Doctoral Dissertation

Studies on natural products containing nitrogen-nitrogen bond

（窒素-窒素結合を含む天然物に関する研究）

Atina Rizkiya Choirunnisa

Laboratory of Natural Products Chemistry

Graduate School of Life Science, Hokkaido University

March 2024

博士学位論文

Studies on natural products containing nitrogen-nitrogen bond

（窒素-窒素結合を含む天然物に関する研究）

Atina Rizkiya Choirunnisa

北海道大学大学院

生命科学院 生命科学専攻 生命医薬科学コース

天然物化学研究室

2024 年3 月

TABLE OF CONTENT

Abbreviation and Symbol

Ala	Alanine
APS	Ammonium persulfate
ATP	Adenosin triphosphate
BCG	Biosynthetic Gene Cluster
CBB	Coomassie brilliant blue
$CHCl_3$	Chloroform
DAP	Diaminopropionic acid
DAB	Diaminobenzaldehyde
DABA	2,4-Diaminobutyric acid
DMSO	Dimethyl sulfoxide
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
E. coli	*Eschericia coli*
Gly	Glycine
Glu	Glutamic acid
HMM	Hidden Markov Model
HPLC	High performance liquid chromatography
HS	Hydrazine synthetase
ISP2	International Streptomyces Project 2
LB	Lysogeni Broth
LC	Liquid chromatography
Lys	Lysine
MeCN	Acetonitrile
MeOH	Methanol
MESO-1	Human Lung Mesothelia Cell
MetRS	Methionyl tRNA synthetase
MQ	Milli-Q
NaCl	Sodium chloride

NBRC	NITE Biological Research Centre
NMO	*N*-hydroxylating monooxygenase / *N*-hydroxylase
N-N	Nitrogen-Nitrogen
NP	Natural product
Orn	Ornithine
PBS	Phosphate buffer saline
QG buffer	Qiagen buffer
SDS PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SFM	Soya Flour Mannitol
SMMS	Supplemented liquid minimal media
TAE buffer	Tris acetate EDTA buffer
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
tRNA	Transfer ribonucleic acid
TSB	Tryptic Soy Broth
UPLC	Ultra performance liquid chromatography
β	beta
p-	para
2xYT	2x Yeast extract tripton

Chapter 1 General Introduction

1.1 Natural Products

Natural products (NPs) are metabolites produced by living organisms. The complex structures and unique physicochemical properties of natural products play significant roles for their diverse biological activities, making them prolific sources of lead compounds in drug discovery. In general, natural products are also referred to as secondary or special metabolites.[1–4] Natural products do not have an essential role for the survival of organisms. They mostly have roles for adaptation to environment, protection, chemical communication, etc. Notable examples include polyketide compounds, phenylpropanoids, alkaloids, and terpenes[2] (Figure 1.1).

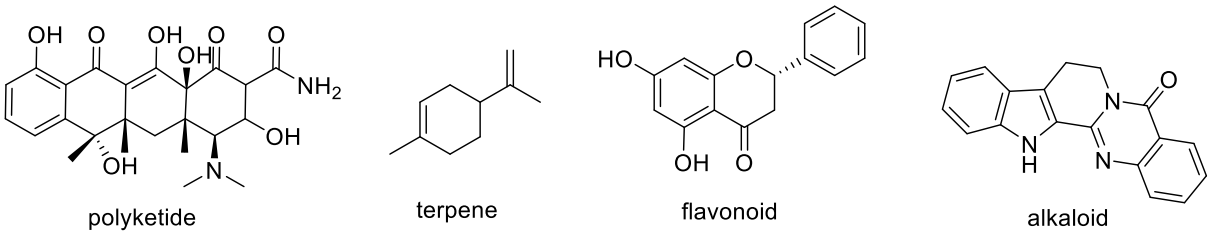


Figure 1.1 Natural products

1.2 Natural Products in Drug Discovery

According to Newman and Cragg (2020), over the last 4 decades natural products have made a significant contribution for drug development. Among 1881 new approved drugs during period 1981-2019, most of them were originally derived from NPs and NP-inspired compounds (Figure 1.2).[1] For example, in drug development for cancer treatment during period 1946-1981, 40 of 75 molecules are NPs or NPs derivatives[1]. Currently, about 60% of approved drugs are related to NPs (64 natural products, 299 natural product derivatives, 268 natural product mimics).[3,5]

The earliest written records of using plant-based natural products for therapeutic applications were found in ancient Mesopotamia medicinal system from 2900-2600 BCE.[6] In 1900's, most of NP-based medicine were derived from plants. In the early of 19th century, Friedrich Serturner did comprehensive research on isolation, pharmacological activity, and

crystallization of analgesic agent from opium, which he named morphium (morphine), followed by commercialisation of morphine by H.E.Merck from Germany in 1826.[4,7] The first semi-synthetic NP aspirin, originally derived from salicin (isolated from *Salix alba*) was introduced by Bayer in 1889.[8] Another notable drug discovery is artemisinin, a first line drug for malaria therapy, from the sweet wormwood *Artemisia annua*.[9]

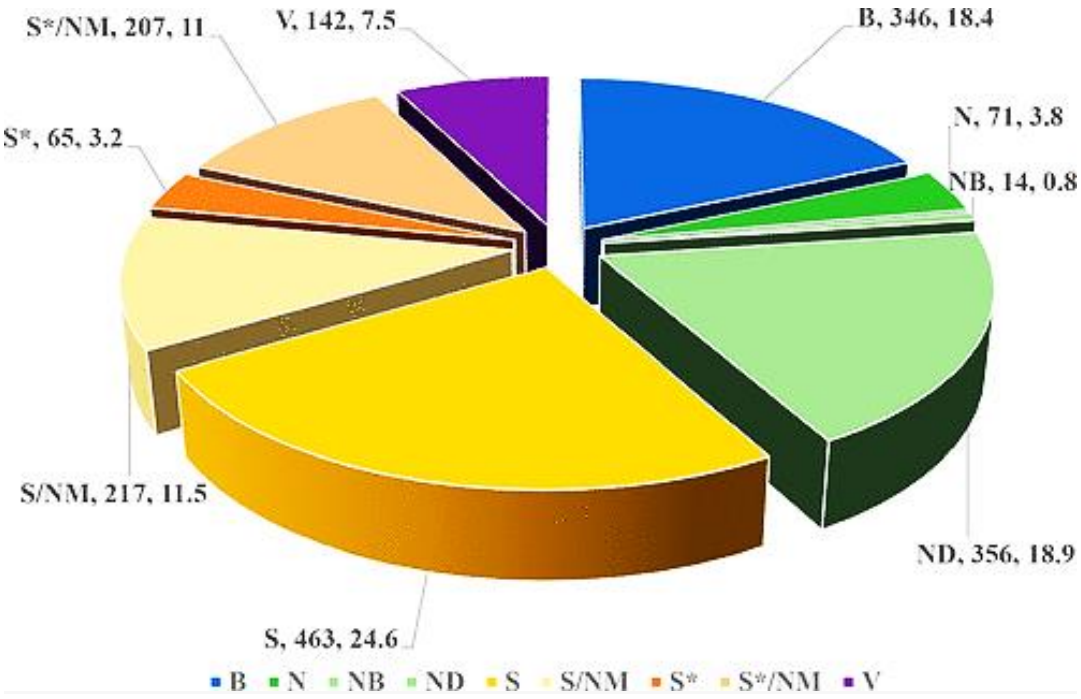


Figure. 1.2 New approved drugs from 1981-2019[1]

B: Biological macromolecule, N: Unaltered natural products, NB: Botanical drugs, ND: Natural product derivative, S: Synthetic drug, S*: Synthetic drug (NPs pharmacophore), V: Vaccine, /NM: NPs mimic

The exploration of NPs from other sources such as microbes was started by discovery of penicillin from *Penicillium notatum* by Alexander Flemming in 1928. Subsequent purification and medicinal applications of penicillin for treating bacterial infection were accomplished by Howard Walter Florey and Ernst Boris Chain, which marked the beginning of antibiotic discovery era.[10] Since then, microbial NPs were extensively used in a wide range of applications including medicines, agriculture, and food industry. Discovery of antibiotics, such as streptomycin, an antituberculosis agent from *Streptomyces griceus*, cephalosporin from *Cephalosporium acremonium*, vancomycin from *Amycolatopsis orientalis* also had contributed significantly to drug development at that time.[11–13]

Table 1. List of Natural Product derived Drugs[14]

	Activity
Ephedrine	An alpha and beta-adrenergic agonist indicated to treat hypotension under anesthesia, allergic conditions, bronchial asthma, and nasal congestion
Levomenthol	Treat mild to moderate muscle and joint pain.
Quinine	Used as antimalarial drug
Echinacoside	Under investigation for the treatment of Parkinson's, Alzheimer's, atherosclerosis, osteoporosis, acute colitis, wound treatment, and hepatitis
Sennosides	Treat constipation
Artemisinin	Treatment of malaria
Penicillin	Antibiotic for gram positive bacteria
Cephalosporin	Wide spectrum antibiotic
Griseofulvin	Dermatophytic fungi
Rifamycin	Treat tuberculosis
Vancomycin	Antibiotic for gram positive bacteria
Amphotericin B	Work as antifungi
Polymixin B	Antibiotic for gram negative bacteria
Morphine	Used for pain medication
Curcumin	Commonly used as immunostimulant
Digoxin	Used to treat various heart conditions. Most frequently it is used for atrial fibrillation, atrial flutter, and heart failure
Paclitaxel	For cancer chemotherapy
Aspirin	Used as analgesic
Reserpin	Drug for treatment hypertension
Atropine	Anticholinergic activity

1.3 Nitrogen-Nitrogen Bond Natural Products

NPs containing Nitrogen-Nitrogen (N–N) bond belong to rare group of secondary metabolites. Since the first discovery of macrozamin from *Macrozamia spiralis*, an endemic

cycad from Australia in 1951, over 300 N–N bond NPs were already discovered.[15] N–N bond containing compounds are widely known for high reactivity, thus contributing to their diverse biological activities including antibacterial, cytotoxic, and antifungal activities.[16]

N–N bond-containing NPs are greatly diverse not only in their bioactivities but also in chemical structures. Some N–N bond-containing NPs discovered recently have diverse structural features such as azoxy, nitrosamine, nitroso hydroxylamine, azo, diazo, hydrazides, hydrazones, hydrazines, azines, pyrazoles and indazoles, pyridazines and cinnolines, triazines, triazoles, and N–N linked heterocycle groups (Figure 1.3).[17]

In 1954, elaiomycin was isolated from *Streptomyces hepaticus*. This azoxy compound was found to be hepatotoxic and carcinogenic to rodents. However it showed antibacterial activity towards *Mycobacterium tuberculosis*.[18] Moreover, elaiomycin E exhibited the strongest antiproliferative activity, which contributed from the 3-hydroxy group in its structure.[19] Another interesting example is streptozotocin isolated from *Streptomyces acromygenes,* which has nitrosamine groups and exhibited anti-cancer activity.[20] Currently, streptozotocin is used for chemotherapy, especially against gastrointestinal and pancreatic cancers. Further investigating N–N bond-containing compound led to the discovery of triacsin A and B from *Streptomyces sp.* SK-1894.[21] It was found that triacsins exhibited cytotoxic activities against Vero, HeLa, and Raji cells as well as anti-malarial and vasodilatory activities[22].

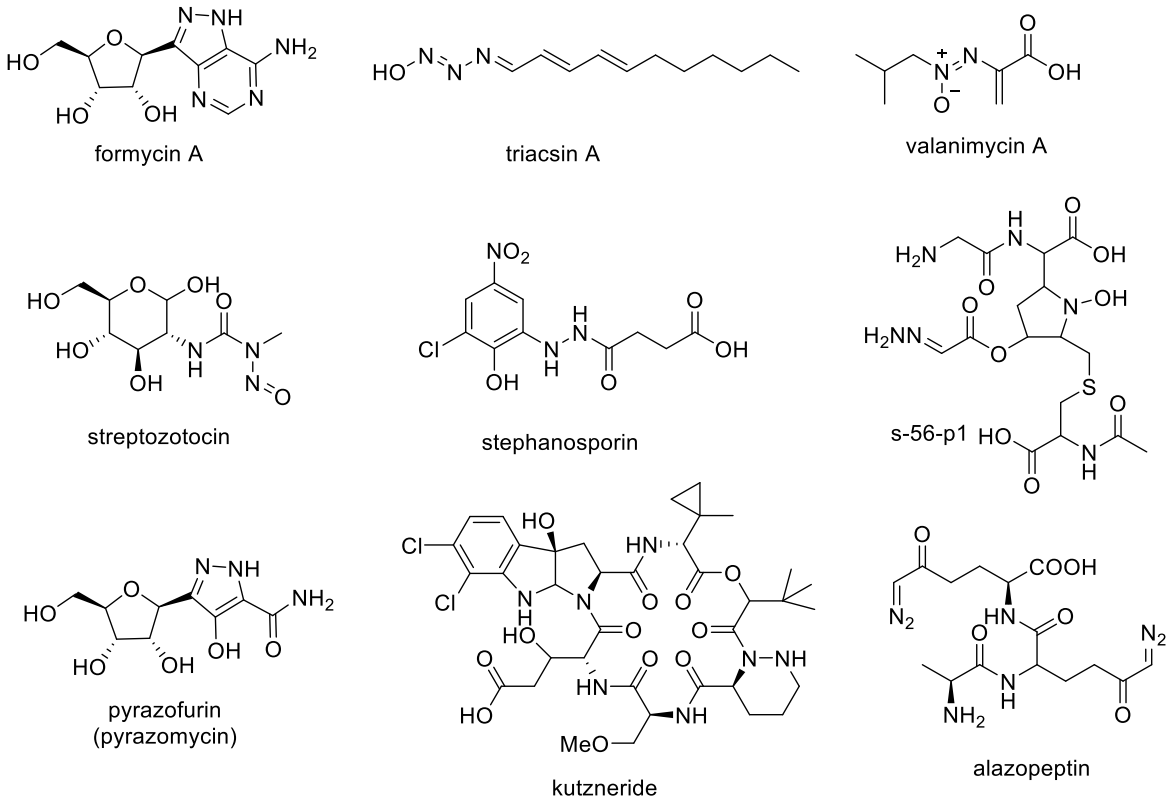


Figure 1.3 Examples of Nitrogen-Nitrogen Natural Products

1.4 Genome Mining for Natural Products Discovery

The exploration of novel natural products has traditionally relied on conventional isolation processes. Organisms are typically selected for isolating target compounds based on empirical or traditional uses, organoleptic characteristics, pharmacological activities, and physicochemical properties.[25] While these classical approaches remain reliable, they present a low probability of discovering novel compounds. Hence, there is a pressing need for a new approach to enhance the novelty in natural product discovery.

In this genomic era, exploiting genome information has significant role for the discovery of NPs especially in efforts to identify secondary metabolites encoded by uncharacterized gene clusters and predict their targets or mechanism of action. Structurally complex and diverse NPs are synthesized by biological systems from various chemical building blocks through catalytic reactions of enzymes encoded on genomes. This allows researchers to mine genetic signatures of enzymes for finding new biosynthetic gene clusters (BGCs) from genome sequence datasets through a new approach popularly called genome mining. The terminology of genome mining corresponds for bioinformatics study to identify biosynthetic pathways of natural products and their possible chemical and functional interactions. Genome mining includes DNA sequence analyses to identify uncharacterized biosynthetic gene clusters, followed with the identification of the BGC-encoded natural products.[23] Some of common bioinformatic tools used for genome mining include AntiSMASH, ClustSCAN, NP searcher, and GNP/PRISM. Among them, AntiSMASH is the most widely used tool, because it is visualized in a user-friendly interactive for predicting a wide range of biosynthetic gene cluster.[24]

In genome mining, several strategies to identify new BGCs include targeting reactive chemical features, ligand-binding features, compound family defining features, and resistance genes.[25] Compound family-based genome mining is aimed at targeting specific enzymes or domains of the biosynthetic pathways that contribute to the formation of certain chemical groups or moieties.[25]

Genome mining that targets bioactive features.

While most of NPs have complex chemical structures, they often have chemical features that contribute to biological activities. Bioactive features can be categorized into two types. The first category is reactive features, including radical, electrophilic, and nucleophilic reactivity which could lead to covalent binding to target protein. For example, previous study by

Wackett group to identify new β-lactones NPs using β-lactone synthetase as starting point. β-lactones are one of electrophilic chemical structure that often contribute to biological activities. In their study, OleC, an enzyme that catalyze the formation of β-lactones in the biosynthetic pathway of olefin hydrocarbon, was used as starting point to find other NPs biosynthesized by similar enzyme family β-lactones synthetases. Using AdenylPred tools, they retrieved 90 β-lactones synthetases from approximately 50,000 BGCs available in the public database. Further functional study enabled them to connect one of the uncharacterized BGCs to a β-lactone compound called nocardiolactone.[26]

Meanwhile, the second category is structural features that make non-covalent binding with target protein, such as ligand binding structure. For example, in genome mining that targeted siderophores containg diazoniumdiolate moiety, an unusual feature for iron binding, Hertweck et al. conducted genome mining using GrbD, one of essential enzymes for the formation of diazonium diolate moieties, led to the identification of 37 candidate BGCs. This subsequently provided a basis for the discovery of a new family of siderophores megapolibactin A-F. [27]

Compound family mining.

Compound families refer to groups of compounds that are structurally related and share the similar biosynthetic origin. In genome mining that targets compound families, genes contributing to chemical structures are used as queries for searching homologues BGCs. A notable example is finding new glycopeptide antibiotics by Gerry Wright and colleagues, where they explored the phylogenetic relationship between known glycopeptide antibiotics such as vancomycin and other glycopeptide related BGCs identified by genome mining. Of 71 BGC that they examined, most of them contained typical vancomycin resistance genes (*vanHax*). Subsequently they focused on the remaining BGCs lacking a known resistance determinant, which led to the identification of a BGC clade encoding the newly discovered corbomycin. The ability of corbomycin to block the activity of multiple autolysins by binding to peptidoglycan, was a novel antibacterial mechanism of action for glycopeptide antibiotics. [28]

Resistance gene-based mining / Targeted genome mining.

Genome mining targeting chemical features or compound family requires prior information on NP structures of interest. However, in case chemical structures are unavailable or not required, genome mining could also be directed to target BGCs harboring certain genes, such

as resistance genes to prioritize certain BGCs for further genome mining study. This so-called prioritization strategy relied on the fact that bioactive molecule-producing organisms have resistance mechanisms to deal with their own toxic products, preventing them from unexpected suicide. A notable example of using resistance gene as a guide to find novel bioactive analogs was reported by Moore group in 2015, where they performed targeted genome mining using *tlmE*, a gene that showed high similarity to another self-resistance gene found in another BGC encoding FASII (Fatty acid synthetase) inhibitor. Using this approach, they identified 912 orthologous group, which subsequently led to the discovery of several thiolactomycin analogs.[29]

1.5 Biosynthetic Pathway of Nitrogen-Nitrogen Bond Natural Products

The unique chemical structures and promising biological activities of N–N bond-containing NPs have attracted researchers to reveal the biosynthetic mechanisms on how this unique family of compounds are made by biological systems. Until recently, there are several known mechanisms of N–N bond formation in nature, such as through nitrous acid, metalloenzymes, and radical recombination.

A. Formation of N–N bond through Nitrous Acid Pathway (Aspartate Nitro Succinate Pathway)

Although nitrous acid has been used for the chemical synthesis of compounds containing N–N bond since long time, the involvement of nitrous acid in the NP biosynthetic pathway in living organism has just been recently described through the discovery of azamerone biosynthetic pathway reported for the first time in 2009 by Moore group.[30] Further discovery and heterologous expression of cremeomycin BGC represent the first experimental success in connecting a diazo-containing metabolite to its encoded genes. In principle, the cremeomycin biosynthetic pathway that consist of two enzymes CreD dan CreE (later called aspartate/nitro succinate pathway abbreviated as ANS), in which CreE mediated the oxidation of aspartate to nitro succinate, and CreD catalyzed the formation of fumarate with the release of nitrous acid. Then an ATP-dependent ligase like enzyme CreM catalyzes diazo formation by coupling the nitrous acid and an aniline.[31] Enzymatic formation of N–N bonds from nitrous acid was also found in the biosynthetic pathway of triacsin, where disruption of *tri21* (encoding flavin dependent enzyme) and *tri16* genes (encoding lyase) abolished the production of triacsin.[32] Recent genome mining of ANS enzymes resulted in the discovery of

desferrioxamine, a potent multidentate iron binder characterized with five rings at the terminal core structures.[33]

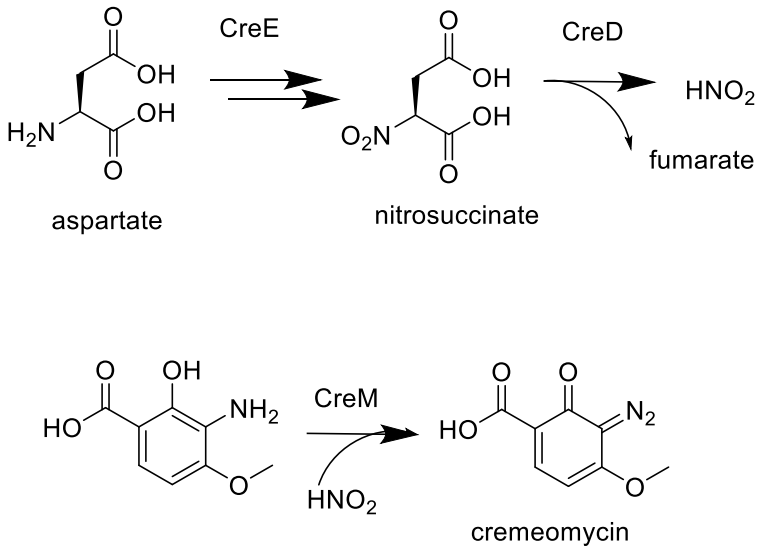


Figure 1.4 Formation of N-N bond by nitrous acid in cremeomycin biosynthesis pathway

B. Formation of N–N bond Using Metalloenzymes

N–N bond formation through metalloenzymes are involved in the biosynthetic pathway of streptozotocin, as reported by Balskus group and Boal group.[34] Their studies concluded that the two nitrogen atoms in the *N*-nitroso group of streptozotocin were derived from arginine, and the N–N bond formation were catalysed by two enzymes SznE and SznF[35]. In the early reaction, SznE catalyzed the *N*-methylation of arginine and subsequently SznF catalysed electron *N*-oxidation of *N*-methyl arginine followed by an oxidative C–N bond cleavage and subsequent N–N bond formation, resulting in *N*-hydroxy-*N*-methyl-*N*-nitroso-citrulline. Based on the X-Ray analysis, SznF was homodimer, which each of monomer contains an *N*-terminal domain, a seven-helix bundle domain and a *C*-terminal cupin domain. Especially, in the cupin domain, Fe^{2+} ion was attached by three histidine residues. The co-crystal of SznF and *N*-hydroxy-*N*-methylarginine showed the binding of SznF to the metal ion, indicating that the importance of Fe^{2+} for the intramolecular rearrangement to generate *N*-nitroso group.[35]

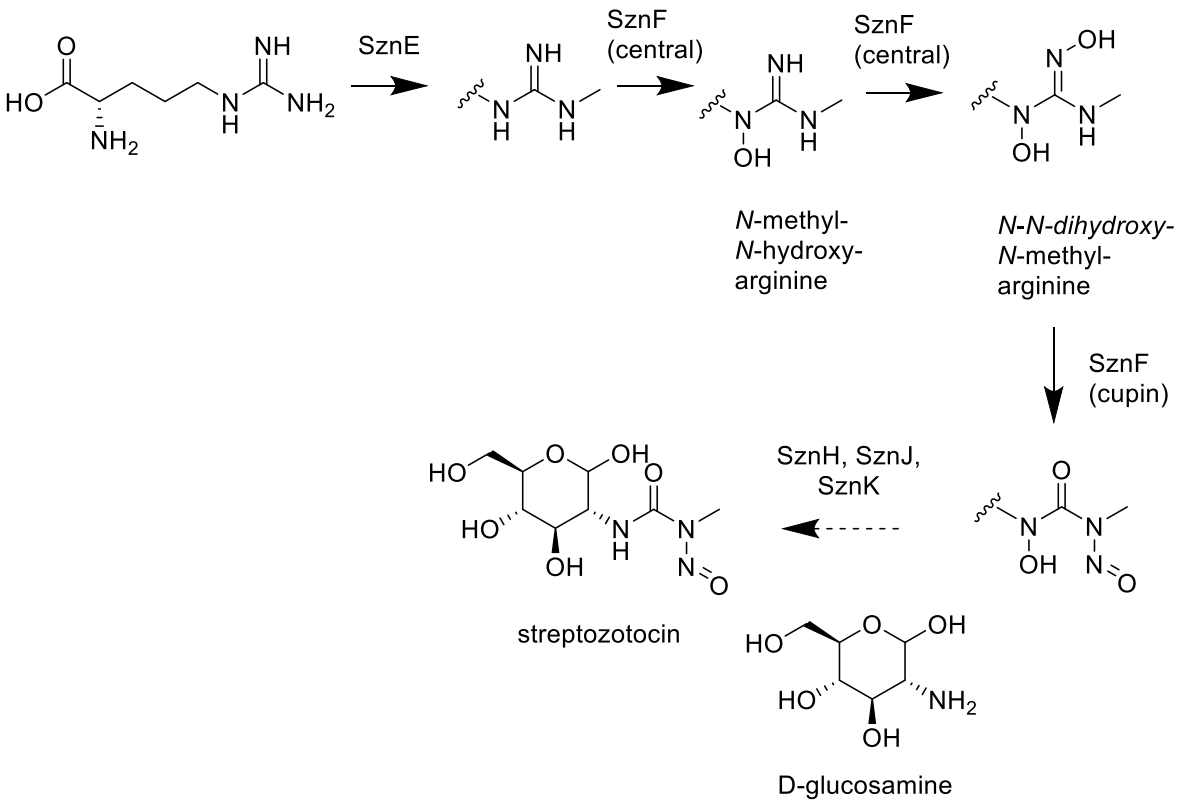


Figure 1.5 N–N bond formation in streptozotocin

Another enzymatic reaction that involves a metal binding enzyme in N–N bond formation is found in the biosynthetic pathway of piperazic acid. In this pathway, KtzT was responsible for converting N^5-OH L-ornithine (synthesized by an FAD-dependent monooxygenase KtzI) to piperazic acid known as a building block for non-ribosomal peptides, such as kutzneride. KtzT has affinity to heme, making it capable of catalysing the reactions in two different iron states (Fe^{2+} and Fe^{3+}) in the absence of molecular oxygen. [36]

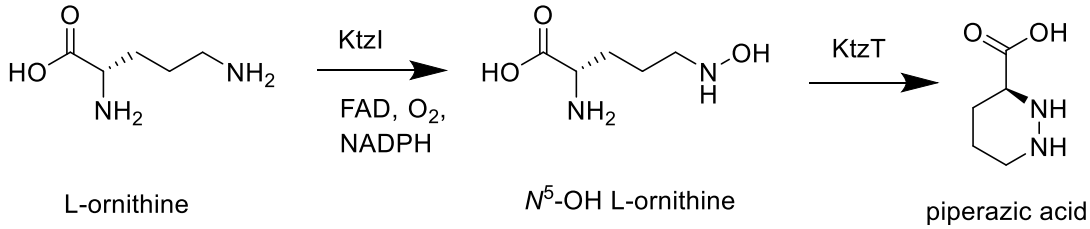


Figure 1.6 N-N bond formation in piperazic acid

Another example of metalloenzymes contributing to the formation of N–N bound is a group of fusion proteins composed of a metal-binding cupin domain and a methionyl-tRNA synthetase (MetRS)-like domain. These di-domain proteins catalyze N−N bond formation

between a hydroxylamine and another amino acid to form N−N bonds. This is initiated by the installation of hydroxy to amino acid catalyzed by *N*-hydroxylase (NMO). Then, MetRS-like domain activates the second amino acid and conjugate it to hydroxylamine, thereby generating *O*-acyl hydroxylamine. Then, cupin domain enables oxidative rearrangement to form a N–N bond-containing NPs.[37] This fusion of MetRS-like domain and cupin domain is hereafter referred to as "hydrazine synthetase".

