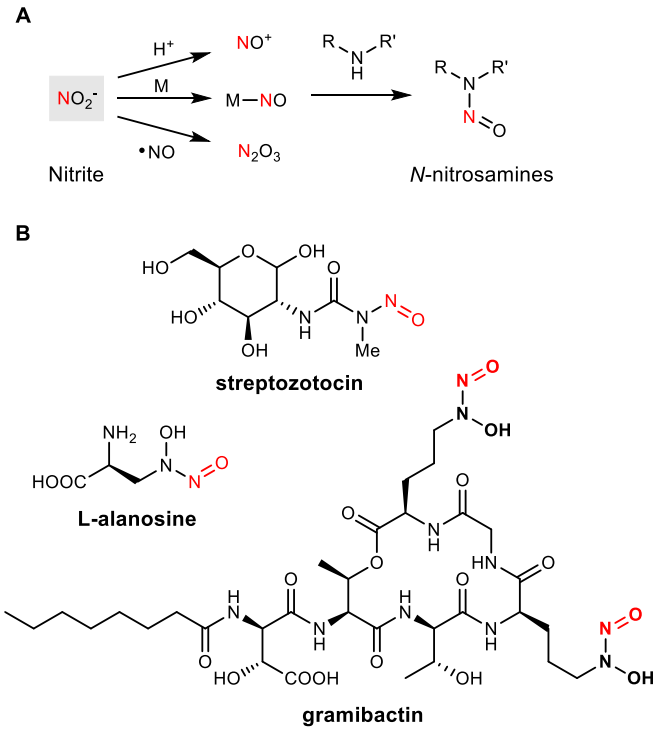


Figure 1.18: *N*-nitrosamines in primary and secondary metabolism. A) Nitrite is central to all previously characterized biological *N*-nitrosation reactions. M = metal. B) Structures of selected *N*-nitroso-containing natural products.

1.4 Non-heme mononuclear iron and diiron metalloenzymes in natural product biosynthesis

In natural product biosynthesis, metalloenzymes play important functions in installing challenging oxidation reactions. This section will briefly review iron-containing metalloenzymes. Oxidoreductases containing a non-heme diiron metallocofactor catalyze a diverse range of challenging oxidative reactions, including hydroxylation of unactivated carbon centers, desaturation of alkanes, *N*-oxygenation, and epoxidation of alkenes **(Figure 1.19, path b)**.[127-128] These enzymes utilize an oxygen-bridged diiron cofactor, with each iron subsite ligated by histidine and acidic residues (aspartate and glutamate), to activate molecular oxygen and affect subsequent oxidation reactions.

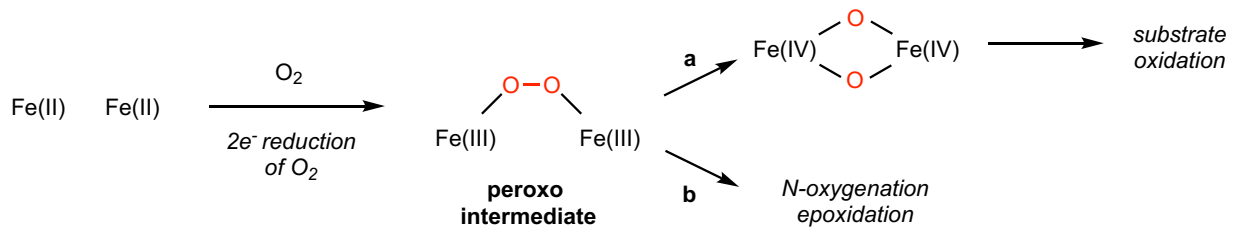


Figure 1.19: Diiron enzymes activate molecular oxygen to form a peroxo intermediate that is competent for diverse oxidation reactions.

Oxygen activation occurs upon oxidative addition of O_2 to the $Fe_2^{II/II}$ center with the diiron center delivering two electrons to O_2 to generate a bridging, μ-(hydro)peroxo-$Fe_2^{III/III}$ species.[129] Interestingly, spectroscopic characterization of this intermediate in several systems has revealed structural differences that may be important for catalyzing these diverse oxidative transformations. In some cases, this μ-(hydro)peroxo-$Fe_2^{III/III}$ species has been shown to be a competent oxidation catalyst, however, in other systems this intermediate is further transformed to high-valent iron-oxo species, including the $Fe_2^{IV/IV}$ complex of soluble methane monooxygenase (**Figure 1.19, path a**) Both types of intermediates subsequently catalyze oxidation of the substrate, completing a net four-electron reduction of O_2 in which two electrons came from the substrate and two from the initial $Fe_2^{II/II}$ center.

The reaction with substrate typically generates a stable $Fe_2^{III/III}$ species. Thus, to complete the catalytic cyclic and regenerate the oxygen-reactive $Fe_2^{II/II}$ species, two external electrons must be provided. The electron delivery system varies between enzymes, but generally uses reduced nicotinamide cofactors as the external electron source. Characterization of the non-heme diiron nitro-forming *N*-oxygenases AurF and CmlI has suggested two novel mechanisms that differ from typical diiron oxidoreductases in terms of their reaction stoichiometry and requirement for external electrons. (**Figure 1.20**). In depth biochemical, spectroscopic, and structural investigations of these enzymes have provided the most detailed insights of nitro group biosynthesis obtained to date.

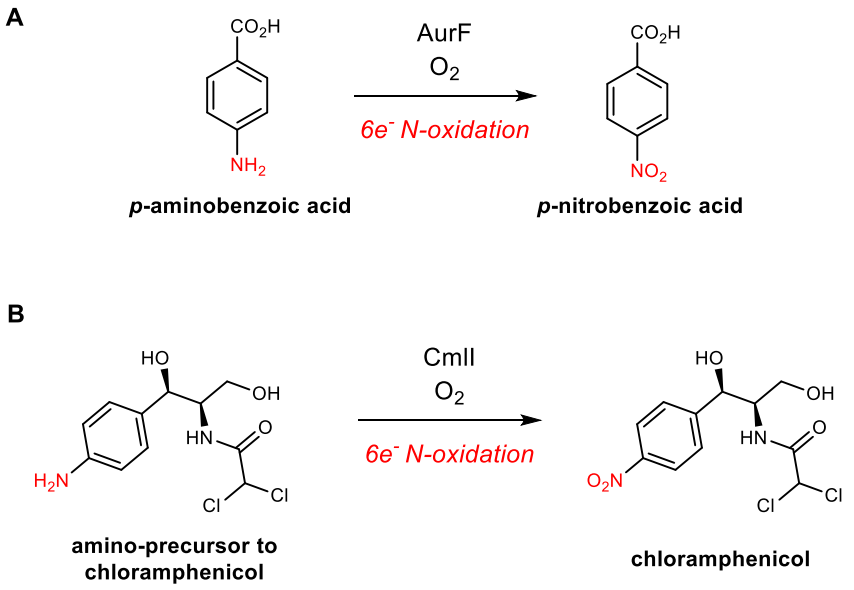


Figure 1.20: Reactions catalyzed by diiron-containing arylamine *N*-oxygenases **A**) AurF and **B**) CmlI.

In 2004, the structure of a new class of diiron containing protein with a heme-oxygenase-like fold was reported.[130] The founding member, **C**hlamydia **P**rotein **A**ssociating with **D**eath **D**omains (CADD), is produced by the intracellular pathogen *Chlamydia trachomatis* and is known to induce apoptosis in several human cell lines. The function of this enzyme remains undetermined. In recent years, enzymes with this structural fold have been reported in natural product biosynthetic pathways. For example, UndA was determined to be involved in alkane biosynthesis by *Pseudomonas* strains. UndA catalyzes the oxidative decarboxylation of lauric acid to form a terminal alkene product (Chapter 3.1).[131-132] Since the discovery of UndA, reactions catalyzed by enzymes containing this diiron binding domain have expanded to include *N*-oxygenation and oxidative C–N bond cleavage.[133-134] This enzyme class will be further discussed in the introductory section of Chapter 3.

Non-heme mononuclear iron oxygenases also perform challenging oxidative reactions in the context of xenobiotic degradation (discussed in Chapter 3) and natural product biosynthesis.[135]

They are members of the larger non-heme mononuclear iron enzyme family which catalyzes a diverse array of oxidative reactions.[136] Members of this family use a single iron atom coordinated by a common 2-His-1-carboxylate facial triad to activate molecular oxygen for subsequent oxygenation reactions.[137] One subfamily of non-heme mononuclear iron enzymes can couple the two-electron oxidation of α-ketoglutarate to succinate and carbon dioxide to generate a high-valent iron oxo-Fe^{IV} species. This highly reactive intermediate can promote a variety of challenging oxidative reactions, such as hydroxylation of unactivated carbon atoms. (**Figure 1.21**). This enzyme family could provide new opportunities in engineering new biocatalysts catalyzing oxidation reactions.

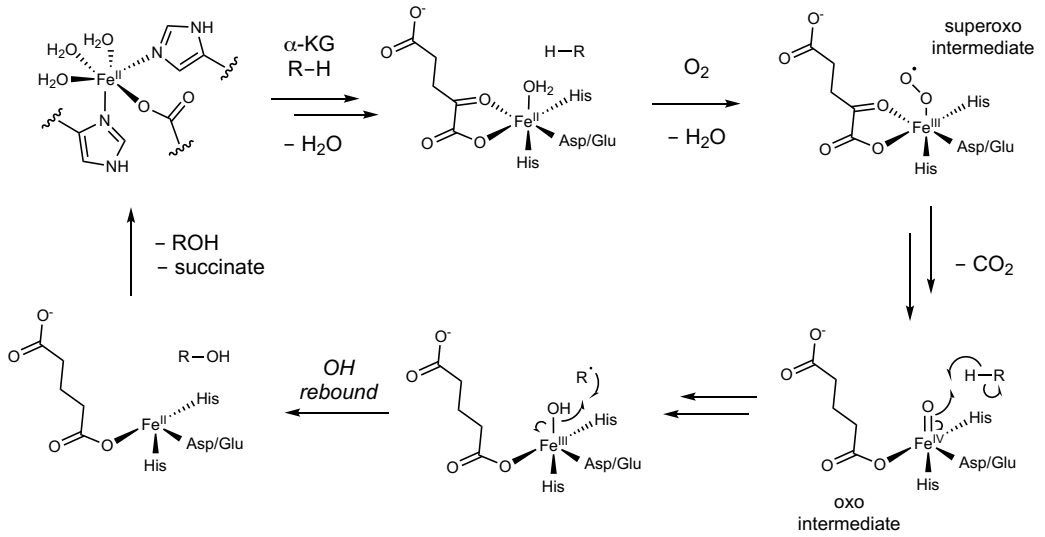


Figure 1.21: The mechanism of substrate hydroxylation by an α-ketoglutarate-dependent monooxygenase.

The Rieske-containing subfamily contain two metal centers: a Rieske [2Fe-2S] cluster (**Figure 1.22A**), which is bound by the conserved sequence motif $CXH_{17}CX_2H$, and a mononuclear iron site anchored by the 2-His-1-carboxylate facial triad. The Rieske [2Fe-2S] center serves to transfer electrons, typically from NADH via ferredoxin partners, to the mononuclear iron site where oxygen activation and catalysis occur. An example of Rieske containing *N*-oxygenase

metalloenzymes is PrnD, a nitro-forming enzyme found in pyrrolnitrin biosynthesis in *Pseudomonas pyrrocinia* (**Figure 1.22B**).[117] *cis*-diol-forming dioxygenases are also well-studied enzymes of this family.[138]

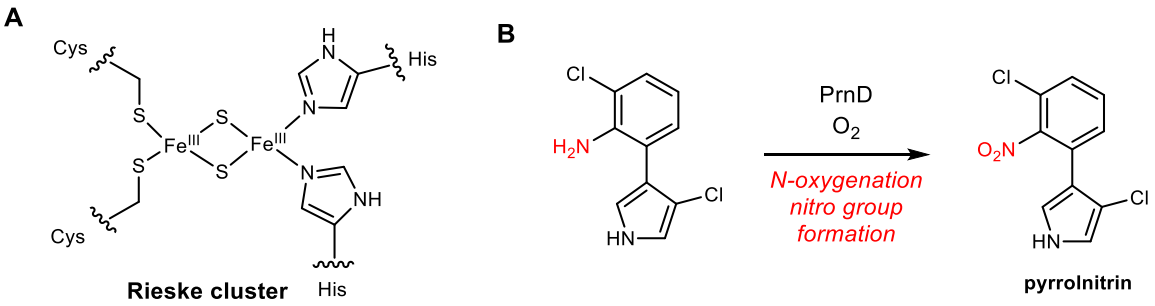


Figure 1.22: **A**) The structure of a Rieske [2Fe-2S] cluster. **B**) The *N*-oxygenation reaction catalyzed by PrnD, a Rieske non-heme mononuclear iron enzyme.

The mechanism of oxygen activation occurs via binding of O_2 to the Fe^{II} site followed by a one-electron transfer from the Rieske site to the mononuclear iron site, and subsequent two-electron reduction of O_2 to afford a side-on (hydro)peroxo-Fe^{III} intermediate. This intermediate has been proposed to act directly as the oxidizing agent, or alternatively, undergo O–O bond cleavage to generate the high-valent HO–Fe^{V}=O species which affects oxygenation.[139-140]

Another well-studied family of non-heme mononuclear iron enzymes are the extra- and intra-diol forming ring cleaving diooxygenases.[141] These enzymes are important in degrading aromatic molecules, such as lignin, in the environment. A typical mechanism involved initial substrate binding to a reduced Fe^{II} metal center (**Figure 1.23**). Subsequent binding and reduction of O_2 result in a Fe^{III}-superoxo intermediate. The Fe^{III} can be reduced by the electron-rich substrate, resulting in a substrate radical which can recombine with the superoxide to form an Fe^{II}-peroxo bridge intermediate. A subsequent Criegee rearrangement and hydrolysis of the lactone afford the cleaved product. Most of these ring-cleaving dioxygenases bear a 2-His-1Asp/Glu facial triad, although several 3-His metal binding sites have been discovered (see below).[142]

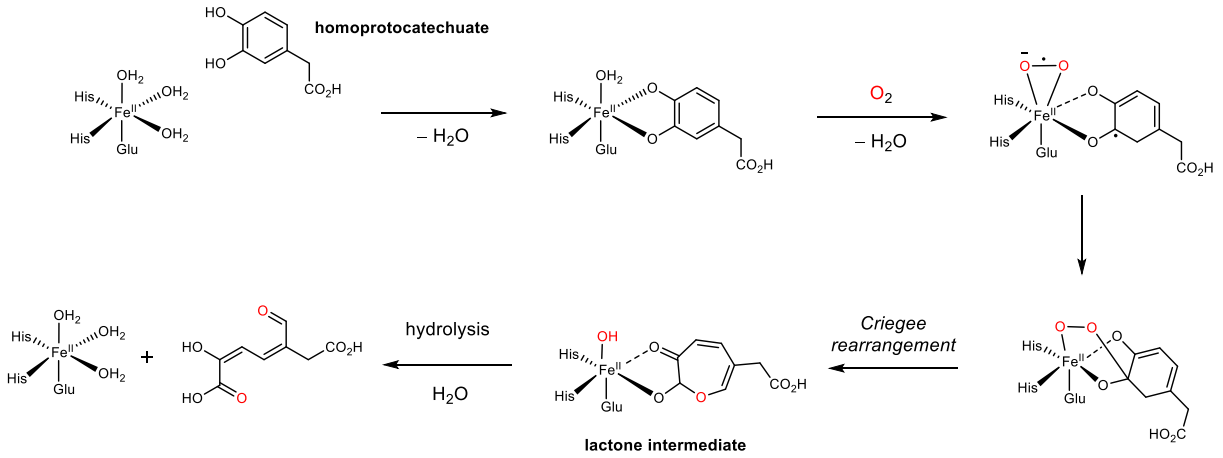


Figure. 1.23. Proposed mechanism of the reaction catalyzed by protocatechuate dioxygenase.[141]

In addition to the more common 2-His-1-carboxylate facial triad, non-heme mononuclear enzymes can also feature a 3-His facial triad motif (**Figure 1.24**).[142] Cysteine dioxygenase is a well-studied enzyme of this family. Cysteine dioxygenase is a cupin-domain containing protein found in both prokaryotes and eukaryotes and is an important enzyme in regulating intracellular levels of L-cysteine.[143] CDO catalyzes the oxidation of four electron oxidation of L-cysteine to L-cysteine sulfinic acid using a starting Fe^{II}-cofactor. Interestingly, the enzyme does not require external electrons, as the four electrons required to reduce O_2 to H_2O is fully provided by the cysteine substrate.

Other 3-His facial triad enzymes include DkeI, which catalyzes the cleavage of the diketone acetylcetone to acetate and methylgloxyl.[142] Several ring cleaving mononuclear enzymes such as gentisate dioxygenase and 1-hydroxy-2-naphthoate dioxygenase also utilize a 3-His metal binding motif.

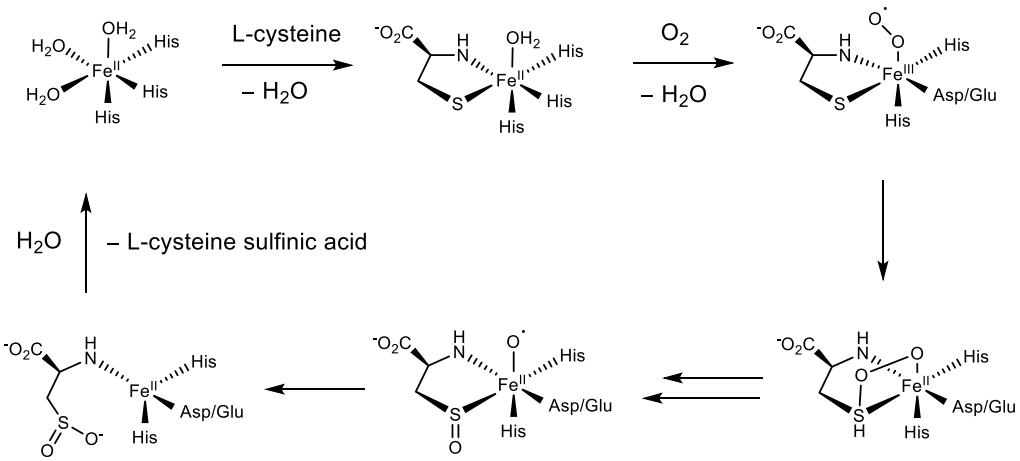


Figure 1.24: The reaction catalyzed by cysteine dioxygenase (CDO), a non-heme mononuclear iron enzyme containing a 3-His motif.

1.5 Chapter Preview

The objective of this thesis research was to discover and characterize the enzymatic chemistry and biosynthetic logic employed to construct the reactive *N*-nitroso functional group in natural product biosynthesis. This dissertation communicated describes our efforts to identify the genes and enzymes involved in streptozotocin biosynthesis. In the course of this work, we discovered and characterized two enzymes that install the critical *N*-nitrosourea functional group, which is responsible for the biological activity of this natural product. With the elucidation of the genetic and biochemical basis for *N*-nitroso biosynthesis, our work now enables investigations of the biosynthesis of additional *N*-nitroso and N–N bond containing natural products and targeted genome mining efforts to identify additional *N*-nitroso compounds. In Chapter 2, we present the identification and validation of the putative streptozotocin biosynthetic gene cluster using a combination of *in vivo* feeding experiments, *in vitro* biochemistry, and gene knockout studies. Chapter 3 of this work describes the biochemical and structural characterization of the *N*-nitrosating metalloenzyme SznF. Finally, in Chapter 4**,** we will describe the roles of the ATP-grasp

enzymes encoded in the streptozotocin biosynthetic gene cluster and assess the distribution of SznF-homologs in microbial genomes.

1.6 References

1) Waldman, A. J.; Ng, T.; Wang, P.; Balskus, E. P. Hetero-Heteroatom bond formation in natural product biosynthesis. *Chem. Rev.* **2017**, *117*, 5784.

2) Davies, J.; Ryan, K. S. Introducing the parvome: bioactive compounds in the microbial world. *ACS Chem. Biol.* **2012**, *7*, 252.

3) Seyedsayamdost, M. R.; Carr, G.; Kolter, R.; Clardy, J. Roseobacticides: Small molecule modulators of an algal-bacterial symbiosis. *J. Am. Chem. Soc.* **2011**, *133*, 18343.

4) Crawford, J. M.; Clardy, J. Bacterial symbionts and natural products. *Chem. Commun.* **2011**, *47*, 7559.

5) Schorn, M. A.; Alanjary, M. M; Aguinaldo, K.; Korobeynikov, A.; Podell, S.; Patin, N.; Lincecum, T.; Jensen, P. R.; Ziemert, N.; Moore, B. S. Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology (Reading, England)* **2016**, *162*, 2075.

6) Neilson, E.; Goodger, J. Q. D.; Woodrow, I. E.; Møller, B. L. Plant chemical defense: at what cost? *Trends Plant Sci.* **2013**, *18*, 250.

7) Sherpa R. T.; Reese C. J.; Montazaeri, A. H. Application of iChip to grow "uncultivable" microorganisms and its impact on antibiotic discovery. *J. Pharm. Pharm. Sci.* **2015**, *18*, 303.

8) Tanner, K.; Vilanova, C.; Porcar, M. Bioprospecting challenges in unusual environments. *Microb. Biotechnol.* **2017**, *10*, 671.

9) Wakimoto, T.; Egami, Y.; Nakashima, Y.; Wakimoto, Y.; Mori, T.; Awakawa, T.; Ito, T.; Kenmoku, H.; Asakawa, Y.; Piel, J.; Abe, I. Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat. Chem. Biol.* **2014**, *10*, 648.

10) Donia, M. S.; Fischbach, M. A. Small molecules from the human microbiota. *Science* **2015**, *349*, 1254766.

11) Wilson, M. R.; Zha, L.; Balskus, E. P. Natural product discovery from the human microbiome. *J. Biol. Chem.* **2017**, *292*, 8546.

12) Robles, O.; Romo, D. Chemo- and site-selective derivatizations of natural products enabling biological studies. *Nat. Prod. Rep.* **2014**, *31*, 318.

13) Brown, D. G.; Lister, T.; May-Dracka, T. L. New natural products as new leads for antibacterial drug discovery. *Bioorg. Med. Chem.* **2014**, *24*, 413.

14) Miller, L. H.; Su, X. Artemisinin: Discovery from the Chinese herbal garden. *Cell* **2011**, *146*, 855.

15) Katz, L.; Baltz, R.; Natural product discovery: past, present, and future. *J. Indust. Microbiol. Biotechnol.* **2016**, *43*, 155.

16) Kinghorn, A.; Chin, Y.; Swanson, S. Discovery of natural product anticancer agents from biodiverse organisms. *Curr. Opin. Drug Discov. Devel.* **2009**, *12*, 189.

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

18) Newman, D.; Cragg, G. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629.

19) Shen, B. A new golden age of natural products drug discovery. *Cell* **2015**, *163*, 1297.

20) Sidebottom, A. M.; Carlson, E. E. A reinvigorated era of bacterial secondary metabolite discovery. *Curr. Opin. Chem. Biol.* **2015**, *24*, 104.

21) Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence–function space and genome context to discover novel functions. *Biochemistry* **2017**, *56*, 4293.

22) Tang, X.; Li, J.; Millán-Aguiñaga, N.; Zhang, J. J.; O'Neill, E. C.; Ugalde, J. A.; Jensen, P. R.; Mantovani, S. M.; Moore, B. S. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **2015**, *10*, 2841.

23) Yan, Y.; Liu, Q.; Zang, X.; Yuan, S.; Bat-Erdene, U.; Nguyen, C.; Gan, J.; Zhou, J.; Jacobsen, S.; Tang, Y. Resistance-gene-directed discovery of a natural product herbicide with a new mode of action. *Nature* **2018**, *559*, 415.

24) Medema, M.; 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. *Nucleic Acids Res.* **2011**, *39*, 339.

25) Hadjithomas, M.; Chen, I. A.; Chu, K.; Huang, J.; Ratner, A.; Palaniappan, K.; Andersen, E.; Markowitz, V.; Kyrpides, N. C.; Ivanova, N. N. IMG-ABC: new features for bacterial secondary metabolism analysis and targeted biosynthetic gene cluster discovery in thousands of microbial genomes. *Nucleic Acids Res.* **2017**, *45*, 560.

26) Katz, L.; Baltz, R. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155.

27) Zerikly, M.; Challis, G. Strategies for the discovery of new natural products by genome mining. *Chembiochem* **2009**, *10*, 625.

28) Walsh, C.; Fischbach, M. Natural Products Version 2.0: Connecting Genes to Molecules. *J. Am. Chem. Soc.* **2010**, *132*, 2469.

29) Chang, F.; Ternei, M.; Calle, P. Y.; Brady, S. F. Targeted metagenomics: finding rare tryptophan dimer natural products in the environment. *J. Am. Chem. Soc.* **2015**, *137*, 6044.

30) Wright, G. Synthetic biology revives antibiotics: re-engineering natural products provides a new route to drug discovery. *Nature* **2014**, *509*, 13.

31) Montiel, D.; Kang, H.; Chang, F.; Charlop-Powers, Z.; Brady, S. F. Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc. Natl. Acad. Soc.* **2015**, *112*, 8953.

32) Bonet, B.; Teufel, R.; Cruesemann, M.; Ziemert, N.; Moore, B. Direct capture and heterologous Expression of Salinispora natural product genes for the biosynthesis of enterocin. *J. Nat. Prod.* **2015**, *78*, 539.

33) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013**, *76*, 1686.

34) Klitgaard, A.; Nielsen, J. B.; Frandsen, R. J. N.; Andersen, M. R.; Nielsen, K. F. Combining stable isotope labeling and molecular networking for biosynthetic pathway characterization. *Anal. Chem.* **2015**, *87*, 6520.

35) Schneider, B. A.; Balskus, E. P. Discovery of small molecule protease inhibitors by investigating a widespread human gut bacterial biosynthetic pathway. *Tetrahedron* **2018**, *74*, 3215.

36) Vila-Farres, X.; Chu, J.; Inoyama, D.; Ternei, M. A.; Lemetre, C.; Cohen, L. J.; Cho, W.; Reddy, B. V.; Zebroski, H. A.; Freundlich, J. S.; Perlin, D. S.; Brady, S. F. Antimicrobials inspired by nonribosomal peptide synthetase gene clusters. *J. Am. Chem. Soc.* **2017**, *139*, 1404.

37) Lau, W.; Sattely, E. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **2015**, *349*, 1224.

38) Brunson, J. K.; Mckinnie, S. M. K.; Chekan, J. R.; Mccrow, J. P.; Miles, Z. D.; Bertrand, E. M.; Bielinski, V. A.; Luhavaya, H.; Oborník, M.; Smith, G. J.; Hutchins, D. A.; Allen, A. E.; Moore, B. S. Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **2018**, *361*, 1356.

39) Reetz, M. T. Biocatalysis in organic chemistry and biotechnology: past present and future. *J. Am. Chem. Soc.* **2013**, *135*, 12480.

40) Narayan, A.; Sherman, D. Re-engineering nature's catalysts. *Science* **2013**, *339*, 283.

41) Clouthier, C. M.; Pelletier, J. N. Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis. *Chem. Soc. Rev.* **2012**, *41*, 1585.

42) Sheldon, R. A.; Wooley, J. M. Role of biocatalysis in sustainable chemistry. *Chem. Rev.* **2017**, *118*, 801.

43) Dutta, A.; Lense, S.; Roberts, J. A. S.; Shaw, W. J. Learning from Nature: Arg-Arg pairing enhances H_2 oxidation catalyst. *Angew. Chem.* **2014**, *53*, 6487.

44) Gormisky, P. E.; White, M. C. Catalyst-controlled aliphatic C-H oxidations with a predictive model for site-selectivity. *J. Am. Chem. Soc.* **2013**, *135*, 14052.

45) Iovan, D. A.; Wilding, M. J. T.; Baek, Y.; Hennessy, E. T.; Betley, T. A. Catalyst-controlled aliphatic C-H oxidations with a predictive model for site-selectivity. *Angew. Chem. Int. Ed.* **2017**, *56*, 15599.

46) Blair L. M.; Sperry, J. Natural Products containing a nitrogen–nitrogen bond. *J. Nat. Prod.* **2013**, *76*, 794.

47) Le Goff, G.; Ouazzani, J. Natural hydrazine-containing compounds: Biosynthesis, isolation, biological activities and synthesis. *Bioorg. Med. Chem.* **2014**, *22*, 6529.

48) Rosen, B. R.; Werner, E. W.; O'Brien, A. G.; Baran P. S. Total Synthesis of Dixiamycin B by Electrochemical Oxidation. *J. Am. Chem. Soc.* **2014**, *136*, 5571.

49) Nawrat C. C.; Moody C. J. Natural products containing a diazo group. *Nat. Prod. Rep.* **2011**, *28*, 1426.

50) Fusari, S. A.; Frohardt, R. P.; Ryder, A.; Haskell, T. H.; Johannes, D. W.; Elder, C. C.; Bartz, Q. R. Azaserine, a new tumor-inhibitory substance. isolation and characterization. *J. Am. Chem. Soc.* **1954**, *76*, 2878.

51) Colis, L. C.; Woo, C. M.; Hegan, D. C.; Li, Z.; Glazer, P. M.; Herzon, S. B. The cytotoxicity of (−)-Lomaiviticin A arises from induction of double-strand breaks in DNA. *Nat. Chem.* **2014**, *6*, 504.

52) Bergy, M. E.; Pyke, T. R. Cremeomycin and process for making. 3350269. US. **1967** Oct 31;

53) Suzuki, H.; Ohnishi, Y.; Furusho, Y.; Sakuda, S.; Horinouchi, S. Novel benzene ring biosynthesis from C3 and C4 primary metabolites by two enzymes. *J. Biol. Chem.* **2006**, *281*, 36944.

54) Waldman, A. J.; Pechersky, Y.; Wang, P.; Wang, J. X.; Balskus, E. P. The cremeomycin biosynthetic gene cluster encodes a pathway for diazo formation. *Chembiochem.* **2015**, *16*, 2172.

55) Sugai, Y.; Katsuyama, Y.; Ohnishi, Y. A nitrous acid biosynthetic pathway for diazo group formation in bacteria. *Nat. Chem. Biol.* **2016**, *12*, 73.

56) Robbel, L.; Helmetag, V.; Knappe, T. A.; Marahiel, M. A. Consecutive enzymatic modification of ornithine generates the hydroxamate moieties of the siderophore erythrochelin. *Biochemistry*. **2011**, *50*, 6073.

57) Winter, J. M.; Jansma, A. L.; Handel, T. M.; Moore, B. S. Formation of the pyridazine natural product azamerone by biosynthetic rearrangement of an aryl diazoketone. *Angew. Chem. Int. Ed.* **2009**, *48*, 767.

58) Katsuyama, Y.; Sato, Y.; Sugai, Y.; Higashiyama, Y.; Senda M.; Senda, T.; Ohnishi, Y. Crystal structure of the nitrosuccinate lyase CreD in complex with fumarate provides insights into the catalytic mechanism for nitrous acid elimination. *FEBS J.* **2018**, *285*, 1540.

59) Waldman, A. J.; Balskus, E. P. Discovery of a diazo-forming enzyme in cremeomycin biosynthesis. *J. Org. Chem.* **2018**, *83*, 7539.

60) Ito, S.; Matsuya, T.; Omura, S.; Otani, M.; Nakagawa, A.; Takeshima, H.; Iwai, Y.; Ohtani, M.; Hata, T. A new antibiotic, kinamycin. *J. Antibiot. (Tokyo)* **1970**, *23*, 315.

61) Smitka, T. A.; Bonjouklian, R.; Perun, T. J.; Hunt, A. H.; Foster, R. S.; Mynderse, J. S.; Yao, R. C. A80316A, a new kinamycin type antibiotic. *J. Antibiot. (Tokyo)* **1992**, *45*, 581.

62) Young, J. J.; Ho, S. N.; Ju, W. M.; Chang, L. R. FL-120A D', new products related to kinamycin from Streptomyces chattanoogensis *subsp.* taitungensis subsp. nov. II. Isolation and Structure Determination. *J Antibiot (Tokyo)* **1994**, *47*, 681.

63) Omura, S.; Nakagawa, A.; Yamada, H.; Hata, T.; Furusaki, A.; Watanabe, T. Structures and biological properties of kinamycin A, B, C, and D. *Chem. Pharm. Bull.* **1973**, *21*, 931.

64) Omura, S.; Nakagawa, A.; Yamada, H.; Hata, T.; Furusaki, A.; Watanabe, T. Structures of kinamycin C, and the structural relationship among kinamycin A, B, C, and D. *Chem. Pharm. Bull.* **1971**, *19*, 2428.

65) Gould, S. J.; Tamayo, N.; Melville C. R.; Cone, M. C. Revised structures for the kinamycin antibiotics: 5-diazobenzo[b]fluorenes rather than benzo[b]carbazole cyanamides. *J. Am. Chem. Soc.* **1994**, *116*, 2207.

66) Mithani, S.; Weeratunga, G.; Taylor, N. J.; Dmitrienko, G. I. The kinamycins are diazofluorenes and not cyanocarbazoles. *J. Am. Chem. Soc.* **1994**, *116*, 2209.

67) Laufer, R. S.; Dmitrienko, G. I. Diazo group electrophilicity in kinamycins and lomaiviticin a: potential insights into the molecular mechanisim of antibacterial and antitumor activity. *J. Am. Chem. Soc.* **2002**, *124*, 1854.

68) Feldman, K. S.; Eastman, K. J. Studies on the mechanism of action of prekinamycin, a member of the diazoparaquinone family of natural products: evidence for both sp2 radical orthoquinonemethide intermediates. *J. Am. Chem. Soc.* **2006**, *128*, 12562.

69) O'Hara KA.; Wu, X.; Patel, D.; Liang, H. Yalowich, J. C.; Chen, N.; Goodfellow, V.; Adedayo, O.; Dmitrienko, G. I.; Hasinoff, B. B. Mechanism of the cytotoxicity of the diazoparaquinone antitumor antibiotic kinamycin F. *Free Radic. Biol. Med.* **2007**, *43*, 1132.

70) Gould, S. J. Biosynthesis of the kinamycins. *Chem. Rev.* **1997**, *97*, 2499.

71) Sato, Y.; Gould, S. J. Biosynthesis of kinamycin D. Incorporation of [1,2-13C]acetate and of [2-2H3,-13C]acetate. *Tetrahedron Lett.* **1985**, *26*, 4023.

72) Sato, Y.; Gould S. J. Biosynthesis of the kinamycin antibiotics by Streptomyces murayamaensis. Determination of the origin of carbon, hydrogen, and oxygen atoms by 13C NMR Spectroscopy. *J. Am. Chem. Soc.* **1986**, *108*, 4625.

73) Ajisaka, K.; Takeshima, H.; Omura, S. Application in biosynthetic studies of 13C isotope shifts in infared spectroscopy. *J. Chem. Soc. Chem. Commun.* **1976**, *14*, 571.

74) Seaton, P. J.; Gould, S. J. Kinamycin biosynthesis. Derivation by excision of an acetate unit from a single-chain decaketide intermediate. *J. Am. Chem. Soc.* **1987**, *109*, 5282.

75) Seaton, P. J.; Gould, S. J. New products related to kinamycin from Streptomyces murayamaensis. II. Structures of pre-kinamycin, keto-anhydrokinamycin, and kinamycins E and F. *J Antibiot (Tokyo)* **1989**, *42*, 189.

76) Gould, S. J.; Melville, C. R. Kinamycin biosynthesis. synthesis, detection, and incorporation of kinobscurinone, a benzo[b]fluorenone. *Bioorg. Med. Chem.* **1995**, *5*, 51.

77) Gould, S. J.; Melville, C. R.; Cone, M. C.; Chen, J.; Carney, J. R. Kinamycin biosynthesis. synthesis, isolation and incorporation of stealthin C, an aminobenzo[b]fluorene. *J. Org. Chem.* **1997**, *62*, 320.

78) Gould, S. J.; Hong, S. T.; Carney, J. R. Cloning and heterologous expression of genes from the kinamycin biosynthetic pathway of Streptomyces murayamaensis. *J. Antibiot (Tokyo)* **1998**, *51*, 50.

79) Wang, B.; Ren, J.; Li, L.; Guo, F.; Pan, G.; Ai, G.; Aigle, B.; Fan, K.; Yang, K. Kinamycin biosynthesis employs a conserved pair of oxidases for B-ring contraction. *Chem Commun.* **2015**, *51*, 8845.

80) Carney, J. R.; Hong, S.; Gould, S. J. Seongomycin: a new sulfur-containing benzo[*b*]fluorene derived from genes clustered with those for kinamycin biosynthesis. *Tetrahedron Lett.* **1997**, *38*, 3139.

81) Wang, P.; Hong, G. J.; Wilson, M. R.; Balskus, E. P. production of stealthin C involves an S-N-type smiles rearrangement. *J. Am. Chem. Soc.* **2017**, *139*, 2864.

82) He, H.; Ding, W. D.; Bernan, V. S.; Richardson, A. D.; Ireland, C. M.; Greenstein, M.; Ellestad, G. A.; Carter, G. T. Lomaiviticins A and B, potent antitumor antibiotics from Micromonospora lomaivitiensis. *J. Am. Chem. Soc.* **2001**, *123*, 5362.

83) Woo, C. M.; Li, Z.; Paulson, E. K.; Herzon, S. B. Structural basis for DNA cleavage by the potent antiproliferative agent (−)-lomaiviticin A. *Proc. Natl. Acad. Sci. USA.* **2016**, *113*, 2851.

84) Kersten, R. D.; Lane, A. L.; Nett, M.; Richter, T. K. S.; Duggan, B. M.; Dorrestein, P. C.; Moore, B. S. bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in Salinispora tropica. *ChemBioChem.* **2013**, *14*, 955.

85) Janso, J. E.; Haltli, B. A.; Eustaquio, A. S.; Kulowski, K.; Waldman, A. J.; Zha, L.; Nakamura, H.; Bernan, V. S.; He, H.; Carter, G. T.; Koehn, F. E.; Balskus, E. P. Discovery of the Lomaiviticin biosynthetic gene cluster in Salinispora pacifica. *Tetrahedron.* **2014**, *70*, 4156.

86) Waldman, A. J.; Balskus, E. P. lomaiviticin biosynthesis employs a new strategy for starter unit generation. *Org. Lett.* **2014**, *16*, 640.

87) Wang, B.; Guo, F.; Ren, J.; Ai, G.; Aigle, B.; Fan, K.; Yang, K. Identification of Alp1U and Lom6 as epoxy hydrolases and implications for kinamycin and lomaiviticin biosynthesis. *Nat. Commun.* **2015**, *6*, 7674.

88) Gao, J.; Ju, K. S.; Yu, X.; Velasquez, J. E.; Mukherjee, S.; Lee, J.; Zhao, C.; Evans, B. S.; Doroghazi, J. R.; Metcalf, W. W, van der donk, W. A. Use of a phosphonate methyltransferase in the identification of the fosfazinomycin biosynthetic gene cluster. *Angew. Chem. Int. Ed.* **2014**, *53*, 1334.

89) Narang, R.; Narasimhan, B.; Sharma, S. A review on biological activities and chemical synthesis of hydrazide derivatives. *Curr. Med. Chem.* **2012**, *19*, 569.

90) Ogita, T.; Gunji, S.; Fukazawa, Y.; Terahara, A.; Kinoshita, T.; Nagaki, H.; Beppu, T. The structures of fosfazinomycins A and B. *Tetrahedron Lett.* **1983**, *24*, 2283.

91) Schiessl, K.; Roller, A.; Hammerschmidt, F. Determination of absolute configuration of the phosphonic acid moiety of fosfazinomycins. *Org. Biomol. Chem.* **2013**, *11*, 7420.

92) Kuroda, Y.; Tanaka, H.; Okamoto, M.; Goto, T.; Kohsaka, M.; Aoki, H.; Imanaka, H. FR-900137. A new antibiotic II. Structure determination of FR-900137. *J Antibiot (Tokyo)* **1980**, *33*, 280.

93) Huang, Z.; Wang, K. K.; Lee, J.; van der Donk, W. A. Biosynthesis of fosfazinomycin is a convergent process. *Chem. Sci.* **2015**, *6*, 1282.

94) Huang, Z.; Wang, K. K. A.; van der Donk, W. A. New insights into the biosynthesis of fosfazinomycin. *Chem Sci*. **2016**, *7*, 5219.

95) Wang, K.; Ng, T.; Wang, P.; Huang, Z.; Balskus, E. P. Glutamic acid is a carrier for hydrazine during the biosyntheses of fosfazinomycin and kinamycin. *Nat*. *Commun*. **2018**, *9*, 1.

96) Matsuda, K.; Hasebe, F.; Shiwa, Y.; Kanesaki, Y.; Tomita, T.; Yoshikawa, H.; Shin-Ya, K.; Kuzuyama, T.; Nishiyama, M. Genome mining of amino group carrier protein-mediated machinery: Discovery and biosynthetic characterization of a natural product with unique hydrazone unit. *ACS Chem*. *Biol*. **2017**, *12*, 124.

97) Twigg, F. F.; Cai, W.; Huang, W.; Liu, J.; Sato, M.; Perez, T. J.; Geng, J.; Dror, M. J.; Montanez, I.; Tong, T. L.; Lee, H.; Zhang, W. Identifying the biosynthetic gene cluster for triacsins with an *N*-hydroxytriazene moiety. *ChemBioChem* **2019**, *20*, 1145.

98) LaRue, T. A.; Child, J. J. Pyrazole in Citrullis vulgaris (Cucurbitacae). *Phytochemistry* **1975**, *14*, 2513.

99) Brown, E. G.; Diffin, F. M. Biosynthesis and metabolism of pyrazole by Cucumis sativus: enzymic cyclization and dehydrogenation of 1,3 diaminopropane. *Phytochemistry* **1990**, *29*, 469.

100) Noe, F. F.; Fowden, L. α-Amino-β-(pyrazolyl-N)-propionic Acid: a new amino acid from Citrullus vulgaris (Watermelon). *Nature* **1959**, *184*, 69.

101) Brown, E. G.; Flayeh, K. A. M.; Gallon, J. R. The biosynthetic origin of the pyrazole moiety of β-pyrazol-1-yl-L-alanine. *Phytochemistry* **1982**, *21*, 863.

102) Flayeh, K. A. M.; Najafi, S. I.; Al-Delymi, A. M.; Hajar, M. A. 1,3-diaminopropane and spermidine in Cucumis sativus (Cucumber) *Phytochemistry* **1984**, *23*, 989.

103) Dunnill, P. M.; Fowden, L. The biosynthesis of β-pyrazol-1-ylalanine. *J*. *Exp*. *Bot*. **1963**, *14*, 237.

104) Murakoshi, I.; Kuramoto, H.; Haginiwa, J.; Fowden, L. The enzymic synthesis of β-substituted alanines. *Phytochemistry* **1972**, *11*, 177.

105) Zhao, G.; Yao, S.; Rothchild, K. W.; Liu, T.; Liu, Y.; Lian, J.; He, H.; Ryan, K. S.; Du, Y. The biosynthetic gene cluster of the C-nucleoside antibiotic pyrazomycin with a rare pyrazole moiety. *Chembiochem* **2019**, *ahead of print*.

106) Wang, S. A.; Ko, Y.; Zeng, J.; Geng, Y.; Ren, D.; Ogasawara, Y.; Irani, S.; Zhang, Y.; Liu, H. W. Identification of the formycin a biosynthetic gene cluster from streptomyces kaniharaensis illustrates the interplay between biological pyrazolopyrimidine formation and de novo purine biosynthesis. *J*. *Am*. *Chem*. *Soc*. **2019**, *141*, 6127.

107) Bevan, K.; Davies, J. S.; Hassall, C. H.; Morton, R. B.; Phillips, D. Amino-acids and peptides. Part X. Characterisation of the monamycins, members of a new family of cyclodepsipeptide antibiotics. *J. Chem. Soc. C* **1971**, *0*, 514.

108) Arroyo, V.; Hall, M. J.; Hassall, C. H.; Yamasaki, K. Incorporation of amino acids into the cyclohexadepsipeptide, monamycin. *J. Chem. Soc. Chem. Commun.* **1976**, *21*, 845.

109) Neumann, C. S.; Jiang, W.; Heemstra, J. R.; Gontang, E. A.; Kolter, R.; Walsh, C. T. Biosynthesis of piperazic acid via N5-hydroxy-ornithine in Kutzneria spp. 744. *ChemBioChem* **2012**, *13*, 972.

110) Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T. Cloning and characterization of the biosynthetic gene cluster for Kutznerides. *Proc. Natl. Acad. Sci. USA.* **2007**, *104*, 16498.

111) Du, Y.; He, H.; Higgins, M.; Ryan, K. A heme-dependent enzyme forms the nitrogen-nitrogen bond in piperazate. *Nat. Chem. Biol.* **2017**, *13*, 836.

112) Langley, B. W.; Lythgoe, B.; Riggs, N. V. *J. Chem. Soc.* **1951**, *0*, 2309.

113) Sandler, S. R.; Karo, W. Organic Functional Group Preparations, Volume 2. Elsevier; **1986**.

114) Guo, Y. Y.; Li, H.; Zhou, Z. X.; Mao, X. M.; Tang, Y.; Chen, X.; Jiang, X. H.; Liu, Y.; Jiang, H.; Li, Y. Q. Identification and biosynthetic characterization of natural aromatic azoxy products from Streptomyces chattanoogensis L10. *Org. Lett.* **2015**, *17*, 6114.

115) Rui, Z.; Petrickova, K.; Skanta, F.; Pospisil, S.; Yang, Y.; Chen, C. Y.; Tsai, S. F.; Floss, H. G.; Petricek, M.; Yu, T. W. Biochemical and genetic insights in asukamycin biosynthesis. *J. Biol. Chem.* **2010**, *285*, 24915.

116) Potterat, O.; Zahner, H.; Metzger, J. W.; Freund, S. Metabolic products of microorganisms. 5-Phenylpentadienoic acid derivatives from Streptomyces sp. *Helv. Chim Acta.* **1994**, *77*, 569.

117) Choi, Y. S.; Zhang, H.; Brunzelle J. S.; Nair S. K.; Zhao, H. In vitro reconstitution and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in aureothin biosynthesis. *Proc. Natl, Acad, Sci, USA.* **2008**, *105*, 6858.

118) Guo, Y.; Li, Z.; Xia, T.; Du, Y.; Mao, X.; Li, Y. Molecular mechanism of azoxy bond formation for azoxymycins biosynthesis. *Nat. Commun.* **2019**, *10*, 4420

119) Winkler, R.; Hertweck, C. Sequential enzymatic oxidation of aminoarenes to nitroarenes via hydroxylamines. *Angew. Chem. Int. Ed.* **2005**, *44*, 4083.

120) Lundberg, J. O.; Weitzberg, E.; Gladwin, M. T. The nitrate–nitrite–nitric oxide pathway in physiology and therapeutics. *Nat. Rev. Drug Discov.* **2008**, *7*, 156.

121) Lundberg, J. O.; Weitzberg, E.; Cole, J. A.; Benjamin, N. Nitrate, bacteria and human health. *Nat. Rev. Micro.* **2004**, *2*, 593.

122) Vitturi, D. A.; Minarrieta, L.; Salvatore, S. R.; Postlethwait, E. M.; Fazzari, M.; Ferrer-Sueta, G.; Lancaster, J. R.; Freeman, B. A.; Schopfer, F. J. Convergence of biological nitration and nitrosation via symmetrical nitrous anhydride. *Nat. Chem. Biol.* **2015**, *11*, 504.

123) Binkerd, E. F.; Kolari, O. E. The history and use of nitrate and nitrite in the curing of meat. *Food Chem. Toxicol.* **1975**, *13*, 655.

124) Bouvard, V.; Loomis, D.; Guyton, K. Z.; Grosse, Y.; Ghissassi, F. E.; Benbrahim-Tallaa, L.; Guha, N.; Mattock, H.; Straif, K1.; International Agency for Research on Cancer Monograph Working Group. Carcinogenicity of consumption of red and processed meat. *Lancet Oncol.* **2015**, *16*, 1599.

125) Hermenau, R.; Ishida, K.; Gama, S.; Hoffman, B.; Pfeifer-Leeg, M.; Plass, W.; Mohr, J. F.; Wichard, T.; Saluz, H. P.; Hertweck, C. *Nat. Chem. Biol.* **2018**, *14*, 841.

126) Murthy, Y K. A.; Thiemann, J. E.; Coronelli, C.; Sensi, P. Alanosine, a new antiviral and antitumour agent isolated from a Streptomyces. *Nature* **1966**, *211*, 1198.

127) Leahy, J. G.; Batchelor, P. J.; Morcomb, S. M. Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* **2003**, *27*, 449.

128) Krebs, C.; Bollinger, J. M.; Booker, S. J. Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like 'diiron-carboxylate' proteins. *Curr. Opin. Chem. Biol.* **2011**, *15*, 291.

129) Jasniewski, A. J.; Que, L.; Dioxygen activation by nonheme diiron enzymes: diverse dioxygen adducts, high-valent intermediates, and related model complexes. *Chem. Rev.* **2018**, *118*, 2554.

130) Schwarzenbacher, R.; Stenner-Liewen, F.; Liewen, H.; Robinson, H.; Yuan, H.; Bossy-Wetzel, E.; Reed, J. C.; Liddington, R. C. Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. *J. Biol. Chem.* **2004**, *279*, 29320.

