UNIVERSITY OF CALIFORNIA

Santa Barbara

Genome Mining for Microbial Siderophores:

Improving Structural Predictions

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemistry

by

Zachary Lawrence Reitz

Committee in charge:

Professor Alison Butler, Chair

Professor Kevin Plaxco

Professor Herbert Waite

Professor Irene Chen

December 2020

The dissertation of Zachary Lawrence Reitz is approved.

_____

Kevin Plaxco

_____

Herbert Waite

_____

Irene Chen

_____

Alison Butler, Committee Chair

December 2020

Genome Mining for Microbial Siderophores:

Improving Structural Predictions

Zachary Lawrence Reitz

ACKNOWLEDGEMENTS

First, I want to thank my advisor, Alison Butler. She let me into her group when I was an organic chemist who had never taken a biochemistry class, and she took another risk by letting me pursue genome mining and evolution. Alison's guidance and support has kept me on the path towards becoming a better scientist. Her enthusiasm for siderophores is infectious, perhaps literally so: she has instilled that same enthusiasm in me.

I could not have gotten a doctorate alone. Special thanks goes to Dr. Clifford Hardy and Jeffrey Carmichael. They were among the very first friends I made at UCSB, and they encouraged me to reach out to Alison. Together we have seen labmates come and go: Dr. Stephen Springer, Dr. Greg Maier, Dr. Robert Lewis, Aneta Jelowicki, Christina Makris, Parker Stow, and Emil Thomsen. Each played a role in shaping my graduate career, and many have worked on projects alongside me. I've also worked with several talented undergraduates and visiting researchers: Kalana Dulaney, Andrew Foster, Jaewon Suk, Robert Augustynski, Natthanan "Teety" Vijara, and Jean Bouvet. I thank our external collaborators Dr. Margo Haygood, Dr. Hiroaki Naka, and Dr. Moriah Sandy for lending their knowledge and expertise.

My committee members, Professors Kevin Plaxco, Herbert Waite, and Irene Chen, contributed valuable discussion and insights. My teaching improved under Dr. Petra Van Koppen and Dr. Morgan Gainer, educators who truly care about their students. Both Dr. Dmitriy Uchenik and Dr. Hongjun Zhou contributed their impressive technical expertise. I'm grateful to Dr. Rachel Behrens for her professional and life advice, brightening the many hours I spent at the UPLC-MS. Finally, the Department of Chemistry and Biochemistry administrative, facilities, and laboratory staff made everything happen behind the scenes.

Shortly after joining the Butler lab, I met Alexander Lau, who has been by my side for nearly five years now. Alex is selfless and caring, and has helped me grow into a better person than I was when we first met. He is my best friend, and I can't imagine my time at UCSB without him. Throughout graduate school, I've had a fantastic network of friends who have encouraged me and kept me sane. I am so thankful that there are too many to name: the chemists, sailors, climbers, housemates, trivia-goers, dogs, and anyone else who has spent an evening at a dinner party in our backyard. Finally, all the love and thanks to my parents, Bobbi and Brian Reitz, who fostered my young interest in science and nature, and have supported me every step of the way.

VITA OF ZACHARY LAWRENCE REITZ
December 2020

EDUCATION

Doctor of Philosophy, Chemistry. December 2020
University of California, Santa Barbara, CA

Bachelor of Science, Chemistry. May 2014
The Pennsylvania State University, State College, PA

Certificate, Workshop on Molecular Evolution. August 2018
Marine Biological Laboratory

PUBLICATIONS

Reitz, Z. L.; Butler, A. Precursor-directed biosynthesis of catechol compounds in *Acinetobacter bouvetii* DSM 14964. Chem. Commun. 2020, 56 (81), 12222-12225.

Reitz, Z. L.; Hardy, C. D.; Suk, J.; Bouvet, J.; Butler, A. Genomic Analysis of Siderophore β-Hydroxylases Reveals Divergent Regio- and Stereocontrol and Expands the Condensation Domain Family. Proc. Natl. Acad. Sci. U.S.A. 2019, 116 (40), 19805-19814.

Naka, H.; Reitz, Z. L.; Jelowicki, A. L.; Butler, A.; Haygood, M. G. Amphi-Enterobactin Commonly Produced among *Vibrio campbellii* and *Vibrio harveyi* Strains Can Be Taken up by a Novel Outer Membrane Protein FapA That Also Can Transport Canonical Fe(III)-Enterobactin. J. Biol. Inorg. Chem. 2018, 23 (7), 1009–1022.

Reitz, Z. L.; Sandy, M.; Butler, A. Biosynthetic Considerations of Triscatechol Siderophores Framed on Serine and Threonine Macrolactone Scaffolds. Metallomics 2017, 9 (7), 824–839.

INVITED TALKS

"Genomic analysis of siderophore Fe(II)/α-ketoglutarate-dependent Aminoacyl β-Hydroxylases reveals functional subtypes." National Meeting of the American Chemical Society in San Diego, CA. August 2019.

"Chained by Iron: The Tug of War Between Bacteria and our Body." Grad Slam competition in Santa Barbara, CA. April 2019.

"Genome Mining for Unwinding Iron Binding." Chemical Sciences Student Symposium in Santa Barbara, CA. March 2019.

"An Ironbound Strategy: Genomic Analyses of Siderophore β-Hydroxylases." Southern California BioInorganic Meeting in Pasadena, CA, December 2018.

"Strike Where the Iron is Not: Exploring the Biology and Chemistry of Iron Chelators." OUTreach Seminar Series in Santa Barbara, CA. October 2018.

"Streamlining Siderophore Discovery through Genomic Predictions." Gordon Research Seminar, Bioinorganic Chemistry in Ventura, CA. January 2018.

"Genomics-based Expansion of a Natural Combinatoric Library of Siderophores." Southern California BioInorganic Meeting in Irvine, CA. December 2017.

POSTER PRESENTATIONS

Reitz, Z. L.; Dulaney, K.; Butler, A. "Genomics-based Expansion of a Natural Combinatoric Library of Siderophores." Poster presented at National Meeting of the American Chemical Society, New Orleans, LA, March 2018.

Reitz, Z. L.; Dulaney, K.; Butler, A. "Streamlining Siderophore Discovery through Genomic Predictions" Poster presented at Gordon Research Seminar, Bioinorganic Chemistry, Ventura, CA, January 2018.

Reitz, Z.; Gonzalez, I.; Feldman, K. "Intramolecular [3+4] Additions to Indolidenium Cations." Poster presented at 247th National Meeting of the American Chemical Society, Dallas, TX, March 2014.

PROFESSIONAL EXPERIENCE

Graduate Student Researcher; Advisor: Alison Butler. 2015 - 2020.
University of California, Santa Barbara, CA

Teaching Assistant, Organic Chemistry Laboratory. 2015 - 2016, 2019 - 2020
University of California, Santa Barbara, CA

Lead Teaching Assistant, Organic Chemistry. 2016.
University of California, Santa Barbara, CA

Teaching Assistant, General Chemistry Laboratory. 2014 - 2015.
University of California, Santa Barbara, CA

Undergraduate Student Researcher; Advisor: Kenneth S. Feldman. 2013 - 2014.
The Pennsylvania State University, State College, PA

Undergraduate Student Researcher; Advisor: Katherine M. Masters. 2012 - 2013.
The Pennsylvania State University, State College, PA

Teaching Assistant, Spectroscopy. 2014.
The Pennsylvania State University, State College, PA

Undergraduate Teaching Assistant, Advanced Organic Chemistry Laboratory. 2013.
The Pennsylvania State University, State College, PA

AWARDS

Grad Slam Runner-up, University of California, Santa Barbara, 2019.

Phi Lambda Upsilon Award, University of California, Santa Barbara, 2015.

Organic Division Undergraduate Award, American Chemical Society, 2014.

3M Summer Research Fellowship, Pennsylvania State University, 2013.

Eagle Scout, Boy Scouts of America, 2011.

ABSTRACT

Genome Mining for Microbial Siderophores:

Improving Structural Predictions

by

Zachary Lawrence Reitz

Nearly all life requires iron; however, under aerobic conditions, Fe(III) is nearly insoluble. To meet their metabolic needs, many bacteria synthesize siderophores: small-molecule, high-affinity Fe(III) chelators. The discovery of new siderophores has been streamlined by *genome mining*, where bacterial genomes of interest are scanned for siderophore biosynthetic gene clusters (BGCs). This dissertation is focused on the development of new bioinformatics tools to improve siderophore structural predictions, as well as the application of those tools towards traditional natural product isolation.

Some peptidic siderophores contain the chelating ligand β-hydroxyaspartate (β-OHAsp), which provides bidentate OO′ coordination to Fe(III). Current genome mining methods cannot reliably predict which aspartate residues will be hydroxylated, nor the resulting stereochemistry, leaving structural ambiguities which must be rectified experimentally. Through coupling BGCs with verified structures, the origin of the β-OHAsp diastereomers in siderophores is reported. Two functional subtypes of nonheme Fe(II)/α-ketoglutarate–dependent aspartyl β-hydroxylases were identified in siderophore BGCs. Each aspartyl β-hydroxylase subtype effects distinct diastereoselectivity. A previously

undescribed, noncanonical member of the nonribosomal peptide synthetase (NRPS) condensation domain superfamily is identified, named the *interface* domain, which is proposed to position the β-hydroxylase and the NRPS-bound amino acid prior to hydroxylation. Through mapping characterized β-OHAsp diastereomers to the phylogenetic tree of siderophore β-hydroxylases, methods to predict β-OHAsp stereochemistry *in silico* are realized.

The DHB-CAA-Ser family of triscatechol siderophores each contain a different cationic amino acid (CAA = D-Arg, D-Lys, or L-Orn), hinting at the existence of their diastereomers with L-Arg, L-Lys, and D-Orn. With no knowledge of which bacteria, if any, are capable of their production, a novel high-throughput genome mining workflow was developed, the Catechol Siderophore Cluster Analysis (CatSCAn). Over 11,000 bacterial genomes were scanned for siderophore BGCs, revealing about 1% potentially capable of the biosynthesis of the combinatoric suite of DHB-CAA-Ser siderophores with L- and D-CAA. The NRPS from *Marinomonas* sp. TW1 was found to be unusually promiscuous, producing a relatively rare example of a suite of siderophores with D-Arg, D-Orn, and L-Orn amino acids.

CatSCAn produced a rich dataset of putative BGCs that may be coupled with structurally- characterized siderophores to study the chemistry and biology of triscatechol siderophores. A detailed analysis of BGCs highlighted key features within siderophore pathways. Mapping known and predicted siderophores onto a phylogenetic tree of NRPS proteins reveals repeated changes in CAA stereochemistry and identity within the DHB-CAA-Ser family. Parsimony suggests that the ancestor of all DHB-CAA-Ser siderophores was trivanchrobactin, (DHB-^{D}Arg-Ser)$_3$. A model of DHB-CAA-Ser siderophore evolution is presented where enzyme promiscuity allows for the biosynthesis of new siderophores.

Genome mining for VibH homologs reveals several species of *Acinetobacter* with a gene cluster that putatively encodes the biosynthesis of catechol siderophores with an amine core. *A. bouvetii* DSM 14964 produces three novel biscatechol siderophores: propanochelin, butanochelin, and pentanochelin. This strain has a relaxed specificity for the amine substrate, allowing for the biosynthesis of a variety of non-natural siderophore analogs by precursor directed biosynthesis. Of potential synthetic utility, *A. bouvetii* DSM 14964 condenses 2,3-dihydroxybenzoic acid (2,3-DHB) to allylamine and propargylamine, producing catecholic compounds which bind iron(III) and may be further modified via thiol–ene or azide–alkyne click chemistry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Introduction: Siderophore-mediated iron acquisition

Sections of this chapter were adapted from Reference 1 with permission from The Royal Society of Chemistry.

1.1. Iron homeostasis and siderophores

The majority of bacteria require iron to grow, yet obtaining iron presents a challenge to microbes in most environments. Iron(III) is insoluble at neutral pH in aerobic environments [K_{sp} of $Fe(OH)_3 = 10^{-39}$], thus severely limiting the availability of this essential nutrient in marine and terrestrial environments. Iron availability is also limited in human-host environments, due to sequestration of cellular Fe(III) by the iron transport and storage proteins transferrin, lactoferrin, and ferritin. One strategy bacteria use to acquire iron is to produce siderophores, low molecular weight organic compounds that bind Fe(III) with high affinity. Siderophores solubilize, capture, and deliver Fe(III) to the cells (Figure 1.1). Several hundred structures of siderophores are known, all of which contain key functional groups for Fe(III) coordination. The most common bidentate coordinating groups are catechols, hydroxamates, α-hydroxycarboxylates, salicylates, and N-diazeniumdiolates (Figure 1.2). Most siderophores contain three chelating groups, providing hexadentate coordination around the Fe(III) center. A siderophore's affinity for ferric iron is often expressed in terms of its proton-independent Fe(III) stability constant (K_f or β_{110}), a formal convention calculated based on competition experiments with EDTA and estimated K_a values of the coordinating atoms. In reality, a siderophore's affinity for Fe(III) is *not* proton-independent, but varies with pH.[2] A more physiologically relevant (albeit still artificial) measure of the ability to bind ferric iron is pFe(III), the equilibrium free iron concentration in a pH 7.4 solution of 10^{-6} M Fe(III) and 10^{-5} M ligand.

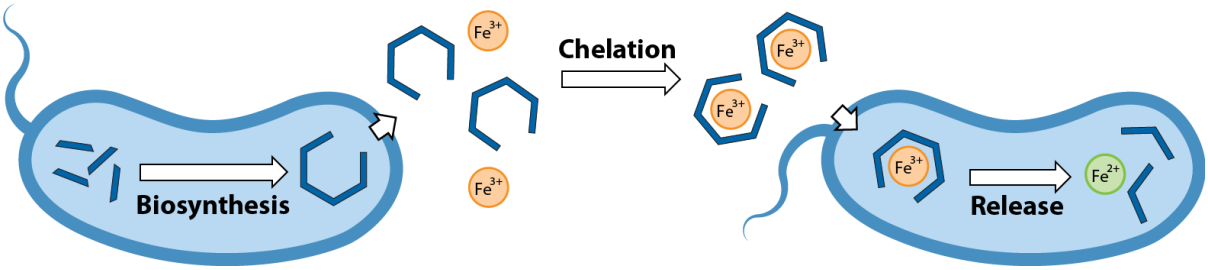


Figure 1.1. Siderophore-mediated iron acquisition in bacteria.

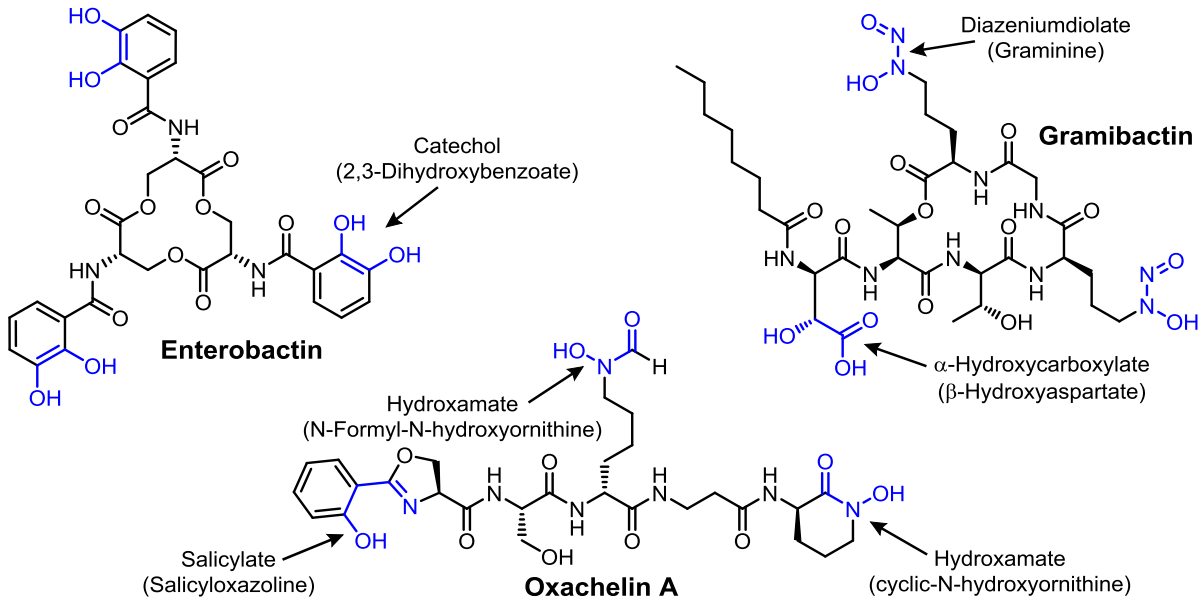


Figure 1.2. Common siderophore iron(III) coordinating groups. The coordinating groups of representative siderophores are highlighted in blue and labeled. The names of the specific monomers are given in parentheses.

1.1.1. Metal-center chirality of ferric siderophore complexes

Upon Fe(III) coordination, siderophores may gain an additional stereocenter, *i.e.*, the metal center itself, with configurations named Δ or Λ (Figure 1.3). Crucially, the "wrong" Fe(III)-center chirality has been implicated in reducing siderophore efficacy by hindering receptor recognition.[3-7] The metal center chirality of several ferric siderophore complexes has been determined by circular dichroism (CD) spectroscopy and X-ray crystallography,

allowing for the relationship between siderophore structure and metal center chirality to be explored.

A pair of enantiomeric ligands will each bind a metal with equal affinities and opposite metal-center chiralities. Transport studies of unnatural siderophore stereoisomers provided the first evidence for stereoselective siderophore-based iron acquisition. Synthetic enantiomers enantio-ferrichrome,[3,8] D-parabactin,[9] and enantio-rhodotorulic acid[4] are not imported into bacteria or fungi. Iron(III) bound to enantio-enterobactin likewise doesn't support growth in *E. coli*.[10] However, the outer and inner membrane receptors of *E. coli* are able to transport both enantiomers of enterobactin; instead, enantioselectivity lies at the esterase Fes, which is unable to cleave the D-Ser macrolactone of ferric enantio-enterobactin, preventing iron release and utilization.[11,12]

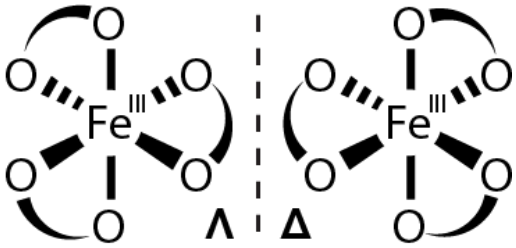


Figure 1.3. Enantiomers of Fe(III) coordinated by three bidentate ligands.