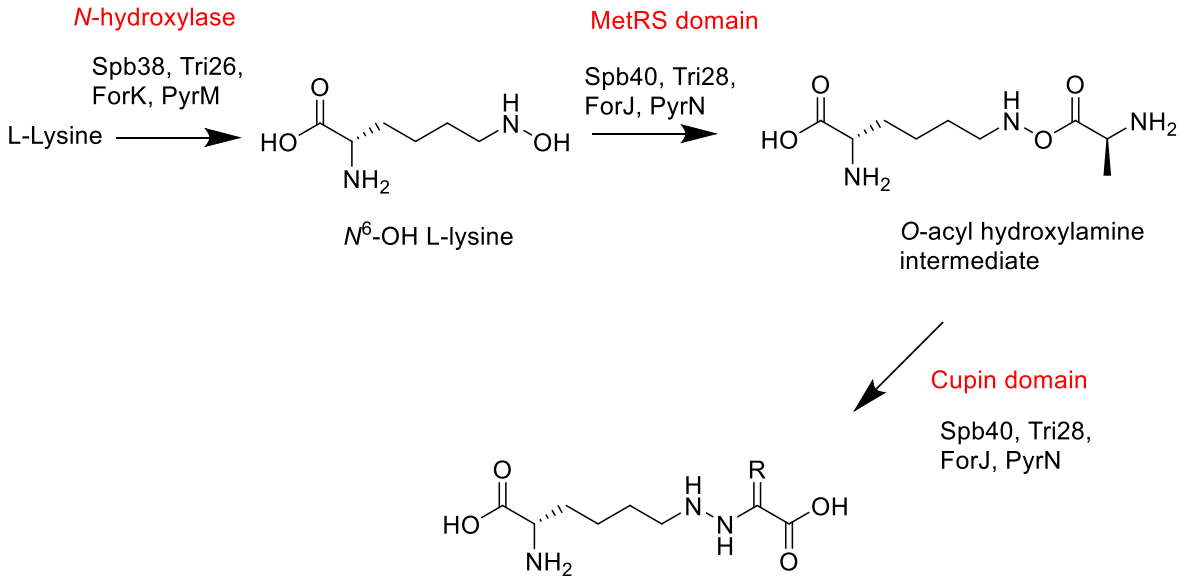


Figure 1.7 N-N bond formation catalyzed by di-domain MetRS and cupin enzymes

C. Nitrogen Transient Radical Recombination Reaction

N–N bonds can be formed without the involvement of any enzymatic reaction. For example, nitroso analogue is initially generated through amine oxidation catalyzed by non-heme di-iron *N*-oxygenase AzoC in azoxymycin biosynthetic pathway. Then, redox coenzymes catalyze the conversion between nitroso group and hydroxylamine via radical transient, followed by dimerization to form azoxy moiety.

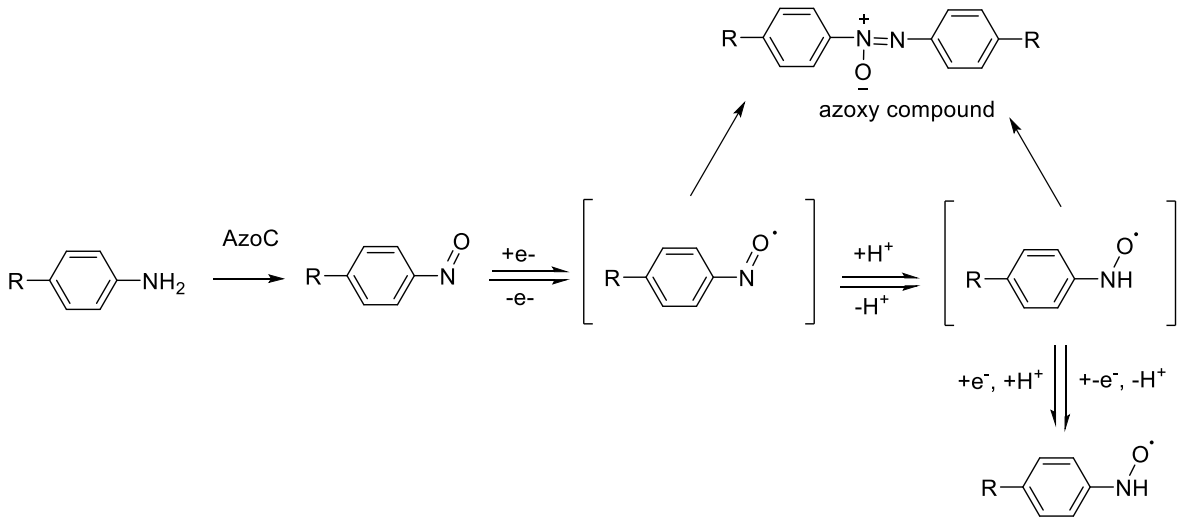


Figure 1.8 N-N bond formation in azoxymycin biosynthesis pathway

1.6 Research Objectives

In this study, we focused on two different genomic based approaches. In the first part, we used reactivity-based genome mining aiming to discover N–N bond compounds. For the reactivity-based screening, we used *p*-dimethylaminobenzaldehyde (DAB) assay, which is a reagent that can detect the presence of hydrazine compound generated from hydrolysis of azoxy compound by acid.

In the second part, we used phylogeny-based genome mining to explore the specific family enzyme contributed for the formation of N–N bond in bacteria. We aim to explore the substrate recognition characteristic and final product resulting from the reaction catalyzed by the enzyme.

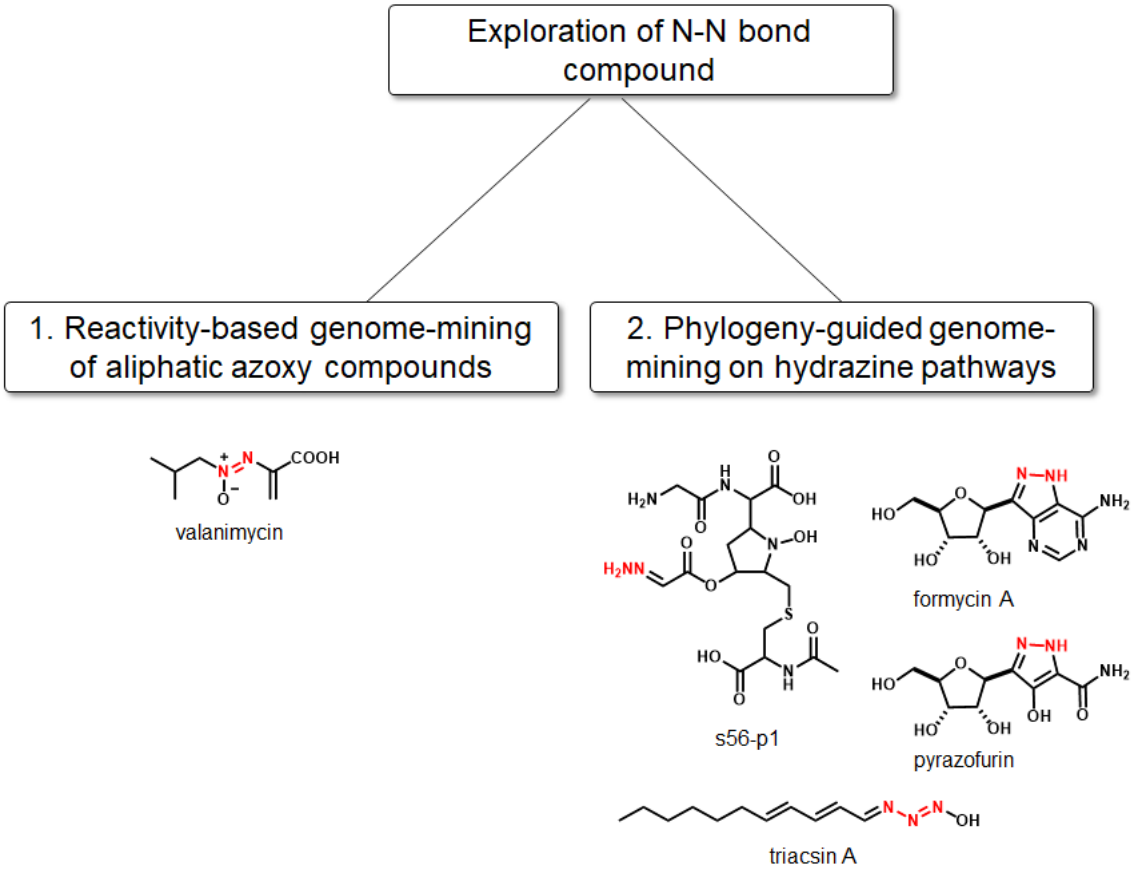


Figure. 1.5 Road map of the research

As mentioned in the previous part, there are 3 related enzymes: *N*-hydroxylase, MetRS homolog, and cupin protein. Even though the catalytic mechanism of these enzymes

might be similar among bacteria, the characteristics of each enzyme might be different, including their substrate specificity. Using genome mining, we would like to explore the diversity of NMO, and 'hydrazine synthetase' (fusion between MetRS and cupin protein) and identify their substrate specificity and reaction products using *in vitro* biochemical reaction approach and chemical bioconversion on *Escherichia coli.*

1.7 Outline of Dissertation Book

This dissertation book is divided into four chapters. Chapter 1 provides a comprehensive introduction as the background of this research. The characteristics, activities, and the known pathway of formation N-N bond containing compound are discussed in this chapter. The general explanation about genome mining and its purpose in this research is also described in this section.

The discovery of new producer of azoxy compound azodyrecin, isolation and the structure activity relationship of the biological activity are discussed in the second chapter.

In the third chapter, the explanation will focus on exploration of enzymes that contributed to the formation of N-N bond, which are NMO and 'hydrazine synthetase'. Starting from genome mining study to explore diversity of NMO and 'hydrazine synthetase' homolog, and study to uncover the substrate specificity of each enzyme and identify structural diversity of the hydrazine products produced by each enzyme.

The last chapter will describe the general conclusion from the research and planning for future research.

Chapter 2 New Azodyrecins Biosynthetic Gene Cluster Identified by a Genome Mining-Directed Reactivity-Based Screening.

2.1 Introduction

Actinomycetes are procaryotic organisms that belong to gram positive bacteria and are well-known as prolific sources of antibiotics.[38] These group have filamentous morphological feature, diameter 1-2 μm, aerobic, and live in soil or marine habitat[38]. The research about actinomycetes started in 1920, continued by discovery of streptomycin from *Streptomyces griseus* as the first anti-tuberculosis bacteria in 1943.[11] Nowadays, about 61% of natural antibiotics are derived from Actinomycetes.[39] One of the most abundance species from Actinomycetes is *Streptomyces* sp., which represents about 95% of total actinobacteria found from terrestrial source and distinguished of its non-motile and spore features.[39]

During the growth phase of *Streptomyces,* these bacteria produce arrays of secondary metabolites. Among those secondary metabolites, some bacteria produce nitrogen–nitrogen (N–N) bond containing compounds. N–N bonds often act as center of the bioactivities or provide essential interactions with molecular targets. N–N bonds are embedded into NPs in various forms of functional groups including azoxy, nitrosamine, nitroso hydroxylamine, azo, diazo, hydrazides, hydrazones, hydrazine, azines, pyrazoles and indazoles, pyridazines and cinnolines, triazines, triazoles, and other N–N bonds-containing heterocycles.[17]

Azoxy-containing natural products (Figure 2.1), which are characterized by $RN{=}N^{+}(O^{-})R$ basic chemical structure and their diverse pharmacological activities, attract the attention of natural product researchers. The first azoxy compound macrozamin was discovered from cycad in 1951[40], but most recently more compounds were discovered from bacteria. Elaiomycin, isolated in 1954 from *Streptomyces hepaticus* showed potent *in vitro* antibacterial activity against *Mycobacterium tuberculosis*.[19] Meanwhile, the aliphatic novel azoxy compound, maniwamycin, was isolated from culture *Streptomyces prasinopilosus* and has been proven as antifungal against *Candida albicans.* Other maniwamycin-like aliphatic azoxy NPs have also been discovered, and they were named as LL-BH872α and KA57A. Another microbial azoxy NP was detected from the culture of *Streptomyces viridifaciens* and was named valanimycin.[14]

Until recently, about 50 azoxy compounds were identified from various sources, including bacteria, fungi, plants, and marine sponges.[16] The conventional isolation methods based on biological activity or physicochemical characteristics are not selective to azoxy functional group. Consequently, there are only a few examples of azoxy natural products, despite their notable biological activities. Therefore, the use of reactivity-based screening method selective to an azoxy group would enable targeted isolation of azoxy NPs and would eventually lead to the identification of new azoxy-containing NPs. The azoxy group is prone to hydrolysis under acidic conditions and generate inorganic hydrazine N_2H_4.[40] Indeed, the generation of N_2H_4 upon acid hydrolysis was exploited as a proof of N–N bonds in the cycad-derived methyl azoxy compound macrozamin in 1951, which was discovered as the first N–N bond containing natural product. Therefore, N_2H_4 can be exploited as an indicator for azoxy compounds. Hydrazine reacts with two equivalents of *p*-dimethylaminobenzaldehyde (DAB) to yield *p*-dimethylaminobenzaldazine, which can be sensitively detected with HPLC by monitoring UV absorption at 485 nm (Scheme1).[41] Although this assay is quite easy and specific, it has not been applied for the screening of azoxy bonds in the context of natural products discovery.

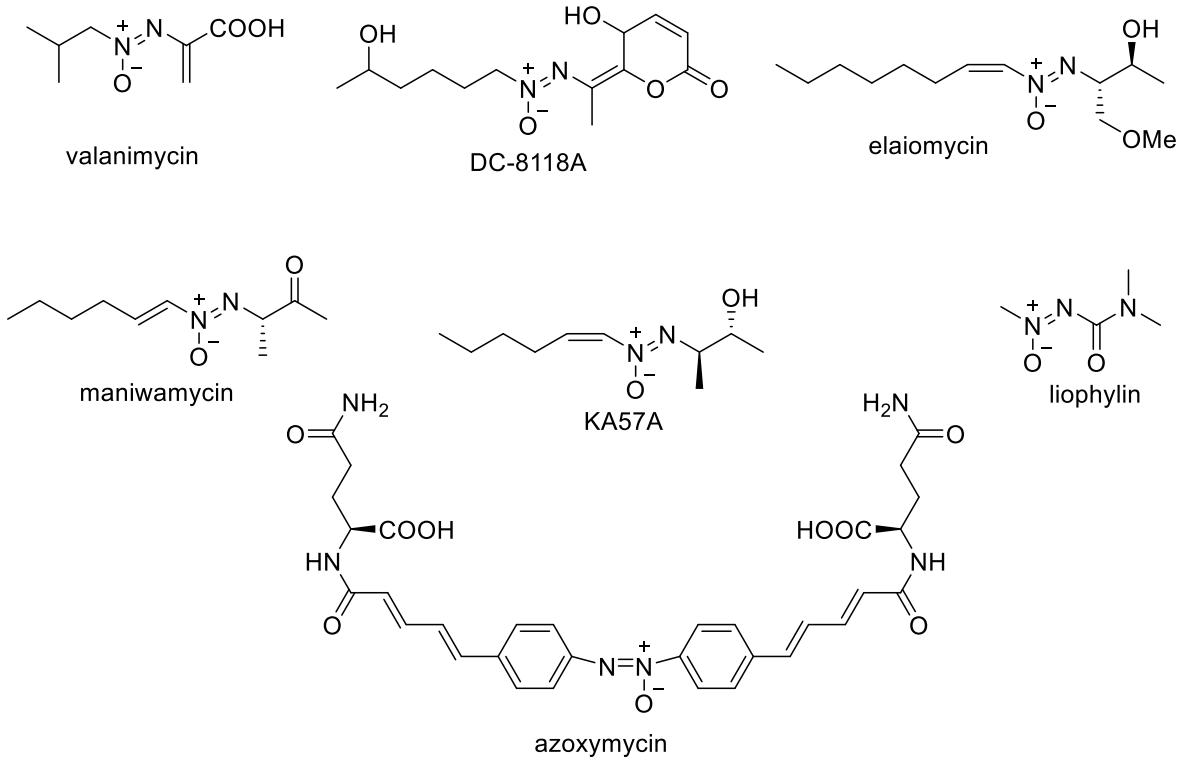


Figure 2.1 Representative azoxy-containing natural products

A distinct mechanism is employed in the biosynthesis of valanimycin, an antibacterial aliphatic azoxy natural product isolated from *Streptomyces viridifaciens* MG456-hF10.[42–44] Valanimycin has been exploited as a model compound for studying the enzymatic basis of azoxy formation in nature, and its biosynthesis has been under intensive investigations for more than two decades by Parry and co-worker. Very recently, the enzymes responsible for azoxy formation were identified, and the entire biosynthetic pathway of valanimycin was fully revealed (Figure 2.2). The biosynthesis initiates with decarboxylation of valine to give isobutylamine, which undergoes *N*-hydroxylation mediated by the flavin-dependent monooxygenase VlmH. *N*-hydroxylisobutylamine is coupled with serine to give *O*-(L-seryl)-isobutylhydroxylamine by function of the VlmA. Interestingly, VlmA specifically utilize serine molecule that is loaded on seryl-tRNA catalyzed by VlmL, thus representing a unique tRNA-dependent biosynthetic enzyme which phylogenetically different from seryl-tRNA synthase in the primary metabolism.[45] The fate of *O*-hydroxylamine intermediate has long been elusive, however, very recently, membrane-bound enzyme VlmO was revealed to catalyze the transformation of *O*-(L-seryl)-isobutylhydroxylamine to a hydrazine intermediate isobutylamino-L-serine.[43,46] Successively, non-heme diiron dioxygenase VlmB catalyze four-electron oxidation of the hydrazine to give an azoxy product.[43,46] Finally, VlmK/VlmJ catalyze dehydration of the serine portion to give *exo* olefin moiety to complete valanimycin biosynthesis. These studies collectively showed that *N*-hydroxylase VlmH, acyltransferase VlmA, membrane-bound hydrazine synthase VlmO, and non-heme diiron dioxygenase VlmB cooperate to install azoxy functionality. Therefore, the gene cassettes coding for these four enzymes could potentially be exploited as the signature of biosynthetic gene cluster (BGCs) of aliphatic azoxy natural products[42,44,45,47,48].

In this study, we combined genome mining approach and reactivity-based screening approach to discover novel azoxy-containing natural products. To explore potential producer of azoxy compounds, we exploited the co-occurring *vlmA, vlmH, vlmB*, and *vlmO* genes as a genetic beacon. The potential producers were cultured and subjected to reactivity-based screening, utilizing an N_2H_4 detecting-colorimetric assay, to investigate the actual production of azoxy-containing natural products under laboratory culture conditions.

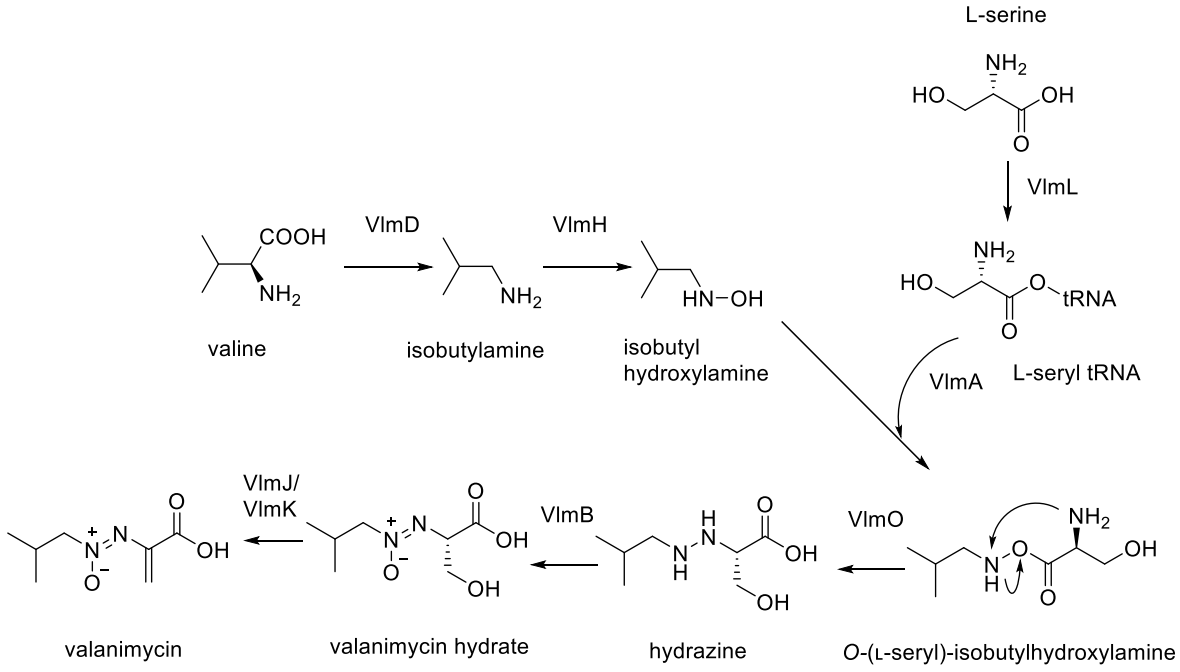


Figure 2.2 Valanimycin biosynthesis pathway

2.2 Method

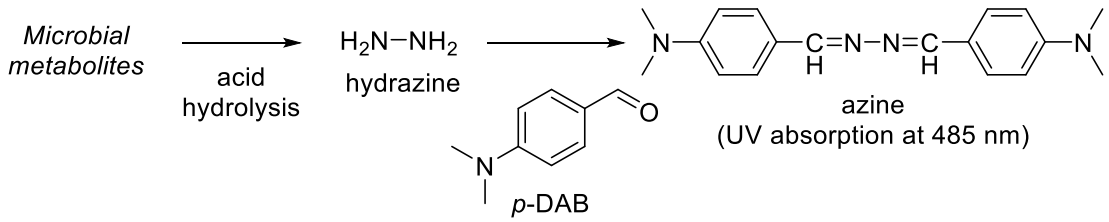
a. Genome Mining of *vlmA*

NCBI non-redundant protein database was searched by phmmer[49] using *vlmA* as a query. After the removal of redundancy, 20 kb genetic neighboring region of each "*vlmAs*" was retrieved. Genetic regions were further searched for homologous genes of "*vlmH*", "*vlmO*", and "*vlmB*", then finally, 813 genetic regions were selected as candidates of BGCs for azoxy-containing natural products. The similarity of the potential BGCs were investigated by a networking program BiGSCAPE[50] and the resultant network was visualized by Cytoscape[51].

b. N_2H_4-Detecting Colorimetric Assay-based Screening of Actinobacteria

Out of 813 potential producers of azoxy compounds, 96 strains were purchased and subjected for N_2H_4-detecting colorimetric assay. Strains were inoculated in TSB agar media for 2 days, then pre-cultured in TSB liquid media for 2 days at 30 °C. 100 μL of the pre-cultured were spread to SFM agar, and incubated for 7 days at 30 °C. After 7 days, plates were extracted using methanol overnight and debris were removed by filtrations, then solvent was evaporated. Resultant residue was dissolved into methanol and subjected to the colorimetric assay.

The colorimetric assay was performed by mixing 10 mM DAB reagent and methanol extract of samples (ratio 1:1) and incubated for 10 min at room temperature. Mixtures were diluted with equal volume of methanol, centrifuged at 15000 rpm for 10 minutes and injected to Shimadzu HPLC system equipped with SPD-M20A. 5 μL of the samples were loaded onto COSMOSIL 5C_{18}-MS-II 2.0 × 100 mm (nacalai tesque), using mobile phase H_2O/MeCN containing 0.1% formic acid with a linear gradient: 2–98% for MeCN + 0.1% formic acid over 5 min with a flow rate at 0.4 mL/min. Column eluates were monitored by UV absorption at wavelength 485 nm.



Scheme 1. N_2H_4-detecting colorimetric assay.

c. Isolation of Azoxy-containing Natural Products from *Streptomyces* sp. A1C6

Glycerol stock of *Streptomyces* sp. A1C6 was inoculated to TSB agar and incubated for 2 days at 30 °C. A single colony was transferred to TSB liquid media and incubated for 2 days at 30 °C. 100 µL of pre-cultured was spread to SFM agar plates and incubated at 30 °C for 7 days. After 7 days, 180 SFM agar plates were extracted using methanol for overnight. Debris were filtered off and extract was evaporated using rotary evaporator. Viscous extract was diluted using Milli-Q water and was fractionated using Diaion® HP20 column chromatography. Sample was eluted using methanol in gradient concentration (20-100%), and the fraction were collected and evaporated by rotary evaporator. The fractions were monitored by DAB assay, and the positive fractions were combined and subsequently used for the next step.

The positive fractions were diluted and applied to Silica gel column chromatography and eluted using gradient concentration of Hexane-Chloroform. Fractions were collected and evaporated, then subsequently subjected to DAB assay. Gel filtration chromatography Sephadex LH20 was chosen for the third fractionation step, using methanol as mobile phase. The elution was conducted overnight, and sub-fractions were collected using automatic fraction collector. Sub-fractions were evaluated for DAB assay activity, and the positive fractions were combined and evaporated.

The isolation of target compounds was performed by applying positive sub-fractions onto preparative HPLC system COSMOSIL 5C18-MS-II, mobile phase: H_2O/MeOH (70-98%) in 30 minutes, flow rate 3 ml/s.

d. Cytotoxic Activity Assay

Isolated compounds were tested for the cytotoxic activities against human ovarian adenocarcinoma SKOV-3 cells, malignant pleural mesothelioma MESO-1 cells, and immortalized T lymphocyte Jurkat cells. SKOV-3 cells were cultured in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, penicillin (50 U/mL), and streptomycin (50 µg/mL). MESO-1 cells were cultured in RPMI1640 medium supplemented with 10% fetal bovine serum, penicillin (50 U/mL), and streptomycin (50 µg/mL). Jurkat cells were cultured in RPMI1640 medium supplemented with 10% fetal bovine serum, penicillin (50 U/mL), streptomycin (50 µg/mL), and GlutaMAX. All cell lines were seeded in a 384-well plate at a density of 1000 cells/well in 20 µL of media and

incubated at 37 °C in a humidified incubator with 5% CO_2. After 4 h, 2-fold serial dilution samples dissolved in DMSO were added to the cell cultures at the concentration of 0.5% (0.1 μL) and incubated for 72 h. Cell viabilities were measured using a CellTiter-Glo luminescent cell viability assay and envision multilabel plate reader.

P388 murine leukaemia cells were cultured in DMEM, supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum in 5% of CO_2 cell incubator at 37 ℃. The cells were placed a 96-well cell culture plate at a density of 1×10^4 cells/well, then 1 μL of test solution in various concentrations (samples were dissolved in DMSO) added to cell plates and incubated for 48 h. Doxorubicin hydrochloride was used as a positive control. Finally, 50 μL of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) solution (1 mg/mL dissolved in PBS buffer) were added to each well and the plates were incubated for 4 h. After the medium was removed, the precipitated dye was solubilized by DMSO, and measured by a microplate reader with the absorbance at 570 nm**.**

2.3 Results

- Exploring Potential BGCs of Aliphatic Azoxy Natural Products

The biosynthetic pathway of azoxy compound had been studied for decades, but the mechanism of formation of azoxy moiety was not fully understood until very recently. Valanimycin, an antibiotic isolated from *Streptomyces viridifaciens* MG456-hF10, is a model compound for the study of azoxy biosynthetic pathway. In the biosynthetic pathway of valanimycin, the *N*-hydroxylation catalyzed by VlmH, acyl transfer by VlmA, and series of intramolecular rearrangement and oxidation by VlmB and VlmO are the important steps for the formation of azoxy bond. Therefore, the presence of *vlmH*, *vlmA*, *vlmB*, and *vlmO* can be the signature of azoxy BGC. In this study, we aimed to expand the chemical diversity of azoxy-containing natural products through genome-mining approach. To this end, gene cassettes coding for "*vlmA*", "*vlmH*", "*vlmB*", and "*vlmO*" in valanimycin biosynthesis were used as queries. As a result of searching database with HMM-based method, 813 genetic regions that harbor *"vlmA/B/O/H"* genes were identified as potential BGCs of azoxy-natural products. To investigate the diversity of potential BGCs, these were subjected to BiGSCAPE analysis.[50] As a result, 813 potential BGCs were categorized into several gene cluster families (GCFs). Three referential BGCs of azoxy natural products (i.e. KA57-A, azodyrecin, and valanimycin) were in three distinct GCFs, indicating that the network threshold is stringent enough to distinguish structural difference between these natural products (Figure 2.3). Considering that most GCFs in the network has no links to their biosynthetic products, a substantial fraction of the chemical diversity in aliphatic natural azoxides likely remains untapped.