131) Zhang, B.; Rajakovich, L. J.; Van Cura, D.; Blaesi, E. J.; Mitchell, A. J.; Tysoe, C. R.; Zhu, X.; Streit, B. R.; Rui, Z.; Zhang, W.; Boal, A. K.; Krebs, C.; Bollinger, J. M. Substrate-triggered formation of a peroxo-Fe(III/III) intermediate during fatty acid decarboxylation by UndA. *J. Am. Chem. Soc.* **2019**, *141*, 14510.

132) Manley, O. M.; Fan, R.; Guo, Y.; Makris, T. M. Oxidative decarboxylase UndA utilizes a dinuclear iron cofactor. *J. Am. Chem. Soc.* **2019**, *141*, 8684.

133) Marchand, J. A.; Neugebauer, M. E.; Ing, M. C.; Lin, C.; Pelton, J. G.; Chang, M. C. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **2019**, *567*, 420.

134) Hedges, J. B.; Ryan, K S. In vitro reconstitution of the biosynthetic pathway to the nitroimidazole antibiotic azomycin. *Angew. Chem. Int. Ed.* **2019**, *58*, 11647.

135) Barry, S. M.; Challis, G. L. Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal.* **2013**, *3*, 2362.

136) Bruijnincx, P. C.; van Koten, G.; Klein Gebbink, R. J. Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: Recent developments in enzymology and modeling studies. *Chem. Soc. Rev.* **2008**, *37*, 2716.

137) Koehntop, K. D.; Emerson, J. P.; Que, L. The 2-His-1-carboxylate facial triad: A versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem.* **2005**, *10*, 87.

138) Kovaleva, E. G.; Lipscomb, J. D. Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nat. Chem. Biol.* **2008**, *4*, 186.

139) Goncharenko, K. V.; Vit, A.; Blankenfeldt, W.; Seebeck, F. P. Structure of the sulfoxide synthase EgtB from the ergothioneine biosynthetic pathway. *Angew. Chem. Int. Ed.* **2015**, *54*, 2821.

140) Braunshausen, A.; Seebeck, F. P. Identification and characterization of the first ovothiol biosynthetic enzyme. *J. Am. Chem. Soc.* **2011***, 133,* 1757.

141) Lipscomb, J.D. Mechanism of extradiol aromatic ring-cleaving dioxygenases. *Curr. Opin. Struct. Biol.* **2008**, *18*, 644.

142) Buongiorno, D.; Straganz, G. D. Structure and function of atypically coordinated enzymatic mononuclear non-heme-Fe(II) centers. *Coord. Chem. Rev.* **2013**, *257*, 541.

143) Dominy, J. E.; Simmons, C. R.; Hirschberger, L. L.; Hwang, J.; Coloso, R. M.; Stipanuk, M. H. Discovery and characterization of a second mammalian thiol dioxygenase, cysteamine dioxygenase. *J. Biol. Chem.* **2007**, *282*, 25189.

Chapter 2. Discovery and validation of the streptozotocin (*szn*) biosynthetic gene cluster

Parts of this chapter are adapted from previous publications.[1] The discovery of the *szn* cluster and the results of the stable isotope feeding experiments, gene deletion experiments and *in vitro* biochemical characterization of SznE have been corroborated by an independent study performed by Professor Katherine Ryan's lab (University of British Columbia).[2]

2.1 Introduction

Naturally occurring *N*-nitroso natural products include streptozotocin (SZN, **Figure 2.1A)**.[2] Streptozotocin is an *N*-nitrosourea-containing sugar produced by soil-dwelling bacteria such as *Streptomyces achromogenes* subsp. *streptozoticus* NRRL 2697.[3] SZN is a potent DNA-methylating agent that releases diazomethane upon hydrolysis of the *N*-methyl-*N*-nitrosourea moiety.[4] The *N*-nitroso group is also susceptible to denitrosation, generating nitric oxide (NO) as a byproduct. Owing to its bioactivity, SZN is marketed as a pancreatic cancer drug marketed under the brand name Zanosar®.[5] The utility of SZN as a cancer chemotherapeutic inspired the development of other *N*-nitrosoureas for therapeutic purposes (**Figure 2.1B**).[6] Because of its structural similarity to sugars, SZN selectively kills pancreatic beta cells, which synthesize and secrete insulin, owing to the numerous glucose transporters on the cells' surface. Today, SZN-based chemotherapy is still the standard of care for metastatic pancreatic cancer.[6] SZN is also a widely used inducer of diabetes in animal models.[7]

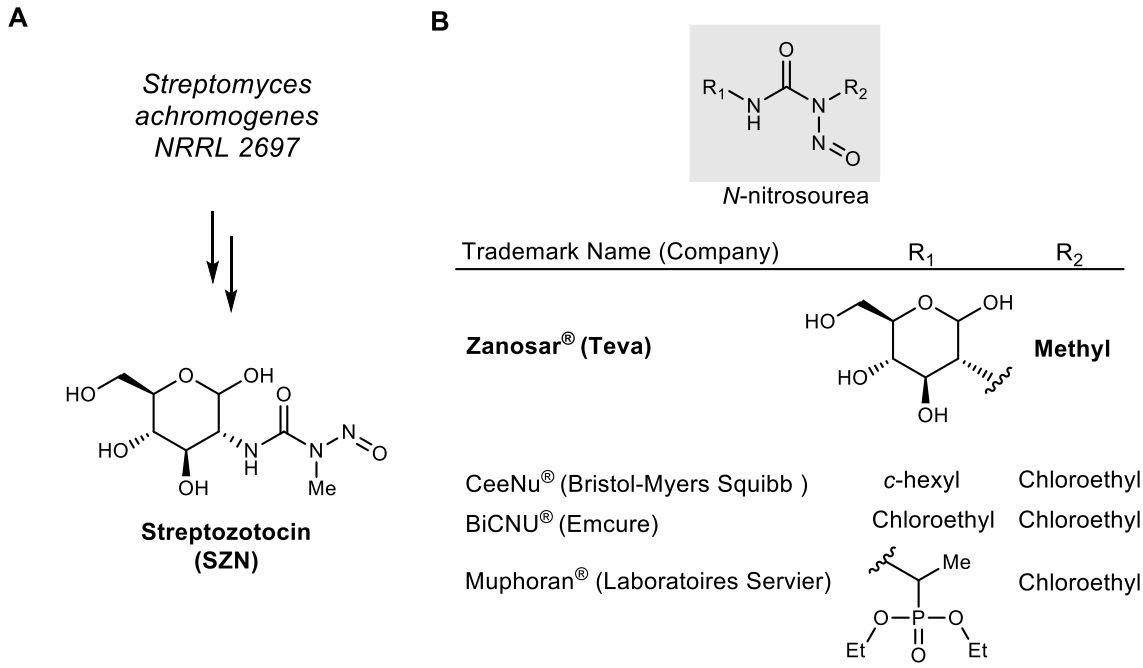


Figure 2.1: A) The structure of streptozotocin, an *N*-methyl-*N*-nitrosourea natural product produced by *Streptomyces achromogenes* NRRL 2697. **B**) The structures of selected FDA-approved *N*-nitrosourea-containing drugs.

As mentioned in Chapter 1, the biosynthesis of *N*-nitroso compounds in secondary metabolism has not been studied extensively. We decided to study the biosynthesis of SZN, a clinically important molecule, to shed light into the *N*-nitrosation reaction in secondary metabolism. Before this work, no gene clusters have been linked to SZN biosynthesis. Previous feeding experiments by Hornemann et al. with radioactively labeled amino acids and sugars indicated that D-glucosamine, L-methionine, and either L-citrulline or L-arginine are biosynthetic precursors to SZN (**Figure 2.2**).[8] Specifically, they observed that the ureido carbon of SZN was labeled when ^{14}C-ureido-L-citrulline or ^{14}C-guanido-L-arginine was fed to the producer *S. achromogenes* NRRL 2697. They also found labeling of the methyl group of SZN when ^{14}C-methyl-L-methionine was fed. The authors also proposed that the distal *N*-nitroso nitrogen of SZN may derive from nitrous acid (or a biochemical equivalent). Based on these feeding results, we formulated a hypothesis in which D-glucosamine would be carbamoylated with L-citrulline-derived carbamoyl phosphate, followed by *N*-methylation by methionine-derived *S*-adenosyl-L-

methionine (SAM). In this hypothesis, the unknown source of nitrite could be activated to an electrophilic nitrosating agent (e.g. nitrous acid), which would nitrosate the *N*-methylurea in the final stages of SZN biosynthesis. One outstanding question this hypothesis raises is whether this *N*-nitrosating process is mediated by an enzyme. To address this question, we began by first identifying a putative streptozotocin gene cluster, which is the subject of this Chapter. We first used a resistance gene-guided genome mining strategy as described in Chapter 1 to identify a putative gene cluster. We then linked this gene cluster to SZN production using gene knockouts, *in vitro* biochemistry, and stable isotope feeding experiments. The results presented in this Chapter will provided a roadmap to uncovering the enzymatic process of *N*-nitrosation, which will be presented in Chapter 3.

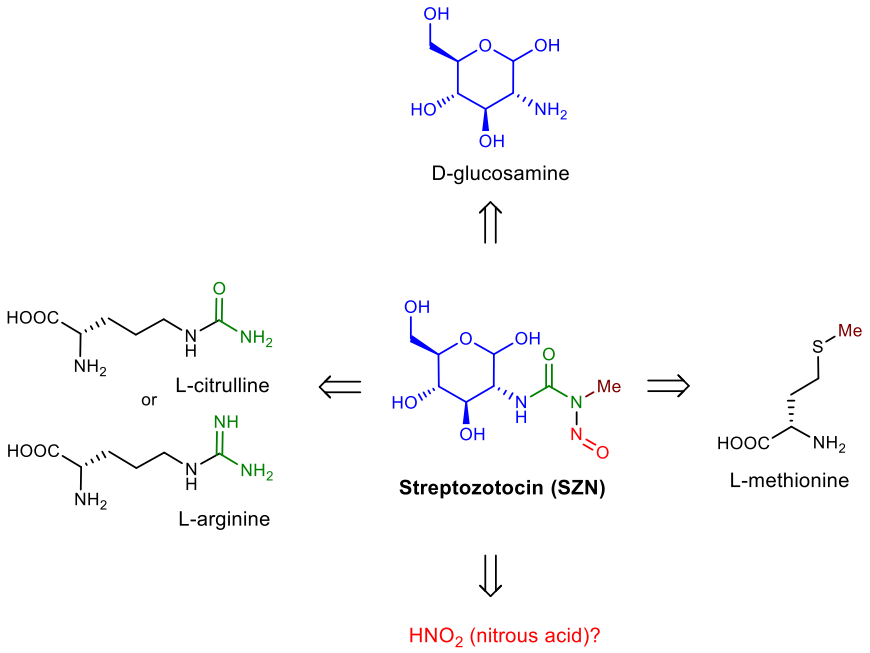


Figure 2.2: Previous feeding experiments suggested that D-glucosamine, L-citrulline or L-arginine, and L-methionine-derived *S*-adenosylmethionine (SAM) would be building blocks for SZN biosynthesis.[8]

2.2 Experimental Results

2.2.1 Identification of the *szn* biosynthetic gene cluster

We first sequenced the genome of *S. achromogenes* var. *streptozoticus* NRRL 2697. Since we had no reference sequences for enzymes that mediate N–N bond formation, we could not use this reactivity to guide genome mining. From the structure of SZN, we hypothesized that its biosynthetic pathway would require an *N*-methyltransferase. However, we had difficulty locating a putative *szn* gene cluster due to the large number of methyltransferases (>100) encoded in the genome. We then searched the genome neighborhoods around putative L-arginine biosynthetic genes and nitric oxide detoxifying enzymes. However, these genomic regions also either lacked a methyltransferase and/or amide-forming enzymes. We have originally proposed that a carbamoyltransferase, which transfers carbamoyl phosphate to amines to form ureas, would be required in SZN biosynthesis.[9] However, we inspected the genome neighborhoods around the genes predicted to encode carbamoyltransferases, and we did not observe genes that encode methyltransferases and enzymes predicted to perform redox chemistry. We then sought alternative strategies for identifying the *szn* gene cluster.

As mentioned in Section 2.1, SZN can hydrolyze to form diazomethane and damage the host's DNA. We hypothesized that bacterial DNA repair enzymes might be clustered with the *szn* biosynthetic genes to protect the producing host from SZN's cytotoxic effects. One such protein is the *O*6-alkylguanine alkyltransferase (*ogt*), which removes alkyl groups from methylated nucleobases with a nucleophilic cysteine residue.[10] When we queried the SZN producer's genome for homologs of *ogt* (from *E. coli*, NCBI Accession: AUG16819.1), we identified a putative 14-kb biosynthetic gene cluster in the *S. achromogenes* genome (**Figure 2.3A**). We also found a homolog of Ada, which is an α-ketoglutarate-dependent enzyme that oxidatively removes methyl

groups from methylated DNA bases. Notably, this gene neighborhood encodes a methyltransferase gene, *sznE* (**Table 2.1**). The putative *szn* gene cluster also encodes a putative iron-containing redox enzyme (SznF) and a ferredoxin (SznG) and carboxylate amine ligases (SznH and SznK). These enzymes could potentially perform the redox and carbamoylation chemistry required in assembling the molecule. Finally, this gene cluster is absent in the genome of a closely related *Streptomyces achromogenes* subsp. *achromogenes*, a strain that does not produce SZN, providing additional support for our assignment (**Figure 2.3B**).[11]

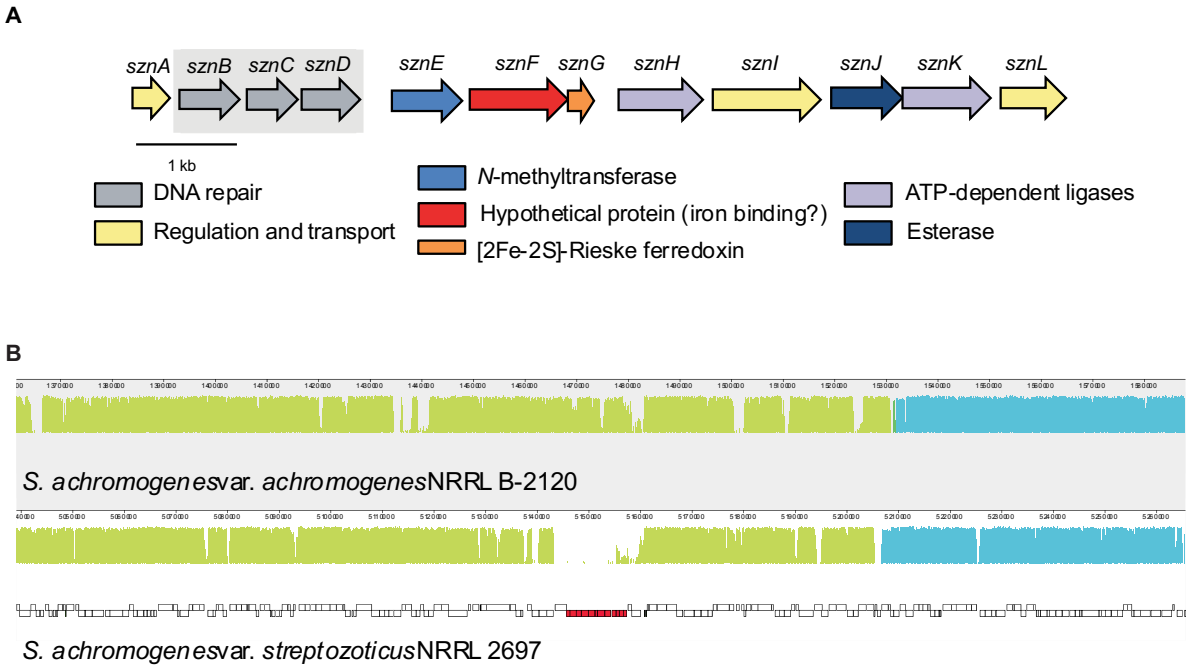


Figure 2.3: **A**) The *szn* biosynthetic gene cluster. Genes encoding DNA repair enzymes are highlighted in grey. **B**) Comparative genomic analysis of *S. achromogenes* var. *achromogenes* NRRL B-2120, a nonproducer of SZN and *S. achromogenes* var. *streptozoticus* NRRL 2697. The *szn* gene cluster is colored in red.

Table 2.1: Annotations of the *szn* biosynthetic genes.

Protein	***Size, aa***	***Predicted function***	***Closest homolog***	***Accession number of homolog***	***% Identity/Similarity***
SznA	197	RNA polymerase sigma-70 factor	*Streptomyces* sp. *di50b*	SCD88966.1	85/90
SznB	266	*O6*-methylguanine-DNA--protein-cysteine methyltransferase (MGMT)	*Streptomyces* sp. *di50b*	SCD88943.1	82/87
SznC	181	Methylated-DNA-[protein]-cysteine S-methyltransferase	*Streptomyces* sp. *di50b*	SCD88925.1	81/85
SznD	255	Proline hydroxylase (Fe(II)-2-oxoglutarate-dependent oxygenase family, AlkB homolog)	*Streptomyces* sp. *NRRL F-2580*	WP_078892722.1	76/81
SznE	361	methyltransferase domain-containing protein	*Streptomyces virginiae*	WP_078908984.1	72/82
SznF	471	Iron-containing redox enzyme	*Streptomyces* sp. *di50b*	SCD88894.1	94/97
SznG	124	Rieske [2Fe-2S] domain-containing protein	*Streptomyces* sp. *di50b*	SCD88886	82/87
SznH	416	ATP-grasp domain-containing protein	*Streptomyces* sp. *SM1*	WP_103542806.1	86/92
SznI	431	MFS transporter	*Streptomyces* sp. *SM1*	WP_103542807.1	86/90
SznJ	318	Alpha/beta hydrolase	*Streptomyces* sp. *SM1*	WP_103542817.1	89/92
SznK	424	ATP-grasp domain-containing protein	*Streptomyces* sp. *SM1*	WP_103542808.1	87/91
SznL	210	LysE family translocator	*Streptomyces* sp. *SM1*	WP_103542809.1	75/82

To experimentally link this gene cluster to SZN biosynthesis, we performed gene inactivation experiments. When we knocked out *sznE, F, H, J,* and *K* individually, we detected either trace or complete loss of SZN production in each mutant (**Figure 2.4**), strongly suggesting that these genes are essential to the production of SZN. The role of SznF in SZN biosynthesis will be further discussed in Chapter 3, while the biochemical functions of SznH and SznK will be discussed in Chapter 4.

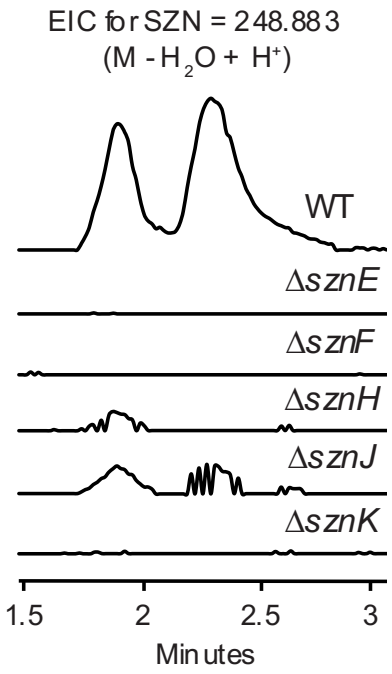


Figure 2.4: Gene inactivation studies demonstrate a link between the *szn* gene cluster and SZN biosynthesis. Extracted ion chromatograms (EICs) for SZN normalized to the same intensity. The two peaks correspond to the α- and β-anomers of SZN.

2.2.2 Biochemical characterization of SznE

To further support this gene cluster's involvement in SZN biosynthesis and begin to understand the order of events in the pathway, we focused on characterizing the putative *N*-methyltransferase enzyme. We set out to (1) identify the SznE substrate with *in vitro* biochemical assays and (2) test if the preferred substrate is a biosynthetic precursor to SZN. Bioinformatic analysis indicated that SznE has strong predicted structural homology to protein arginine methyltransferases (PRMT) found in eukaryotic cells, which are responsible for modifying histones in mammalian cells. Notably, all the basic active site residues predicted to bind a cationic guanidine moiety are present in SznE, further supporting the enzyme can bind L-arginine (**Figure 2.5**).[12] Based on this result, we hypothesized that SznE could methylate L-arginine using *S*-adenosyl-L-methionine (SAM) as a methyl donor.

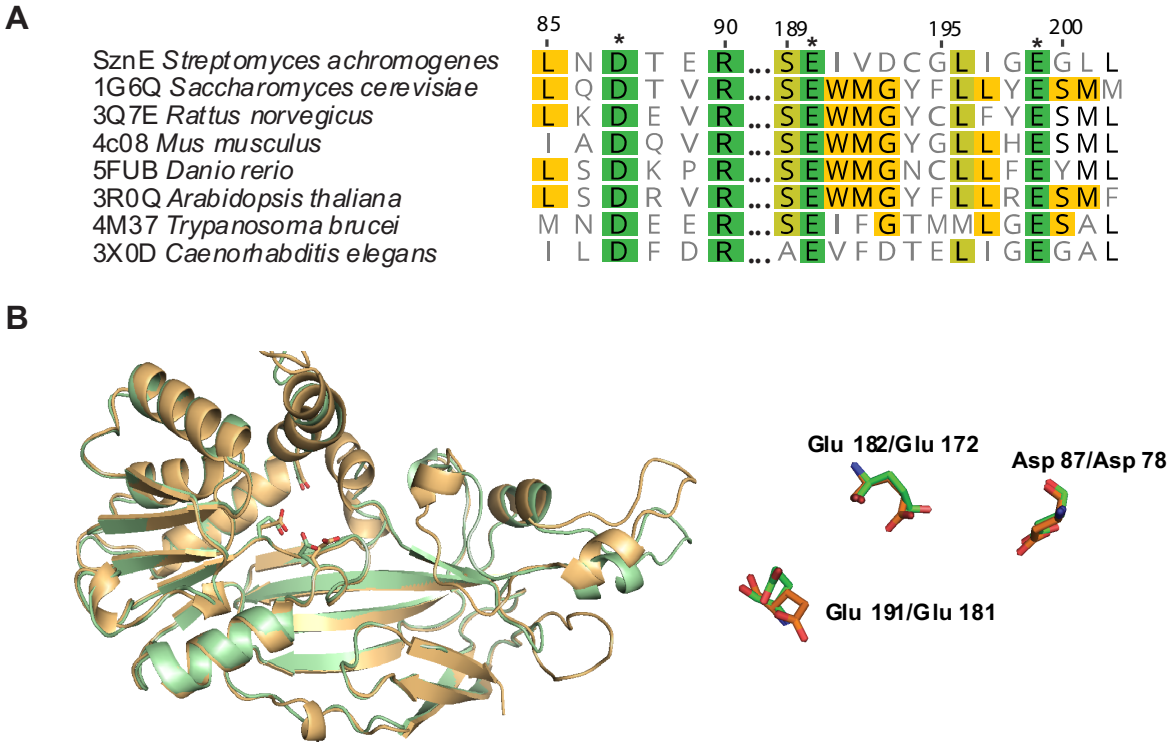


Figure 2.5: Bioinformatic analysis of SznE indicates that the enzyme likely binds L-arginine. **A**) Multiple sequence alignment of SznE with structurally characterized protein arginine methyltransferases (PRMT) from eukaryotes. Conserved residues involved in binding L-arginine are marked with an asterisk. **B**) Overlay of a homology model of SznE (green) with the crystal structure of PRMT7 from *T. brucei* (PDB accession code 4M37) (orange).[12] The highlighted amino acid side chain carboxylates are involved in binding of the basic guanidine group of protein-bound L-arginine.

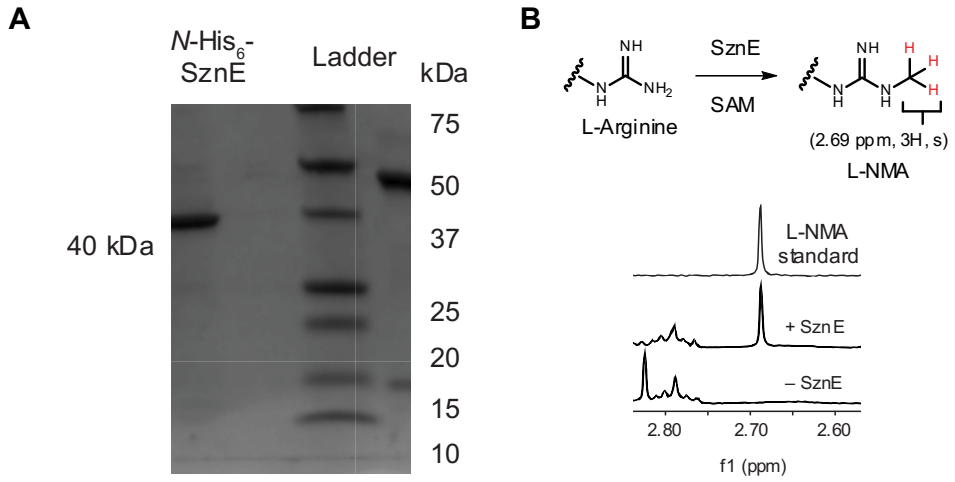


Figure 2.6: SznE catalyzes methylation of L-arginine. **A**) SDS-PAGE of purified SznE. The expected molecular weight is 40 kDa. Ladder = Precision Plus Protein All Blue Standards (BioRad). **B**) ^{1}H NMR assay shows that SznE uses S-adenosyl-L-methionine (SAM) as a methyl donor to convert L-arginine to N^{ω}-methyl-L-arginine (L-NMA) *in vitro*.

We overexpressed and purified *N*-His$_6$-tagged SznE from *E. coli* BL21(DE3) (**Figure 2.6A**). Incubating L-arginine and SAM in the presence of purified SznE resulted in the conversion of L-arginine to L-*N*-monomethylarginine (L-NMA) *in vitro* (**Figure 2.6B**). To demonstrate that L-NMA is an on-pathway intermediate, we chemically synthesized d_3-methyl-L-NMA and fed this amino acid to *S. achromogenes*. We observed 50% deuterium incorporation into SZN (**Figure 2.7A**). This result provided evidence that the methyl group of SZN is derived from L-NMA. We also fed L-NMA to the Δ*sznE* mutant and observed restoration of SZN production (**Figure 2.7B**). This chemical complementation experiment demonstrated that the knockout of *sznE* did not lead to polar mutations. Overall, our *in vitro* and *in vivo* analyses confirm that SznE methylates L-arginine to form L-NMA, a biosynthetic precursor to SZN.

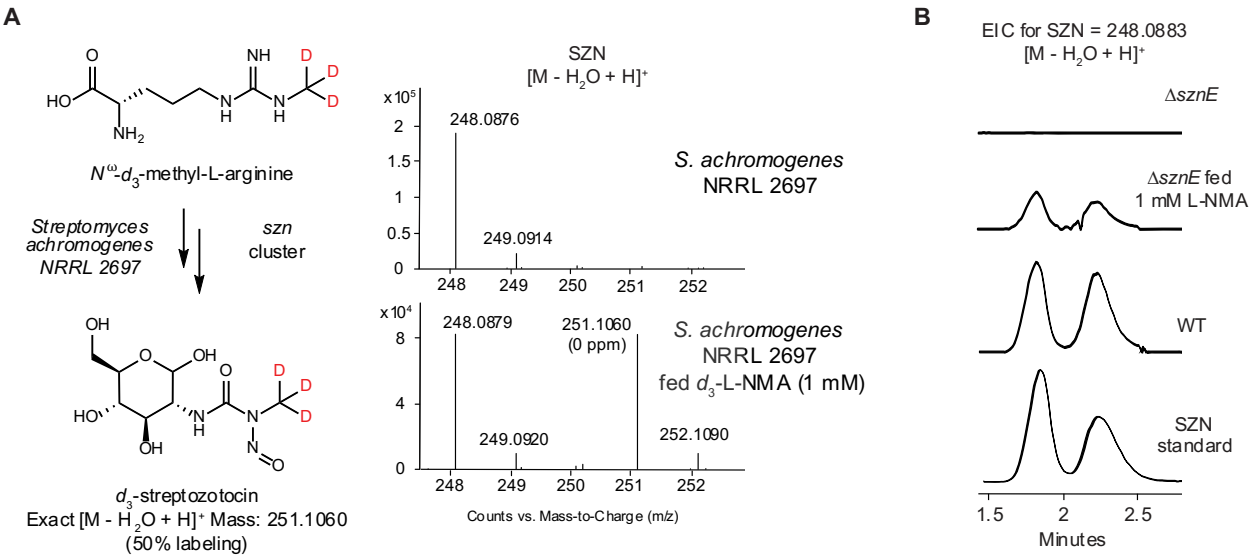


Figure 2.7: L-NMA is an on-pathway intermediate to SZN. **A**) Mass spectra of SZN produced when feeding *S. achromogenes var. streptozoticus* NRRL with d_3-methyl-L-NMA. The expected masses $[M–H_2O+H]^+$ for SZN and d_3-methyl-SZN are 248.0993 and 251.1060, respectively. **B**) LC-MS traces demonstrating restoration of SZN production by the Δ*sznE* mutant upon chemical complementation with L-NMA.

2.2.3 Feeding studies with ^{15}N-nitrogen sources

While we have identified the SZN biosynthetic gene cluster, the origin of the *N*-nitroso group remained unknown. As discussed in Chapter 1, when we began this work all characterized routes for *N*-nitrosation *in vivo* employed nitrite and were not known to be catalyzed by dedicated enzymes. We hypothesized that nitrite might be the source of the distal *N*-nitroso nitrogen atom. To test this hypothesis, we fed ^{15}N-nitrite, nitrate, and ammonium to *S. achromogenes var. streptozoticus* NRRL 2697. Surprisingly, when we fed ^{15}N-labeled nitrite, nitrate, or ammonium salts to *S. achromogenes*, we did not observe isotopically labeled SZN (**Figure 2.8**). We therefore concluded that *N*-nitrosation in SZN assembly likely employed a distinct biosynthetic logic.

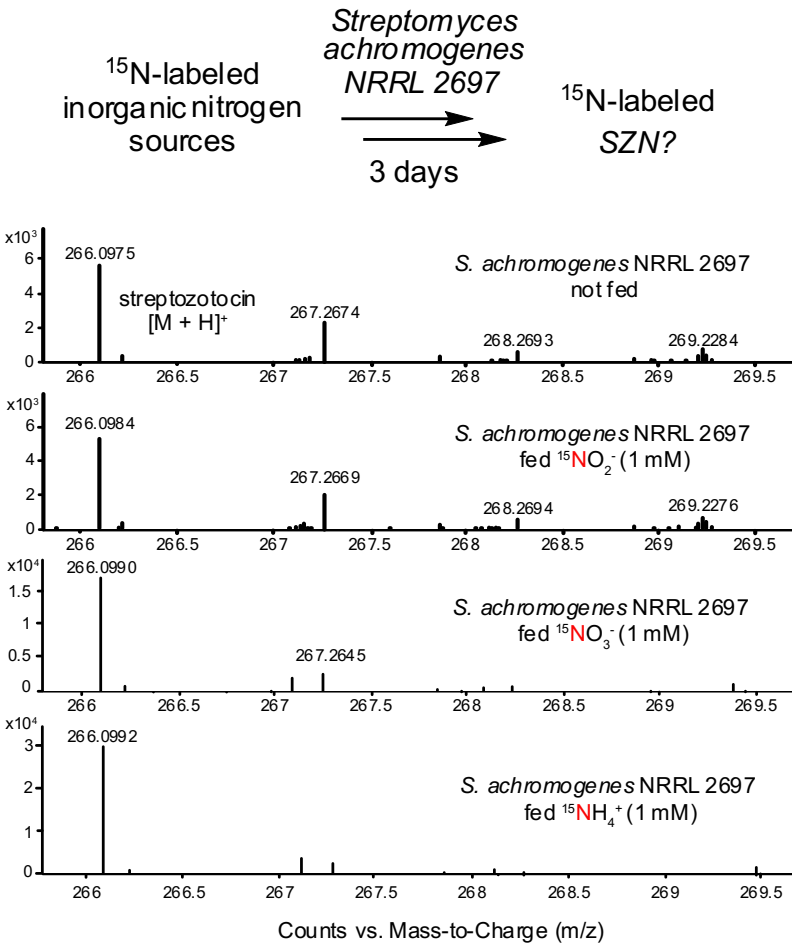


Figure 2.8: Mass spectra of culture extracts in which 1 mM of ^{15}N-nitrate, ^{15}N-nitrite, or ^{15}N-ammonium chloride were fed to *S. achromogenes var. streptozoticus* NRRL 2697. The expected masses ($[M+H]^+$) for SZN, ^{15}N-SZN, [$^{15}N_2$]-SZN, and [$^{15}N_3$]-SZN are 266.0983, 267.0953, 268.0923, and 269.0894, respectively.

Knowing that L-arginine plays an important role in SZN biosynthesis, we fed [$^{15}N_4{}^{13}C_6$]-L-arginine to *S. achromogenes* and detected the generation of [$^{15}N_2{}^{13}C$]-SZN (40%) and [$^{15}N^{13}C$]-SZN (7%) (**Figure 2.9**). Tandem mass spectrometry (MS/MS) of SZN can generate MS2 fragments with different numbers of nitrogen atoms, which allow us to determine which nitrogen atom of the labeled SZN is isotopically enriched (See caption in **Figure 2.9)**. Our result demonstrated both of the nitrogens in the *N*-nitroso group of SZN are labeled in this experiment. Specifically, the high abundance of [$^{15}N_2{}^{13}C$]-SZN relative to [$^{15}N^{13}C$]-SZN and the absence of

[^{15}N]-SZN suggests that the nitrogen and carbon atoms of the *N*-methyl-*N*-nitrosourea derive intact from two of the four nitrogens in L-arginine. We suspected that the two terminal nitrogen atoms in the guanidine group of L-arginine were responsible for the labeling as they are known precursors of reactive nitrogen species (e.g. nitric oxide). This proposal was tested by feeding [Guanido-$^{15}N_2$]-L-arginine, which resulted in the analogous major product [$^{15}N_2$]-SZN (42%) and minor product [^{15}N]-SZN (14%). This result confirms the *N*-nitroso nitrogens in SZN derive from the two terminal nitrogens of a single L-arginine guanidine group.

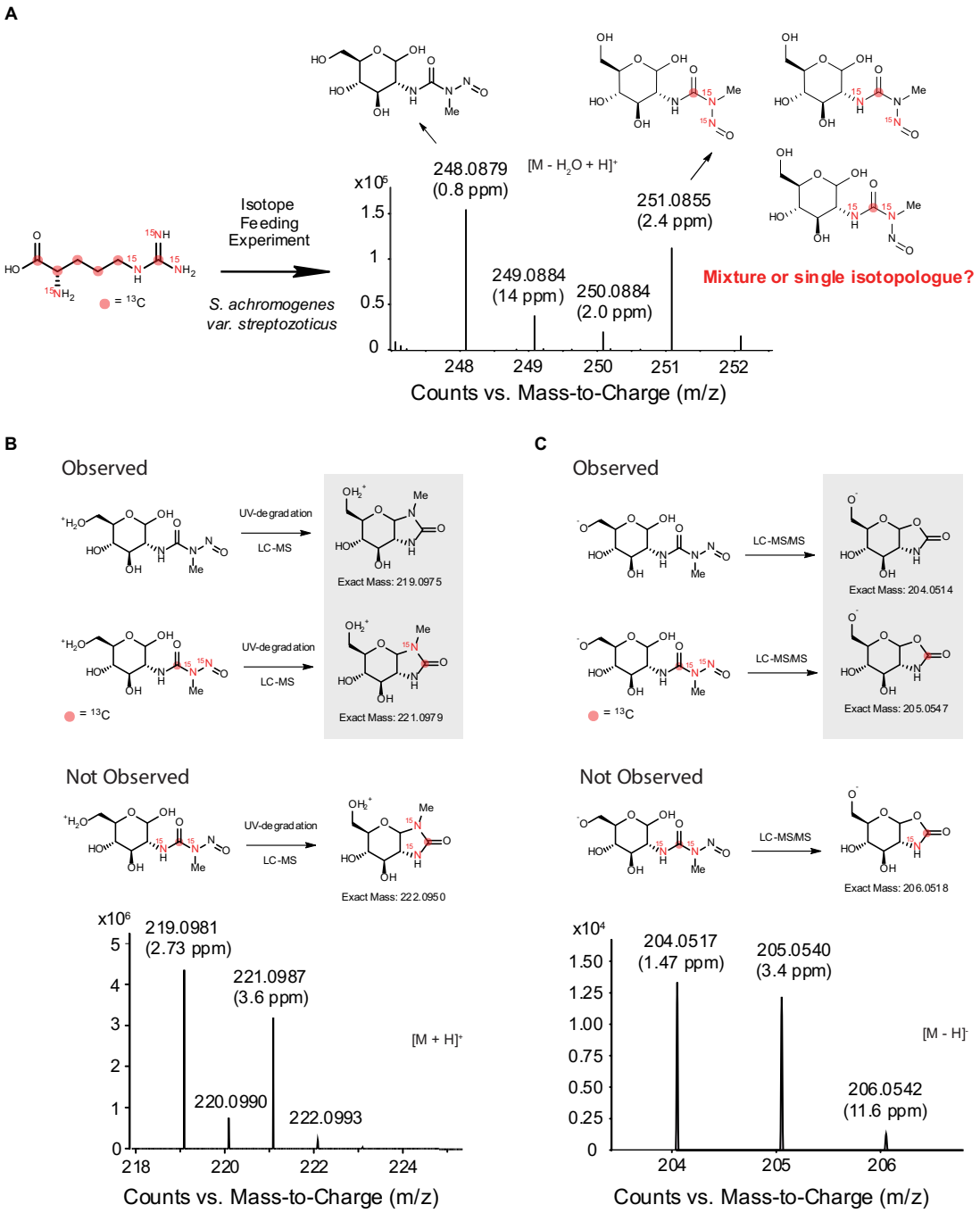


Figure 2.9: LC-MS/MS analysis of ^{15}N- and ^{13}C-labeled SZN. **A**) The mass spectrum of SZN $[M–H_2O+H]^+$ when 1 mM of $[^{15}N_4{}^{13}C_6]$-L-arginine was added to fermentation culture. To determine whether the labeled SZN was a single or a mixture of isotopologues, degradation (panel **B**) and MS/MS (panel **C**) experiments were performed. **B**) Exposing SZN to UV-light generated a one-carbon and one-nitrogen labeled cyclic urea previously reported to be a denitrosated SZN product, indicating that the distal nitroso nitrogen is labeled.[7] **C**) MS/MS fragmentation of SZN revealed a one-carbon labeled cyclic carbamate fragment, indicating that both the *N*-nitroso nitrogens are labeled.

With this information, we generated a biosynthetic hypothesis for SZN production that involves an oxidation and subsequent rearrangement of the guanidine group of L-arginine to yield the *N*-methyl-*N*-nitrosourea moiety (**Figure 2.10**). This reaction would require oxidation of the guandine group. Since SznF is annotated as a metal-binding enzyme, we hypothesized that it could oxidize L-NMA to form the *N*-nitrosourea-containing amino acid. In Chapter 3, we will discuss the predicted annotations of SznF in more detail and provide examples of reaction catalyzes by this enzyme family. Finally, this functional group could be transferred to D-glucosamine or an analogous intermediate by the putative C–N bond-forming enzymes SznH and/or SznK. The proposed intramolecular donation of a guanidine nitrogen atom to form an N–N bond with an amine is unprecedented in the literature and would provide a novel mechanism of biological *N*-nitrosation.

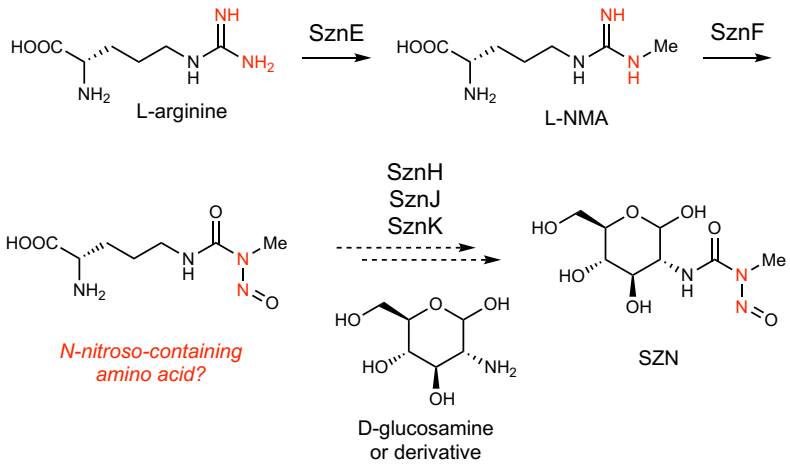


Figure 2.10: The biosynthetic pathway of SZN likely involves an *N*-nitrosourea containing amino acid intermediate.

2.3 Discussion

We identified a putative gene cluster responsible for the biosynthesis of SZN using comparative genomics and a resistance gene-guided genome mining strategy. In this Chapter, we also presented data linking these genes to the production of SZN using gene knockout methods.

Our feeding studies provided evidence that nitrite is not the distal nitrogen source of the *N*-nitroso group in this natural product. Notably, the result of feeding ^{15}N-nitrite contrasts sharply with the strong labeling (>75%) observed in studies of pathways that use nitrite for diazo biosynthesis.[13-14] This result could indicate that nitrite is not required as a co-substrate in the *N*-nitrosation reaction. Instead, we hypothesized the *N*-nitrosourea moiety of SZN arises from an oxidative rearrangement of a guanidine group. Specifically, we identified L-NMA as a key intermediate in the biosynthetic pathway, and SznF could perform the N–N bond-forming rearrangement reaction with this substrate. This intramolecular rearrangement reaction that results in N–N bond formation has not been previously reported. This logic also departs from the canonical nitrite-dependent *N*-nitrosation reactions. Further *in vitro* assays with purified SznF would be required to confirm this reactivity.

As discussed in the introduction, Prof. Katherine Ryan's group (University of British Columbia) has also discovered this gene cluster and observed many of the results presented in this Chapter.[2] The authors first used stable isotope feeding experiments to demonstrate that D-glucosamine is incorporated into streptozotocin. They then fed $^{15}N_2$-L-arginine and observed that both the *N*-nitroso nitrogens of streptozotocin are enriched and similarly concluded that L-arginine is important for building the *N*-nitrosourea group. Based on this information, they recognized that a gene encoding an arginine methyltransferase must be present in a gene cluster that encodes for streptozotocin production. They searched for predicted methyltransferase genes in the genome of the streptozotocin producer and identified a methyltransferase gene annotated as arginine methyltransferase (closest homolog: protein arginine methyltransferase from *Stigmatella aurantiaca* DW4/3-1). The authors then compared the genome of the producing strain to the nonproducing strain *Streptomyces achromogenes subsp. achromogenes* NRRL B-2120 and

identified the same putative *szn* (*stz*) cluster presented in this Chapter. Ryan and coworkers also included an extra gene, a predicted regulatory gene *stzM*, in their *szn* cluster based on the observation that the thirteen genes (*stzA-M*) are conserved across ten *Streptomyces* species. Congruent with our results, they were able to purify SznE and observed production of L-NMA when L-arginine is incubated with SznE and SAM. Their Δ*stzE* mutant also did not produce streptozotocin. In conclusion, two independent groups arrived at the same set of genes responsible for streptozotocin biosynthesis and presented the same evidence that L-NMA is an important biosynthetic intermediate.

In our process of discovering the *szn* biosynthetic gene cluster, we have leveraged our knowledge of the natural product's bioactivity and used a resistance-gene guided genome mining strategy to priories putative clusters for further study. This strategy is distinct from other genome mining strategies since it does not rely solely on searching for biosynthetic enzymes or redundant copies of target genes. In the future, it could be exciting to apply our resistance gene-directed method to uncover gene clusters encoding novel DNA-damaging natural products. In addition to DNA repair enzymes, resistance genes found in other DNA damaging natural product biosynthetic clusters (*e.g.* calicheamicin) could be employed in searching sequenced bacterial genomes.[15] One could search for homologs of these resistance genes in sequenced microbial genomes and examine the gene neighborhood for known biosynthetic genes (e.g. genes encoding NRPS). The resulting candidates could be further studied with *in vivo* heterologous expression to aid identification and isolation of bioactive natural products.

In this chapter we have identified L-NMA as an early biosynthetic precursor to SZN biosynthesis. We have also identified three genes, *sznF, H,* and *K*, as essential genes in SZN production. However, we cannot conclude that they are minimally required. Heterologous

expression of the *szn* gene cluster in a variety of bacterial hosts may confirm that this genetic locus encodes all the essential genes. However, if the *szn* gene cluster is insufficient for SZN production, additional examination of the host genome and/or transcriptomics analyses might be necessary to locate the missing biosynthetic enzymes.

2.4 Materials and Methods

Cultivation of *Streptomyces achromogenes* var. *streptozoticus* NRRL 2697 and detection of streptozotocin

Streptomyces achromogenes var. *streptozoticus* NRRL 2697 was obtained from the Agricultural Research Service (ARS) Culture Collection. The organism was grown on mannitol-soy (MS) agar plates (20 g/L soy flour, 20 g/L mannitol and 20 g/L Bacto agar (Difco)) at 30 °C for five days. The spores were scraped and inoculated into 5 ml of tryptic soy broth medium (Difco) and allowed to shake at 30 °C until saturation. 3 ml of this starter culture was inoculated into 100 ml production medium in a 250 ml baffled flask (40 g/L yellow cornmeal (Arrowhead Mills), 15 g/L potato starch (Bob's Red Mill), 1 g/L glucose, 3 g/L peptone (Bacto) and 14 g/L ammonium sulfate). The fermentation was carried out for 5 days at 30 °C with shaking at 200 r.p.m. 750 µl of the culture was centrifuged at 16,100*g* for 1 min, and an equivalent volume of acetone was added to the supernatant to further precipitate proteins and starches. After vortexing, the precipitate was removed by centrifugation at 16,100*g*, and the acetone was removed under vacuum. The samples were analyzed with liquid chromatography coupled with high-resolution mass spectrometry (LC–HRMS).

Streptozotocin was detected as the molecular ions $[M+H]^+$ or $[M–H_2O+H]^+$. Mass spectra data were collected only between 1.4 and 5 min, with the LC stream diverted to waste before 1.4

min and after 5 min. The LC column was an Acclaim Polar Advantage II C18 column (3 μm, 120 Å, 2.1 × 150 mm, Thermo Fisher Scientific). The flow rate was 0.3 ml/min. The LC conditions were: 95% solvent A, hold for 2 min; 95% to 50% solvent A in 6 min; 50% solvent A, hold for 1 min; 50% to 95% solvent A in 3 min; and 4 min equilibration at 95% solvent A (solvent A = 0.1% formic acid in water; solvent B = 0.1% formic acid in acetonitrile).

Genome sequencing of *Streptomyces achromogenes* var. *streptozoticus* NRRL 2697 and genomic DNA library construction

Genomic DNA was purified using an UltraClean Microbial DNA Isolation Kit (MoBio). Library construction from genomic DNA, sequencing, and assembly were performed by Cofactor Genomics. Next-generation sequencing used HiSeq Illumina reads of two short-insert paired-end libraries (300 bp insert and 500 bp insert) and a long-insert mate-pair library (1–2 kb insert). Assembly of the reads resulted in 8.6 Mb of non-redundant sequence distributed over 24 contigs. Annotations were carried out using PROKKA's database of other *Streptomyces* species. The assembled data were converted into a local BLAST database using Geneious Pro Version 7.1.6 (Biomatters). The *Streptomyces achromogenes* var. *streptozoticus* NRRL 2697 fosmid library was prepared using the CopyControl HTP Fosmid Library Production Kit (Epicentre) following the manufacturer's protocol. A library of 2,000 clones was picked into 96-well plates and stored at −80 °C as 50% glycerol stocks.

Feeding experiments with inorganic nitrogen salts

S. achromogenes var. *streptozoticus* NRRL 2697 was grown in 25 ml of fermentation medium in 250 ml baffled flasks while shaking at 30 °C as described above. After 16 h, stock solutions of [^{15}N]calcium nitrate, [^{15}N]sodium nitrite and [^{15}N]ammonium chloride (Cambridge

Isotope Laboratories) in water were sterilized by passing through a 0.22-µm filter membrane. Each of these nitrogen sources was added to the fermentation cultures to a final concentration of 1 mM. After four more days of fermentation at 30 °C, the presence of labeled and unlabeled SZN was determined by LC–HRMS after removing cell debris as described above. These feeding experiments were performed at least twice on different days and the same result was obtained.

Identification of the putative streptozotocin (*szn*) biosynthetic gene cluster

A BLAST search using *E. coli* DNA repair enzymes AlkA (annotation: DNA-3-methyladenine glycosylase 2, National Center for Biotechnology Information (NCBI) GenBank accession: AUG16819.1) revealed three homologues in the *S. achromogenes* var. *streptozoticus* genome. Comparison of the *S. achromogenes* var. *streptozoticus* and *S. achromogenes* var. *achromogenes* NRRL B-2120 (GenBank Assembly accession number: GCA_000720835.1) genomes was performed using Mauve alignment software.[16] The region encoding DNA repair enzyme homologues and putative biosynthetic enzymes that is absent from the non-producing strain is depicted in **2.3**

Homology modelling and sequence alignment of SznE

Structural homologues of SznE were identified using HHPred (http://toolkit.tuebingen.mpg.de/#/tools/hhpred). The sequences of the closest structural homologues were aligned with SznE using ClustalW. The SznE homology model was generated with MODELLER (Max Planck Institute for Developmental Biology) using the structure of TbPRMT7, a protein arginine methyltransferase that exclusively monomethylates arginine

residues, as a template (PDB: 4M37).[12] The resulting file was aligned with the TbPRMT7 PDB file using PyMOL Molecular Graphics System, Version 1.8 (Schrödinger).