Synthetic siderophore enantiomers provided the first evidence for the impact of iron center chirality on the uptake of siderophores; however, they lack the true ecological relevance of natural siderophore stereoisomers paired with matching uptake machinery. Three naturally occurring pairs of enantiomeric siderophores have been reported to date: *R,R*-rhizoferrin *versus S,S*-rhizoferrin (from Mucorales fungi and γ-proteobacteria, respectively),[6,13] pyochelin *versus* enantio-pyochelin (from *Pseudomonas aeruginosa* and *P. fluorescens*, respectively),[14,15] and madurastatin C1 *versus* (–)-madurastatin C1 (each from

different *Actinomadura* spp.).[16,17] *S,S*-rhizoferrin producing bacterium *Ralstonia pickettii* was able to take up *R,R*-rhizoferrin with equal efficiency, while the fungus *Rhizopus arrhizus* imported the native *R,R*-rhizoferrin at a two-fold faster rate compared to *S,S*-rhizoferrin.[6] Pyochelin and enantio-pyochelin are only recognized as siderophores and signalling molecules in their respective strains.[15] X-ray crystallography revealed that the outer membrane receptors FptA and FetA show high structural similarities, but share no homology in the siderophore binding sites.[18] Additional enantioselectivity occurs at the inner membrane receptors and transcriptional regulators.[19,20]

In contrast to enantiomeric pairs, two siderophore diastereomers do not necessarily have equal affinities for iron(III). Indeed, salicyl thiazoline-containing siderophores pyochelin and yersiniabactin are isolated as rapidly equilibrating mixtures of epimers; only one diastereomer is able to bind iron(III).[21,22] Other thiazoline/oxazoline-containing siderophores isolated as epimers include watasemycins,[23] thiazostatins,[24] spoxazomicins,[25] and tetroazolemycins.[26] Similar epimeric siderophores agrochelin and massiliachelin were each reported as single diastereomers (from *Agrobacterium* sp. and *Massilia* sp. NR 4-1, respectively), although no iron(III) chelation experiments were performed.[27,28]

1.1.2. Transport of ferric siderophore complexes

In Gram-negative bacteria, ferric siderophore [Fe(III)-Sid] complexes are recognized and imported by cognate outer membrane receptors (OMRs). These large proteins are composed of a β-barrel domain, which forms the channel, and a plug domain, which prevents the unintended transit of other molecules.[29] Upon Fe(III)-Sid binding, the TonB-ExbB-ExbD inner membrane complex transduces energy from proton motive force to the receptor, moving the plug and allowing passage of the Fe(III)-Sid complex.[29] A periplasmic

binding protein (PBP) then binds the siderophore complex, escorting it either to an ATP-binding cassette (ABC) inner membrane transporter complex, or directly to a periplasmic iron release mechanism (see below). Gram-positive bacteria lack an outer membrane; instead, a membrane-bound PBP homolog called a siderophore binding protein (SBP) recognizes the Fe(III)-Sid complex and facilitates import to the cytoplasm through an ABC transporter.[30] The representative ferric siderophore uptake systems of pyoverdine of *Pseudomonas aeruginosa* and enterobactin in *Escherichia coli* are shown in Figure 1.4.

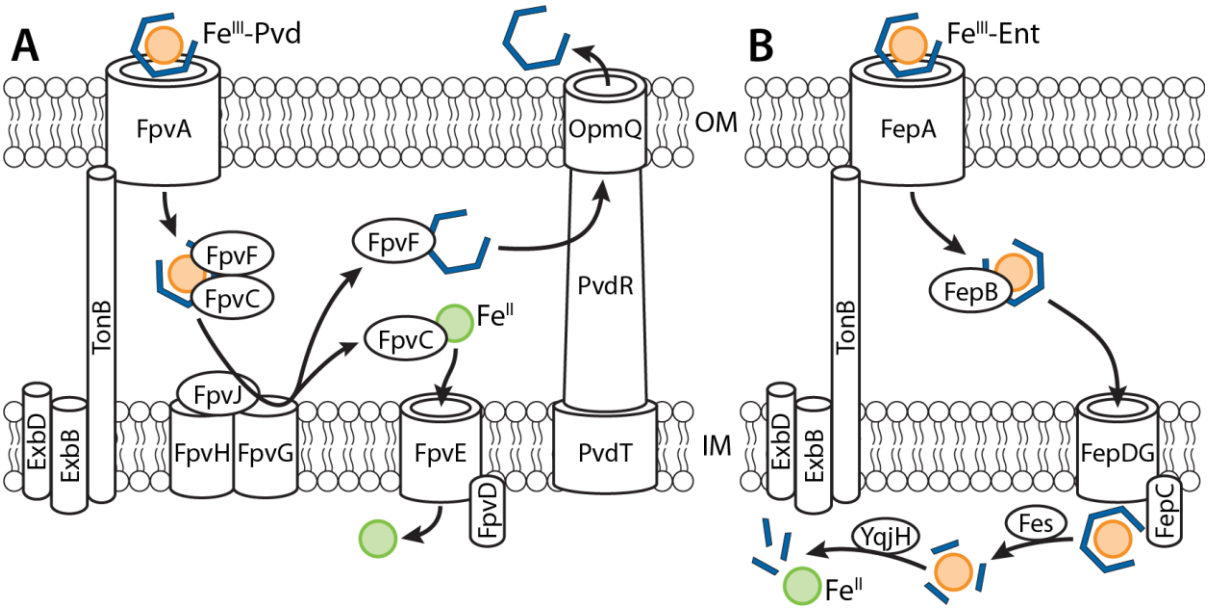


Figure 1.4. Mechanisms of ferric siderophore transport and iron release in (A) *Pseudomonas aeruginosa* and (B) *Escherichia coli*. Both ferric siderophores are transported across the outer membrane (OM) by a TonB-ExbB-ExbD-powered outer membrane receptor. (A) Ferric pyoverdine (Fe^{III}-Pvd) is escorted through the periplasm by FpvCF. FpvGHJ reduces complexed Fe(III), liberating Fe(II) and apo-Pvd. FpvC escorts Fe(II) to FpvDE for inner membrane (IM) transport to the cytoplasm, while FpvF escorts apo-Pvd to the OmpQ-PvdRT efflux pump to be recycled into the environment. (B) Ferric enterobactin (Fe^{III}-Ent) is escorted through the periplasm by FepB and transported intact through the IM by FepCDG. Cytoplasmic Fes cleaves the Fe^{III}-Ent backbone prior to Fe(III) reduction by YqjH.

1.1.3. Mechanisms of iron release

Iron must be released from the siderophore before it can be incorporated into metalloenzymes. Bacterial metalloenzymes and iron storage proteins cannot compete with the high Fe(III) affinity of siderophores; however, a siderophore's affinity for Fe(II) is generally much lower (ex: $K_f(Fe^{III}) = 30.8$, $K_f(Fe^{II}) = 9.0$ for pyoverdine).[31] In *Pseudomonas aeruginosa*, pyoverdine-bound Fe(III) is reduced by inner membrane protein complex FpvGHJ (Figure 1.4A).[32] Iron(II) is then transferred to periplasmic binding protein FpvC, and apo-pyoverdine is recycled back into the environment.[32] The low reduction potential of Fe(III)-enterobactin (–750 mV at pH 7) prevents biological reduction.[33] The macrolactone backbone of Fe(III)-enterobactin must first be cleaved by esterase Fes, bringing the reduction potential to –350 mV and allowing NADPH-dependent reduction by YqjH to occur (Figure 1.4).[33,34]

1.2. Common siderophore biosynthesis pathways

1.2.1. Nonribosomal peptide synthetases (NRPSs)

Many microbial secondary metabolites, including siderophores, are produced by nonribosomal peptide synthetases (NRPSs). A NRPS is a large, multifunctional enzyme that utilizes an assembly line approach to synthesize a specific peptide product. NRPSs are organized into one or more catalytic modules, each of which is responsible for incorporation of a specific monomer unit (often an L-amino acid) into the peptide. A NRPS may work alone or function in tandem with other NRPSs or polyketide synthase (PKS) enzymes. Iterative NRPSs use each module more than once during the synthesis of a single peptide, producing multimeric products. In addition to L- amino acids, nonribosomal peptides can include D-, non-proteinogenic,

and modified amino acids, as well as fatty acids, aryl acids, and diamines.[35] In the NRPS-based synthesis of siderophores, this monomer flexibility allows for the control of chelating moieties, solubility, and conformation needed to effectively coordinate Fe(III). The representative NRPS-mediated biosynthesis of serobactin A is shown in Figure 1.5.

A NRPS comprises three core catalytic domains: adenylation (A) domains, thiolation (T) domain (also known as a peptidyl carrier protein, PCP), and condensation (C) domains. The A domain selectively binds and adenylates an amino acid substrate, activating it for further processing. Stachelhaus *et al.* identified 10 key residues that interact with the substrate.[36] These residues compose a "specificity-conferring code" that can be used to predict A domain specificity based on peptide sequence.[37] A variety of other sequence-based predictive methods have since been developed.[38-40] The resulting aminoacyl-adenylate is transferred to a thiol moiety of 4'-phosphopantetheine (Ppant), which is tethered to a T domain, to form an aminoacyl-thioester. A Ppant transferase is responsible for posttranslational insertion of the Ppant moiety onto a conserved serine within the active site of the T domain.[41] The T domain functions as a shuttle, carrying its thioester-bound substrate to and from various catalytic domains of the NRPS. The C domain catalyzes peptide-bond formation between specific aminoacyl substrates.

In addition to the core A, T, and C domains, NRPSs may also contain domains responsible for substrate modifications, including cyclization, N-methylation, N-formylation, oxidation, and reduction. C-terminal NRPS modules typically contain a thioesterase (Te) domain.[35] The Te domain catalyzes the release of the covalently tethered product in the final step of peptide biosynthesis. In this step, the full-length

product is transferred from the T domain to a conserved Ser in the Te active site. The Te domain catalyzes release of the acyl-O-Te intermediate either via hydrolysis, leading to a linear peptide product, or through intramolecular nucleophilic attack, leading to a cyclic peptide product.[42]

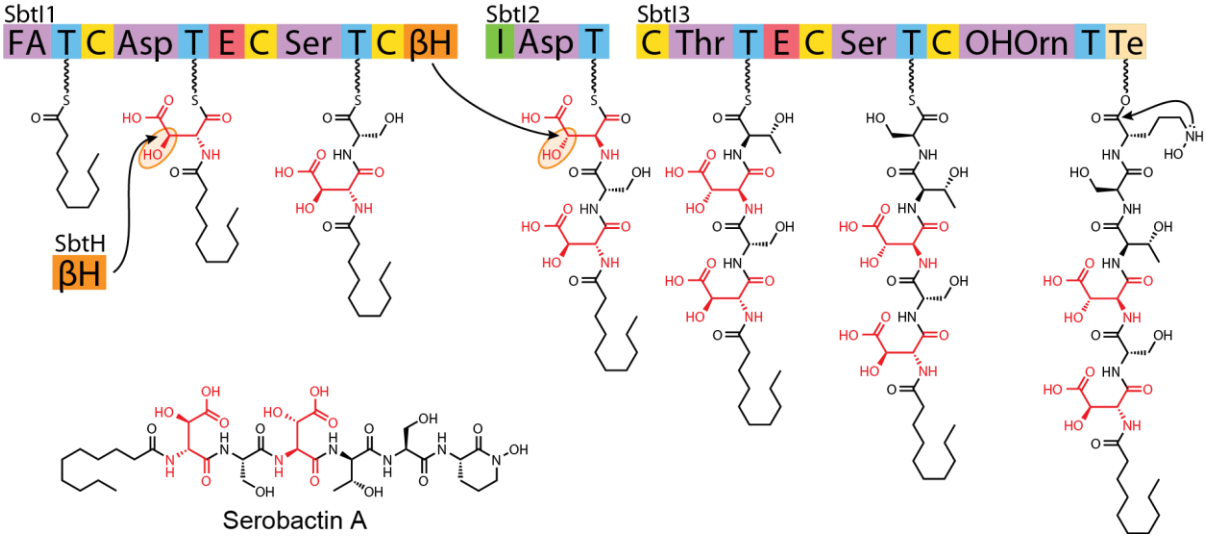


Figure 1.5. Proposed nonribosomal biosynthesis of serobactin A. Aspartyl β-hydroxylases (βH) and interface (I) domains are described in Chapter 2. Adenylation (A) domains are labelled by the substrate they activate and incorporate. C, condensation domain; E, epimerization domain; T, thiolation domain; Te, thioesterase domain.

D-amino acids are generally not directly activated and incorporated into bacterial non-ribosomal peptides; rather, an L-amino acid is first adenylated and condensed before conversion to the D-amino acid by an epimerization (E) domain, which deprotonates and reprotonates the α-carbon with catalytic His and Glu residues.[43,44] This process favors the D-amino acid, but still provides a mixture of D- and L-isomers.[45] The high stereoselectivity of NRPS modules is therefore provided by the

downstream C domain.[45,46] These C domains, downstream from an E domain and selective for catalyzing the formation of an amide bond between D- and L-amino acids, are termed $^{D}C_{L}$. By contrast, $^{L}C_{L}$ domains are selective for an upstream L-amino acid and follow T domains directly.[46,47]

As condensation domains require an upstream and downstream residue, the N-terminal initiation module of a NRPS generally lacks a C domain.[48] However, Rausch *et al*. delineated a class of condensation domains found only in the initiation module, called Starter C (C_{start}) domains. A C_{start} domain acylates the first amino acid with an acyl group, often a β-hydroxy fatty acid or a hydroxylated aryl acid derivative such as salicylate or 2,3-dihydroxybenzoate (DHB).[47] Fatty acids are activated by a long-chain fatty acid CoA ligase, and react directly with the C_{start} domain,[49-52] while aryl acids are adenylated by a standalone A domain and transferred to an external thiolation domain.[53-55] This T domain, also called an aryl carrier protein (ArCP), interacts with the C_{start} domain during condensation.[56] These domain subtypes — $^{L}C_{L}$, $^{D}C_{L}$, and C_{start} — each have unique structural motifs. Thus, sequence analysis can be used to predict reactivity.[47]

1.2.2. Biosynthetic origins of various chelating groups

1.2.2.1. 2,3-dihydroxybenzoate (2,3-DHB)

2,3-Dihydroxybenzoate (2,3-DHB) is synthesized from chorismate, part of the shikimate amino acid synthesis pathway. The three-step biosynthesis is catalyzed by an isochorismate synthase, an isochorismatase, and a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (Figure 1.6). Synthesized 2,3-DHB is activated via a DHB-AMP ligase and transferred to the a NRPS aryl carrier protein (ArCP) for incorporation into peptidyl siderophores (see below).

1.2.2.2. Hydroxamate

Hydroxamates are all produced by the hydroxylation and acylation of a primary amine (Figure 1.6). In peptidic siderophores, the amine is usually ornithine, which is N^5-hydroxylated by an flavin-dependent monooxygenase.[57] An acyltransferase may then catalyze the formation of δ-N-acyl-N-hydroxyornithine, most often δ-N-formyl- (fOHOrn) or δ-N-acetyl-δ-N-hydroxyornithine (AcOHOrn). Lactamization of δ-N-hydroxyornithine forms cyclic N-hydroxyornithine (cOHOrn). Alternatively, a NRPS condensation domain can ligate N-hydroxyornithine to an upstream amino acid.

1.2.2.3. β-Hydroxyaspartate and β-hydroxyhistidine

Aspartic acid and histidine may be hydroxylated by a family of non-heme Fe(II)/ α-ketoglutarate dependent β-hydroxylases (Figure 1.6). Based on homology to the syringomycin aspartyl β-hydroxylase SyrP, these enzymes are believed to act on amino acids tethered to the thiolation domain of a NRPS.[58] New insights into β-hydroxylation are presented in Chapter 2.

1.2.2.4. Salicylate

Like 2,3-DHB, the biosynthesis of salicylate in siderophores branches from chorismate of the shikimate amino acid pathway (Figure 1.6). In pyochelin biosynthesis, the isochorismate synthase PchA isomerizes chorismate to isochorismate. PchB, an isochorismate pyruvate lyase (IPL), then produces salicylate, releasing pyruvate.[59,60] The salicylate synthases Irp9 and MbtI (of yersiniabactin and mycobactin biosyntheses, respectively) are each single bifunctional domains that carry out both transformations.[61-63]

1.2.2.5. Diazeniumdiolate

The novel siderophore iron binding group diazeniumdiolate was recently discovered in gramibactin (Figure 1.2).[64] Gene knockouts have revealed two enzymes responsible for the synthesis of the diazeniumdiolate-containing amino acid graminine, GrbD and GrbE.[65] The latter is homologous to SznF, a ferroenzyme that catalyzes N-N bond formation in the biosynthesis of streptozotocin.[66] Several more diazeniumdiolate siderophores have since been isolated from Burkholderiaceae based on the presence of *grbDE* homologs in the genomes.[65]

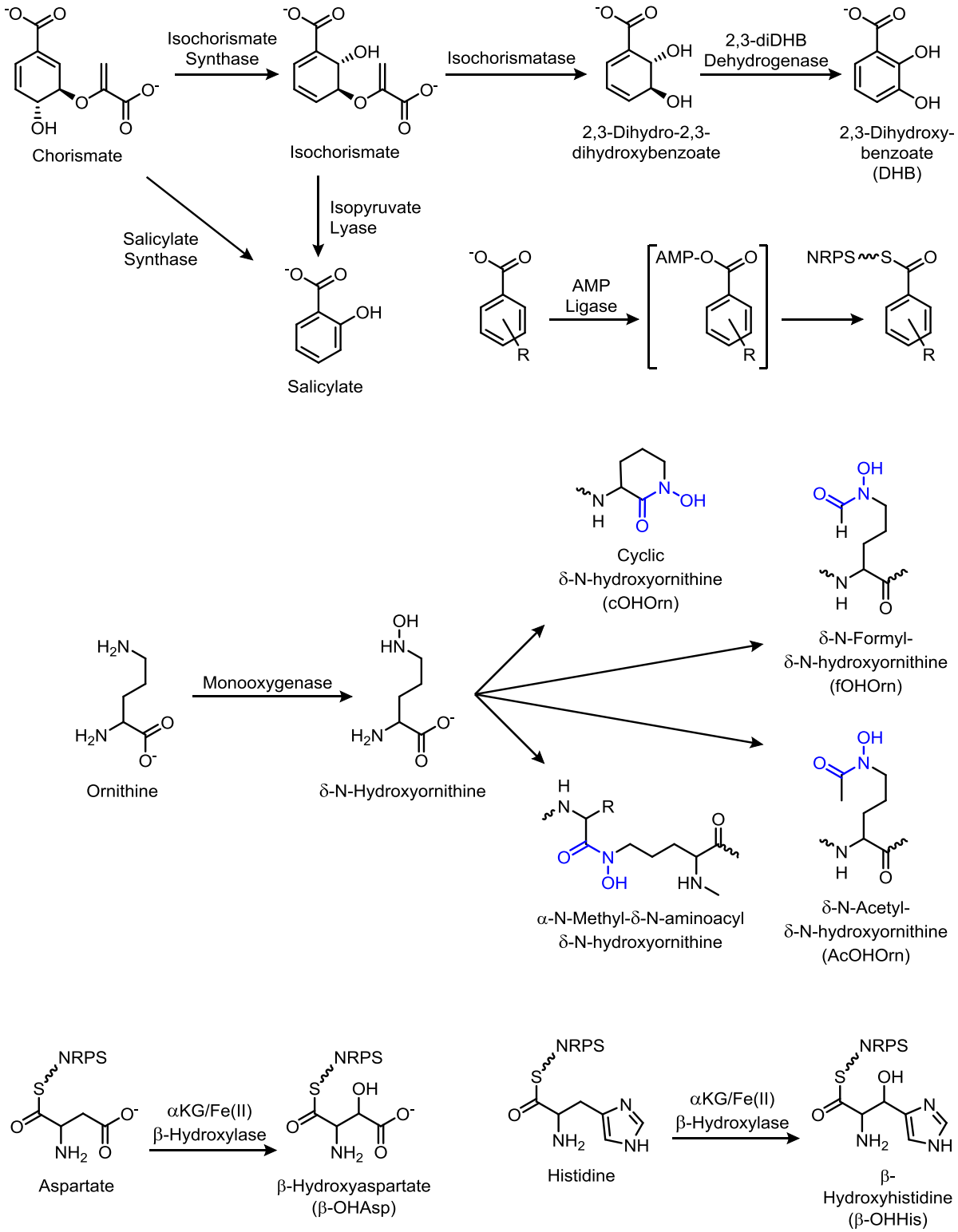


Figure 1.6. Biosyntheses of 2,3-dihydroxybenzoate, salicylate, hydroxamate, β-hydroxyaspartate, and β-hydroxyhistidine chelating groups. β-hydroxyasparagine has also been reported in one siderophore.[67]

1.3. Structural and biosynthetic considerations of catechol siderophores

1.3.1. Serine- and threonine-based triscatechol siderophores

Enterobactin, the macrolactone trimer of N-2,3-dihydroxybenzoyl-L-serine (Figure 1.7), coordinates Fe(III) with three 2,3-dihydroxybenzoate (DHB) catechol groups. As one of the first discovered siderophores, much is known about microbial production of enterobactin, as well as how it functions in microbial iron uptake. The exceptional affinity of enterobactin for Fe(III) has been attributed, in part, to pre-organization of the DHB groups for Fe(III) coordination as a result of macrolactonization of the tri-serine backbone.