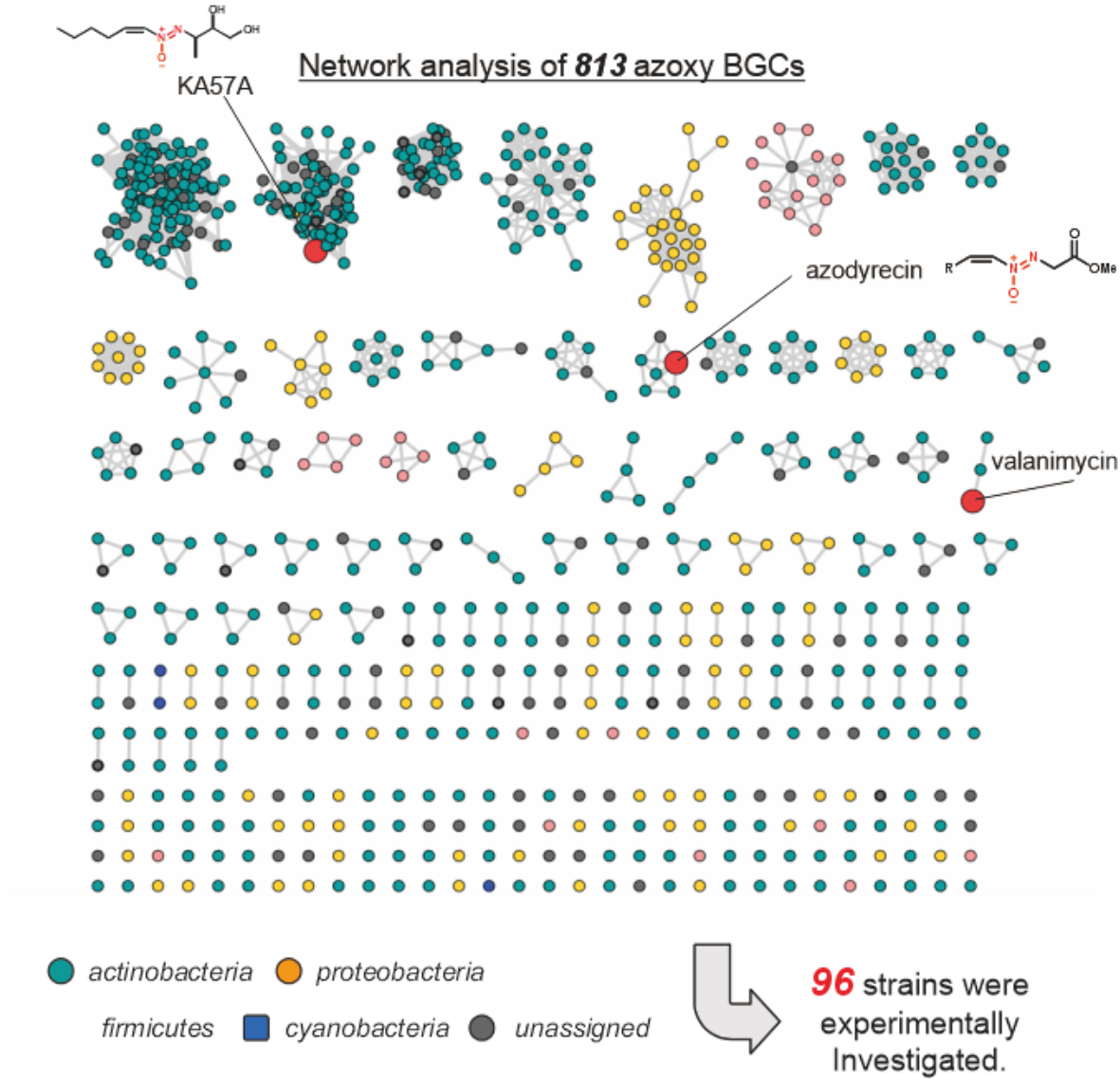


Figure 2.3 Genome mining of the producer candidates of azoxy compounds

Reactivity-Based Screening of Azoxy Natural Products.

Among 813 BGCs investigated in the previous section, 96 BGCs were encoded in the genome of commercially available actinobacteria. To test whether these strains produce azoxy natural products under the laboratory culture conditions, these strains were purchased, and their culture broths were subjected to N_2H_4-detecting colorimetric assay. Strains were cultured on Soya Flour Mannitol Agar (SFM) for 7 days at 30 °C, then the solid agars were extracted by methanol. The resultant methanol extracts were subjected to the colorimetric assay. As a result, the generation of N_2H_4 were detected upon acid hydrolysis of 19 strains, suggesting that these strains produce azoxy natural products when cultivated on SFM solid media. A

significant amount of N_2H_4 was detected in the acid hydrolysate of metabolites from strains with BGCs similar to that of KA57-A. As these BGCs are likely to produce azoxy natural products similar to KA57-A, these strains were excluded from further investigation. Meanwhile, N_2H_4 was also detected from the acid hydrolysate of a strain with a unique BGC (singletons in the network), such as *Streptomyces* sp. A1C6 (Figure 2.4). With the expectation of novelty in chemical structure, I focused on *Streptomyces* sp. A1C6 for further investigation.

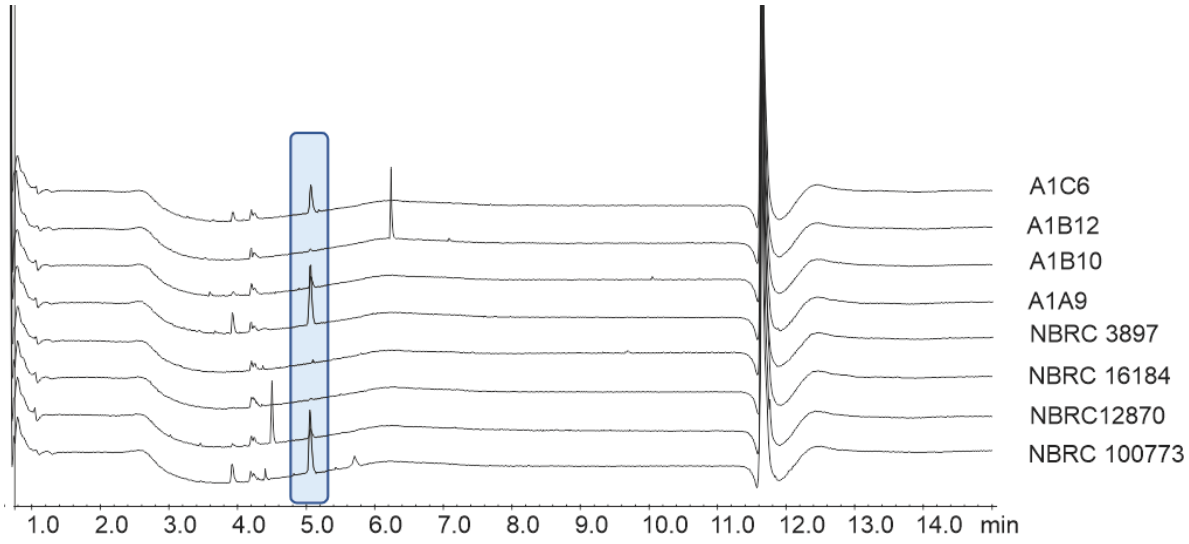


Figure 2.4 DAB assay screening result of several *in-house* collection Actinomycetes strains

UPLC system: COSMOSIL 5C18-MS-II 2.0 × 100 mm (nacalai tesque), using mobile phase H_2O/MeCN containing 0.1% formic acid with a linear gradient: 2–98% for MeCN + 0.1% formic acid over 5 min with a flow rate at 0.4 mL/min. Column eluates were monitored by UV absorption at wavelength 485 nm. The peak corresponding to the azine product was highlighted by blue box.

Reactivity-guided Purification of Azoxy Natural Products.

Reactivity-guided purification of azoxy natural products. The purification process of the target azoxy natural products were performed in the colorimetric assay-guided manner. *Streptomyces* sp. A1C6 was cultivated on SFM solid media, then extracted with methanol. After the removal of solvent, residue was fractionated using synthetic adsorbent Diaion® HP20. The fractions that generate N_2H_4 upon acid hydrolysis were combined and further subjected to silica gel column chromatography. Resultant fractions were subjected to colorimetric assay, and positive fractions were further fractionated by gel filtration chromatography using Sephadex LH20 with methanol as mobile phase. LC-MS analysis of assay-positive fractions showed that the positive fractions contain several hydrophobic

compounds **1**-**4**, **6**, and **8**. Compounds **1** and **4** exhibit identical *m/z* value at 327.3 in positive mode, suggesting that these two are a pair of isomers. Similarly, **3** and **6** that have molecular ion peaks at *m/z* 355.3 are likely to be another pair of isomers. Additionally, compound **1**, **2**, and **3** showed molecular ion peaks differ by 14 atomic mass units, which is likely to corresponds to a different number of methylene carbons (Figure 2.6i). Considering that these compounds are highly hydrophobic, these are likely to be a group of related compounds with different lengths of the fatty acid chain.

When searched for the databases, a group of compounds azodyrecin fit well to above mentioned characteristics. Azodyrecins are a group of aliphatic azoxy natural products derived from *Streptomyces* sp. P8-A2[52] and recently isolated from *Streptomyces* sp. RM72 by our group. Azodyrecins possess a long unsaturated aliphatic chain, which is conjugated with L-alanine methylester *via* a characteristic azoxy group. Series of congeners with different length of aliphatic chains (azodyrecin A-C) and a pair of geometric isomers (e.g. azodyrecin A and 1′-*trans* azodyrecin A) have been isolated. Additionally, our group previously isolated saturated derivatives azodyrecin D-G from *Streptomyces* sp. RM72. LC-MS comparison of the assay-positive fraction with authentic standards of azodyrecins showed that compounds **1**-**4**, **6**, and **8**, are identical to azodyrecin A, azodyrecin B, azodyrecin C, 1′-*trans* azodyrecin A, 1′-*trans* azodyrecin C, and azodyrecin E, respectively (Figure 2.6iii). These results showed that *Streptomyces* sp. A1C6 is a new producer of azodyrecins.

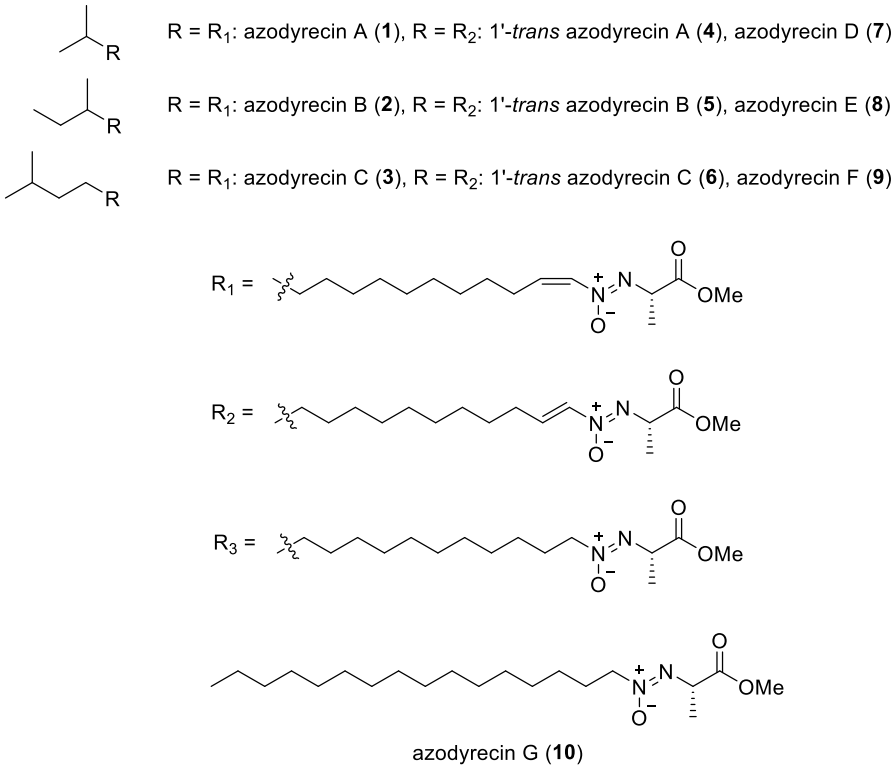


Figure. 2.5. Structures of azodyrecins.

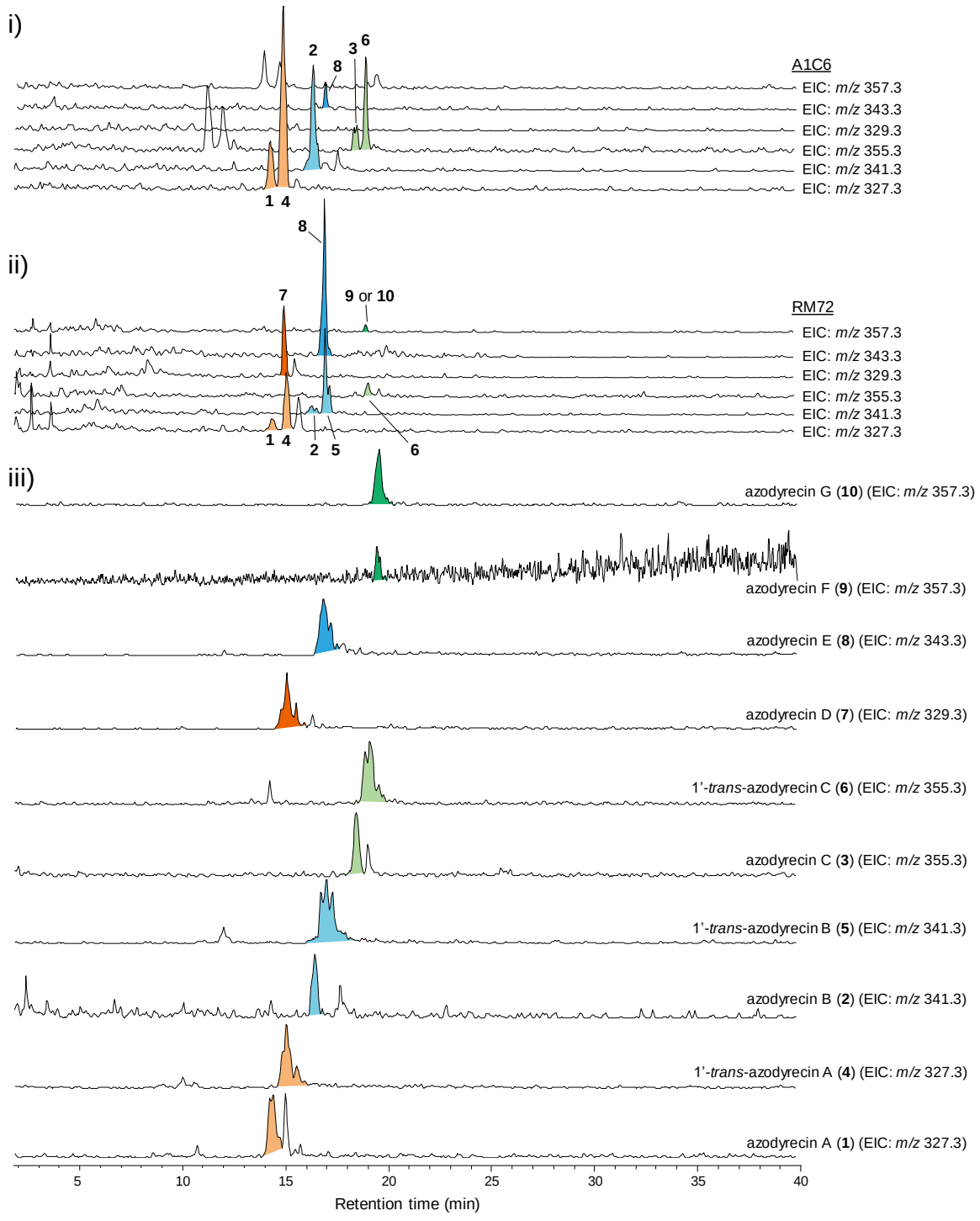


Figure. 2.6. Production of azodyrecins in *Streptomyces* strains. (i) LC-MS analysis of crude metabolite from *Streptomyces* sp. A1C6. (ii) LC-MS analysis of crude metabolite from *Streptomyces* sp. RM72 in previous study. (iii) Authentic standards of azodyrecins.

Cytotoxic Activity Profile of Isolated Compound and Structure Activity Relationship. Although cytotoxicity of azodyrecin have been evaluated, the structure activity relationship of azodyrecin has not been elucidated. Therefore, in this study, in collaboration with National Institute of Advance Industrial Science and Technology (AIST), the cytotoxic activity of azodyrecin congeners was elucidated. To this end, four type of cell lines were tested: human ovarian adenocarcinoma SKOV-3 cells, malignant pleural mesothelioma MESO-1 cells, immortalized T lymphocyte Jurkat cells, and P388 cell. SK-OV-3 is a human ovarian cancer cell line with epithelial-like morphology. These cells are resistant to tumour necrosis factor and to other cytotoxic drugs such as diphtheria toxin, cisplatin. This cell line was established in 1973 from the ascites of a 64-year-old Caucasian female with adenocarcinoma of the ovary.[53] Other cell, MESO-1 cell, is human malignant pleural mesothelia cell line. It is epitheal like cell isolated from Caucasian men.[54] Another cell used for cytotoxic assay was immortalized T lymphocyte Jurkat cells. Jurkat cell is originally from the peripheral blood of a 14-year-old boy with T cell leukemia and now commonly used to study acute T-cell leukemia, T-cell signalling, and the expression of various chemokine receptors susceptible to viral entry, particularly HIV.[55] The last cell that we used is P388 cell line, a mouse cell line derived from lymphoid tumor.

As results, unsaturated azodyrecins (azodyrecin B and 1′-trans-azodyrecin B) showed moderate cytotoxic activity against all types of tumour cells with IC_{50} 7.37-9.7 μM for azodyrecin B and 3.36-8.24 μM for 1′-trans-azodyrecin B, respectively (Table 2.1). On the other hand, saturated azodyrecin D and E had almost no or negligible cytotoxic activity (IC > 50 μM). These results highlights the importance of the double bond adjacent to the azoxy moiety for exerting the cytotoxic activity of azodyrecins.

Table 2.1: IC_{50} of cytotoxic assay

Compound	IC_{50} (μM)			
	SKOV3	**MESO1**	**Jurkat**	**P388**
azodyrecin B (2)	**7.37**	**9.70**	**8.72**	**11.6**
1′-*trans*-azodyrecin B (5)	**8.24**	**6.70**	**3.36**	**4.72**
azodyrecin D (7)	> 50	43.2	> 50	> 50
azodyrecin E (8)	> 50	> 50	> 50	> 50

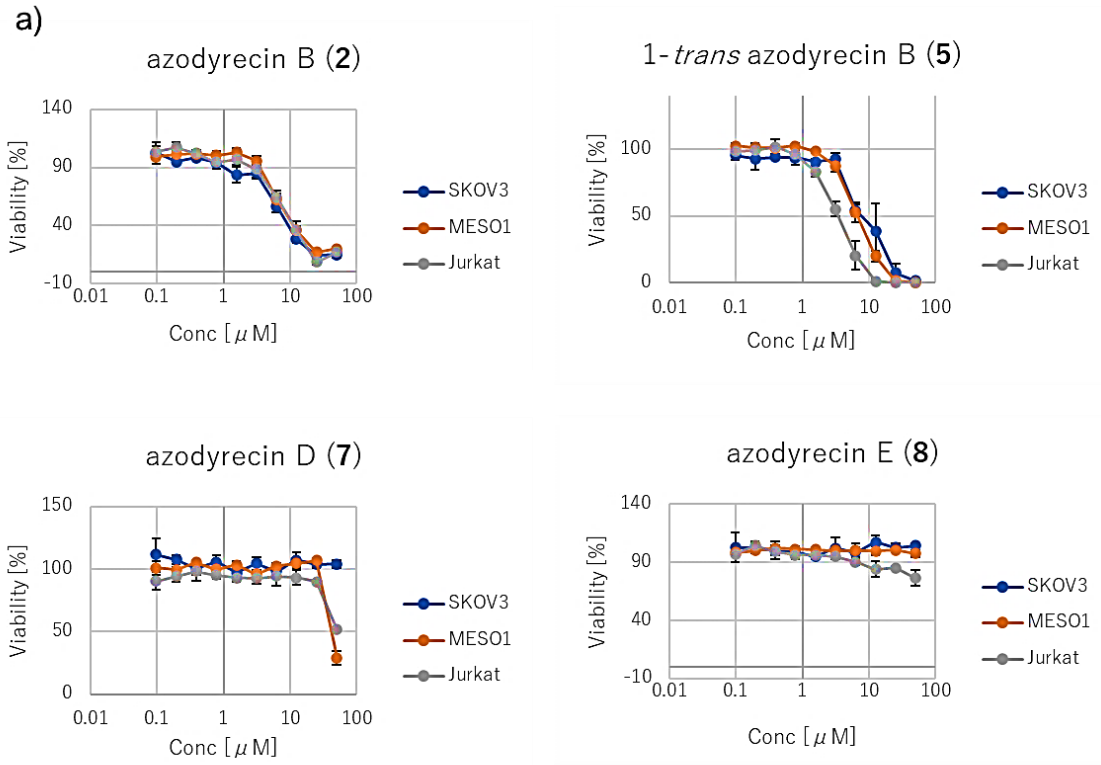


Figure 2.7 Cytotoxic activity of azodyrecin compounds

2.4 Discussion

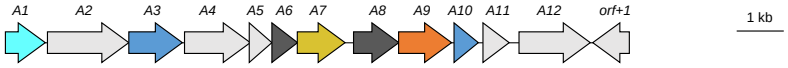
Rapid accumulation of genomic sequence information and fast development of genome sequencing technologies have revealed the abundant, as-yet unexploited biosynthetic potential of natural products with unique functional groups, such as those containing nitrogen-nitrogen (N–N) covalent bonds. Since the first discovery of macrozamin from *Macrozamia spiralis*, an endemic cycad from Australia in 1951[40], over 300 N–N bonds-containing natural products have been discovered. Recent advances in biosynthetic studies have revealed various enzymatic machineries for the biogenesis of N–N bond-containing functional groups, thereby opening more opportunities for the discovery of novel N-N bond-containing natural products through genome mining approach.

To explore the diversity of azoxy compound BGCs, key genes for azoxy compound biosynthesis, *vlmA*, *vlmH*, *vlmO*, and *vlmB*, were used for the queries on genome mining study. From the genome mining result, 813 BGCs were classified into more than 10 clades, in which most of them were distinct from the references KA-57A and valanimycin BGCs. This indicates that there were numerous potential producers of azoxy compounds that might lead to the discovery of novel azoxy compounds. Then, colorimetric assay using *p-diaminobenzaldehyde* (DAB) were used to confirm the presence of hydrazine generated from the acid hydrolysis of azoxy compound, which led to the discovery of a novel azoxy compound producer, *Streptomyces* sp. A1C6.

Reactivity assay-based isolation in this work led to the identification of several azodyrecin congeners in *Streptomyces* sp. A1C6, revealing this strain as a new azodyrecin producer. Further analysis revealed that azodyrecin BGC from *Streptomyces* sp. A1C6 was distinct from known BGCs in *Streptomyces* sp. RM72 and *Streptomyces mirabilis* P8-A2. The differences are in the gene organization, especially the order of the genes in the BGCs because there are several inversion and insertion between these two BGCs. However, both BGCs still have similarity, especially the presence of several core genes such as methyl transferase, desaturase, v*lmA*-like, and *vlmH*-like genes.

Previously, azodyrecin congeners were identified in *Streptomyces* sp. strain RM72, another in-house collection strain in our laboratory, followed by characterization of their encoded BGC. Azodyrecin BGC from *Streptomyces* sp. A1C6 strain was easily found by BLAST search using VlmA as a query (Figure 2.8). By comparing two azodyrecin BGCs, we found several similarities including the presence of gene coding for desaturase (responsible

for double bond installation), methyltransferase gene for methyl esterification, *vlmH*-like gene for *N*-hydroxylation, and *vlmA*-like gene for acyl transferase (Figure 2.9). However, the genes organization between these two BGCs are different that may had occurred through several insertion and inversion events. In the RM72 strain, the core genes were located on the upstream part, while the additional genes were in the downstream part. Meanwhile, in the A1C6 strain, the core genes were positioned between the additional genes. Furthermore, the order of genes between the two BGcs are also different. Taken together, the BGC of azodyrecin from *Streptomyces* sp. A1C6 is considered as a new type of azodyrecin BGC.



ORF	aa length	homologous protein [organism] (sequence id)	proposed function	accession no.
orf-1	117	PRC domain containing protein [*Streptomyces angustmyceticus*] WP_086720082.1	Unknown	
adyA1	280	SAM-dependent methyltransferase [*Streptomyces angustmyceticus*] WP_223659814.1	Methyltransferase	
adyA2	575	putative aldehyde dehydrogenase [*Streptomyces angustmyceticus*] GES28483.1	Dehydrogenase	
adyA3	377	acyl-CoA dehydrogenase family protein [*Streptomyces angustmyceticus*] WP_152104274.1	*N*-hydroxylase	
adyA4	462	aminotransferase class III-fold pyridoxal phosphate-dependent enzyme [*Streptomyces angustmyceticus*] WP_152104273.1	Aminotransferase	
adyA5	157	SRPBCC family protein [*Streptomyces angustmyceticus*] WP_152104272.1	Unknown	
adyA6	179	hypothetical protein [*Streptomyces angustmyceticus*] WP_223659821.1	Unknown	
adyA7	338	DUF2156 domain-containing protein [*Streptomyces angustmyceticus*] WP_223659823.1	VlmA-like	
adyA8	320	VlmB-like protein [*Streptomyces angustmyceticus*] WP_086717659.1	Unknown	
adyA9	373	acyl-CoA desaturase [*Streptomyces angustmyceticus*] WP_086717661.1	Desaturase	
adyA10	172	flavin reductase family protein [*Streptomyces angustmyceticus*] WP_086717663.1	Flavin reductase	
adyA11	185	hypothetical protein [*Streptomyces angustmyceticus*] WP_086717665.1	Unknown	
adyA12	503	AarF/UbiB family protein [*Streptomyces angustmyceticus*] WP_086717668.1	Protein kinase	
orf+1	256	alpha/beta fold hydrolase [*Streptomyces angustmyceticus*] WP_086717669.1	Hydrolase	

Figure 2.8 Functional annotation of azodyrecin gene cluster

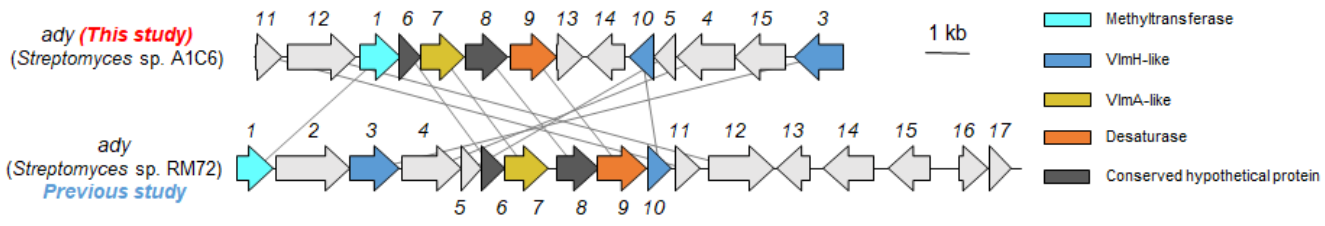


Figure 2.9 Gene cluster of azodyrecin in *Streptomyces* sp. A1C6 and RM72

Functional annotation of azodyrecin BGC suggest the following biosynthetic pathways: the biosynthesis is likely initiated by Ady2, a putative dehydrogenase that recruits fatty acids from primary metabolism to generate an aldehyde, which would be converted to an aliphatic amine by the pyridoxal phosphate (PLP)-dependent transaminase Ady4. The amine would be *N*-hydroxylated by the two-component flavin-dependent monooxygenase Ady3/Ady10, as in the valanimycin biosynthesis mediated by VlmH/VlmR. The hydroxylamine would be conjugated to alanyl-tRNA to form an ester intermediate by the function of the tRNA-utilizing enzyme Ady7, which is homologous to VlmA. In valanimycin biosynthesis, the substrate seryl-tRNA is provided by VlmL, an additional seryl-tRNA synthetase (SerRS) encoded within the *vlm* cluster. However, no aminoacyl-tRNA synthetase gene is present in the ady cluster, suggesting that the alanyl-tRNA is directly provided from the cellular tRNA pool in the case of azodyrecin biosynthesis. The mechanism for the subsequent rearrangement of the ester intermediate for azoxy bond formation remains unclear; however, the conservation of the two hypothetical proteins Ady6/Ady8 among the biosynthetic gene clusters of valanimycin, KA57-A, and azodyrecins may suggest their participation in this step. Ady6 showed low homology to DUF4260 (PF14079.9), a family of integral membrane proteins with unknown functions, while Ady8 was similar to the ferritin-like superfamily protein (IPR009078). Ady6/Ady8 are homologous to VlmO/VlmB and SRO_1835/SRO_1837 in the biosynthetic gene clusters of valanimycin and KA57-A, respectively. Very recently, Liu et al. and Du et al. revealed that VlmO catalyzes the intramolecular rearrangement of *O*-acylhydroxylamine intermediate to produce hydrazine.[43] The resultant hydrazine is oxidized by VlmB via hydrazine-azo-azoxy pathway to furnish the azoxy moiety. These findings further support that Ady6/Ady8 are involved in azoxy formation in azodyrecin biosynthesis.

The azoxy bond formation would be followed by the Ady1-mediated methyl esterification to form saturated azodyrecins. Subsequent installation of a *cis*-olefin on the 1,2-positions of the alkyl side chain would afford azodyrecins A–C (**1**–**3**), in a reaction likely to be mediated by Ady9, a putative fatty acid desaturase. Nevertheless, the possibility that the desaturation occurs prior to the methyl esterification could not be excluded. Importantly, following study in our group elucidated that recombinant protein of Ady1, a putative methyltransferase catalyzes methylesterification of dimethyl-azodyrecin D (**7**) *in vitro*, using SAM as a methyl donor.