Cloning, overexpression and purification of SznE

PCR reaction mixtures contained 22.75 µl of water, 0.125 µl of forward and reverse primers (sznE-F: TATCATATGCGGCACGTACAAGAGGCACG, sznE-R: ATAAAGCTTTCAGATGCTCACGCGGAGGG), 0.5 µl of genomic DNA, 1.5 µl of DMSO and 25 µl of Q5 High-Fidelity 2x Master Mix (New England Biolabs) in a final volume of 50 µl. Thermocycling was carried out in a MyCycler gradient cycler (Bio-Rad) using the following parameters: denaturation for 1 min at 98 °C; 35 cycles of 0.5 min at 98 °C, 0.5 min at 70 °C, 2 min at 72 °C; and a final extension time of 10 min at 72 °C. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide or SYBR safe staining, pooled, and purified. Amplified fragments were digested with Fastdigest NdeI and Fastdigest HindIII (Thermo Fisher Scientific) for 2 h at 37 °C. Digests contained 16.5 µl purified PCR product, 2 µl of FastDigest Buffer (10×), 0.4 µl of NdeI (20 U/µl) and 0.4 µl of HindIII (20 U/µl). Restriction digests for PCR were purified directly using Zymo DNA Clean & Concentrator. To prepare a linearized pET28a vector, 50 ng of the vector was digested with NdeI and HindIII as described with the exception of the addition of 1 µl calf-intestinal alkaline phosphatase in the reaction after 2 h and further incubation for 1 h. The linearized DNA was purified first by agarose gel electrophoresis and further purified with Zymoclean Gel DNA Recovery Kit. The insert digest was ligated into linearized expression vector pET28a using T4 DNA ligase (New England Biolabs). Ligations were run overnight at 16 °C and contained 1 µl T4 Ligase Buffer (10×), 1.5 µl digested vector, 5.5 µl digested insert DNA and 2 µl T4 DNA Ligase (400 U/µl). 5 µl of each ligation was used to transform a single tube of *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pET28a–SznE construct was confirmed

by sequencing of purified plasmid DNA. These constructs were used to transform chemically competent *E. coli* BL21 (DE3) cells (Invitrogen) and stored at –80 °C as frozen glycerol stocks.

A 50 ml starter culture of pET28a–SznE BL21 *E. coli* was inoculated from a single colony and grown overnight at 37 °C in LB medium supplemented with 50 μg/mL kanamycin. 7 ml of the saturated starter culture was inoculated into 700 ml of LB. The cultures were incubated at 37 °C with shaking for around 2.5 h, induced with 250 μM IPTG at an OD_{600} of around 0.5, and then incubated at 15 °C for 16 h. Cells were pelleted by centrifugation (6,700*g* for 10 min) and resuspended in 40 ml of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM $MgCl_2$, pH 8). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 10,000 psi, and the lysate was clarified by centrifugation (20,000*g* for 40 min). The supernatant was incubated with 2 ml of Ni-NTA resin for 1 h at 4 °C. The mixture was then loaded into a glass column and washed with 20 ml of wash buffer (50 mM HEPES, 20 mM imidazole, 10 mM $MgCl_2$, pH 8.0). The protein was eluted with approximately 20 ml of elution buffer (50 mM HEPES, 200 mM imidazole, 10 mM $MgCl_2$, pH 8.0). SDS–PAGE analysis (4–15% Tris-HCl gel) confirmed the presence and purity of SznE. The eluted protein was concentrated to around 1 ml using a Corning Spin-X UF 20 ml Centrifugal Concentrator (30,000 molecular weight cut-off membrane) after centrifugation at 1,145*g*. 12 ml of exchange buffer (20 mM HEPES, 50 mM NaCl, 10% glycerol, pH 8.0) was added, and the sample was concentrated again to 1 ml. This process was repeated once more before the concentrated, desalted solution containing purified SznE was frozen in liquid N_2 and stored at –80 °C.

^{1}H NMR assay of SznE activity

In a 500 µl reaction mixture, 50 mM potassium phosphate pH 8.0, 10 mM $MgCl_2$, 2 mM L-arginine, 4 mM SAM and 5 µM of SznE were mixed and incubated overnight at room temperature. Negative control assays omitting SAM, L-arginine or SznE in the reaction were also performed. The reaction mixtures were flash-frozen with liquid N_2 and lyophilized. The residues were resuspended in D_2O (Cambridge Isotope Laboratories) and analyzed with ^{1}H NMR spectroscopy using an Agilent DD2-600 NMR spectrometer (600 MHz). Chemical shifts (δ) are reported in parts per million (ppm) downfield from tetramethylsilane using the solvent resonance as an internal standard for ^{1}H (D_2O = 4.79 ppm). This experiment was performed in triplicate.

Feeding experiments with labeled arginine and degradation studies

Streptomyces achromogenes var. *streptozoticus* NRRL 2697 was grown in 10 ml of fermentation medium in 50 ml baffled flasks while shaking at 30 °C as described above. Stock solutions of [$^{15}N_4{}^{13}C_6$]-L-arginine, [$^{15}N_2$]-L-arginine or d_3-L-N^{ω}-methyl-L-arginine in water were sterilized by passing through a 0.22-µm filter membrane and added to the fermentation culture 16 h after inoculation to give a final concentration of 1 mM. After two more days of fermentation at 30 °C, the presence of labeled and unlabeled SZN was determined by LC–HRMS as described above. The positions of the labeled nitrogen and carbon atoms were elucidated using degradation experiments as described in the text above.

Gene disruption and chemical complementation experiments

Gene inactivation in *Streptomyces achromogenes* var. *streptozoticus* NRRL 2697 was performed according to standard protocols.[19] In brief, primers used to amplify *sznE* were used to screen the fosmid library to obtain a fosmid containing the *szn* gene cluster. The fosmid was

transformed into *E. coli* BW25113/pkD46 (Coli Genetic Stock Center Collection) by electroporation. The *aac(3)IV*-oriT cassette was amplified by PCR from pIJ773 using the primers listed in **Table 2.2**. Each of the biosynthetic genes was replaced with the *aac(3)IV*-oriT cassette PCR product using PCR targeting and λ-red-mediated recombination. The mutant fosmid was used to transform into *E. coli* WM6026 for conjugation with *S. achromogenes* var. *streptozoticus* NRRL 2697. Double crossover mutants were selected with apramycin resistance on mannitol-soy agar, and the exconjugants were grown in tryptic soy broth with apramycin added (50 μg/mL final concentration). After the seed culture reached saturation, the genomic DNA was isolated, and insertion of the resistance marker was confirmed by PCR.

Table 2.2: Primers used for generating knockout cassettes from pIJ773.

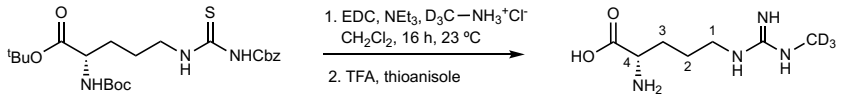
Oligo	***Nucleotide Sequence (5' to 3')***	***Targeted gene***
sznEKO-F	CGGCGCTGTTGCAGTCGCTCGCGCAGGAGATGAACGCGCATTCCGGG GATCCGTCGACC	*sznE*
sznEKO-R	TGCATCCAGTGCGAGCGGGGGTTGTCCGGCGGGTTGGACTGTAGGCT GGAGCTGCTTC	*sznE*
sznFKO-F	TCGACTGGCTGAACCACGACAACCCCTACCGCCGGCAGCATTCCGGG GATCCGTCGACC	*sznF*
sznFKO-R	CCCTCGCCGGCCTGGAGGGTCAGCGAGGAGCCCAGGCCG TGTAGGCTGGAGCTGCTTC	*sznF*
sznHKO-F	TCTGGCTGCTCCAGCCGGCACCGGTGACCTGGGAGGAGCATTCCGGG GATCCGTCGACC	*sznH*
sznHKO-R	ATGACCCGGGCGTACCGGGAGATGTAGCCGCGCGGCGGCTGTAGGCT GGAGCTGCTTC	*sznH*
sznJKO-F	TACGACTCGCTCACCGAGGTCCTGCCCATCCCCGTCCCCATTCCGGGG ATCCGTCGACC	*sznJ*
sznJKO-R	GCGCACCGGGGCGTGCACCATCCCCTCCTCGCGCACATATGTAGGCT GGAGCTGCTTC	*sznJ*
sznKKO-F	CGGGGCGTCCCGGTCGCGGCGGTGGTGAGCAGTCCGCGCATTCCGG GGATCCGTCGACC	*sznK*
sznKKO-R	GAAGTCCAGGCTGTCGACGAGGTCCTTGGTGACGGGGACTGTAGGCT GGAGCTGCTTC	*sznK*

For the chemical complementation experiments, wild-type and mutant strains were grown in 10 ml of fermentation medium in 50 ml baffled flasks while shaking at 30 °C as described above. Stock solutions of L-NMA in water were sterilized by passing through a 0.22-μm filter membrane

and added to the fermentation to a final concentration of 1 mM after 16 h. After two more days of fermentation at 30 °C, the presence of SZN was determined by LC–HRMS as described above.

For LC-HRMS detection of L-NMA from culture supernatant extracts, negative ion mode was used. The LC column was a Cogent Diamond Hydride column (4 µm, 100 Å, 3 × 150 mm, Microsolv Technology Corp.). The flow rate was 0.5 ml/min. The LC conditions were: 10% solvent A, hold for 1 min; 10 to 70% solvent A in 19 min; 70% solvent A, hold for 1 min; 70 to 10% solvent A in 4 min; and 4 min equilibration at 10% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). The extracted ion chromatograms were generated using the $[M-H]^-$ molecular ions with a 5 ppm window.

Synthesis of N^{ω}-d_3-methyl-L-arginine (d_3-methyl-L-NMA)

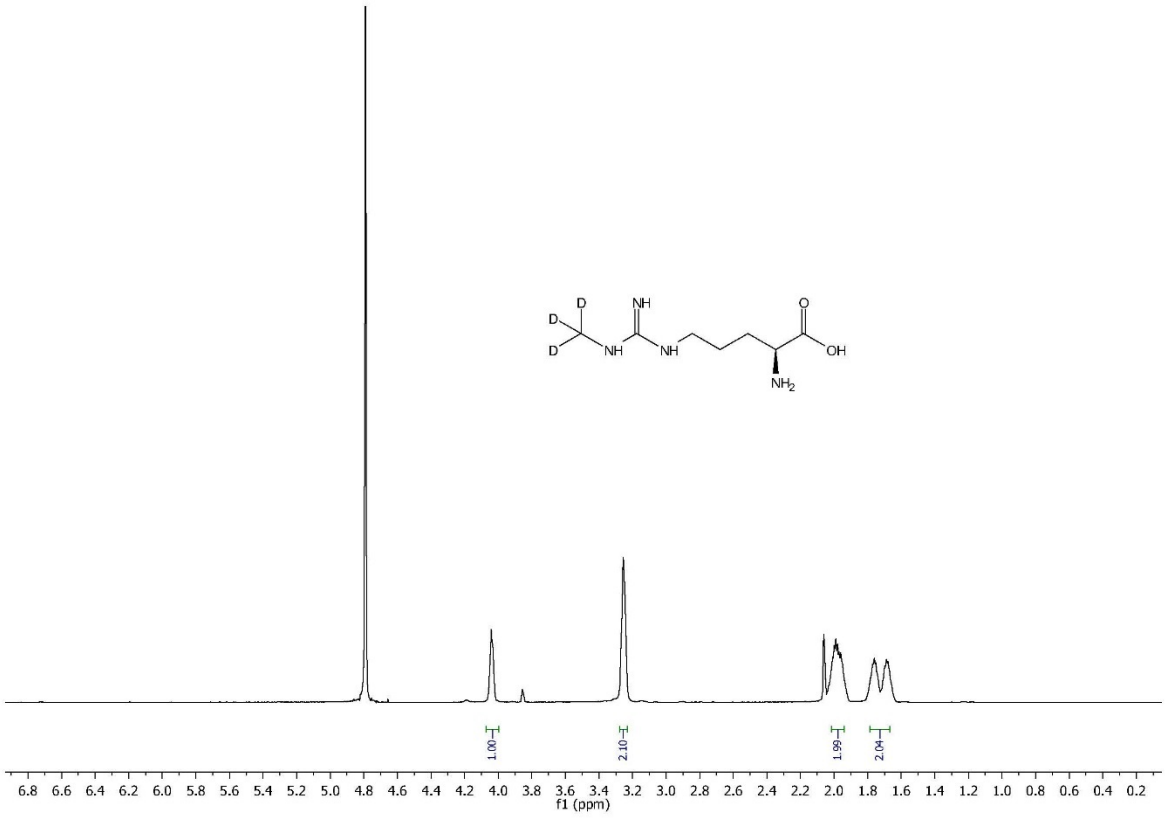


2.1 **d_3-L-NMA**

The carboxybenzyl (Cbz)-protected thiourea **2.1** was prepared according to Jansen Labby et al.[20] In a round bottom flask containing **2.1** (200 mg, 415 µmol, 1.00 equiv) dissolved in 20 mL of anhydrous dichloromethane (CH_2Cl_2) at 0 ºC was added 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC, 88.8 mg, 457 µmol, 1.10 equiv), triethylamine (NEt_3, 116 µL, 830 µmol, 2.00 equiv), and d_3-methylamine hydrochloride ($D_3C–NH_3^+Cl^-$, Cambridge Isotope Inc.; 30.0 mg, 415 µmol, 1.00 equiv) in one portion. After 16 h of stirring at room temperature, the reaction was washed with 1 x 10 mL of brine, and the collected organic layer was dried over anhydrous sodium sulfate (Na_2SO_4), filtered, and concentrated *in vacuo* to afford an oil. The crude oil was dissolved in 5 mL of trifluoroacetic acid (TFA) and 0.80 mL of thioanisole and stirred at room temperature for 16 h. The reaction was then concentrated *in vacuo* and 20 mL of

toluene was added. The solution was concentrated *in vacuo*, and the residue was dissolved in 5 mL of water. The aqueous solution was then washed with 2 x 5 mL of diethyl ether to remove nonpolar impurities. The crude d_3-methyl-L-NMA in the aqueous layer was purified by reverse-phase semi-preparative HPLC. The LC column was Kromasil 100 C18 (250 x 10 μm, 5mm). The flow rate was 3 mL/min. The LC conditions were: a linear gradient from 5 to 10% solvent B in 6 min; 10 to 30% solvent B in 1 min; an isocratic hold at 30% solvent B for 1 min; a linear gradient from 30 B to 5% solvent B in 2 min; and an isocratic hold at 5% solvent B for 2 min. (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). d_3-methyl-L-NMA was monitored at 200 nm and eluted between 4 – 4.5 min. The collected fractions were lyophilized to afford 34 mg of d_3-L-NMA as a white solid (45% yield). ^{1}H NMR of d_3-L-NMA (600 MHz D_2O): δ (ppm) = 4.07 – 4.01 (m, 1H, H_4), 3.28 – 3.23 (m, 2H, H_1), 2.05 – 1.92 (m, 2H, H_2), 1.79 – 1.67 (m, 2H, H_3). HRMS: Calc'd for $C_7H_{14}D_3N_4O_2$ $[M+H]^+$ = 192.1534, found = 192.1526.

Figure 2.11: NMR spectrum of d_3-L-NMA. ^{1}H NMR of d_3-methyl-L-NMA (recorded in D_2O at 600 MHz).



2.5 References

1) Ng, T.; Rohac, R.; Mitchell, A. J.; Boal, A. K.; Balskus, E. P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **2019**, *566*, 94.

2) He, H.; Henderson, A. C.; Du, Y.; Ryan, K. S. Two-Enzyme Pathway Links l-Arginine to Nitric Oxide in N-Nitroso Biosynthesis. *J. Am. Chem. Soc.* **2019**, *141*, 9.

3) Vavra, J. J.; Deboer, C.; Dietz, A.; Hanka, L. J.; Sokolski W. T. Streptozotocin, a new antibacterial antibiotic. *Antibiot. Annu.* **1959**, *7*, 230.

4) Pathak, S.; Dorfmueller, H. C.; Borodkin, V. S.; van Aalten D. M. Chemical dissection of the link between streptozotocin, O-GlcNAc, and pancreatic cell death. *Chem. Biol.* **2008**, *15*, 799.

5) Weiss, R. B.; Issell, B. F. The nitrosoureas: carmustine (BCNU) and lomustine (CCNU). *Cancer Treat. Rev.* **1982**, *9*, 313.

6) Krug, S.; Boch, M.; Daniel, H.; Nimphius, W.; Müller, D.; Michl, P.; Rinke, A.; Gress, T. M. Streptozocin-based chemotherapy in patients with advanced neuroendocrine neoplasms–predictive and prognostic markers for treatment stratification. *PLoS One* **2015**, *10*, e0143822.

7) Wu, K. K.; Huan, Y. Streptozotocin-induced diabetic models in mice and rats. *Curr. Protoc. Pharmacol* Chapter 5, Unit 5.47, **2008**.

8) Singaram, S.; Lawrence, R. S.; Hornemann, U. Studies on the biosynthesis of the antibiotic streptozotocin (streptozocin) by Streptomyces achromogenes var. streptozoticus. Feeding experiments with carbon-14 and tritium labelled precursors. *J. Antibiot. (Tokyo)* **1979**, *32*, 379.

9) Sankaranarayanan, R.; Cherney, M. M.; Cherney, L. T.; Garen, C. R.; Moradian, F.; James, M. N. The crystal structures of ornithine carbamoyltransferase from Mycobacterium tuberculosis and its ternary complex with carbamoyl phosphate and L-norvaline reveal the enzyme's catalytic mechanism. *J. Mol. Biol.* **2008**, *375*, 1052.

10) Fu, D.; Calvo J. A.; Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **2012,** *12*, 104.

11) Bergy, M. E. et al. Streptozotocin and its production. U.S. Patent 3027300A (**1962**).

12) Wang, C.; Zhu, Y.; Caceres, T. B.; Liu, L.; Peng, J.; Wang, J.; Chen, J.; Chen, X.; Zhang, Z.; Zuo, X.; Gong, Q.; Teng, M.; Hevel, J. M.; Wu, J.; Shi, Y. Structural determinants for the strict monomethylation activity by trypanosoma brucei protein arginine methyltransferase 7. *Structure* **2017**, *22*, 756.

13) Winter J. M.; Jansma, A. L.; Handel, T. M.; Moore, B. S. Formation of the pyridazine natural product azamerone by biosynthetic rearrangement of an aryl diazoketone. *Angew. Chem. Int. Ed.* **2009**, *48*, 767.

14) Wang, K. A.; Ng, T. L.; Wang, P.; Huang, Z.; Balskus, E. P.; van der Donk, W. A. Glutamic acid is a carrier for hydrazine during the biosyntheses of fosfazinomycin and kinamycin. *Nat. Commun*. **2018**, *9*, 3687.

15) Biggins, J. B.; Onwueme, K. C.; Thorson, J. S. Resistance to enediyne antitumor antibiotics by CalC self-sacrifice. *Science* **2003**, *301*, 1537.

16) Darling, A. C.; Mau, B.; Blattner, F. R.; Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. **2004**, *14*, 1394.

17) Soding, J.; Biegert, A.; Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. **2005**, *33*, W244.

18) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947.

19) Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA*. **2003***, 100*, 1541.

20) Jansen Labby, K.; Li, H.; Roman, L. J.; Martásek, P.; Poulos, T. L.; Silverman, R. B. Methylated Nω-Hydroxy-l-arginine analogues as mechanistic probes for the second step of the nitric oxide synthase-catalyzed reaction. *Biochemistry* **2013**, *52*, 3062.

Chapter 3. Discovery and biochemical characterization of the *N*-nitrosating metalloenzyme SznF

Part of this Chapter is adapted from previous publication.[1] The structural characterization and subsequent analyses were performed by Dr. Andrew J. Mitchell, Dr. Roman Rohac, and Professor Amie K. Boal (Penn State University). Part of the mechanistic proposal in Section **3.2.7** was contributed by Professor Amie K. Boal. The *in vitro* reaction that converts N^{ω}-monomethyl-L-arginine to form N^{ω}-hydroxyl-N^{ω}-monomethyl-L-arginine (**3.1**) and N^{ω}-hydroxyl-N^{ω}-monomethyl-L-citrulline (**3.4**) has been corroborated by an independent study performed by Professor Katherine Ryan's lab (University of British Columbia).[2] The study also performed ^{18}O-labeling experiments and arrived at the same conclusion as this thesis work.

3.1 Introduction

In Chapter 2, we described the discovery of the streptozotocin (*szn*) gene cluster and the identification of N^{ω}-L-monomethylarginine (L-NMA) as a key intermediate in SZN biosynthesis. From feeding studies, we formulated a hypothesis that the *N*-methyl-*N*-nitroso urea moiety of SZN results from an oxidative rearrangement reaction of L-NMA catalyzed by an iron-containing redox enzyme SznF (**Figure 2.10**). No enzymes have been characterized to date to perform this type of reaction. In Chapter 3, we will first describe our bioinformatics analysis of SznF, which will support SznF's role in binding a metal cofactor. The Chapter will then detail our biochemical characterization of SznF's activity *in vitro* and *in vivo*. Part of this work has been informed by a crystal structure of SznF, solved by Dr. Andrew J. Mitchell, Dr. Roman Rohac, and Prof. Amie Boal at Penn State University. Proposed mechanisms and future avenues for exploring SznF's reactivity will be presented in the end of the Chapter.

3.2 Experimental Results

3.2.1 Bioinformatic analysis of SznF

SznF is predicted to contain three domains – an N-terminal domain of unknown function, a central heme-oxygenase-like domain, and a C-terminal cupin domain (**Figure 3.1**).

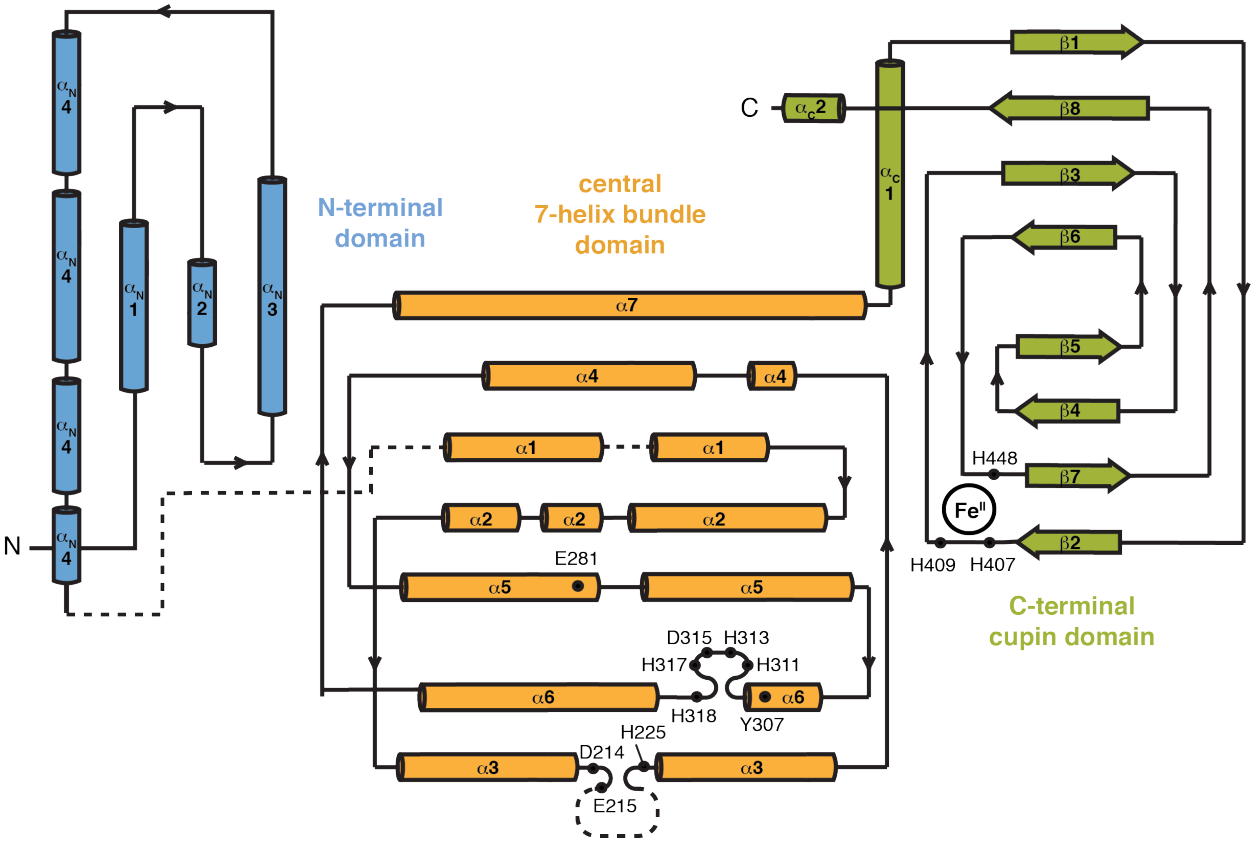


Figure 3.1: Topology diagram of SznF highlighting the secondary structures in the N-terminal domain (blue), the central heme-oxygenase-like domain (orange), and the C-terminal cupin domain (green). This figure was prepared by Prof. Amie K. Boal.

The central domain is part of a larger structural family that includes the *Pseudomonas aeruginosa* fatty acid decarboxylase UndA (**Figure 3.2A**), and a protein of unknown function from *Chlamydia trachomatis* termed Chlamydia protein associating with death domains (CADD). UndA catalyzes the oxidative decarboxylation of lauric acid to form 1-undecane using Fe^{II} and O_2 as cofactors.[3] X-ray structures of CADD and UndA contain di- and mononuclear iron centers,

respectively, but the structure and oxidation state of the active metallocofactor remain undefined until recent studies (see next paragraph).[4-5]

Since the beginning of this thesis work, several published studies have improved our understanding of the chemistry performed by enzymes with a HO-domain. Recent spectroscopic evidence using stopped-flow and Mossbauer spectroscopy of UndA assigned a diiron peroxo intermediate in the CADD domain.[6-8] At the time of this work, this domain was not known to catalyze *N*-oxidation. However, a very recent study demonstrated that a CADD domain-containing enzyme, RhoS, uses Fe^{II} to catalyze nitro formation in the biosynthesis of azomycin (**Figure 3.2 C**). Biochemical experiments revealed that RhoS requires a source of electrons to reduce the iron and activates O_2 to perform two successive oxygenation reactions on 2-aminoimidazole.[3] Another heme-oxygenase domain containing enzyme, BesC, from the biosynthetic pathway of the alkyne-containing amino acid β-ethynylserine, catalyzes an unusual oxidative cleavage of a carbon-nitrogen bond of 4-Cl-lysine to generate a terminal alkene product. (**Figure 3.2B**) Altogether, these reports demonstrate that this enzyme family can catalyze a diverse set of reactions.

To test if SznF also contains the metal binding residues of the HO-domain, we compared its sequence to the protein sequences of CADD and UndA. Our sequence alignment demonstrated that SznF conserves all of the metal-binding His, Asp, and Glu ligands present in CADD and UndA (**Figure 3.3**), supporting SznF's potential to bind metal cofactor(s) and perform oxidative chemistry. Based on our hypothesis, SznF could utilize a diiron cofactor to activate O_2 and perform the initial oxygenation of the guanidine group of L-NMA. However, the rearrangement reaction that results in N–N bond formation would be an unprecedented reaction catalyzed by this domain.

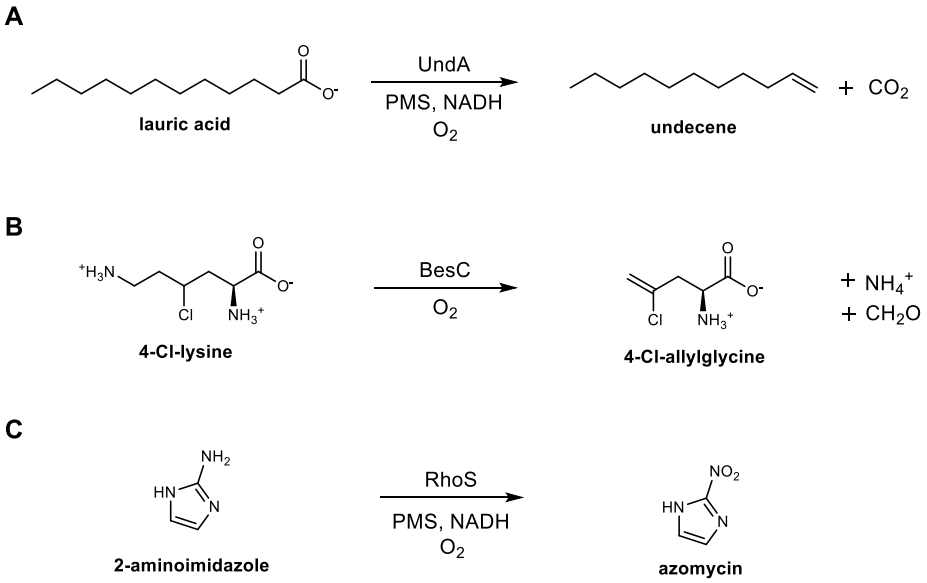


Figure 3.2: Reactions catalyzed by enzymes that contain the heme-oxygenase-like domain: **A**) UndA, **B**) BesC, and **C**) RhoS.

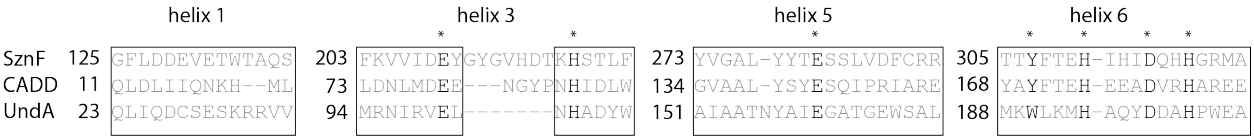


Figure 3.3: Sequence alignment of SznF with CADD and UndA reveals that the residues predicted to bind a ferric diiron cofactor are present. The predicted metal-binding residues are highlighted with asterisks.

The C-terminal domain of SznF is predicted to contain a cupin-like fold. The cupin superfamily is a diverse set of enzymes that also include cysteine dioxygenase, isomerases, and α-ketoglutarate-dependent dioxygenases (**Figure 3.4**).[9-12] The reactions this superfamily performs include C–H hydroxylation, sulfur oxidation, and isomerization reactions, some of which were described in **Chapter 1.4**. Crystal structures of cupin-containing enzymes have revealed that the metals zinc, nickel, and iron could bound.[13] However, the chemistry performed by this domain does not include N–N bond formation. Generally, our bioinformatic analyses support that both the central HO-domain and C-terminal cupin domain could bind iron. However, the functions of each

domain, if any, remain obscure without *in vitro* biochemical characterization. The roles of these two domains in *N*-oxygenation and N–N bond formation will be discussed in this Chapter.

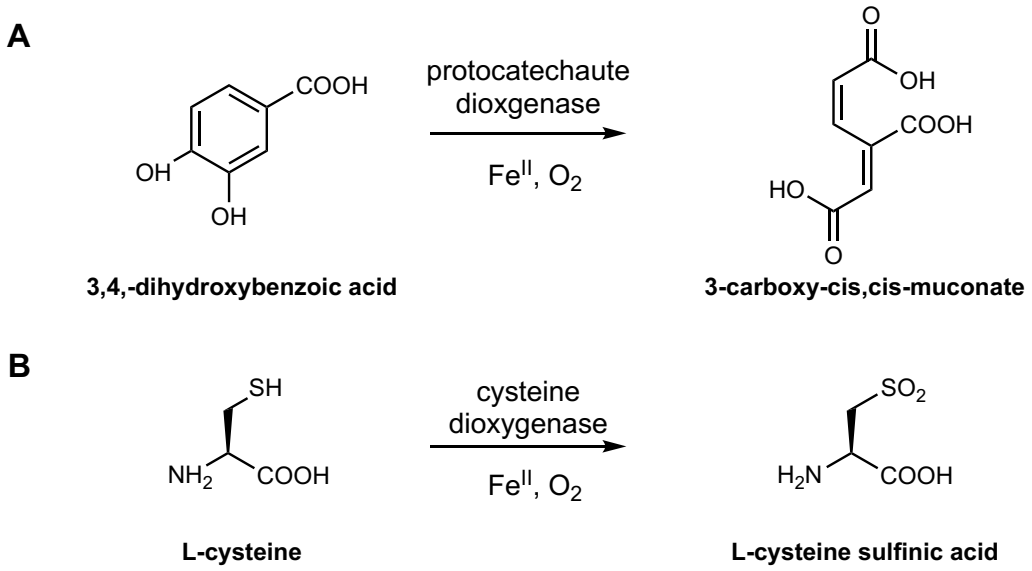


Figure 3.4: Examples of reactions catalyzed by enzymes that contain the cupin domain: **A**) protocatechuate dioxygenase and **B**) cysteine dioxygenase.

3.2.2. Biochemical characterizations of SznF

In Chapter 2, we described how genetic knockout of *sznF* abolished the production of SZN, confirming SznF is required for SZN biosynthesis. To begin characterizing the activity of SznF, we overexpressed *N*-His_{6x}-tagged SznF in *Escherichia coli* BL21(DE3) and purified the enzyme aerobically (**Figure 3.5A**). Purified contained around 0.9 Fe per monomer based on the Ferene-S assay. Because iron-dependent oxygenases typically require initial reduction to the Fe^{II} state to bind and activate dioxygen, we incubated purified SznF in air with L-NMA, phenazine methosulfate (PMS), and NADH as a reductant. We observed SznF-dependent production of nitrite using the Griess assay, indicating production of nitrite (**Figure 3.5B**).

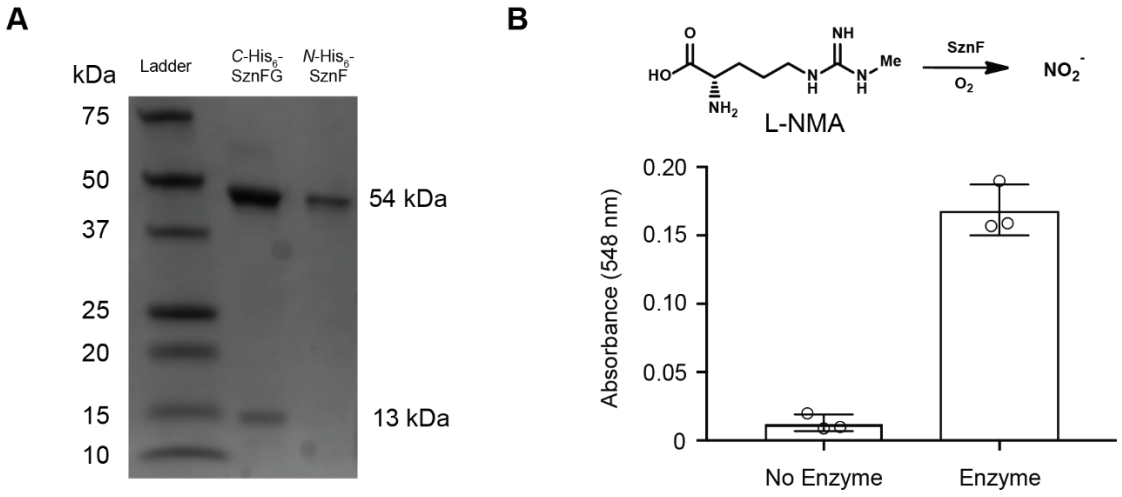


Figure 3.5: SznF catalyzes *N*-oxidation. **A**) SDS-PAGE of purified SznF and SznFG. The molecular weights of SznF and SznG are 54 kDa and 13 kDa, respectively. Ladder = Precision Plus Protein All Blue Standards (BioRad). **B**) Nitrite was detected when Fe^{II}-SznF and L-NMA were incubated together. Nitrite was detected with the Griess reagent, and absorbance was measured at 548 nm.[14] Error bars represent standard deviation of the mean of three replicates.

To determine whether nitrite could be derived from a more reactive nitrogen species nitric oxide (NO), we repeated the enzymatic reaction in the presence of *N*-methyl-D-glucamine dithiocarabamate (MGD) and iron, which forms an $Fe(MGD)_2$ complex that coordinates nitric oxide to form an EPR-active complex.[15] Indeed, we observed production of a nitric oxide radical by electron paramagnetic resonance (EPR) spectroscopy (**Figure 3.6**). To confirm that the nitric oxide is derived from the terminal nitrogen atoms of the guanidine ring, we used isotopically-labeled $^{15}N_2$-L-arginine and generated $^{15}N_2$-L-NMA *in situ* using SznE and SAM. Incubating this labeled substrate with SznF and trapping the resulting NO resulted in hyperfine splitting in the EPR spectrum suggesting that ^{15}NO was generated. These results confirm that SznF catalyzes *N*-oxidation of the guanidine group of L-NMA.

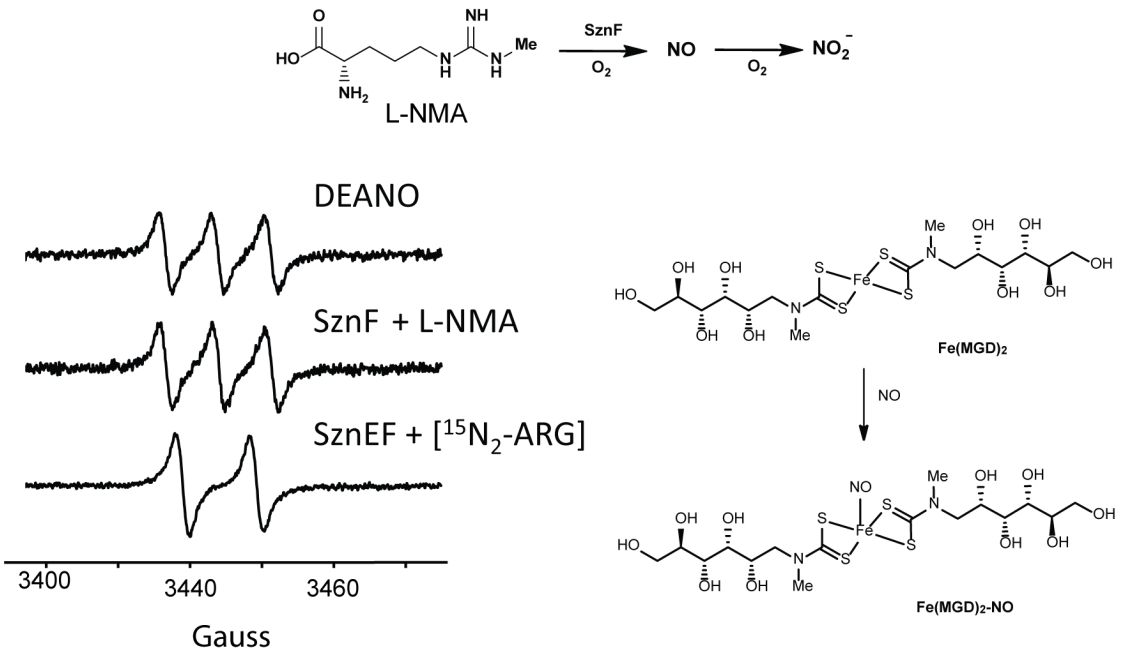


Figure 3.6: NO is generated from the SznF-catalyzyed N-oxidation of L-NMA. NO was trapped with Fe^{II}-*N*-methyl-D-glucamine dithiocarbamate (MGD) and analyzed by electron paramagnetic resonance (EPR) spectroscopy at room temperature.[15] Sodium 2-(*N*, *N*-diethylamino)-diazenolate-2-oxide (DEANO) was used as a positive control for NO detection. Assays using [guanidino-$^{15}N_2$]-L-NMA as a substrate revealed changes in hyperfine splitting by EPR spectroscopy, indicating that NO is derived from the terminal guanidine group of L-NMA.

The NO observed in the SznF-mediated reaction could be derived from the degradation of an *N*-nitrosourea reaction product or may be formed as part of the N–N bond forming step. To identify additional amino acid product(s) generated by SznF, we analyzed assay mixtures LC-MS and characterized four new compounds by comparison to synthetic standards (**Figure 3.7** and **Figure 3.8**). We observed that SznF first hydroxylates L-NMA at the N^{δ}-position to afford N^{δ}-hydroxy-N^{ω}-methyl-L-arginine (**3.1**). **3.1** is further hydroxylated to give N^{δ}-hydroxy-N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (**3.2**). Subsequently, an oxidative rearrangement converts **3.2** to N^{δ}-hydroxy-N^{ω}-methyl-N^{ω}-nitroso-L-citrulline (**3.3**). This proposed biosynthetic intermediate degrades in a non-enzymatic manner to the denitrosated product **3.4** as detected by LC-MS.

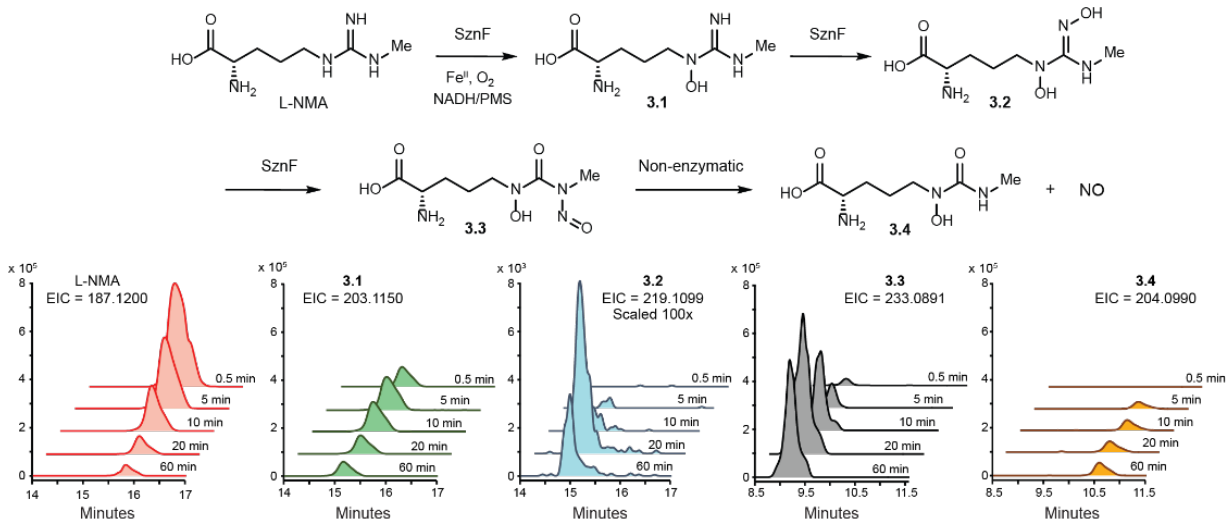


Figure 3.7: LC-MS time course of the SznF-mediated transformation of L-NMA to *N*-nitrosourea **3.3**. EICs are shown, and the chromatograms were generated with a 5 ppm window. The structures of **3.1**–**3.4** were confirmed by comparison to synthetic standards (**Figure 3.8**).

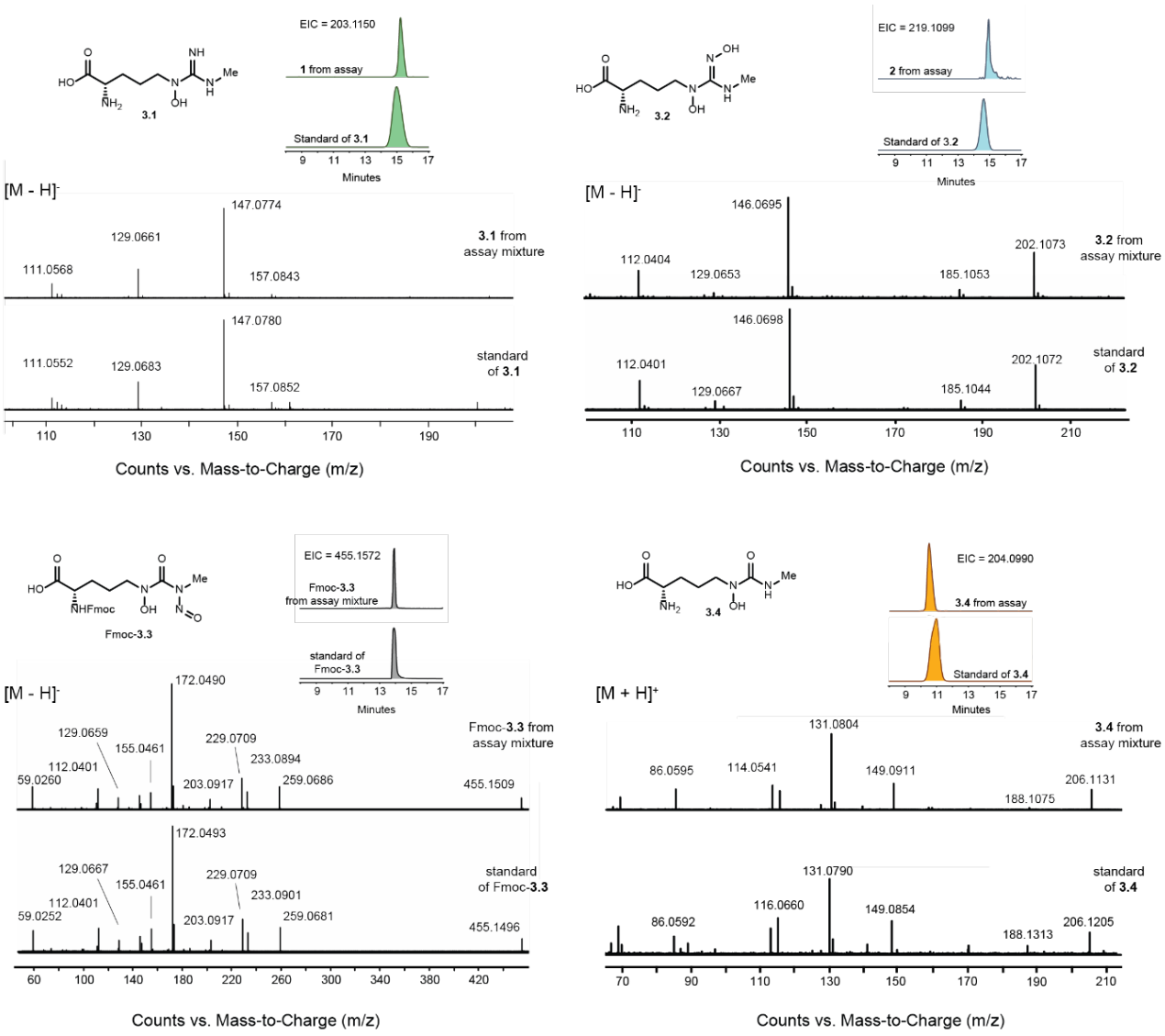


Figure 3.8: Comparison of retention times and MS/MS fragmentation patterns of **3.1**, **3.2**, Fmoc-**3.3**, and **3.4** generated in SznF assay mixtures with the corresponding synthetic standards. EIC traces were generated with a 5 ppm window. NMR characterization and synthetic procedures are reported in the methods section of this Chapter.

Aspects of the SznF-mediated reaction resembles nitric oxide synthase, a heme-dependent enzyme that oxidizes L-arginine to form nitric oxide and L-citrulline (**Figure 3.9**).[16] However, there is no evidence of heme-binding in SznF, as there is no heme absorption observed by UV-vis spectroscopy. Furthermore, L-arginine was not accepted as a substrate. When 1 mM L-arginine, 80 μM SznF, 20 μM phenazine methosulfate (PMS), and 5 mM NADH were incubated at room

temperature for one hour, only trace amounts of a mass corresponding to L-hydroxyarginine (EIC ($[M–H]^-$) = 189.0993) was observed. No masses corresponding to L-hydroxycitrulline (EIC ($[M–H]^-$) = 190.0883), L-dihydroxyarginine (EIC ($[M–H]^-$) = 205.0942), or the L-nitrosocitrulline (EIC($[M–H]^-$) = 203.0786) were observed (**Figure 3.10**). The N^{δ}-hydroxylation in this reaction sequence was unanticipated, as this nitrogen atom is not incorporated into SZN. The role of this hydroxyl group may be involved in substrate binding as discussed in the following section.

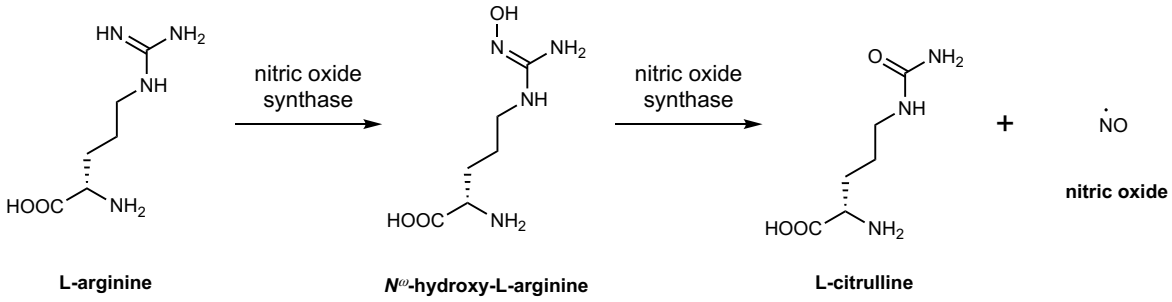


Figure 3.9: Nitric oxide synthase (NOS) catalyzes the oxidation of L-arginine to generate L-citrulline and nitric oxide (NO).