Given the exceptional affinity of triscatecholate complexes for Fe(III), especially of those which pre-orient the catechol for octahedral Fe(III) coordination, it is not surprising that bacteria have evolved variants of enterobactin which retain both the pre-organized ester scaffold and three catechol groups.[68-72] Salmochelin and bacillibactin (Figure 1.7) are other well-known tri-Ser and tri-Thr macrolactone siderophores, respectively. Recently, several additional triscatechol macrolactone siderophores based on L-Ser and L-Thr scaffolds have been reported, i.e., amphi-enterobactins,[51] cyclic trichrysobactin,[73] paenibactin,[74] and griseobactin (also called streptobactin),[75,76] as well as linear versions with L-Ser, i.e., trivanchrobactin and turnerbactin (Figure 1.7).[77,78] Other distinct siderophores may well be discovered, with further variation in cyclic versus linear structures, D- versus L-amino acids, and Ser versus Thr scaffolds. The biosynthetic gene clusters for many of the siderophores in Figure 1.7 have been sequenced, facilitating investigations of the biosynthesis of this class of siderophore, as reviewed herein.

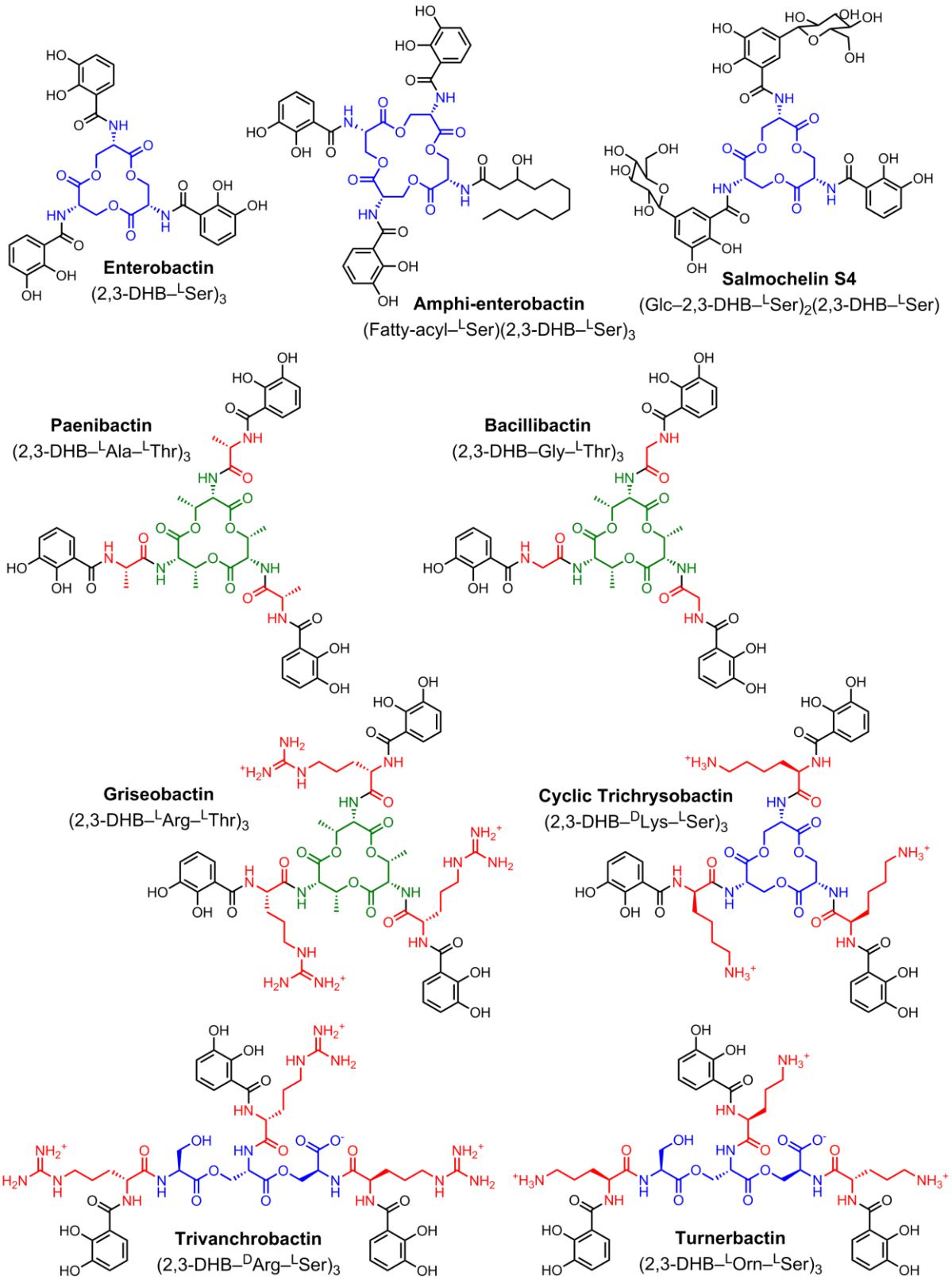


Figure 1.7. Triscatechol siderophores framed on a trimeric L-Ser or L-Thr backbone.

1.3.1.1. Enterobactin, salmochelins, and microcins

The biosynthesis of enterobactin by EntABCDEF has been studied extensively.[35] 2,3-Dihydroxybenzoate (DHB) is first synthesized and activated by EntABCE (Figure 1.8A). EntC is an isochorismate synthase, the N-terminus of EntB is an isochorismatase, and EntA is a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase. Synthesized 2,3-DHB is activated via EntE, a stand-alone A domain, and transferred to the C-terminal T domain of EntB. EntF, an iterative NRPS with a domain architecture of C-A-T-Te, is the core of enterobactin biosynthesis, catalyzing the formation of each ester and amide bond (Figure 1.8B). The A domain recognizes and adenylates L-Ser, which is transferred to the 4'-phosphopantetheinyl (Ppant) arm of the T domain. The N-terminal starter C domain of EntF binds DHB-bound EntB and catalyzes formation of the DHB–^{L}Ser amide bond. DHB–^{L}Ser is transferred to the thioesterase domain, which carries the monomer, dimer, and trimer, with each subsequent iteration, before catalyzing the macrocyclization of the trimer and release of the completed enterobactin siderophore. EntD is a 4'-phosphopantetheinyl transferase required to activate the T domains of EntF and EntB.

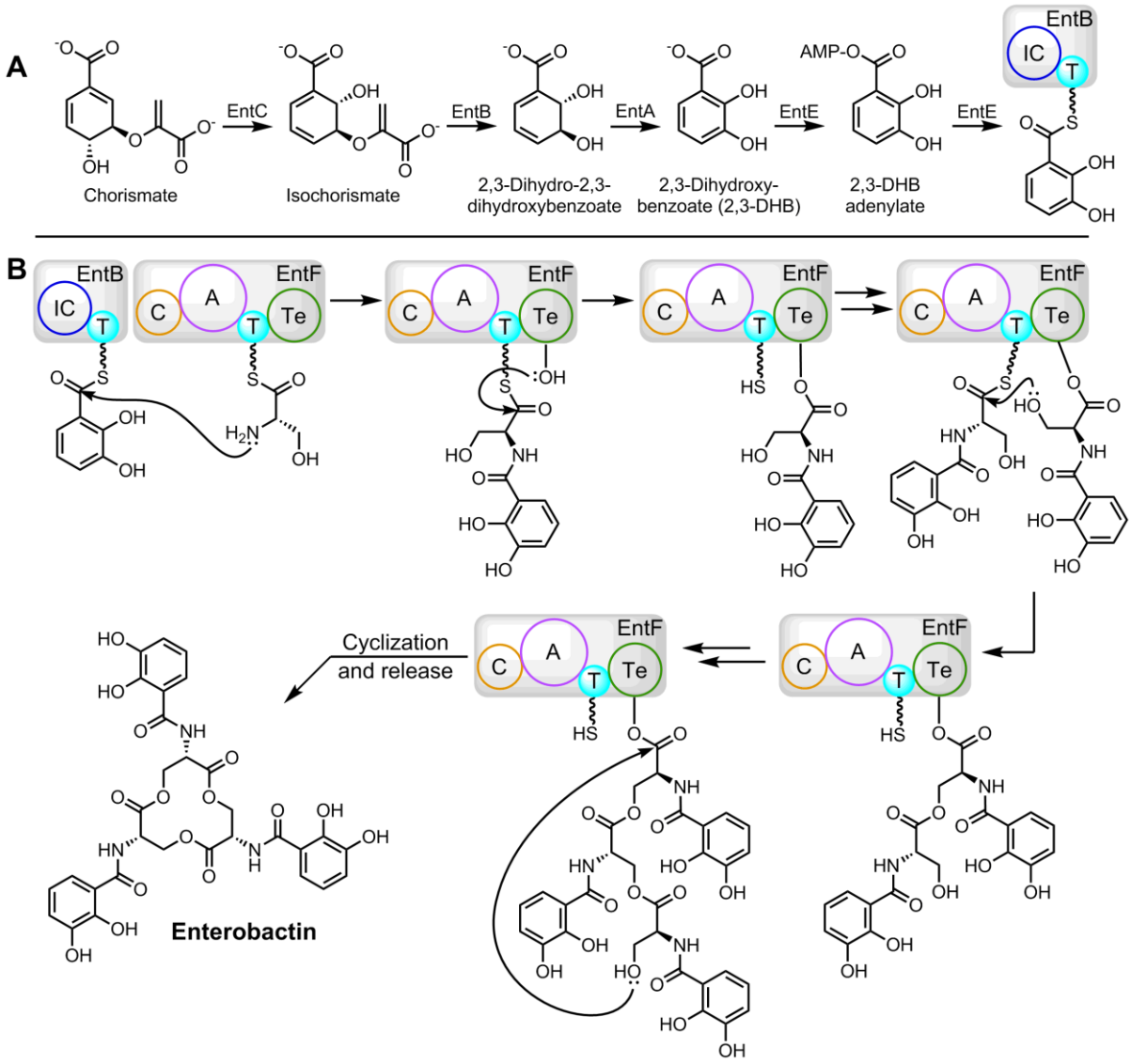


Figure 1.8. Biosynthesis of enterobactin. (A) Synthesis and activation of 2,3-dihydroxybenzoate (2,3-DHB) by EntABCE. (B) Assembly of enterobactin by NRPS EntF. IC – isochorismatase; C – condensation domain; A – adenylation domain; T – thiolation domain; Te – thioesterase domain.

Salmochelin S4, diglucosyl enterobactin, is produced by *Salmonella* spp. and certain strains of *E. coli* and *Klebsiella* spp.[79] Biosynthesis of salmochelin requires IroB, a C-glucosyl-transferase, in addition to the *ent* cluster. IroB catalyzes the formation of a C–C bond between the C5 carbanion of enterobactin and the anomeric C1' of UDP-glucose.[80,81] Glucosylation is distributive, with the successive buildup of mono-, di-, and triglucosyl enterobactin (MGE, DGE (salmochelin S4), and TGE, respectively) *in vitro* (Figure 1.9).[80] *In vivo*, MGE is only seen in certain strains, and TGE has not been reported, a consequence of differing relative rates of glucosylation and export.[80,82-84] Salmochelin S4 may be linearized by two esterases: periplasmic IroE cleaves once to yield salmochelin S2, while cytoplasmic IroD will continue to hydrolyze S2 to the monomers Glc–2,3-DHB–^{L}Ser (salmochelin SX) and 2,3-DHB–^{L}Ser (Figure 1.9).[85,86]

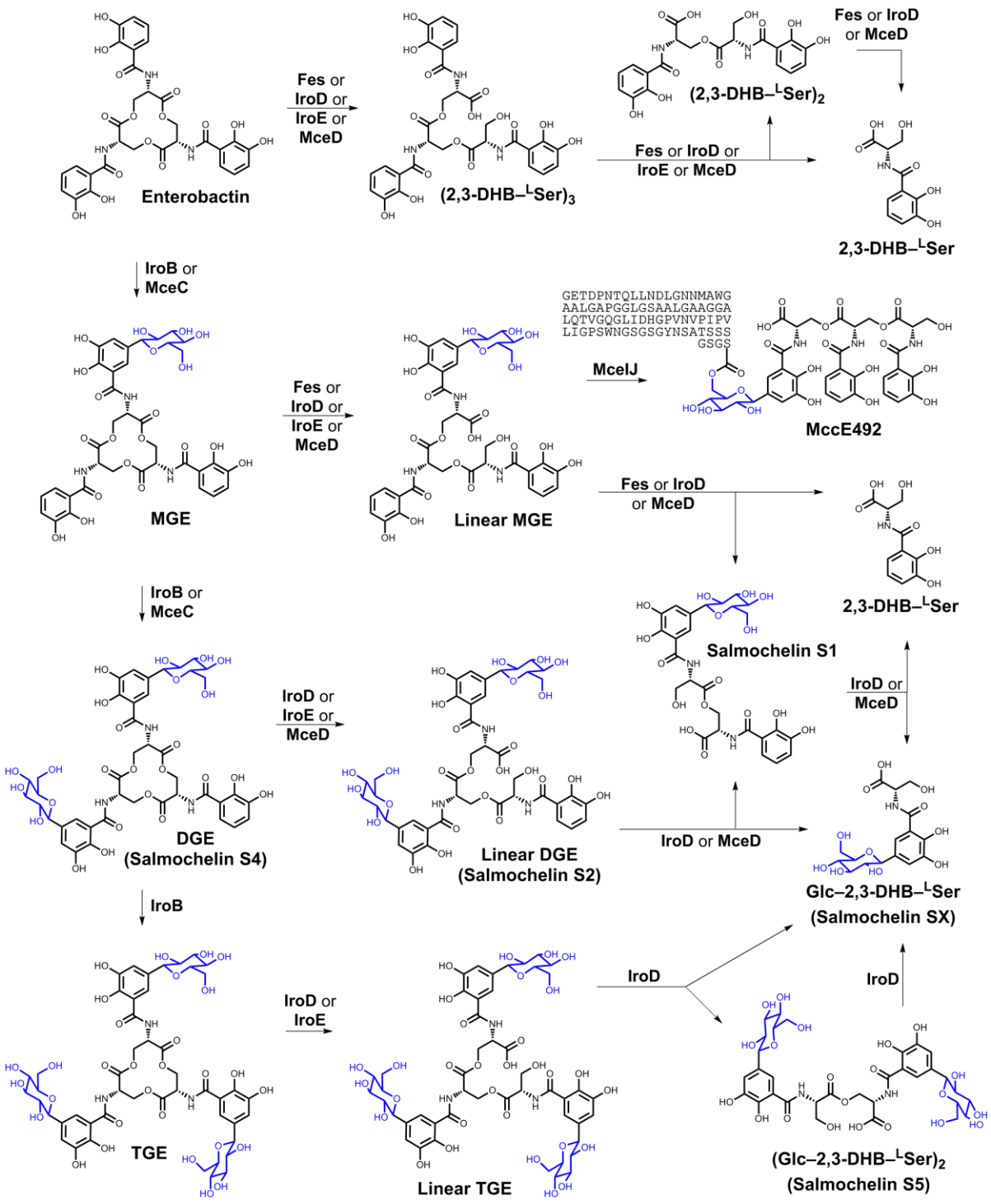


Figure 1.9. Post-synthetic modifications made to enterobactin.

Salmochelin S4 is believed to have evolved to evade host immune response. Siderocalin (Scn, also known as Lcn2, neutrophil-gelatinase-associated lipocalin, 24p3, and uterocalin) is a mammalian protein produced in response to bacterial infection to sequester siderophores and deprive bacteria of iron.[87,88] Although enterobactin and bacillibactin have two of the highest known Fe(III) affinities, neither is able to contribute to virulence because they are intercepted by Scn.[89] The shallow, positively charged calyx (binding site) of Scn can accommodate both enterobactin and bacillibactin, but salmochelin S4 evades sequestration through steric bulk.[87] Glucosylation of enterobactin thereby allows salmochelin-producing strains to retain virulence in the presence of Scn without sacrificing iron acquisition. [87,88,90]

Homologs of IroB and IroD (MceC and MceD) are found in the gene clusters of siderophore-microcins MccE492, MccM, MccH47, and MccI47 (Figure 1.9).[91] These pore-forming toxins feature a small (5-10 kDa) peptide post-translationally modified with an O-glycosidic bond to linear MGE. The linear enterobactin moiety, capable of binding iron, allows entry into the target cell through catecholate receptors Fiu, Cir, IroN, and FepA.[92,93] Biosynthesis of MccE492, like other siderophore-microcins, requires enterobactin.[94] C-glucosyltransferase MceC synthesizes MGE, but MGE is not an efficient substrate for glucosylation, and DGE (salmochelin S4) is not seen in MccE492 producing strain *Klebsiella pneumoniae RCY492*.[95] IroB or external salmochelin can complement MceC.[96] MGE is linearized by esterase MceD, and MceI and MceJ ligate the precursor peptide MceA to MGE.[95]

1.3.1.2. Amphi-enterobactin

Vibrio campbellii BAA-1116, formerly *V. harveyi*, produces an amphiphilic derivative of enterobactin known as amphi-enterobactin (Figure 1.7).[51] As in enterobactin, iron chelation is provided by three 2,3-DHB groups; however, amphi-enterobactin contains a fourth L-Ser in the cyclic backbone, which serves as an attachment point for a fatty acid tail. 3-hydroxydodecanoate (shown in Figure 1.7) is the most prevalent fatty acid, but 6 others, ranging C_{10} to C_{14}, were observed.[51] The biosynthetic cluster is similar to that of enterobactin, but requires AebG, a long-chain fatty acid CoA ligase (FACL, Figure 1.10). After activation, fatty acyl–CoA is condensed onto L-serine in the first iteration of the NRPS. The next three iterations proceed as in EntF, with three DHB–^{L}Ser monomers successively added to the thioesterase-bound fatty acyl–^{L}Ser, followed by the release of the cyclic tetralactone.[51]

The AebF condensation domain is therefore unique in recognizing two different donors: fatty acyl–CoA and 2,3-DHB–Ppant–AebB. *In vitro* studies confirmed that AebF reacts directly with fatty acyl–CoA in the absence of an acyl carrier protein. Similar reactivity is predicted for the N-terminal condensation domains of acyl peptidic siderophores amphibactin and cupriachelin,[97] and has been demonstrated *in vitro* for SrfAA, the first NRPS in the biosynthesis of surfactin, a cyclic lipopeptide antibiotic.[49] Surfactin biosynthesis does not have a dedicated FACL, but instead seems to use fatty acyl-CoA produced by primary fatty acid metabolism.[49] Knockout mutation of *aebG* disrupts amphi-enterobactin biosynthesis,[51] but the gene's effect on fatty acid metabolism has not been investigated. Recombinant AebF will not react with DHB–Ppant–AebB in the absence of fatty acyl–CoA, suggesting that fatty acyl–

CoA must be condensed in the first iteration before DHB–^{L}Ser condensation can occur.[51] A structural basis for this selectivity is not known, though the loading of fatty acyl–^{L}Ser on the thioesterase domain may cause a conformational shift that allows for the protein-protein interaction between AebF and AebB.