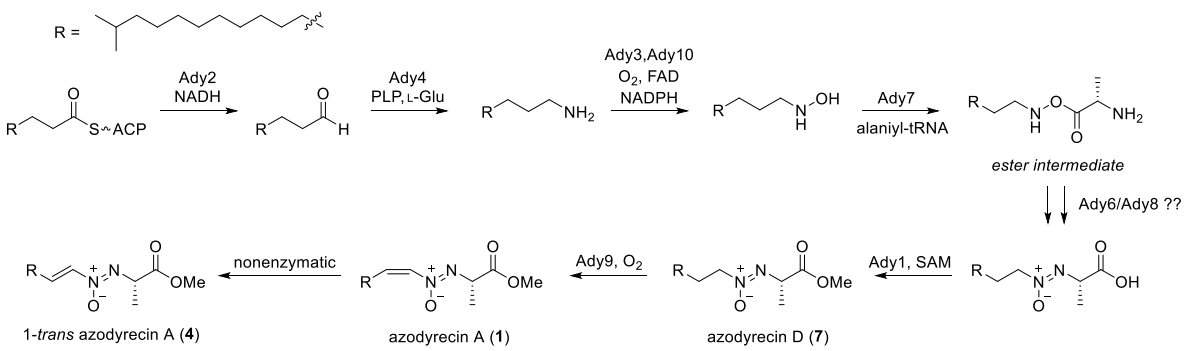


Figure. 2.10 Proposed biosynthetic pathway of azodyrecin**.**

Another study by Wibowo et al. revealed that azodyrecin was also discovered in *Streptomyces mirabilis* P8-A2.[52] Comparing azodyrecin BGC from P8-A2 to azodyrecin BGCs in our study, it was found that both BGCs shared average nucleotide identity (ANI) of 95.7% and gene protein identities of 81 to 98 %.[56] Although BGC from *Streptomyce*s sp. RM72 showed low similarity to *S. mirabilis* P8-A2, it is still able to encode the production of azodyrecins A-F. This suggests even though those strains contain different gene organization for azodyrecins, those strains could produce the same compounds as long as their BGCs contain similar core genes.

Biosynthetic pathway of azodyrecin was discovered very recently by Matiss Maleckis and co-workers.[56] In their study, they disrupted 14 genes in the azodyrecin BGC of *Streptomyces* mirabilis P8-A2 through CRISPR-Cas9 application. The results suggested that the biosynthesis pathway started from conversion of fatty acid to generate iso-fatty alcohol by AzdP, then followed by conversion to fatty amine by AzdQ. Then, AzdH install hydroxylamine to fatty amine. The role of AzdR is related to prevent formation of off product and direct the reaction towards *N*-hydroxylation. The function of AzdA is catalyzing alanine transfer from $tRNA^{Ala}$ to a hydroxylated fatty amine. The role of AzdO is still unclear, but it is likely related to formation of hydrazine. Then, AzdB responsible for the formation of azoxy bond in azodyrecin.[56]

In this study, azodyrecin congeners showed cytotoxic activity against SKOV3, MESO1, Jurkat cell, and P388. This activity corresponds to the presence of double bond adjacent to the azoxy moiety. Meanwhile, the geometric of the double bond did not have correlation to the activity. Further molecular study is needed to understand about the mechanism of cytotoxic activity of azodyrecin.

This study revealed the diversity of unexplored azoxy compound BGCs, which led to the discovery of a novel producer of azodyrecins. Meanwhile, most of BGCs of azoxy compounds are still poorly investigated. Therefore, biosynthetic studies of these untapped BGCs could reveal novel azoxy compounds, thereby expanding our understanding on the structural diversity of azoxy compounds in nature.

Chapter 3 Diversity of Cupin-mediated Hydrazine Formation in Bacteria

3.1 Introduction

Natural products containing nitrogen-nitrogen (N–N) bond belong to a rare group of natural products with only 0.1% having this kind of functional groups.[17] However, the structural diversity of these compounds are extremely large. N–N bond can be found in a wide range of natural product classes including polyketide, peptide, nucleoside, amino acid, and fatty acid derivatives.[36,37,56,57] The unique chemical characteristics of N-N bond-containing NPs contributed to the potent pharmacological activities, making these groups become one of important source for drug lead compounds.[17]

Recent advances in genomics supported by rapid development of DNA sequencing technologies has contributed greatly to the discovery of the biosynthetic machineries of N–N bond-containing compounds, such as the discovery of nitrous acid biosynthetic pathway in secondary metabolism and metalloenzyme responsible for the formation of N–N bond.[58] Some metalloenzymes were found to catalyze the formation of N–N bond. Notable examples are CreM that contributed to formation of N–N bond in cremeomycin, SznF for *N*-nitroso formation in streptozotocin, and KtzT that converted N^5-OH L-ornithine to piperazic acid. [36]

Hydrazine synthetases constitute a group of proteins characterized by a *N*-terminal zinc-binding cupin domain and a *C*-terminal domain resembling a methionyl-tRNA synthetase (MetRS-like domain). Operating in the early stages of biosynthetic pathways, hydrazine synthetases facilitate the coupling of N^{ω}-hydroxylamine substrates and amino acid substrates through N–N bonds, resulting in the production of amino acid-based hydrazines.[37] Figure 3.1 illustrates our current comprehension of the hydrazine synthetase reaction mechanism.

The reaction catalyzed by hydrazine synthetase can be divided in two distinct steps. In the initial half-reaction, the MetRS-like domain activates an amino acid substrate by utilizing ATP, analogous to canonical amino acyl-tRNA synthetases.[59] This activated amino acid is subsequently joined with an *N*-hydroxylamine substrate to create a reactive *O*-acylhydroxylamine intermediate (MetRS domain in Figure 3.1). Remarkably, hydrazine synthetase is a tRNA-independent enzyme, coupling the activated amino acid with *N*-hydroxylamine rather than tRNA.[59] In the second half-reaction, zinc-binding cupin domains

catalyze N–N bond formation, transforming the reactive *O*-acylhydroxylamine intermediates into hydrazine molecules through intramolecular rearrangement (cupin domain in Figure 3.1).[59]

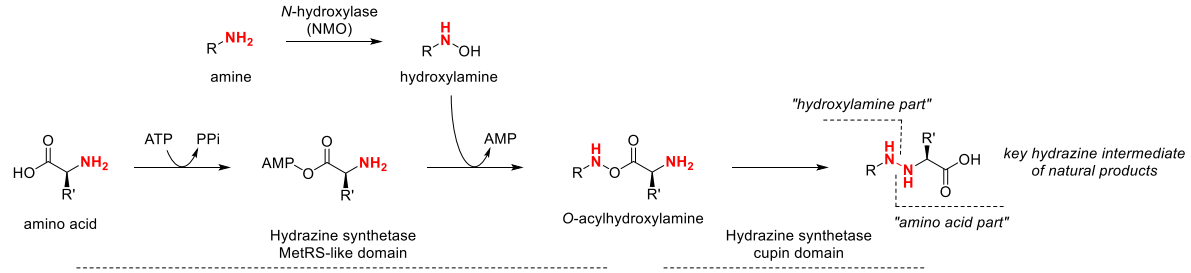


Figure 3.1 N-N bond formation mediated by *N*-hydroxylases and hydrazine synthetases**.**

The resultant hydrazine products serve as initial intermediates featuring N–N bonds, undergoing diverse biosynthetic modifications to yield functionalities such as hydrazone, diazo, and *N*-hydroxyl triazene groups, as well as N–N bond-containing heterocycles like pyrazole and dihydropyridazinone. To date, hydrazine synthetases have been implicated in the biosynthesis of six natural products, some of which hold pharmaceutical significance (Figure 3.2). These encompass a non-ribosomal peptide (s56-p1)[60], amino acid derivatives (azaserine and actinopyridazinone)[37], a fatty acid derivative (triacsin)[32], and a *C*-nucleoside (formycin/pyrazomycin)[57]. While not experimentally validated, a putative hydrazine synthetase is also encoded in the biosynthetic gene cluster (BGC) of the N–N bond-containing xanthone-type polycyclic polyketide albofungin[61]. These examples underscore the diverse array of biosynthetic pathways that leverage hydrazine synthetases.

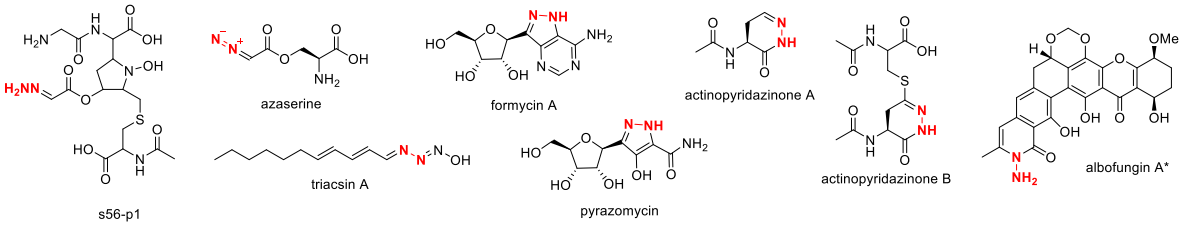


Figure 3.2 N–N bond containing natural products biosynthesized by hydrazine synthetases. *Involvement of hydrazine synthetase in albofungin biosynthesis is not experimentally validated.

The hydrazine structure can be dissected into two constituent parts: a *N*-hydroxylamine-derived segment and an amino acid-derived segment (Figure 3.1). Recent investigations have elucidated the diverse combinations of these components, wherein *N*-hydroxylamine precursors include N^4-OH L-diaminobutyric acid (N^4-OH L-DABA)[37], N^5-OH L-ornithine[37], and N^6-OH L-lysine[59], while the amino acid counterparts encompass glycine, L-alanine[37], L-serine[59], L-glutamic acid[57], and L-tyrosine[59]. Presently, seven distinct hydrazines have been identified as products of the hydrazine synthetase (HS) family. Among these, the biosynthetic products of three hydrazines (Lys-Gly[60], Lys-Glu[57], DABA-Ala[37]) have been established (Figure 3.3), whereas the remaining four hydrazines (Lys-Ala[59], Lys-Ser[59], Lys-Tyr[59], Orn-Gly[37]), discovered through an enzyme-mining approach, await association with their respective biosynthetic products.

A bioinformatic analysis has highlighted that the characterized NMO/HS pairs to date represent only a fraction of the complete sequence diversity within this protein family[60]. Furthermore, the hydrazine biosynthetic pathways mediated by NMO/HS are found across diverse bacterial phyla, including actinobacteria, proteobacteria, firmicutes, bacteroidetes, deinococcus-thermus, and cyanobacteria[60]. Notably, the experimentally characterized NMO/HS pairs are predominantly derived from actinobacteria, with the exception of Por9/Por11 originating from proteobacteria.[37] These observations underscore the unexplored chemical diversity inherent in the hydrazines generated by the HS family.

In this study, I conducted a phylogeny-guided enzyme-mining of hydrazine-forming pathways to explore the structural diversity of amino acid-based hydrazine in nature. Characterizations of NMOs and HSs from previously unexplored clades provided an extensive view of the structural diversity exhibited by hydrazine products. Our global survey identified several hydrazines with novel combinations of building blocks which were not previously recognized as hydrazine building blocks. The array of hydrazine intermediates suggests the immense structural diversity of N–N bond-containing natural products synthesized through HS-mediated hydrazine formation.

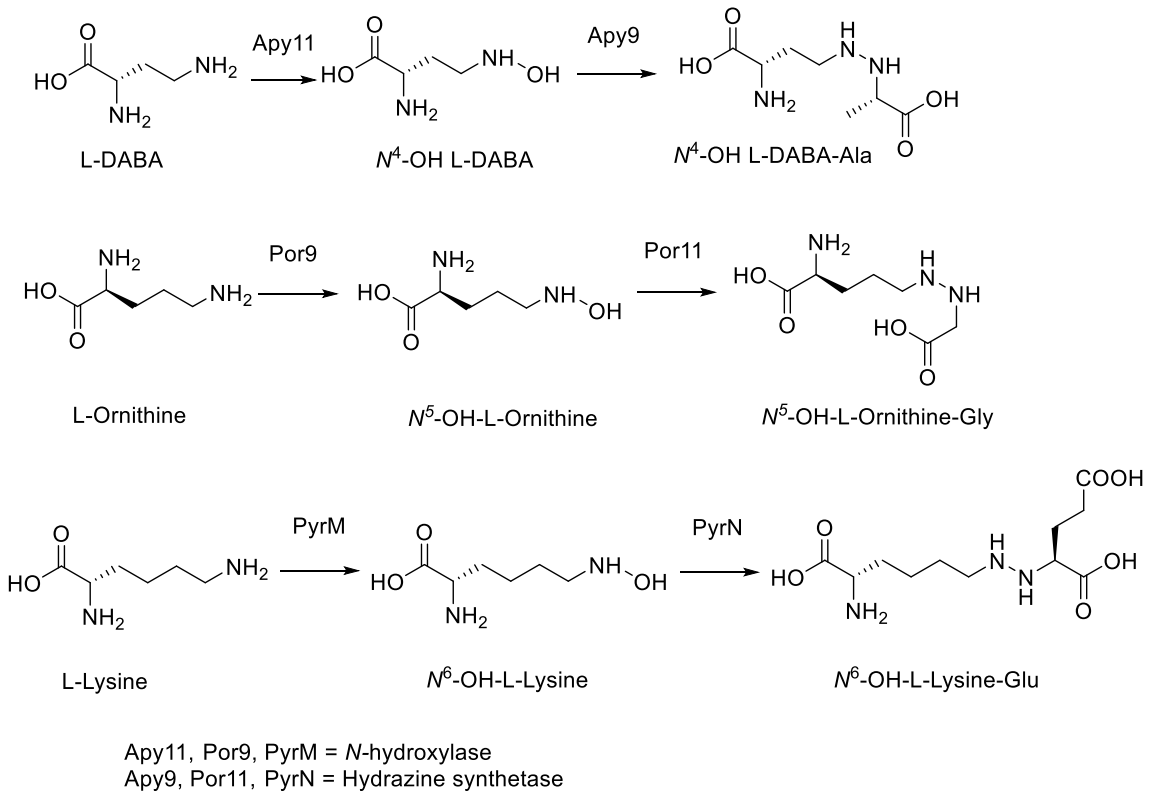


Figure. 3.3 Biosynthetic pathway of some knowns *N*-hydroxylases and hydrazine synthetases

3.2 Method

Bioinformatics

839 HS sequences in the previous dataset were searched by phmmer program using Apy9 (BDC79915.1) as a query. 20-kb flanking regions of each HS were retrieved from GenBank using entrez programming utilities, then subsequently, NMOs were searched using phmmer program using Apy11 (BDC79917.1) as a query to give resultant 422 NMOs. Taxonomic information was retrieved using TaxonKit[62] (v0.15.1) based on Genbank IDs. Multiple sequence alignments were generated by muscle (v3.8) with default parameters. Maximum-likelihood phylogenetic trees were constructed using MEGA X[63], then visualized, and edited by webtool iTOL v6[64].

Preparation of Plasmid *bac3*-pColdII

bac3-pColdII construct was prepared by digesting *bac3* synthetic gene and pColdII plasmid (Takara) using NdeI/HindIII (Takara) restriction enzymes followed by overnight incubation at 37 °C. The digestion product was separated on the electrophoresis gel for about 20 min. The target DNA band was cute and extracted using the gel extraction method (in procedure e). Ligation was done by mixing the gel-extracted *bac3* and pColdII in Solution 1 (Takara) containing ligase enzyme. The ligation mixture was incubated at 4 °C overnight, then transformed into *E. coli* DH5α on LB agar containing ampicillin, and incubated overnight at 37 °C. The colonies were individually cultured overnight in 2xYT liquid media containing ampicillin. Recombinant plasmids were extracted from the cultures and checked for the presence of correct-sized insert by NdeI/HindIII digestion.

Preparation of Plasmid *bac1/2*-pRSF duet

bac2-pRSF duet construct was prepared by digest *bac2* synthetic gene (MetRS) and pRSF duet plasmid (Takara) using NdeI/XhoI (Takara) restriction enzymes and incubated them at 37 °C overnight. The next day, the digestion result was loaded on electrophoresis gel for about 20 min, the band was cut and DNA was extracted using gel extraction method. Ligation was done by mixing *bac2* and pRSF duet from gel extraction, ethanol precipitation

was done, and solution 1 (Takara) contains ligase enzyme was added. The mixture solution was put at 4 °C for overnight. The next day, the ligation result was transformed to *E. coli* DH5α in LB agar containing kanamycin and was incubated overnight at 37 °C. The colonies were cultured overnight in 2xYT media containing kanamycin, plasmid extraction was done and plasmid quality was checked by NdeI/XhoI digestion. Then, the digestion result was added on electrophoresis gel.

To prepare *bac1/2*-pRSF duet construct, *bac1*-pTwist (cupin) and *bac2*-pRSF duet were digested using EcoRI/HindIII (Takara) restriction enzymes and incubated them at 37 °C overnight. The next day, the digestion result was loaded on electrophoresis gel for about 20 min, the band was cut, and DNA was extracted using gel extraction method. Ligation was done by mixing *bac2*-pRSF and *bac1* from gel extraction, ethanol precipitation was done, and solution 1 (Takara) contains ligase enzyme was added. The mixture solution was put at 4 °C for overnight. The next day, the ligation result was transformed to *E. coli* DH5α in LB agar containing kanamycin and was incubated overnight at 37 °C. The colonies were cultured overnight in 2xYT media containing kanamycin, plasmid extraction was done, and plasmid quality was checked by EcoRI/HindIII (Takara) and NdeI/XhoI (Takara) digestion. Then, the digestion result was added on electrophoresis gel.

Preparation of Plasmid *kit3*-pColdII

kit3-pColdII construct was prepared by digest *kit3*-pTwist synthetic gene and pColdII plasmid (Takara) using NdeI/HindIII restriction enzymes and incubated them at 37 °C overnight. The digestion product was separated on the electrophoresis gel for about 20 min. The target DNA band was cute and extracted using the gel extraction method (in procedure e). Ligation was done by mixing the gel-extracted *kit3* and pColdII in Solution 1 (Takara) containing ligase enzyme. The ligation mixture was incubated at 4 °C overnight, then transformed into *E. coli* DH5α on LB agar containing ampicillin, and incubated overnight at 37 °C. The colonies were individually cultured overnight in 2xYT liquid media containing ampicillin. Recombinant plasmids were extracted from the cultures and checked for the presence of correct-sized insert by NdeI/HindIII digestion.

Preparation of Plasmid *kit1/2*-pRSF duet

kit2-pRSF duet construct was prepared by digest *kit2*-pTwist synthetic gene and pRSF duet plasmid (Takara) using NdeI/XhoI (Takara) restriction enzymes and incubated them at 37 °C overnight. The next day, the digestion result was loaded on electrophoresis gel for about 20 min, the band was cut and DNA was extracted using gel extraction method. Ligation was done by mixing *kit2* and pRSF duet from gel extraction, ethanol precipitation was done, and solution 1 (Takara) contains ligase enzyme was added. The mixture solution was put at 4 °C for overnight. The next day, the ligation result was transformed to *E. coli* DH5α in LB agar containing kanamycin and was incubated overnight at 37°C. The colonies were cultured overnight in 2xYT media containing kanamycin, plasmid extraction was done and plasmid quality was checked by NdeI/XhoI digestion. Then, the digestion result was added on electrophoresis gel.

To prepare *kit1/2*-pRSF duet construct, *kit1*-pTwist (cupin) and *kit2*-pRSF duet were digested using EcoRI/HindIII (Takara) restriction enzymes and incubated them at 37 °C overnight. The next day, the digestion result was loaded on electrophoresis gel for about 20 min, the band was cut, and DNA was extracted using gel extraction method. Ligation was done by mixing *kit2*-pRSF and *kit1* from gel extraction, then ethanol precipitation was done, and solution 1 (Takara) contains ligase enzyme was added. The mixture solution was put at 4 °C for overnight. The next day, the ligation result was transformed to *E. coli* DH5α in LB agar containing kanamycin and was incubated overnight at 37 °C. The colonies were cultured overnight in 2xYT media containing kanamycin, plasmid extraction was done, and plasmid quality was checked by EcoRI/HindIII (Takara) and NdeI/XhoI (Takara) digestion. Then, the digestion result was added on electrophoresis gel.

Preparation of Plasmid *cor1/2*-pRSF duet

cor2-PET28a synthetic genes and PRSF duet vector were digested using NdeI/XhoI (Takara) restriction enzymes for overnight at 37 °C. Then loaded to gel electrophoresis, cut the band, and extracted the gel. pRSF duet and *cor2* were mixed and ethanol precipitation was done. MQ water was added to the mixture, then add solution 1 (Takara) containing ligase and incubated overnight at 4 °C. Ligation mixture was transformed to E. coli DH5α and was spreaded to LB agar containing kanamycin. The insert was confirmed by plasmid extraction and restriction enzyme digestion.

Cupin primer (*cor1*) was amplified by PCR using KAPA Hifi kit. PCR result was loaded to gel electrophoresis, and bands were cut and extracted, and ethanol precipitation was conducted. MQ water was added to the tube and digest the cupin using restriction enzymes EcoRI/HindIII (Takara) at 37 °C for overnight. At the same time, plasmid *cor2*-pRSF duet was also treated by EcoRI/HindIII (Takara) digestion. The next day, the digestion result was loaded on electrophoresis gel for about 20 min, the band was cut, and DNA was extracted using gel extraction method. Ligation was done by mixing *cor2*-pRSF and *cor1* from gel extraction, then ethanol precipitation was done, and solution 1 (Takara) contains ligase enzyme was added. The mixture solution was put at 4 °C for overnight. The next day, the ligation result was transformed to *E. coli* DH5α in LB agar containing kanamycin and was incubated overnight at 37 °C. The colonies were cultured overnight in 2xYT media containing kanamycin, plasmid extraction was done, and plasmid quality was checked by EcoRI/HindIII (Takara) and NdeI/XhoI (Takara) digestion. Then, the digestion result was added on electrophoresis gel.

Preparation of Plasmid *col1/2*-pRSF duet

col2-PET28a synthetic genes and PRSF duet vector were digested using NdeI/XhoI (Takara) restriction enzymes for overnight at 37 °C. Then loaded to gel electrophoresis, cut the band, and extracted the gel. pRSF duet and *col2* were mixed and ethanol precipitation was done. MQ water was added to the mixture, then add solution 1 (Takara) containing ligase and incubated overnight at 4 °C. Ligation mixture was transformed to *E. coli* DH5α and was spreaded to LB agar containing kanamycin. The insert was confirmed by plasmid extraction and restriction enzyme digestion.

Cupin primer (*col1*) were amplified by PCR using KAPA Hifi kit. PCR result was loaded to gel electrophoresis, and bands were cut and extracted, and ethanol precipitation was conducted. MQ water was added to the tube and digest the cupin using restriction enzymes EcoRI/HindIII (Takara) at 37 °C for overnight. At the same time, plasmid *cor2*-pRSF duet was also treated by EcoRI/HindIII (Takara) digestion. The next day, the digestion result was loaded on electrophoresis gel for about 20 min, the band was cut, and DNA was extracted using gel extraction method. Ligation was done by mixing *col2*-pRSF and *col1* from gel extraction, then ethanol precipitation was done, and solution 1 (Takara) contains ligase enzyme was added. The mixture solution was put at 4 °C for overnight. The next day, the ligation result was transformed to *E. coli* DH5α in LB agar containing kanamycin and

was incubated overnight at 37°C. The colonies were cultured overnight in 2xYT media containing kanamycin, plasmid extraction was done, and plasmid quality was checked by EcoRI/HindIII (Takara) and NdeI/XhoI (Takara) digestion. Then, the digestion result was added on electrophoresis gel.

In vitro Assay of NMO

Purified protein was transferred to centrifugal filtration tube, lysis buffer was added, and centrifuged at 8000 xg. *In vitro* assay was conducted by adding Tris HCl 1M, FAD, nicotinamide adenine dinucleotide phosphate (NADPH), amino acid (L-lysine, L-ornithine, DAB, DAP), and putative NMO enzyme. Reaction mixture was incubated at 30 °C for 30 minutes.

Fmoc derivatization

Derivatization was conducted by adding Fmoc Cl (20 mM in acetonitrile), and borate buffer 200 mM to reaction mixture. Then reaction mixture was incubated at 30 °C for 30 minutes, and reaction was quenched with methanol and put at -30 °C for about 10 minutes. Centrifuged the reaction mixture at 15,000 rpm 10 minutes, and injected to UPLC MS system using C18 cholester column, mobile phase H_2O/MeCN (Trifluoroacetic acid additive) 40-98% in 20 minutes.

Bioconversion of Hydrazine Synthetase

Bacteria were cultured at 2xYT liquid media containing suitable antibiotic overnight. 1% of this culture was transferred to M9 media, then incubated at 37 °C for 6 hours, and feed with IPTG 0.1 mM and suitable hydroxy amino acid. Continue the incubation at 30 °C for 16 hours. Culture was centrifuged at 10,000 rpm for 10 minutes and transferred the supernatant to another tube.

Dansyl chloride derivatization

Derivatization of the supernatant was conducted by adding acetonitrile, sodium bicarbonate buffer 0.5 M pH 9.5, and dansyl chloride (50 mM in acetonitrile) and incubated at 30 °C for 1 hour. Reaction mixture was quenched with methanol, centrifuged at 15,000 rpm, and injected to Shimadzu HPLC prominence system coupled with a Shimadzu LCMS-

2020 spectrometer or an Amazon SL-NPC system (Bruker Daltonics) using C18 MSII column, mobile phase H_2O/MeCN (formic acid additive) 2-98% for 20 min.

3.3 Results and Discussion

To explore the structural diversity of hydrazine intermediates, we identified the substrate specificity of the enzymes responsible for the formation of hydrazines. The hydrazine structure consists of two parts, namely hydroxylamine part generated by flavin-dependent N^{ω}-hydroxylating monooxygenases (NMOs) enzyme, and the amino acid part installed by MetRS domain of the hydrazine synthetase.[37] Therefore, to get better insight into the structural diversity of hydrazine intermediates, we examined the substrate specificity of the *N*-hydroxylases and hydrazine synthetases identified in this work.

3.3.1 Diverse Substrate Specificities of *N*-hydroxylase (NMO)

Putative *N*-hydroxylase enzymes was searched from public database by HMM-based method. This resulted in the identification of 422 putative *N*-hydroxylases. Those members were subsequently subjected to phylogenetic analysis, and the result showed that all of the identified NMOs were classified into eleven clades likely based on their substrate specificity towards certain amino acids. Then, to predict the substrate specificity of each clade, I integrated information of known *N*-hydroxylases to the phylogenetic tree (Figure 3.4).

Clade V.

A unique hydroxylase, Apy11, that is specific to diamino butyric acid (DABA), belongs to clade V. This hydroxylase is involved in the biosynthesis of DABA-based hydrazine, a precursor to the N–N bond-containing heterocycle dihydropyridazinone.[37]

Clade VI.

N-hydroxylases, such as SidA, PvdA, and KtzT, specifically catalyze the hydroxylation of ornithine.[36,65] The resultant hydroxylated ornithine is used as a precursor for the hydroxamate group in ornithine-based siderophores or undergoes intramolecular cyclization to form piperazate. Piperazate is then integrated into various piperazate-containing peptide natural products.[36] These ornithine-specific hydroxylases are confined to clade VI, therefore, clade VI represents a group of ornithine hydroxylases.

Clade IX.

In another example, the hydroxylases DesB, GorA, and AlcA were specific for diamine substrates, such as putrescine and cadaverine, thereby producing hydroxylated diamines. The resulting hydroxylated diamines serve as precursors for hydroxamate moieties in diamine-containing siderophores.[66] These hydroxylases belonged to clade IX, which represents a group of diamine-specific hydroxylases.

Clade XI.

Known lysine-N^6-hydroxylases involved in the biosynthesis of hydroximate moieties in lysine-based siderophore such as LucD, VbsO, BibB are categorized into clade XI[67]. Also, known lysine-N^6-hydroxylases Spb38[60], Tri26[32], AzaG[68], PyrM[57], ForK[69] were classified into clade XI in the phylogenetic tree. Therefore, other uncharacterized NMO enzymes in the clade XI are likely to possess specificity toward L-lysine.

Above mentioned phylogenetic analysis suggested the specificity of NMOs in some clades: clade V (L-DABA) , clade VI (L-ornithine), clade IX (diamine), and clade XI (L-lysine). At the same time, the substrates of NMOs in clades I-IV, VII, VIII, and X remain unpredictable. Therefore, I selected several NMOs mainly from unpredictable clades for experimental characterization.

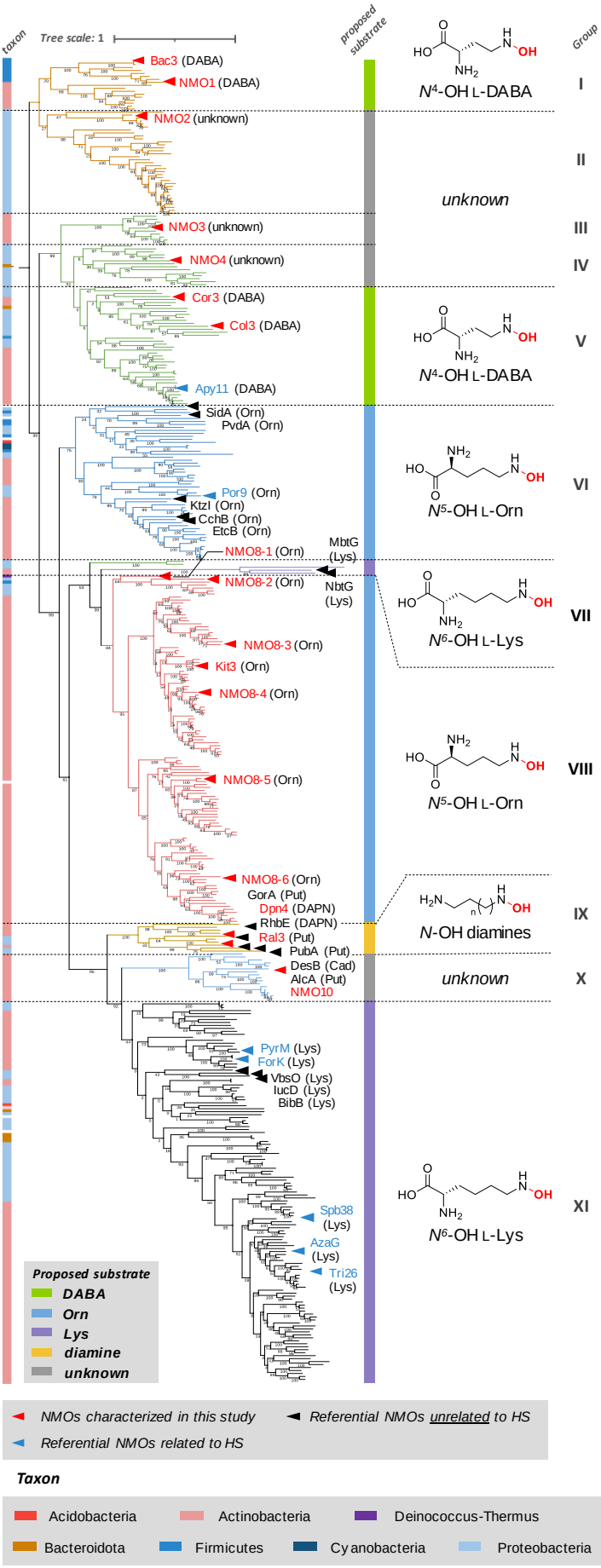


Figure 3.4 Phylogenetic tree of *N*-hydroxylase. Maximum-likelihood phylogenetic tree of 444 NMOs. Colored branches represent different NMO groups. Known NMOs and those characterized in this study are highlighted by colored arrows. Substrates are described for each NMO. Taxonomic information is shown in the color strip on the left. Proposed substrates for each clade of NMOs are shown on the right. Color schemes are described at the bottom of the tree.