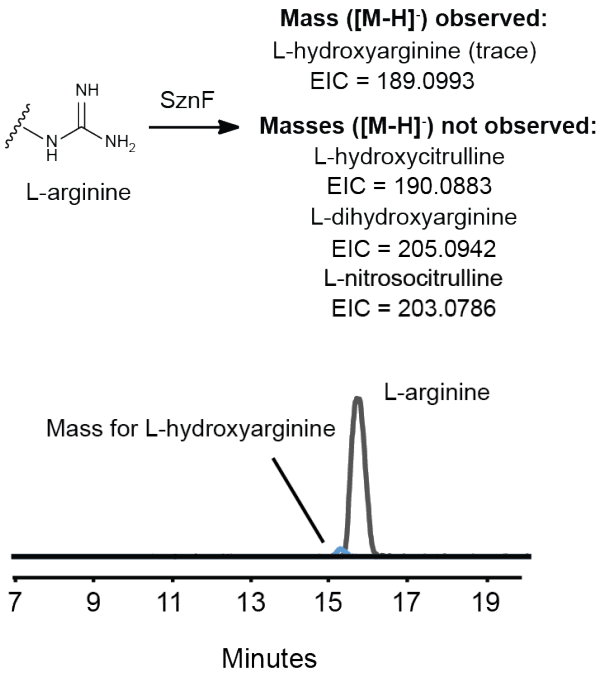


Figure 3.10: L-arginine is not processed by SznF to generate oxidized guanidine or citrulline products. This result suggested that the methyl group is important for substrate recognition and catalysis by SznF. EIC traces were generated with a 5 ppm window.

We further tested the requirements of SznF's activity. Incubating SznF with ethylenediamine tetraacetic acid (EDTA), a metal chelator, abolished activity, indicating a metal cofactor is essential (**Figure 3.11A**). We also overexpressed and purified *apo*-SznF from *E. coli* and observed this enzyme preparation has no activity. Mixing *apo*-SznF with Fe restores activity (**Figure 3.11B**), while other divalent metals did not result in active enzyme. This result suggested that Fe is a necessary and sufficient cofactor for converting L-NMA to **3.3**. SznF also requires oxygen and external reductants to perform this reaction (**Figure 3.11A**).

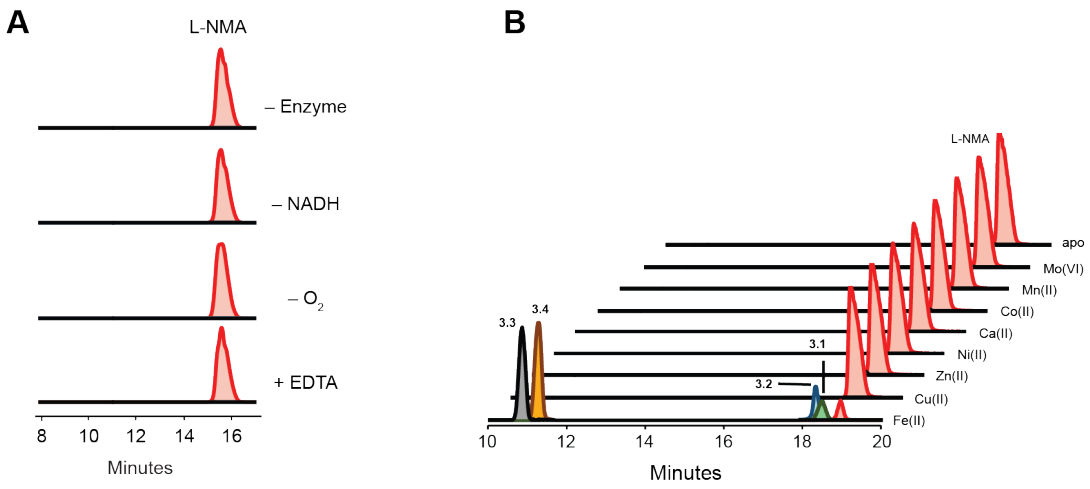


Figure 3.11: Control assays identify the cofactors required for SznF's activity. **A**) LC-MS traces of SznF *in vitro* assays defining the requirements for activity. EICs are shown. **B**) Testing the metal dependence of SznF. 80 µM of apo-SznF was incubated with 200 µM of various divalent metals, 20 µM PMS, 1 mM L-NMA, and 5 mM NADH for one hour at room temperature. The EIC traces were generated with a 5 ppm window.

We confirmed that **3.3** is derived from an intact guanidine group using a crossover experiment. We observed no scrambling of the product nitrogen atoms in assays containing both unlabeled and ^{15}N-labeled L-NMA (**Figure 3.12**) This supports the results obtained in the *in vivo* feeding experiments presented in Chapter 2 which indicated that the carbon and nitrogen atoms of the guanidine group of L-NMA are transferred as an intact unit into the *N*-nitrosourea group of SZN.

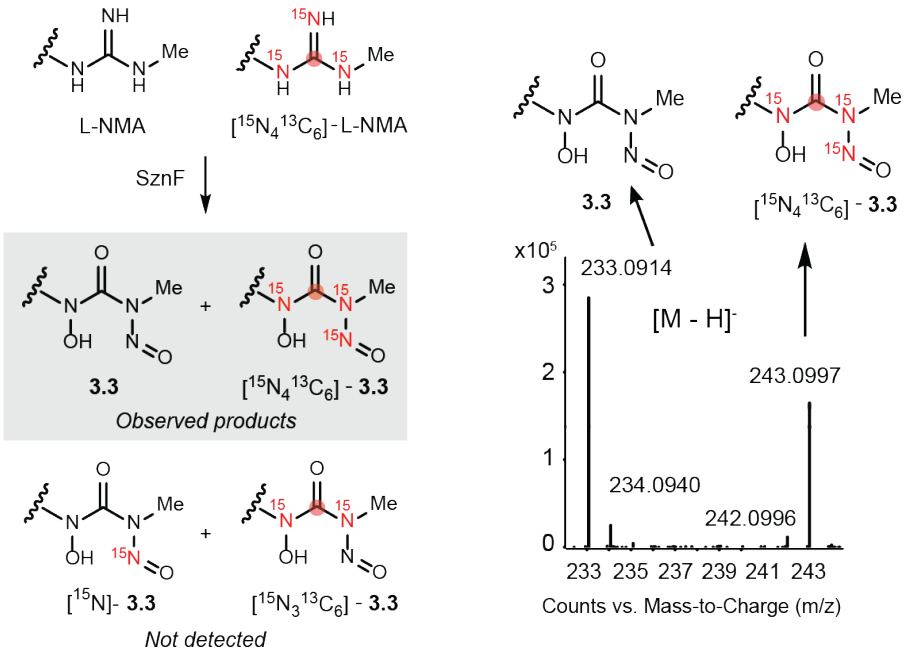


Figure 3.12: Crossover experiment showed that SznF catalyzes rearrangement of an intact guanidine group. Mass spectrum of **3.3** is shown.

We also co-expressed SznF with Rieske ferredoxin SznG. Ferrodoxins utilize iron-sulfur clusters to shuttle electrons to and from metal-binding sites in a metalloenzyme. SznG may complex with SznF and improve assembly of the metal cofactors, which could lead to better activity. Although these proteins co-purified, SznG did not improve product conversion *in vitro*. Together, these data showed that SznF catalyses the oxidative rearrangement of a guanidine group to an *N*-nitrosourea in a reaction dependent on both non-heme iron and oxygen.

3.2.3 The role of oxygen in the SznF-mediated reaction

Next, we further defined the role of oxygen in this transformation. The addition of L-NMA to SznF assay mixtures qualitatively accelerated the consumption of O_2 (**Figure 3.13A**) as assessed by an oxygen-sensing optode. This confirms that O_2 is used as the co-substrate. All three *N*-nitrosourea oxygen atoms in **3.3** were labeled when assays were performed in $^{18}O_2$ (**Figure**

3.13BC), whereas reactions performed in $H_2{}^{18}O$ did not yield any enrichment (**Figure 3.13D**). Therefore, the ureido oxygen atom does not arise from hydrolytic chemistry. This isotope tracer experiment also indicated that there is no scrambling of the oxygenating species invoked by SznF with bulk solvent.

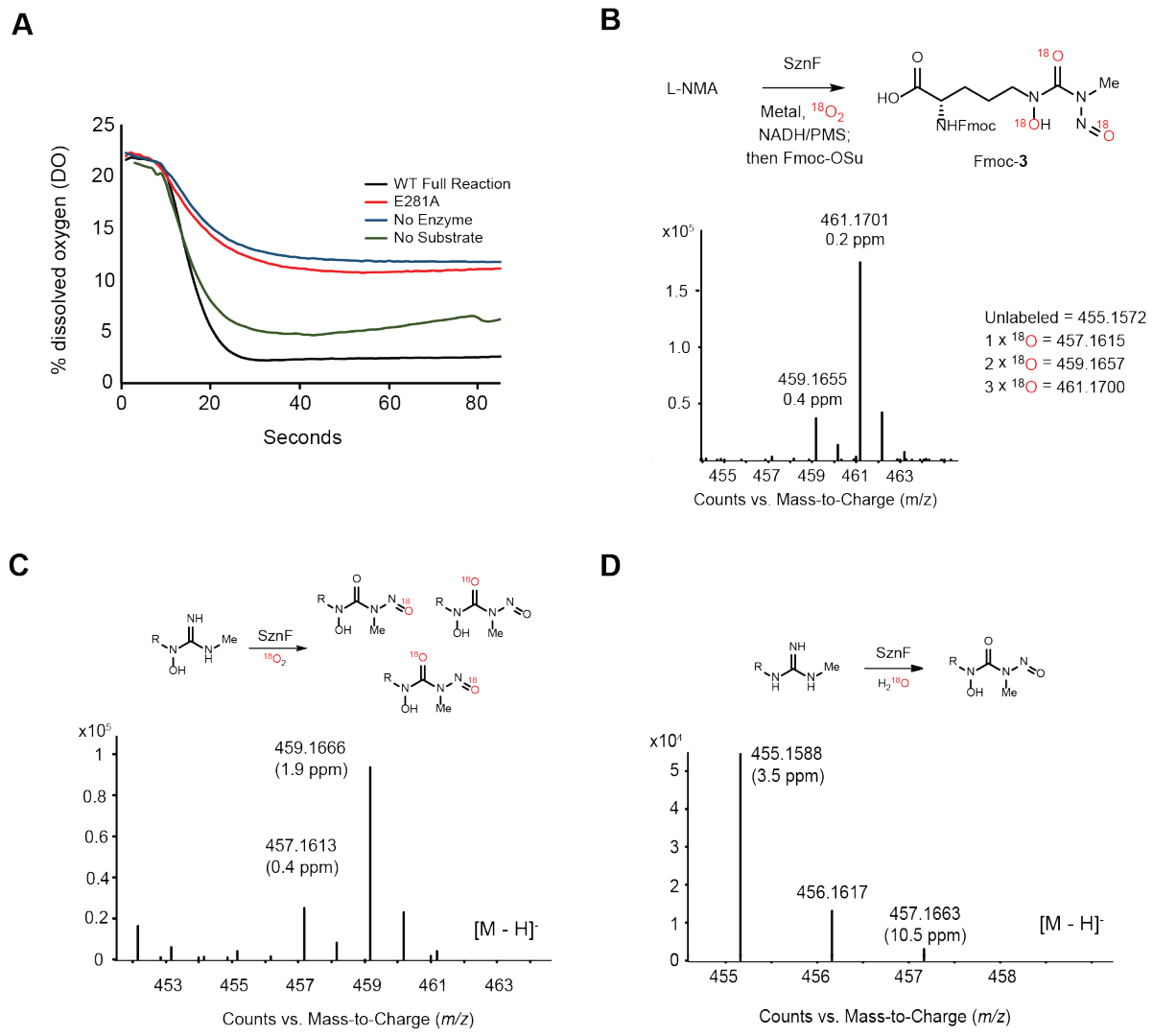


Figure 3.13: SznF uses O_2 as a cofactor. **A**) O_2 consumption by SznF in the presence of L-NMA, PMS, and NADH. **B)** Mass spectrum of Fmoc-**3.3** when L-NMA and SznF is incubated with PMS, NADH, and $^{18}O_2$. **C**) Same as panel **B** but **3.1** is used as substrate. **D**) Mass spectrum of Fmoc-**3.3** when L-NMA, PMS, NADH, and $H_2{}^{18}O$ are used.

Finally, the addition of catalase or superoxide dismutase did not affect SznF activity (**Figure 3.14**). This precluded the possibility that SznF is a peroxidase that uses hydrogen peroxide as an oxidant for *N*-oxygenation reactions. These experiments establish that SznF is a monooxygenase. The two new oxygen atoms in **3.2** arise from sequential *N*-hydroxylations, each involving one molecule of O_2, with a third equivalent of O_2 providing the ureido oxygen in **3.3**.

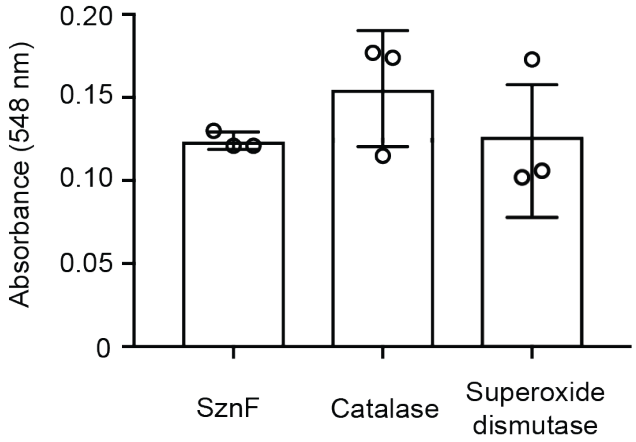


Figure 3.14: Addition of one unit of catalase or superoxide dismutase to the assay mixtures does not affect SznF-catalyzed *N*-oxygenation as measured by the Griess assay. Error bars represent standard deviation of the mean of three replicates.

3.2.4 Connecting SznF reactivity and streptozotocin biosynthesis

To connect the *in vitro* activity of SznF to SZN production *in vivo*, we fed monohydroxylated, SznF-derived intermediate **3.1** to Δ*sznE* and Δ*sznF* mutant strains. Feeding **3.1** to the Δ*sznE* strain restored SZN production, confirming that this compound is an on-pathway intermediate (**Figure 3.14A**). However, SZN was not detected when **3.1** was fed to the Δ*sznF* mutant, which indicates that additional processing of this intermediate by SznF is required. We currently cannot synthesize or isolate **3.3** since the compounds is too unstable. We have also reexamined the mutant strains that lacked predicted downstream biosynthetic enzymes (Δ*sznH*,

Δ*sznJ* and Δ*sznK*). Although the unstable *N*-nitrosourea **3.3** did not accumulate in these strains, we observed increased levels of decomposition product **3.4** in extracts from mutants compared to the wild type (**Figure 3.15**). Together, these results implicate the activity of SznF in the biosynthesis of SZN, and suggest that SznH, SznJ and SznK may transfer the *N*-nitrosourea group of **3.3** to the D-glucosamine scaffold (**Figure 3.16**) The biochemical characterization of SznH and SznK will be presented in Chapter 4.

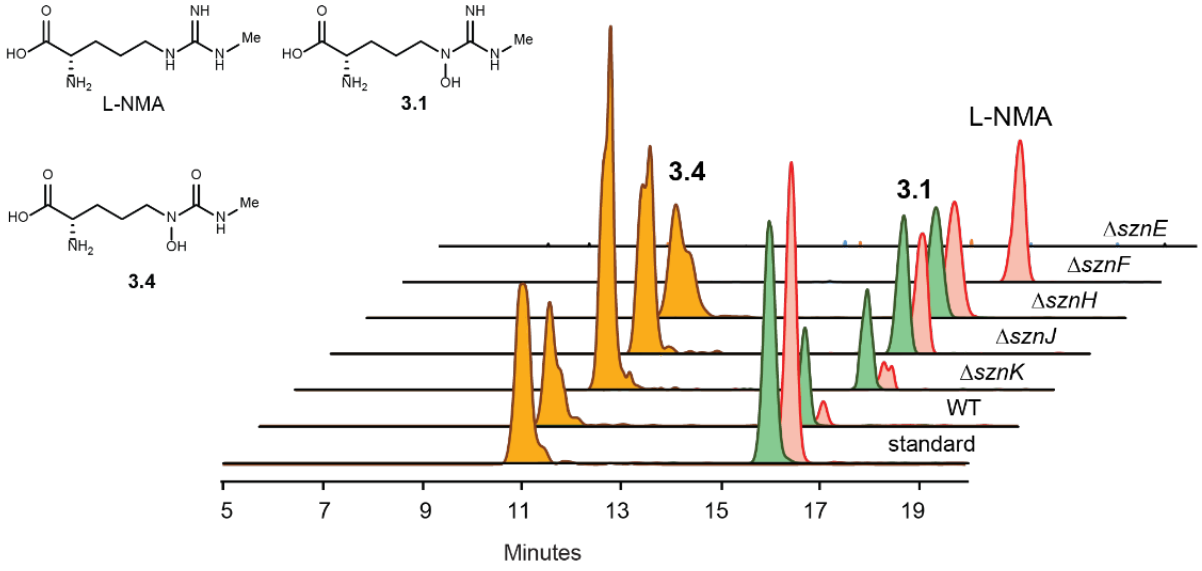


Figure 3.15: LC-MS analysis of culture supernatant extracts from each mutant. EICs of L-NMA and SznF-generated intermediates and products are shown. This result supports the hypothesis that **3.3** is likely processed by SznH and SznK as the Δ*sznH* and Δ*sznK* slightly accumulated **3.4**.

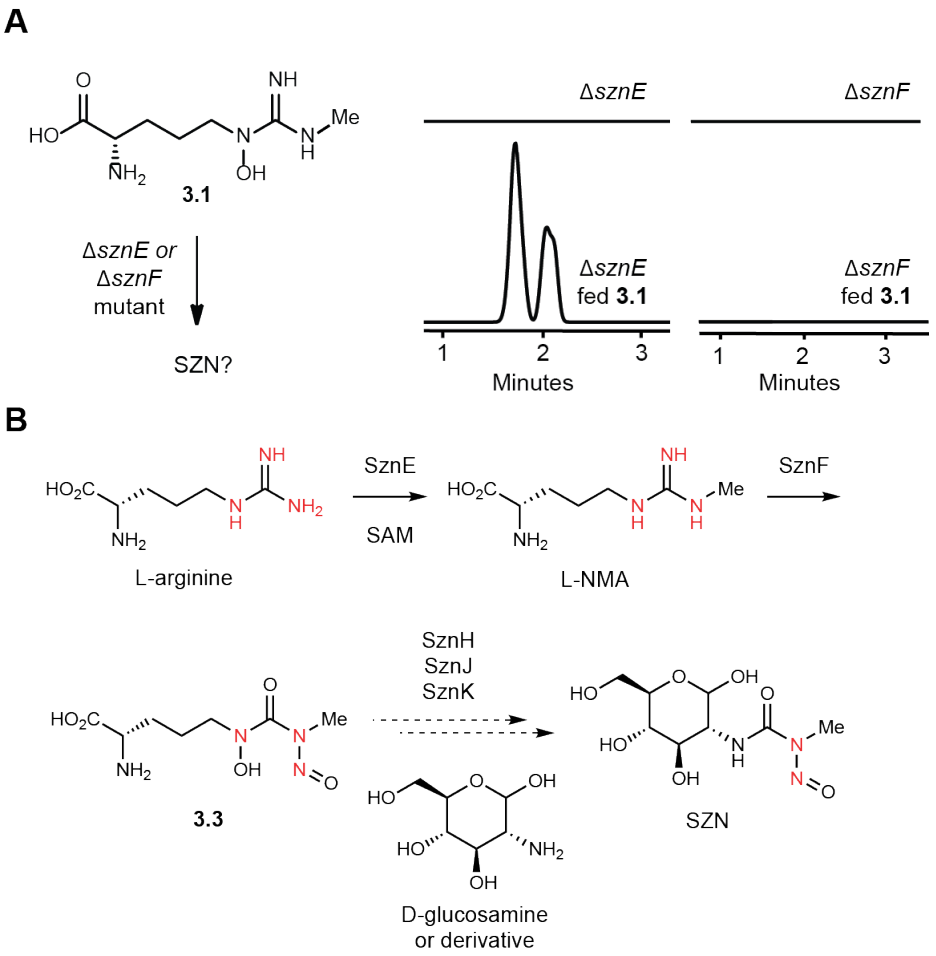


Figure 3.16: The reaction catalyzed by SznF is essential for SZN production *in vivo*. **A**) SznF-derived intermediate **3.1** complements the Δ*sznE* mutant but not the Δ*sznF* mutant. EICs for SZN ($[M–H_2O+H]^+$ = 248.0883) are shown. **B**) Proposed pathway for SZN biosynthesis.

3.2.5 Structural characterization of SznF

To gain further insight into the reactivity of SznF, we characterized its crystal structure with and without **3.1** added. The detailed methods for solving the x-ray crystal structure, and in-depth analysis of the SznF structure can be found in Chapter 7 of Reference 17. In brief, a 2.08-Å-resolution dataset revealed SznF to be a dimer with three distinct domains (**Figure 3.17A**). N-terminal helical motifs are flanked by C-terminal β-barrel cupin domains, each of which features a 3-His Fe-binding site (**Figure 3.17B**). The central domain is a seven-helix bundle, devoid of full

occupancy metal ions in these crystals, but similar to dinuclear metalloproteins related to the HO-fold. (**Figure 3.17C**).[18] SznF contains a large cavity in the middle of its central helical bundle domain (orange). As predicted, the central domain of SznF is similar in topology to HO and resembles closely to the CADD superfamily.[5]

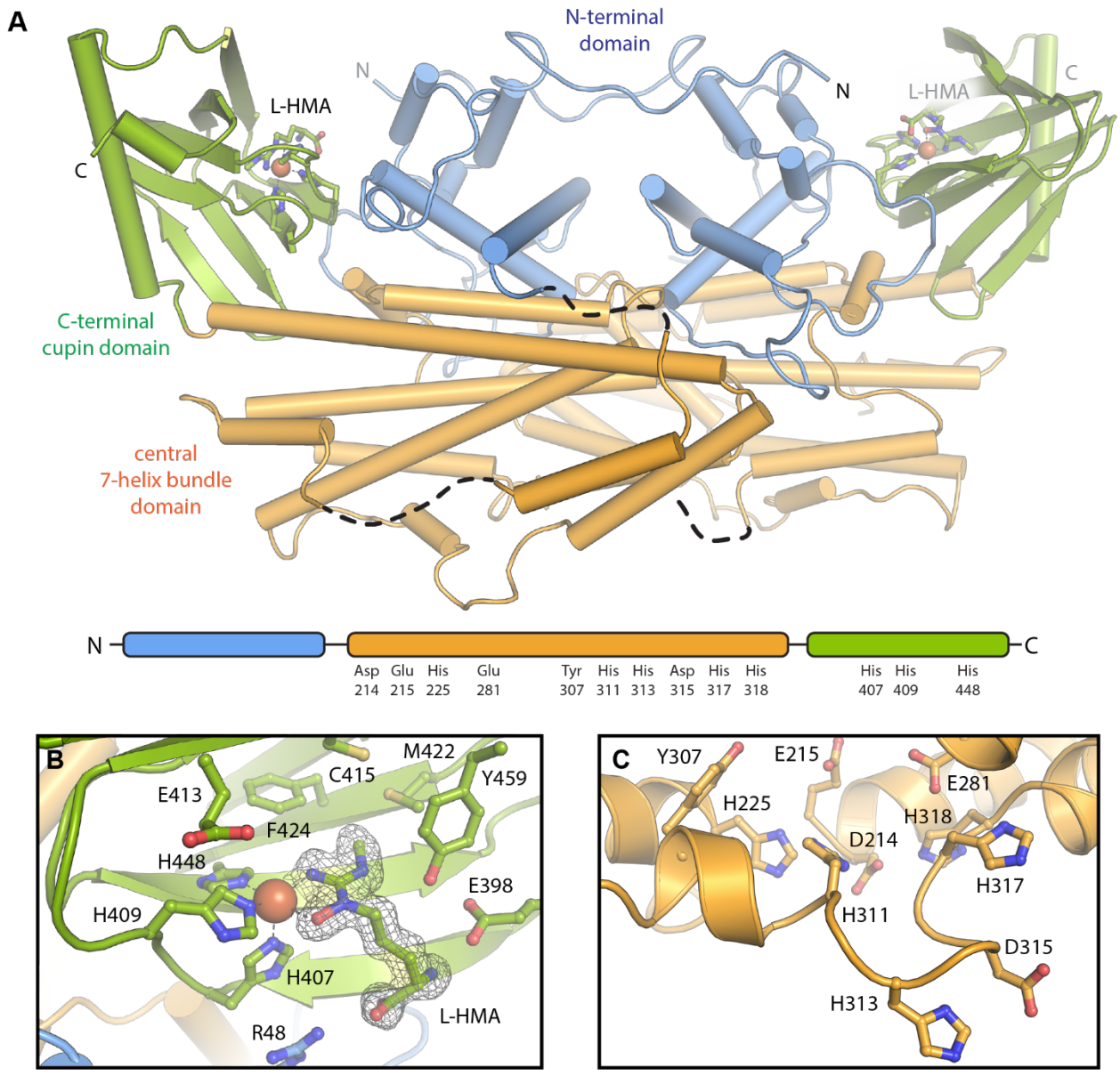


Figure 3.17: The x-ray crystal structure of SznF with **3.1** bound. A) The overall dimeric structure of SznF. B) The mononuclear His-coordinated Fe^{II} site (orange sphere) in the cupin domain. This structure also contains the intermediate L-HMA (**3.1**) (green sticks) as a bidentate ligand. **C**) The central domain of SznF, which lacks full occupancy metal ions in the crystals but conserves a series of His/carboxylate ligands found in related diiron proteins. This figure was prepared by Prof. Amie K. Boal, Dr. Andrew J. Mitchell, and Dr. Roman Rohac.

Most of the secondary structures in this central domain contain loop disruptions and disordered regions, suggesting considerable refolding upon binding/release of the L-NMA

substrate and/or assembly of the iron-based cofactor. Not unique to SznF, these heme-oxygen like domain containing proteins share a propensity for distorted secondary structure motifs that perhaps enable complex formation with large and polar substrates for oxidative transformations. One such enzyme is the aforementioned fatty acid decarboxylase UndA (Chapter 3.1). The published crystal structure of UndA contains only a single iron ion. A mechanism was initially proposed utilizing a mononuclear cofactor, and to date the dinuclear form of UndA has remained refractory to crystallographic characterization due to a propensity for disorder in the helix containing the site 2 metal ion ligands.[19] This crystal structure confirms that SznF conserves all of the metal-binding His/Asp/Glu residues, as predicted bioinformatically, but it fails to stably incorporate iron in this domain in the current preparations. Substitution of any one of these residues (E215, H225, E281, H311, D315 or H318) with alanine abolished all SznF activity (**Figure 3.16**), which is consistent with the involvement of these residues in binding an essential multinuclear metallocofactor.

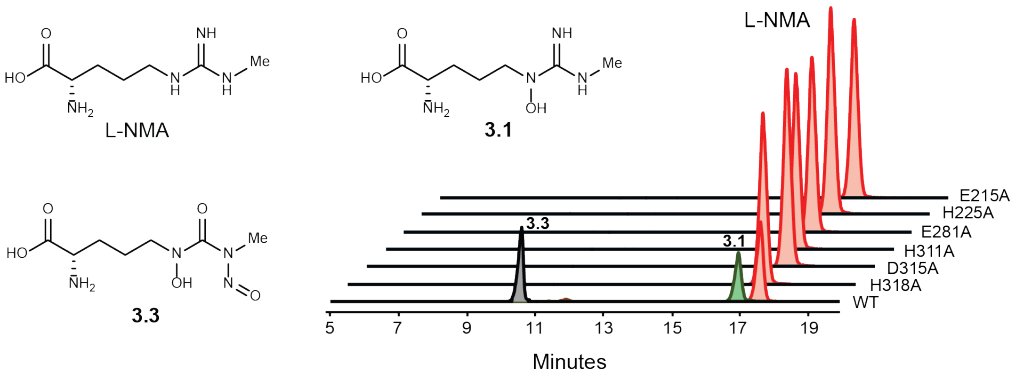


Figure 3.18: Mutation of the predicted iron-binding residues in the SznF central bundle helix domain abolished *N*-oxygenation activity. EICs are shown, and the chromatograms were generated with a 5 ppm window.

The C-terminal cupin domain shows a single metal ion bound by three histidine ligands (**Figure 3.17B**). Anomalous difference maps showed that the cupin domain is fully occupied by Fe. An extended water-mediated H-bonding network tethers the non-metabolizable ligand **3.1**

(green sticks, black lines) in the active site via its Me-Nω, Nδ-O(H), and backbone amine/carboxylate functional groups were present (**Figure 3.19**). Analysis of the network suggests a potential mechanism for deprotonation of **3.1** Me-N$^{\omega}$ via Y459 and E98. The cupin active site also contains an open hydrophobic pocket near the unmethylated Nω position. Apart from the aforementioned Y459 interaction, there are no H-bonds between the substrate functional groups and residues in the active site.

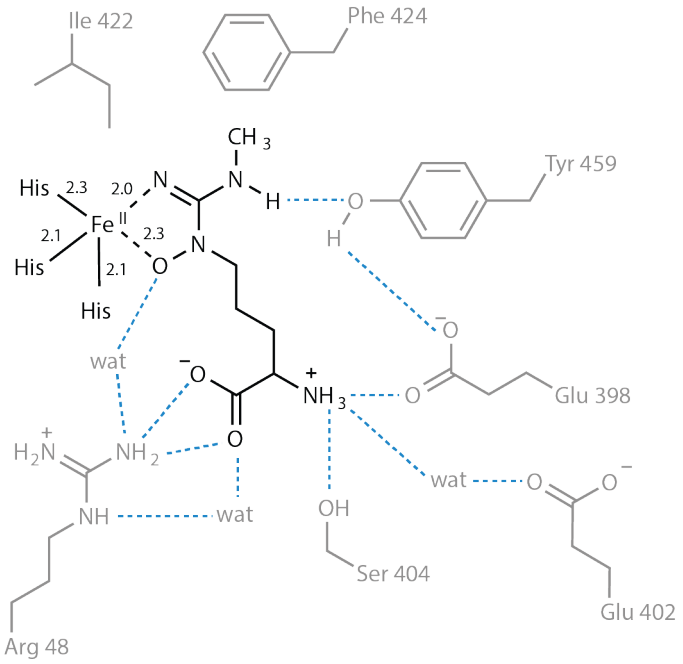


Figure 3.19: Ligand interaction map showing Fe^{II}-coordination interactions (distances in Å) and H-bonding interactions with selected side chains and water molecules. H-bonding/ionic interactions are shown as gray (panel a) or blue (panel b) dashed lines.

3.2.6 Site-directed mutagenesis of SznF

As mentioned in the Section 3.2.1, we cannot predict based on bioinformatics the functions of the two domains. We next sought to define the biochemical reaction catalyzed by each of the two metal-binding domains. To investigate the cupin domain, we substituted its three Fe^{II}-binding histidine residues with alanine. In reactions with L-NMA or **3.1** as the substrate, the

H407A/H409A/H448A variant accumulated intermediate **3.2** but did not generate **3.3** (**Figure 3.20AB**). Conversely, an SznF variant (E215A) with an intact cupin domain and disrupted central domain converted synthetic intermediate **3.2** to *N*-nitrosourea **3.3** (**Figure 3.20C**). The two metallocofactors of SznF are therefore functionally distinct, with the central domain catalyzing two sequential *N*-hydroxylations of L-NMA and the cupin domain enabling oxidative rearrangement and N–N bond formation to yield the *N*-nitrosourea product (**Figure 3.20D**). Although **3.1** is not processed by the cupin domain, the coordination mode in the crystal structure suggested that the N^{δ}-hydroxyl group is important for Fe^{II} binding. Indeed, assays with N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (**3.5**) did not result in consumption of substrate (**Figure 3.21**). Overall, this indicate that the N^{δ}-OH is required for the oxidative rearrangement.

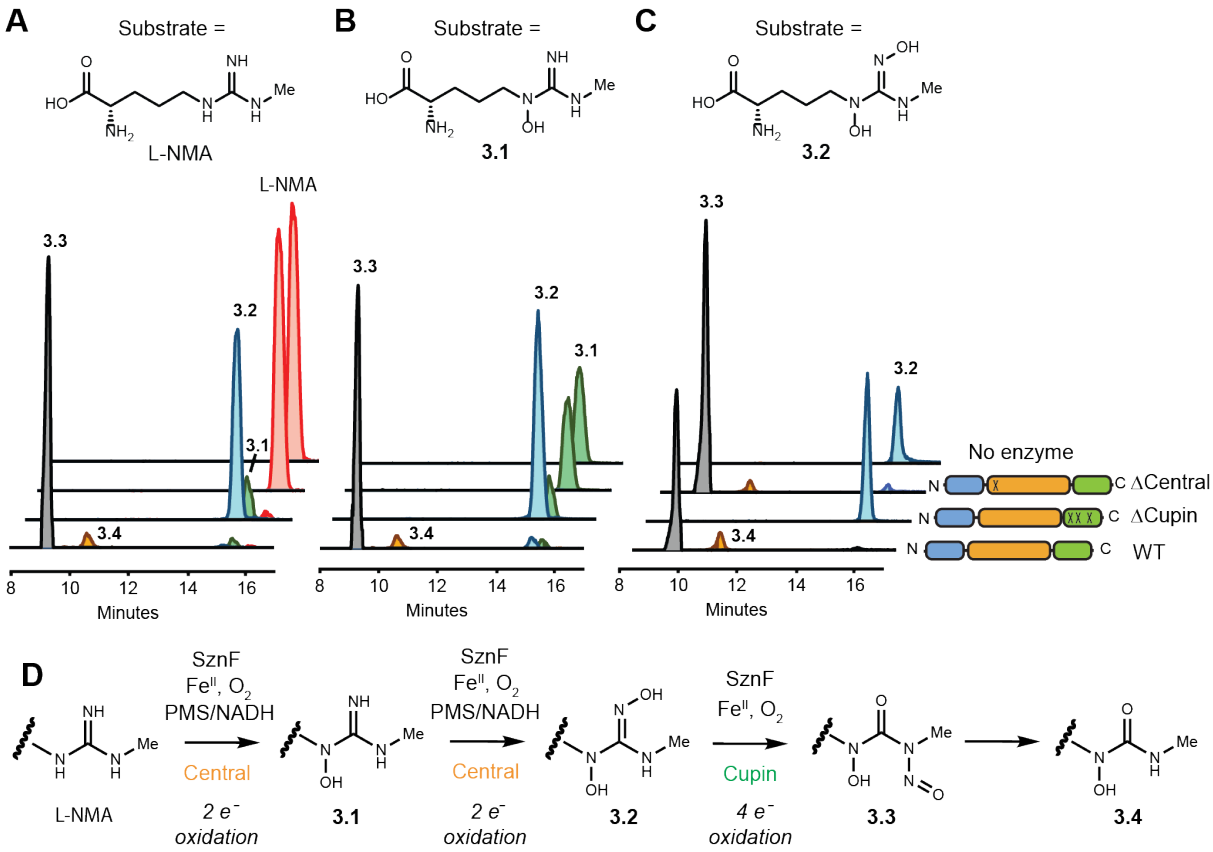


Figure 3.20: The two metallocofactor sites in SznF are functionally distinct. To test the reactivity of the separate domains, L-NMA (panel **A**), **3.1** (panel **B**), or **3.2** (panel **C**) were incubated with no enzyme, an SznF mutant with an inactive central domain (E215A, Δcentral), an SznF mutant with an inactive cupin domain (H407A H409A H448A, Δcupin), or WT SznF. EICs were generated within 5 ppm. **D**) The reactions catalyzed by the two metallocofactor binding sites of SznF.

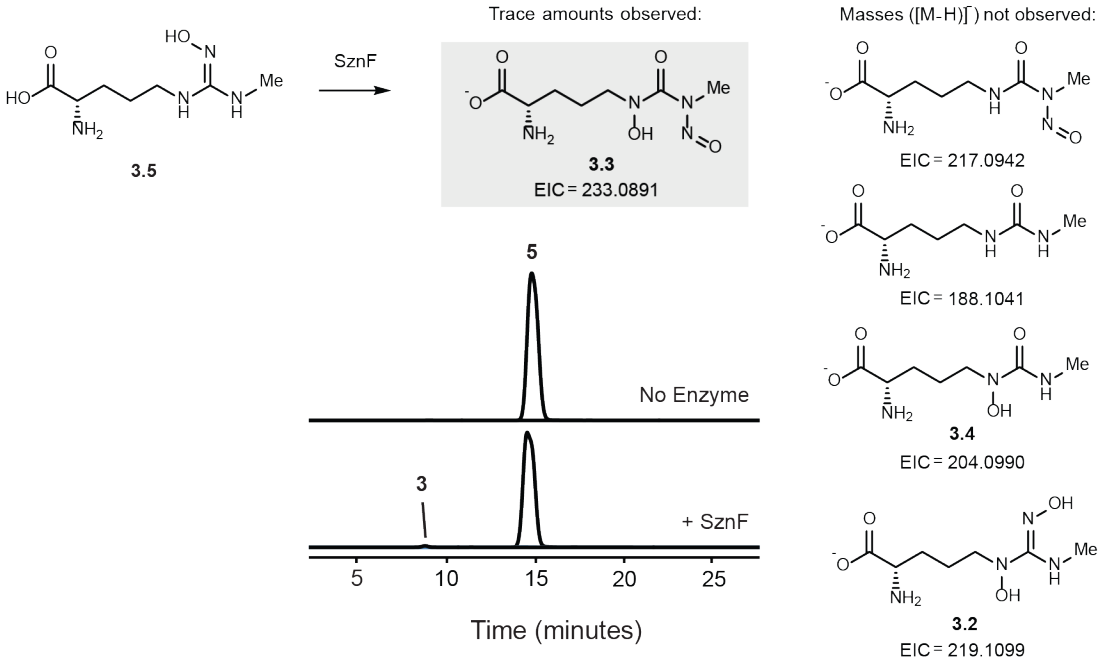


Figure 3.21: N^{ω}-hydroxy-N^{ω}-methyl-L-arginine is not a substrate for SznF. This result suggested that the N^{δ}-hydroxyl group is important for binding of **3.2** to the metal cofactor and subsequent oxygen activation by SznF. EICs were generated within 5 ppm.

Finally, we determined the requirement of external electrons in the SznF-catalyzed oxygenation reactions. While the central domain of SznF requires NADH, the cupin domain completely converted dihydroxylated intermediate **3.2** to *N*-nitrosourea **3.3** without an external reductant (**Figure 3.22**). This result indicates that the four-electron oxidation necessary for the final rearrangement is fully coupled to a single four-electron reduction of O_2. As briefly described in Chapter 1.4, the 3-His enzyme cysteine dioxygenase utilizes a starting Fe^{II} and couples the oxidation of L-cysteine to reduction of O_2 to H_2O. However, we hypothesized that a starting Fe^{III} may also be possible, and the implication for the mechanism will be discussed below in the next section.

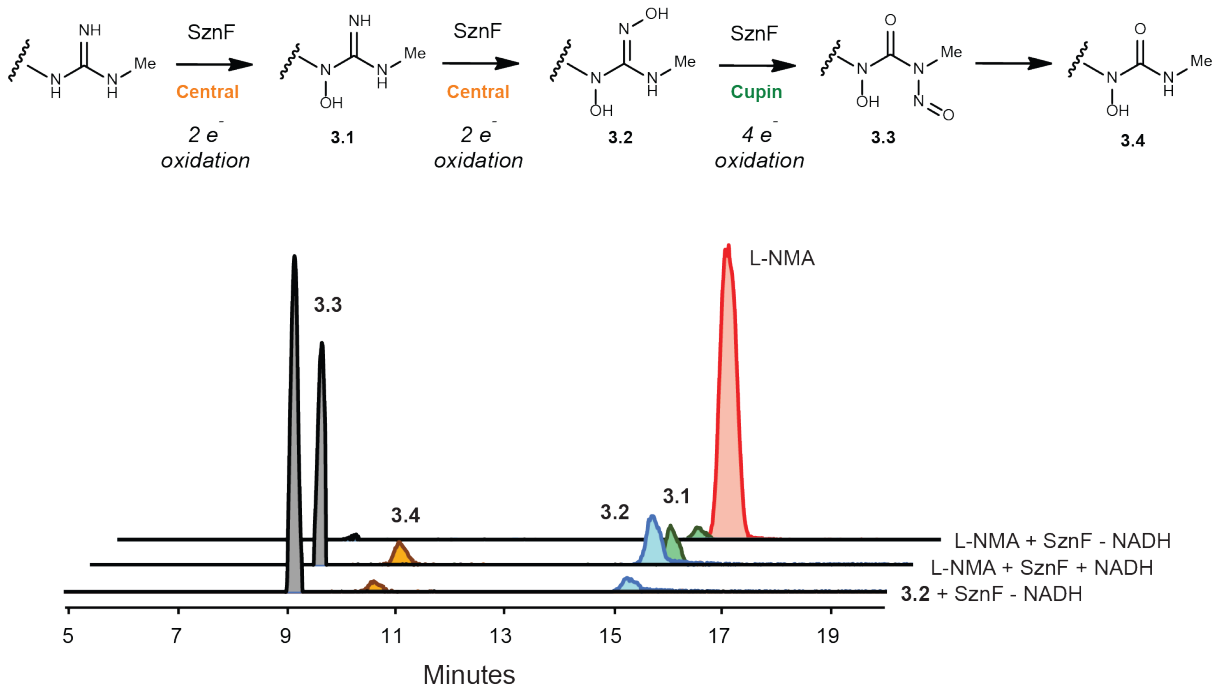


Figure 3.22: External reductant is not required for the oxidative rearrangement reaction. When L-NMA is incubated with SznF without NADH, no reaction progress was observed. However, when **3.2** was used as a substrate without NADH, we observed conversion of **3.2** to **3.3**. EICs are shown, and the chromatograms were generated with a 5 ppm window.

3.2.7 Proposed mechanism of SznF

Part of this mechanistic proposal is written by Professor Amie K. Boal (Penn State University).

Having confirmed that SznF catalyzes an unusual oxidation reaction of L-NMA to form an *N*-nitrosourea product, we proceeded to formula a mechanism for this rearrangement reaction. In light of the recent publications on UndA and RhoS, we can propose one possible mechanism for SznF's *N*-oxygenation reaction by the central heme-oxygenase-like domain.[6-7] First, the diiron cofactor in its di-ferrous state could bind L-NMA and trigger O_2 activation. This reaction is also the first step of the *N*-oxygenation reaction catalyzed by the diiron enzyme AurF.[20] Two electron reduction of O_2 would generate a di-ferric peroxo species, which could be attacked by the Nω nitrogen atom of L-NMA to generate **3.1**. After substrate release, the diferric cofactor would need

to be reduced again to generate the active di-ferrous species. SznF would then re-bind **3.1** and repeat the reaction with N^{ω} of **3.1** to generate **3.2**. This mechanistic proposal raises many opportunities for further investigation. For example, what is the structural basis for distinguishing L-NMA and intermediate **3.1**? The order of diiron cofactor assembly, substrate binding, and O_2 activation is also unclear. Given the propensity of the central domain to become disordered, the diiron cofactor might need to be reassembled after the first round of *N*-oxygenation.

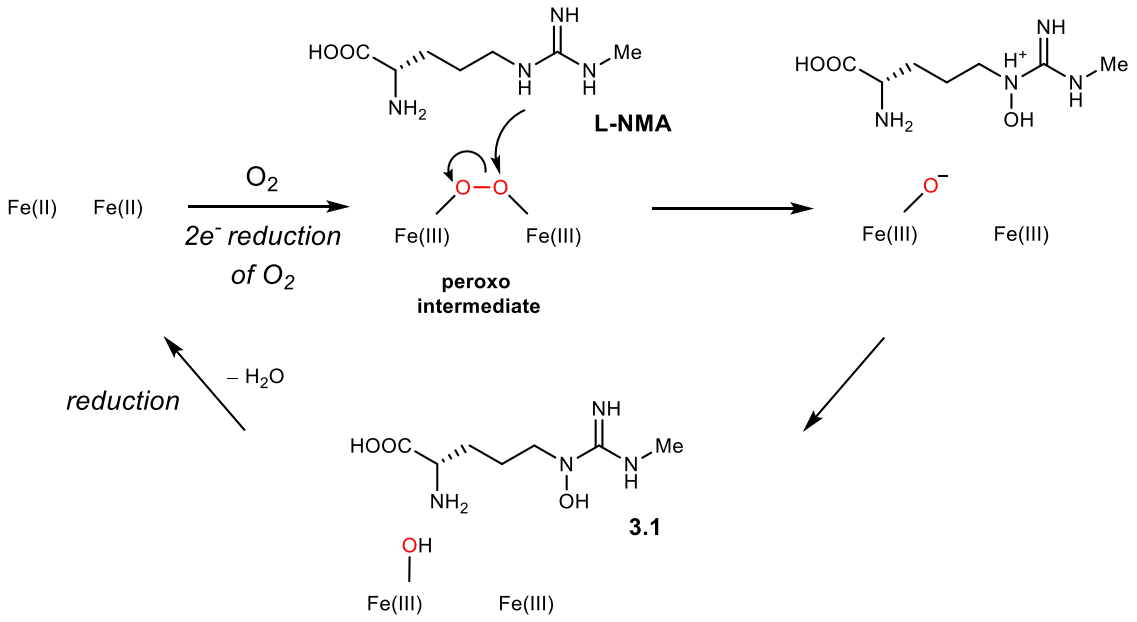


Figure 3.23: One mechanistic proposal for the SznF-catalyzed *N*-oxygenation of L-NMA.

We have attributed the novel oxidative rearrangement to the cupin domain of SznF. This reaction, as mentioned earlier in this Chapter, has not been reported in the literature. Elucidating this mechanism would improve our understanding of the chemistry performed by cupin-domain containing enzymes. We do not yet have an experimentally determined view of the true rearrangement substrate, **3.2**, bound to the cupin Fe^{II} cofactor. However, analog studies and *in vitro* biochemical characterization support N^{δ}-OH chelation of the iron center as observed

crystallographically in **3.1**•Fe^{II}• SznF. Shown below are several possible orientations in which **3.2** could bind (**Figure 3.24**). The C–N bond cleaved during the reaction is indicated by a wavy line.

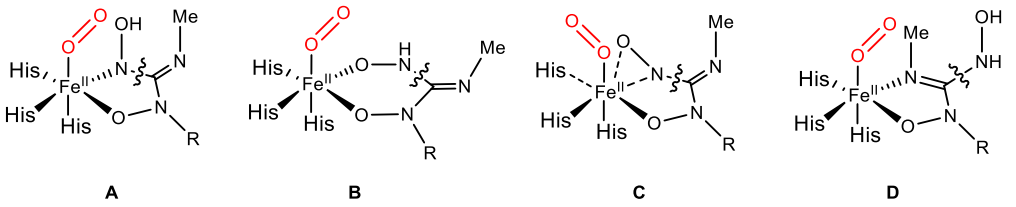


Figure 3.24: Possible binding orientations of **3.2** to the Fe^{II} center of the SznF cupin domain.

One possible mechanism using orientation A as the starting point because it yields a five-membered chelate similar to reactant complexes in α-ketoglutarate-dependent oxygenases[21] and extradiol ring-cleaving Fe^{II} dioxygenases.[22] In orientation B, the C^{ε} of **3.2** could be too far from the distal oxygen atom of the metal-O_2 adduct for the attack required to form the citrulline product. Another possibility is a side-on coordination of the hydroxylamine moiety of **3.2** to the metal center (orientation C). This binding mode would require a dissociation of one of the histidine ligands. Orientation D is disfavored because the methylated N^{ω} of **3.2** acts as a nucleophile in both mechanistic routes, and it will be more reactive if it is not coordinated to the Fe^{II} cofactor. We propose initial Fe^{II} coordination of substrate **3.2** (via mode A) in a fully deprotonated state to achieve charge balance with a neutral protein ligand set (**Figure 3.24**). Loss of the proton from the methylated N^{ω} atom may be facilitated by the Y459/E398 pair in the second-sphere of the SznF cupin domain active site.