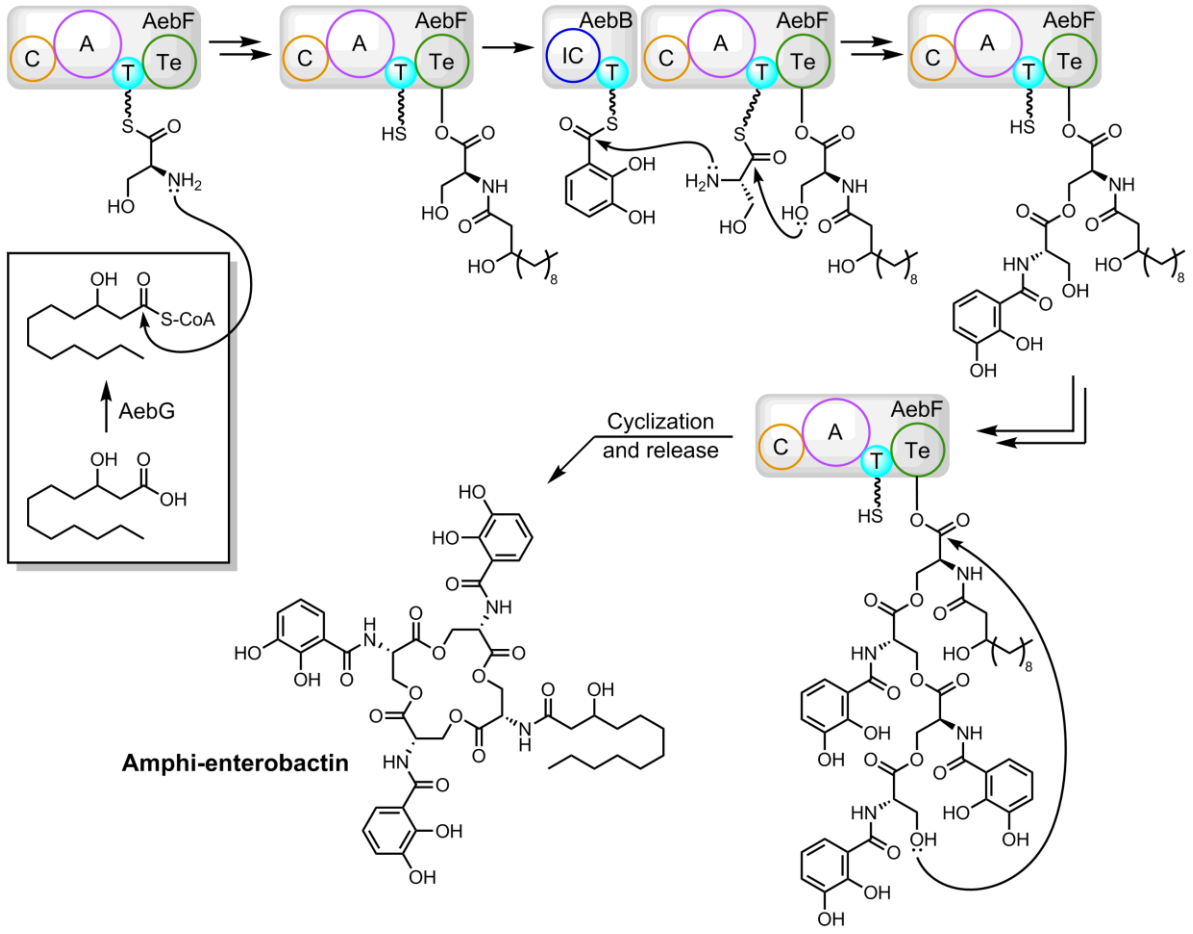


Figure 1.10. Proposed biosynthesis of amphi-enterobactins. AebG activates a fatty acid, which is condensed with L-serine in the first iteration of AebF. The next three iterations condense 2,3-DHB to L-serine as in enterobactin, followed by cyclization and release. IC – isochorismatase; C – condensation domain; A – adenylation domain; T – thiolation domain; Te – thioesterase domain.

1.3.1.3. Spacer-containing triscatechol siderophores

Bacillibactin (Figure 1.7), produced by several *Bacillus* species, including anthrax agent *B. anthracis*, was the first of a family of trimeric siderophores with elongated arms created by adding a spacer amino acid between 2,3-dihydroxybenzoate and the tri-serine or tri-threonine core, $(2,3\text{-DHB–XXX–Ser/Thr})_3$.[89,98] The cyclic trimer of 2,3-DHB–Gly–Thr was originally reported in 1997 from *Corynebacterium glutamicum* under the name "corynebactin"[99] and was rediscovered in 2001 from *B. subtilis* and named "bacillibactin".[100] Genomic and ^{55}Fe studies later showed that *C. glutamicum* does not produce a catechol siderophore.[101] Thus, "bacillibactin" is preferred, while the name "corynebactin" is now used for an unrelated citrate-based siderophore.[102]

A decade later, five related siderophores were discovered (Figure 1.7): griseobactin (also called streptobactin) from *Streptomyces spp.*,[75,76] trivanchrobactin from *Vibrio campbellii* DS40M4,[77] cyclic trichrysobactin from *Dickeya chrysanthemi* EC16,[73] paenibactin from *Paenibacillus elgii* B69,[74] and turnerbactin from *Teredinibacter turnerae* T7901,[78] all of which share the $(2,3\text{-DHB–XXX–Ser/Thr})_3$ structural motif.

The biosyntheses of these amino acid spacer-containing triscatechol siderophores are similar to the biosynthesis of enterobactin and require homologs of *entABCDEF*. These genes are clustered alongside genes required for siderophore export and uptake (Figure 1.11). In griseobactin producer *Streptomyces sp.* ATCC 700974, a premature stop codon separates EntB bifunctionality into two separate proteins: the isochorismatase DhbB, and the acyl carrier protein DhbG.[75] The structural variation in these triscatechol siderophores is provided by the homologs of EntF: each contains two domains that determine the spacer and the ester core (Figure 1.6). The

biosynthetic clusters have undergone significant rearrangement; whole cluster phylogenetic analysis may shed light on their evolutionary history.

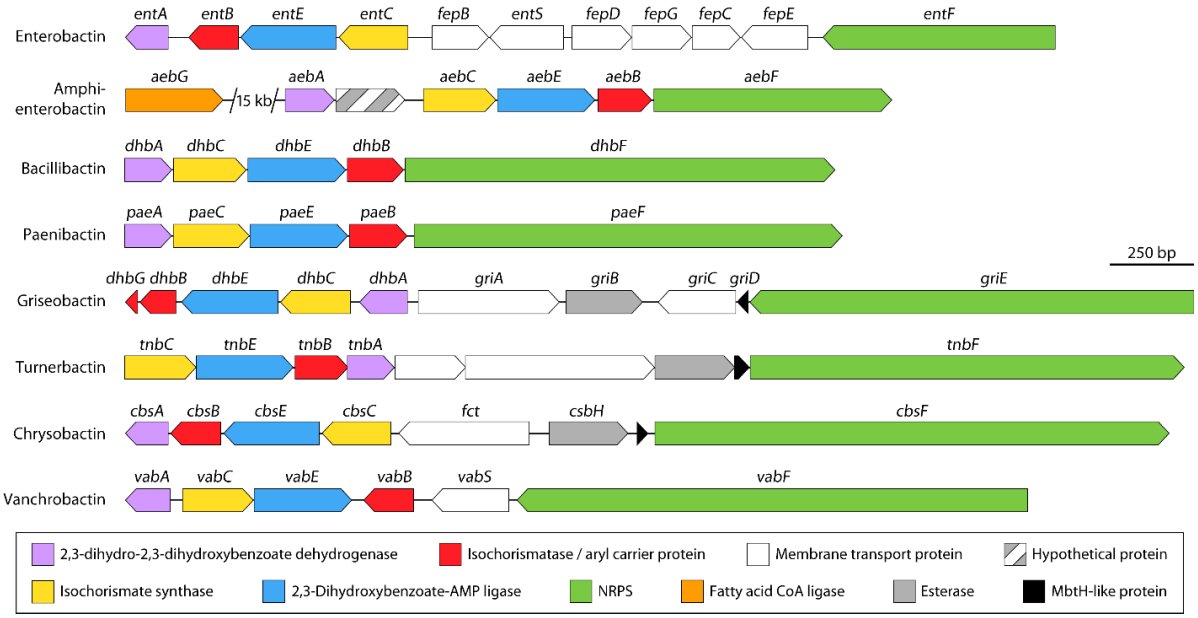


Figure 1.11. Triscatechol siderophore gene clusters. Each cluster is drawn to scale and truncated to core biosynthesis genes. Arrows represent the direction of transcription. Genomic data were retrieved from GenBank: enterobactin – *Escherichia coli* K12, amphi-enterobactin – *Vibrio campbellii* BAA-1116, bacillibactin – *Bacillus subtilis* 168, paenibactin – *Paenibacillus elgii* B69, griseobactin – *Streptomyces* sp. ATCC 700974, turnerbactin – *Teredinibacter turnerae* T7901, chrysobactin – *Dickeya dadantii* 3937, vanchrobactin – *Vibrio anguillarum* RV22. Chrysobactin producer *D. dadantii* 3937 and vanchrobactin producer *V. anguillarum* RV22 were used in lieu of published complete genomes of cyclic trichrysobactin producer *D. chrysanthemi* EC16 and trivanchrobactin producer *V. campbellii* DS40M4.

Each biosynthesis also requires a homolog of Ppant transferase EntD. However, NRPS-related Ppant transferases are not substrate specific, and one gene may be used for multiple secondary metabolites.[103,104] The biosynthesis of bacillibactin requires the Ppant transferase gene *sfp,* located in the gene cluster of non-ribosomal peptide surfactin.[103,105,106] *Bacillus subtilis* type strain 168 and derived strains carry a frameshift mutation in *sfp*, and are negative for both bacillibactin and surfactin production.[103,104] Of the spacer-containing siderophores, only the vanchrobactin locus contains a Ppant transferase.

The biosynthesis of cyclic trichrysobactin by the *cbs* operon is shown in Figure 1.12. 2,3-DHB is synthesized and activated by CsbABCE and loaded onto the Ppant arm of CbsB. CbsF adenylation domain A_1 activates the spacer amino acid L-Lys, and the N-terminal condensation domain forms the DHB–^{L}Lys amide bond. L-Lys is then converted to D-Lys by an epimerization domain, which is found only in cyclic trichrysobactin and trivanchrobactin NRPSs. Domain A_2 activates L-Ser, and the internal condensation domain catalyzes the formation of the 2,3-DHB–^{D}Lys–^{L}Ser monomer. The monomer is transferred to the C-terminal thioesterase domain, and this process is repeated twice more to yield thioesterase-bound (2,3-DHB–^{D}Lys–^{L}Ser)$_3$. Intramolecular addition forms the trilactone core and releases cyclic trichrysobactin. In some strains, the siderophores are instead released as the DHB–XXX–Ser/Thr monomer or the (DHB–XXX–Ser/Thr)$_3$ linear trimer (*vide infra*).

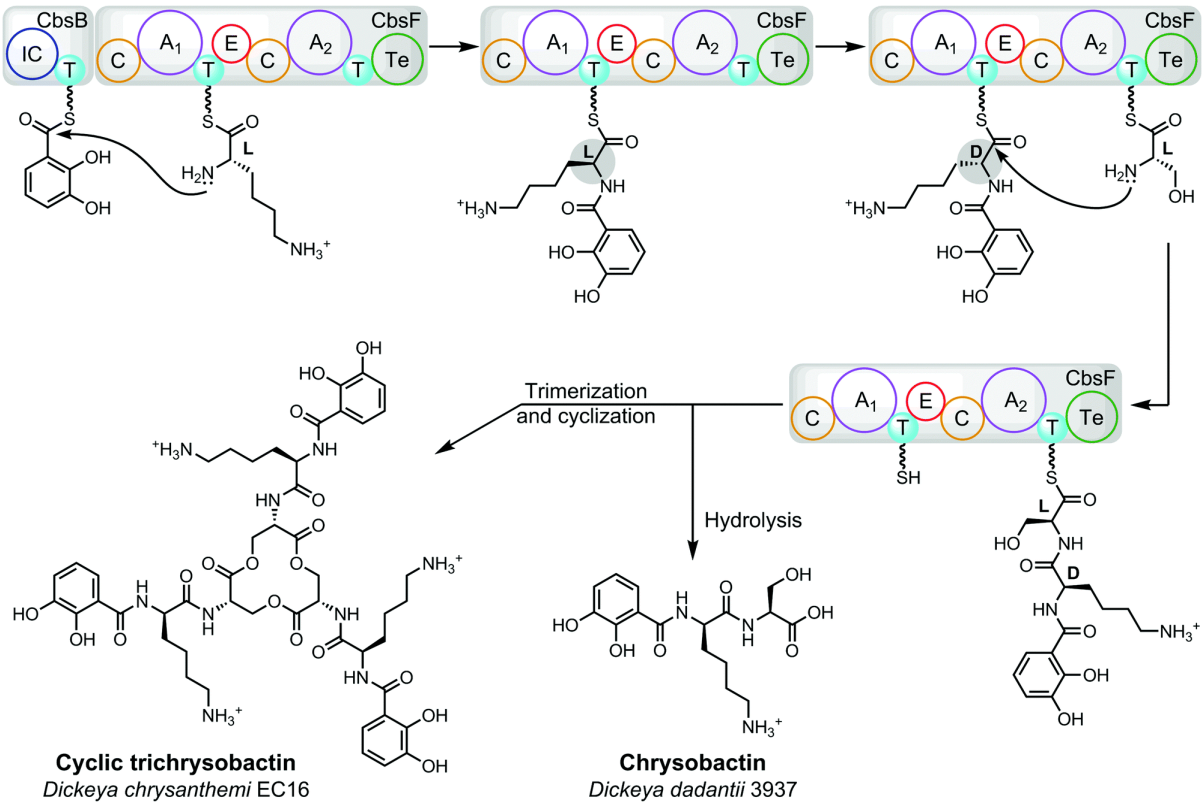


Figure 1.12. Proposed biosyntheses of cyclic trichrysobactin and chrysobactin. IC – isochorismatase; C – condensation domain; A – adenylation domain; T – thiolation domain; E – epimerization domain; Te – thioesterase domain.

NRPS domain analysis[47,107] shows L-threonine-based siderophores bacillibactin, paenibactin, and griseobactin NRPSs each contain an $^{L}C_{L}$ domain, while serine-based siderophores turnerbactin, cyclic trichrysobactin, and trivanchrobactin NRPSs each contain a $^{D}C_{L}$ domain (Figure 1.13). However, the turnerbactin NRPS, TnbF, lacks an epimerization domain; accordingly, only cyclic trichrysobactin and trivanchrobactin include D-amino acids.[73,77,78] Evidence of a lost epimerization domain can also be seen in the turnerbactin thiolation domain, which contains a conserved GGDSI motif (Figure 1.14A) that is required for epimerization domain interaction.[108] Thiolation domains followed by $^{L}C_{L}$, including those of bacillibactin, paenibactin, and

griseobactin, share the motif GGHSL.[108] Condensation domains have been found to be highly stereoselective,[43,109] but the TnbF domain has lost this specificity. Similar convergent evolution of $^{D}C_{L}$ domains toward $^{L}C_{L}$ activity has previously been seen in glycopeptide NRPSs.[47]

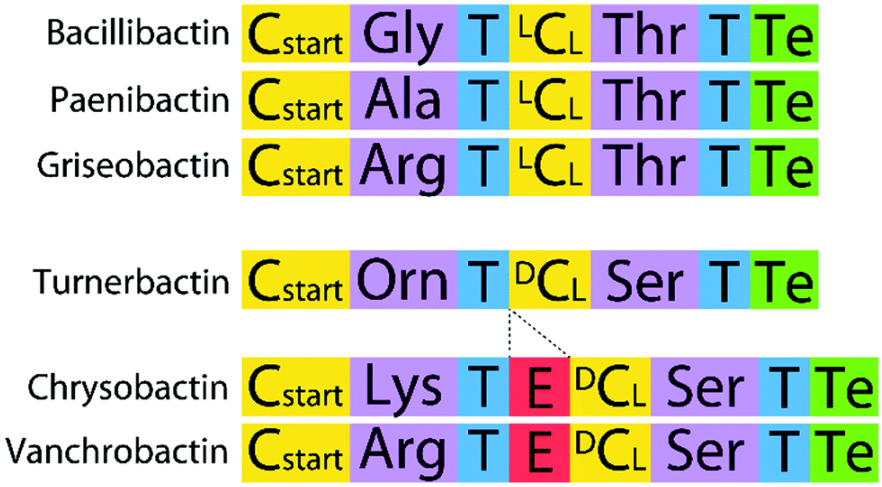


Figure 1.13. Domain architecture of NRPSs involved in the biosynthesis of spacer-containing triscatechol siderophores. Each adenylation domain is represented by the selected amino acid. Domains were identified using the PKS/NRPS Analysis Web Server. See Figure 1.11 caption for a list of strains.

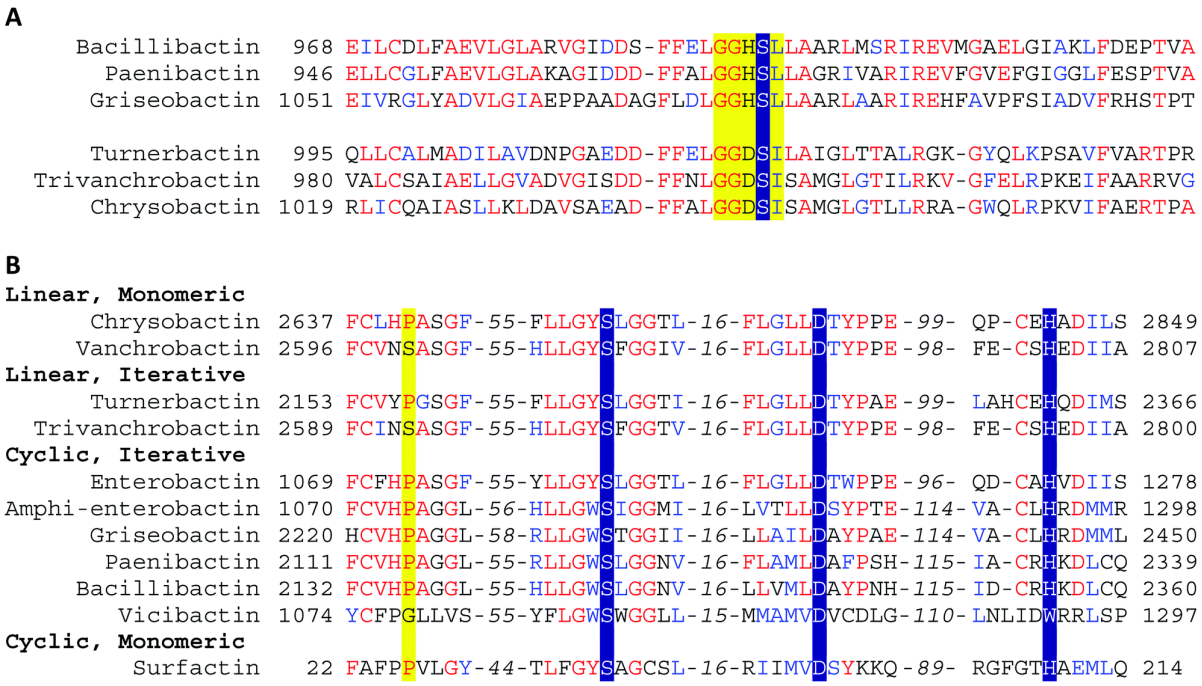


Figure 1.14. Multisequence alignments of (A) NRPS T_1 domains, and (B) NRPS Te domains. Blue backgrounds indicate catalytic residues; yellow backgrounds indicate residues found to influence reactivity, as described in the text. Numbering is relative to the whole protein, with the exception of surfactin, which refers to the Te domain excised by Bruner *et al.*[110] Modules were extracted using the PKS/NRPS Analysis Web Server;[107] alignments were made using MUSCLE (EMBL-EBI).[111,112]

1.3.1.4. Monomeric and dimeric spacer-containing siderophores

The EntF thioesterase domain is an efficient catalyst for trimerization and cyclization, and recombinant EntF produced only enterobactin *in vitro*.[113,114] This result suggests that any 2,3-DHB–Ser monomer, dimer, or linear trimer produced *in vivo* is a product of enzymatic or non-enzymatic hydrolysis. The same is presumed to be true of the other trilactone siderophores (bacillibactin, griseobactin, cyclic trichrysobactin, and paenibactin), though selectivity for cyclization has not been confirmed *in vitro*.[73-75,100] By contrast, cyclic forms of turnerbactin and

trivanchrobactin have not been found;[77,78] moreover, in certain strains, only 2,3-DHB–XXX–Ser/Thr monomers have been found. The best studied of these monocatechol siderophores is chrysobactin (2,3-DHB–^{D}Lys–^{L}Ser, Figure 1.15),[115-119] a virulence factor *in planta*.[120-124] Chrysobactin was first found in plant pathogen *Dickeya dadantii* 3937 (formerly *Erwinia chrysanthemi*) and later in *Serratia marcescens*, *Pseudomonas luteola*, and *D. carotovora* subsp. *carotovora*.[125-128] Three other 2,3-DHB–XXX–Ser/Thr monomers have been found without oligomeric counterparts: vanchrobactin (2,3-DHB–^{D}Arg–^{L}Ser) from strains of marine fish pathogen *Vibrio anguillarum,*[129-133] bacillibactin monomer S_{VK21} (2,3-DHB–Gly–^{L}Thr) from thermoresistant *Bacillus licheniformis* VK21,[134] and griseobactin monomer benarthin (2,3-DHB–^{L}Arg–^{L}Thr) from *Streptomyces xanthophaeus* MJ244-SF1 (Figure 1.15).[135,136] Benarthin was only investigated as a pyroglutamyl peptidase inhibitor,[135] but both vanchrobactin and S_{VK21} bind Fe(III) and promote growth in their respective strains.[129,134]

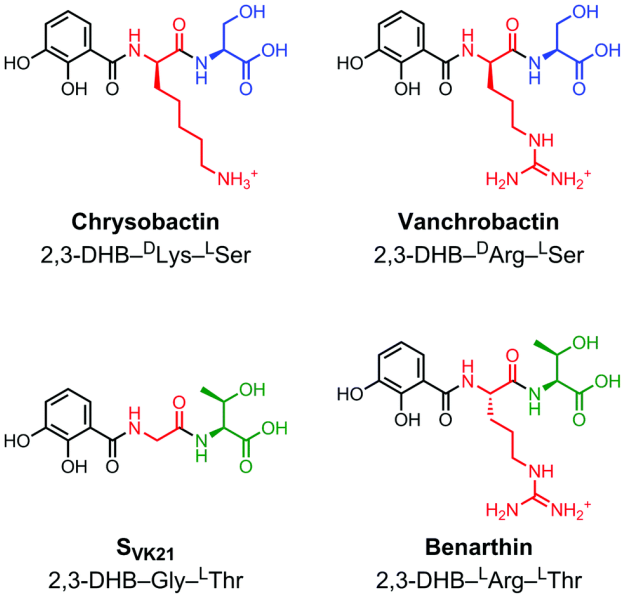


Figure 1.15. Reported monocatechol siderophores of the motif DHB–XXX–^{L}Ser/^{L}Thr.