Characterization of NMOs in Clade VIII

Kit3, a Putative NMO in Clade VIII.

First enzyme that we characterized was Kit3 – a *N*-hydroxylase from *Kitasatospora gansuensis*, whose NMO (i.e. kit3) belongs to clade VIII in phylogenetic tree. This strain has three interesting genes that were also found in hydrazine-producing organisms[37]: Kit 3 (putative N-hydroxylase), Kit1/2 (putative hydrazine synthetase), and Kit4 (putative FAD binding oxidoreductase). For the in vitro assay of *N*-hydroxylase, we prepare the recombinant protein in a large-scale (200 mL culture) supplemented with IPTG to induce its production. The His-tagged recombinant protein was purified using Ni-NTA column. In the purification step, non-target proteins were removed by adding wash buffer containing low concentrations of imidazole thereby removing non-target proteins with low affinity to column. Finally, the His-tagged target protein was eluted using elution buffer containing high concentration of imidazole. The eluted protein was then concentrated with an appropriate buffer for the *in vitro* assay. In the assay, FAD and NADPH were added as cofactors for the reaction using one of the following amino acids as the substrates: DAP, L-DABA, L-lysine, and L-ornithine. The mixture reaction was then subjected to Fmoc-Cl derivatization. Fmoc-Cl is a reagent used both in *N*-protection reactions of α-amino acids and in quantification and detection of amino acids using HPLC-PDA or fluorescence analysis.[70] The benefit of Fmoc derivatization including stable derivative products, complete amino-acid reaction, low detection limit and good reproducibility.[71]

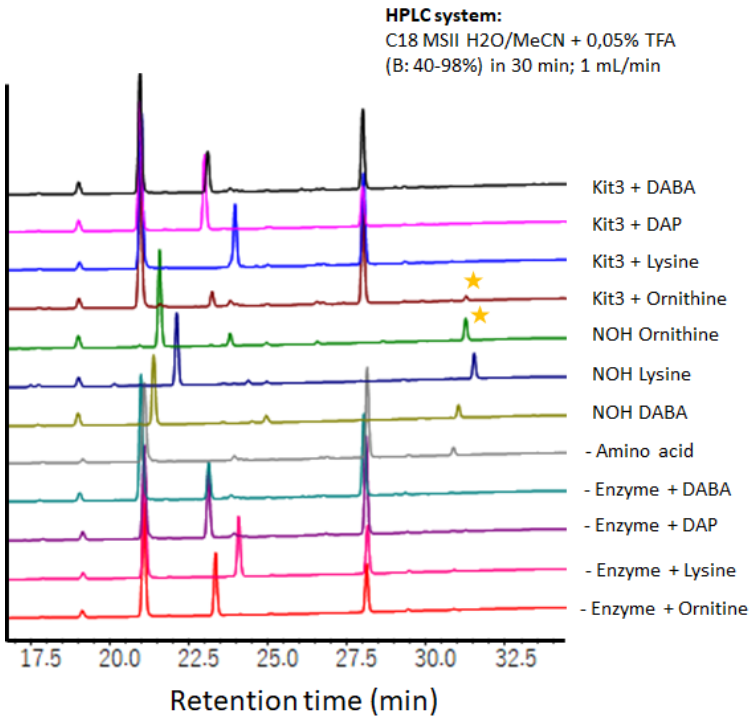


Figure. 3.5 Kit3 in vitro assay result

The result from in vitro assay showed that Kit3 could accept L-ornithine as substrate and convert it to N^5-OH L-ornithine (Figure 3.5). However, this enzyme could not accept other substrates such as L-lysine, DAP (L-2,3-diaminopropionic acid), and DABA (L-2,4-diaminobutyric acid), suggesting that the specificity of Kit3 toward L-ornithine. Then, we eliminated cofactors in the reaction to know which cofactor has essential role for the in vitro reaction of Kit3. The result revealed that omitting NADPH abolished the formation of N^5-OH L-ornithine (Figure 3.6). Therefore, NADPH is indispensable for Kit3 *N*-hydroxylation.

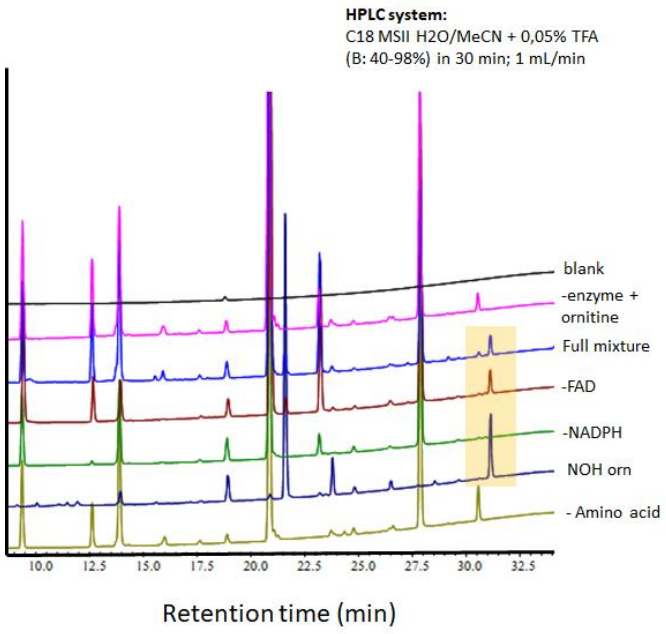


Figure 3.6 Kit3 is NADPH dependent enzyme.

***Nmo8-4*, a Putative NMO in Clade VIII.**

Another *N*-hydroxylase from *Saccharothrix variisporea* that belonged to clade VIII, labelled as Nmo8-4, were tested for *in vitro* assay. It was found that recombinant Nmo8-4 could recognize L-ornithine (Figure 3.7). However, Nmo8-4 did not accept L-lysine, DAP, and DABA, suggesting that this enzyme is ornithine-specific *N*-hydroxylase.

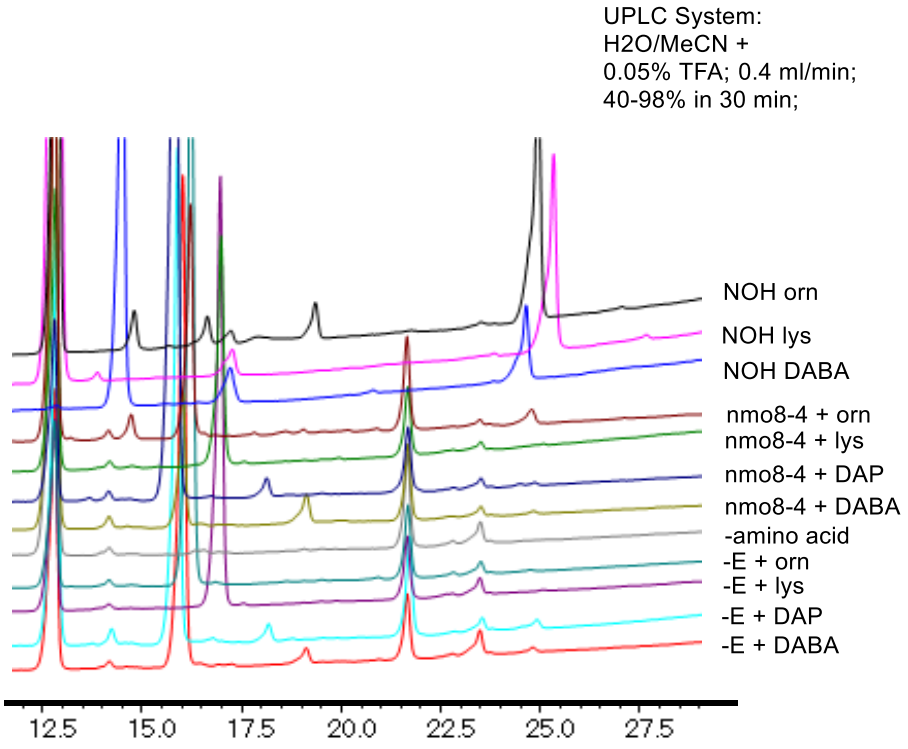


Figure 3.7. HPLC analysis of enzymatic reaction mixtures of Nmo8-4. Column elutes were monitored by UV absorption at 263 nm.

Nmo8-6, a Putative NMO in Clade VIII.

The investigation of clade VIII specificity continued with Nmo8-6, a putative *N*-hydroxylase from *Actinokineospora fastidiosa*, which was predicted to accept L-ornithine, the same substrate shown by Kit3 and Nmo8-4. To confirm this prediction, Nmo8-6 protein was purified using Ni-NTA column, and subjected to *in vitro* assays using several different amines as the substrates: L-lysine, L-ornithine, DAP, and DABA. The results showed that Nmo8-6 could accept L-ornithine, consistent with its phylogenetic tree affiliation (Figure 3.8).

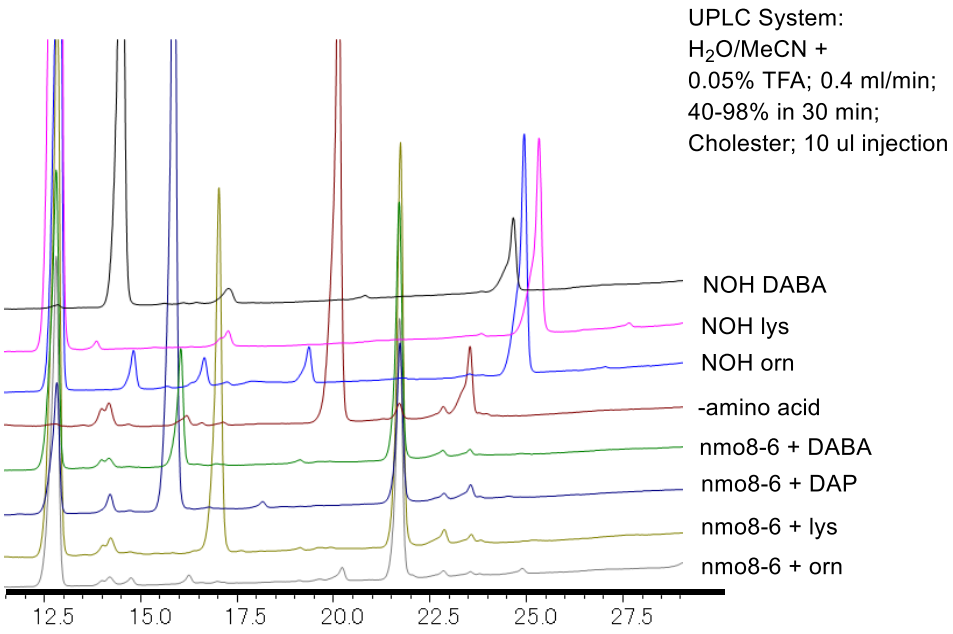


Figure 3.8. HPLC analysis of enzymatic reaction mixtures of Nmo8-6. Column elutes were monitored by UV absorption at 263 nm.

While Nmo8-6 could recognize L-ornithine, we could see that the peak generated from N^5-OH L-ornithine was very low, as shown in Figure 3.8. It was predicted that the product of reaction underwent decomposition. We optimized the reaction duration and found that the optimum incubation time of Nmo8-6 was 20 minutes (Figure 3.9).

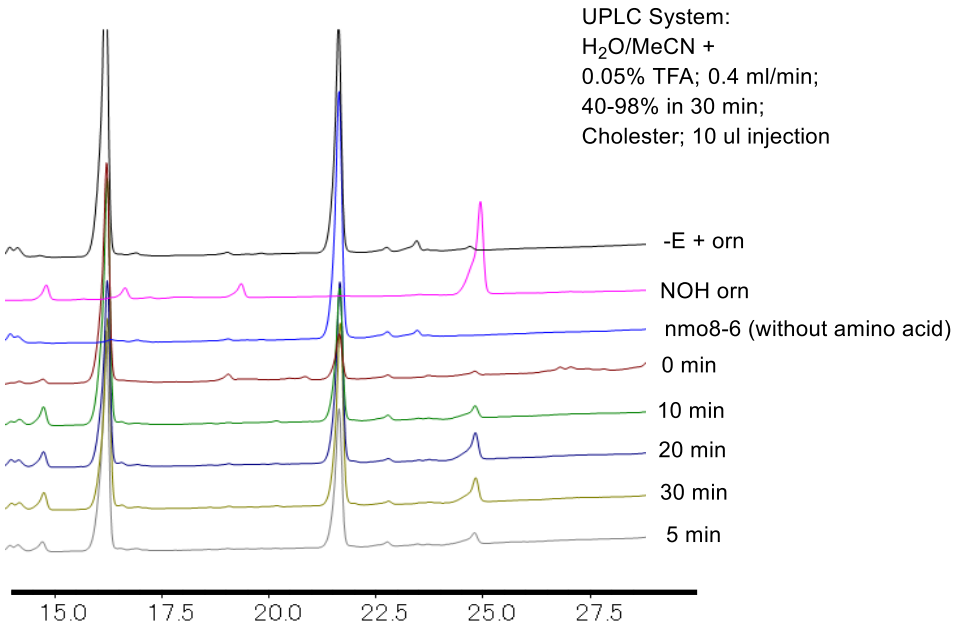


Figure. 3.9. Optimization of incubation time for Nmo8-6 activity. Column elutes were monitored by UV absorption at 263 nm.

Nmo8-1, a Putative NMO in Clade VIII.

Another clade VIII member, Nmo8-1 from *Deinococcus ficus* was tested for the *in vitro* assay. Subjecting the purified recombinant protein Nmo8-1 to i*n vitro* assay using L-lysine, L-ornithine, DAP, and DABA as the substrates revealed that Nmo8-1 possessed specificity to L-ornithine, the same specificity as other characterized members of the clade VIII (Figure 3.10).

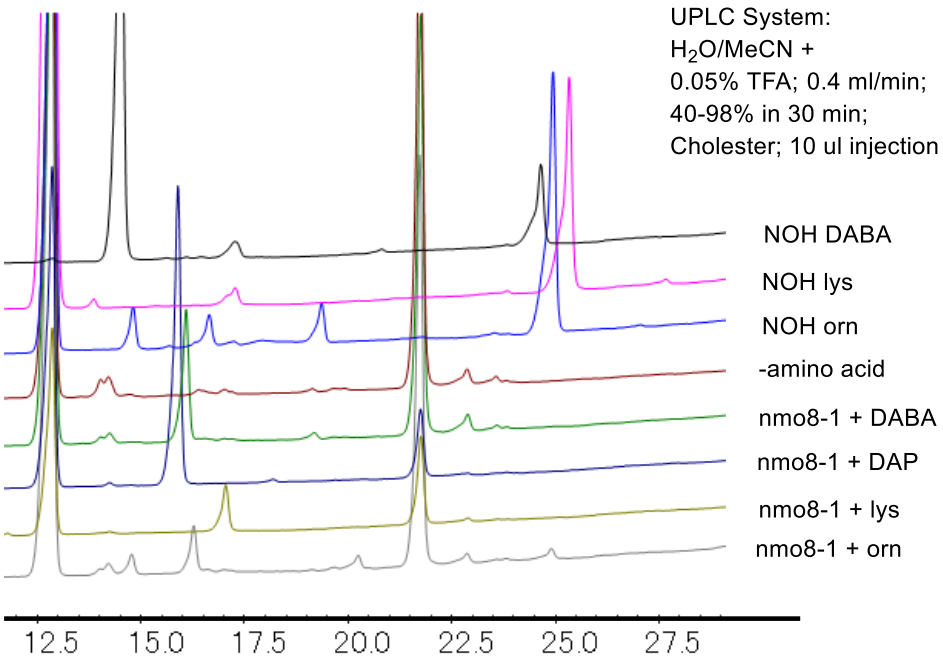


Figure 3.10 HPLC analysis of enzymatic reaction mixtures of Nmo8-1. Column elutes were monitored by UV absorption at 263 nm**.**

Nmo8-2, a Putative NMO in Clade VIII.

The last enzyme in the clade VIII is Nmo8-2 from *Brochothrix thermosphacta*. *In vitro* activity assay was conducted using purified Nmo8-2 by mixing it with an amine (L-lysine, L-ornithine, DAP, or DABA). The result showed its specificity toward L-ornithine (Figure 3.11). Based on the consistent results shown by Kit3, Nmo8-1, Nmo8-2, Nmo8-4, and Nmo8-6, we conclude that *N*-hydroxylases that belong to clade VIII have specificity to L-ornithine.

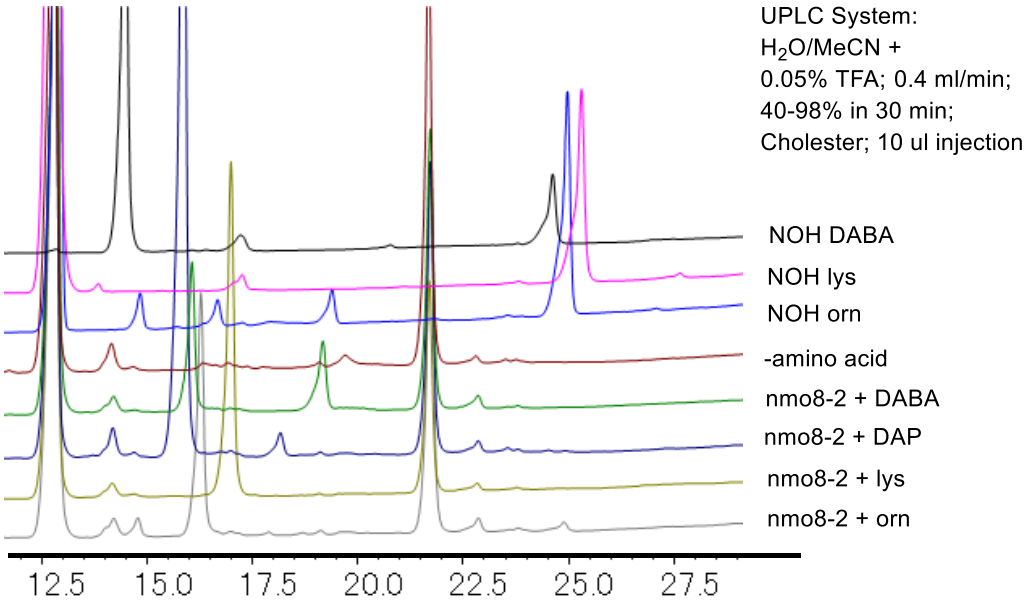


Figure 3.11 HPLC analysis of enzymatic reaction mixtures of Nmo8-2. Column elutes were monitored by UV absorption at 263 nm**.**

Characterization of NMOs in clade I

Bac3, a Putative NMO in Clade I.

Next, I move to another uncharacterized clade, which is clade I. From this clade, we chose two *N*-hydroxylases named as Bac3 and Nmo1. Bac3 is *N*-hydroxylase from *Bacillus cereus* of Firmicutes group. As almost all characterized *N*-hydroxylase are derived from Actinobacteria, except Por9 from Proteobacteria, this enzyme is the first characterized *N*-hydroxylase from Firmicutes. *In vitro* activity assay of Bac3 was conducted by applying purified protein to *in vitro* assay using the same manner as other *N*-hydroxylases. After derivatization and injection to UPLC system, there was a peak coming from mixture of protein and L-DABA, showing that this enzyme has specificity to L-DABA (Figure 3.12). Previously, Apy11, an *N*-hydroxylase that responsible for formation N–N bond in actinopyridazinone, is the only enzyme that accept L-DABA. Therefore, Bac3 is the second *N*-hydroxylase that accept L-DABA. Notably, Bac3 is the only known *N*-hydroxylase related to hydrazine biosynthesis from Firmicutes.

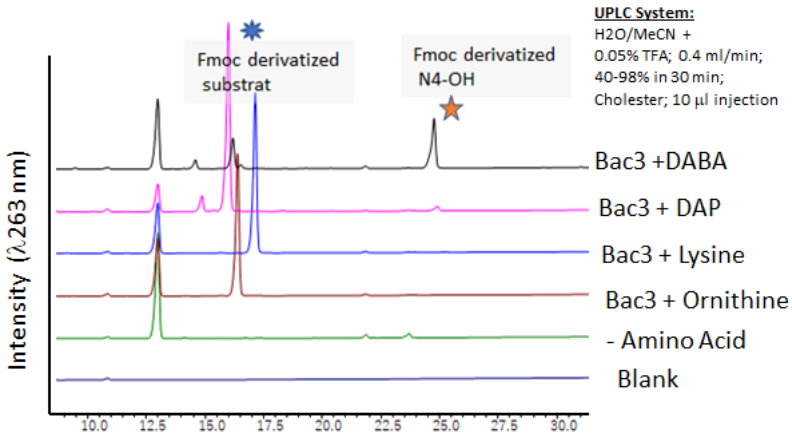


Figure 3.12 HPLC analysis of enzymatic reaction mixtures of Bac3. Column elutes were monitored by UV absorption at 263 nm.

Nmo1, a Putative NMO in Clade I.

Another enzyme that belongs to clade I, Nmo1 was tested for *in vitro* activity assay. We assumed that this enzyme would exhibit specificity toward L-DABA, as shown by the close phylogenetic relation with Bac3. To confirm this prediction, Nmo1 enzyme was purified and tested *in vitro* to several amines. The result revealed that this enzyme had specificity to L-DABA (Figure 3.13), consistent with its closeness position to Bac3 in the phylogenetic tree. Therefore, based on the result from Bac3 and Nmo1 *in vitro* assays, we concluded that clade I consist of *N*-hydroxylases that accept L-DABA.

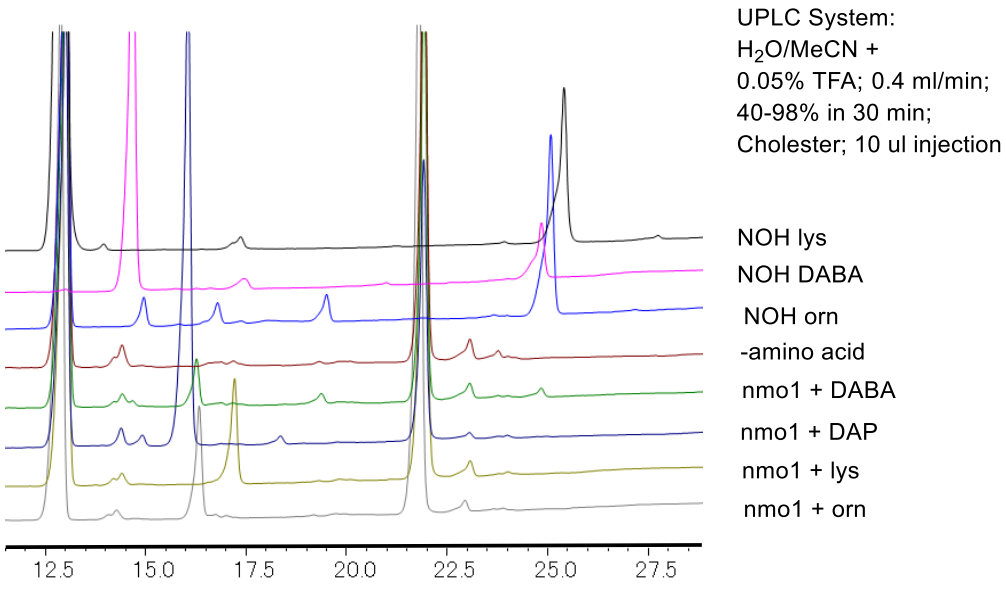


Figure 3.13 HPLC analysis of enzymatic reaction mixtures of Nmo1. Column elutes were monitored by UV absorption at 263 nm.

NMOs from Other Putative Clades.

In vitro characterization of other NMOs from clade II (Nmo2), III (Nmo3), IV (Nmo4), X (Nmo10) were also conducted, however, these four NMOs showed no *N*-hydroxylating activities toward any amines tested, leaving certain NMO sequence spaces uncharacterized (Figure S29-33). These NMOs may hydroxylate specialized amines biosynthesized through dedicated pathways. This possibility is currently under investigation. Overall, the phylogeny guided global enzyme-mining offered a comprehensive perspective

on the functional diversity of NMOs. This information will facilitate the prediction of the hydroxylamine substrates of cognate HSs.

3.3.2 Diverse Substrate Specificities of Hydrazine Synthetase

The amino acid part of hydrazine structure is determined by the specificity of MetRS domain. *C*-terminal MetRS-like domain activates amino acid in ATP dependent manner, then couple it with hydroxylamine to give *O*-acylhydroxylamine intermediate. Subsequent rearrangement of the intermediate by cupin domain generates hydrazine structure.[37] MetRS domain has important role for the amino acid substrate recognition. Therefore, to explore the diversity of substrate specificity among hydrazine synthetase (HS) family, we retrieved 401 MetRS-like proteins from public database and subjected them to phylogenetic analysis. Phylogenetic analysis showed more than thirteen groups of HS enzymes with potentially distinct substrate recognition. (Figure 3.14).

Then, we integrated known HS enzymes to the phylogenetic tree. Some of the HS enzymes such as Spb40, Tri26, AzaE could accept glycine[32,60,68]. Their MetRS domain couple N^6-OH lysine and glycine in an ATP-dependent manner to give ester intermediate, followed by the cupin domain-catalyzed formation of glycine-based hydrazine. Glycine-based hydrazine is precursor for some N–N bond-containing compounds such as hydrazone, diazo, and triazene.[32,60,68] According to the analysis, these enzymes belong to clade XIII in the phylogenetic tree. In another enzyme, MetRS domain in Apy9 (Clade XI) has unique specificity to alanine and convert it to *O*-acylhydroxylamine. This ester intermediate is converted to alanine-based hydrazine (i.e. DABA-Ala) by cupin domain, which is precursor for N–N bond-containing natural products such as dihydropyridazinone.[37] In the phylogenetic tree, Apy9 belong to clade XI. The other known enzymes, ForJ and PyrN, were able to activate glutamate and catalyzed the condensation between N^5-OH ornithine and glutamate to generate Orn-Glu.[43,57] This is further converted to pyrazole moiety in *C*-nucleosides.