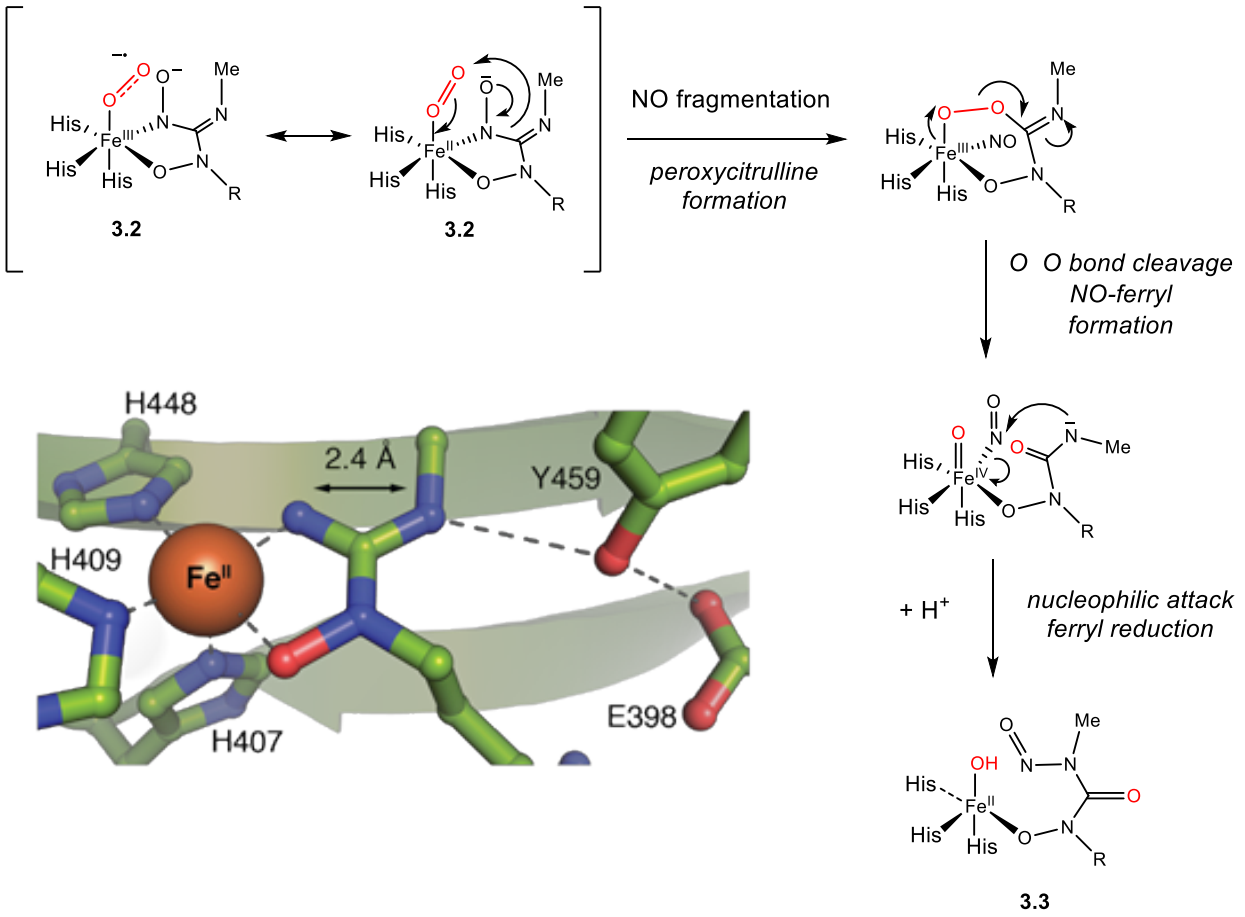


Figure 3.25: Proposed mechanism of SznF-catalyzed oxidative rearrangement that invokes a ferryl-nitrosyl intermediate.

The initial stages of this proposed mechanism resemble the first steps in well-established scission of α-ketoglutarate to a ferrylsuccinate complex and CO_2 by mononuclear iron enzymes in the α-ketoglutarate-dependent dioxygenase superfamily (Chapter 1.4)[21]. We propose that Fe^{II} -binding of **3.2** triggers addition of O_2 to yield an end-on six-coordinate adduct. Fragmentation of **3.2** at the N^{ω}-C^{ε} bond generates both an {FeNO}8 complex and a cyclic peroxycitrulline intermediate. The latter complex is analogous to the peroxysuccinate intermediate proposed computationally and detected crystallographically in α-ketoglutarate enzymes. In those systems, O–O bond cleavage yields the first accumulating intermediate in solution, a high-spin Fe^{IV}=O complex[21]. In SznF, we propose that a similar transformation generates a distinctive anionic NO

coordinated ferryl, defined here in Enemark-Feltham notation[23-24] as an {FeNO}6 complex. Formation of the carbonyl at C^{ε} also yields an amidate nucleophile at Me-N^{ω}.

At this stage, the SznF reaction pathway may diverge from that of α-ketoglutarate-dependent enzymes. We propose that that the Me-N^{ω} amidate nitrogen attacks the coordinated nitrogen of the {FeNO}6 unit in SznF. Importantly, this nitrosyl transfer event returns the catalyst to the Fe^{II} state, consistent with experimental observation of full coupling of the oxidizing equivalents in co-substrate O_2 to the four-electron oxidation of substrate **3.2**. We believe that the SznF {FeNO}6 intermediate is suited for such reactivity based on structural considerations and comparison to enzymatic heme and iron model-complex NO reactivity. In the **3.1**•Fe^{II}•SznF complex, the two N^{ω} atoms involved in N–N bond formation are located 2.4 Å apart. This proximity is optimal for nucleophilic attack and is likely maintained in the NO-coordinated intermediate via substrate coordination at the O-N^{δ} position. Many iron-nitrosyl complexes are susceptible to nucleophilic attack at the Fe-coordinated nitrogen[24] and, in a particularly relevant biological reaction, a heme {FeNO}6 complex in a cytochrome P460 implicated in nitrogen cycle chemistry undergoes N–N bond formation to generate N_2O via reaction with hydroxylamine.[25-26] Together, these examples provide important precedent for the novel SznF-dependent N–N bond formation in a distinctive arginine-derived substrate.

Interestingly, the citrulline derivative and FeNO complex generated transiently in this reaction pathway are similar to the products of the second reaction of heme-dependent nitric oxide synthase (NOS), in which a hydroxylated L-Arg intermediate is fragmented at the C^{ε} –N^{ω} bond to serve as the biological source of NO.[27] The cofactor requirements of NOS, a 5-coordinate His-ligated Fe-porphyrin complex, preclude coordination of the substrate L-Arg and the hydroxylated L-Arg intermediate. In SznF, we speculate that a non-heme iron site with fewer requirements for

tethering ligands allows for direct interaction with substrate and subsequent evolution of new N–N coupling activity.

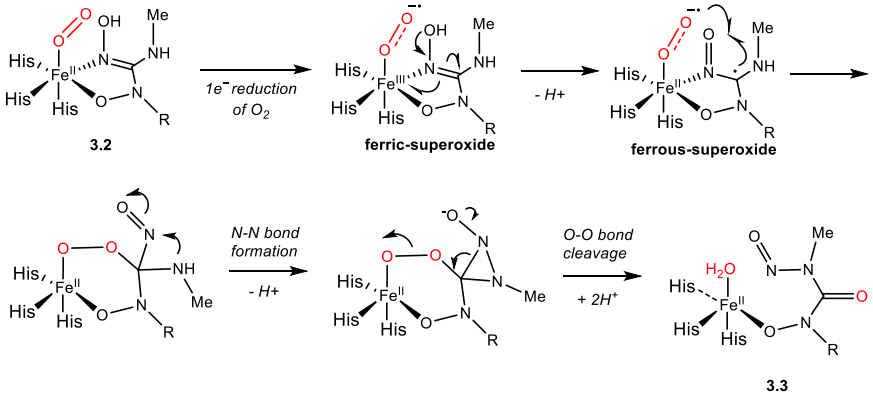


Figure 3.26: Proposed mechanism of SznF-catalyzed oxidative rearrangement that invokes a diaziridine intermediate.

In order for proposal **3.1** to be consistent with our crossover experiment (**Figure 3.12**), the transient {FeNO}6 complex cannot exchange prior to N–N bond formation. This phenomenon could be promoted by the close proximity of the functional groups enforced by initial bidentate Fe^{II} coordination of the substrate. However, with the result of the crossover experiment in mind, we have also proposed a second mechanism in which transient NO formation is not invoked (**Figure 3.26**). Bidentate coordination of a redox-active substrate by SznF is a feature shared in common with non-heme iron catechol dioxygenases (**Figure 3.26**).[22] In the second proposal, binding of substrate **3.2** triggers addition of O_2 to yield an initial Fe^{III}-superoxo adduct[28]. But, as in the proposed mechanism of ring-cleaving catechol dioxygenases, direct coordination of substrate promotes electron transfer to generate an Fe^{II}-superoxo and a substrate-centered radical.[22]

This radical would be stabilized by the adjacent nitrogen atoms. Recombination between the superoxide radical and the Cε • of **3.2** forms an Fe^{II}-peroxo bridge at Cε of **3.2**.[28-30] In the catechol dioxygenases, a similar structure undergoes a Criegee rearrangement and hydrolysis to ultimately incorporate both O atoms of the cosubstrate O_2 into the ring-opened product. Here, we

propose that the alkylperoxide intermediate generated by SznF breaks down following N–N bond formation to generate a diaziridine intermediate. Rearrangement of the diaziridine could result in O–O bond cleavage, generating *N*-nitrosourea **3.3** and water as a co-product.

3.3 Discussion

This chapter presented data that confirmed the role of SznF in catalyzing the proposed oxidative rearrangement of L-NMA to form an amino acid containing an *N*-nitrosourea moiety. During the course of biochemical characterization experiments, we identified at least four amino acid intermediates and products of this reaction. We have also defined the essential cofactor(s) required for this reaction. Site-directed mutagenesis and x-ray crystallography identified key residues important in binding metal cofactors and substrate. These experiments also provided evidence that the two metal-binding domains are functionally distinct.

Some of our results were not reported in the aforementioned study performed by Ryan and coworkers.[2] For example, we were able to detect formation of both **3.2** and **3.3** by SznF, while Ryan and coworkers did not report detection of these two products. In their enzyme assay they observed production of **3.1**, **3.4**, and a dehydroxylated **3.4** congener we did not identify in this thesis work. This discrepancy between the two studies could be attributed to differences in sample handling, such as their derivatization methods, or differences in the instrumentation sensitivity. Aside from these differences, our studies both concluded that SznF contains two metal-binding active sites and that the enzyme can generate nitric oxide from L-NMA. Using site-directed mutagenesis, both studies also arrived at the same conclusion that the two sites are functionally distinct. However, Ryan and coworkers proposed one possible mechanism in which the central

diiron domain generates NO and acknowledged that the cupin domain may or may not be involved in *N*-oxidation. Results obtained in our study did not support the hypothesis that the role of the diiron domain is to generate NO. In our work, we experimentally demonstrated that the cupin domain utilizes a mononuclear iron to perform the oxidation and rearrangement of **3.2**, while the central domain performs two hydroxylation reactions. This difference in the mechanisms proposed in these two studies could be attributed to Ryan et al's inability to detect products **3.2** and **3.3**. Our mechanistic proposal is also partially informed by the crystal structure of SznF and extensive biochemical characterization of SznF mutants. Finally, the ^{18}O-labeling experiments performed by Ryan and coworkers also support our findings that the oxygen atoms installed in **3.4** are derived from $^{18}O_2$ and not $H_2{}^{18}O$. Overall, the two studies confirm that SznF is a multi-domain iron-binding enzyme that leverages a nitrogen-rich L-NMA substrate to generate an *N*-nitrosourea-containing compound.

Future efforts to characterize SznF will be focused on elucidating both the *N*-oxygenation and the oxidative rearrangement mechanisms. Stopped-flow UV-vis analysis of a single turnover reaction catalyzed by the CADD-domain of SznF and detailed spectroscopic studies analogous to those performed with UndA should provide useful information about the active oxygenating species from a metal cofactor focused perspective. This may provide evidence for the binding mode of the peroxo species to the proposed diiron cofactor. Assays with analogues of L-NMA and **3.1** might reveal features of the substrate that are important for catalysis. While we have identified four compounds produced from the SznF-catalyzed *N*-oxygenation, other as-yet unidentified amino acid products may be produced and offer insights into the reaction mechanism. These products could be detected via NMR analysis of the enzyme assay or derivatization with Fmoc during different time points of the SznF reaction.

A detailed study of reaction outcomes using the various SznF mutants and assessing the distribution and rates of product formation may provide insight into how each residue within the first coordination sphere of the active site contribute to catalysis. In order to start investigating the rearrangement reaction, binding constants of each substrate to the respective domain should be elucidated. In addition, the starting oxidation state of the metal should be unequivocally determined with steady state kinetics using reduced or oxidized SznF and L-dHMA. While we presented evidence that Fe is necessary and sufficient for the entire eight electron oxidation of L-NMA, the question remains whether other divalent metals, such as manganese or cobalt, might be used as a metallocofactor for the rearrangement reaction. While we hypothesized that an {FeNO}6 intermediate might be involved in catalysis, freeze quench EPR should be employed to observe transient states. If the reaction is too rapid, mutations and substrate analogues should be used to slow down kinetics in order to capture and characterize intermediates.

Both of these approaches would benefit from additional structural information. Co-crystallization of L-dHMA within the cupin domain active site would provide our first insight into substrate binding and help prioritize what residues would be important in initiating catalysis of the rearrangement. Given the solvent-exposed nature of this active site and promiscuous binding to hydroxy guanidine substrates, it would be interesting to investigate how these substrates trigger oxygen activation. Currently, this site lacks full occupancy. Occupying both the iron-binding sites in the CADD central domain would also likely lead to improved understanding of this class of diiron enzymes. This reaction is unique because the central domain catalyzes two different oxygenating reactions on different nitrogen atoms of the substrate. Finally, in crystallo capture of L-NMA and L-HMA within the diiron site may shed light on the difference in substrate binding that needs to occur for these different outcomes.

3.4 Materials and Methods

Overexpression and purification of SznF

The *sznF* gene was amplified from the cosmid encoding the *szn* gene cluster using the primers:(sznF-F:TATCATATGAGTCACGTCCCCCCGCACGT; sznF-R:ATAAAGCTTCTACAGGCACTTTCGGTAGT) The amplified gene was digested with NdeI and HindIII and ligated into the pET28a vector as described above for pET28a–SznE. The plasmid was transformed into *E. coli* BL21 (DE3) cells (Invitrogen), and the protein was expressed and purified using the same procedure as described for SznE with the exception that the exchange buffer contained 20 mM MOPS pH 7.5 instead of 20 mM HEPES pH 8.

LC–MS assay of SznF activity

In a total volume of 50 μL, 50 mM MOPS pH 7.5, 20 μM PMS, 80 μM of SznF and 1 mM of amino acid substrate were mixed. NADH (Merck Millipore) was added at a final concentration of 5 mM to initiate the reaction. Reaction mixtures were quenched by adding 50 μL of LC–MS grade methanol. After 10 min on ice, the reaction mixtures were centrifuged to remove precipitated protein and analyzed using normal phase LC–HRMS as described in section 'Gene disruption and chemical complementation experiments' in which L-NMA, **3.1** and **3.2** were detected from cell extracts.

For analysis of Fmoc derivatives, a 10 mM 9-fluorenylmethyl *N*-succinimidyl carbonate (Fmoc-OSu) acetonitrile solution was first prepared. An additional 50 μL of acetonitrile was added to the quenched reaction mixtures, followed by 5 μL of the 10 mM Fmoc-OSu solution. This reaction mixture was incubated for 1 h, centrifuged to remove precipitates, and analyzed by reversed phase LC–HRMS. The LC column was an Acclaim Polar Advantage II C18 column (3

μm, 120 Å, 2.1 × 150 mm, Thermo Fisher Scientific). The flow rate was 0.3 mL/min. The LC conditions were: 95% to 5% solvent A in 15 min; 5% solvent A, hold for 5 min; 5% to 95% solvent A in 4 min; and 5 min equilibration at 95% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile).

To access labeled L-NMA, a large-scale SznE assay was prepared. The assay mixture contained 50 mM potassium phosphate buffer pH 8, 10 mM $MgCl_2$, 10 mM L-arginine, 40 mM SAM and 10 μM SznE in a final volume of 1 mL. The reaction was allowed to proceed overnight at room temperature and was quenched with 2 mL of MeOH. After centrifugation to remove precipitates, the enzyme reaction was evaporated to dryness, resuspended in 1 mL of water, and labeled L-NMA was purified using semi-preparative scale HPLC. The LC column was Kromasil 100 C18 (5 mm, 100 Å, 10 × 250 μm). The flow rate was 3 mL/min. The LC conditions were: a linear gradient from 5% to 50% solvent B in 12 min; an isocratic hold at 50% solvent B for 2 min; a linear gradient from 50% to 5% solvent B in 2 min; and an isocratic hold at 5% solvent B for 2 min (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). d_3-L-NMA was monitored at 200 nm and eluted between 4–4.5 min. The collected fractions were lyophilized to dryness, and the residue was resuspended to make a 100 mM stock solution, which was kept at –20 °C.

Testing the metal dependence of SznF

apo-SznF was prepared by inoculating 1 l of M9 medium with 1% of a saturated LB culture of *E. coli* BL21 containing pET28a–SznF. At an OD_{600} of around 0.4, 0.2 mM of IPTG was added and the culture was incubated at 15 °C overnight. The protein purification procedure was identical to that described above for SznF, except that after elution from the Ni-NTA column, a 100 mM

aqueous solution of EDTA pH 7 was added to a final concentration of 0.5 mM to remove any residual Ni bound to SznF. This solution was concentrated to about 400 µL and diluted into 12 mL of exchange buffer containing 0.5 mM EDTA. The protein solution was concentrated to around 400 µL and diluted again into 12 mL of exchange buffer without EDTA. The protein sample was concentrated once more to obtain a desalted EDTA-treated SznF sample. An aliquot of this solution was submitted to the University of Georgia Center for Applied Isotope Studies for inductively-coupled plasma-mass spectrometry (ICP-MS) analysis. This SznF sample had <0.04 ppm ^{56}Fe, ^{65}Cu, ^{59}Co and ^{55}Mn by ICP-MS analysis was referred to as *apo*-SznF.

To screen the effects of different metals on the SznF-mediated *N*-oxygenation reactions, 100 mM stock solutions of $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$, $MnCl_2 \cdot 4H_2O$, $CaCl_2 \cdot 2H_2O$, $ZnCl_2$, $CuCl_2 \cdot 2H_2O$, $CoCl_2 \cdot 6H_2O$, $Na_2MoO_4 \cdot 2H_2O$ and $NiSO_4 \cdot 6H_2O$ were prepared in 1 M aqueous HCl. Assay mixtures contained 50 mM MOPS pH 7.5, 1 mM L-NMA, 200 µM metal, 20 µM PMS, 5 mM NADH and 100 µM *apo*-SznF in a final volume of 30 µl. Assay mixtures were incubated for 4 h before quenching with methanol and analyzing with LC–HRMS.

Griess assay for nitrite production and nitric oxide detection by electron paramagnetic resonance spectroscopy

To characterize nitrite liberated from the SznF-mediated *N*-nitrosation, solutions of 0.5% sulfanilic acid in 30% acetic acid (solution A) and 0.1% (napthyl)ethylenediamine dihydrochloride in 30% acetic acid (solution B) were prepared. 25 µL of solution A, 25 µL of solution B, 10 µL 6 M HCl and 10 µL of the enzyme assay mixtures were combined and heated at 60 °C for 3 min. The samples were centrifuged to remove precipitated proteins, and the absorbance at 548 nm was recorded with a BioTek Gen5 Microplate Reader.

Nitric oxide detection was carried out by adapting the work of reference 15. In brief, a 10x stock aqueous solution containing 50 mM iron ammonium sulfate and 10 mM *N*-methyl-D-glucamine dithiocarbamate (MGD) was freshly prepared. 25 μL of this $(MGD)_2Fe^{II}$ solution was then added to 225 μL of an SznF assay mixture. After 30 min at room temperature, the reaction mixture was transferred to a quartz capillary and analyzed with a Bruker ElexSysE500 EPR instrument with a 100 K–600 K Digital Temperature Control at room temperature. The following parameters were used: 9.75 GHz microwave frequency; 100 kHz modulation frequency; 10 mW microwave power, 3.81 G modulation amplitude; 3,418 G centre field; 82 ms time constant. An enzyme-free sample containing 1 mM sodium 2-(*N*,*N*-diethylamino)-diazenolate-2-oxide (DEANO) was used as a positive control. To demonstrate that the nitric oxide was generated from the terminal nitrogen atoms of L-NMA, 1 mM of $[^{15}N_2]$-L-arginine (Cambridge Isotope Laboratories) was first converted to $[^{15}N_2]$-L-NMA with SznE as described in the section '^{1}H NMR assay of SznE activity'. After 1 h, SznF, PMS and NADH were added directly to the reaction mixture to the same final concentrations as described in the section 'LC–MS assay of SznF activity'. $(MGD)_2Fe^{II}$ was then added, and the SznEF assay mixture was analyzed with EPR spectroscopy.

Addition of superoxide dismutase and catalase to SznF reaction mixtures

Stock solutions of superoxide dismutase (from bovine, recombinant from *E. coli*, Sigma-Aldrich) and catalase (from bovine liver, Sigma-Aldrich) were prepared by adding 100 μL of buffer (20 mM MOPS pH 8, 50 mM NaCl) to the lyophilized powders. After centrifugation to remove particulates, the enzyme solutions were added to the SznF reaction mixtures at a final concentration of 1U per 50 μL before addition of NADH. The reactions were allowed to proceed

for 1 h before activity was measured using the Griess assay. The experiments were performed in triplicate on three different days.

Oxygen consumption assay

Dissolved oxygen was monitored using a MultiFrequency phase fluorometer equipped with a FOXY optode (Ocean Optics). The optode was calibrated using a two-point calibration method with air-saturated 20 mM MOPS buffer pH 7.7 (280 μM) and the same solution saturated with sodium dithionite (0 μM). 500 μL of an SznF reaction mixture was prepared in a sealed pear-shaped flask with stirring as described above. The data was analyzed using NeoFox Viewer version 2.40. The different experimental conditions tested are described in **Figure 3.13** The experiments were performed at least twice on different days.

$^{18}O_2$ and $H_2{}^{18}O$ labelling experiments

Solutions of 1 M HEPES buffer, 1 M $MgCl_2$ and water were degassed by sparging with argon for 1 h. SznF was made anaerobic by repeated cycles of vacuum and purging with argon. The enzyme, degassed buffer, degassed solution of $MgCl_2$, degassed water, PMS and NADH were brought into an MBraun glove box under an atmosphere consisting of 99.997% nitrogen (N_2) with less than 0.1 ppm O_2. 50 mM HEPES pH 8.0, 10 mM $MgCl_2$, 1 mM L-NMA or 1 mM **3.1**, 200 μM PMS and 100 μM of SznF were mixed in a vial in a total volume of 100 μL. The reaction mixture and a degassed, aqueous solution of NADH were brought into a glovebag purged with nitrogen. The reaction in the vial was charged with 5 mL of $^{18}O_2$ using a gas-tight syringe, and NADH was added to a final concentration of 5 mM. After 1 h, methanol briefly sparged with argon was drawn into a gas-tight syringe previously filled with N_2 inside the glove bag and added to the

reaction mixture. The reactions were derivatized with Fmoc and analyzed with HR–LCMS and tandem MS (MS/MS). For the reaction performed in $H_2{}^{18}O$, the reaction mixture was prepared as above except $H_2{}^{18}O$ was used to a final volume of 50 μL instead of 100 μL. The reaction mixture was quenched, derivatized, and analyzed with LC–HRMS.

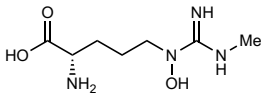
Site-directed mutagenesis

The SznF point mutants were constructed by adapting the Quikchange protocol.[31] In brief, each PCR reaction contained 41.65 μL H_2O, 1.5 μL DMSO, 0.2 μL dNTP (50 mM stock), 0.4 μL of pET28a–SznF (50 ng/μL stock), 0.125 μL of each mutagenesis primer (200 μM stock, nucleotide sequences in **Table 3.1**), 1 μL of Pfu Turbo AD and 5 μL of Buffer (Turbo AD). Thermocycling was carried out in a MyCycler gradient cycler (Bio-Rad) using the following parameters: denaturation for 2 min at 95 °C; 18 cycles of 0.5 min at 95 °C, 1 min at 71 °C, and an extension time of 13 min at 72 °C. 40 μL of the PCR reaction mixture was digested with 2 μL of DpnI for 2 h at 37 °C. Afterwards, the reaction was transformed into chemically competent *E. coli* Top10 cells, and the constructs containing the desired point mutants were confirmed by sequencing. The mutant SznF enzymes were expressed, purified, and assays using the procedures described above for wild-type SznF.

Table 3.1: Primers used for generating SznF variants

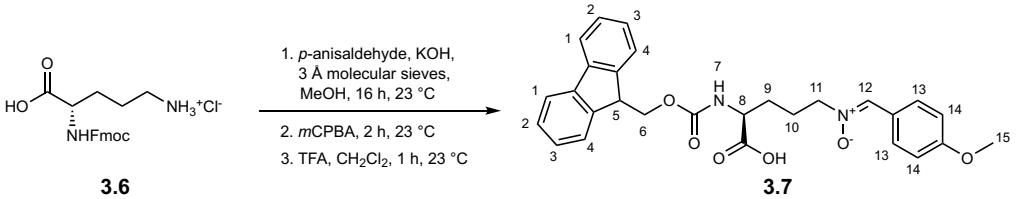
Oligo	Nucleotide Sequence (5' to 3')
E215A-F	CATCGAC**GCC**TACGGCTACGGCGTCCACGACACC
E215A-R	CCGTA**GGC**GTCGATGACGACCTTGAACCACTCGG
H225A-F	GTCCACGACACCAAG**GCC**AGCACCCTGTTCGAG
H225A-R	CTCGAACAGGGTGCT**GGC**CTTGGTGTCGTGGAC
E281A-F	GCGCTGTACTACACG**GCC**AGCTCCCTGGTCGAC
E281A-R	GTCGACCAGGGAGCT**GGC**CGTGTAGTACAGCGC
H311A-F	CACCTACTTCACCGAG**GCC**ATCCACATCGACCAG
H311A-R	CTGGTCGATGTGGAT**GGC**CTCGGTGAAGTAGGTG
D315A-F	CATC**GCC**CAGCACCACGGCCGGATGGCCCGCG
D315A-R	GCTG**GGC**GATGTGGATGTGCTCGGTGAAGTAGG
H318A-F	GCAC**GCC**GGCCGGATGGCCCGCGAGAAGATCATC
H318A-R	CGGCC**GGC**GTGCTGGTCGATGTGGATGTGCTCG
H407A-F,	GAGCTGTCCAACACC**GCC**TGCCACGACGGTGAC
H407A-R	GTCACCGTCGTGGCA**GGC**GGTGTTGGACAGCTC
H409A-F	GTCCAACACCCACTGC**GCC**GACGGTGACGAGCTG
H409A-R	CAGCTCGTCACCGTC**GGC**GCAGTGGGTGTTGGAC
H448A-F	CAAGCGCAACCGGCTG**GCC**GGCGCGAACATCGAG
H448A-R	CTCGATGTTCGCGCC**GGC**CAGCCGGTTGCGCTTG
H408A_triple_F1	CTGC**GCC**GACGGTGACGAGCTGTGCCACATCG
H408A_triple_R1	CGTC**GGC**GCAGGCGGTGTTGGACAGCTCGCCG

Synthesis of N^{δ}-hydroxy-N^{ω}-methyl-L-arginine (3.1)



3.1

Step 1:

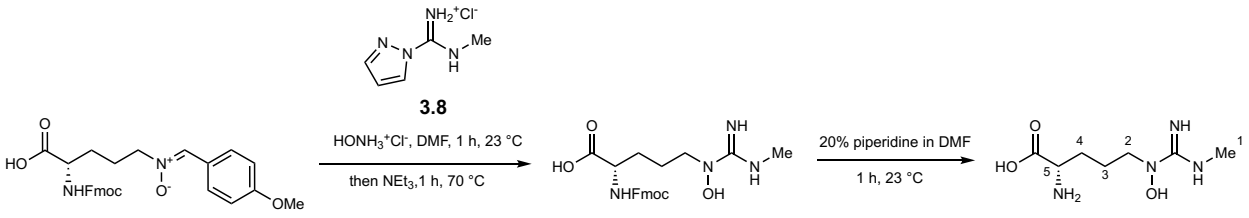


3.6 **3.7**

The nitrone **3.7** was prepared by adapting the procedure from Yun-Ming Lin et al.[32] To an oven-dried flask was first added 3 Å activated molecular sieves followed by 12 mL of anhydrous methanol. *N*-alpha-(9-Fluorenylmethyloxycarbonyl)-L-ornithine hydrochloride (Fmoc-L-ornithine, **3.6**, Chem-Impex International, Inc.; 433 mg, 1.11 mmol, 1.00 equiv), *p*-anisaldehyde (149 μL, 1.22 mmol, 1.10 equiv), and potassium hydroxide (KOH, 68.4 mg, 1.22 mmol, 1.10 equiv) were added in one portion and stirred overnight at room temperature under N_2. Then the reaction mixture was cooled to 0 ºC, and *m*-chloroperoxybenzoic acid (*m*CPBA, 191.5 mg, 1.11 mmol, 1.00 equiv) was added to the stirring reaction mixture in two portions over 5 min. The reaction mixture was stirred at 0 ºC for 1 h before another equiv of *m*CPBA was added to drive the reaction to completion. After 1 h of stirring at 0 ºC, the reaction mixture was filtered to remove the molecular sieves and precipitated salts, and the sieves were thoroughly washed with methanol. The combined filtrate was concentrated to dryness *in vacuo*, and the residue was dissolved in 10 mL of CH_2Cl_2 and 1 mL of TFA. The resulting reaction mixture was allowed to stir at room temperature for 1 h. The reaction mixture was then concentrated *in vacuo* to afford crude nitrone **3.7**, which was purified with a 10 g C18 Sep-Pak cartridge (Waters). The cartridge was first washed with 100 mL of water, then 100 mL of 30% acetonitrile in water to

remove **3.6**, and the nitrone **3.7** was eluted with 200 mL of 50% acetonitrile in water. The desired fraction was concentrated *in vacuo* to remove the acetonitrile, affording an aqueous suspension containing **3.7**. The aqueous solution was then extracted with 2 x 50 mL of ethyl acetate. The combined organic layer was dried over Na_2SO_4, filtered, and evaporated to afford **3.7** as a light brown solid (160 mg, 30% yield). ^{1}H NMR of **3.7** (600 MHz $CDCl_3$): δ (ppm) = 8.18 (d, J = 8.5 Hz, 2H, H_1), 7.73 (d, J = 7.5 Hz, 2H, H_{13}), 7.58 (t, J = 8.4 Hz, 2H, H_4), 7.47 (s, 1H, H_{12}), 7.40 – 7.34 (m, 2H, H_2), 7.30 – 7.26 (m, 2H, H_3), 6.90 (d, J = 8.4 Hz, 2H, H_{14}), 5.94 (d, J = 6.9 Hz, 1H, NH_7), 4.41 – 4.31 (m, 3H, H_5 and H_6), 4.17 (t, J = 7.3 Hz, 1H, H_8), 4.13 – 3.93 (m, 2H, H_{11}), 3.81 (s, 3H, H_{15}), 2.20 – 2.01 (m, 2H, H_9), 1.98 – 1.88 (m, 2H, H_{10}). ^{13}C NMR of **3.7** (125 MHz $CDCl_3$): δ (ppm) = 173.4 (**C**), 162.4 (**C**), 155.9 (**C**), 143.8 (2 x **C**), 141.3 (**C**H), 139.3 (2 x **C**), 132.4 (2 x **C**H), 127.7 (2 x **C**H), 127.1 (2 x **C**H), 125.1 (2 x **C**H), 121.7 (**C**), 120.0 (2 x **C**H), 114.1 (2 x **C**H), 67.0 (**C**H_2), 65.2 (**C**H_2), 55.5 (**C**H_3), 53.2 (**C**H), 47.1 (**C**H), 29.6 (**C**H_2), 23.1 (**C**H_2). HRMS: Calc'd for $C_{28}H_{28}N_2O_6Na$ $[M+Na]^+$ = 511.1845, found = 511.1818.

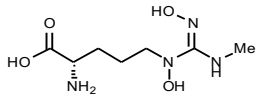
Step 2 – 3:



To nitrone **3.7** (100 mg, 0.205 mmol, 1.00 equiv) in 20 mL of dimethylformamide (DMF) was added hydroxylamine hydrochloride ($HONH_3^+Cl^-$, 17.0 mg, 0.246 mmol, 1.20 equiv). The reaction mixture was stirred at room temperature for 1 h under N_2. Then, NEt_3 (57.0 μL, 0.410 mmol, 2.00 equiv) and 1*H*-pyrazole-1-(*N*-methylcarboxamidine) hydrochloride (**3.8**, Alfa Aesar) (32.8 mg, 0.205 mmol, 1.00 equiv) were added to the reaction mixture in one portion, followed by

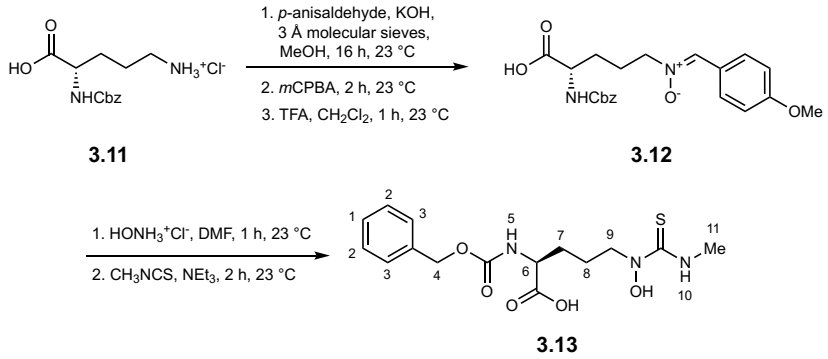
heating to 70 ºC with an oil bath, and stirring for 1 h under N_2. The reaction mixture was then allowed to cool to room temperature, and 4 mL of piperidine were added. After an additional 1 h of stirring at room temperature, the reaction mixture was concentrated *in vacuo*, and the residue was dissolved in 2 mL of water. The aqueous solution was washed with 2 x 2 mL of ethyl acetate to remove nonpolar impurities, and the aqueous layer was concentrated *in vacuo* to remove residual organic solvents. Crude **3.1** was purified by preparative HPLC using a Thermo Hypersil Gold aQ C18 (250 × 20 mm, 5 μm, Thermo Fisher Scientific). The flow rate was 8 mL/min. The LC conditions were: a linear gradient from 5 to 95% solvent B in 15 min; an isocratic hold at 95% solvent B for 5 min; a linear gradient from 95 to 5% solvent B in 2 min; and an isocratic hold at 5% solvent B for 3 min (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). **3.1** was monitored at 200 nm and eluted between 7.2 – 8.2 min. The collected fractions were lyophilized to afford 12 mg of **3.1** as a white solid (28% yield). ^{1}H NMR of **3.1** (500 MHz D_2O): δ (ppm) = 4.06 (t, *J* = 6.2 Hz, 1H, H_5), 3.63 (t, *J* = 6.7 Hz, 2H, H_2), 2.91* (s, 2.3H, H_1), 2.89* (s, 0.6H, H_1), 2.09 – 1.78 (m, 4H, H_3 and H_4). ^{13}C NMR of **3.1** (125 MHz D_2O): δ (ppm) = 172.3 (**C**), 158.7 (**C**), 52.9 (**C**H), 51.5 (**C**H_2), 27.7 (**C**H_2), 27.1* (**C**H_3), 26.8* (**C**H_3), 21.5 (**C**H_2). HRMS: Calc'd for $C_7H_{15}N_4O_3$ $[M-H]^-$ = 203.1150, found = 203.1153. *rotamers

Synthesis of N^{δ}-hydroxy-N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (3.2)



3.2

Steps 1 – 2:



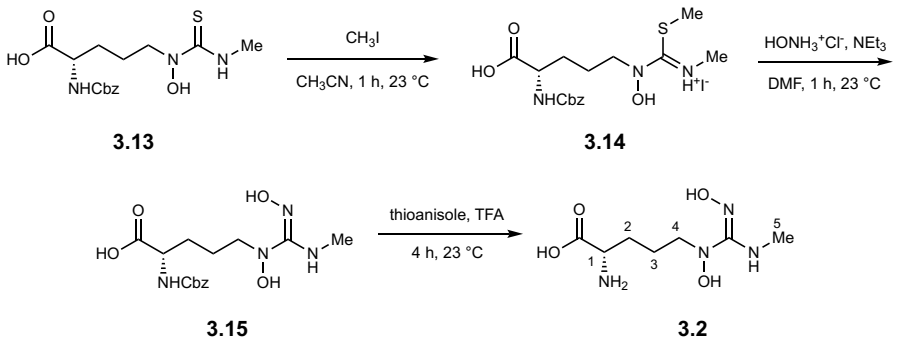
3.11 **3.12**

3.13

Cbz-protected nitrone **3.11** was prepared analogously to **3.7** with L-Cbz-ornithine hydrochloride (**11**, Chem-Impex International) as the starting material. To nitrone **3.11** (847 mg, 2.12 mmol, 1.00 equiv) dissolved in 20 mL of DMF was added hydroxylamine hydrochloride (161 mg, 2.33 mmol, 1.10 equiv) in one portion and stirred for 1 h under N_2. Then *N*-methyl isothiocyanate (CH_3NCS, 170 mg, 2.33 mmol, 1.10 equiv) dissolved in 5 mL of anhydrous CH_2Cl_2 was added to the reaction mixture in one portion, and NEt_3 (325 μL, 2.33 mmol, 1.10 equiv) was added dropwise. After stirring at room temperature for 2 h, the reaction mixture was quenched with 10 mL of 1 M HCl and extracted with ethyl acetate (3 x 10 mL). The combined organic layers were washed with 10 mL of brine, dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford crude thiourea **3.13** as a yellow oil. This compound was first purified with the CombiFlash Rf system to remove major impurities using a 150 g C18 column (RediSep Rf, Teledyne Isco). The flow rate was 75 mL/min. The LC conditions were: an isocratic hold at 5% solvent B for 2 min; a linear gradient from 5 to 100% solvent B in 11 min; an isocratic hold at 100% solvent B for 3 min; and an isocratic hold at 50% solvent B for 3 min (solvent A = 0.1% TFA in water, solvent B = 0.1% TFA in acetonitrile). **3.13** was monitored at 254 nm. The collected fractions containing **3.13** were combined and extracted with ethyl acetate (3 x 20 mL). The combined organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford the an impure thiourea **3.13** (310 mg, 41% yield) for subsequent steps.

To obtain an analytically pure sample for spectral characterization, a few miligrams of **3.13** were further purified using preparative HPLC and a Thermo Hypersil Gold aQ C18 (250 × 20 mm, 5 μm, Thermo Fisher Scientific). The flow rate was 8 mL/min. The LC conditions were: a linear gradient from 50 to 95% solvent B in 15 min; an isocratic hold at 95% solvent B for 5 min; a linear gradient from 95 to 50% solvent B in 4 min; and an isocratic hold at 50% solvent B for 4 min (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). **3.13** was monitored at 264 nm and eluted between 10.9 – 11.3 min. The collected fractions were lyophilized to afford **3.13** as a white powder. ^{1}H NMR of **3.13** (400 MHz, $CDCl_3$): δ (ppm) = 7.41 – 7.30 (m, 5H, $H_{1\text{-}3}$), 7.16 (br s, 1H, NH_{10}), 5.85 (d, J = 8.2 Hz, 1H, NH_5), 5.16 – 5.04 (m, 2H, H_4), 4.75 – 4.61 (m, 1H, H_9), 4.41 (t, J = 9.3 Hz, 1H, H_6), 3.88 – 3.79 (m, 1H, H_9), 3.10* (s, 2.5H, H_{11}), 3.01* (s, 0.5H, H_{11}), 2.24 –1.63 (m, 4H, H_7 and H_8). ^{13}C NMR of **3.13** (100 MHz $CDCl_3$): δ (ppm) = 181.0 (**C**), 174.1 (**C**), 157.4 (**C**), 135.6 (**C**), 128.7 (2 x **C**H), 128.6 (**C**H), 128.1 (2 x **C**H), 67.8 (**C**H_2), 54.2 (**C**H), 53.6 (**C**H), 32.2 (**C**H_2), 31.7 (**C**H_3), 22.1 (**C**H_2). HRMS: Calc'd for $C_{15}H_{21}N_3O_5SNa$ $[M+Na]^+$ = 378.1100, found = 378.1092. *rotamers

Step 3 – 5:

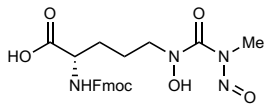


Thiourea **3.13** (297 mg, 836 μmol, 1.00 equiv) was dissolved in 10 mL of acetonitrile (CH_3CN), and methyl iodide (CH_3I, 57.3 μL, 917 μmol, 1.10 equiv) was added dropwise to the stirring solution. The reaction mixture was stirred for 1 h under N_2 at room temperature. The

reaction was then concentrated *in vacuo* and resuspended in 10 mL of acetonitrile. Hydroxylamine hydrochloride (63.9 mg, 917 μmol, 1.10 equiv) was added to the solution in one portion followed by NEt_3 (128 μL, 917 μmol, 1.10 equiv) and 2 mL of DMF. This reaction mixture was stirred under N_2 for 2 h until conversion to the guanidine **3.15** was confirmed by mass spectrometry. The reaction was acidified with 100 μL of TFA and concentrated *in vacuo*. The residue was applied directly to a C18 Sep-Pak cartridge. After washing the cartridge with 100 mL of water + 0.1% formic acid and 100 mL of 5% acetonitrile in water + 0.1% formic acid, **3.15** was eluted with 200 mL of 15% acetonitrile in water + 0.1% formic acid. The fractions were combined, concentrated *in vacuo* to remove acetonitrile, and the resulting aqueous layer was lyophilized to dryness to afford **3.15** as a white solid (13 mg, 5% yield), which was used immediately without characterization.

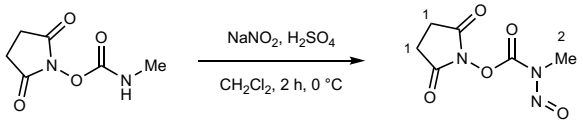
The Cbz group was removed by dissolving **3.15** in 2 mL of 10% (*v*/*v*) thioanisole in TFA. After stirring for 4 h at room temperature, 5 mL of water was added to the reaction mixture. The aqueous solution containing **3.2** was first washed with 2 x 5 mL of diethyl ether, concentrated, and purified with preparative HPLC using the same column, solvents, and method as were used to purify amino acid **3.1**. This afforded 4 mg of amino acid **3.2** (49% yield from **3.15**) as a white solid. ^{1}H NMR of **3.2** (500 MHz D_2O): δ (ppm) = 3.80 (t, J = 6.0 Hz, 1H, H_1), 3.53 (t, J = 6.9 Hz, 2H, H_4), 2.98 (s, 3H, H_5), 2.03 – 1.74 (m, 4H, H_2 and H_3).^{13}C NMR of **3.2** (150 MHz D_2O): δ (ppm) = 174.3 (**C**), 162.3 (**C**), 54.5 (**C**H), 54.4 (**C**H_2), 29.5 (**C**H_2), 27.3 (**C**H_3), 21.7 (**C**H_2). HRMS: Calc'd for $C_7H_{15}N_4O_4$ $[M-H]^-$ =219.1099, found = 219.1101.

Synthesis of Fmoc-N^{δ}-hydroxy-N^{ω}-methyl-N^{ω}-nitroso-L-citrulline (3.18)



3.18

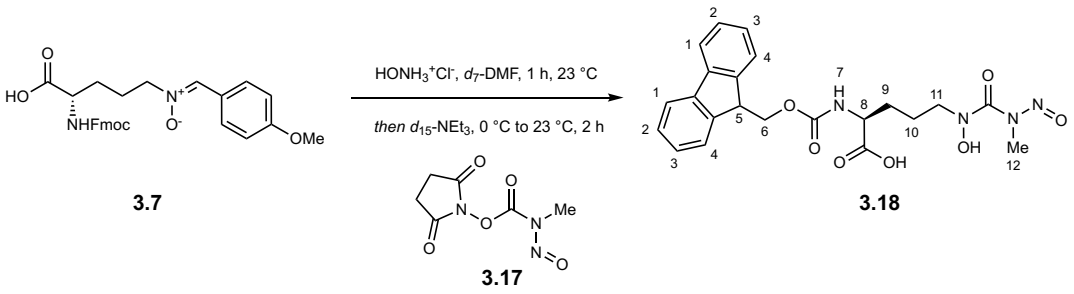
Step 1:



3.16 **3.17**

To *N*-succinimidyl *N*-methylcarbamate (**3.16**, 3.31 mmol, 570 mg, 1.00 equiv) in 25 mL of CH_2Cl_2 was added 200 µL of concentrated sulfuric acid (H_2SO_4). The flask was cooled to 0 ºC, and sodium nitrite ($NaNO_2$) (343 mg, 4.97 mmol, 1.50 equiv) dissolved in 5 mL of water was added dropwise over 5 min into the stirring solution. After 1 h, another equivalent of $NaNO_2$ and H_2SO_4 were added, and stirring was continued at 0 ºC for 1 h. 10 mL of water was then added to the reaction mixture, and N_2 was bubbled through the solution for 5 min to remove NO_2. The biphasic solution was extracted with 3 x 50 mL of CH_2Cl_2. The combined organic layers were dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford crude **3.17**, which was purified using the CombiFlash Rf system with a 40 g silica column (RediSep Rf Gold, Teledyne Isco). The flow rate was 40 mL/min. The LC conditions were: an isocratic hold at 0% solvent B for 2 min; a linear gradient from 0 to 100% solvent B in 13 min; an isocratic hold at 100% solvent B for 3 min (solvent A = hexanes, solvent B = ethyl acetate). **3.17** was monitored at 254 nm. The fractions containing **3.17** were pooled and concentrated *in vacuo* to obtain 32 mg of **3.17** as a light yellow powder (22% yield). ^{1}H NMR (600 MHz $CDCl_3$): δ (ppm) = 3.22 (s, 3H, H_2), 2.94 (s, 4H, H_1). ^{13}C NMR (125 MHz $CDCl_3$): δ (ppm) = 168.3 (**C**), 151.0 (**C**), 28.5 (**C**H_3), 25.6 (2 x **C**H_2).

Step 2:

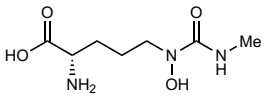


While we can synthesize **3.18** from **3.8** and **3.17**, we have been unsuccessful in isolating analytically pure **3.18** for characterization. Despite numerous attempts to isolate **3.18** by semi-preparative HPLC using different purification conditions, **3.18** readily degraded to the corresponding denitrosated product **3.19** and other uncharacterized products. We have obtained an exact mass, a diagnostic MS/MS spectrum, and used labeled materials to confirm the structure of **3.18** generated in biochemical assays with SznF (**Figure 3.7**). We have also obtained NMR data for **3.18**, by performing the coupling reaction in an NMR tube using deuterated base and solvent.

To a solution of nitrone **3.7** (6.00 mg, 12.3 μmol, 1.00 equiv) dissolved in 450 μL of d_7-DMF (Sigma Aldrich) was added 6.66 μL of a 1 mM stock solution of hydroxylamine hydrochloride in d_7-DMF (1.0 mg, 14.8 μmol, 1.20 equiv). After incubating the reaction mixture in a 5 mm NMR tube for 1 h at room temperature, d_{15}-NEt_3 (Cambridge Isotope Inc.; 2.07 μL,14.8 μmol, 1.20 equiv) was added, followed by **3.17** (2.47 mg, 12.3 μmol, 1.00 equiv) dissolved in 50 μL of d_7-DMF. The reaction mixture was incubated for 30 min at room temperature, and the conversion from **3.17** to **3.18** was analyzed by ^{1}H NMR. The integration of product peaks cannot be accurately obtained since the conversion was not complete, and the overlapping peaks from **3.8**, **3.18**, and other intermediates made integration assignments difficult. ^{1}H NMR of **3.18** (400 MHz d_7-DMF): δ (ppm) = 7.96 – 7.91 (m, H_1), 7.82 – 7.75 (m), 7.49 – 7.32 (m), 4.32 – 4.01 (m, H_5, H_6, and H_8), 3.74 (t, J = 5.5 Hz, H_{11}), 3.15 (s, H_{12}), 2.02 – 1.59 (m, H_9 and H_{10}). ^{13}C NMR of **3.18**

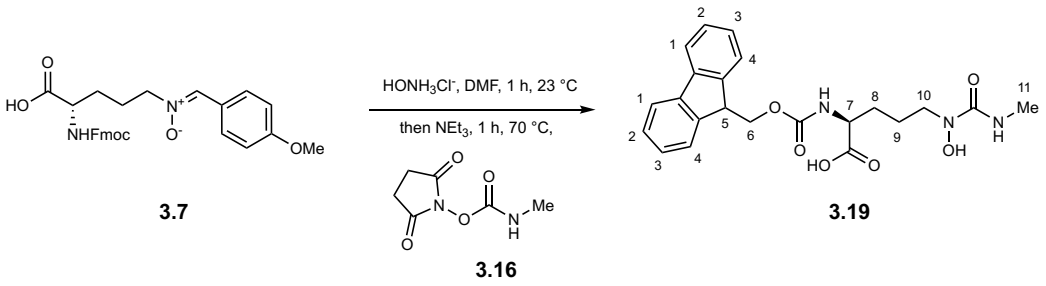
(100 MHz d_7-DMF): δ (ppm) = 175.5 (**C**), 163.5 (**C**), 155.5 (**C**), 144.4 (2 x **C**), 141.3 (2 x **C**), 127.8 (2 x **C**H), 127.3 (2 x **C**H), 126.0 (2 x **C**H), 120.3 (2 x **C**H), 66.4 (**C**H_2), 55.1 (**C**H), 50.7 (**C**H_2), 47.3 (**C**H), 30.5 (**C**H_3), 29.6 (**C**H_2), 23.4 (**C**H_2). HRMS for **3.18**: Calc'd for $C_{22}H_{23}N_4O_7$ $[M-H]^-$ = 455.1572, found = 455.1588.