Given the superior Fe(III) affinity of trilactone triscatechol siderophores and the fidelity of EntF, it is tempting to suggest that monomers and linear trimers are merely hydrolysis products. However, nonenzymatic hydrolysis during siderophore extraction is unlikely to be the cause in most cases. The threonine trilactone core of bacillibactin was found to be highly resistant to hydrolysis; the cyclic ester remained mostly intact after stirring in 6M HCl for 12 hours.[12] Under the same conditions, enterobactin hydrolyzed, but with significant amounts of dimer and trimer remaining.[12] This precludes the possibility that any reasonable workup would completely hydrolyze the trilactones to their monomers. To determine if turnerbactin was hydrolyzed to the linear trimer during workup, Han *et al*. performed a side-by-side extraction of *T. turnerae* T7901 and *D. chrysanthemi* EC16 under identical conditions (XAD-2 adsorption, MeOH/H_2O extraction, HPLC purification).[78] Monomers, dimers, and trimers were seen in both cultures, but only trichrysobactin was found in cyclic form.[78] Such a control experiment has not yet been performed with a monomer-producing strain, but a similar extraction of *S. marcescens* with XAD-2 resin yielded only chrysobactin and serratiochelin.[126]

These "precursor" siderophores could be the final product of their respective biosyntheses if oligomerization or cyclization is interrupted (Figure 1.12). The factors that determine different thioesterase activities—iteration, hydrolysis, and cyclization—are not well understood, but certain features have been shown to play a role. Thioesterase activity requires a conserved Ser–His–Asp catalytic triad (in EntF: Ser1138, His1271, Asp1165; Figure 1.14B), where Ser1138 is the nucleophile that holds the DHB–Ser precursors.[114] Mutation of this serine (Ser1138→Ala) stops all oligomerization, and DHB–Ser monomer is slowly released with a ~1500–3000-fold

drop in turnover.[113,114] With the loss of Ser1138, DHB–Ser is trapped on the Ppant arm of the thiolation domain and is slowly released by hydrolysis. A His1271→Ala mutant was still able to produce enterobactin with a ~10,000-fold drop in turnover.[113,114] Catalytic base His1271 is required for efficient turnover, but the precursors held in the thioesterase domain are sufficiently shielded from water to allow oligomerization and cyclization to occur.

To prevent hydrolysis, thioesterases of EntF and other cyclizing NRPSs contain a "lid" structure that excludes water from the active site (Figure 1.16).[137] The lid encloses a hydrophobic pocket, which accommodates the growing tripeptide and holds it in a bent conformation that encourages macrocyclization.[110,137] To investigate the cyclization of surfactin, a cyclic non-ribosomal peptide produced by *B. subtilis*, Tseng *et al.* focused on a proline residue adjacent to the oxyanion hole.[42] The residue is conserved among NRPS thioesterases, while lipases, which only catalyze hydrolysis, contain a conserved glycine.[42] Mutation of this proline in the surfactin thioesterase (Pro26→Gly) reversed selectivity from cyclization to hydrolysis (12-fold change in ratio).[42] When N,N-dimethylformamide is used as a solvent, the Pro26→Gly mutant and the wildtype thioesterase catalyze cyclization with similar efficiencies, and no hydrolysis is seen.[138] Pro26 is not required for cyclization *sensu stricto*, but likely provides rigidity to the thioesterase, whereas the conformational flexibility of glycine allows water to bypass the lid and enter the active site, resulting in the release of a linear peptide. A multisequence alignment of triscatechol siderophore thioesterase domains (Figure 1.14B) shows that serine replaces the conserved proline in the NRPS VabF responsible for linear trivanchrobactin, providing a possible explanation for the absence of cyclized product.[139] On the

contrary, proline is conserved in TnbF (linear turnerbactin). Furthermore, vicibactin is a trilactone hydroxamate siderophore synthesized by an iterative NRPS (Figure 1.17)[140] that contains the glycine supposedly indicative of hydrolysis. Thus, while this residue may influence cyclization, other factors are in play, and further mutational studies are needed help identify the structural basis for iteration and cyclization in these triscatechol NRPS. Mutating the conserved proline of EntF (Pro1073) to glycine may result in the release of linear DHB–Ser trimer, which has not been reported in other EntF mutants.[114,141]

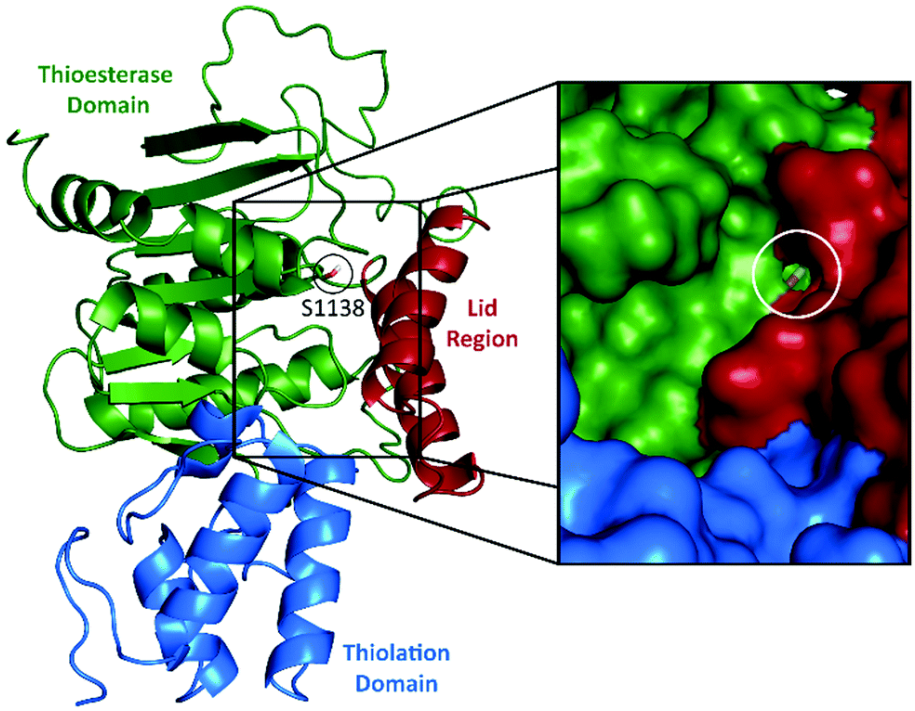


Figure 1.16. Structure of the T–Te portion of EntF. The Te domain (green) contains a lid region (red) that shields the active site from water. Catalytic Ser1138 is circled.

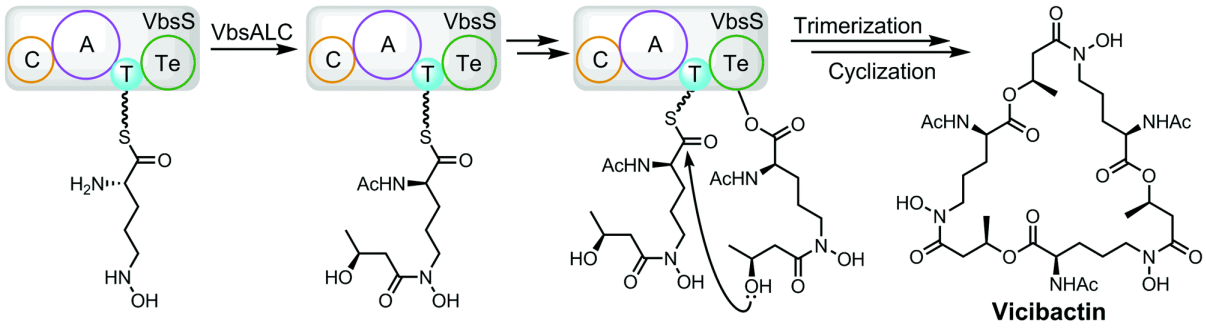


Figure 1.17. Proposed iterative NRPS biosynthesis of the trilactone vicibactin, a hydroxamate siderophore. IC – isochorismatase; C – condensation domain; A – adenylation domain; T – thiolation domain; Te – thioesterase domain.

1.3.1.5. Perspective on Ser- and Thr-based triscatechol siderophores

Enterobactin is a seemingly perfect Fe(III) chelator, but three factors limit its effectiveness in certain environments: unfavorable partitioning into membranes,[90] sequestration by mammalian siderocalin,[87] and piracy by competing bacteria.[142-144] The family of triscatechol siderophores in Figure 1.7 represents a natural biocombinatoric library based on the enterobactin scaffold. Divergent evolution has modified the trilactone serine backbone, glucosylated the catechol binding groups, and even inserted additional amino acids, creating an array of siderophores with unique chemical and biological properties.

Linearizing the trilactone core, whether through hydrolysis or biosynthesis, increases hydrophilicity at the expense of Fe(III) affinity,[90,145] while expansion to a tetralactone scaffold allows for further tailoring of hydrophobicity by the addition of a fatty acid tail.[51] Incorporation of threonine in place of serine in the macrolactone core both provides resistance to non-enzymatic hydrolysis and prevents enzymatic hydrolysis by competing bacteria.[12] C5-glucosylation of the catechol prevents

sequestration of salmochelins by siderocalin, while simultaneously improving Fe(III) acquisition rates in membrane-rich environments.[87,90]

Structural expansion by the insertion of a spacer amino acid between the macrolactone scaffold and DHB provides a handle for further variation. Though the glycine spacer of bacillibactin slightly decreases Fe(III) affinity by increasing conformational freedom,[146] the elongated ferric complex cannot be transported through enterobactin-specific receptor FepA, preventing piracy;[12] other amino acid spacers likely provide the same effect. The significance of D or L spacer configuration is currently not known, but it may prevent piracy by inhibiting siderophore uptake or by preventing Fe(III) release; cross-feeding experiments will provide more insight here. Siderocalin is able to bind bacillibactin, but larger side chains may provide enough steric bulk to evade sequestration. Notably, while enterobactin and bacillibactin are neutral compounds which gain a -3 charge upon Fe(III) complexation, cationic-spacer containing siderophores form neutral iron complexes. In addition to affecting solubility and membrane partitioning, this charge difference may disrupt the coulombic and cation−π interactions required for siderocalin binding.[147]

The array of siderophore structures described herein hints at potential undiscovered variations on the enterobactin scaffold. Despite differences in structure, all of the known triscatechol siderophores share key biosynthetic elements: DHB biosynthesis genes homologous to *entABCE* and a NRPS homolog of *entF* that begins with a C_{start} domain and ends with a module selective for serine or threonine (Figures 1.11 and 1.13). These conserved elements are prime targets for genome mining. Indeed, genomic analysis has led to the discoveries of paenibactin,[74] turnerbactin,[78]

and amphi-enterobactins.[51] While genome mining certainly confers powerful predictive advantages, current genome mining tools are not able to replace the process of isolation and characterization of siderophores or other natural products: adenylation domain analysis failed to predict the spacer amino acids of paenibactin and turnerbactin,[74,78] the amphi-enterobactin locus *aeb* was predicted to encode for enterobactin,[51] and the factors determining thioesterase activity remain unknown. Triscatechol siderophores, though structurally diverse, share a conserved genetic scaffold; therefore, they can provide unique insights into how siderophores are synthesized and evolve.

1.3.2. Other catechol siderophores

Three classes of enzymes in the condensation domain superfamily have been reported to be responsible for the incorporation of NRPS-bound 2,3-DHB into siderophore structures: C_{start} domains, heterocyclization (Cy) domains, and VibH-like domains. C_{start} domains are involved not only in serine- and threonine-based triscatechol siderophore biosynthesis (see above), but also in the biosyntheses of a variety of other siderophores where 2,3-DHB is condensed to an amino acid, the ε-nitrogen of Lys, or even the Ω-nitrogen of Arg (Figure 1.18). A smaller set of reported siderophore biosyntheses features a Cy domain, which catalyzes the condensation of 2,3-DHB to Ser, Thr, or Cys, as well as the subsequent cyclodehydration to form a oxazoline, methyloxazoline, or thiazoline, respectively (Figure 1.18).[148] A third class of catechol siderophores is based on a diamine or polyamine core. DHB–amine bond formation is catalyzed by a family of standalone NRPS C domains typified by VibH of vibriobactin biosynthesis, which transfers Ppant-bound DHB to a primary amine of norspermidine.[149]

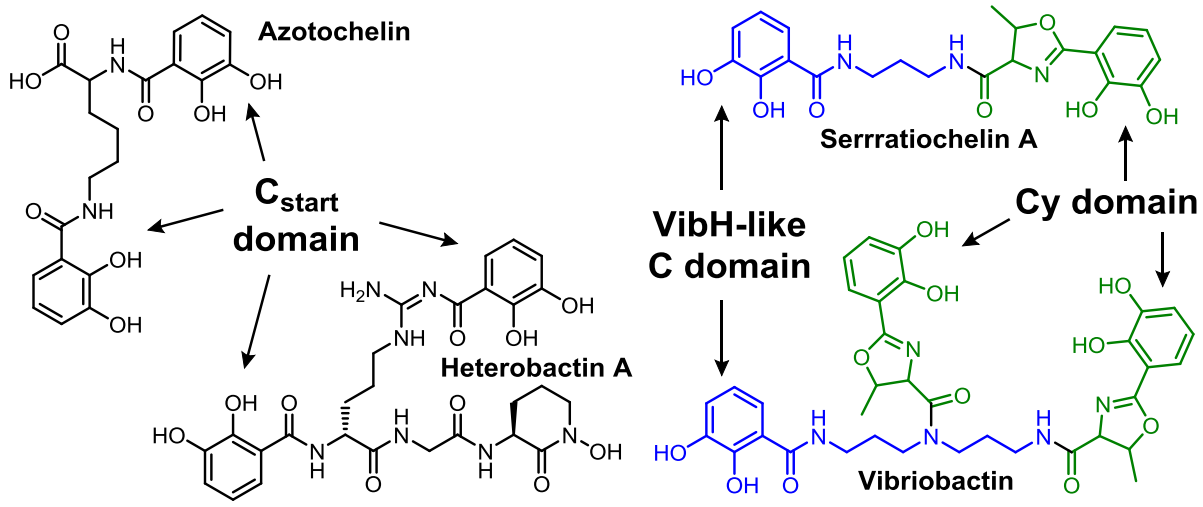


Figure 1.18. Representative siderophores produced by each of the three condensation domain subtypes responsible for siderophore DHB incorporation.

1.4. A primer to select bioinformatic techniques

The work herein depends on several publicly-available bioinformatic tools. Full explanations of their utility can be found in their respective publications, as well as in bioinformatics tutorials aimed at genome miners.[150-152]

1.4.1. BLAST: Pairwise alignments

Two genomic sequences are considered homologous if they share a common ancestor.[153] Homologs are created during either a speciation event (orthologs) or during a gene duplication event (paralogs). Over evolutionary time, the two sequences diverge as mutations accumulate. Homology can be inferred when sequences have a degree of similarity in their sequences that is unlikely to have arisen by chance. Pairwise alignment tools such as BLAST (Basic Local Alignment Search Tool) determine if the "excess similarity" between two sequences is statistically significant.[154]

1.4.2. MUSCLE: Multiple Sequence Alignments

Not all residues within an enzyme are tolerant of mutations: loss of a catalytic base, metal binding site, or structural disulfide bond may disrupt enzyme function. These crucial residues are therefore expected to be strictly conserved across all functional representatives of the protein family. In contrast, residues distant from the catalytic site(s) may be free to drift over time. A multiple sequence alignment of three or more homologous sequences not only allows for the visualization of conserved residues, but also serves as a starting point for more advanced enzyme profiles and phylogenetic reconstruction (see below). Multiple sequence alignments are much more computationally intensive than pairwise alignments, and different alignment algorithms make different trade-offs of time and accuracy. The current work primarily uses MUSCLE (MUltiple Sequence Comparison by Log-Expectation), a popular algorithm that generates alignments with good accuracy and speed.[111]

1.4.3. HMMER and PFAM: Profile hidden Markov Models

The Universal Protein Resource Knowledgebase (UniProtKB)[155] contains a staggering 180 million proteins (October 2019).[156] Studying every protein of interest *in vitro* or *in vivo* is simply not possible; indeed, only ~150,000 proteins in the UniProtKB database have experimental evidence supporting their existence.[156] Instead, the majority of protein function is proposed using amino acid sequence homology. Pairwise alignment tools like BLAST are not suited to finding distant homologs or classifying diverse protein families, as they cannot place any emphasis on conserved residues. Profile hidden Markov models (pHMMs) are position-specific probabilistic models that can be generated from multiple

sequence alignment with HMMER.[157] The resulting pHMM can then be compared to an amino acid sequence of interest using HMMER to determine the probability that the sequence belongs to the family. Several large databases of pHMMs have been built to aid in protein classification and automated genome annotation; the work herein uses Pfam,[158] which integrates well with the EMBL-EBI HMMER Web Server.[159] Other protein family databases include the Conserved Domain Database (CDD),[160] TIGRFAMs,[161] and InterPro.[162]

1. Z. L. Reitz, M. Sandy and A. Butler, Biosynthetic considerations of triscatechol siderophores framed on serine and threonine macrolactone scaffolds, *Metallomics*, 2017, **9**, 824-839.
2. M. Miethke and M. A. Marahiel, Siderophore-based iron acquisition and pathogen control, *Microbiol. Mol. Biol. Rev.*, 2007, **71**, 413-451.
3. G. n. Winkelmann and V. Braun, Stereoselective recognition of ferrichrome by fungi and bacteria, *FEMS Microbiol. Lett.*, 1981, **11**, 237-241.
4. B. F. Matzanke, G. I. Müller and K. N. Raymond, Hydroxamate siderophore mediated iron uptake in E. coli: stereospecific recognition of ferric rhodotorulic acid, *Biochem. Biophys. Res. Commun.*, 1984, **121**, 922-930.
5. R. J. Bergeron, J. B. Dionis, G. T. Elliott and S. J. Kline, Mechanism and stereospecificity of the parabactin-mediated iron-transport system in Paracoccus denitrificans, *J. Biol. Chem.*, 1985, **260**, 7936-7944.
6. M. Münzinger, K. Taraz, H. Budzikiewicz, H. Drechsel, P. Heymann, G. Winkelmann and J.-M. Meyer, S, S-rhizoferrin (enantio-rhizoferrin) – a siderophore of Ralstonia (Pseudomonas) pickettii DSM 6297 – the optical antipode of R, R-rhizoferrin isolated from fungi, *Biometals*, 1999, **12**, 189-193.
7. Z. A. Youard, N. Wenner and C. Reimmann, Iron acquisition with the natural siderophore enantiomers pyochelin and enantio-pyochelin in Pseudomonas species, *Biometals*, 2011, **24**, 513-522.
8. G. Winkelmann, Evidence for stereospecific uptake of iron chelates in fungi, *FEBS Lett.*, 1979, **97**, 43-46.
9. R. J. Bergeron and W. R. Weimar, Kinetics of iron acquisition from ferric siderophores by Paracoccus denitrificans, *J. Bacteriol.*, 1990, **172**, 2650-2657.
10. J. B. Neilands, T. J. Erickson and W. H. Rastetter, Stereospecificity of the ferric enterobactin receptor of Escherichia coli K-12, *J. Biol. Chem.*, 1981, **256**, 3831-3832.
11. P. Thulasiraman, S. M. Newton, J. Xu, K. N. Raymond, C. Mai, A. Hall, M. A. Montague and P. E. Klebba, Selectivity of ferric enterobactin binding and cooperativity of transport in gram-negative bacteria, *J. Bacteriol.*, 1998, **180**, 6689-6696.