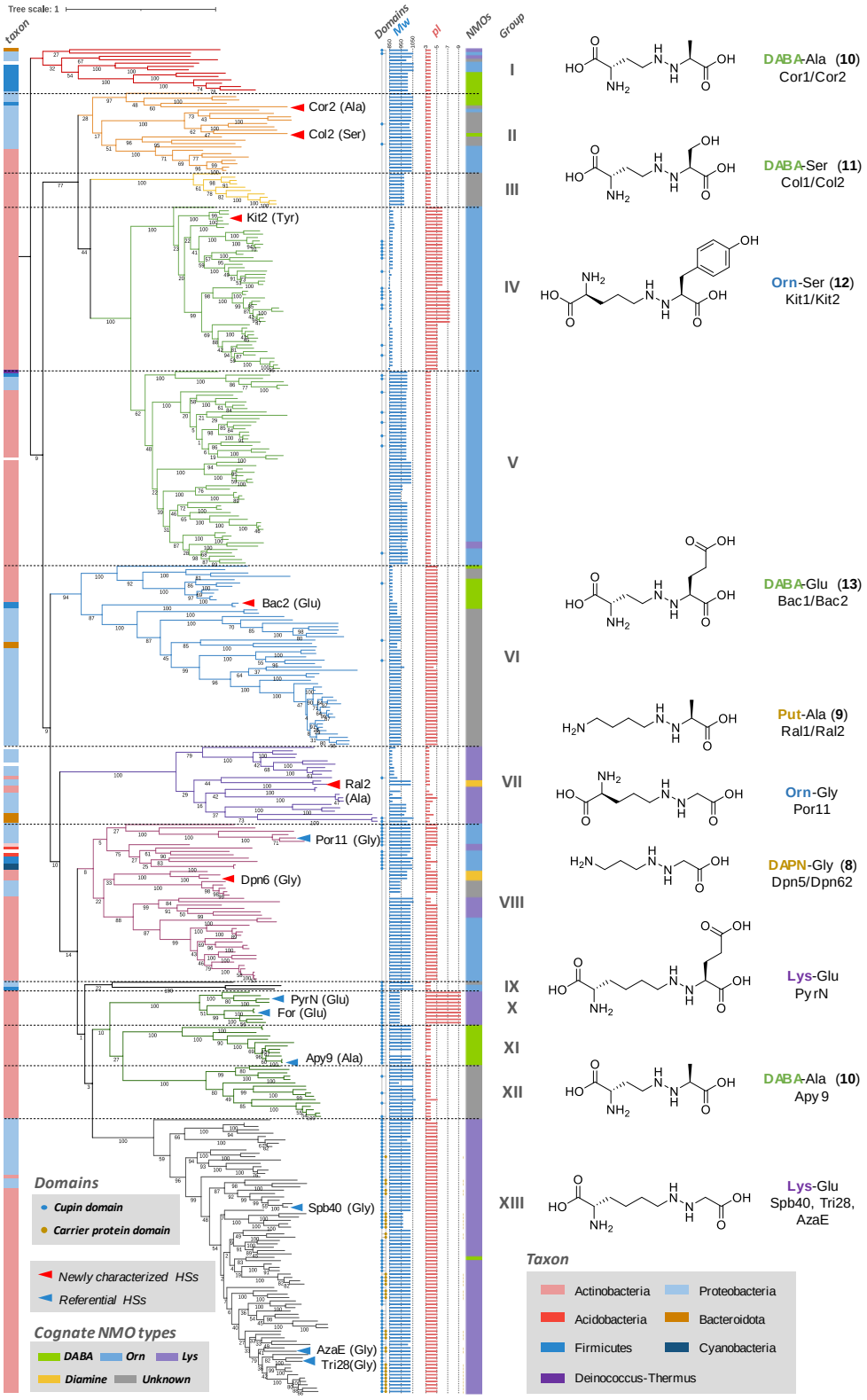


Figure 3.14 HS phylogenetic tree Maximum-likelihood phylogenetic tree of 401 HSs. Colored branches represent different HS groups. Known HSs and those characterized in this study are highlighted by colored arrows. Taxonomic information is shown in the color strip on the left side. On the right side of the tree, molecular weights (Mw) and isosteric points (pI) of pocket-forming residues in MetRS-like domains are shown. Fused cupin and carrier protein domains are represented by blue and brown dots, respectively (Domains). The types of cognate NMOs are shown as a colored strip on the right side (NMOs).

Then, we investigated genomic neighboring regions of HSs to search cognate NMO genes. The cognate NMOs were subjected to phylogenetic analysis and their substrate amine was predicted based on the clade affiliation. The resultant information was integrated into the phylogenetic tree of HSs to get better insight about the combination of building blocks for hydrazine (see NMOs color strip in Figure 3.14). In the Figure 3.14, glycine accepting HS in clade XIII are connected to NMOs that give L-lysine-N^6-hydroxylase (purple strip), therefore the hydrazine product generated by clade XIII HS can be predicted as lysine-glycine based hydrazine (Lys-Gly). Indeed, three known HSs Spb40, AzaG, and Tri28 all gives Lys-Gly hydrazine as biosynthetic intermediates of s56-p1, azaserine, and triacsin, respectively. [32,60,68] Meanwhile, alanine accepting HSs in clade XI are connected to L-DABA-N^4-hydroxylase (green strip), therefore, the product of clade XI HS can be predicted to be DABA-Ala. Indeed, Apy9, a representative HS in clade XI gives DABA-Ala as a biosynthetic intermediate of actinopyridazinone.[37] In this way, the combinations of building blocks of hydrazine were predicted throughout the hydrazine biosynthetic pathways in nature, and several HS enzymes were chosen based on predicted novelty of hydrazine products and subjected to experimental investigation. In this study, I selected four putative HS systems for characterization: Kit1/2 (clade IV), Cor1/2 (clade II), Col1/2 (clade II), and Bac1/2 (clade VI).

The characterization was conducted using *E. coli* bioconversion experiment. HS synthetic gene was transformed to *E. coli*, then culture broths were fed with synthetic hydroxylamine, which was predicted based on phylogenetic analysis in Figure 3.14. For example, when the HS of interest was coupled with lysine-N^6-hydroxylase, synthetic N^6-OH lysine was fed to culture broth. After overnight incubation, supernatant was derivatized using dansyl chloride.

Kit1/kit2, a Putative Discrete HS in Clade IV.

First, Kit1/2 hydrazine synthetase from *Kitasatospora gansuensis* was investigated. In *kit* pathway, HS was encoded in two genes; cupin domain as *kit1*, MetRS domain as *kit2*. Therefore, *kit1* and *kit2* genes were co-expressed in *E. coli* host for the functional characterization. From previous experiment, Kit3 (*N*-hydroxylase) was known to accept L-ornithine to generate N^5-OH L-ornithine. Therefore, *E. coli* expressing Kit1 and Kit2 (*E. coli*::*kit1*/*kit2*) was cultured in M9 media and fed with N^5-OH L-ornithine. The supernatant was derivatized by dancyl-Cl and analyzed by LC-MS. As a result, a new peak was generated only in the culture fed with N^5-OH L-ornithine (Figure 3.15). Further MS/MS analysis revealed that this peak derived from L-ornithine and L-tyrosine based hydrazine (Figure S23), which is a new amino acid combination for hydrazine precursor. Further analysis revealed that L-tyrosine feeding increases the production of Orn-Tyr (Figure S21).

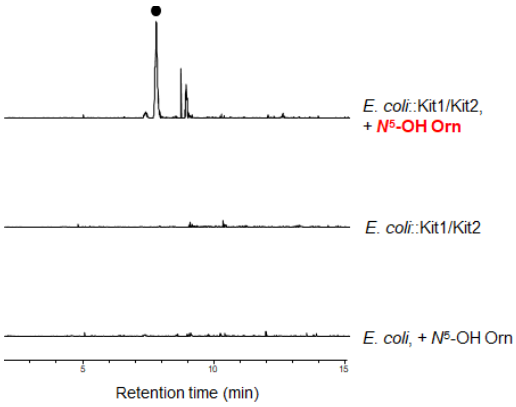


Figure 3.15 Biotransformation of N^5-OH L-ornithine by *E. coli*::*kit1/kit2*. Extracted ion chromatograms of *m/z* 545.2 are shown. The peak corresponding to Orn-Tyr was highlighted.

Bac1/Bac2, a Putative Discrete HS in Clade VI.

Another enzyme Bac1/Bac2 was also tested for the bioconversion experiment. Previously, we tested Bac3 (*N*-hydroxylase) from *Bacillus cereus* and this enzyme has specificity to N^4-OH L-DABA. Therefore, we fed *E. coli::bac1/bac2* with N^4-OH L-DABA. As a result, a new compound was generated (Figure 3.15). MS/MS analysis revealed that this compound was L-DABA and L-glutamic acid-based hydrazine (i.e. DABA-Glu), which is a new combination for hydrazine precursor (Figure S36). Feeding with glutamic acid decrease the generation of this product (Figure S22). With this study, Bac1/Bac2 is the first characterized hydrazine synthetase from Firmicutes.

Cor1/Cor2, a Putative Discrete HS in Clade II.

Hydrazine synthetase from *Corallococcus ilansteffanensis*, named as Cor1/Cor2, was also characterized in this work. As Cor1/Cor2 was coupled with DABA-N^4-hydroxylase (named *cor3* in Figure 3.4), N^4-OH DABA was fed to *E. coli*::*cor1/cor2*. As a result, a new peak corresponding to DABA and L-alanine based hydrazine (DABA-Ala) was generated (Figure 3.16).

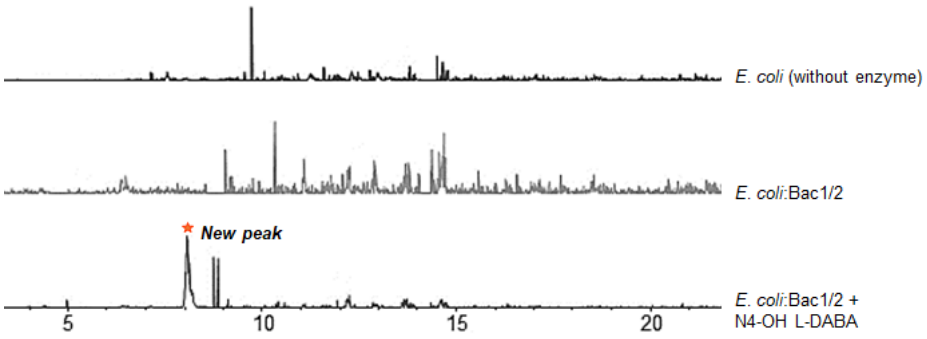


Figure 3.15 Biotransformation of N^4-OH L-DABA by *E. coli*::*bac1/bac2*. Extracted ion chromatograms of *m/z* 497.2 are shown. The peak corresponding to DABA-Glu was highlighted.

Col1/Col2, a Putative Discrete HS in Clade II.

The last enzyme, Col1/Col2 from *Colwellia* sp. was transformed to *E. coli* BL21 (DE3). As *col1/col2* were also coupled with DABA-N^4-hydroxylase (named *col3* in

Figure.3.4), N^4-OH DABA was fed to *E. coli*:: Col1/Col2. After derivatization with dansyl-Cl and injection to UPLC system, a new peak generated (Figure 3.17), and MS/MS analysis (Figure S28) confirmed the presence of DABA and L-serine-based hydrazine, which represents a new hydrazine with a new combination of building block.

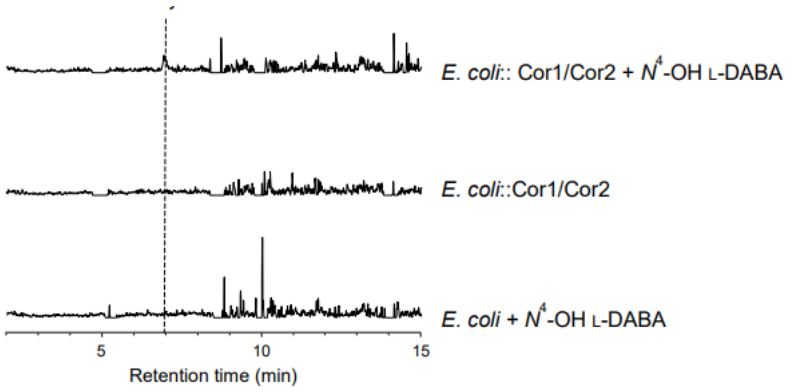


Figure 3.16 Biotransformation of N^4-OH DABA by *E. coli*:: *cor1/cor2* . Extracted ion chromatograms of *m/z* 439.2 are shown. The peak corresponding to DABA-Ala was highlighted.

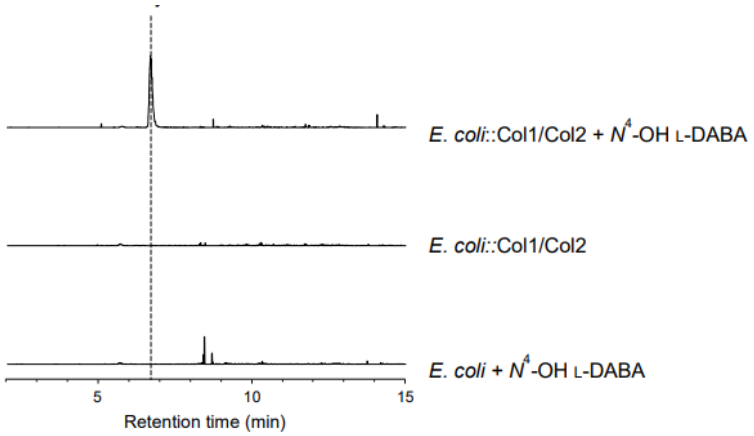


Figure 3.17 Biotransformation of N^4-OH DABA by *E. coli:: col1/col2* . Extracted ion chromatograms of *m/z* 455.2 are shown. The peak corresponding to DABA-Ser was highlighted.

Through this phylogeny-guided enzyme-mining approach, I identified three new hydrazines: Orn-Tyr, DABA-Glu, DABA-Ser. Considering that seven hydrazines were known before, this study substatially expanded the diversity of hydrazines derived from hydrazine synthetase-mediated pathways. Nevertheless, only three hydrazines were linked to

their final products (i.e. Lys-Gly for s56-p1, azaserine, triacsin; Lys-Glu for formycin, pyrazomycin; DABA-Ala for actinopyridazinone). Since the HSs with links to their final products are limited to only a minor fraction (HS groups X, XI, and XIII), the diversity of HS-mediated N–N bondforming pathways remains largely unexplored. This study paved the way for further phylogeny-guided enzyme-mining to pursue novel bacterial hydrazines as well as for future genome-mining efforts to identify the N–N bond-containing natural products biosynthesized by unexplored groups of HSs.

Chapter 4 Conclusion and Future Perspectives

4.1 General Conclusion

Genome mining approach combined with reactivity-based screening led to the identification of the new producer of azodyrecin, *Streptomyces* sp. A1C6. Furthermore, azodyrecin B and *trans*- azodyrecin B had cytotoxic activity against SKOV3, MESO1, Jurkat cancer cell lines and promising for future exploration as drug lead compound candidate. The importance of the double bond adjacent to azoxy moiety in azodyrecin's cytotoxicity suggest that it may act as a Michael acceptor that covalently bound to its molecular target. This result sets the stage for the investigation of the mode of action of azodyrecin.

Genome mining approach combined with phylogenetic analysis led to the identification of a new sub-type of N–N bond forming enzymes, especially NMO, and hydrazine synthetase from untapped source. The characterization of enzymes in this study is also revealed the diverse and unique substrate specificity of each enzyme. Understanding the substrate specificity of the enzymes will contribute to biocatalytic production of hydrazine molecules in future.

4.2 Future Perspectives

Some groups of N–N containing compounds, such as azoxy compound are still underexplored. However, looking at other known nitrogen containing compounds, these groups have a bright future for the application as drug lead compound. Therefore, using combination of genome mining and chemical screening as tools, we can efficiently explore azoxy compounds, not only from bacteria but also from other sources such as fungi and plants. As mentioned in the introduction part of this thesis, the first azoxy compound was discovered from cycad plants, so it is possible that we could find other types of N–N containing compound in Cycadaceae or other Gymnspermae family. As Indonesia has big diversity of organisms, there are a lot of unexplored native cycad plants. Currently, cycad plants are just ornamenting plant in Indonesian culture, but if we could explore the potency of this plant, we could add value for this plant as source of active compounds.

Exploration of N–N bond forming enzymes is the first important step to understanding the full pathway of N–N bond formation in nature. Therefore, genome mining tools could help us to find novel enzymes that might have good prospects in the future. Also, understanding the substrate characteristics of the enzyme might become the first gate to explore and engineer the enzyme, thus could maximize the future use of the enzymes as biocatalyst. For some enzymes that have promiscuous substrate recognition, it is very interesting for further research such as enzymatic engineering.

Furthermore, it is also interesting to characterize the activity of other enzymes in the Bac cluster, Kit cluster, NMO, and HS, to uncover the full pathway for formation of N–N bond-containing natural products in bacteria and identify the final products from those pathways.

Supplementary Information

Supplementary method

Media preparation

M9 media composition:

Solution A

Ingredients	Amount
Na_2HPO_4	6 gr
KH_2PO_4	3 gr
NaCl	0.5 gr
NH_4Cl	1 gr
H_2O	ad to 500 mL

Solution B

Ingredients	Amount
1 M $MgSO_4$	1 mL
2 M glucose	5.6 mL
1% Vitamin B1	1 mL
1M $CaCl_2$	0.1 mL

M9 media was prepared by mixing 2 mL solution A and 30.8 µL solution B.

2xYT media composition

Trypton 1.6%

Yeast extract 1%

NaCl 0.5%

Distilled water to 100 mL

LB agar composition

Tryptone 1%

Yeast extract 0.5%

NaCl 0.5%

Agar 1.5%

MQ water to 100 mL

Plasmid extraction method

DNA was extracted using Fast Gene™ Plasmid Mini Kit method. About 5 mL bacterial culture was centrifuged at 3500 rpm for 10 min, and then the supernatant was removed. Suspension buffer, lysis buffer and neutralization buffer were added to the pellet, then centrifuged at 15,000 rpm for 10 min. Lysate was transferred to plasmid mini spin tube, and spinned down at 15,000 rpm 1 min. First wash buffer was added, then re-spin down and second wash buffer (contains ethanol) was added. Tube was spinned down at 15000 rpm 1 min or until ethanol evaporated and DNA was eluted using elution buffer. Quality of DNA was checked using Nanodrop spectrophotometer.

Electrophoresis

1% Agarose S gel was prepared in 1x TAE buffer, then put in microwave for about 30 seconds. Ethidium bromide was added to the gel solution, then gel solution was poured to the gel template and comb was putted. Solid gel was placed into the gel box and 1xTAE buffer was poured. Loading buffer was added to the samples, then samples were loaded into each well and electrophoresis was run until the band separated.

Gel extraction and ethanol precipitation method

The tube containing gel was put in the heat block, QG buffer (Tris-Cl Buffer pH 6.6 and Guanidine thiocyanate) was added and heat block set at 55 °C for 5 min. Then, liquid gel was transferred to Econospin® column and spin down at 15,000 rpm 1 min. The column was washed using wash buffer (contains ethanol), spin down at 15,000 rpm for 2 mins. Then elution was done by adding MQ water to the column and spin down at 15,000 rpm for 1 min. Ethanol precipitation was done by adding 0.1x volume of CH_3COONa 3 M and 2.5x volume of ethanol 99.5%. Mix and centrifuged for 10 min at 15,000 rpm. The liquid was removed and placed on heat block at 55 °C for 1 min.

Transformation to *E. coli*

Transformation to *E. coli* was initiated by mixing the ligation product or DNA insert with *E. coli* competent cell aliquot and let on ice for about 20 min. The heat block was set up at 42 °C, the tube containing *E. coli* and insert DNA mixture was put in the heat block for 30 seconds, and transferred to the ice for around 3 min. The LB liquid media was added into the transformation mixture followed with incubation at 37 °C for about 20 mins. The transformation mixture was then spread on the specific solid agar media containing appropriate antibiotic(s).

Protein purification and SDS PAGE

Culture preparation

1% bacterial liquid culture were transferred to 200 mL 2xYT media and was incubated at 37 °C for 3 hours. 0.1 mM IPTG was added to media and incubation for overnight at 16 °C. Liquid culture was centrifuged at 4000 rpm for 10 mins, then the supernatant was removed.

Purification step

Buffer A (lysis buffer), buffer B (wash buffer), buffer C (elution buffer) were prepared. Buffer A contains 20 mM tris HCl pH 8 and 150 mM NaCl, buffer B contains 20 mM tris HCl pH 8, 150 mM NaCl, and 20 mM imidazole, while buffer C contains 20 mM tris HCl pH 8, 150 mM NaCl, and 500 mM imidazole. Buffer A was added to the bacterial pellet, then was sonicated, and centrifuged at 18000 xg for 15 minutes. Supernatant was collected and ready for purification process.

Prepare Ni-NTA column at 4°C, equilibrate the column using buffer C, buffer B, and buffer A. Sample was loaded to the column, then eluted using buffer B to wash non target protein. In the last step, target protein was eluted by buffer C.

SDS PAGE

Gel composition (For 15% gel)

	Separating gel	Stacking gel
MQ	3.55 mL	2.1 mL
1,5 M tris-HCl (pH 8.8)	2.5 mL	-
0,5 M tris-HCl (pH 6.8)	-	760 μL
40% acrylamide	3.75 mL	375 μL
10% SDS	100 μL	30 μL
10% APS	100 μL	30 μL
TEMED	8 μL	4 μL

Purified protein was denaturized at 98 °C for 5 minutes, then loading buffer was added. Protein sample was loaded into each well, then SDS PAGE was turned on and the sample will be separated. For visualization, gel was macerated with CBB stain blue and was heated at microwave for 20 seconds. Gel was washed using MQ water several times and the band seen will be seen.

Figure S1. LC MS analysis of Azodyrecin standard

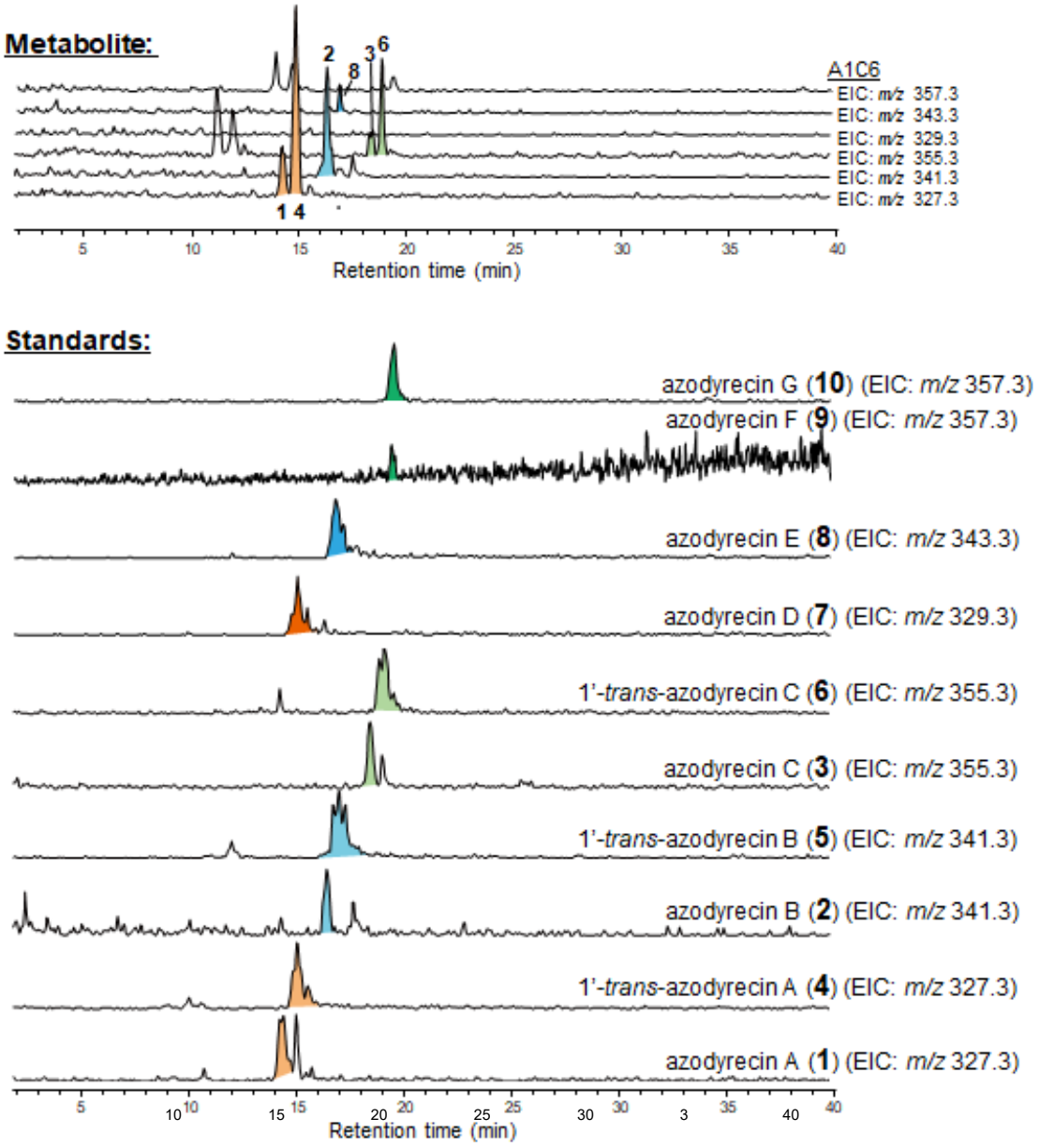
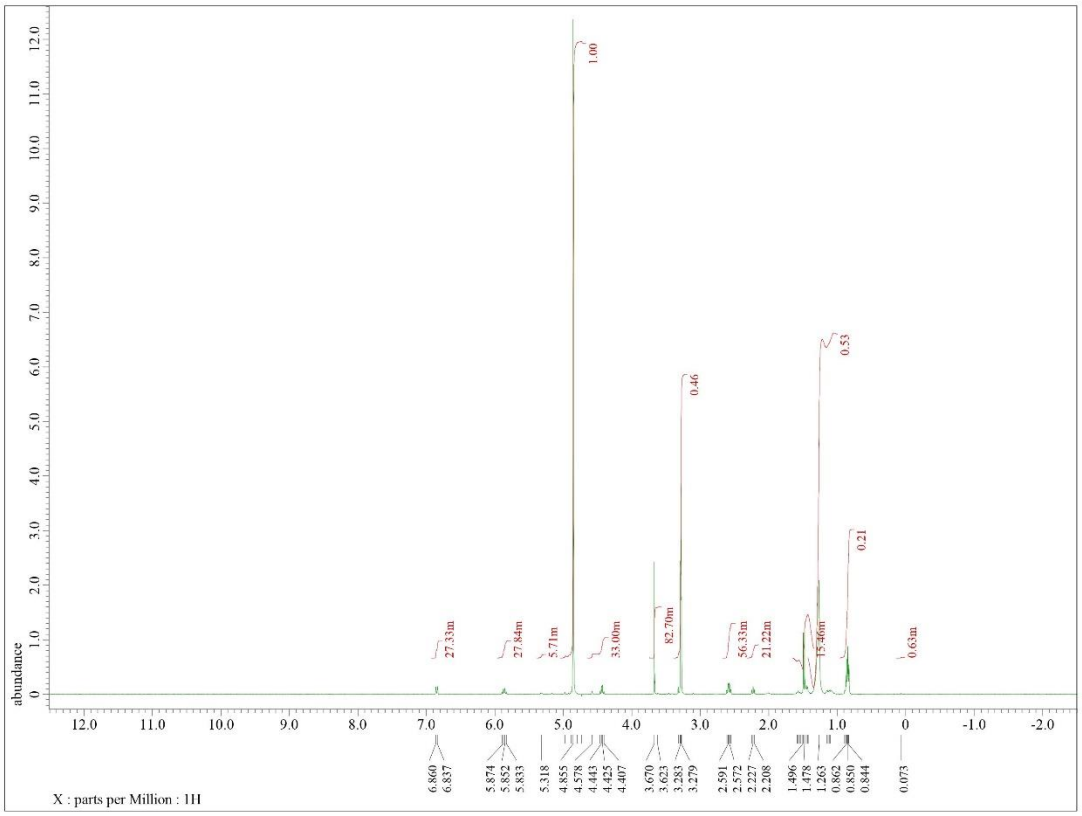


Figure S2. H-NMR of Isolate 1



olefinic olefinic =CH- O-CH_3 CH_3 CH_3

Table S1. List of Plasmids used in this study.

Plasmid	Description
bac3-pColdII	Expression plasmid for *bac3*. DNA fragment was inserted into NdeI/HindIII site.
bac1/2-pRSF duet	Co-expression vector for *bac1/2*. DNA fragments coding for cupin and MetRS-like enzyme were inserted into EcoRI/HindIII site and NdeI/XhoI site, respectively.
kit1/2-pRSF duet	Co-expression vector for *kit1/2*. DNA fragments coding for cupin and MetRS-like enzyme were inserted into EcoRI/HindIII site and NdeI/XhoI site, respectively.
kit3-pColdII	Expression plasmid for *kit3*. DNA fragment was inserted into NdeI/HindIII site.
nmo8-4-pColdII	Expression plasmid for *nmo8-4*. DNA fragment was inserted into NdeI/XhoI site.
nmo8-6-pColdII	Expression plasmid for *nmo8-6*. DNA fragment was inserted into NdeI/XhoI site.
nmo8-1-pColdII	Expression plasmid for *nmo8-1*. DNA fragment was inserted into NdeI/XhoI site.
nmo8-2-pColdII	Expression plasmid for *nmo8-2*. DNA fragment was inserted into NdeI/XhoI site.
nmo1-pColdII	Expression plasmid for *nmo1*. DNA fragment was inserted into NdeI/XhoI site.
hs1-pRSFduet	Co-expression vector for *hs1*. DNA fragments coding for cupin and MetRS-like enzyme were inserted into EcoRI/HindIII site and NdeI/XhoI site, respectively.
cor1/cor2-pRSF	Co-expression vector for *cor1/cor2*. DNA fragments coding for cupin and MetRS-like enzyme were inserted into EcoRI/HindIII site and NdeI/XhoI site, respectively.

col1/col2-pRSF	Duet expression vector for *col1/col2*. DNA fragments coding for cupin and MetRS-like enzyme were inserted into EcoRI/HindIII site and NdeI/XhoI site, respectively.