Synthesis of N^{δ}-hydroxy-N^{ω}-methyl-L-citrulline (3.4)



3.4

Step 1:

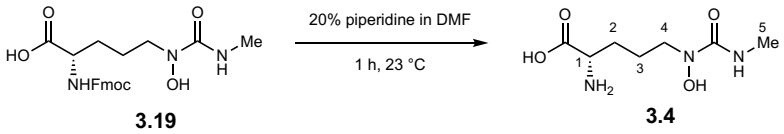


3.7 **3.16** **3.19**

To nitrone **3.7** (100 mg, 0.205 mmol, 1.00 equiv) dissolved in 20 mL of DMF was added hydroxylamine hydrochloride (17.0 mg, 0.246 mmol, 1.20 equiv) in one portion. The resulting reaction mixture was stirred for 1 h under N_2. Then, *N*-succinimidyl-*N*-methylcarbamate (**3.16**, 38.7 mg, 2.26 μmol, 1.10 equiv) and NEt_3 (57.0 μL, 0.410 mmol, 2.00 equiv) were added in one portion, and stirring was continued for 16 h at room temperature. 50 mL of ethyl acetate were added to the reaction mixture, the layers were separated, and the organic layer was washed with 10 mL of 1 M HCl, 10 mL of water, and 10 mL of brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to dryness. The residue was dissolved in methanol and purified using a C18 Sep-pak cartridge. After an initial wash with 100 mL of water + 0.1% formic acid to remove salts, crude **3.19** was eluted with 200 mL of 30% acetonitrile + 0.1% formic acid. The

desired fraction was concentrated *in vacuo* to remove the acetonitrile, affording an aqueous solution containing **3.19**. The aqueous solution was then extracted with 2 x 50 mL of ethyl acetate. The combined organic layer was dried over Na_2SO_4, filtered, and evaporated to afford a reaction mixture of **3.19** and *p*-anisaldeyhde (75:25), which was carried forward without further purification. ^{1}H NMR (600 MHz d_7-DMF): δ (ppm) = 7.93 (d, J = 7.5 Hz, 2H, H_1), 7.61 – 7.58 (m, 2H, H_4), 7.44 (t, J = 7.4 Hz, 2H, H_2), 7.38 – 7.33 (m, 2H, H_3), 4.35 – 4.27 (m, 3H, H_5 and H_6), 4.24 – 4.19 (m, 1H, H_7), 3.51 (t, J = 6.4 Hz, 2H, H_{10}), 2.75 (d, J = 4.7 Hz, 3H, H_{11}), 2.00 – 1.72 (m, 4H, H_8 and H_9). ^{13}C NMR (600 MHz d_7-DMF): δ (ppm) = 174.2 (**C**), 160.8 (**C**), 156.7 (**C**), 144.4 (2 x **C**), 144.3 (2 x **C**), 128.1 (2 x **C**H), 127.8 (2 x **C**H), 127.3 (2 x **C**H), 120.3 (2 x **C**H), 66.5 (**C**H_2), 55.1 (**C**H_3), 54.59 (**C**H), 50.4 (**C**H_2), 47.2 (**C**H), 29.1 (**C**H_2), 23.9 (**C**H_2). HRMS for **3.19**: Calc'd for $C_{22}H_{24}N_3O_6$ $[M-H]^-$ = 426.1671, found = 426.1679.

Step 2:

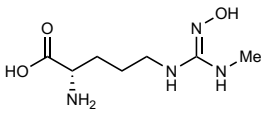


3.19 **3.4**

Crude urea **3.19** from the previous step was dissolved in 10 mL of 20% (*v/v*) piperidine in DMF and stirred for 4 h at room temperature. The solvent was removed *in vacuo*, and the solids were dissolved in 5 mL of water. The aqueous solution containing crude **3.4** was washed with 2 x 5 mL of diethyl ether to remove nonpolar impurities and then applied to reverse phase semi-preparative HPLC using a Kromasil 100 C18 column (250 x 10 μm, 5mm). The flow rate was 3 mL/min. The LC conditions were: 5 to 12% solvent B in 10 min; a linear gradient from 12 to 30% solvent B in 2.5 min; a linear gradient from 30 to 5% solvent B in 1.5 min; and an isocratic hold at 5% solvent B for 3 min (solvent A = 0.1 % formic acid in water, solvent B = 0.1% formic acid

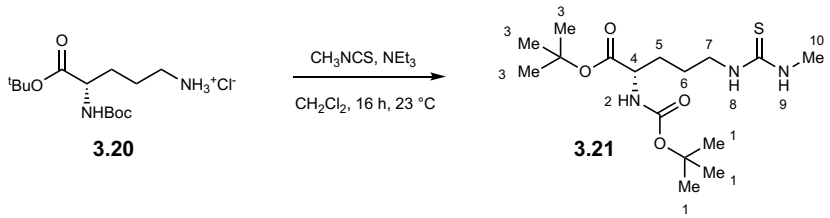
in acetonitrile). **3.4** was monitored at 200 nm and eluted between 4.6 – 5.2 min. The collected fractions were lyophilized to afford 14 mg of **3.4** (33% yield from **3.8**). ^{1}H NMR of **3.4** (600 MHz D_2O): 3.97 (t, J = 6.3 Hz, 1H, H_1), 3.52 (t, J = 6.7 Hz, 2H, H_4), 2.74 (s, 3H, H_5), 2.02 – 1.88 (m, 2H, H_2), 1.78 – 1.67 (m, 2H, H_3). ^{13}C NMR of **3.4** (125 MHz DMSO): δ (ppm) = 171.4 (**C**), 161.8 (**C**), 52.3 (**C**H), 50.1 (**C**H_2), 27.9 (**C**H_2), 26.9 (**C**H_3), 22.4 (**C**H_2). HRMS: Calc'd for $C_7H_{14}N_3O_4$ $[M-H]^-$ = 204.0990, found = 204.0991.

Synthesis of N^{ω}-hydroxy-N^{ω}-methyl-L-arginine (3.5)



3.5

Step 1:

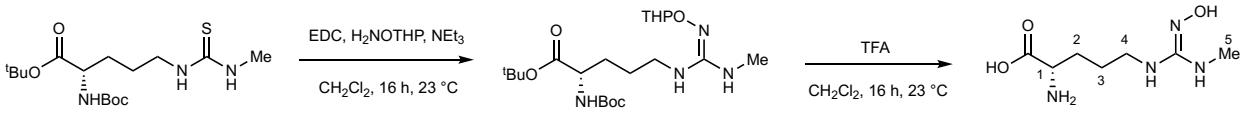


3.20 **3.21**

Compound **3.5** was prepared by adapting the procedures of Jansen Labby *et al*.[33] The protected amino acid **3.20** (Bachem, 200 mg, 617 μmol, 1.00 equiv) was dissolved in 50 mL of anhydrous CH_2Cl_2 at 0 ºC, and NEt_3 (172 μL, 1.23 mmol, 2.00 equiv) was added to the solution in one portion. A solution of methyl isothiocyanate (CH_3NCS, 45.1 mg, 617 μmol, 1.00 equiv) in 5 mL of CH_2Cl_2 was added dropwise to the reaction mixture, and the resulting reaction mixture was allowed to warm to room temperature while stirring overnight. The reaction mixture was washed with 1 x 10 mL of 1M HCl, 1 x 10 mL of saturated sodium bicarbonate, 1 x 10 mL of water, and 1 x 10 mL of brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo*

to afford crude **3.21**. The compound was purified with silica gel chromatography (70:30 hexanes: ethyl acetate) to afford **3.21** (155 mg, 69.5% yield). ^{1}H NMR of **3.21** (600 MHz $CDCl_3$): δ 6.40 (br s, 2H, NH_8 and NH_9), 5.24 (d, J = 7.9 Hz, 1H, NH_2), 4.17 – 4.08 (m, 1H, H_4), 3.57 – 3.43 (m, 2H, H_7), 3.02 (s, 3H, H_{10}), 1.85 – 1.60 (m, 4H, H_5 and H_6), 1.46 (d, J = 1.9 Hz, 9H), 1.43 (d, J = 2.0 Hz, 9H). Calc'd for $C_{16}H_{32}N_3O_4S$ $[M+H]^+$ = 362.2108, found = 362.2086.

Step 2 – 3:



3.21 **3.22** **3.5**

To thiourea **3.21** (96.0 mg, 169 μmol, 1.00 equiv) dissolved in 20 mL of CH_2Cl_2 was added EDC (28.9 mg, 186 μmol 1.10 equiv), NEt_3 (47.1 μL, 338 μmol, 2.00 equiv), and 4-(aminooxy)tetrahydro-2*H*-pyran (H_2NOTHP, Santa Cruz Biotechnology; 19.8 mg, 169 μmol, 1.00 equiv) in one portion. After 16 h of stirring at room temperature, the reaction mixture was washed with 1 x 10 mL of brine. The layers were separated, and the organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to afford a light yellow oil. The crude oil was dissolved in 20 mL of CH_2Cl_2 and 5 mL of TFA, and the resulting reaction mixture was stirred at room temperature for 16 h. The reaction mixture was then concentrated *in vacuo*. The residue was dissolved in 5 mL of water and the aqueous solution was washed with 2 x 5 mL of diethyl ether to remove nonpolar impurities. The crude product in the aqueous layer was purified by preparative HPLC using the same column and method as described for purifying **3.1**. Lyophilizing the collected fractions afforded **3.5** as a white solid (11 mg, 32% yield). ^{1}H NMR of **3.5** (500 MHz, D_2O): δ 3.78 (t, J = 6.1 Hz, 1H, H_1), 3.26 (t, J = 6.9 Hz, 2H, H_4), 2.84 (s, 3H, H_5), 1.96 – 1.87 (m, 2H, H_2), 1.80 – 1.62 (m, 2H, H_3). ^{13}C NMR of **3.5** (100 MHz, D_2O): δ 174.4 (**C**), 157.8 (**C**), 54.3 (**C**H), 40.1 (**C**H_2),

27.6 (**C**CH_2), 27.0 (**C**H_3), 23.8 (**C**H_2). HRMS: Calc'd for $C_7H_{15}N_4O_3$ $[M-H]^-$ = 203.1150, found = 203.1159.

Figure 3.27: NMR spectra of d_3-L-NMA. ^{1}H NMR of d_3-L-NMA (recorded in D_2O at 600 MHz).

A

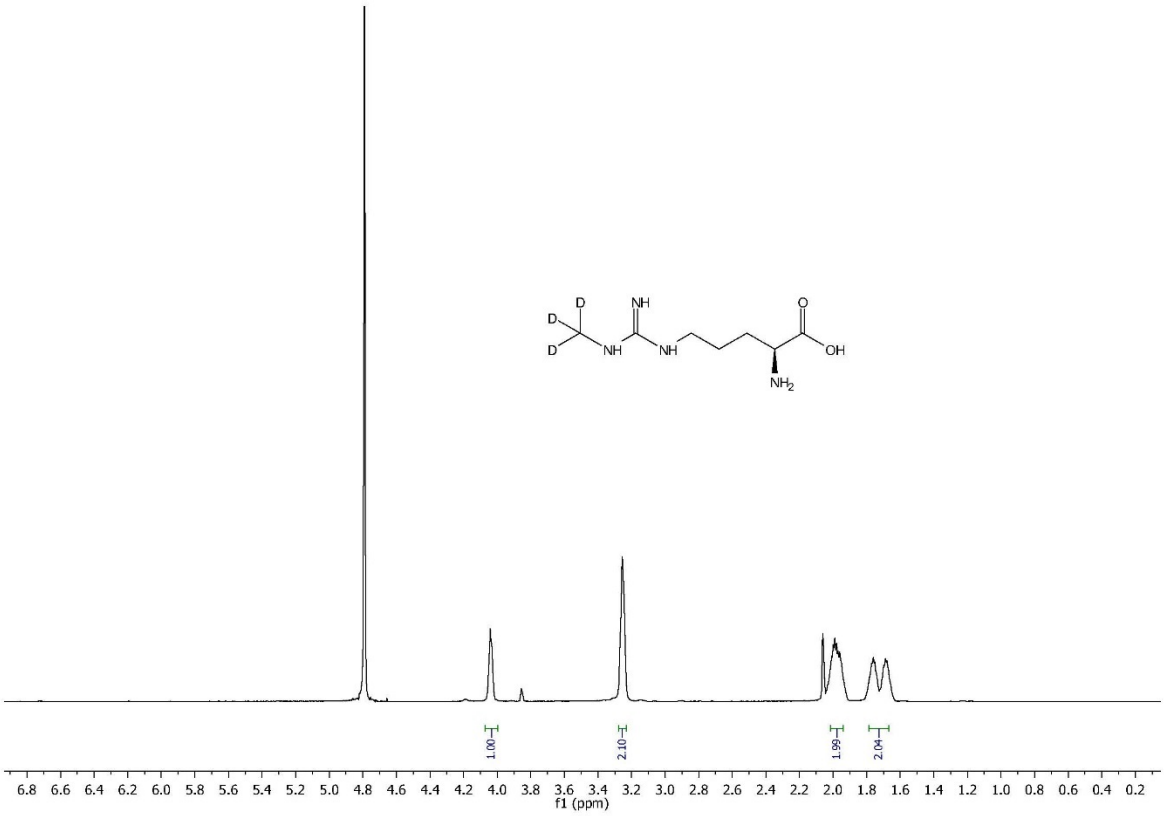
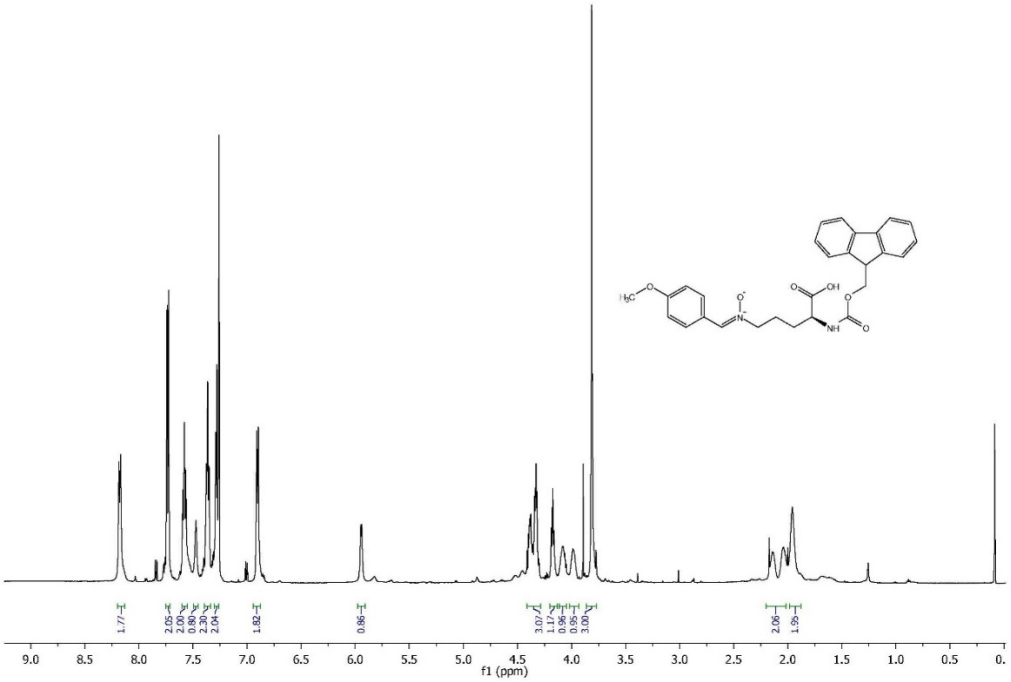


Figure 3.28: NMR spectra of **3.7**. **A**) ^{1}H NMR spectrum of **3.7** (recorded in $CDCl_3$ at 600 MHz). **B**) ^{13}C NMR spectrum of **3.7** (recorded in $CDCl_3$ at 125 MHz).

A



B

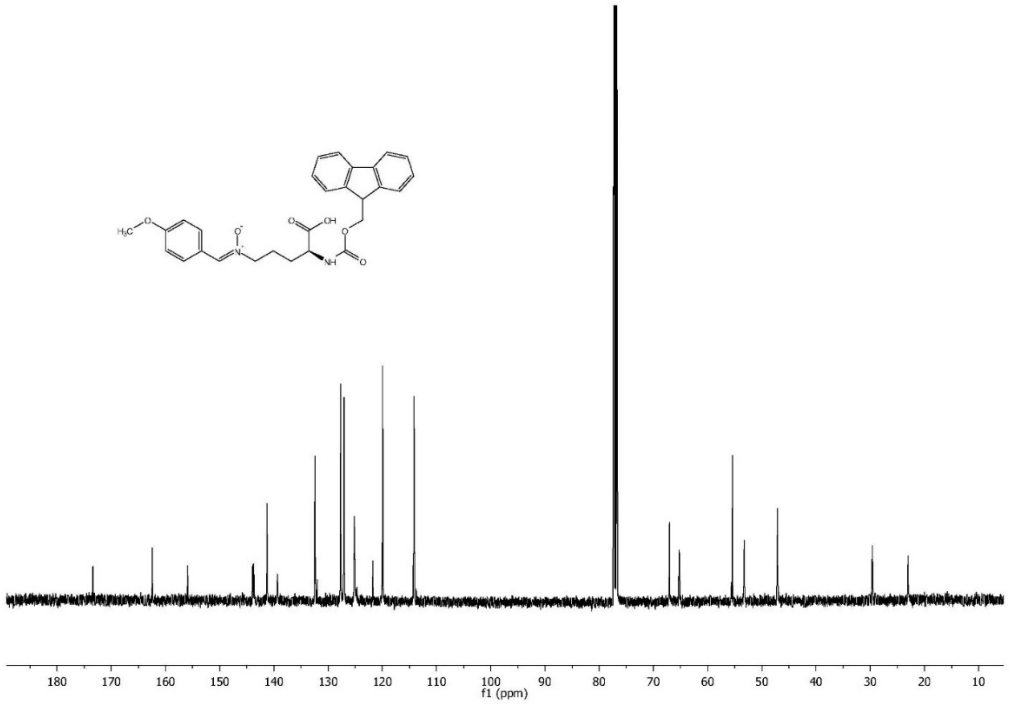
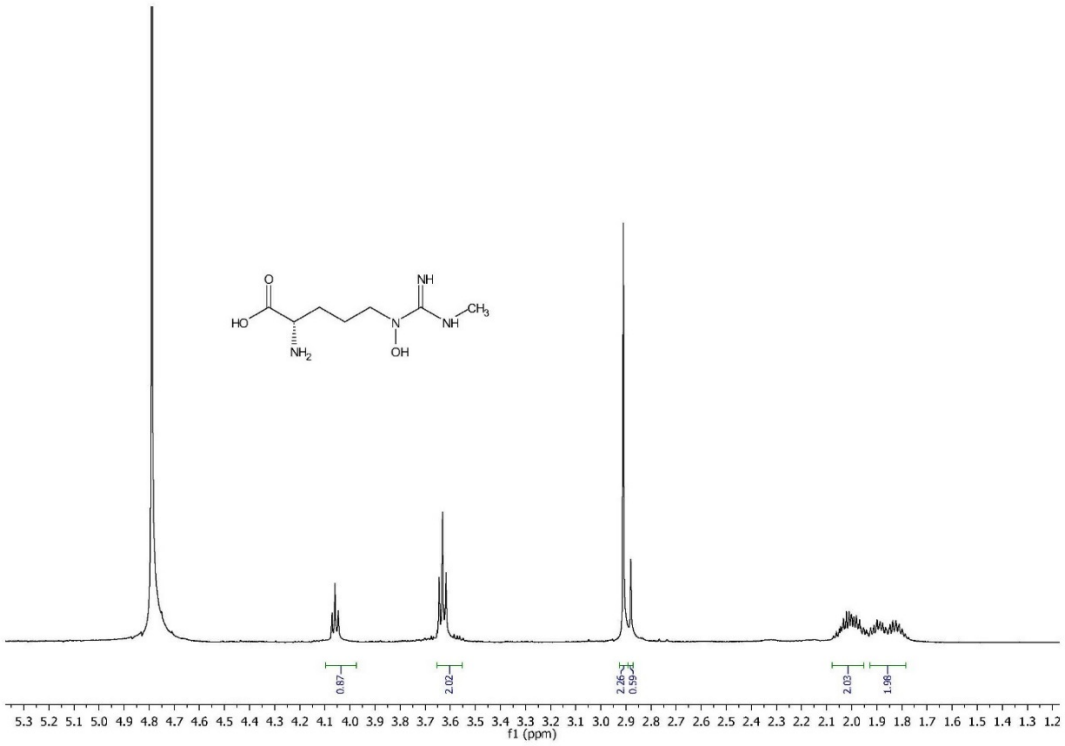


Figure 3.29: NMR spectra of **3.1**. **A**) ^{1}H spectrum of **3.1** (recorded in D_2O at 500 MHz). **B**) ^{13}C spectrum of **3.1** (recorded in D_2O at 125 MHz).

A



B

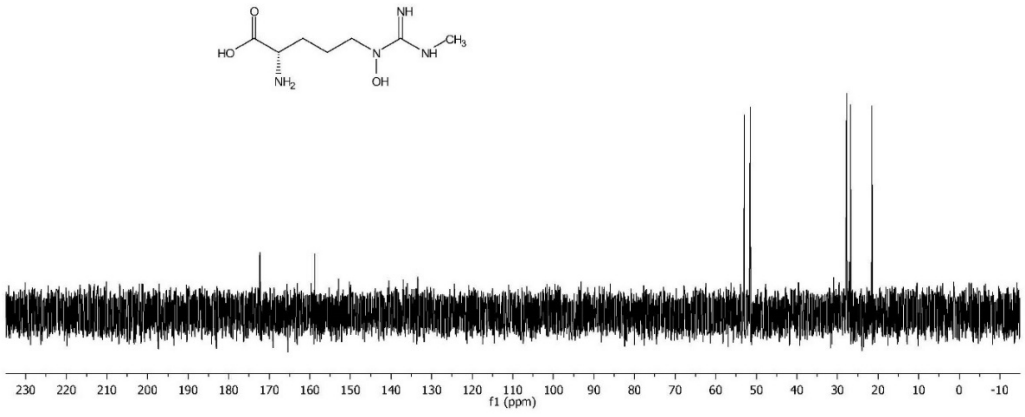
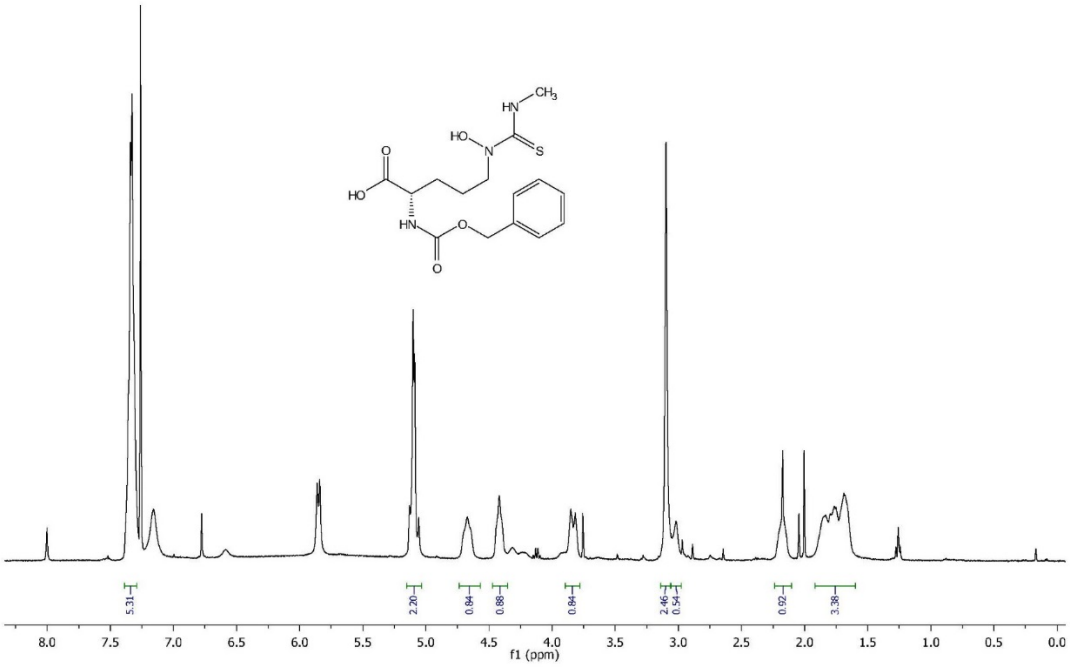
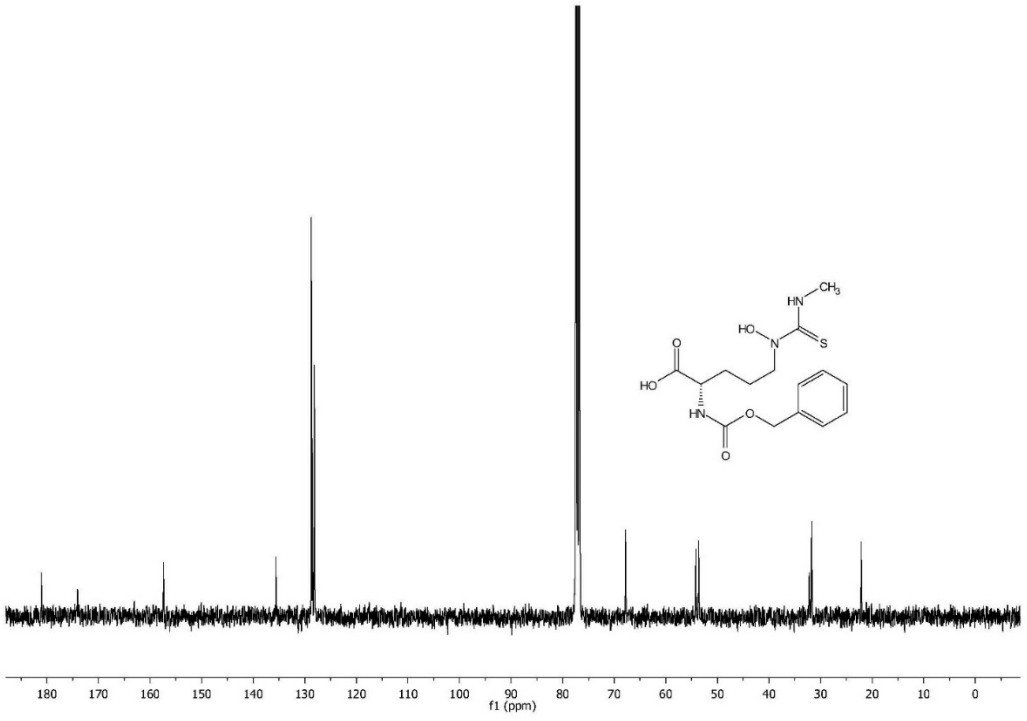


Figure 3.30: NMR spectra of **3.13**. **A**) ^{1}H spectrum of **3.13** (recorded in $CDCl_3$ at 400 MHz). **B**) ^{13}C NMR spectrum of **3.13** (recorded in $CDCl_3$ at 100 MHz). **C**) ^{1}H–^{13}C HSQC spectrum of **3.13** (recorded in $CDCl_3$ at 400 MHz and 100 MHz).

A



B


HN
CH3
HO
N
S
O
HO
HN
O
O
180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
f1 (ppm)

C

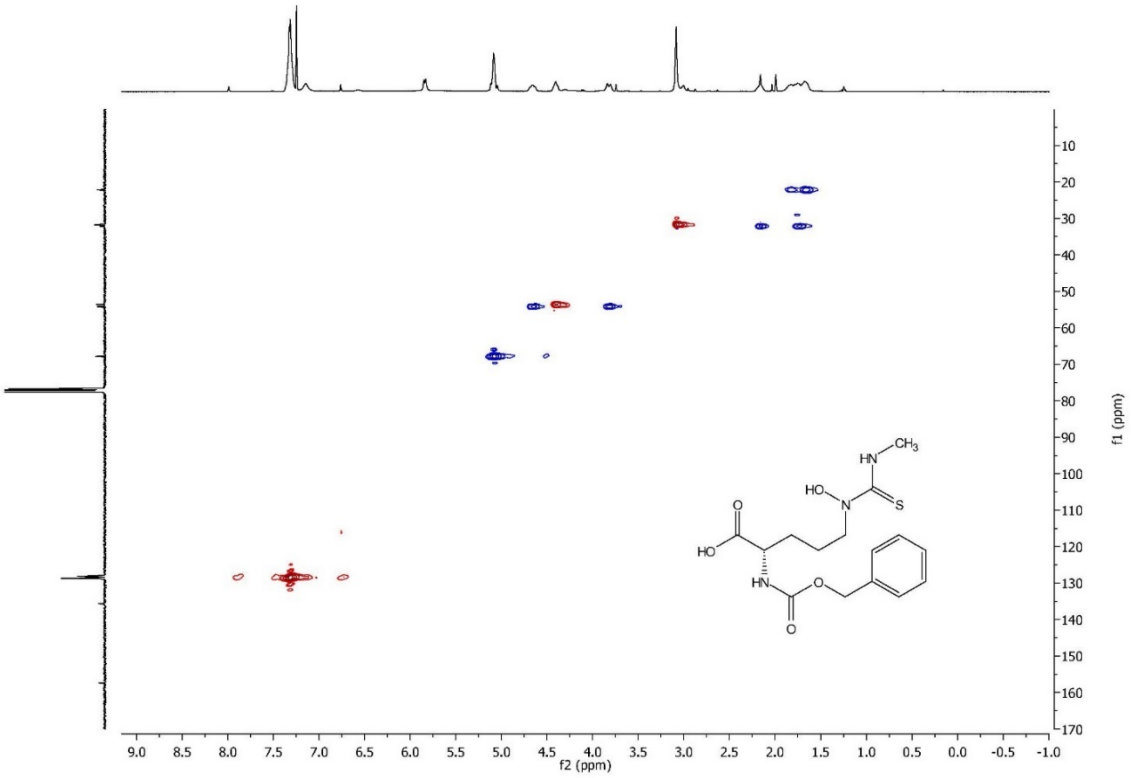
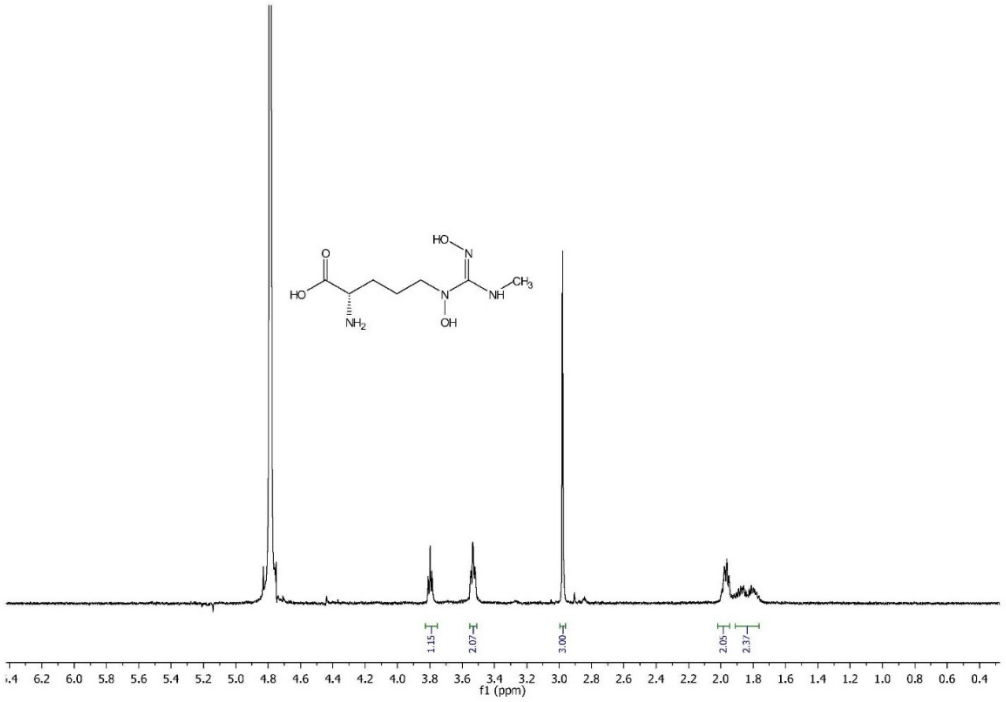

HN
CH3
HO
N
S
O
HO
HN
O
O
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170
f1 (ppm)
9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f2 (ppm)

Figure 3.31: NMR spectra of **3.2**. **A**) ^{1}H spectrum of **3.2** (recorded in D_2O at 500 MHz). **B**) ^{13}C spectrum of **3.2** (recorded in D_2O at 150 MHz).

A



B

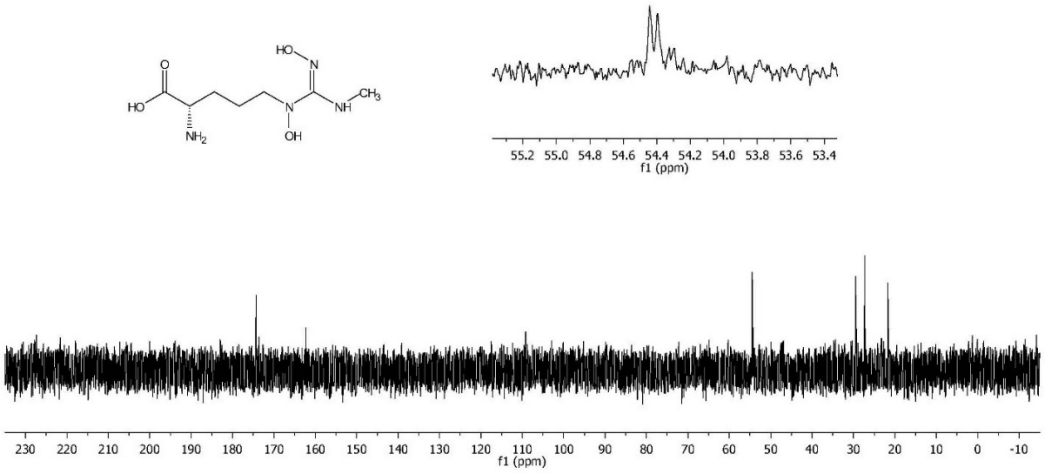
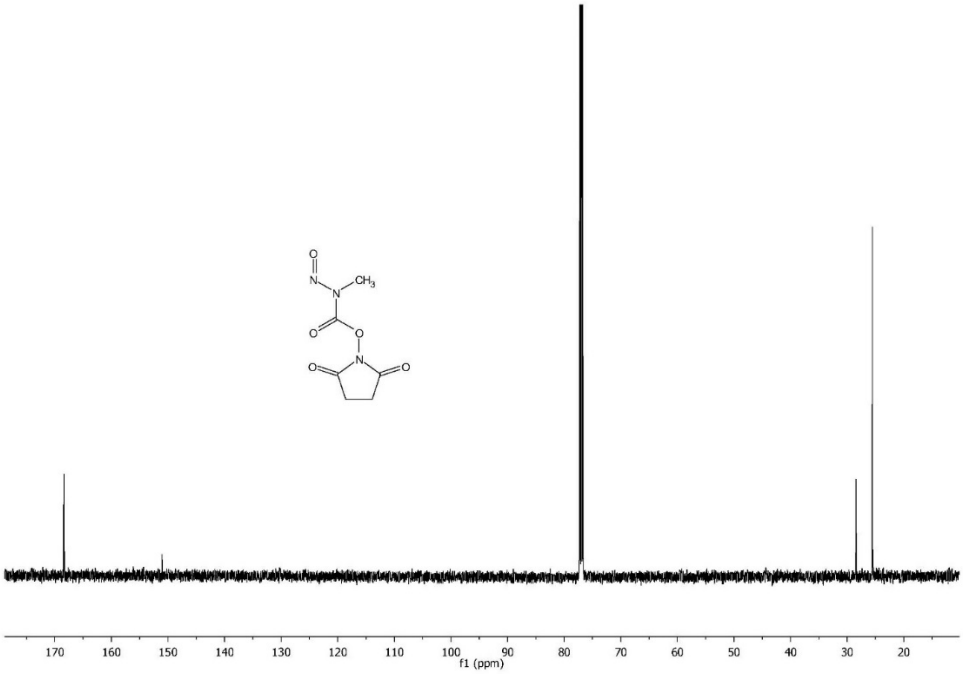


Figure 3.32: NMR spectra of **3.17**. **a**, ^{1}H spectrum of **3.17** (recorded in $CDCl_3$ at 600 MHz). **b**, ^{13}C spectrum of **3.17** (recorded in $CDCl_3$ at 125 MHz).

A



B

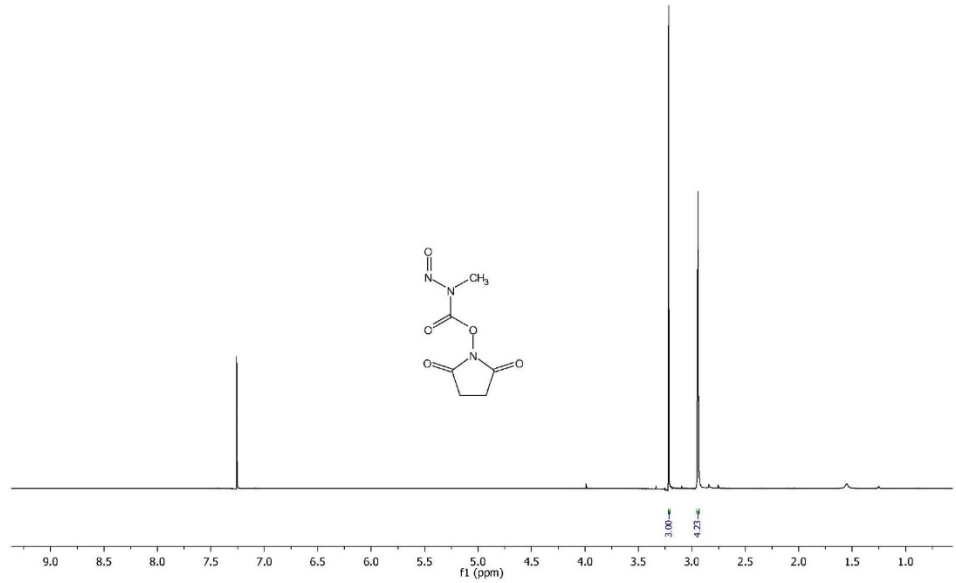
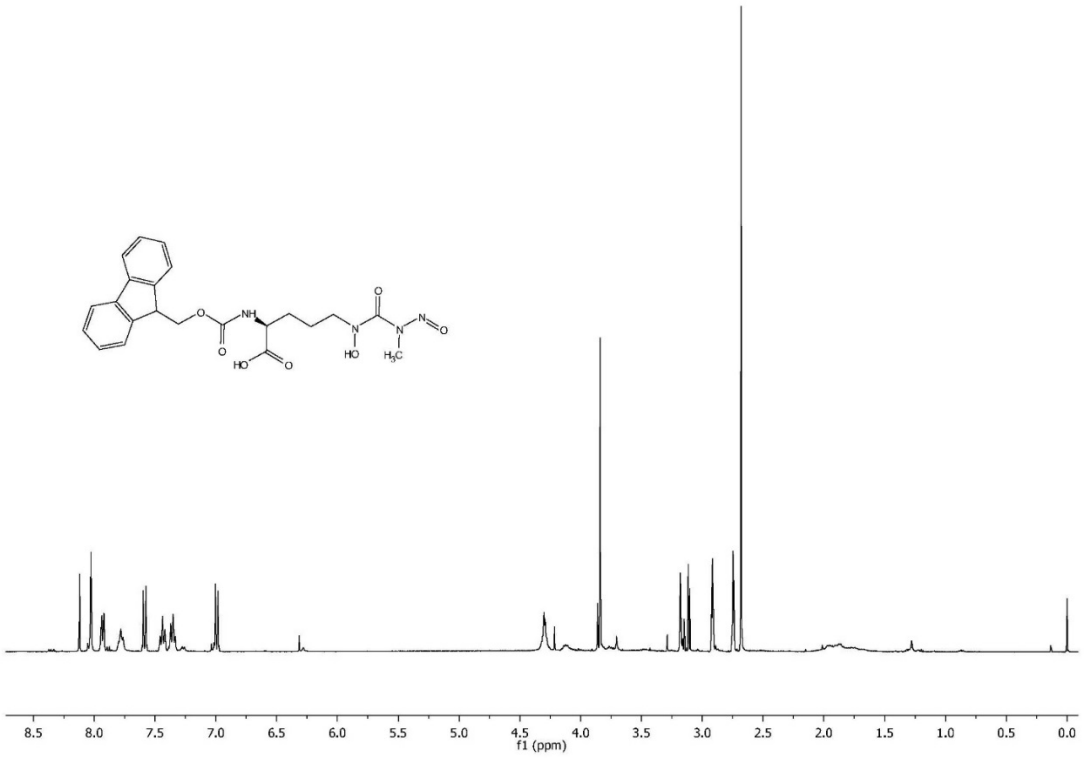
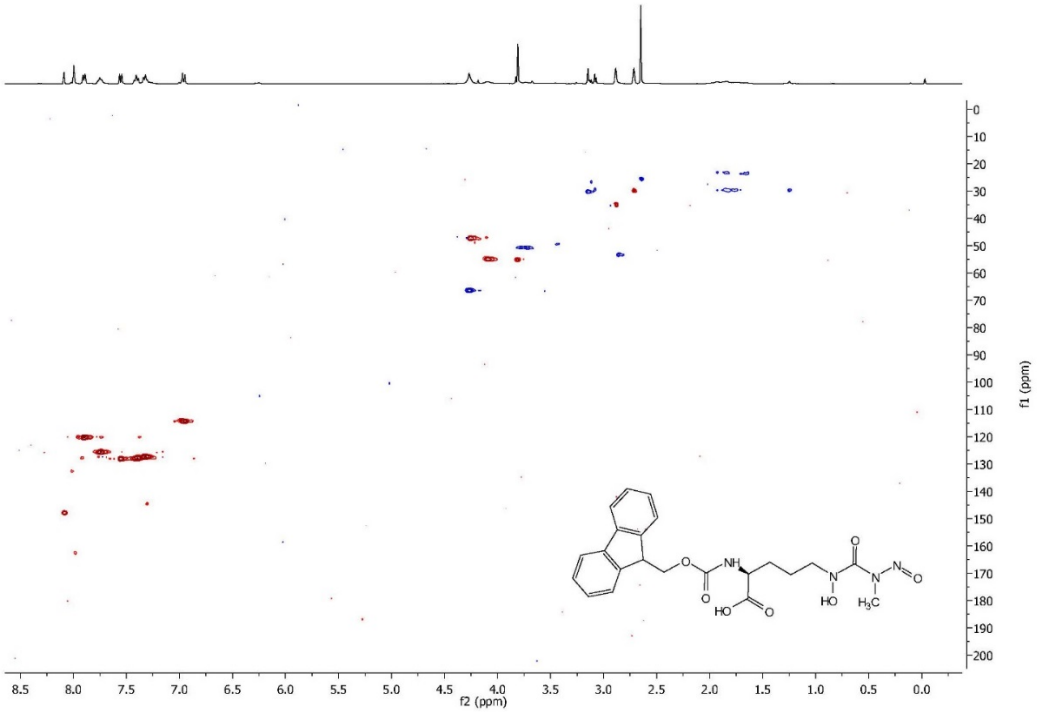


Figure 3.33: NMR spectra of **3.18**. **A**, ^{1}H spectrum of **3.18** (recorded in d_7-DMF at 400 MHz). **B**, ^{1}H–^{13}C HSQC spectrum of **3.18** (recorded in d_7-DMF at 400 MHz and 100 MHz). **C**, ^{1}H–^{13}C HMBC spectrum of **3.18** (recorded in d_7-DMF at 400 MHz and 100 MHz). **D**, ^{1}H–^{1}H COSY spectrum of **3.18** (recorded in d_7-DMF at 400 MHz).