12. R. J. Abergel, A. M. Zawadzka, T. M. Hoette and K. N. Raymond, Enzymatic hydrolysis of trilactone siderophores: where chiral recognition occurs in enterobactin and bacillibactin iron transport, *J. Am. Chem. Soc.*, 2009, **131**, 12682-12692.
13. H. Drechsel, G. Jung and G. Winkelmann, Stereochemical characterization of rhizoferrin and identification of its dehydration products, *Biometals*, 1992, **5**, 141-148.
14. C. D. Cox, K. L. Rinehart, Jr., M. L. Moore and J. C. Cook, Jr., Pyochelin: novel structure of an iron-chelating growth promoter for Pseudomonas aeruginosa, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 4256-4260.
15. Z. A. Youard, G. L. A. Mislin, P. A. Majcherczyk, I. J. Schalk and C. Reimmann, Pseudomonas fluorescens CHA0 Produces Enantio-pyochelin, the Optical Antipode of the Pseudomonas aeruginosa Siderophore Pyochelin, *J. Biol. Chem.*, 2007, **282**, 35546-35553.
16. A. R. Tyler, H. Mosaei, S. Morton, P. G. Waddell, C. Wills, W. McFarlane, J. Gray, M. Goodfellow, J. Errington, N. Allenby, N. Zenkin and M. J. Hall, Structural Reassignment and Absolute Stereochemistry of Madurastatin C1 (MBJ-0034) and the Related Aziridine Siderophores: Madurastatins A1, B1, and MBJ-0035, *J. Nat. Prod.*, 2017, **80**, 1558-1562.
17. J.-X. Yan, M. G. Chevrette, D. R. Braun, M. K. Harper, C. R. Currie and T. S. Bugni, Madurastatin D1 and D2, Oxazoline Containing Siderophores Isolated from an Actinomadura sp, *Org. Lett.*, 2019, **21**, 6275-6279.
18. K. Brillet, C. Reimmann, G. L. A. Mislin, S. Noël, D. Rognan, I. J. Schalk and D. Cobessi, Pyochelin Enantiomers and Their Outer-Membrane Siderophore Transporters in Fluorescent Pseudomonads: Structural Bases for Unique Enantiospecific Recognition, *J. Am. Chem. Soc.*, 2011, **133**, 16503-16509.
19. Z. A. Youard and C. Reimmann, Stereospecific recognition of pyochelin and enantio-pyochelin by the PchR proteins in fluorescent pseudomonads, *Microbiology*, 2010, **156**, 1772-1782.
20. C. Reimmann, Inner-membrane transporters for the siderophores pyochelin in Pseudomonas aeruginosa and enantio-pyochelin in Pseudomonas fluorescens display different enantioselectivities, *Microbiology*, 2012, **158**, 1317-1324.
21. H. Drechsel, H. Stephan, R. Lotz, H. Haag, H. Zähner, K. Hantke and G. Jung, Structure elucidation of yersiniabactin, a siderophore from highly virulentYersinia strains, *Liebigs Ann.*, 1995, **1995**, 1727-1733.
22. H. Hayen and D. A. Volmer, Different iron-chelating properties of pyochelin diastereoisomers revealed by LC/MS, *Anal. Bioanal. Chem.*, 2006, **385**, 606-611.
23. O. Sasaki, Y. Igarashi, N. Saito and T. Furumai, Watasemycins A and B, new antibiotics produced by Streptomyces sp. TP-A0597, *J. Antibiot.*, 2002, **55**, 249-255.
24. K. Shindo, A. Takenaka, T. Noguchi, Y. Hayakawa and H. Seto, Thiazostatin A and thiazostatin B, new antioxidants produced by Streptomyces tolurosus, *J. Antibiot.*, 1989, **42**, 1526-1529.
25. Y. Inahashi, M. Iwatsuki, A. Ishiyama, M. Namatame, A. Nishihara-Tsukashima, A. Matsumoto, T. Hirose, T. Sunazuka, H. Yamada, K. Otoguro, Y. Takahashi, S. Omura and K. Shiomi, Spoxazomicins A-C, novel antitrypanosomal alkaloids produced by an endophytic actinomycete, Streptosporangium oxazolinicum K07-0460(T), *J. Antibiot.*, 2011, **64**, 303-307.

26. N. Liu, F. Shang, L. Xi and Y. Huang, Tetroazolemycins A and B, two new oxazole-thiazole siderophores from deep-sea Streptomyces olivaceus FXJ8.012, *Mar. Drugs*, 2013, **11**, 1524-1533.
27. L. M. Cañedo, J. A. de la Fuente, C. Gesto, M. J. Ferreiro, C. Jiménez and R. Riguera, Agrochelin, a new cytotoxic alkaloid from the marine bacteria Agrobacterium sp, *Tetrahedron Lett.*, 1999, **40**, 6841-6844.
28. J. Diettrich, H. Kage and M. Nett, Genomics-inspired discovery of massiliachelin, an agrochelin epimer from Massilia sp. NR 4-1, *Beilstein J. Org. Chem.*, 2019, **15**, 1298-1303.
29. N. Noinaj, M. Guillier, T. J. Barnard and S. K. Buchanan, TonB-dependent transporters: regulation, structure, and function, *Annu. Rev. Microbiol.*, 2010, **64**, 43-60.
30. F. Peuckert, A. L. Ramos-Vega, M. Miethke, C. J. Schwörer, A. G. Albrecht, M. Oberthür and M. A. Marahiel, The siderophore binding protein FeuA shows limited promiscuity toward exogenous triscatecholates, *Chem. Biol.*, 2011, **18**, 907-919.
31. A.-M. Albrecht-Gary, S. Blanc, N. Rochel, A. Z. Ocaktan and M. A. Abdallah, Bacterial Iron Transport: Coordination Properties of Pyoverdin PaA, a Peptidic Siderophore of Pseudomonas aeruginosa, *Inorg. Chem.*, 1994, **33**, 6391-6402.
32. G. Ganne, K. Brillet, B. Basta, B. Roche, F. Hoegy, V. Gasser and I. J. Schalk, Iron Release from the Siderophore Pyoverdine in Pseudomonas aeruginosa Involves Three New Actors: FpvC, FpvG, and FpvH, *ACS Chem. Biol.*, 2017, **12**, 1056-1065.
33. S. R. Cooper, J. V. McArdle and K. N. Raymond, Siderophore electrochemistry: relation to intracellular iron release mechanism, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 3551-3554.
34. M. Miethke, J. Hou and M. A. Marahiel, The siderophore-interacting protein YqjH acts as a ferric reductase in different iron assimilation pathways of Escherichia coli, *Biochemistry*, 2011, **50**, 10951-10964.
35. E. S. Sattely, M. A. Fischbach and C. T. Walsh, Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways, *Nat Prod Rep*, 2008, **25**, 757-793.
36. T. Stachelhaus, H. D. Mootz and M. A. Marahiel, The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chem. Biol.*, 1999, **6**, 493-505.
37. G. L. Challis, J. Ravel and C. A. Townsend, Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains, *Chem. Biol.*, 2000, **7**, 211-224.
38. C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben and D. H. Huson, Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs), *Nucleic Acids Res.*, 2005, **33**, 5799-5808.
39. M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity, *Nucleic Acids Res.*, 2011, **39**, W362-W367.
40. G. Aguero-Chapin, G. Perez-Machado, A. Sanchez-Rodriguez, M. M. Santos and A. Antunes, Alignment-Free Methods for the Detection and Specificity Prediction of Adenylation Domains, *Methods Mol Biol*, 2016, **1401**, 253-272.

41. R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, A new enzyme superfamily - the phosphopantetheinyl transferases, *Chem. Biol.*, 1996, **3**, 923-936.
42. C. C. Tseng, S. D. Bruner, R. M. Kohli, M. A. Marahiel, C. T. Walsh and S. A. Sieber, Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase, *Biochemistry*, 2002, **41**, 13350-13359.
43. L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel and C. T. Walsh, Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B, *Biochemistry*, 2002, **41**, 9184-9196.
44. S. A. Samel, P. Czodrowski and L.-O. Essen, Structure of the epimerization domain of tyrocidine synthetase A, *Acta Crystallogr. D Biol. Crystallogr.*, 2014, **70**, 1442-1452.
45. T. Stachelhaus and C. T. Walsh, Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase, *Biochemistry*, 2000, **39**, 5775-5787.
46. S. L. Clugston, S. A. Sieber, M. A. Marahiel and C. T. Walsh, Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst, *Biochemistry*, 2003, **42**, 12095-12104.
47. C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution, *BMC Evol. Biol.*, 2007, **7**, 78.
48. G. L. Challis and J. H. Naismith, Structural aspects of non-ribosomal peptide biosynthesis, *Curr. Opin. Struct. Biol.*, 2004, **14**, 748-756.
49. F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker and M. A. Marahiel, Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation, *Chem. Biol.*, 2010, **17**, 872-880.
50. M. F. Kreutzer, H. Kage and M. Nett, Structure and Biosynthetic Assembly of Cupriachelin, a Photoreactive Siderophore from the Bioplastic Producer Cupriavidus necator H16, *J. Am. Chem. Soc.*, 2012, **134**, 5415-5422.
51. H. K. Zane, H. Naka, F. Rosconi, M. Sandy, M. G. Haygood and A. Butler, Biosynthesis of amphi-enterobactin siderophores by Vibrio harveyi BAA-1116: identification of a bifunctional nonribosomal peptide synthetase condensation domain, *J. Am. Chem. Soc.*, 2014, **136**, 5615-5618.
52. Z. Hojati, C. Milne, B. Harvey, L. Gordon, M. Borg, F. Flett, B. Wilkinson, P. J. Sidebottom, B. A. M. Rudd, M. A. Hayes, C. P. Smith and J. Micklefield, Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor, *Chem. Biol.*, 2002, **9**, 1175-1187.
53. F. Rusnak, W. S. Faraci and C. T. Walsh, Subcloning, expression, and purification of the enterobactin biosynthetic enzyme 2,3-dihydroxybenzoate-AMP ligase: demonstration of enzyme-bound (2,3-dihydroxybenzoyl)adenylate product, *Biochemistry*, 1989, **28**, 6827-6835.
54. E. J. Drake, D. A. Nicolai and A. M. Gulick, Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain, *Chem. Biol.*, 2006, **13**, 409-419.

55. U. Keller, M. Lang, I. Crnovcic, F. Pfennig and F. Schauwecker, The actinomycin biosynthetic gene cluster of Streptomyces chrysomallus: a genetic hall of mirrors for synthesis of a molecule with mirror symmetry, *J. Bacteriol.*, 2010, **192**, 2583-2595.
56. J. R. Lai, M. A. Fischbach, D. R. Liu and C. T. Walsh, Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection, *J. Am. Chem. Soc.*, 2006, **128**, 11002-11003.
57. L. Ge and S. Y. K. Seah, Heterologous expression, purification, and characterization of an l-ornithine N(5)-hydroxylase involved in pyoverdine siderophore biosynthesis in Pseudomonas aeruginosa, *J. Bacteriol.*, 2006, **188**, 7205-7210.
58. G. M. Singh, P. D. Fortin, A. Koglin and C. T. Walsh, beta-Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: characterization of two candidate hydroxylases AspH and SyrP in Pseudomonas syringae, *Biochemistry*, 2008, **47**, 11310-11320.
59. L. Serino, C. Reimmann, H. Baur, M. Beyeler, P. Visca and D. Haas, Structural genes for salicylate biosynthesis from chorismate in Pseudomonas aeruginosa, *Mol. Gen. Genet.*, 1995, **249**, 217-228.
60. C. Gaille, P. Kast and D. Haas, Salicylate biosynthesis in Pseudomonas aeruginosa. Purification and characterization of PchB, a novel bifunctional enzyme displaying isochorismate pyruvate-lyase and chorismate mutase activities, *J. Biol. Chem.*, 2002, **277**, 21768-21775.
61. C. Pelludat, D. Brem and J. Heesemann, Irp9, encoded by the high-pathogenicity island of Yersinia enterocolitica, is able to convert chorismate into salicylate, the precursor of the siderophore yersiniabactin, *J. Bacteriol.*, 2003, **185**, 5648-5653.
62. O. Kerbarh, A. Ciulli, N. I. Howard and C. Abell, Salicylate biosynthesis: overexpression, purification, and characterization of Irp9, a bifunctional salicylate synthase from Yersinia enterocolitica, *J. Bacteriol.*, 2005, **187**, 5061-5066.
63. A. J. Harrison, M. Yu, T. Gårdenborg, M. Middleditch, R. J. Ramsay, E. N. Baker and J. S. Lott, The structure of MbtI from Mycobacterium tuberculosis, the first enzyme in the biosynthesis of the siderophore mycobactin, reveals it to be a salicylate synthase, *J. Bacteriol.*, 2006, **188**, 6081-6091.
64. R. Hermenau, K. Ishida, S. Gama, B. Hoffmann, M. Pfeifer-Leeg, W. Plass, J. F. Mohr, T. Wichard, H.-P. Saluz and C. Hertweck, Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system, *Nat. Chem. Biol.*, 2018.
65. R. Hermenau, J. L. Mehl, K. Ishida, B. Dose, S. J. Pidot, T. P. Stinear and C. Hertweck, Genomics-Driven Discovery of NO-Donating Diazeniumdiolate Siderophores in Diverse Plant-Associated Bacteria, *Angew. Chem. Int. Ed Engl.*, 2019, **58**, 13024-13029.
66. T. L. Ng, R. Rohac, A. J. Mitchell, A. K. Boal and E. P. Balskus, An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin, *Nature*, 2019, **566**, 94-99.
67. K. Scholz, T. Tiso, L. M. Blank and H. Hayen, Mass spectrometric characterization of siderophores produced by Pseudomonas taiwanensis VLB120 assisted by stable isotope labeling of nitrogen source, *Biometals*, 2018, **31**, 785-795.
68. T. B. Karpishin and K. N. Raymond, The First Structural Characterization of a Metal–Enterobactin Complex:[V(enterobactin)]2−, *Angewandte Chemie International Edition in English*, 1992, **31**, 466-468.

69. T. B. Karpishin, T. M. Dewey and K. N. Raymond, Coordination chemistry of microbial iron transport. 49. The vanadium(IV) enterobactin complex: structural, spectroscopic, and electrochemical characterization, *J. Am. Chem. Soc.*, 1993, **115**, 1842-1851.
70. T. D. P. Stack, Z. Hou and K. N. Raymond, Rational reduction of the conformational space of a siderophore analog through nonbonded interactions: the role of entropy in enterobactin, *J. Am. Chem. Soc.*, 1993, **115**, 6466-6467.
71. Z. Hou, T. D. P. Stack, C. J. Sunderland and K. N. Raymond, Enhanced iron(III) chelation through ligand predisposition: syntheses, structures and stability of tris-catecholate enterobactin analogs, *Inorganica Chim. Acta*, 1997, **263**, 341-355.
72. K. N. Raymond, E. A. Dertz and S. S. Kim, Enterobactin: an archetype for microbial iron transport, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3584-3588.
73. M. Sandy and A. Butler, Chrysobactin siderophores produced by Dickeya chrysanthemi EC16, *J. Nat. Prod.*, 2011, **74**, 1207-1212.
74. Y. Wen, X. Wu, Y. Teng, C. Qian, Z. Zhan, Y. Zhao and O. Li, Identification and analysis of the gene cluster involved in biosynthesis of paenibactin, a catecholate siderophore produced by Paenibacillus elgii B69, *Environ. Microbiol.*, 2011, **13**, 2726-2737.
75. S. I. Patzer and V. Braun, Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of Streptomyces sp. ATCC 700974, *J. Bacteriol.*, 2010, **192**, 426-435.
76. Y. Matsuo, K. Kanoh, J.-H. Jang, K. Adachi, S. Matsuda, O. Miki, T. Kato and Y. Shizuri, Streptobactin, a tricatechol-type siderophore from marine-derived Streptomyces sp. YM5-799, *J. Nat. Prod.*, 2011, **74**, 2371-2376.
77. M. Sandy, A. Han, J. Blunt, M. Munro, M. Haygood and A. Butler, Vanchrobactin and anguibactin siderophores produced by Vibrio sp. DS40M4, *J. Nat. Prod.*, 2010, **73**, 1038-1043.
78. A. W. Han, M. Sandy, B. Fishman, A. E. Trindade-Silva, C. A. G. Soares, D. L. Distel, A. Butler and M. G. Haygood, Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont Teredinibacter turnerae T7901, *PLoS One*, 2013, **8**, e76151.
79. S. I. Müller, M. Valdebenito and K. Hantke, Salmochelin, the long-overlooked catecholate siderophore of Salmonella, *Biometals*, 2009, **22**, 691-695.
80. M. A. Fischbach, H. Lin, D. R. Liu and C. T. Walsh, In vitro characterization of IroB, a pathogen-associated C-glycosyltransferase, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 571-576.
81. D. Foshag, C. Campbell and P. D. Pawelek, The C-glycosyltransferase IroB from pathogenic Escherichia coli: identification of residues required for efficient catalysis, *Biochim. Biophys. Acta*, 2014, **1844**, 1619-1630.
82. B. Bister, D. Bischoff, G. J. Nicholson, M. Valdebenito, K. Schneider, G. Winkelmann, K. Hantke and R. D. Süssmuth, The structure of salmochelins: C-glucosylated enterobactins of Salmonella enterica§, *Biometals*, 2004, **17**, 471-481.
83. M.-L. V. Crouch, M. Castor, J. E. Karlinsey, T. Kalhorn and F. C. Fang, Biosynthesis and IroC-dependent export of the siderophore salmochelin are essential for virulence of Salmonella enterica serovar Typhimurium, *Mol. Microbiol.*, 2008, **67**, 971-983.