Table S2. List of Strain used in this study

Strain	Description
E. coli DH5α	Host for cloning
E.coli BL21 (DE3)	Host for protein expression
E.coli: *bac3*-pColdII	E.coli BL21 (DE3) contains plasmid *bac3*-pColdII
E.coli: *bac1/2*-pRSF duet	E.coli BL21 (DE3) contains plasmid *bac1/2*-pRSF duet
E.coli: *kit1/2*-pRSF duet	E.coli BL21 (DE3) contains plasmid *kit1/2*-pRSF duet
E.coli: *kit3*-pColdII	E.coli BL21 (DE3) contains plasmid *kit3*-pColdII
E.coli: *nmo8-4*-pColdII	E.coli BL21 (DE3) contains plasmid *nmo8-4*-pColdII
E.coli: *nmo8-6*-pColdII	E.coli BL21 (DE3) contains plasmid *nmo8-6*-pColdII
E.coli: *nmo8-1*-pColdII	E.coli BL21 (DE3) contains plasmid *nmo8-1*-pColdII
E.coli: *nmo8-2*-pColdII	E.coli BL21 (DE3) contains plasmid *nmo8-2*-pColdII
E.coli: *nmo1*-pColdII	E.coli BL21 (DE3) contains plasmid *nmo1*-pColdII
E.coli: *hs1*-pRSFduet	E.coli BL21 (DE3) contains plasmid *hs1*-pRSFduet
E.coli: *cor1/2*-pRSFduet	E.coli BL21 (DE3) contains plasmid *cor1/2*-pRSFduet
E.coli: *col1/2*-pRSFduet	E.coli BL21 (DE3) contains plasmid *col1/2*-pRSFduet

Table S3. Protein used in this study

Protein	GenBank ID	Amino azid sequence	Function
Bac1	WP_076855485.1	MGSSHHHHHHSQDPNSHVKTLDRENLKFEYDLYAQRNFP WKDVVDPPFGGAWCVLHPGQTSAPHNHDEKEVFIIVKGKG LARGDEKTREVTTGDVIYFTPFTEHCITNI GEEPLEFFSLWWDGQPQNNENEMEYQI	Cupin
Bac2	WP_076855484.1	MSKIKHHHHHHTSDKKKYLVTAAPPTSNGNLHIGHLAGPF LGADVFSRIQRMLENDVLYISYGDDYQDYVERKALETNKT PEDISTFYTNEMKETLRIAHMLPDHFPGSKNIPMHPQIMQEI FLDLYKNNQIQFMNVTTFYCEKCEKFEYEAYARGNCHYCG SSSDANYCESCGQPQEAGMLLNATCHTCKNELSLYIEDRAY FPLSKHKEILEKYYENKNWRPDAKELCKKILDNEIDAPVTR NSNWGIKIPLENREKHILDTWYGGPAGYISATMDWSSLQSD KDTWKEYWQSEDTKWIAFVGYDCTFSHGVFYPAMLKAHG KYNAPDTVVTNKFYGLDGEKISTSRNHAIWGNDIFNLVDS DAVRFYLSSTAPESTEMNFNLDDFQKFVNEDLAESWNKWL LSLFNEAAALGITSIESILKKPTTKLEKKAIKCLEQIQVYYDA ENFSLQKITSLILELAKEAMDSYWRFDRTVHESNLAMIKEN LSVVRMLAVITSPIMTEFSIKIMDILGEKNTEDRKIYLWTDL KNKTQILPKKPLEPFFSKVPQDVFE KLSSYKKEVVK	MetRS
Bac3	WP_076855482.1	MNHKVHHHHHHHMINCSEEIYDLIGIGFGPAGIALGVAIDDH NEGLSEEKSIKALFLEKANDSTWQGNMLLPRAHIQHHFFR DLVTPYNPRSKYTFANYLKEKGRLFQFGLLGLTPGRIEWND YISWVAQQLNQYVIYQQNVIEINPIKDAEGNIDLVEVITTDV VSNEIKKFKAKNITVNTGRTPYIPEQYKPLMGEKVFHPHFF KKKIENFKAEDKPNFAVIGSGQNAIEVILSISNKFPEADVYSI NRNLGFKTYDLGQFSNEVYHPEFVDYFYGLTKEERQKLFD EQMKATNYSAVGPHVSTDLYWKVYEQSIEGINKIHVMRCQ EVLDVKDVENGYELHLKDNYTKEELSINVDGIVVCTGFTE EKIPRPLDSILNYLELHEDGDIDISRNYEIATTNQFNVGVYA NGITERTHGISDATSFSMMALKAKHIFDHVQESKKANQ ELLRV	*N*-hydroxylase
Kit1	WP_184911088.1	MGSSHHHHHHSQDPNSEIRRLDRANLTPAYGILGERLQPWD ALNAPFEGAWCVISGGGESTPHSHHEYEIFIAMAGRSTLVV DGTDHEFIAGDIVRLPPGCTHKVVNNHQEDFEYYGIWWDT DMSDAFLARHQKEKA	Cupin

Kit2	WP_184911086.1	MSKIKHHHHHHTSMSKRVAVISPAPTANGDLHLGHLAGPFL AADVYTRYQRATGRETVFGTGAQDTSTFVVTTARRLGTSP EALVADSTAKIEATLAAMGIEVDGFSRDEERFTKLVLDFVG RLHSAGKVVLKPARFPYLPSTGEYLVDGFVKGGCPVCLTEG CAGLCESCGHPSAAGDLIDPRSSEHPDEPVEIREVPVLVLPL EEYRAQLKEYFASSGAALRPHMAQAVQEMLDRPLPDLPIT YPISWGIPAPFPEVSGQTINPNAETMAWSIYTTALGAEKRGE VLAADDQLWWPESETEVVYFLGFDNTFPFAIAGAAMLLAL GDRYALPARFVTNEFYELEHQKFSTSRGHLVWGQELADRL PRDVGRFHLAATSPENQRTNFGWGALEAVTGARLIQPWNR VAAKVEQWAGRTLPVSAASRFAADRITQRFSAAYQMPAFSL GRAAETLAEQLGRLDRWEVTEELAGDFCHQVEAVLRGAA PILIDLAAAALDDTAVRDRDETTEITPKILPRLTGGAV	MetRS
Kit3	WP_184911090.1	MTNHEVEVLAIGAGPSNLALAVALEESGSTGLANNTLILEQ YPDVKWQRNLLLPFARSQVSFVKDLVTLRNPQSRFSFLSFL HAKGRLDEFVNLGTFNPFRRELSEYLQWVANNLDHVGIRY NAKGQSVAPRRADDGTVIGWRVTLTNGDTVDTRDLVVGT GRDANVPKEFAHLPADRVIHSTQYSTRIAQYPIDRPVRVVVI GAAQSAAEMFMAVHKDLPLSQPTLIHRSIGLQNYQTSKFV NELFFPSFVNEFHDSPAESRKQILEEIRLTNYAGLAAPFLDET YSMLYEQKLSGSQRSAVRSMTEVLEAREEDGEVVLELRDR RTGKTDTVRCDLVLLGTGYDQRMPAMVRHLAGQVGIEDIA VNRRYRVDLGESTRAGLYLQGFNEATHGISDSLLSVLAHRS HEITTDLLDRRTVSAERSL	*N*-Hydroxylase
HS1 (MetRS)	WP_013209038.1	MNARKTYVLIPVMPTPNGPLHLGHVAGPFLKMDMLARHL RRNGETVALVSATDPYETHVLPRADEQNKPVEAICAENHR AIHRCLQALDIRYDAFIDPLASPYRARLSGITREVLDDLNAQ GLLHARNEPVHVSRRTGRMIVGSRIVGTCPCCGVEMGGYH CEGCGMEVWPRDLIAPRAEPADDTVEVEERASVFLDADIPA LKRRMLEARVPADVRRIAERFMHATGGAVRLSNPGEWGVA WHDGQATVPSVVFTYTALFMLSVLCGEVARERLALDHNPF DQRSDAFTVTSFGFDNTVPFCVGVETLAQHSRRYRGFDRY LTNFFYTLDGRKFSTSRQHCIWADQAVQELGVTSDVLRYFL AKTSPEGGPSDFSRDGFDAFRRAVEPRLAQMNAAVESGIGG DAADPYADTLARLAADMEAAFDVDHFSLRAATRCIDQWL ALEWQPAHAAAYVRGFCMLAYPVMPELALTLWGKLGHEG LPRYLA VTELETAE	MetRS

HS1 (Cupin)	WP_013209037.1	MGSSHHHHHHSQDPNSTRSPIRLAFQAPQVSETGLSVEVID FGFLRAAATFEASRFEVAVGGFSPPEKHPEREIWMIAAGSGR LDYRGEVTEVRSGDCLTFDPNVEHSVHNTGTEPMRIFSVW WTERAP	Cupin
Cor2 (MetRS)	WP_120641657.1	MSRMTLIIAPPPTPNGDLHVGHMSGPYLAADIYKRYIGQHG RTARYTVSTDDNQSYVDTTARRLKTDVPSLLAKSRAEVRD SLLTYSIGVDHFGEQDAAYPGFVTAFFEKLLAAGFVENADV EVFYDTKRHTYPVEAYISGRCPTCLDSTCGGICEACGHPNA CTDLLEVNTPDLEVRREARLIFRLEPFRAELEEHLLRVMKT HRPALKALIASMLASPLAPFVLSYKTGRGISAGFAGLPDQQ LNVWAEMYPGHFYFLTKAAGKVQGDDDYVQAMGFDNSY FYVFVHVALALAGRRCGVDWPLPKAFITNQFYNLDSAKFS TSKGHLVWARDLAREYNTDVIRLYLALHGPEFQEANFSLG AFKEGAAELTGRVNGLVAAFNTARKTAREGVGDGGMLAR LNAVLSTELPAGDYAASTLARRALNGIALMKDSLDKGYLG LVPFVPALLALALEPFCPLYAQAIRRQHRLTAETWNDLAPTS EDQALPELLV KK	MetRS
Cor1 (Cupin)	WP_120641656.1	MGSSHHHHHHSQDPNSHIQRKSDATVKFEYGTDLRRIYPW KGIQDPLYWGSAIASVRPGECTTPHSHDEFETFIVLSGKGE MQIEAEREVMTAGDIVFIPKDHHHQFQNLSNEEPLVFISIYW DSPEAREQILGHLLQKKAG	Cupin
Col2 (MetRS)	MBL4898055.1	MGKSPTPPDFIVTLPPPTPNGGLHVGHMAGPFLAADIFTKA AQASGKTVHVLSYSDSNQSYVRATAEKQNRDPKELASFWT QDILETLDIYGCQVDNYFEPHPASDGFVRDTLLRLYEQGLL KKKSYPFFYLPEQGVFLDEAGASGHCPRCLDDCKCGICEA CGFPTQADTLINPRATNDPSLEIELRHVEVLVLELEAWRGSL RRFHDFSDAIRPKYRWLVDDVLAGPLPDFPITVPGSWGIAA DHPDLPNQVINAWPELVLDFVYGYQKVAPQGGPTPKIVNFF GYDNSYFYAIVHVALLHAAELQRFLPHATMVNEFYNLENA KFSTSRGHLIWARDMAQSYPPDLIRFYTALSSPGFEQGNFSE TDMREVIQRRLLAPFQSIAAAAPVLDEAIEPSATVRQIMHGL STRVAESYQLHRFNLRHAAEDALKGLNAIQTCLSKGDLSRG ELVLLLDCWGQSIAPLMPAVATSLRHSLVMDIIPEALPFKPV KSQ APRRVEEANA	MetRS

Col1 (Cupin)	MBL4898056.1	MFSLNDRMSFDIETLVRRGSDAETSREYNADLRRLLPGPGY GNTKRALTEFGAILVTVRPQEEVDAHHHDEEECFIVVSGQG DLQVEDQTTTLHPGDVVYMPRFWMHQLRNPNDTAFTFVD LYWDDKGRSFDQYVQSQRAEAL	Cupin
Nmo8-4	WP 121229554.1	MGSSHHHHHHSSGLVPRGSHMTETRVGLLAIGAGPANLAL AAAIEESGRPDLAEQTMLLEQGPDVTWQRDLLMPWARSQ VSFLKDLVTLRNPRSRFSFLNFLHDQGRLDEFVNLATFNPFR WEISDYLQWVARSLERVRIRFDSRAASIEPTRGGDGTITGW AVTLTDGHVIRCRDLVLGGGRDPHVPEVFGDLPPERIIHSTR YRTRVAALPADAPLRAVVVGGAQSAAEMFMALHDNLPNST CTMVVRSVGPQNYQTSKFVNELFFPSFVDRFYDSPPEVRRQ LLAEMHLTNYAGMAPPFLEHMYTALYQERGLGTPRSQVRT LTEVVGARVEDDEVVLDLRDRTTGKVEPLRCDVVLLGTGF AQRMPAMVSDLAERIGLSEITVNRRYRLDLGEPAWGAVYL QGVNEATHGIADSLISVLAHRSQDILDDL VARQGTTELRSA	*N*-hydroxylase
Nmo8-6	WP 189212134.1	MGSSHHHHHHSSGLVPRGSHMAESRAPDRDVELLAIGAGP ANLALAAAVDEAGPADLAANSLVIEQADRVVWQRGMLLP GAQSQVSFLKDLATLRDPRSRFSFLNYLHSVGRLNEFVNLG TFQPYRTEVSDYFQWAADSMANVSVEYRRRCVGIDAERGP SGEPVAWVTRLADGSAIRSRYLVVGAGRDPYVPAEYATLPR ERVVHSTEYVQRIAELPRDRAYRVAVIGGAQSAAEMFDAVA TDLPVTTRTLVMRSIAPKTYENGKFSNEYFYPSAVDEFFTAR PEAREQILREMHTTNYSGLAPDLMDRLYRQRYVDRLGGAD RMRVATMTDVTAAYEDGGDVVLELTDRRTGRTEELRCDLV LLGTGFVRGMPALVRGLAGALGLPRVEVTRDYRLRVGPAA CYLQGVNEATHGISDSLLSVMSVRAADILADITADRAGHGN RRTTEKETSWRSADWTGIDWAPTTTTRSG*	*N*-hydroxylase
Nmo8-1	WP 022802429.1	MGSSHHHHHHSSGLVPRGSHMTIQTTELLALGAGPSNLAL AVALEELAPHLAKRCVVAEQHGDVRWHRGTLMPGTLSQV SFLKDLATQRNPRSRFTFLNYLREQNLLDAFINLGTFTPYRQ EISDYLQWVARQFDHVQVRYHARAEAITPDYDPAGRVRGW TTRFTDGQTLTSRHLVMGVGRDPNVPAVFRGVRTDRVVHS SAYLHQTRDLLGRPALRIAVVGGAQSAAELYRAALQDWPD ADVRMIMRSIGLVAYEGSRFTNELFYASFVDDFYSCEPGTRE VILAEMHRTNYGGVAPSLLDELYRLRYQRIITGQGDGDMLT MTDIVDARDSADGVLLTLKDKRTGAVQTLEVDLVLLGTGF SPQLPQLLAPLAQELGLDALDVSRHYRVQFGQFAGAGLYV	*N*-hydroxylase

		QGINEATHGIADSLLSVLAARSEEIVMDLL ALQDDPALPPARDLIGA	
Nmo8-2	HCZ39338.1	MGSSHHHHHHSSGLVPRGSHMHSVDLIMIGCGPSNLAVSV ALEEKNKVSNIKTSIILEKDNTVCWHRGMLFPEAQSQVSFL KDLVTQRDPTSRFSFLNFLHKTQRLDQFINLQTFNPFRKEIS NYLQWVADNLKKTEVVYGFKVIDITPEIDKHGKISEWRVTS ENGEEYRAKKIIYGGGRDLNIPEVFAGISEERVIHSVNYLSFL NRIDKSNIRNIAVIGGAQSSAEIYQSCLEQFPEARISMIMRSIG LVNYEGSQFTNTLFQNEYINTHFEKNDEIK AGVLKAMRTTNYAGVAPATLSGLYRFHYLQDMNGKKRAF MHTQCEILEAEELSDGVYIKWKDNISDKICSEKFGLVFLGT GFKNRTPALFEKLQKAMKLEKIKVSRKYRAELPYAEGASL HLQGVNEETHGIADSLLSVLALRSREIVDDLLVKQ	*N*-hydroxylase
Nmo1	WP 022802429.1	MGSSHHHHHHSSGLVPRGSHMYRKFYDVVGVGFGPAGIA LETAIRDMEDINNKSVYKRVFLEKNNDSAWMPEMLLPGTD IQHHFLRDFATPRNPRSQYTFSNYLYEKGRLFSFGLLGRPSR TEWSDYVQWVSKQIDDKAHYNEEVKSVKPIVVNDDIEGLE VHTVNNVSKEHNVFHTTNLILNSGRKPYIPSQFKDKVSDKV FHASKFRSSVQSINKDDNPTFTVIGSGQNSVEIILYLANKFPK SKINSIVRHTGFRLYELGHFTNEVFFPEFTNYFYSLNQNERD KLFEQIKHTNYSAVDDDVSEALYWKMYEDKIRGTEQIELH RCKEIVDVANTQKDYLLNIKDIYNKQEDSINTDYIILCTGFF EEEFPEVLQPMKKYFNYNKNNSLKVSENYKVDTKSNVKA NIYLNGLTEKTHGIGDAASFTMMATKAQRILDSMNNITSMK DGVLS	*N*-hydroxylase

Electrophoresis and SDS PAGE result

Figure S3. *bac3* digestion check and SDS page result

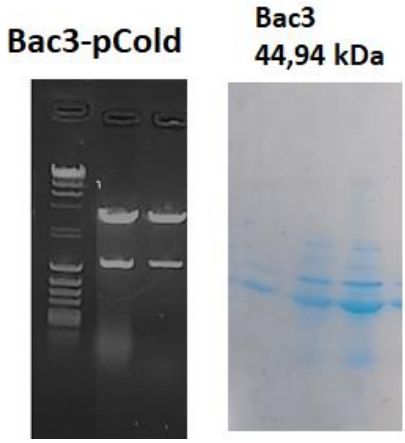
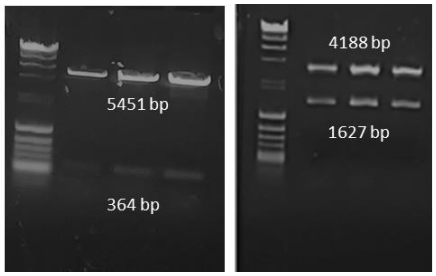


Figure S4. *bac1/2* digestion check



Digest with EcoRI/HindIII

Digest with NdeI/XhoI

Figure S5. *kit1,2,3* digestion check

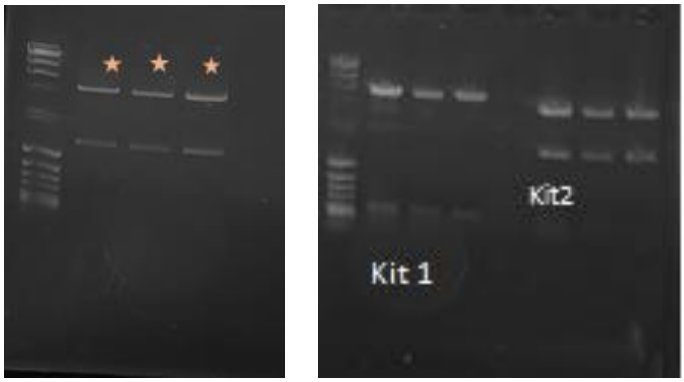
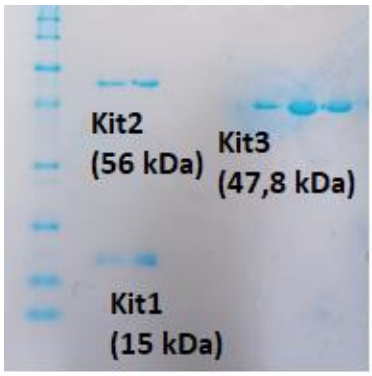


Figure S6. SDS PAGE Kit 1,2,3



SDS PAGE

Figure S7. Nmo8-4

Fig S8. Nmo8-6

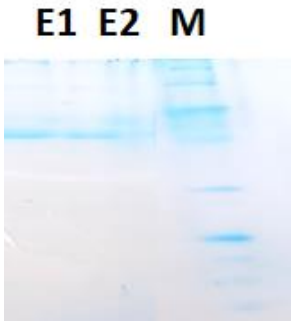


Figure S9. Nmo8-1

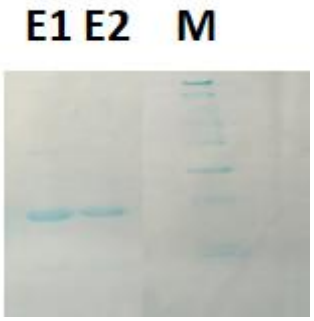


Figure S10. Nmo8-2

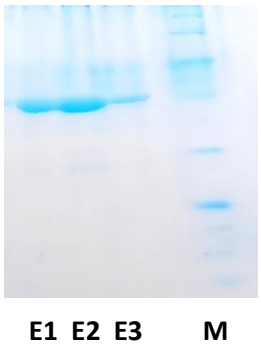


Figure S11. Nmo1

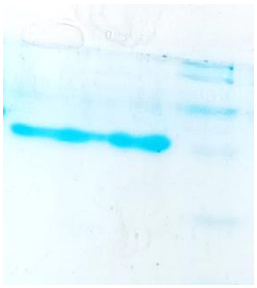


Figure S12. *hs1* digestion check

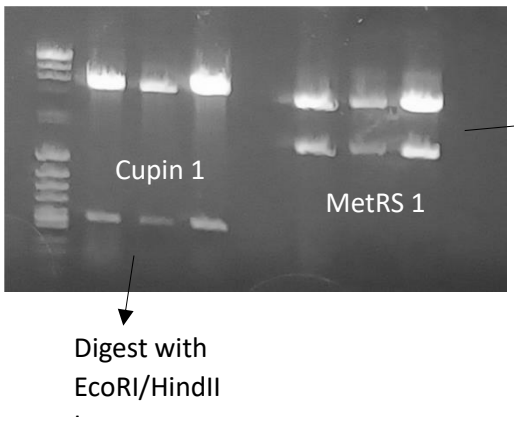


Figure S13. *cupin 1* (PCR with KOD one)

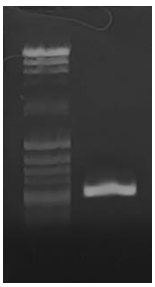


Figure S14 and S15. *cor1/2* construct

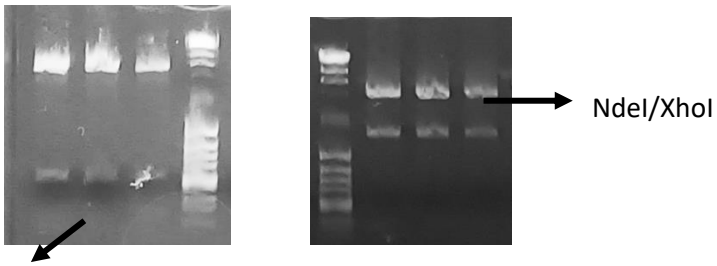


Figure S16. PCR *cor1* (KAPA HiFi)

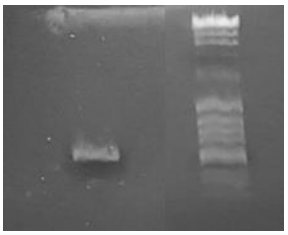


Figure S17. and S18. *col 1/col2* construct

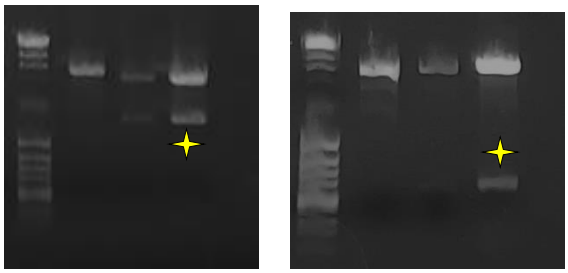


Figure S19. PCR *col1* (KAPA HiFi)

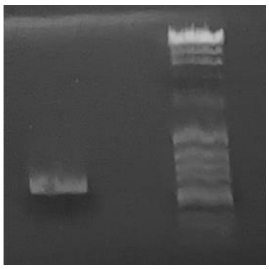


Figure S20. Genome mining Vlm O and VlmB

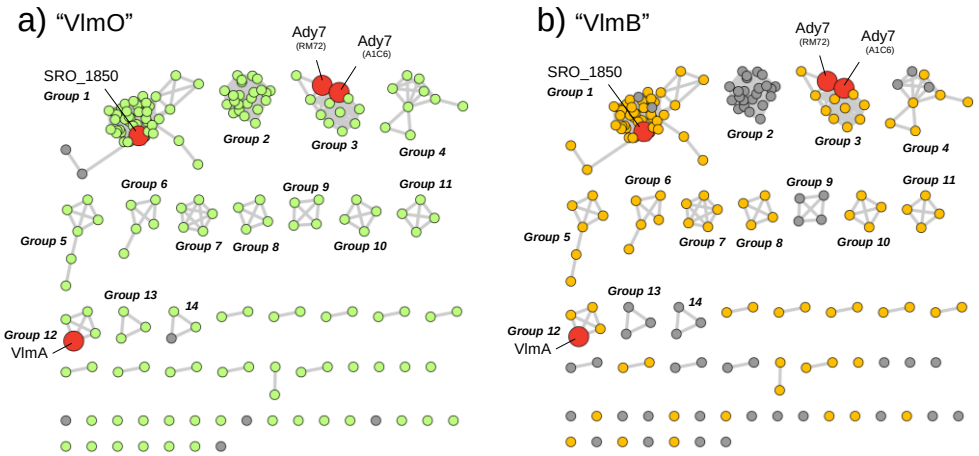
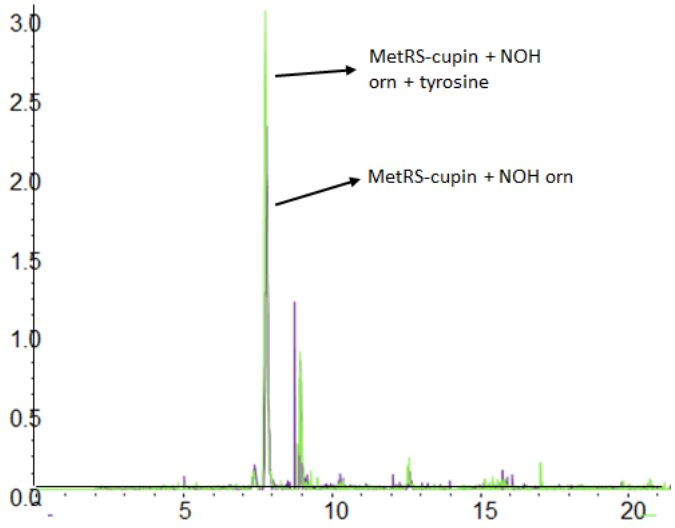
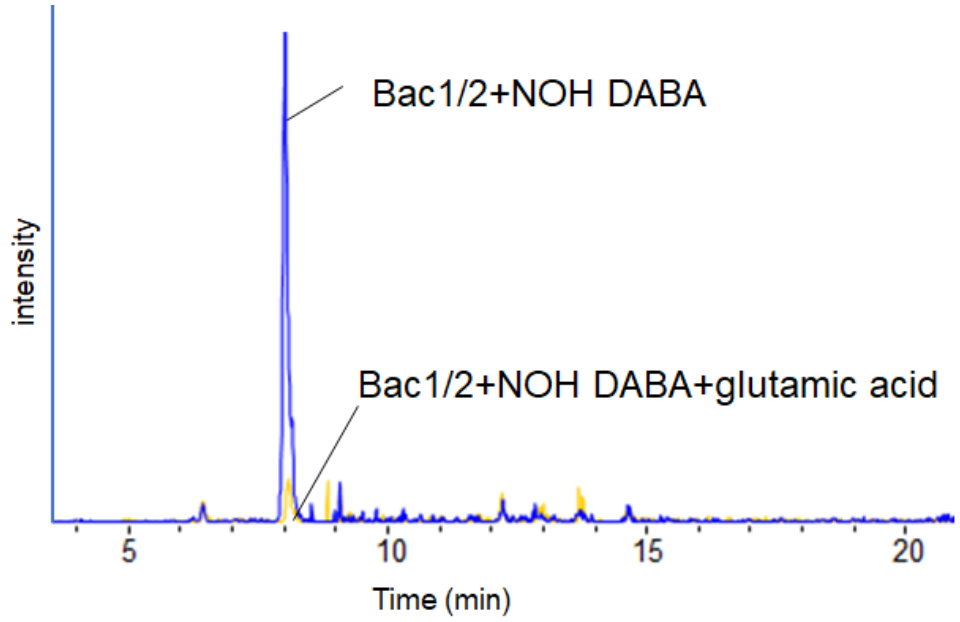


Figure S21. Bioconversion Kit1/2 with tyrosine feeding



FigS22. Feeding L-glutamic acid to *E.coli*: Bac1/2



MS-MS analysis

Figure S23. MS-MS of N^5-OH L-ornithine-tyrosine (Kit1/2)

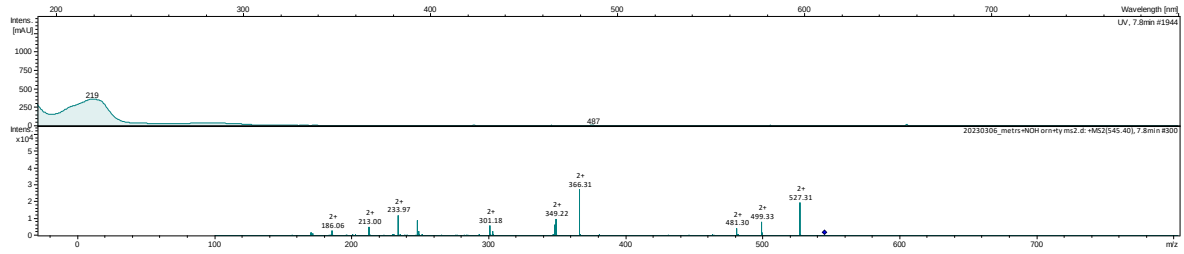


Figure S24. MS-MS of N^4- OH L-DABA-glutamic acid (Bac1/2)