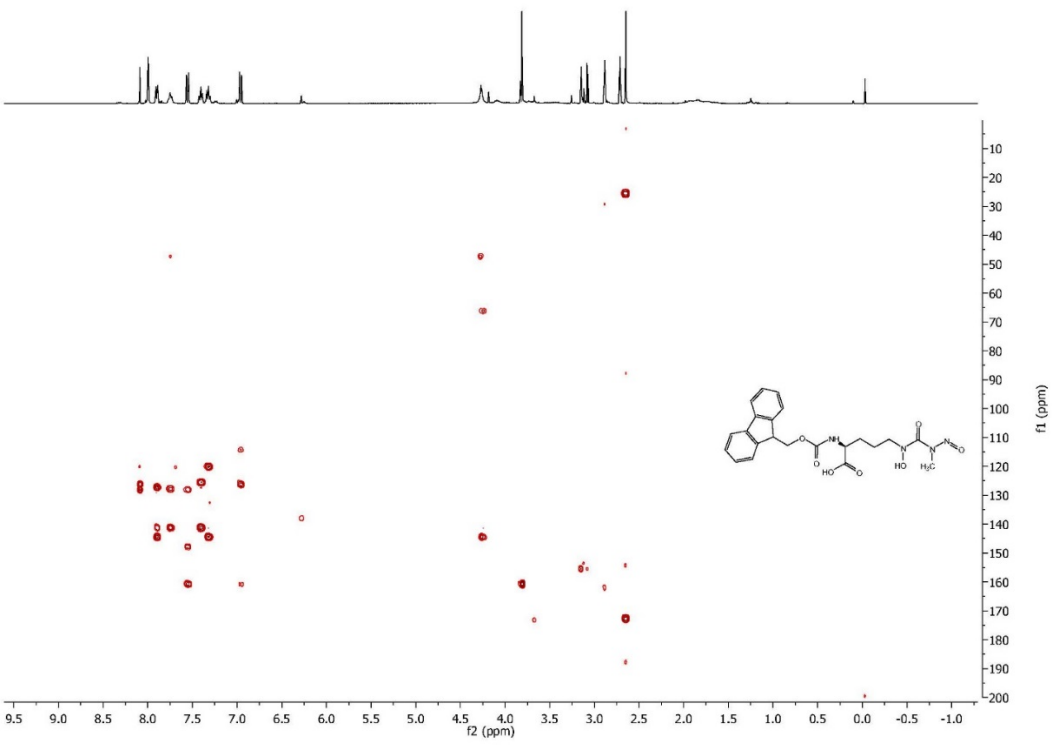
A



B



C



D

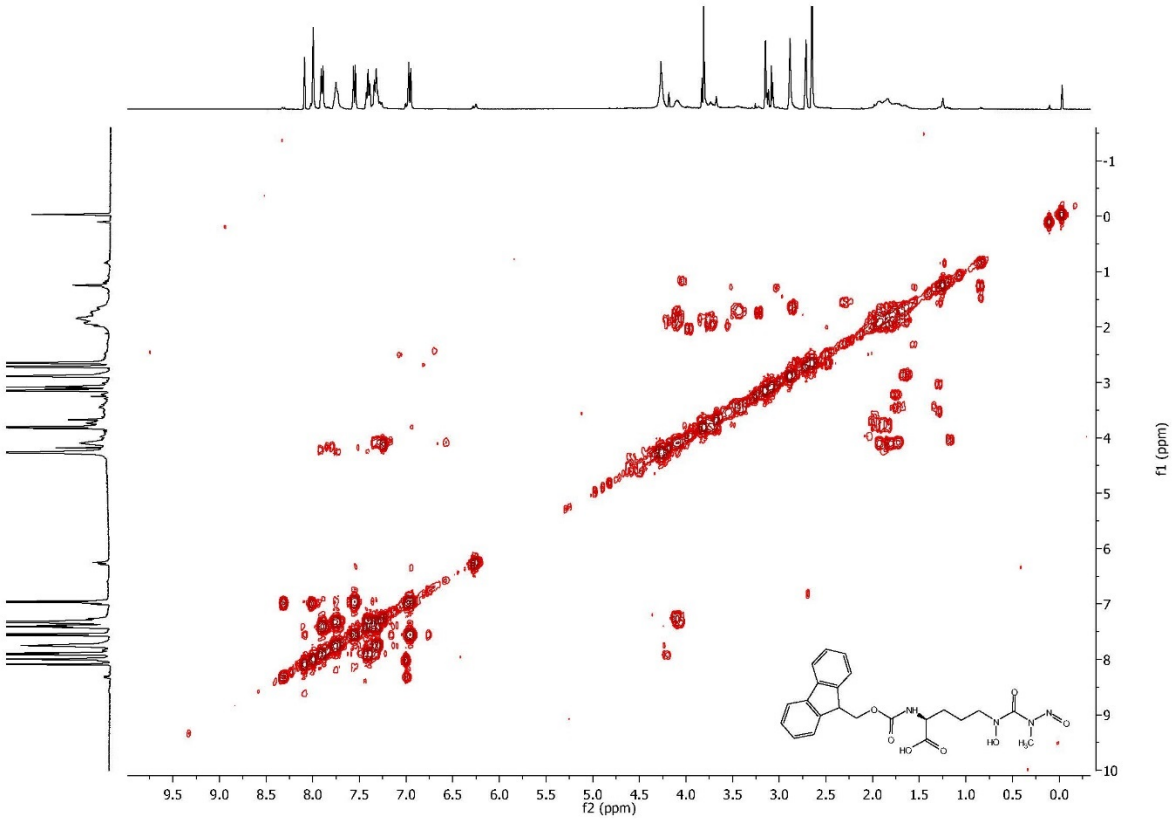
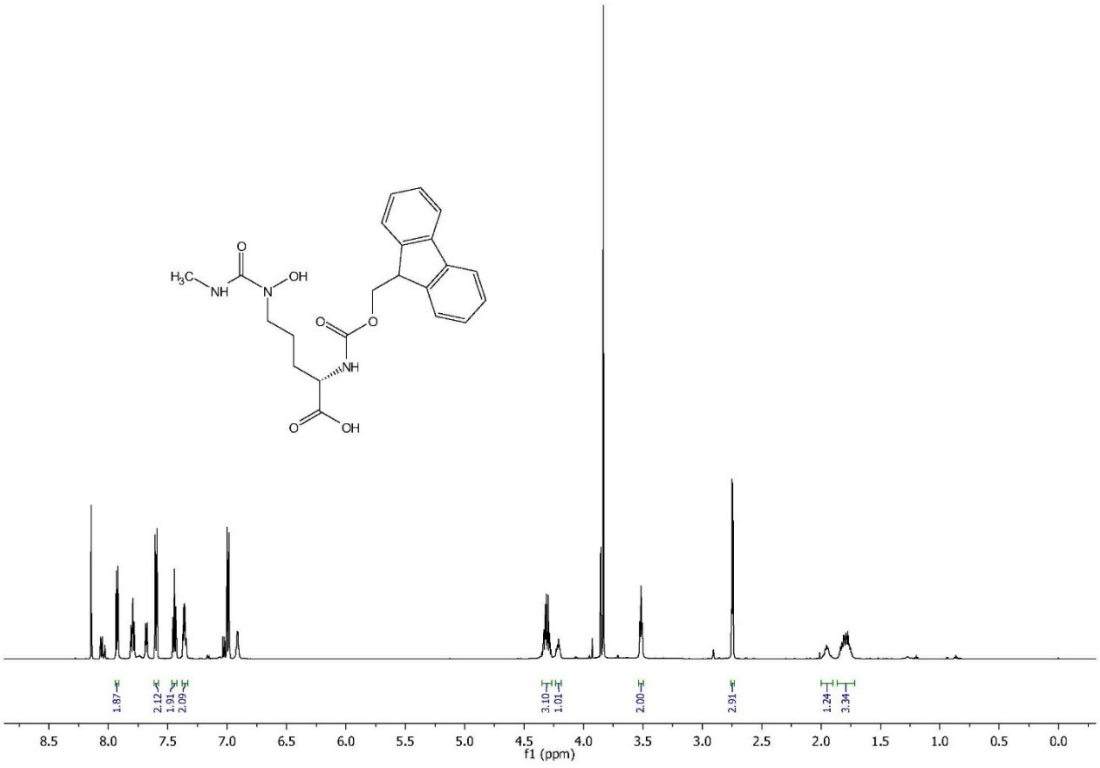
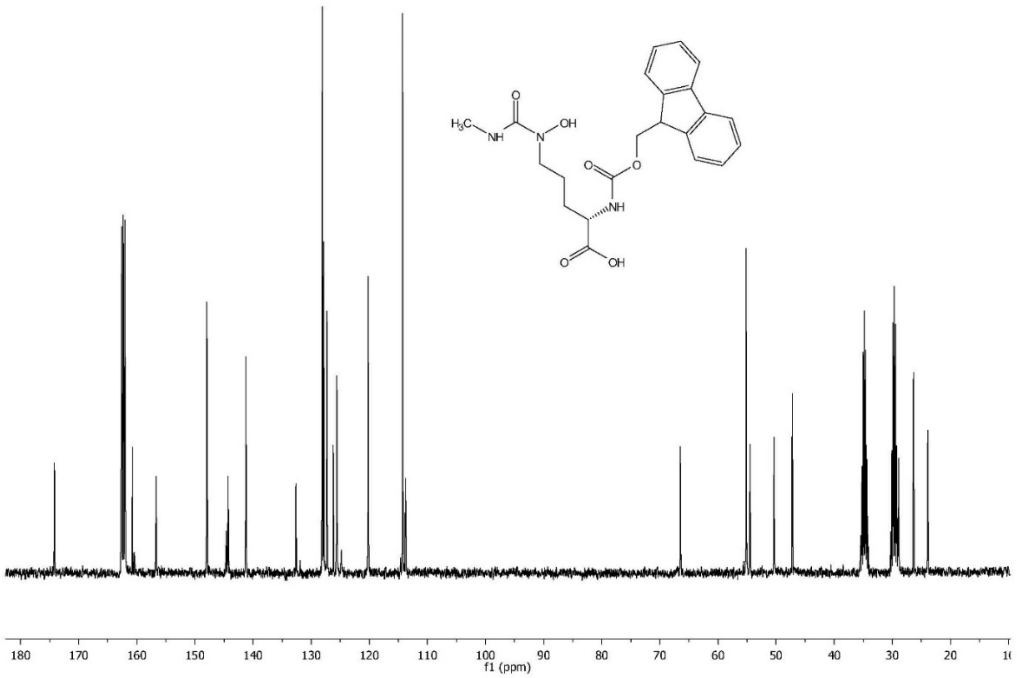


Figure 3.34: NMR spectra of **3.19**. **A**) ^{1}H spectrum of **3.19** (recorded in d_7-DMF at 600 MHz). **B**) ^{13}C spectrum of **3.19** (recorded in d_7-DMF at 100 MHz). **C**) ^{1}H–^{13}C HSQC spectrum of **3.19** (recorded in d_7-DMF at 400 MHz and 100 MHz).

A



B



C

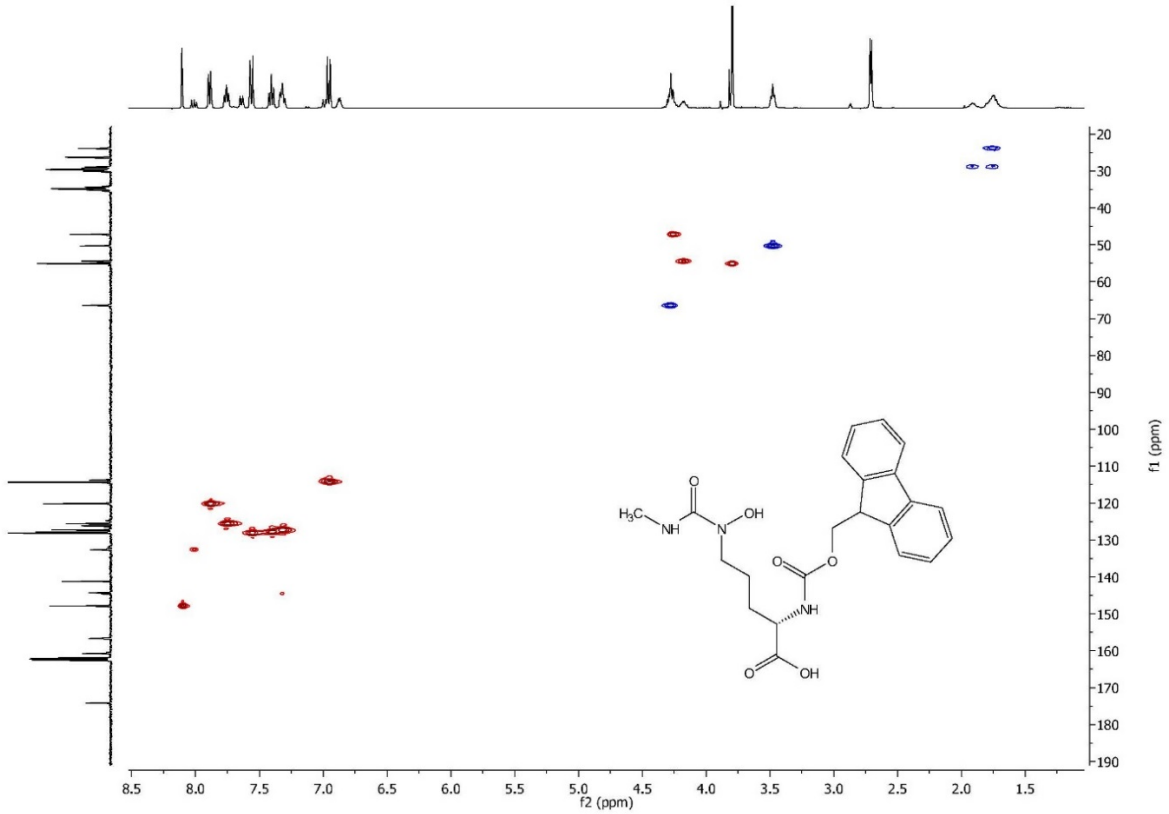
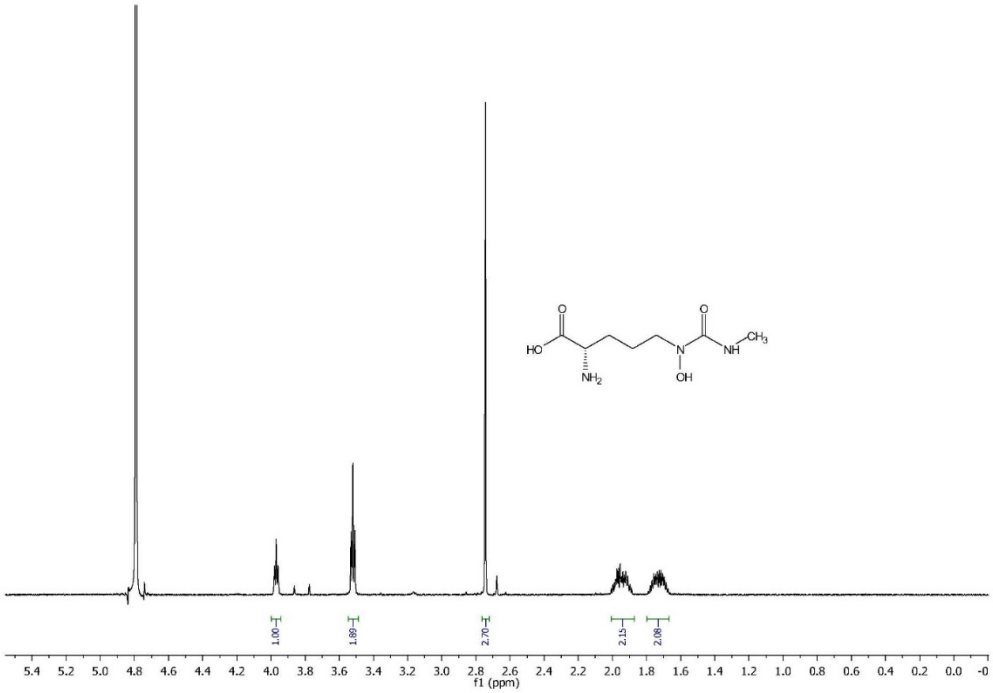


Figure 3.35: NMR spectra of **3.4**. **A**) ^{1}H spectrum of **3.4** (recorded in D_2O at 600 MHz). **B**) ^{13}C spectrum of **3.4** (recorded in d_6-DMSO at 125 MHz).

A



B

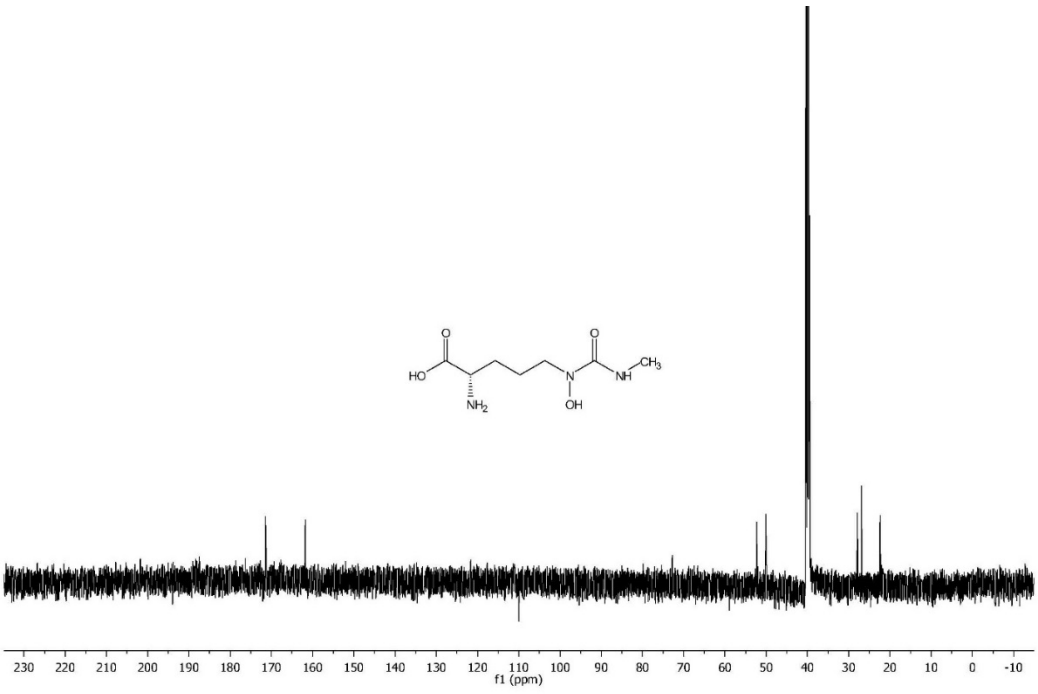


Figure 3.36: NMR spectra of **3.21**. ^{1}H spectrum of **3.21** (recorded in $CDCl_3$ at 400 MHz).

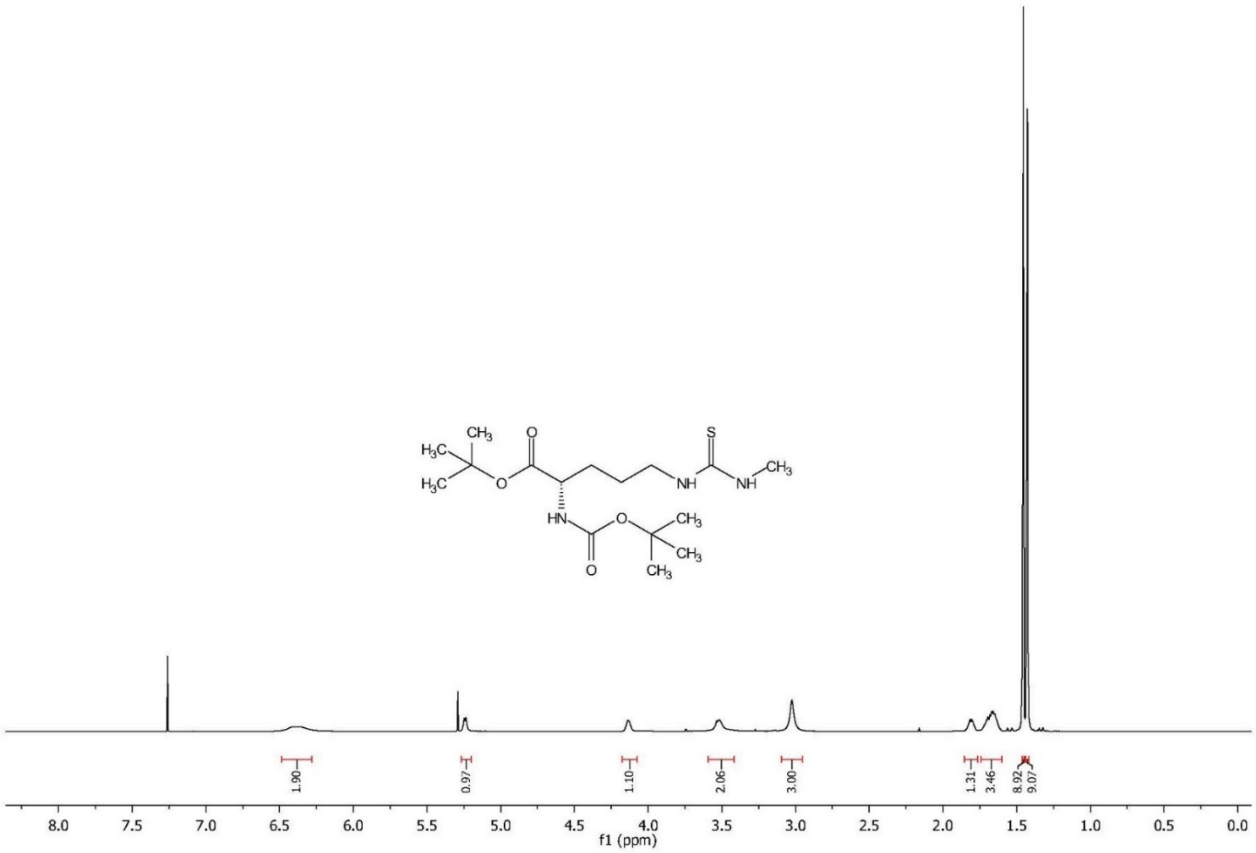
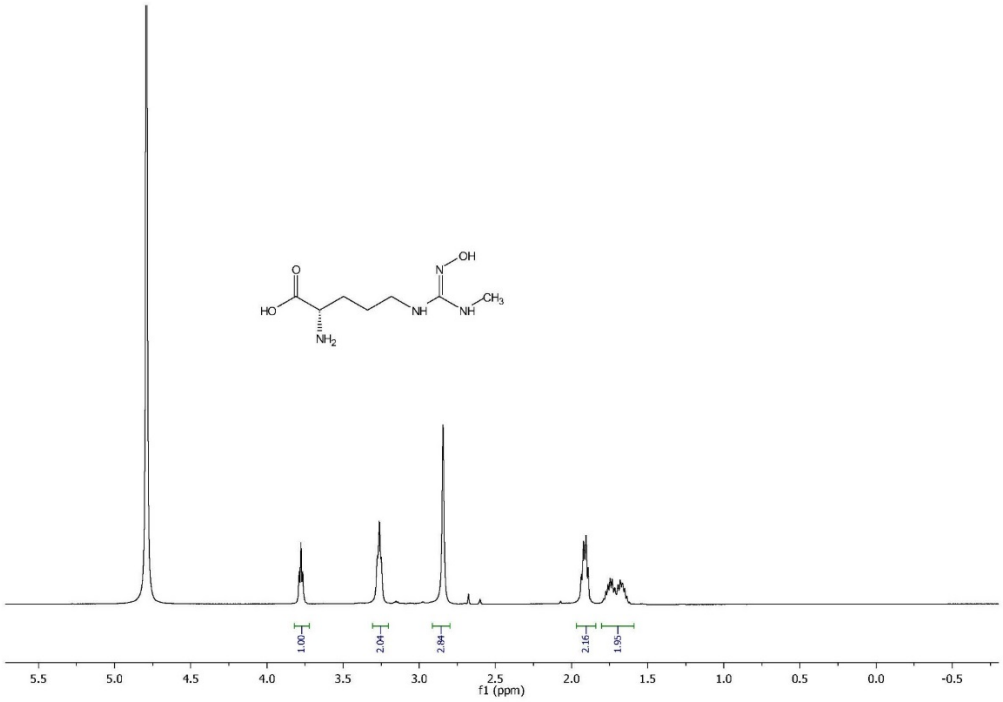
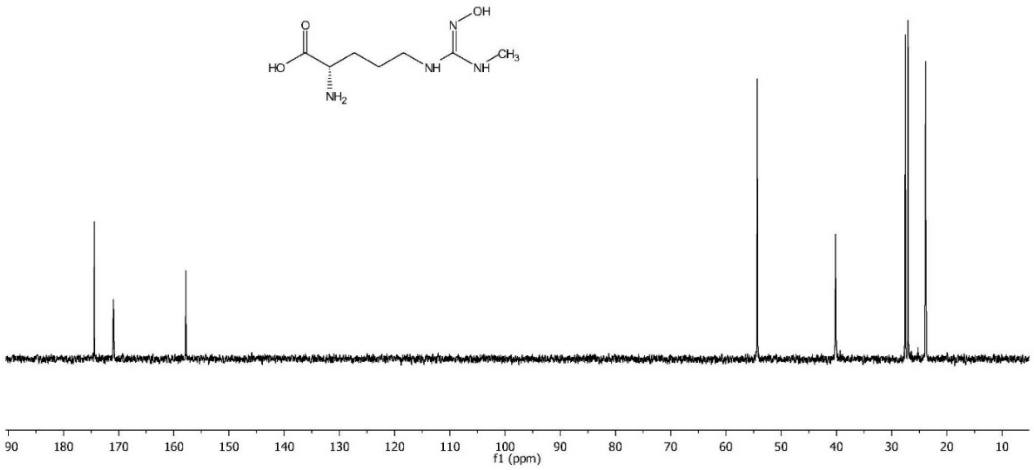


Figure 3.37: NMR spectra of **3.5**. **A**) ^{1}H spectrum of **3.5** (recorded in D_2O at 500 MHz). **B**) ^{13}C spectrum of **3.5** (recorded in D_2O at 100 MHz).

A



B



3.5 References

1) Ng, T.; Rohac, R.; Mitchell, A. J.; Boal, A. K.; Balskus, E. P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **2019**, *566*, 94.

2) He, H.; Henderson, A. C.; Du, Y.; Ryan, K. S. Two-Enzyme Pathway Links l-Arginine to Nitric Oxide in N-Nitroso Biosynthesis. *J. Am. Chem. Soc.* **2019**, *141*, 9.

3) Hedges, J. B.; Ryan, K. S. In vitro reconstitution of the biosynthetic pathway to the nitroimidazole antibiotic azomycin. *Angew. Chem. Int. Ed.* **2019**, *58*, 11647.

4) Rui, Z.; Zhu, X.; Liu, J.; Domigan, B.; Barr, I.; Cate, J. H. D.; Zhang, W. Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proc. Natl Acad. Sci. USA* **2014**, *111*, 18237.

5) Schwarzenbacher, R.; Stenner-Liewen, F.; Liewen, H.; Robinson, H.; Yuan, H.; Bossy-Wetzel, E.; Reed, J. C.; Liddington, R. C. Structure of the *Chlamydia* protein CADD reveals a redox enzyme that modulates host cell apoptosis. *J. Biol. Chem.* **2004**, *279*, 29320.

6) Zhang, B.; Rajakovich, L. J.; Van Cura, D.; Blaesi, E. J.; Mitchell, A. J.; Tysoe, C. R.; Zhu, X.; Streit, B. R.; Rui, Z.; Zhang, W.; Boal, A. K.; Krebs, C.; Bollinger, J. M. Substrate-triggered formation of a peroxo-Fe(III/III) intermediate during fatty acid decarboxylation by UndA. *J. Am. Chem. Soc.* **2019**, *141*, 14510.

7) Manley, O. M.; Fan, R.; Guo, Y.; Makris, T. M. Oxidative decarboxylase UndA utilizes a dinuclear iron cofactor. *J. Am. Chem. Soc.* **2019**, *141*, 8684.

8) Marchand, J.A.; Neugebauer, M.E.; Ing, M. C.; Lin, C.; Pelton, J. G.; Chang, M. C. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **2019**, *567*, 420.

9) Dominy, J. E.; Simmons, C. R.; Hirschberger, L. L.; Hwang, J.; Coloso, R. M.; Stipanuk, M. H. Discovery and characterization of a second mammalian thiol dioxygenase, cysteamine dioxygenase. *J. Biol. Chem.* **2007**, *282*, 25189.

10) Leitgeb, S.; Straganz, G. D.; Nidetzky, B. Biochemical characterization and mutational analysis of the mononuclear non-haem Fe2+ site in Dke1, a cupin-type dioxygenase from Acinetobacter johnsonii. *Biochem. J.* **2009**, *418*, 403.

11) Lee, J.; Simurdiak, M.; Zhao, H. Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation. *J. Biol. Chem.* **2005**, *280*, 36719.

12) Kal, S.; Que, L. Dioxygen activation by nonheme iron enzymes with the 2-His-1-carboxylate facial triad that generate high-valent oxoiron oxidants. *J. Biol. Inorg. Chem.* **2017**, *22*, 339.

13) Dunwell, J. M.; Purvis, A.; Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **2004**, *65*, 7.

14) Forist, A. A. Spectrophotometric determination of streptozotocin. *Anal. Chem.* **1964**, *36*, 1338.

15) Xia, Y.; Zweier, J. L. Direct measurement of nitric oxide generation from nitric oxide synthase. *Proc. Natl. Acad. Sci. USA.* **1997***, 94*, 12705.

16) Crane, B. R.; Sudhamsu, J.; Patel, B. A. Bacterial nitric oxide synthases. *Annu. Rev. Biochem.* **2010**, *79*, 445.

17) Mitchell, A. J. *Structural basis for alternative reaction outcome by the iron- and 2-oxoglutarate-dependent oxygenases*. Ph.D. thesis, The Pennsylvania State University, (**2017**).

18) Merkx, M.; Kopp, D. A.; Sazinsky, M. H.; Blazyk, J. L.; Müller, J.; Lippard, S. J. Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: a tale of two irons and three proteins. *Angew. Chem. Int. Ed.* **2001**, *40*, 2782.

19) Pei, J.; Kim, B. H.; Tang, M.; Grishin, N. V. PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res.* **2007**, *35*, W649.

20) Li, N.; Korboukh, V. K.; Krebs, C.; Bollinger, J. M. Four-electron oxidation of p-hydroxylaminobenzoate to p-nitrobenzoate by a peroxodiferric complex in AurF from Streptomyces thioluteus. *Proc. Natl. Acad. Sci. USA.* **2010**, *107*,15722.

21) Bollinger, J. M. et al. Mechanisms of 2-Oxoglutarate-Dependent Oxygenases: The Hydroxylation Paradigm and Beyond (eds R. P. Hausinger & C. J. Schofield) 95–122 (The Royal Society of Chemistry, **2015**).

22) Kovaleva, E. G.; Lipscomb, J. D. Crystal structures of Fe(II) dioxygenase superoxo, alkylperoxo, and bound product intermediates. *Science* **2017**, 316, 453.

23) Enemark, J. H.; Feltham, R. D. Principles of structure, bonding, and reactivity for metal nitrosyl complexes. *Coord. Chem. Rev.* **1974**, *13*, 339.

24) Richter-Addo, G. B.; Legzdins, P. *Metal Nitrosyls*. 281–285 (Oxford University Press, **1992**).

25) Vilbert, A. C.; Caranto, J. D.; Lancaster, K. M. Influences of the heme-lysine crosslink in cytochrome P460 over redox catalysis and nitric oxide sensitivity. *Chem. Sci.* **2017**, *9*, 368.

26) Caranto, J. D.; Vilbert, A. C.; Lancaster, K. M. *Nitrosomonas europaea*cytochrome P460 is a direct link between nitrification and nitrous oxide emission. *Proc. Natl Acad. Sci. USA* **2016**, *113*, 14704.

27) Li, H.; Poulos, T. L. Structure-function studies on nitric oxide synthases. *J. Inorg. Biochem.* **2005**, *99*, 293.

28) Tamanaha, E.; Zhang, B.; Guo, Y.; Chang, W. C.; Barr, E. W.; Xing, G.; St Clair, J.; Ye, S.; Neese, F.; Bollinger, J. M.; Krebs, C. Spectroscopic Evidence for the two C–H-cleaving intermediates of Aspergillus nidulans isopenicillin N synthase. *J. Am. Chem. Soc.* **2016**, *138*, 8862.

29) Ye, S.; Riplinger, C.; Hansen, A.; Krebs, C.; Bollinger, J. M.; Neese, F. Electronic structure analysis of the oxygen-activation mechanism by Fe(II)- and alpha-ketoglutarate-dependent dioxygenases. *Chemistry* **2012**, *18*, 6555.

30) Mitchell, A. J.; Dunham, N. P.; Martinie, R, J.; Bergman, J. A.; Pollock, C. J.; Hu, K.; Allen, B. D.; Chang, W. C.; Silakov, A.; Bollinger, J. M.; Krebs, C.; Boal, A. K. Visualizing the reaction cycle in an iron(II)- and 2-(oxo)-glutaratedependent hydroxylase. *J. Am. Chem. Soc.* **2017**, *139*, 13830.

31) Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA.* **1985***, 82*, 488.

32) Lin, Y.; Miller, M. J. Practical synthesis of hydroxamate-derived siderophore components by an indirect oxidation method and syntheses of a DIG−siderophore conjugate and a biotin−siderophore conjugate. *J. Org. Chem.* **1999**, *64*, 7451.

33) Jansen Labby, K.; Li, H.; Roman, L. J.; Martásek, P.; Poulos, T. L.; Silverman, R. B. Methylated Nω-Hydroxy-l-arginine analogues as mechanistic probes for the second step of the nitric oxide synthase-catalyzed reaction. *Biochemistry* **2013**, *52*, 3062.

Chapter 4. Characterization of the L-amino acid ligases SznH and SznK and distribution of SznF homologs in microbial genomes

Part of this Chapter is adapted from previous publication.[1]

4.1 Introduction

Having elucidated the roles of the methyltransferase SznE (Chapter 2) and the metalloenzyme SznF (Chapter 3) in the streptozotocin (*szn*) biosynthetic cluster, we next characterized the two ATP-grasp enzymes, SznH and SznK. We described in Chapter 3 that SznH and SznK might be involved in forming the amide linkage between D-glucosamine and the *N*-nitrosourea-containing amino acid **3.3**. To test this hypothesis, we performed *in vivo* experiments analyzing the extracts of the various *szn* mutants and searched for accumulating on-pathway metabolites or shunt products. We also performed *in vitro* experiments using purified SznH and SznK to understand their biochemical functions. Contrary to our hypothesis, we found that SznH and SznK are L-amino acid ligases. Specifically, SznK uses ATP to ligate L-serine and **3.3** to form a dipeptide. Subsequently, SznH ligates L-alanine to the dipeptide to afford a tripeptide natural product. These results provided evidence that biosynthetic enzymes have evolved to further elaborate *N*-nitrosourea-containing natural products. Nonetheless, we have concluded that further *in vivo* analysis may be required for understanding the biosynthesis of SZN. In the last part of this chapter, we will discuss our bioinformatic analysis of SznF and genome neighborhood analysis of gene clusters containing SznF homologs. These analyses should pave the road for uncovering additional *N*-nitroso-containing natural products. The bioinformatic analyses presented in this section were performed with Dr. Grace E. Kenney.

4.2 Experimental Results

4.2.1 Biochemistry of ATP-grasp enzymes

An ATP-grasp enzyme catalyzes the ATP-dependent ligation of a carboxylate and an amine, thiol, or alcohol nucleophile to afford the corresponding amide, thioester, or ester product (**Figure 4.1A**).[2] This enzyme superfamily, which includes biotin carboxylase, L-amino acid ligase, D-alanine-D-alanine ligase (**Figure 4.1B**), and phosphoribosylamine-glycine ligase, can be found in both prokaryotic and eukaryotic cells.[3]

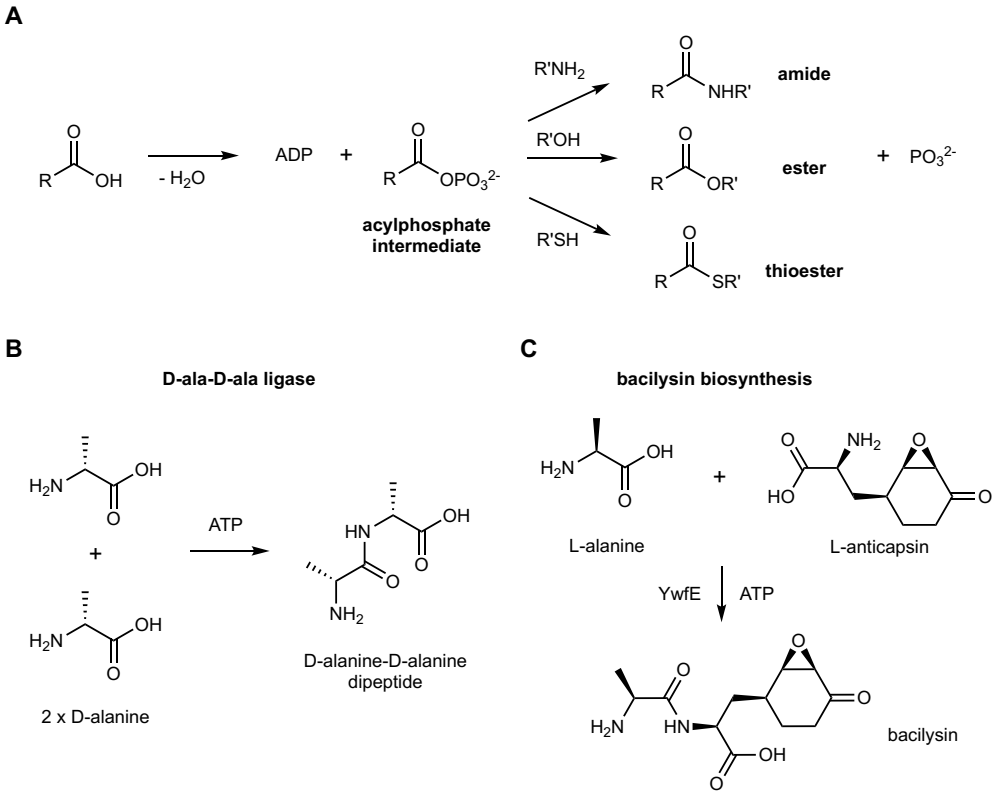


Figure 4.1: The reactions catalyzed by ATP-grasp enzymes. **A**) A carboxylic acid can be converted into an amide, ester, or thioester by an ATP-grasp enzyme. Some examples include **B**) D-ala-D-ala ligase found in cell-wall biosynthesis[4] and **C**) YwfE, an L-amino acid ligase found in bacilysin biosynthesis.[5]

These enzymes play crucial roles in primary metabolism, including *de novo* purine biosynthesis and cell wall biosynthesis. In addition to primary metabolism, ATP-grasp enzymes in natural product biosynthetic pathways are used to cyclize natural products to form

macrolactams, functionalize natural product scaffolds[6], or introduce protecting groups in the biosynthesis of natural products containing reactive functional groups (**Figure 4.1C**).[7]

This superfamily contains a distinct "ATP-grasp" fold, which comprises of two interacting α and β domains in which ATP is bound in the cleft between the two subunits.[8] Structural alignments of a large number of ATP-grasp enzymes have revealed several signature motifs to coordinate the requisite magnesium ion (Mg^{2+}) and the negatively charged phosphate groups of ATP. For example, two basic residues, usually arginine or lysine, bind the α- and β-phosphates of ATP.[9] Also, a triad of acidic residues and asparagine are typically involved in binding Mg^{2+}.[10] A typical mechanism catalyzed by an ATP grasp enzyme involves initial activation of a substrate carboxylate and adenosine triphosphate (ATP) to form an acyl-phosphate intermediate and adenosine diphosphate (ADP) (**Figure 4.1A**). In a separate step, nucleophilic attack by an amine, thiol, or alcohol converts the acyl-phosphate intermediate to release phosphate (PO_4^{3-}) and the corresponding amide, thioester, or ester.

Our *in vivo* studies have indicated that the two predicted ATP-grasp enzymes SznH and SznK are important in SZN biosynthesis (Chapter 2). One biosynthetic hypothesis might involve initial cleavage of **3.3** to afford a transient *N*-methyl-*N*-nitroso carbamic acid, which is phosphorylated to form *N*-methyl-*N*-nitroso carbamoyl phosphate (**Figure 4.2**). The enzyme would then ligate the unstable intermediate to D-glucosamine to form SZN. This hypothesis, however, would require one of the ATP grasp enzymes to cleave an amide bond, an unprecedented reaction of this superfamily. This hypothesis also highlights the challenge of identifying the correct sugar substrate for *in vitro* analysis. For example, D-glucosamine can be acetylated, phosphorylated, or converted into a sugar nucleotide *in vivo* before the transfer reaction occurs.

These considerations mandate both an *in vivo* and *in vitro* approach to understanding SZN biosynthesis.

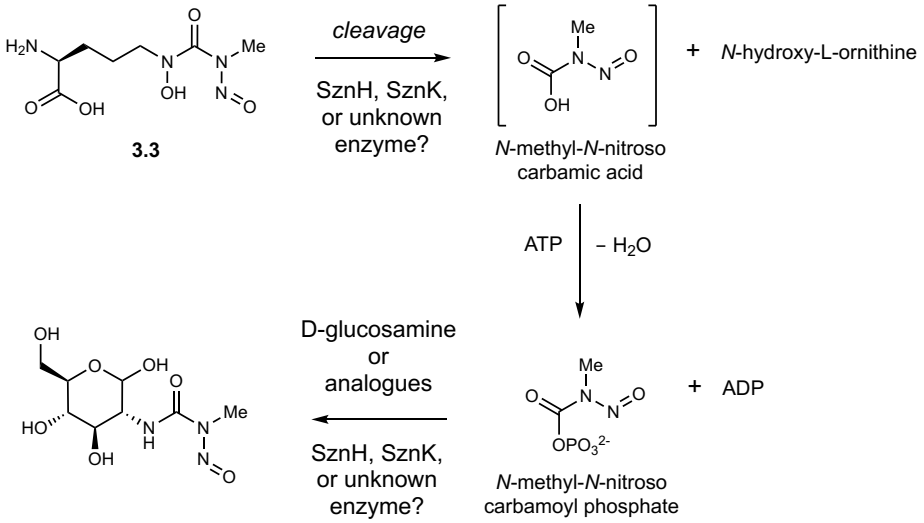


Figure 4.2: One biosynthetic proposal for the transfer of the *N*-nitrosourea moiety from **3.3** to D-glucosamine to afford SZN.

4.2.2 *In vivo* metabolomics of the Δ*sznH* and Δ*sznK* mutants

To investigate the functions of SznH and SznK, we first analyzed the fermentation extracts of the Δ*sznH* and Δ*sznK* mutants. *In vivo* knockouts often provide insights into the order of events in a biosynthetic pathway. Presumably, the Δ*sznH* and Δ*sznK* mutants would accumulate the substrates of their corresponding enzymes, SznH and SznK, which can inform *in vitro* assay design. The masses of these substrates could be detected with liquid chromatography-mass spectrometry (LC-MS), and the differences in mass abundance could be detected via metabolomics tools such as XC-MS.[11]

When we compared the fermentation extracts from the Δ*sznH* and Δ*sznK* mutants against the extracts from wild-type (WT) using XC-MS, we observed accumulation of a *m*/*z* ($[M+H]^+$ = 338.1035) in the mutants. One chemical formula within 5 ppm error is $C_{11}H_{19}N_3O_7S$ ($[M+H]^+$ = 338.1016), which could be the *N*-acetylcysteine adduct **4.1** (**Figure 4.3**). We suspected this

compound is derived from mycothiol reacting with **3.3**. Mycothiol is a protective thiol produced by actinobacteria that scavenges electrophiles with a thiol warhead.[12] The resulting adduct is cleaved by an amidase that forms a mercapturic acid **4.1**, which is exported outside the cell. This result suggested that **3.3** is likely processed by SznH and SznK. Aside from **4.1**, we did not reproducibly observe accumulation of any other metabolites, including **3.3**, in the extracts produced by the Δ*sznH* and Δ*sznK* mutants.

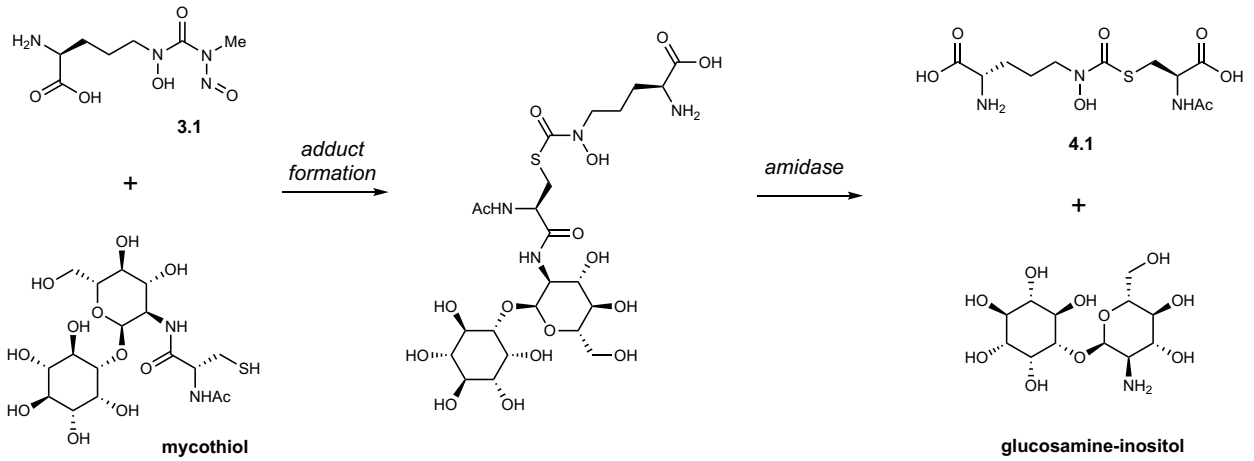
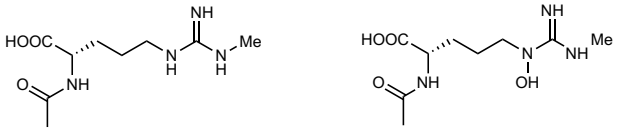


Figure 4.3: Comparative metabolomics analysis of WT *S. achromogenes*, Δ*sznH*, and Δ*sznK* mutants. **A**) XC-MS analysis of the fermentation extracts between WT *S. achromogenes* and Δ*sznH*. The overlaid chromatograms demonstrate that a metabolite with a $[M+H]^+$ *m*/*z* of 338.1 is produced by Δ*sznH* (red traces) and not WT (black traces). This unknown metabolite is also produced by Δ*sznK* (panel **B**). **C**) Mycothiol might be scavenging **3.2**, a metabolite accumulating in the Δ*sznH* and Δ*sznK* mutants. The resulting adduct can be cleaved intracellularly to afford the mercapturic acid **4.1**, which has a $[M+H]^+$ *m*/*z* of 338.1016.

In addition to searching for differences in mass abundance, we also fed unlabeled L-NMA and d_3-methyl-L-NMA to the Δ*sznH* and Δ*sznK* mutants and analyzed the fermentation extracts with LC-MS and XC-MS. We hypothesized that any on-pathway intermediates or off-pathway de-nitrosated shunt products would also contain the *N*-methyl group. While we observed labeling of *N*-methylated intermediates in samples fed with d_3-methyl-L-NMA, these masses likely

correspond to *N*-acetylated L-NMA or *N*-acetylated-**3.3** (**Figure 4.4**). These metabolites might arise from promiscuous amino acid *N*-acetyltransferases *in vivo* that are irrelevant to SZN biosynthesis.[13] Aside from these masses, we did not observe new metabolites with a mass shift of +3 Da when d_3-methyl-L-NMA was fed. Based on these results, we hypothesized that SznH and SznK may work in concert and not release intermediates. An alternative explanation is that the intermediates are too low in abundance or too unstable for detection using our MS-based methods. Overall, our *in vivo* analysis did not reveal new insights into the biosynthesis of SZN. Additional metabolomics method, such as NMR-based detection methods, may provide new information about the metabolites produced during the late stages of SZN biosynthesis. These additional avenues for exploration will be described in the conclusion section of this chapter.



N-acetyl-L-NMA
$[M+H]^+$ = 231.1452

N-acetyl-**3.1**
$[M+H]^+$ = 247.1401

Figure 4.4: Structures of *N*-acetylated-L-NMA (left) and *N*-acetylated-**3.1** (right).

4.2.3 *In vitro* characterization of SznK

We next focused on *in vitro* characterization of SznH and SznK and their enzymatic chemistry on **3.3**. This approach enables interrogation of the function of the two enzymes in a less complex matrix. However, this effort can be impeded by our lack of knowledge of the biologically relevant substrates. We first confirmed that the amino acid residues involved in binding ATP and Mg^{2+} are present by aligning the protein sequences of SznH and SznK to protein sequences of biochemically and structurally characterized ATP-grasp enzymes (**Figure 4.5**).[14-16] We then purified the two N-His_{6x}-tagged recombinant enzymes using Ni-NTA resins. However, we did not

observe production of SZN when we incubated SznH and SznK with **3.3**, D-glucosamine, Mg^{2+}, and ATP. Therefore, additional substrates and/or cofactors may be required.

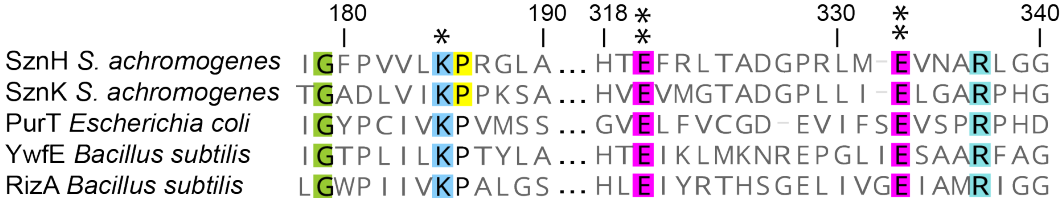


Figure 4.5: Sequence alignment of SznH and SznK with biochemically and structurally characterized ATP-grasp enzymes. PurT is a glycinamide ribonucleotide transformylase (PDB: 1KJQ) involved in de novo purine biosynthesis. YwfE (PDB: 3WNZ) is an L-amino acid ligase from bacilysin biosynthesis (**Figure 4.1C**). RizA (PDB: 4WD3) is also an L-amino acid ligase in the biosynthetic pathway of rhizoticin, an oligopeptide natural product. The lysine residue highlighted with a single asterisk is involved in ATP-binding. The two glutamate residues highlighted with two asterisks are involved in binding Mg^{2+}. This alignment provides evidence that both SznH and SznK likely bind ATP and Mg^{2+}.

We then decided to investigate whether SznH and SznK exhibit L-amino-acid ligase activity. Screening several proteinogenic L-amino acids, such as L-serine, L-alanine, and L-cysteine, revealed that ATP is rapidly consumed only in the presence of **3.3,** L-serine, and SznK (**Figure 4.6**). These results suggested that a productive coupling reaction involving ATP hydrolysis occurs when L-serine and SznK are present.

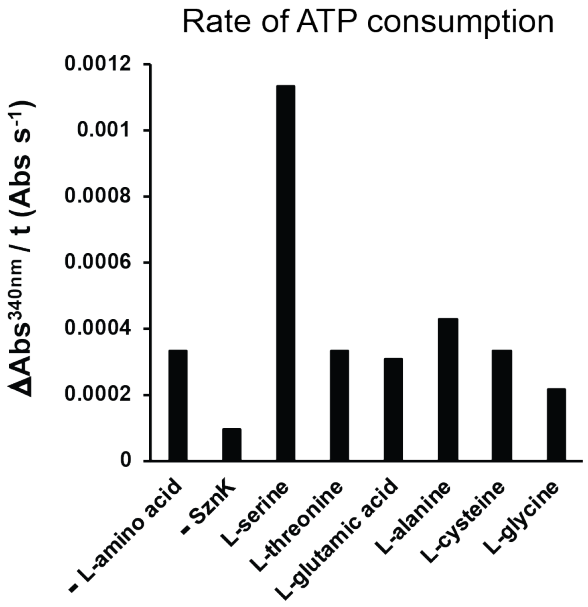


Figure 4.6: Rate of ATP consumption using a pyruvate kinase/lactate hydrogenase coupled assay (See Methods). In the presence of L-serine and SznK, ATP is consumed at a faster rate compared to assays in which other proteinogenic L-amino acids are used.

To confirm that a dipeptide natural product is produced, we analyzed the SznFK assay mixture with LC-MS. We observed a mass corresponding to a dipeptide **4.2** and its degraded product **4.3** (**Figure 4.7**). Excluding ATP, **3.3**, or Mg^{2+} abolished enzyme activity, confirming the involvement of ATP. (**Figure 4.8**).