84. M. Caza, F. Lépine, S. Milot and C. M. Dozois, Specific roles of the iroBCDEN genes in virulence of an avian pathogenic Escherichia coli O78 strain and in production of salmochelins, *Infect. Immun.*, 2008, **76**, 3539-3549.
85. M. Zhu, M. Valdebenito, G. Winkelmann and K. Hantke, Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization, *Microbiology*, 2005, **151**, 2363-2372.
86. N. A. Larsen, H. Lin, R. Wei, M. A. Fischbach and C. T. Walsh, Structural characterization of enterobactin hydrolase IroE, *Biochemistry*, 2006, **45**, 10184-10190.
87. M. Valdebenito, S. I. Müller and K. Hantke, Special conditions allow binding of the siderophore salmochelin to siderocalin (NGAL-lipocalin), *FEMS Microbiol. Lett.*, 2007, **277**, 182-187.
88. M. A. Fischbach, H. Lin, L. Zhou, Y. Yu, R. J. Abergel, D. R. Liu, K. N. Raymond, B. L. Wanner, R. K. Strong, C. T. Walsh, A. Aderem and K. D. Smith, The pathogen-associated iroA gene cluster mediates bacterial evasion of lipocalin 2, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16502-16507.
89. R. J. Abergel, M. K. Wilson, J. E. L. Arceneaux, T. M. Hoette, R. K. Strong, B. R. Byers and K. N. Raymond, Anthrax pathogen evades the mammalian immune system through stealth siderophore production, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 18499-18503.
90. M. Luo, H. Lin, M. A. Fischbach, D. R. Liu, C. T. Walsh and J. T. Groves, Enzymatic tailoring of enterobactin alters membrane partitioning and iron acquisition, *ACS Chem. Biol.*, 2006, **1**, 29-32.
91. G. Vassiliadis, D. Destoumieux-Garzón, C. Lombard, S. Rebuffat and J. Peduzzi, Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47, *Antimicrob. Agents Chemother.*, 2010, **54**, 288-297.
92. S. I. Patzer, M. R. Baquero, D. Bravo, F. Moreno and K. Hantke, The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir, Fiu and IroN, *Microbiology*, 2003, **149**, 2557-2570.
93. X. Thomas, D. Destoumieux-Garzón, J. Peduzzi, C. Afonso, A. Blond, N. Birlirakis, C. Goulard, L. Dubost, R. Thai, J.-C. Tabet and S. Rebuffat, Siderophore peptide, a new type of post-translationally modified antibacterial peptide with potent activity, *J. Biol. Chem.*, 2004, **279**, 28233-28242.
94. M. F. Azpiroz and M. Laviña, Involvement of enterobactin synthesis pathway in production of microcin H47, *Antimicrob. Agents Chemother.*, 2004, **48**, 1235-1241.
95. E. M. Nolan, M. A. Fischbach, A. Koglin and C. T. Walsh, Biosynthetic tailoring of microcin E492m: post-translational modification affords an antibacterial siderophore-peptide conjugate, *J. Am. Chem. Soc.*, 2007, **129**, 14336-14347.
96. G. Mercado, M. Tello, M. Marín, O. Monasterio and R. Lagos, The Production In Vivo of Microcin E492 with Antibacterial Activity Depends on Salmochelin and EntF, *J. Bacteriol.*, 2008, **190**, 5464-5471.
97. M. P. Kem and A. Butler, Acyl peptidic siderophores: structures, biosyntheses and post-assembly modifications, *Biometals*, 2015, **28**, 445-459.
98. M. K. Wilson, R. J. Abergel, K. N. Raymond, J. E. L. Arceneaux and B. R. Byers, Siderophores of Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis, *Biochem. Biophys. Res. Commun.*, 2006, **348**, 320-325.

99. H. Budzikiewicz, A. Bössenkamp, K. Taraz, A. Pandey and J. M. Meyer, Corynebactin, a cyclic catecholate siderophore from Corynebacterium glutamicum ATCC 14067 (Brevibacterium sp. DSM 20411), *Zeitschrift für Naturforschung C*, 1997, **52**, 551-554.
100. J. J. May, T. M. Wendrich and M. A. Marahiel, The dhb Operon of Bacillus subtilisEncodes the Biosynthetic Template for the Catecholic Siderophore 2,3-Dihydroxybenzoate-Glycine-Threonine Trimeric Ester Bacillibactin, *J. Biol. Chem.*, 2001, **276**, 7209-7217.
101. E. A. Dertz, A. Stintzi and K. N. Raymond, Siderophore-mediated iron transport in Bacillus subtilis and Corynebacterium glutamicum, *J. Biol. Inorg. Chem.*, 2006, **11**, 1087-1097.
102. S. Zajdowicz, J. C. Haller, A. E. Krafft, S. W. Hunsucker, C. T. Mant, M. W. Duncan, R. S. Hodges, D. N. M. Jones and R. K. Holmes, Purification and Structural Characterization of Siderophore (Corynebactin) from Corynebacterium diphtheriae, *PLoS One*, 2012, **7**, e34591.
103. M. M. Nakano, N. Corbell, J. Besson and P. Zuber, Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in Bacillus subtilis, *Mol. Gen. Genet.*, 1992, **232**, 313-321.
104. J. Ollinger, K.-B. Song, H. Antelmann, M. Hecker and J. D. Helmann, Role of the Fur regulon in iron transport in Bacillus subtilis, *J. Bacteriol.*, 2006, **188**, 3664-3673.
105. T. H. Grossman, M. Tuckman, S. Ellestad and M. S. Osburne, Isolation and characterization of Bacillus subtilis genes involved in siderophore biosynthesis: relationship between B. subtilis sfpo and Escherichia coli entD genes, *J. Bacteriol.*, 1993, **175**, 6203-6211.
106. L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases, *Biochemistry*, 1998, **37**, 1585-1595.
107. B. O. Bachmann and J. Ravel, Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data, *Methods Enzymol.*, 2009, **458**, 181-217.
108. U. Linne, S. Doekel and M. A. Marahiel, Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases, *Biochemistry*, 2001, **40**, 15824-15834.
109. P. J. Belshaw, C. T. Walsh and T. Stachelhaus, Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis, *Science*, 1999, **284**, 486-489.
110. S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh and M. T. Stubbs, Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE, *Structure*, 2002, **10**, 301-310.
111. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, 2004, **32**, 1792-1797.
112. W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso and R. Lopez, The EMBL-EBI bioinformatics web and programmatic tools framework, *Nucleic Acids Res.*, 2015, **43**, W580-584.

113. A. M. Gehring, I. Mori and C. T. Walsh, Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF, *Biochemistry*, 1998, **37**, 2648-2659.
114. C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg and C. T. Walsh, Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization, *Chem. Biol.*, 1999, **6**, 385-400.
115. T. Franza, C. Enard, F. van Gijsegem and D. Expert, Genetic analysis of the Erwinia chrysanthemi 3937 chrysobactin iron-transport system: characterization of a gene cluster involved in uptake and biosynthetic pathways, *Mol. Microbiol.*, 1991, **5**, 1319-1329.
116. M. Persmark and J. B. Neilands, Iron(III) complexes of chrysobactin, the siderophore ofErwinia chrysanthemi, *Biometals*, 1992, **5**, 29-36.
117. L. Rauscher, D. Expert, B. F. Matzanke and A. X. Trautwein, Chrysobactin-dependent Iron Acquisition inErwinia chrysanthemi : FUNCTIONAL STUDY OF A HOMOLOG OF THE ESCHERICHIA COLIFERRIC ENTEROBACTIN ESTERASE, *J. Biol. Chem.*, 2002, **277**, 2385-2395.
118. V. Tomišić, S. Blanc, M. Elhabiri, D. Expert and A.-M. Albrecht-Gary, Iron(III) Uptake and Release by Chrysobactin, a Siderophore of the Phytophatogenic Bacterium Erwinia chrysanthemi, *Inorg. Chem.*, 2008, **47**, 9419-9430.
119. P. Vukosav, L. Frkanec and M. Mlakar, Voltammetric investigation of iron(III) complexes with siderophore chrysobactin in aqueous solution, *Electrochim. Acta*, 2012, **59**, 479-484.
120. C. Enard, A. Diolez and D. Expert, Systemic virulence of Erwinia chrysanthemi 3937 requires a functional iron assimilation system, *J. Bacteriol.*, 1988, **170**, 2419-2426.
121. C. Neema, J. P. Laulhere and D. Expert, Iron Deficiency Induced by Chrysobactin in Saintpaulia Leaves Inoculated with Erwinia chrysanthemi, *Plant Physiol.*, 1993, **102**, 967-973.
122. C. Sauvage and D. Expert, Differential regulation by iron of Erwinia chrysanthemi pectate lyases: pathogenicity of iron transport regulatory (cbr) mutants, *Mol. Plant. Microbe. Interact.*, 1994, **7**, 71-71.
123. C. Masclaux and D. Expert, Signalling potential of iron in plant—microbe interactions: the pathogenic switch of iron transport in Erwinia chrysanthemi, *Plant J.*, 1995, **7**, 121-128.
124. T. Franza, B. Mahé and D. Expert, Erwinia chrysanthemi requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection, *Mol. Microbiol.*, 2005, **55**, 261-275.
125. M. Persmark, D. Expert and J. B. Neilands, Isolation, characterization, and synthesis of chrysobactin, a compound with siderophore activity from Erwinia chrysanthemi, *J. Biol. Chem.*, 1989, **264**, 3187-3193.
126. G. Ehlert, K. Taraz and H. Budzikiewicz, Serratiochelin, a New Catecholate Siderophore from Serratia marcescens, *Zeitschrift für Naturforschung C*, 1994, **49**, 11-17.
127. M. Adolphs, K. Taraz and H. Budzikiewicz, Catecholate siderophores from Chryseomonas luteola, *Zeitschrift für Naturforschung C*, 1996.

128. H. H. Barnes and C. A. Ishimaru, Purification of catechol siderophores by boronate affinity chromatography: Identification of chrysobactin from Erwinia carotovora subsp. carotovora, *Biometals*, 1999, **12**, 83-87.
129. M. L. Lemos, P. Salinas, A. E. Toranzo, J. L. Barja and J. H. Crosa, Chromosome-mediated iron uptake system in pathogenic strains of Vibrio anguillarum, *J. Bacteriol.*, 1988, **170**, 1920-1925.
130. R. G. Soengas, C. Anta, A. Espada, V. Paz, I. R. Ares, M. Balado, J. Rodríguez, M. L. Lemos and C. Jiménez, Structural characterization of vanchrobactin, a new catechol siderophore produced by the fish pathogen Vibrio anguillarum serotype O2, *Tetrahedron Lett.*, 2006, **47**, 7113-7116.
131. R. G. Soengas, C. Anta, A. Espada, R. M. Nieto, M. Larrosa, J. Rodríguez and C. Jiménez, Vanchrobactin: absolute configuration and total synthesis, *Tetrahedron Lett.*, 2007, **48**, 3021-3024.
132. E. Iglesias, I. Brandariz, C. Jiménez and R. G. Soengas, Iron(iii) complexation by Vanchrobactin , a siderophore of the bacterial fish pathogen Vibrio anguillarum, *Metallomics*, 2011, **3**, 521-528.
133. R. G. Soengas, M. Larrosa, M. Balado, J. Rodríguez, M. L. Lemos and C. Jiménez, Synthesis and biological activity of analogues of vanchrobactin, a siderophore from Vibrio anguillarum serotype O2, *Org. Biomol. Chem.*, 2008, **6**, 1278-1287.
134. Y. V. Temirov, T. Z. Esikova, I. A. Kashparov, T. A. Balashova, L. M. Vinokurov and Y. B. Alakhov, A Catecholic Siderophore Produced by the Thermoresistant Bacilluslicheniformis VK21 Strain, *Russ. J. Bioorganic Chem.*, 2003, **29**, 542-549.
135. T. Aoyagi, M. Hatsu, F. Kojima, C. Hayashi, M. Hamada and T. Takeuchi, Benarthin: a new inhibitor of pyroglutamyl peptidase. I. Taxonomy, fermentation, isolation and biological activities, *J. Antibiot.*, 1992, **45**, 1079-1083.
136. M. Hatsu, H. Naganawa, T. Aoyagi and T. Takeuchi, Benarthin: a new inhibitor of pyroglutamyl peptidase. II. Physico-chemical properties and structure determination, *J. Antibiot.*, 1992, **45**, 1084-1087.
137. D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh and G. Wagner, Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase, *Nature*, 2008, **454**, 903-906.
138. B. Wagner, S. A. Sieber, M. Baumann and M. A. Marahiel, Solvent engineering substantially enhances the chemoenzymatic production of surfactin, *Chembiochem*, 2006, **7**, 595-597.
139. A. W. Han, Oregon Health & Science University, 2011.
140. R. A. Carter, P. S. Worsley, G. Sawers, G. L. Challis, M. J. Dilworth, K. C. Carson, J. A. Lawrence, M. Wexler, A. W. B. Johnston and K. H. Yeoman, The vbs genes that direct synthesis of the siderophore vicibactin in Rhizobium leguminosarum: their expression in other genera requires ECF σ factor RpoI, *Mol. Microbiol.*, 2002, **44**, 1153-1166.
141. Y. Liu, T. Zheng and S. D. Bruner, Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases, *Chem. Biol.*, 2011, **18**, 1482-1488.
142. X. Zeng, Y. Mo, F. Xu and J. Lin, Identification and characterization of a periplasmic trilactone esterase, Cee, revealed unique features of ferric enterobactin acquisition in Campylobacter, *Mol. Microbiol.*, 2013, **87**, 594-608.

143. E. E. Wyckoff, B. E. Allred, K. N. Raymond and S. M. Payne, Catechol Siderophore Transport by Vibrio cholerae, *J. Bacteriol.*, 2015.
144. W. Rabsch, W. Voigt, R. Reissbrodt, R. M. Tsolis and A. J. Bäumler, Salmonella typhimurium IroN and FepA proteins mediate uptake of enterobactin but differ in their specificity for other siderophores, *J. Bacteriol.*, 1999, **181**, 3610-3612.
145. R. C. Scarrow, D. J. Ecker, C. Ng, S. Liu and K. N. Raymond, Iron(III) coordination chemistry of linear dihydroxyserine compounds derived from enterobactin, *Inorg. Chem.*, 1991, **30**, 900-906.
146. E. A. Dertz, J. Xu and K. N. Raymond, Tren-Based Analogues of Bacillibactin: Structure and Stability1, *Inorg. Chem.*, 2006, **45**, 5465-5478.
147. T. M. Hoette, R. J. Abergel, J. Xu, R. K. Strong and K. N. Raymond, The role of electrostatics in siderophore recognition by the immunoprotein Siderocalin, *J. Am. Chem. Soc.*, 2008, **130**, 17584-17592.
148. K. Bloudoff, C. D. Fage, M. A. Marahiel and T. M. Schmeing, Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis, *Proceedings of the National Academy of Sciences*, 2017, **114**, 95-100.
149. T. A. Keating, C. G. Marshall, C. T. Walsh and A. E. Keating, The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains, *Nat. Struct. Biol.*, 2002, **9**, 522-526.
150. I. Schmitt and F. K. Barker, Phylogenetic methods in natural product research, *Nat. Prod. Rep.*, 2009, **26**, 1585-1602.
151. N. Ziemert and P. R. Jensen, in *Methods in Enzymology*, ed. D. A. Hopwood, Academic Press, 2012, vol. 517, pp. 161-182.
152. M. Adamek, M. Alanjary and N. Ziemert, Applied evolution: phylogeny-based approaches in natural products research, *Nat. Prod. Rep.*, 2019, **36**, 1295-1312.
153. W. R. Pearson, An introduction to sequence similarity ("homology") searching, *Curr. Protoc. Bioinformatics*, 2013, **Chapter 3**, Unit3.1.
154. M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T. L. Madden, NCBI BLAST: a better web interface, *Nucleic Acids Res.*, 2008, **36**, W5-9.
155. C. UniProt, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.*, 2019, **47**, D506-D515.
156. C. The UniProt, UniProtKB/TrEMBL 2019_09, https://www.uniprot.org/statistics/TrEMBL, (accessed 2019/11/10).
157. S. R. Eddy, Accelerated Profile HMM Searches, *PLoS Comput. Biol.*, 2011, **7**, e1002195.
158. S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto and R. D. Finn, The Pfam protein families database in 2019, *Nucleic Acids Res.*, 2019, **47**, D427-D432.
159. S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez and R. D. Finn, HMMER web server: 2018 update, *Nucleic Acids Res.*, 2018, **46**, W200-W204.
160. A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng and S. H. Bryant, CDD: NCBI's conserved domain database, *Nucleic Acids Res.*, 2015, **43**, D222-226.

161. D. H. Haft, J. D. Selengut and O. White, The TIGRFAMs database of protein families, *Nucleic Acids Res.*, 2003, **31**, 371-373.
162. A. L. Mitchell, T. K. Attwood, P. C. Babbitt, M. Blum, P. Bork, A. Bridge, S. D. Brown, H.-Y. Chang, S. El-Gebali, M. I. Fraser, J. Gough, D. R. Haft, H. Huang, I. Letunic, R. Lopez, A. Luciani, F. Madeira, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, G. Nuka, C. Orengo, A. P. Pandurangan, T. Paysan-Lafosse, S. Pesseat, S. C. Potter, M. A. Qureshi, N. D. Rawlings, N. Redaschi, L. J. Richardson, C. Rivoire, G. A. Salazar, A. Sangrador-Vegas, C. J. A. Sigrist, I. Sillitoe, G. G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, S.-Y. Yong and R. D. Finn, InterPro in 2019: improving coverage, classification and access to protein sequence annotations, *Nucleic Acids Res.*, 2019, **47**, D351-D360.

Chapter 2. Genomics and stereochemistry of β-hydroxylation

This chapter was adapted from Reference (1) with permission from The National Academy of Sciences.

2.1. Introduction

Many peptidic siderophores contain β-hydroxyaspartate (β-OHAsp), which provides bidentate OO' coordination to Fe(III) (2). The first structural determination of a β-OHAsp-containing siderophore came with the crystallization of ferric pyoverdine from *Pseudomonas* B10 (Figure 2.1) in 1981 (3). Since then, a variety of peptidic siderophores with β-OHAsp have been characterized from both marine and terrestrial bacteria. Like other α-hydroxycarboxylate ligands, β-OHAsp bound to Fe(III) can undergo photo-induced reduction of Fe(III) to Fe(II) accompanied by oxidative decarboxylation of the ligand (2, 4).

Far fewer siderophores contain the chelating group β-hydroxyhistidine (β-OHHis). The first reported example is pyoverdine pf244 of *Pseudomonas fluorescens* 244 (Figure 2.1) (5). β-OHHis has since been identified in the peptide of pyoverdines from a variety of pseudomonads (6-8). Some *Pseudomonas* strains produce the fatty-acyl peptidic siderophores corrugatin, ornicorrugatin, or histicorrugatin (Figure 2.1), which contain both β-OHHis and β-OHAsp (9-11). One β-hydroxyasparagine-containing siderophore has been reported, pyoverdine VLB120 from *Pseudomonas taiwanensis* VLB120 (Figure 2.1) (12, 13).

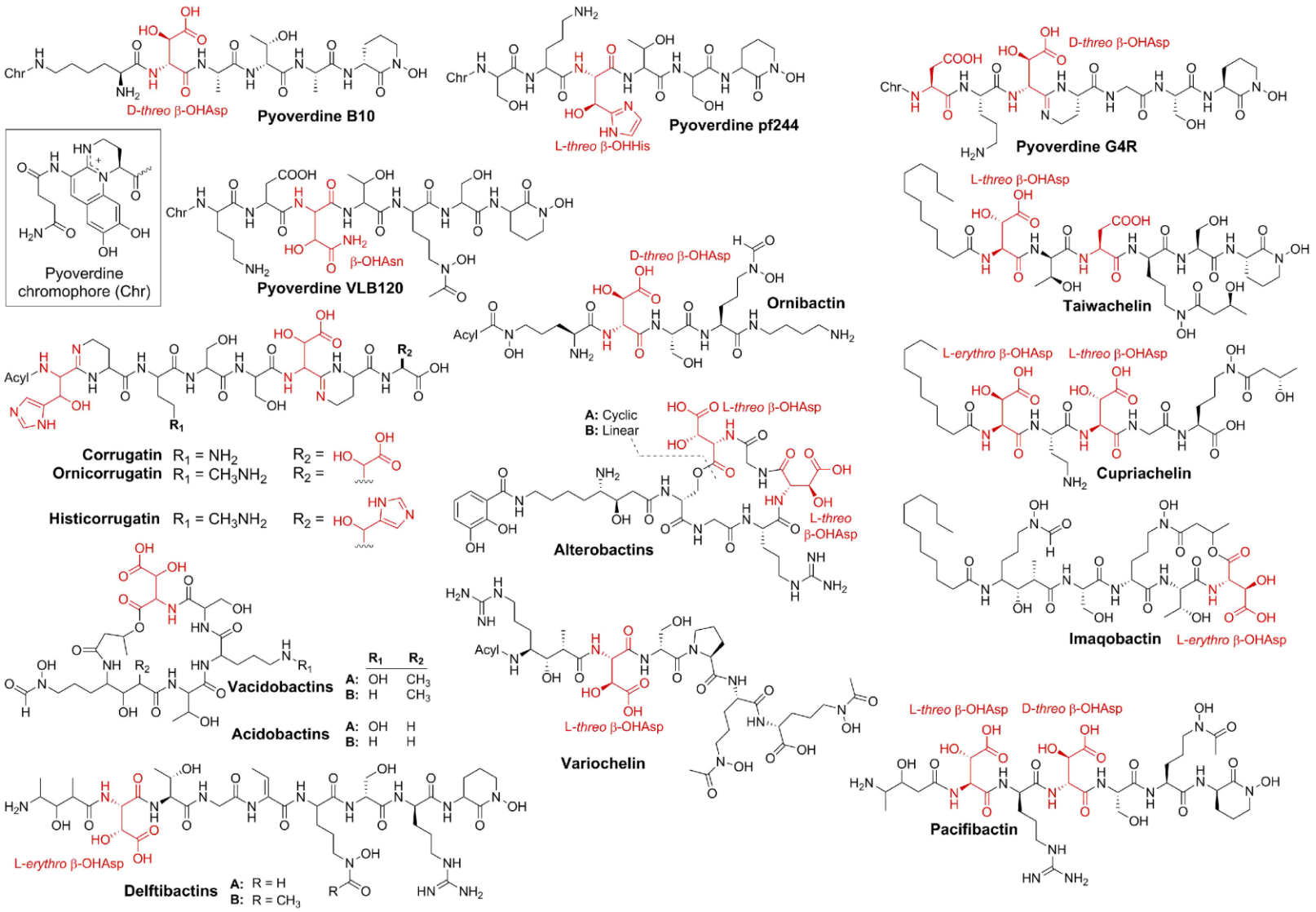


Figure 2.1. Reported siderophores with β-OHAsp, β-OHHis, or β-OHAsn discussed in the text.