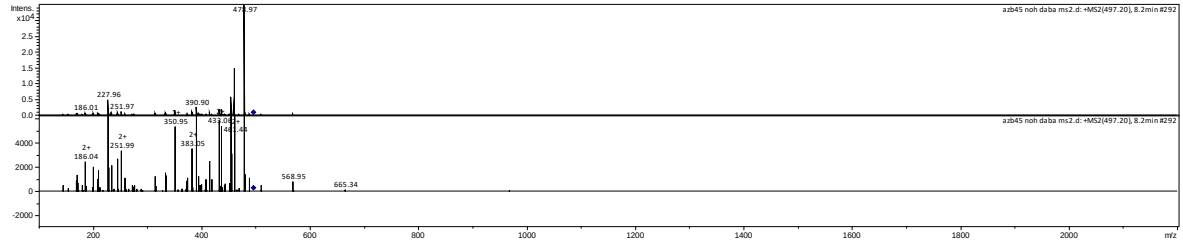
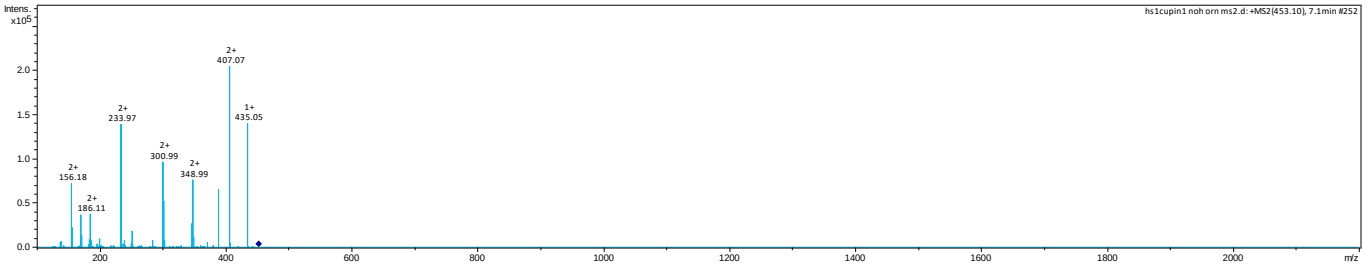


Figure S25. MS-MS of of N^5-OH L-ornithine -alanine (HS1)



9

Figure S26. MS-MS of N^4- OH L-DABA-alanine (Cor1/Cor2)

Figure S27. MS-MS of N^5-OH L-ornithine -alanine (Cor1/Cor2)

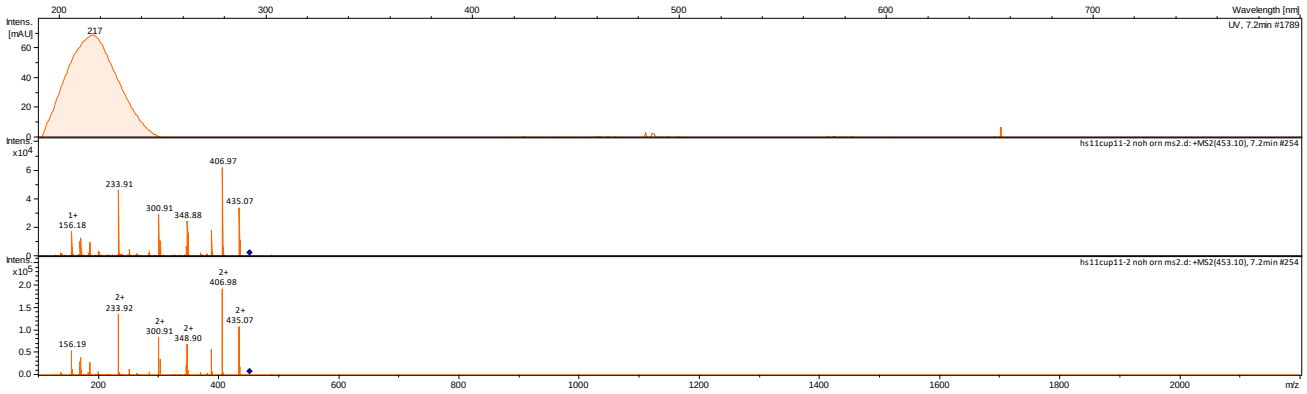


Figure S28. MS-MS of N^4-OH L-DABA -serine (Col1/Col2)

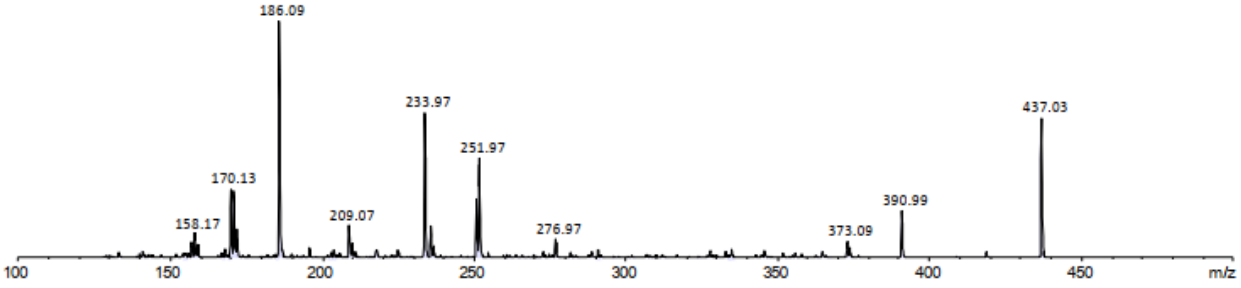


Figure S29. ***In vitro*** **assay NMO4**

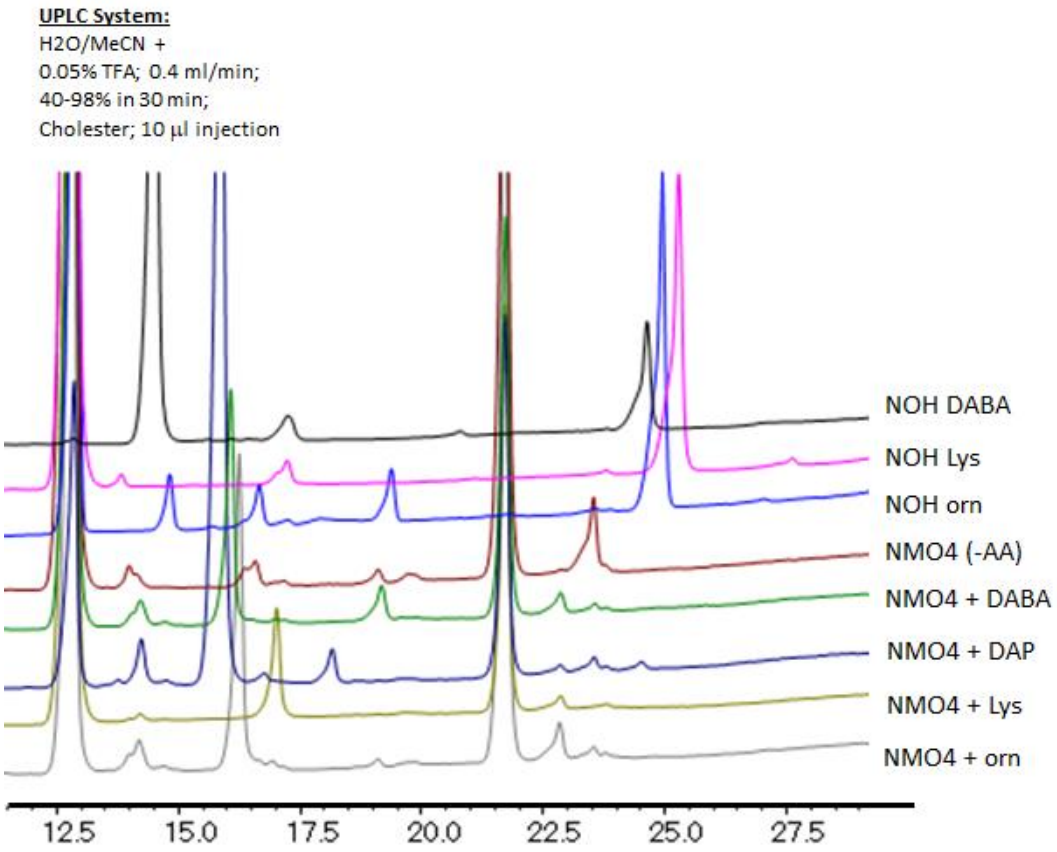


Figure S30. ***In vitro*** **assay NMO6**

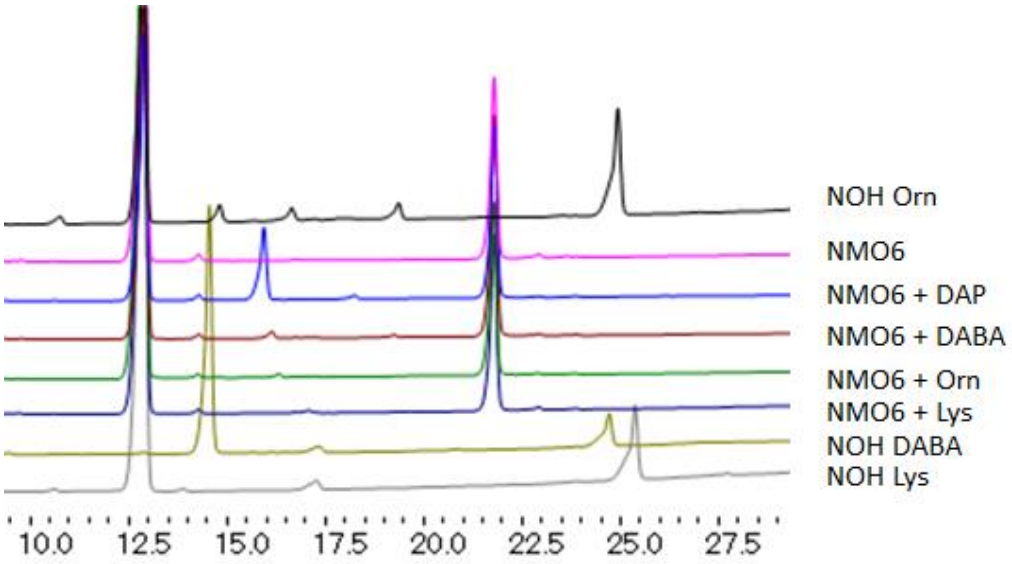


Figure S31. ***In vitro*** **assay NMO8**

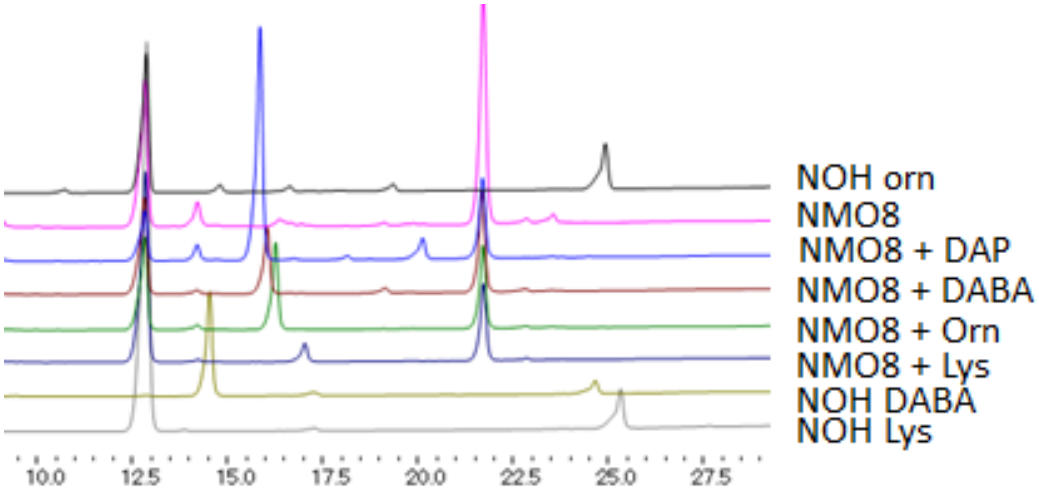


Figure S32. ***In vitro*** **assay NMO9**

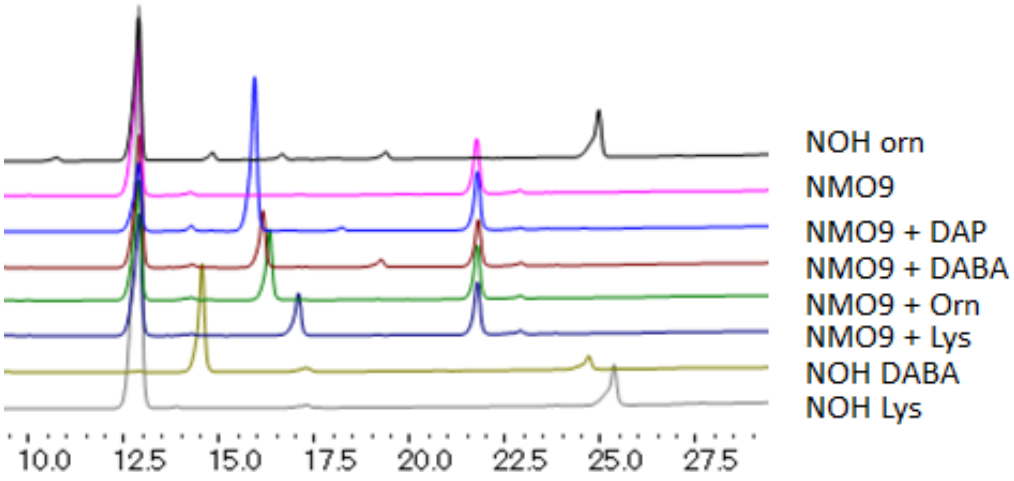
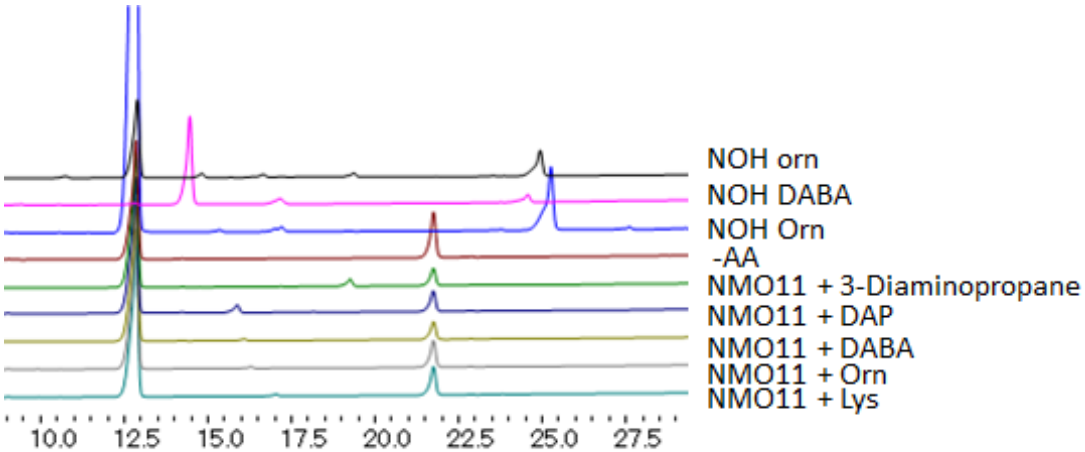


Figure S33. ***In vitro*** **assay NMO11**



REFERENCES

1. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

2. Reshi, Z. A., Ahmad, W., Lukatkin, A. S. & Javed, S. Bin. From Nature to Lab: A Review of Secondary Metabolite Biosynthetic Pathways, Environmental Influences, and In Vitro Approaches. *Metabolites* **13**, 895 (2023).

3. Patridge, E., Gareiss, P., Kinch, M. S. & Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today* **21**, 204–207 (2016).

4. Atanasov, A. G. *et al.* Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnol. Adv.* **33**, 1582–1614 (2015).

5. Matsumura, E. *et al.* Microbial production of novel sulphated alkaloids for drug discovery. *Sci. Rep.* **8**, 7980 (2018).

6. Borchardt, J. K. The beginnings of drug therapy: Ancient mesopotamian medicine. *Drug News Perspect.* **15**, 187 (2002).

7. Veeresham, C. Natural products derived from plants as a source of drugs. *J. Adv. Pharm. Technol. Res.* **3**, 200 (2012).

8. Mahdi, J. G., Mahdi, A. J., Mahdi, A. J. & Bowen, I. D. The historical analysis of aspirin discovery, its relation to the willow tree and antiproliferative and anticancer potential. *Cell Prolif.* **39**, 147–155 (2006).

9. Su, X.-Z. & Miller, L. H. The discovery of artemisinin and the Nobel Prize in Physiology or Medicine. *Sci. China Life Sci.* **58**, 1175–1179 (2015).

10. Gaynes, R. The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerg. Infect. Dis.* **23**, 849–853 (2017).

11. Rawlins, M. The disputed discovery of streptomycin. *Lancet* **380**, 207 (2012).

12. Bo, G. Giuseppe Brotzu and the discovery of cephalosporins. *Clin. Microbiol. Infect.* **6 Suppl 3**, 6–9 (2000).

13. Allen, N. From Vancomycin to Oritavancin: The Discovery and Development of a

Novel Lipoglycopeptide Antibiotic. *Antiinfect. Agents Med. Chem.* **9**, 23–47 (2010).

14. Dias, D. A., Urban, S. & Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, 303–336 (2012).

15. He, H.-Y., Niikura, H., Du, Y.-L. & Ryan, K. S. Synthetic and biosynthetic routes to nitrogen–nitrogen bonds. *Chem. Soc. Rev.* **51**, 2991–3046 (2022).

16. Wibowo, M. & Ding, L. Chemistry and Biology of Natural Azoxy Compounds. *J. Nat. Prod.* **83**, 3482–3491 (2020).

17. Blair, L. M. & Sperry, J. Natural Products Containing a Nitrogen–Nitrogen Bond. *J. Nat. Prod.* **76**, 794–812 (2013).

18. KARLSON, A. G. Specific inhibitory effect of elaiomycin in vitro upon Mycobacterium tuberculosis. *Antibiot. Chemother. Fortschritte. Adv. Progrès* **12**, 446–449 (1962).

19. HASKELL, T. H., RYDER, A. & BARTZ, Q. R. Elaiomycin, a new tuberculostatic antibiotic; isolation and chemical characterization. *Antibiot. Chemother. (Northfield, Ill.)* **4**, 141–4 (1954).

20. VAVRA, J. J., DEBOER, C., DIETZ, A., HANKA, L. J. & SOKOLSKI, W. T. Streptozotocin, a new antibacterial antibiotic. *Antibiot. Annu.* **7**, 230–5 (1959).

21. OMURA, S., TOMODA, H., XU, Q. M., TAKAHASHI, Y. & IWAI, Y. Triacsins, new inhibitors of acyl-CoA synthetase produced by Streptomyces sp. *J. Antibiot. (Tokyo).* **39**, 1211–1218 (1986).

22. TOMODA, H., NAMATAME, I. & OMURA, S. Microbial metabolites with inhibitory activity against lipid metabolism. *Proc. Japan Acad. Ser. B* **78**, 217–240 (2002).

23. Albarano, L., Esposito, R., Ruocco, N. & Costantini, M. Genome Mining as New Challenge in Natural Products Discovery. *Mar. Drugs* **18**, 199 (2020).

24. Ren, H., Shi, C. & Zhao, H. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. *iScience* **23**, 100795 (2020).

25. Bauman, K. D., Butler, K. S., Moore, B. S. & Chekan, J. R. Genome mining methods to discover bioactive natural products. *Nat. Prod. Rep.* **38**, 2100–2129

(2021).

26. Robinson, S. L. *et al.* Global analysis of adenylate-forming enzymes reveals β-lactone biosynthesis pathway in pathogenic Nocardia. *J. Biol. Chem.* **295**, 14826–14839 (2020).

27. Hermenau, R. *et al.* Genomics-Driven Discovery of NO-Donating Diazeniumdiolate Siderophores in Diverse Plant-Associated Bacteria. *Angew. Chemie Int. Ed.* **58**, 13024–13029 (2019).

28. Xu, M., Wang, W. & Wright, G. D. Glycopeptide antibiotic discovery in the genomic era. in *Methods in Enzymology* 325–346 (2022). doi:10.1016/bs.mie.2021.11.009.

29. Tang, X. *et al.* Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).

30. 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. Chemie Int. Ed.* **48**, 767–770 (2009).

31. Waldman, A. J. & Balskus, E. P. Discovery of a Diazo-Forming Enzyme in Cremeomycin Biosynthesis. *J. Org. Chem.* **83**, 7539–7546 (2018).

32. Del Rio Flores, A. *et al.* Biosynthesis of triacsin featuring an N-hydroxytriazene pharmacophore. *Nat. Chem. Biol.* **17**, 1305–1313 (2021).

33. Hagihara, R., Katsuyama, Y., Sugai, Y., Onaka, H. & Ohnishi, Y. Novel desferrioxamine derivatives synthesized using the secondary metabolism-specific nitrous acid biosynthetic pathway in Streptomyces davawensis. *J. Antibiot. (Tokyo).* **71**, 911–919 (2018).

34. Ng, T. L., Rohac, R., Mitchell, A. J., Boal, A. K. & Balskus, E. P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **566**, 94–99 (2019).

35. McBride, M. J. & Boal, A. K. SznF, a Metalloenzyme Employed in the Biosynthesis of Streptozotocin. in *Encyclopedia of Inorganic and Bioinorganic Chemistry* 1–11 (Wiley, 2021). doi:10.1002/9781119951438.eibc2775.

36. Du, Y.-L., He, H.-Y., Higgins, M. A. & Ryan, K. S. A heme-dependent enzyme forms the nitrogen–nitrogen bond in piperazate. *Nat. Chem. Biol.* **13**, 836–838 (2017).

37. Matsuda, K. *et al.* A Natural Dihydropyridazinone Scaffold Generated from a Unique Substrate for a Hydrazine-Forming Enzyme. *J. Am. Chem. Soc.* **144**, 12954–12960 (2022).

38. Selim, M. S. M., Abdelhamid, S. A. & Mohamed, S. S. Secondary metabolites and biodiversity of actinomycetes. *J. Genet. Eng. Biotechnol.* **19**, 72 (2021).

39. Takahashi, Y. & Nakashima, T. Actinomycetes, an inexhaustible source of naturally occurring antibiotics. *Antibiotics* (2018) doi:10.3390/antibiotics7020045.

40. Langley, B. W., Lythgoe, B. & Riggs, N. V. 512. Macrozamin. Part II. The aliphatic azoxy structure of the aglycone part. *J. Chem. Soc.* 2309 (1951) doi:10.1039/jr9510002309.

41. Gamble, D. S. & Hoffman, I. Photometric analysis of trace amounts of hydrazine with p -dimethylaminobenzaldehyde. Chemical equilibria. *Can. J. Chem.* **45**, 2813–2819 (1967).

42. Garg, R. P., Alemany, L. B., Moran, S. & Parry, R. J. Identification, Characterization, and Bioconversion of a New Intermediate in Valanimycin Biosynthesis. *J. Am. Chem. Soc.* **131**, 9608–9609 (2009).

43. Zheng, Z. *et al.* Reconstitution of the Final Steps in the Biosynthesis of Valanimycin Reveals the Origin of Its Characteristic Azoxy Moiety. *Angew. Chemie* **136**, (2024).

44. Garg, R. P., Qian, X. L., Alemany, L. B., Moran, S. & Parry, R. J. Investigations of valanimycin biosynthesis: Elucidation of the role of seryl-tRNA. *Proc. Natl. Acad. Sci.* **105**, 6543–6547 (2008).

45. Garg, R. P., Gonzalez, J. M. & Parry, R. J. Biochemical Characterization of VlmL, a Seryl-tRNA Synthetase Encoded by the Valanimycin Biosynthetic Gene Cluster. *J. Biol. Chem.* **281**, 26785–26791 (2006).

46. Shi, J. *et al.* Conserved Enzymatic Cascade for Bacterial Azoxy Biosynthesis. *J. Am. Chem. Soc.* **145**, 27131–27139 (2023).

47. Garg, R. P., Ma, Y., Hoyt, J. C. & Parry, R. J. Molecular characterization and analysis of the biosynthetic gene cluster for the azoxy antibiotic valanimycin. *Mol. Microbiol.* **46**, 505–517 (2002).

48. Ma, Y., Patel, J. & Parry, R. J. A novel valanimycin-resistance determinant (vlmF) from Streptomyces viridifaciens MG456-hF10 The GenBank accession number for the sequence in this paper is AF148322. *Microbiology* **146**, 345–352 (2000).

49. Saripella, G. V., Sonnhammer, E. L. L. & Forslund, K. Benchmarking the next generation of homology inference tools. *Bioinformatics* **32**, 2636–2641 (2016).

50. Scobie, D., Hjorleifsson, G., Herron, P., Rogers, S. & Duncan, K. The Missing Link: Developing a pipeline for accelerated antibiotic discovery from Streptomyces through linking ‘omics data. *Access Microbiol.* **2**, (2020).

51. Otasek, D., Morris, J. H., Bouças, J., Pico, A. R. & Demchak, B. Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* **20**, 185 (2019).

52. Wibowo, M. *et al.* Azodyrecins A–C: Azoxides from a Soil-Derived Streptomyces Species. *J. Nat. Prod.* **83**, 3519–3525 (2020).

53. Hua, W., Christianson, T., Rougeot, C., Rochefort, H. & Clinton, G. M. SKOV3 ovarian carcinoma cells have functional estrogen receptor but are growth-resistant to estrogen and antiestrogens. *J. Steroid Biochem. Mol. Biol.* **55**, 279–289 (1995).

54. Reale, F. R. *et al.* Characterization of a human malignant mesothelioma cell line (H-MESO-1): a biphasic solid and ascitic tumor model. *Cancer Res.* **47**, 3199–205 (1987).

55. Yang, T. *et al.* Determination of the Membrane Transport Properties of Jurkat Cells with a Microfluidic Device. *Micromachines* **10**, 832 (2019).

56. Maleckis, M. *et al.* Biosynthesis of the azoxy compound azodyrecin from Streptomyces mirabilis P8-A2. *bioRxiv* 2023.10.13.562208 (2023) doi:10.1101/2023.10.13.562208.

57. Zhao, G. *et al.* The Biosynthetic Gene Cluster of Pyrazomycin—A C-Nucleoside Antibiotic with a Rare Pyrazole Moiety. *ChemBioChem* **21**, 644–649 (2020).

58. Katsuyama, Y. & Matsuda, K. Recent advance in the biosynthesis of nitrogen–nitrogen bond–containing natural products. *Curr. Opin. Chem. Biol.* **59**, 62–68 (2020).

59. Zhao, G. *et al.* Molecular basis of enzymatic nitrogen-nitrogen formation by a family of zinc-binding cupin enzymes. *Nat. Commun.* **12**, 7205 (2021).

60. Matsuda, K. *et al.* Discovery of Unprecedented Hydrazine-Forming Machinery in Bacteria. *J. Am. Chem. Soc.* **140**, 9083–9086 (2018).

61. Wang, Z.-C. *et al.* Genetic and Biochemical Characterization of Halogenation and Drug Transportation Genes Encoded in the Albofungin Biosynthetic Gene Cluster. *Appl. Environ. Microbiol.* **88**, (2022).

62. Shen, W. & Ren, H. TaxonKit: A practical and efficient NCBI taxonomy toolkit. *J. Genet. Genomics* **48**, 844–850 (2021).

63. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

64. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

65. Martín del Campo, J. S. *et al.* Inhibition of the Flavin-Dependent Monooxygenase Siderophore A (SidA) Blocks Siderophore Biosynthesis and Aspergillus fumigatus Growth. *ACS Chem. Biol.* **11**, 3035–3042 (2016).

66. Li, B. *et al.* Functional Identification of Putrescine C - and N -Hydroxylases. *ACS Chem. Biol.* **11**, 2782–2789 (2016).

67. Heemstra, J. R., Walsh, C. T. & Sattely, E. S. Enzymatic Tailoring of Ornithine in the Biosynthesis of the Rhizobium Cyclic Trihydroxamate Siderophore Vicibactin. *J. Am. Chem. Soc.* **131**, 15317–15329 (2009).

68. Van Cura, D. *et al.* Discovery of the Azaserine Biosynthetic Pathway Uncovers a Biological Route for α-Diazoester Production. *Angew. Chemie Int. Ed.* **62**, (2023).

69. Wang, S.-A. *et al.* 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.* **141**, 6127–6131 (2019).

70. Silva, A. *et al.* COMPOSIÇÃO QUÍMICA E ATIVIDADES FOTOPROTETORA E ANTIRADICALAR IN VITRO DOS GALHOS DE Platonia insignis (CLUSIACEAE). *Quim. Nova* **44**, 543–546 (2021).

71. Garside, D. M., Monteiro, P. M. S. & Orren, M. J. A critical evaluation for the determination of amino acids in the marine environment by derivatization using 9-fluorenylmethyl chloroformate (FMOC-Cl) and reversed phase HPLC separation. *South African J. Mar. Sci.* **6**, 47–53 (1988).