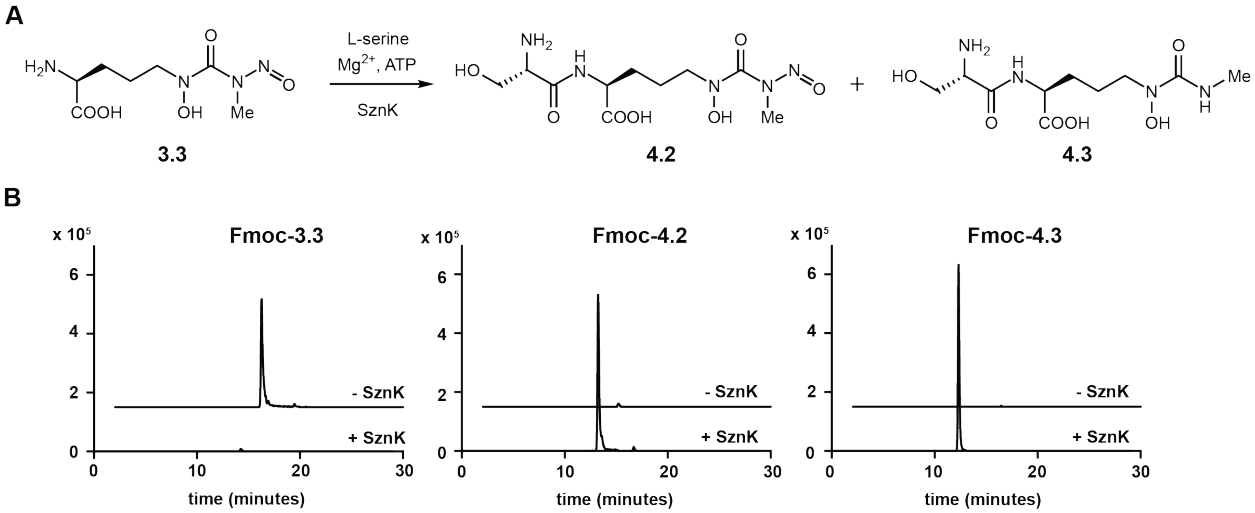


Figure 4.7: SznK is an L-amino acid ligase. **A**) The reaction catalyzed by SznK. **B**) LC-MS traces demonstrating **4.2** and **4.3** are produced when SznK, L-serine, **3.3**, ATP were incubated for 1 h at room temperature. EICs ($[M–H]^-$) of Fmoc-**3.3**, Fmoc-**4.2** and Fmoc-**4.3** are 455.1572, 542.1893 and 513.1991, respectively.

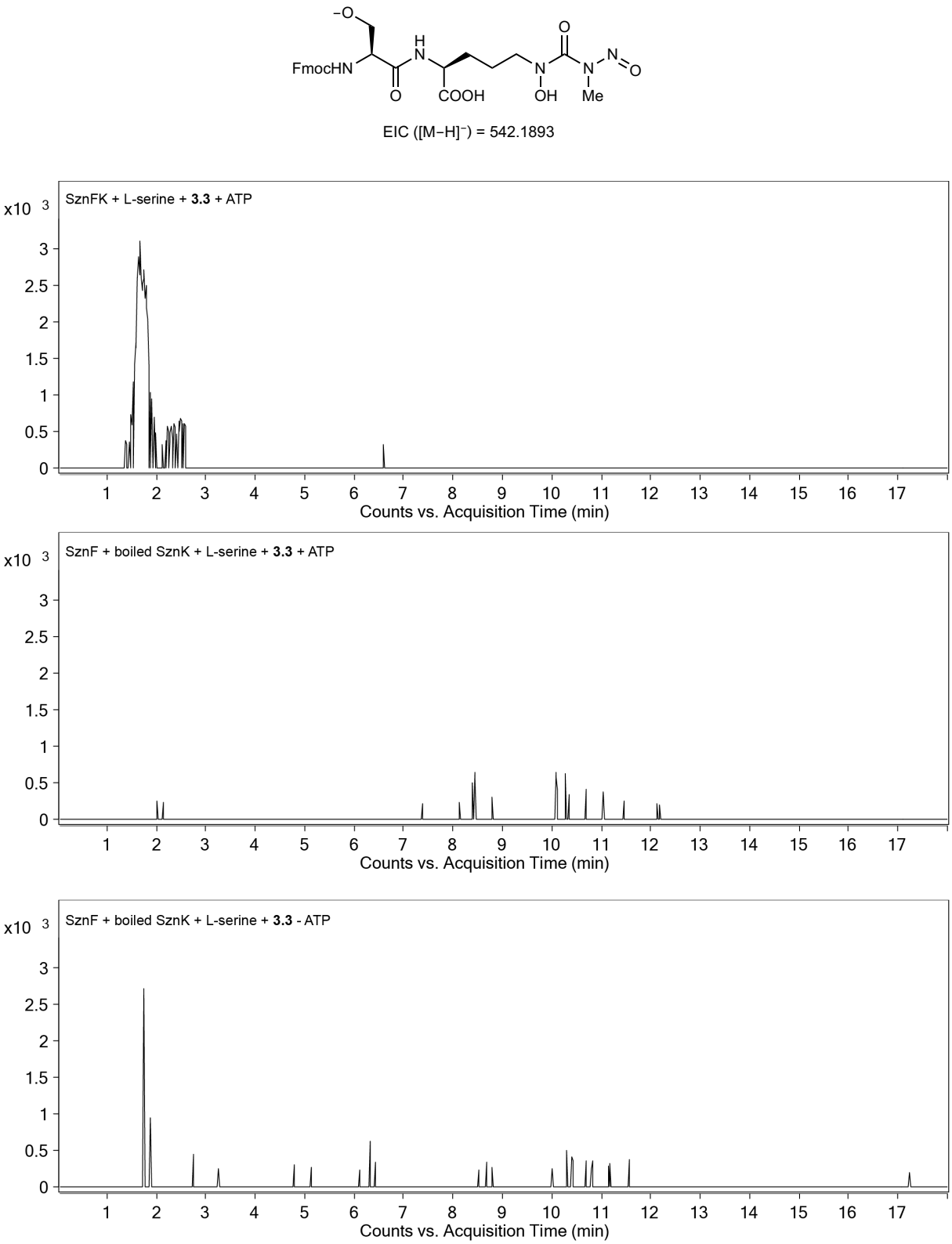


Figure 4.8: EIC chromatograms of Fmoc-**4.2** demonstrating that **4.2** is produced only in the presence of SznK and ATP.

To gain insights into the structure of **4.2**, we derivatized **4.2** with Fmoc and analyzed the resulting derivative with LC-MS/MS. We observed a mass fragment corresponding to Fmoc-L-serine, which suggested the amino group of **3.3** is linked to the carboxylate group of L-serine (**Figure 4.9**). Next, we focused on the biochemistry of SznH. We initially hypothesized SznH can transfer the *N*-methyl-*N*-nitrosourea group from the dipeptide **4.2** to D-glucosamine. However, inclusion of L-serine and ATP, Mg^{2+}, D-glucosamine, and SznH in a reaction mixture did not afford SZN. This result suggested that either D-glucosamine is not accepted as a substrate or additional enzymes are required to reconstitute SZN biosynthesis.

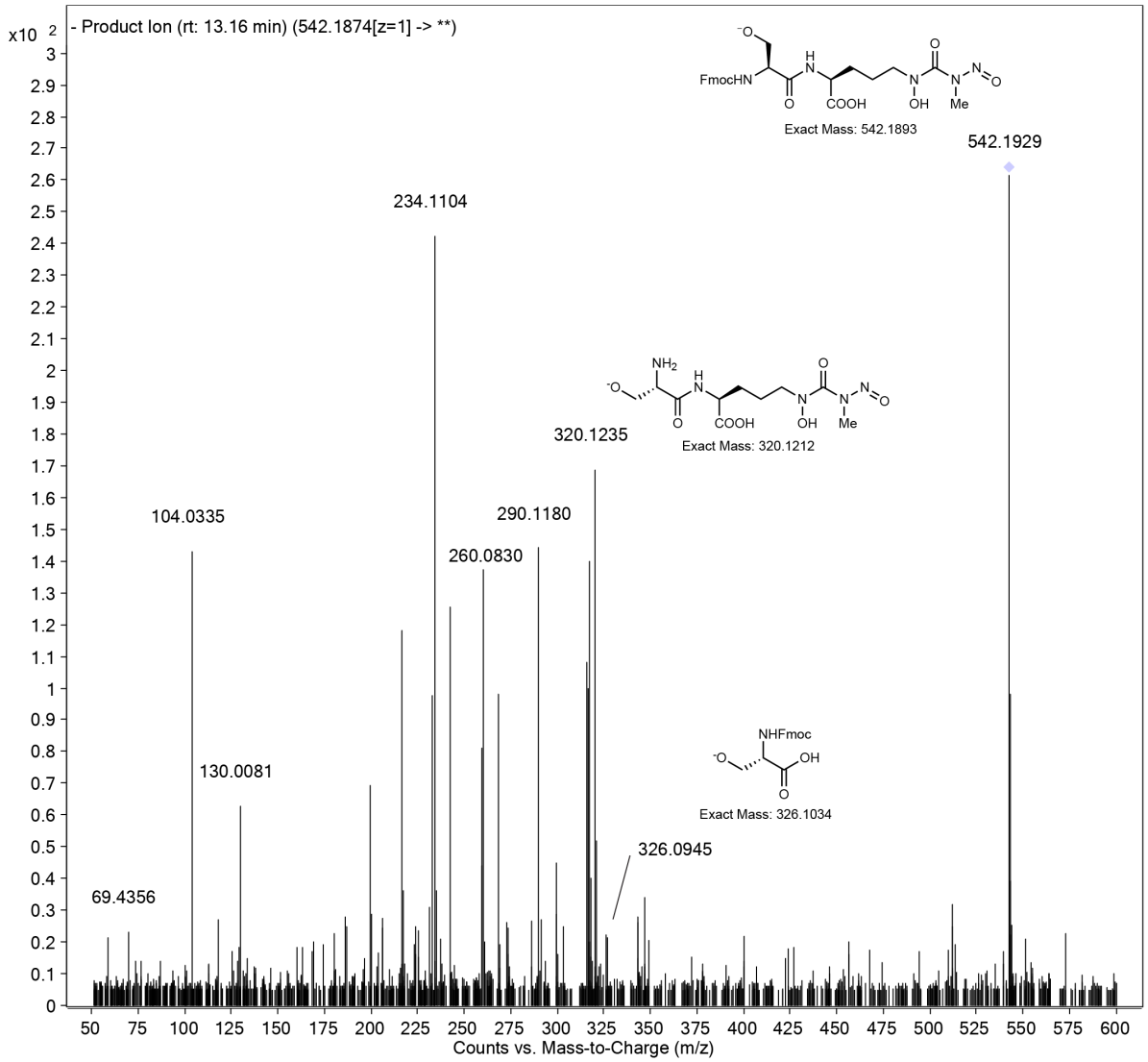


Figure 4.9: MS2 spectrum of Fmoc-**4.2** with the key MS2 fragments highlighted. The MS2 fragment 326.1034 suggests the amino group of L-serine reacted with Fmoc-Osu, therefore the amide linkage formed by SznK is between the carboxylate of L-serine and the amino group of **3.3.**

4.2.4 *In vitro* characterization of SznH

We then hypothesized that SznH is also an L-amino acid ligase that ligates a second amino acid to the dipeptide. Indeed, we screened six pools of proteinogenic amino acids with SznFK-generated dipeptide **4.2**. We observed consumption of dipeptide **4.2** and appearance of a new mass

corresponding to the *N*-nitrosourea-containing tripeptide **4.4** and its degradation product when L-alanine was used as the amino acid (**Figure 4.10**). D-alanine was not accepted as a substrate, demonstrating that only one enantiomer is accepted **(Figure 4.11)**.

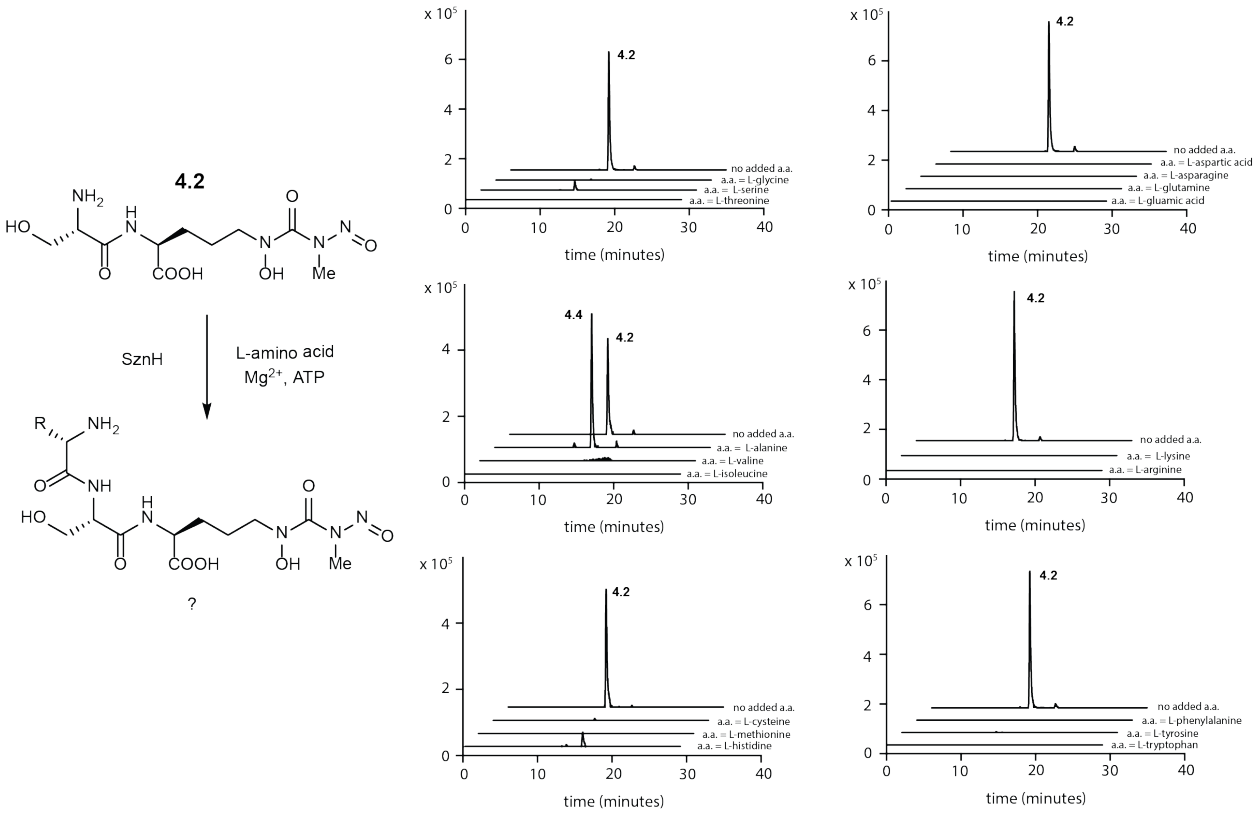
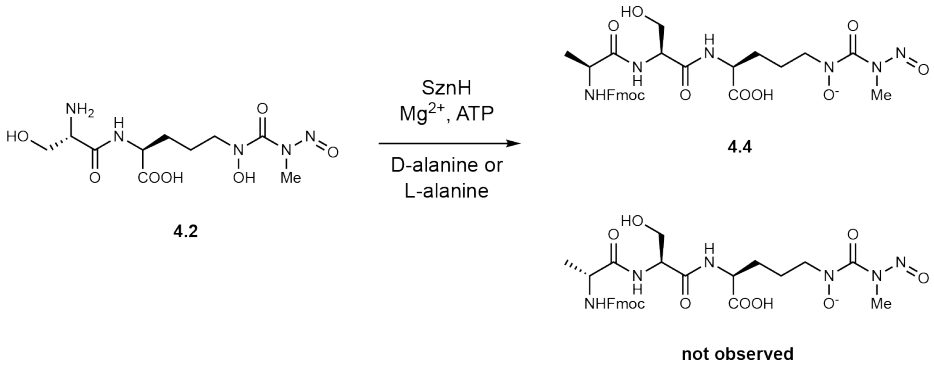
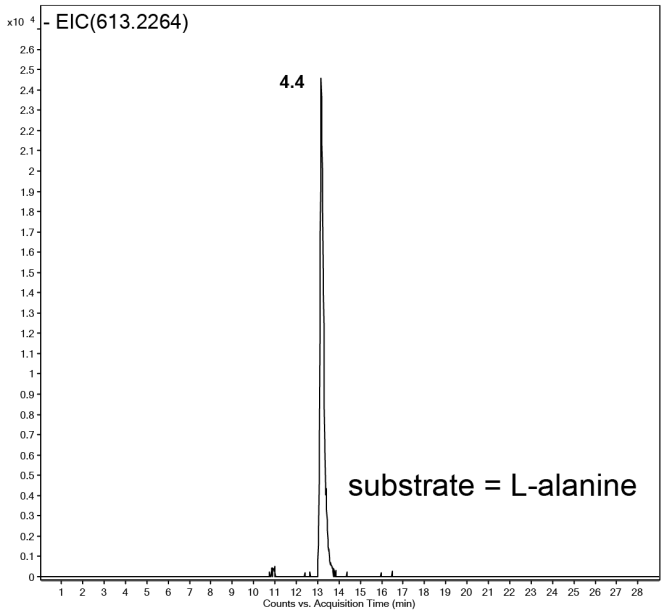


Figure 4.10: EIC chromatograms for tripeptides comprised of **4.4** and an L-amino acid in enzyme assays containing SznH, Mg^{2+}, and ATP. When the added amino acid is L-alanine, modest production of the tripeptide **4.4** was observed.





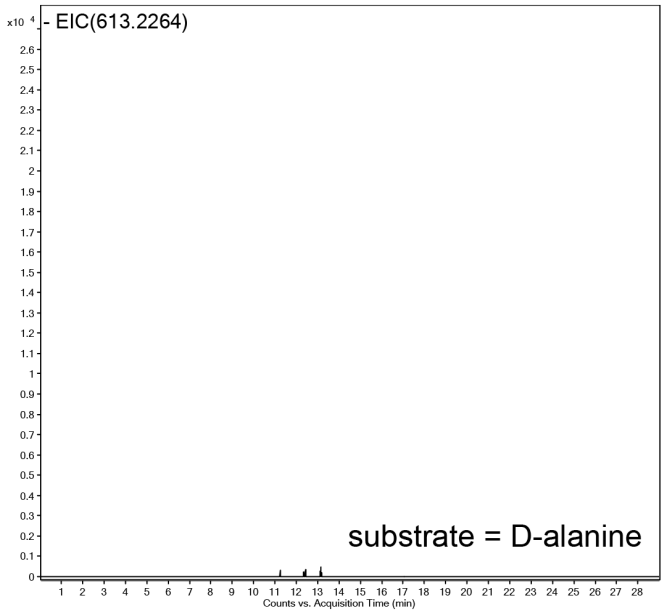


Figure 4.11: EIC chromatograms of Fmoc-**4.4** when L-alanine (top) or D-alanine (bottom) was used as substrate.

We also incubated the amino acid pools with SznH and SznF-generated **3.3** and did not observe any masses of the corresponding dipeptide **(Figure 4.12).** This result confirmed that SznK first generates a dipeptide with **3.3** and L-serine before SznH forms the tripeptide natural product. We reanalyzed our old *in vivo* data, and we were not able to detect these dipeptides or tripeptides in the extracts of WT, Δ*sznK,* and Δ*sznH* mutants. One possible explanation is that *S achromogenes* produces promiscuous peptidases that cleave these small peptides *in vivo*. Therefore, the abundance of these metabolites may not be high enough for our detection methods. Our *in vivo* and *in vitro* analysis of SznH and SznK led us to believe that other enzymes in the *szn* cluster or in the *S. achromogenes* genome might be responsible for the transfer reaction.

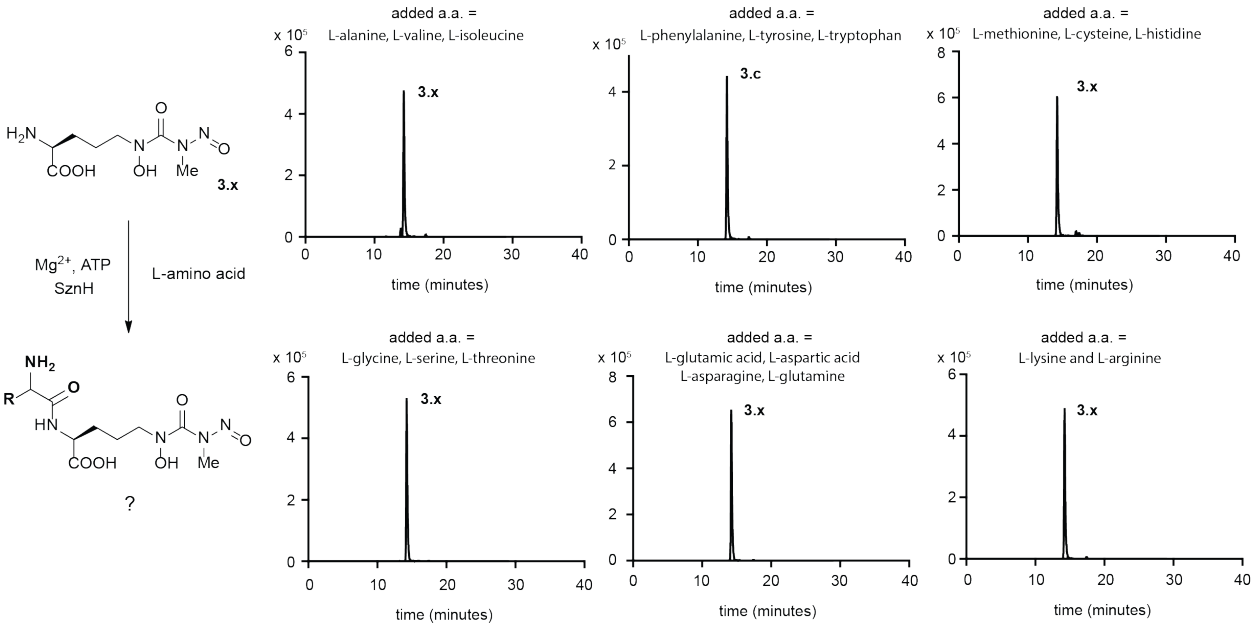


Figure 4.12: EIC chromatograms for dipeptides comprised of **3.3** and an L-amino acid in enzyme assays containing SznH, Mg^{2+}, and ATP. No dipeptide products were observed, indicating SznH only accepts dipeptide substrates.

4.2.5 Conclusion and future direction for SZN biosynthesis

Our study of SznH and SznK revealed further modifications of the *N*-methyl-*N*-nitrosourea containing amino acid **3.3** to form a tripeptide natural product. While unexpected, this result

suggested that additional enzymes are required to transfer the *N*-nitrosourea warhead from **3.3** to D-glucosamine or its analogues. One outstanding question is whether **4.4** is an on-pathway intermediate to SZN biosynthesis. If true, the question that remains is which enzyme(s) is/are responsible for this transformation. One of the challenges to addressing this question is the instability of natural products containing hydroxylated *N*-methyl-*N*-nitrosourea groups. We currently cannot synthesize and feed these unstable compounds to *S. achromogenes* as the compounds will likely degrade rapidly in solution. Finding a route to synthesize these amino acids might be valuable for chemical complementation studies or resting cell assays. Obtaining chemically pure *N*-nitrosourea substrates could also enable further *in vitro* characterization of SznH and SznK and simplify assay conditions by omitting SznF and unreacted **3.2**. Reducing the number of components in the enzyme reaction could lead to less interference in the transfer reaction and improve detection of analytes.

Other methods for analyzing the metabolites from *in vivo* extracts may prove to be useful. This may include molecular networking to detect features such as loss of the D-glucosamine moiety in tandem mass spectrometry (MS/MS). Other methods may involve NMR-based metabolomics analysis. For example, one could feed ^{15}N-D-glucosamine to *S. achromogenes*, fractionate the resulting extracts, and analyze the fractions by ^{15}N-NMR to isolate compounds that are ^{15}N-enriched. These new metabolites could be modified D-glucosamine or degraded biosynthetic intermediates containing D-glucosamine that will provide insight into SZN biosynthesis.

With regards to additional gene candidates for the transfer reaction, one enzyme might be the predicted carboxylesterase SznJ. A typical esterase utilizes a catalytic serine residue to generate an enzyme-tethered ester intermediate and an alcohol. Next, water attacks the enzyme-bound ester

to form a carboxylic acid **(Figure 4.13**). While **3.3, 4.2,** and **4.4** do not contain an ester bond, SznJ may be able to cleave **3.3** to form an enzyme-tethered *N*-methyl-*N*-nitroso carbamate. Afterwards, instead of water, an amine nucleophile (i.e. D-glucosamine) attacks the enzyme-bound ester to afford SZN. This transacylation reaction can be tested *in vitro* with purified SznJ, D-glucosamine, and either **3.3**, **4.2**, or **4.4**.

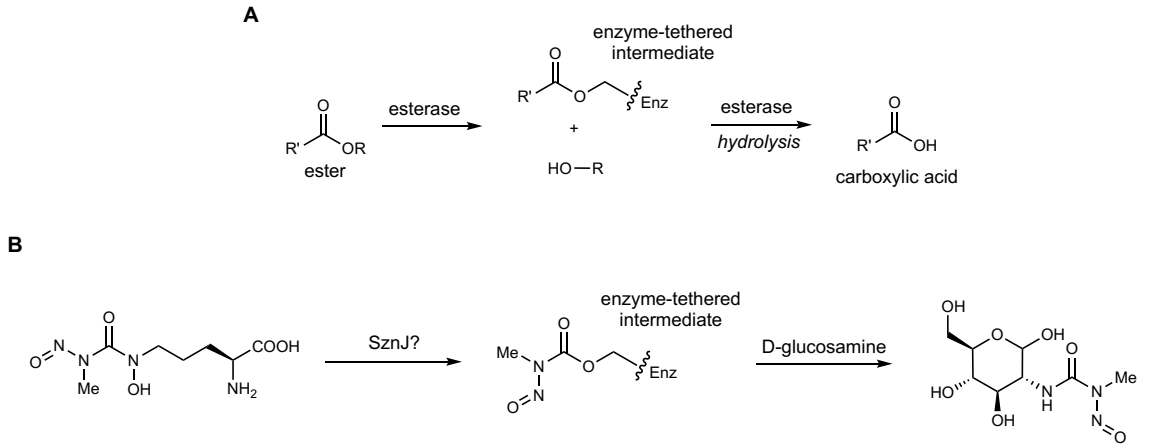


Figure 4.13: **A**) A typical hydrolysis reaction catalyzed by an esterase. **B**) A biosynthetic proposal in which an esterase, such as SznJ, could cleave the *N*-nitrosourea group to generate an enzyme-tethered carbamate before nucleophilic attack by D-glucosamine.

4.2.6 Distribution of *szn* cluster in microbial genomes and future directions

The novelty of SznF motivated us to assess its distribution in microbial genomes. BLASTP searches uncovered SznF homologs containing both the central and cupin domains encoded in numerous cryptic biosynthetic gene clusters (e-value < 1E-5) (**Figure 4.14A**). These gene clusters are widely distributed across bacterial phyla and habitats (**Figure 4.14B**) including soil and rhizosphere *Burkholderia* and *Paraburkholderia* strains, the plant symbiont *Frankia,* and the human pathogen *Legionella pneumophila*. This analysis suggested diverse bacteria can potentially produce *N*-nitroso-containing compounds, highlighting potential new biological roles for these metabolites and enabling genome mining-based natural product discovery. For example, the SznF-

encoding gene cluster from the human pathogen *Legionella pneumophila* could be a target for exploring whether SznF homologs in other bacteria also generate *N*-nitrosourea containing natural products (**Figure 4.14 C**). This organism causes Legionnaire's disease, and its genome encode several biosynthetic gene clusters that could produce a variety of natural products.[17] The *L. pneumophila* gene cluster that contains the *sznF* homolog also encodes a flavin reductase and ATP-dependent enzyme (**Figure 4.14C**). To test whether novel metabolites are produced, one could first ferment the strain under various media and detect production of nitrite or *N*-nitroso compounds in the fermentation extracts with the Griess assay. *In vivo* comparative metabolomics with wild type *L. pneumophila* and the *L. pneumophila* mutant with the *sznF* homolog knocked out could reveal new bioactive *N*-nitroso metabolites.

We observed many clusters in which *sznF* is encoded alongside predicted multifamily facilitator superfamily and methyltransferases. Some clusters contained predicted glutathione-cysteine ligases and alcohol dehydrogenases. In most proteobacteria, including *Legionella pneumophila,* we observed some conservation of *sznG* and ATP-grasp homologs compared to the *szn* cluster. This analysis also revealed numerous clusters in which a methyltransferase gene is lacking (e.g. azomycin biosynthesis) and clusters in which sznF homologs are clustered with NRPS genes (e.g. gramibactin biosynthesis). This result could help guide future efforts to prioritize the type of natural products encode by these gene clusters (e.g. nitro-containing metabolites versus nonribosomal peptides).

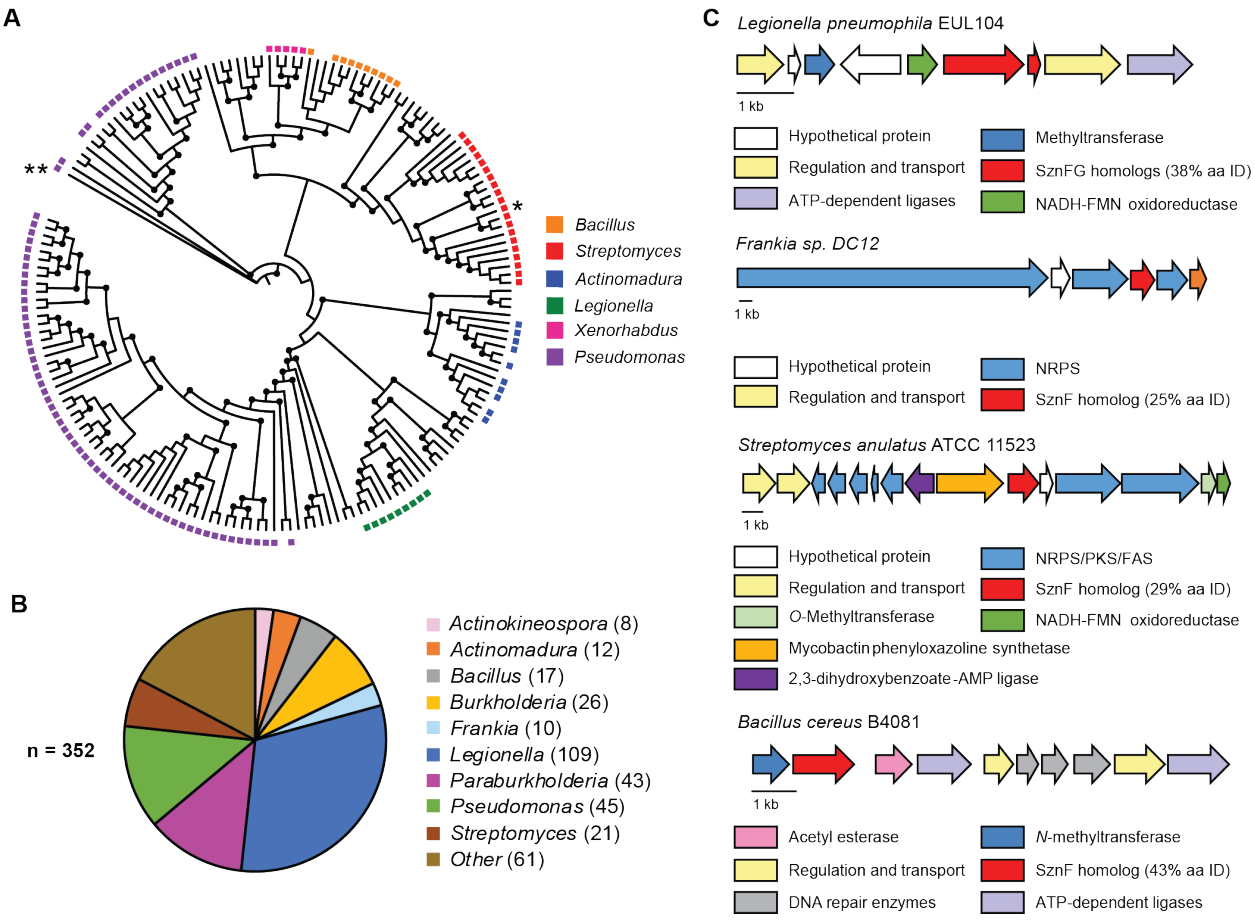


Figure 4.14: Distribution of SznF homologs in microbial genomes. **A**) Maximum-likelihood phylogenetic tree inferred from 50 replicates showing the relationship between selected SznF homologs containing both the central domain and cupin domain (NCBI "non-redundant protein sequences" database, 2018) (e-value < 1E-50). The branch corresponding to *S. achromogenes* SznF is highlighted by a single asterisk. Bootstrap confidence values of >50 are indicated by black circles on the nodes. The amino acid sequence of UndA is used as an outgroup (highlighted as **). **B**) Distribution of 352 SznF homologs that contain both a central domain and a cupin domain in different bacterial genera (IMG/JGI "all isolates" database, 2018) (e-value < 1E-5). **C**) Selected biosynthetic gene clusters encoding homologs of SznF.

Finally, we have also generated a sequence similarity network to help prioritize SznF homologs for *in vitro* studies (**Figure 4.15**). In our network (2700 sequences, evalue cutoff = 1E-70), we obtained several isofunctional clusters containing SznF homologs with only the CADD domain, only the cupin domain, or both the CADD and the cupin domains. This analysis, together with the genome neighborhood analysis, identified clusters containing sequences of protein with unknown functions. These uncharacterized enzymes could perform reactions other than *N*-

oxidation or N–N bond formation. *In vitro* biochemical characterization of these enzymes could be avenues for elucidating new chemistries performed by SznF homologs.

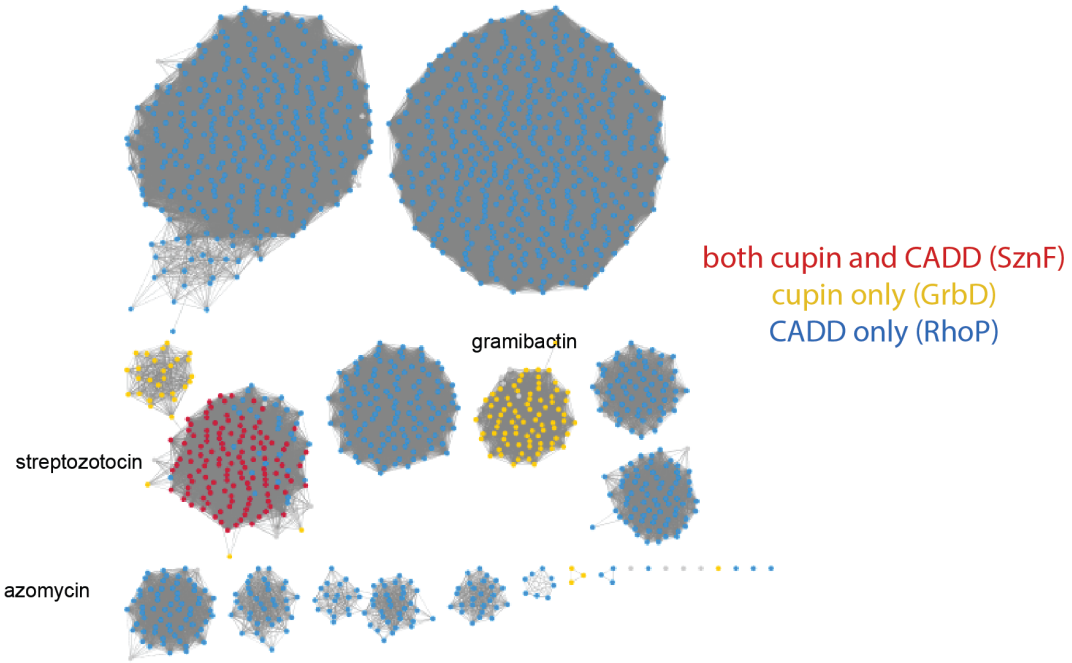


Figure 4.15: Sequence similarity network of SznF (number of sequences = 1613, evalue cutoff = 1E-70). Sequences in which the metal binding residues (three histidine residues for the cupin domain and more than five conserved metal binding residues for the CADD domain) found in SznF are present are colored. This figure was prepared by Dr. Grace E. Kenney.

4.3 Materials and Methods

LC-MS analysis of metabolites and comparative metabolomics

For our comparative metabolomics experiment, fermentation extracts from wild type *S. achromogenes*, Δ*sznH*, and Δ*sznK* mutants were analyzed using LC-MS. The LC column was a Cogent Diamond Hydride column (4 µm, 100 Å, 3 × 150 mm, Microsolv Technology Corp.). The flow rate was 0.5 mL/min. The LC conditions were: 10% solvent A, hold for 1 min; 10 to 70% solvent A in 19 min; 70% solvent A, hold for 1 min; 70 to 10% solvent A in 4 min; and 4 min equilibration at 10% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile). High resolution mass spectrometry data were obtained using an Agilent 1200 series LC system coupled to an Agilent 6530 quadrupole time-of-flight (qTOF) mass spectrometer

with an ESI source. The mass spectra data were recorded in either positive or negative ionization mode with a mass range of 100 to 1700 m/z; spectra rate, 1 spectra/s; capillary voltage, 3500 V; nebulizer pressure, 35 psi; drying gas (N2) flow, 8 L/min; temperature, 275 ºC. For MS2 settings, mass range, 100 to 1700 m/z; collision energy, 10 eV; isolation width MS/MS, ~1.3 amu; spectra rate, 1 spectra/s; nebulizer pressure, 35 psi; drying gas (N2) flow, 8 L/min; precursor threshold, 200 counts; temperature, 275 ºC. The extracted ion chromatograms were generated using the $[M+H]^+$ or $[M-H]^-$ molecular ions with a 10 ppm window.The chromatographic datasets were aligned by retention time and mass, and the aligned data was statistically analyzed using the XCMS online software (https://xcmsonline.scripps.edu).[11] The results were then extracted from XCMS-processed data using the following parameters: $p < 0.05$, maximum intensity > 10,000, and fold change > 5.

For the experiments in which we fed d_3-methyl-L-NMA, the cultures were fed 1 mM d_3-methyl-L-NMA immediately after inoculation and analyzed in a similar manner after 3 days. Triplicates were performed for comparative metabolomics experiments.

Overexpression and purification of SznH and SznK

PCR reaction mixtures contained 22.75 µl of water, 0.125 µl of forward and reverse primers (sznH-F:TATCATATGGGCAGGGACGGGGACAAGCC,
sznH-R:ATAAAGCTTTCATCGGGGCGGCTCCTGCG;
sznK-F:TATCATATGGCGGGCGCTGACCCGGG,
sznK-R: ATAAAGCTTTCATGGCCCGGTGATGACCA),
0.5 µl of genomic DNA, 1.5 µl of DMSO and 25 µl of Q5 High-Fidelity 2x Master Mix (New England Biolabs) in a final volume of 50 µl. Thermocycling was carried out in a MyCycler gradient

cycler (Bio-Rad) using the following parameters: denaturation for 1 min at 98 °C; 35 cycles of 0.5 min at 98 °C, 0.5 min at 70 °C, 2 min at 72 °C; and a final extension time of 10 min at 72 °C. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide or SYBR safe staining, pooled, and purified. Amplified fragments were digested with Fastdigest NdeI and Fastdigest HindIII (Thermo Fisher Scientific) for 2 h at 37 °C. Digests contained 16.5 µl purified PCR product, 2 µl of FastDigest Buffer (10×), 0.4 µl of NdeI (20 U/µl) and 0.4 µl of HindIII (20 /µl). Restriction digests for PCR were purified directly using Zymo DNA Clean & Concentrator. To prepare a linearized pET28a vector, 50 ng of the vector was digested with NdeI and HindIII as described with the exception of the addition of 1 µl calf-intestinal alkaline phosphatase in the reaction after 2 h and further incubation for 1 h. The linearized DNA was purified first by agarose gel electrophoresis and further purified with Zymoclean Gel DNA Recovery Kit. The insert digest was ligated into linearized expression vector pET28a using T4 DNA ligase (New England Biolabs). Ligations were run overnight at 16 °C and contained 1 µl T4 Ligase Buffer (10×), 1.5 µl digested vector, 5.5 µl digested insert DNA and 2 µl T4 DNA Ligase (400 U/µl). 5 µl of each ligation was used to transform a single tube of *E. coli* TOP10 cells (Invitrogen). The identity of the resulting pET28a–SznH and pET28a−SznK constructs were confirmed by sequencing of purified plasmid DNA. These constructs were used to transform chemically competent *E. coli* BL21 (DE3) cells (Invitrogen) and stored at –80 °C as frozen glycerol stocks.

A 50 ml starter culture of pET28a–SznH and pET28a−SznK BL21 *E. coli* was inoculated from a single colony and grown overnight at 37 °C in LB medium supplemented with 50 µg/ml kanamycin. 7 ml of the saturated starter culture was inoculated into 700 ml of LB. The cultures were incubated at 37 °C with shaking for around 2.5 h, induced with 250 µM IPTG at an OD_{600} of around 0.5, and then incubated at 15 °C for 16 h. Cells were pelleted by centrifugation

(6,700*g* for 10 min) and resuspended in 40 ml of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM $MgCl_2$, pH 8). The cells were lysed by passage through a cell disruptor (Avestin EmulsiFlex-C3) twice at 10,000 psi, and the lysate was clarified by centrifugation (20,000*g* for 40 min). The supernatant was incubated with 2 ml of Ni-NTA resin for 1 h at 4 °C. The mixture was then loaded into a glass column and washed with 20 ml of wash buffer (50 mM HEPES, 20 mM imidazole, 500 mM NaCl, 10 mM $MgCl_2$, pH 8.0). The protein was eluted with approximately 20 ml of elution buffer (50 mM HEPES, 200 mM imidazole, 500 mM NaCl, 10 mM $MgCl_2$, pH 8.0). SDS–PAGE analysis (4–15% Tris-HCl gel) confirmed the presence and purity of SznH and SznK. The eluted protein was concentrated to around 1 ml using a Corning Spin-X UF 20 ml Centrifugal Concentrator (30,000 molecular weight cut-off membrane) after centrifugation at 1,145*g*. 12 ml of exchange buffer (20 mM HEPES, 200 mM NaCl, 10% glycerol, pH 8.0) was added, and the sample was concentrated again to 1 ml. This process was repeated once more before the concentrated, desalted solutions containing purified SznH and SznK was frozen in liquid N_2 and stored at –80 °C.

ATP consumption assays with PK/LDH coupled assay

In a total volume of 50 µl, 50 mM MOPS pH 7.5, 1 mM of **3.2** and 80 µM of SznF were mixed and incubated at room temperature for 1 h to generate **3.3** Then, a second reaction was prepared in a UV-STAR® (Greiner) 96-well plate with a total volume of 100 ul. This reaction contained 100 mM MOPS pH 7.5, 10 mM $MgCl_2$, 0.2 mM phosphoenolpyruvate, 0.2 mM NADH, 1 mM L-amino acid, 4 mM ATP, 24 units of rabbit pyruvate kinase/lactate dehydrogenase (Sigma Aldrich), and 50 µl of the SznF-reaction containing **3.3**. The ATP consumption assay was initiated with the addition of 10 µM of SznK. Reaction progress was monitored at 340 nm with a BioTek Gen5 Microplate Reader with a time interval of 10 seconds. Control reactions omitting SznK, ATP, and amino acids were also prepared in the adjacent wells.

LC–MS assay of SznH and SznK activity

In a total volume of 50 µl, 50 mM MOPS pH 7.5, 1 mM of **3.2**, and 80 µM of SznF were mixed and incubated at room temperature for 1 h to generate **3.3.** Then, the reaction mixture was diluted with 25 ul of water. Next, 1 ul of 1 M $MgCl_2$ (final conc. 10 mM), 10 µl of 10 mM L-amino acid (final conc. 1 mM), 10 µl of 200 µM SznH or SznK (Final conc. 20 µM) were added. The reaction was initiated by adding 4 µl of 100 mM ATP (final conc. 4 mM). After 1 h at room temperature, the reaction mixtures were quenched by adding 100 µl of LC–MS grade methanol. After 10 min on ice, the mixtures were centrifuged to remove precipitated protein.

The analytes were then derivatized with 9-fluorenylmethyl *N*-succinimidyl carbonate (Fmoc-OSu). A solution of Fmoc-OSu acetonitrile solution was first prepared. An additional 100 µl of acetonitrile was added to the quenched reaction mixtures, followed by 10 µl of the 10 mM Fmoc-OSu solution. This mixture was incubated for 1 h, centrifuged to remove precipitates, and analyzed by reversed phase LC–HRMS. The LC column was an Acclaim Polar Advantage II C18 column (3 µm, 120 Å, 2.1 × 150 mm, Thermo Fisher Scientific). The flow rate was 0.3 ml/min. The LC conditions were: 95% to 5% solvent A in 15 min; 5% solvent A, hold for 5 min; 5% to 95% solvent A in 4 min; and 5 min equilibration at 95% solvent A (solvent A = 0.1% formic acid in water, solvent B = 0.1% formic acid in acetonitrile).

Phylogeny of SznF and gene cluster analysis

The maximum likelihood phylogenetic tree was generated using sequences of SznF that contain both the central helix domain and cupin domain using the NCBI non-redundant (nr) sequences Database (2018, evalue cutoff < 1E-50). The tree was calculated with MEGA-X with

50 bootstrap replications, LG substitution model, and Nearest Neighbor-Interchange heuristic model.[18] Genome neighborhood analysis of SznF homologs were performed with the software tools available at Integrated Microbial Genomes and Microbiome system (Joint Genomic Institute).[19]

Bioinformatics analysis of *szn* clusters

We initially search for all members of the haem_oxygenase_2 PFAM (*i.e.* the CADD domain), which returned roughly 18,000 genes. Because of the large number of genes, we performed the phylogenetic analysis for the top 2,700 family members ordered by sequence similarity by BLAST. After manually confirming intact scaffolds and choosing representative sequences (1613 sequences at 95% sequence identity), we generated genome neighborhoods consisting of ten genes flanking the SznF homolog in each direction. Some possible gene clusters of interest are highlighted in the **Figure 4.14**.

The sequence similarity network was generated from the 1613 sequences with a 1E-70 evalue cutoff value. Within each cluster, sequences that contain 'conserved' metal centers are highlighted. Our criteria for 'conserved' metal centers for the cupin domain are the three histidine residues; for the CADD domain are more than five conserved metal binding residues. In our analysis, 83% only have a functional CADD domain; about equal numbers are SznF-like (7.6%) or only have a functional cupin domain (7.1%).

4.4 References

1) Ng, T.; Rohac, R.; Mitchell, A. J.; Boal, A. K.; Balskus, E. P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **2019**, *566*, 94.

2) Fawaz, M. V.; Topper, M.; Firestine, S. M. The ATP-Grasp Enzymes. *Bioorg. Chem.* **2011**, *39*, 185.

3) Iyer, L. M.; Abhiman, S.; Burroughs, A. M.; Aravind, L. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: Synthesis of novel metabolites and peptide modifications of proteins. *Mol. BioSyst.* **2009**, *5*, 1636.

4) Walsh, C. T. Enzymes in the D-alanine branch of bacterial cell wall peptidoglycan assembly. *J. Biol. Chem.* **1989**, *264*, 2393.

5) Tsuda, T.; Asami, M.; Koguchi, Y.; Kojima, S. Single Mutation Alters the Substrate Specificity of l-Amino Acid Ligase. *Biochemistry* **2014**, *53*, 2650.

6) Zhang, Y.; Li, K.; Yang, G.; McBride, J. L.; Bruner, S. D.; Ding, Y. A distributive peptide cyclase processes multiple microviridin core peptides within a single polypeptide substrate. *Nat. Commun.* **2018**, *9*, 1780.

7) Walker, M. C.; van der Donk, W. A. The Many Roles of Glutamate in Metabolism J Ind *Microbiol. Biotechnol.* **2016**, *43*, 419.

8) Ye, D.; Wei, M.; McGuire, M.; Huang, K.; Kapadia, G.; Herzberg, O.; Martin, B. M.; Dunaway-Mariano, D. Investigation of the catalytic site within the ATP-grasp domain of Clostridium symbiosum pyruvate phosphate dikinase. *J. Biol. Chem.* **2001**, *276*, 37630.

9) Galperin, M. Y.; Koonin, E. V. A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Science*. **1997**, *6*, 2639.

10) Miller, G. J.; Wilson, M. P; Majerus, P. W.; Hurley, J. H. *Mol Cell*. **2005**, *18*, 201.

11) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **2012**, *84*, 5035.

12) Newton, G. L.; Buchmeier, N.; Fahey, R. C. Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 471.

13) Sevin D.C.; Fuhrer, T.; Zamboni, N.; Sauer, U. "Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli." *Nat. Methods* **2017**, *14*, 187.

14) Tsuda, T.; Asami, M.; Koguchi, Y.; Kojima, S. Single Mutation Alters the Substrate Specificity of l-Amino Acid Ligase. *Biochemistry* **2014**, *53*, 2650.

15) Thoden, J. B.; Firestine, S. M.; Benkovic, S. J.; Holden, H. M. PurT-encoded glycinamide ribonucleotide transformylase. Accommodation of adenosine nucleotide analogs within the active site. *J. Biol. Chem.* **2002**, *277*, 23898.

16) Kagawa, W.; Arai, T.; Ishikura, S.; Kino, K.; Kurumizaka, H. Structure of RizA, an L-amino-acid ligase from Bacillus subtilis. *Acta Crystallogr. F Struct. Biol. Commun.* **2015**, *71*, 1125.

17) Johnston, C. W.; Plumb, J.; Li, X.; Grinstein, S.; Magarvey, N. A. Informatic analysis reveals Legionella as a source of novel natural products. *Synth. Syst. Biotechnol.* **2016**, *1*, 130.

18) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547.

19) Markowitz, V. M.; Chen, I. M.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; Huntemann, M.; Anderson, I.; Mavromatis, K.; Ivanova, N. N.; Kyrpides, N. C. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **2012**, *40*, D115.