Several families of enzymes are responsible for the β-hydroxy amino acids found in nonribosomal peptides, namely Fe-heme monooxygenases (14), diiron monooxygenases (15), and non-heme Fe(II)/α-ketoglutarate-dependent dioxygenases (16). Only this last class of Fe_{NH}/αKG dioxygenases have been found to hydroxylate aspartic acid (16-18). OrbG, encoded in the biosynthetic gene cluster of the siderophore ornibactin (Figure 2.1), was the first enzyme predicted to be an aspartyl β-hydroxylase based on homology to Fe_{NH}/αKG dioxygenases (19). All β-OHAsp-containing siderophores feature at least one homolog of *orbG* in their biosynthetic gene cluster, either as a discrete gene, or as a tailoring domain fused to the C-terminus of a NRPS gene (*e.g.* serobactin, Figure 2.2) (2). No siderophore β-hydroxylases of aspartate or histidine have been characterized, although they have high sequence similarity (47-76%) to SyrP, the only NRPS-associated aspartyl β-hydroxylase characterized *in vitro* (16). SyrP is involved in the biosynthesis of syringomycin (Figure 2.3), a phytotoxin produced by *Pseudomonas syringae* (16). Originally annotated as a regulatory protein, SyrP was found to hydroxylate not free Asp, but Asp tethered to a thiolation domain of the syringomycin NRPS SyrE (16).

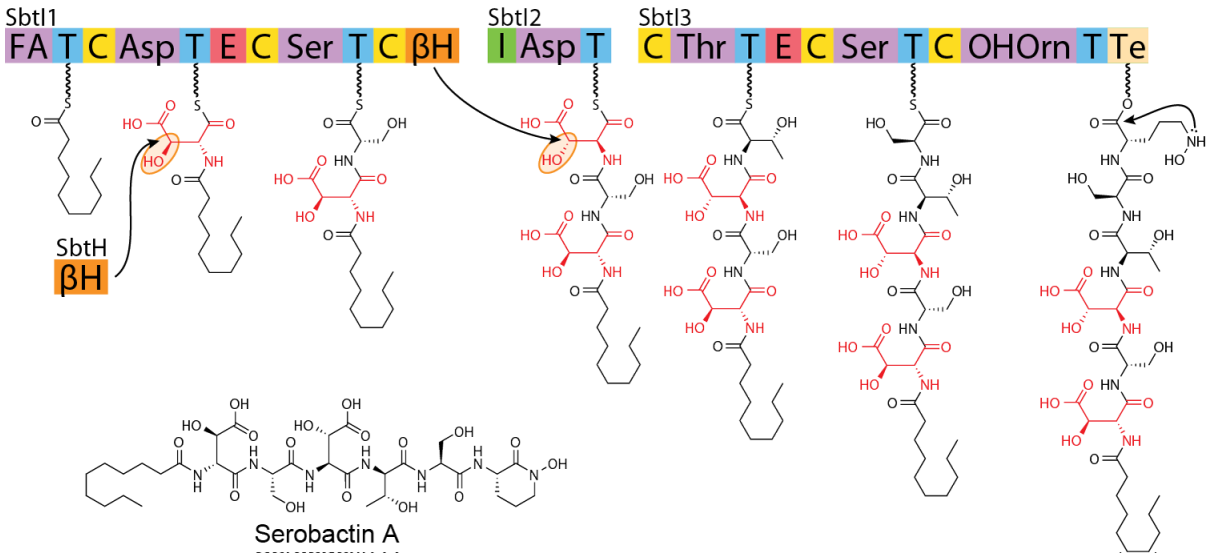


Figure 2.2. Proposed non-ribosomal biosynthesis of serobactin A. β-hydroxylases (βH), which may be standalone enzymes (*e.g.*, SbtH) or fused NRPS tailoring domains (*e.g.*, orange domain on SbtI1), hydroxylate L-Asp residues while they are tethered to the thiolation domain of the NRPS. Some β-hydroxylases are associated with an interface (I) domain (*e.g.*, green domain on SbtI2), newly identified and described in the text. Adenylation (A) domains are labeled by the substrate they activate and incorporate. T – thiolation domain; C – condensation domain; E – epimerization domain; Te – thioesterase domain.

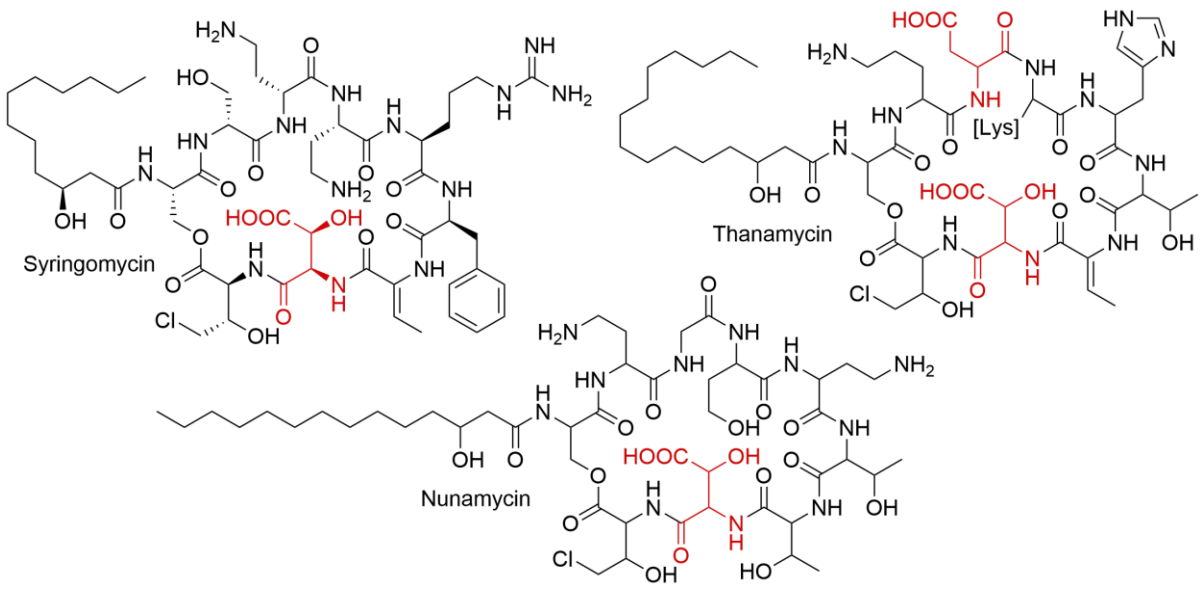


Figure 2.3. Phytotoxins with β-OHAsp residues.

Based on homology to SyrP and OrbG, the existence of a number of β-OHAsp residues in siderophores has been rationalized (7, 11, 20-24) or predicted (25-29). Recently, Kurth and coworkers used *cucF*, a β-hydroxylase domain from the cupriachelin (Figure 2.1) gene cluster, as a handle to scan genomes for photoactive Fe(III)-siderophores, leading to the discovery of variochelin (Figure 2.1) (29). Genome mining for β-OHAsp-containing siderophores has been quite successful, but current techniques can leave ambiguity in the predicted structure. Although the taiwachelin (Figure 2.1) NRPS was correctly predicted to load two Asp residues, predicting that only one would be hydroxylated by TaiD, the putative Asp β-hydroxylase, was not possible at the time (25). Some pyoverdines also retain an unmodified Asp (*e.g.*, pyoverdine G4R, Figure 2.1) (7, 23). In contrast, both Asp residues are hydroxylated in cupriachelin, serobactin, pacifibactin, and alterobactin (Figure 2.1) (21, 26, 30).

Further complicating structural predictions, β-OHAsp has two stereocenters (i.e., at the α- and β-carbons); thus, β-OHAsp potentially exists as any of four diastereomers. All stereochemically-characterized siderophores were reported to contain either D-*threo* (2*R*, 3*R*) or L-*threo* (2*S*, 3*S*) β-OHAsp until 2018, when the L-*erythro* (2*S*, 3*R*) isomer was reported in the amphiphilic siderophore imaqobactin (Figure 2.1) (31). The L- or D- configuration of the α-carbon can easily be predicted by the absence or presence of an E domain; however, no methods exist for predicting the stereochemistry at the β-carbon.

To develop refined genomic tools to predict the reactivity and stereoselectivity of β-hydroxylation in siderophores, gene clusters responsible for the biosynthesis of structurally characterized siderophores containing β-OHAsp or β-OHHis were analyzed. Functional subtypes of β-hydroxylases emerged, which were corroborated with phylogenetic analysis. These subtypes show clear patterns in genomic organization (standalone enzymes or integrated NPRS tailoring domains), amino acid substrate (Asp or His), and reactive NRPS partner. Significantly, the subtypes also exhibit divergent diastereoselectivity, enabling the prediction of the conformation at the C3 stereocenter. A previously undescribed member of the NRPS condensation domain superfamily was also identified, which is associated with certain β-hydroxylase subtypes. This class of enzymes was named the *interface* (I) domain for a proposed role in positioning the β-hydroxylase and the NRPS-bound amino acid substrate prior to hydroxylation.

2.2. Methods

2.2.1. Collection of known siderophore biosynthetic gene clusters

Genomes of bacterial strains with reported β-OHAsp- and β-OHHis-containing siderophores were downloaded as assemblies from NCBI RefSeq (Table 2.5) (69). The

collection covers siderophore structures published through June 2019. Siderophores produced by unsequenced strains were excluded from this analysis, with the exception of alterobactins. Originally isolated from an unsequenced marine isolate of *Pseudoalteromonas luteoviolacea* (70), a putative catechol/β-OHAsp gene cluster consistent with alterobactin is present in all sequenced *P. luteoviolacea* strains. NRPS domain organization was determined by comparing the amino acid sequences to a database of common NRPS domain HMMs using hmmscan (HMMER3 (68)) and confirmed by comparison to the linear structure of the siderophore.

2.2.2. Protein sequence manipulation

In each genome of interest, putative β-hydroxylase domains were identified using hmmsearch (HMMER3 (68)) to find matches to the Pfam TauD family PF02668 (39). The amino acid sequences of these domains were excised from the proteins by trimming to the resulting hmmsearch envelope range. Sequences then were aligned using MUSCLE (71). Interface (I) domains were too poorly conserved to be trimmed by hmmsearch with the PFAM condensation domain family PF00668 (39). Instead, parent NRPS protein sequences were truncated to the 500 N-terminal amino acids and aligned using MUSCLE. Using the profile-profile alignment function from MUSCLE, these sequences were aligned to the condensation domain alignment created by Rausch *et al*. (16), and trimmed to length using SeaView (72). The resulting excised I domains were then realigned with MUSCLE. To visualize the multiple sequence alignments, sequence logos were created using WebLogo (73), trimming any position with >50% gaps. Profile Hidden Markov models (pHMMs) were created for each of the β-hydroxylase subtypes. Representative protein sequences were

collected and aligned with MUSCLE (71), then HMMs were constructed from these multiple sequence alignments using the hmmbuild function from HMMER3 (68).

2.2.3. Phylogenetic analyses

Phylogenetic trees of β-hydroxylases and of the condensation domain superfamily were reconstructed. Multiple sequence alignments of protein sequences were prepared with MUSCLE (71) and trimmed with SeaView (72). The maximum likelihood phylogenetic tree was reconstructed in IQ-TREE 1.6.7 (74), using the best-fit model of protein evolution for each alignment as chosen by ModelFinder (Akaike information criterion) (75). Branch support was assessed by bootstrapping (100 bootstrap replicates). Phylogenies were visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree/). To place the interface domain in the condensation domain superfamily, we recreated the unrooted phylogeny from Rausch et al. 2007, Figure 2.5 (16). Using the profile-profile align function from MUSCLE, we aligned our interface domains to the 203 taxa used by Rausch and coworkers (16), and trimmed the interface domain sequences to the existing alignment before IQ-TREE reconstruction.

2.3. Results and Interpretation

2.3.1. Compilation and organization of β-hydroxylase genes from known siderophores

A comprehensive literature search revealed more than 35 β-OHAsp- and β-OHHis-containing siderophores with reported structures (Tables 2.1 and 2.2), of which 26 were isolated from strains with published genomes (Table 2.5). The associated biosynthetic gene clusters were extracted and annotated, resulting in a final dataset of 30 β-hydroxylases

(Table 2.1). Two families of aspartyl β-hydroxylases emerged with distinct genetic organizations: (1) integrated tailoring domains fused to the core NRPS machinery, and (2) standalone enzymes encoded by discrete genes within the biosynthetic cluster (Table 2.1). The gene clusters for cupriachelin, serobactin, and pacifibactin each encode two putative aspartyl β-hydroxylases (*e.g.*, SbtH and SbtI1 in serobactin; Figure 2.2). All three hydroxylases putatively responsible for β-OHHis synthesis are found in the biosynthetic gene clusters as standalone genes. By coordinating the genomic organization of the β-hydroxylase with both the NRPS architecture and the final peptidic siderophore structure, contrasting reactivity was found, as summarized in Table 2.1 and elaborated below.

2.3.2. Reactivity by the aspartic acid β-hydroxylase NRPS domain

The aspartyl β-hydroxylase domains are found at the C-terminus of a NRPS gene, directly following a C domain (Figure 2.4). These tailoring domains hydroxylate Asp residues loaded by NRPS modules with the domain architecture C*-A-T. While this C* domain is homologous to other C domains, it lacks the typical condensation domain catalytic motif, HHxxxDG (32). The β-hydroxylase-associated C* domains form a distinct phylogenetic clade within the condensation domain superfamily (*vide infra*); we name this previously unreported subtype the *interface* (I) domain for a putative role in positioning the β-hydroxylase and NRPS-bound substrate for hydroxylation, as discussed below. Accordingly, we name the integrated aspartyl β-hydroxylase domain the *interface-associated* Asp β-hydroxylase ($I\beta H_{Asp}$). Taiwachelin has a single $I\beta H_{Asp}$ domain in TaiD and two modules that load Asp; hydroxylation is only found to occur in the TaiE module containing an I domain (Figure 2.4) (25).

Table 2.1. Putative amino acid β-hydroxylases found in siderophore biosynthesis clusters.

	Interface-associated Asp β-Hydroxylases ($I\beta H_{Asp}$) *NRPS domain*		**T_E-Associated Asp β-Hydroxylases ($T\beta H_{Asp}$)** *Standalone enzyme*		**Interface-associated His β-Hydroxylases ($I\beta H_{His}$)** *Standalone enzyme*	
Siderophore	**Name**	**Stereo-isomer***	**Name**	**Stereo-Isomer***	**Name**	**Stereo-isomer***
Alterobactin	AltH	L-*threo*				
PVD 17400	PvdJ	L-*threo*				
PVD Pf0-1	PvdJ	[L-*threo*]				
Taiwachelin	TaiD	L-*threo*				
Variobactin	Var4	[L-*threo*]				
Variochelin	VarG	L-*threo*				
Cupriachelin	CucF	L-*threo*	CucE	**L-*erythro***		
Pacifibactin	PfbH	L-*threo*	PfbF	D-*threo*		
Serobactin	SbtI1	L-*threo*	SbtH	D-*threo*		
Acidobactin			Aave_3734	[L-*erythro*]		
Azotobactin			AVCA_RS11700	D-*threo*		
Crochelin			CroC	[D-*threo*]		
Delftibactin			DelD	**L-*erythro***		
LB400			MbaH	D-*threo*		
Malleobactin			MbaH	D-*threo*		
Marinobactin			ENO16757	D-*threo*		
Ornibactin			OrbG	D-*threo*		
PVD 1448a			Pspph_RS09710	[D-*threo*]		
PVD B10			BXA01_RS25280	D-*threo*		
PVD GB-1			PputGB1_4087	**D-*threo***		
PVD G4R			PP_4222	[D-*threo*]		
PVD W15Oct28			AK81_RS20255	[D-*threo*]		
Vacidobactin			Vapar_3747	[L-*erythro*]		
Histicorrugatin			HcsC	**L-*erythro***	HcsE	[L-*threo*]
PVD pf244					$SyrP_{21245}$	L-*threo*
PVD L48					PSEEN_RS14940	[L-*threo*]

Genomic data were retrieved from RefSeq (Table 2.5).
* Brackets indicate predicted β-OHAsp or β-OHHis stereochemistry, as described in the text. Bold text indicates stereochemistries reported herein. L-*threo*, (2*S*, 3*S*); L-*erythro*, (2*S*, 3*R*); D-*threo*, (2*R*, 3*R*).

Table 2.2. Siderophores with β-Hydroxy amino acids from unsequenced bacteria

Siderophore	Producing Strain	β-OH AA	Stereo-Isomer	Structure Ref
Aquachelins	*Halomonas* aquamarina DS40M3	β-OHAsp	L-*threo*	(33)
Halochelins	*Halomonas* sp. SL01	β-OHAsp	-	(34)
Loihichelins	*Halomonas* sp. LOB-5	β-OHAsp	D-*threo*	(35)
Sodachelins	*Halomonas* sp. SL28	β-OHAsp	-	(36)
Imaqobactin	*Variovorax* sp. RKJM285	β-OHAsp	L-*erythro*	(31)
Pseudoalterobactins	*Pseudoalteromonas* sp. KP20-4	β-OHAsp	-	(37)
Corrugatin	*Pseudomonas corrugata*	β-OHAsp	L-*threo*	(9)
		β-OHHis	L-*threo*	
Ornicorrugatin	*Pseudomonas fluorescens* AF76	β-OHAsp	L-*threo*	(10)
		β-OHHis	L-*threo*	
Pyoverdine VLB120*	*Pseudomonas taiwanensis* VLB120	β-OHAsn	-	(12, 13)
Gramibactin*	*Paraburkholderia graminis* C4D1M	β-OHAsp	D-*threo*	(38)

* Genomes of *P. taiwanensis* VLB120 and *P. graminis* C4D1M have been reported, but no Fe(II)/αKG β-hydroxylase was identified in the siderophore gene clusters.

Table 2.3. Putative amino acid β-hydroxylases found in phytotoxin biosynthetic gene clusters

Phytotoxin	Sequenced Producing Strain	Hydroxylase	NCBI Accession	Structure Ref	Cluster ID
Syringomycin	*Pseudomonas syringae* pv. syringae B301D	SyrP	AKF46133.1	(39)	(16, 40)
Nunamycin	*Pseudomonas fluorescens* In5	NupP	KPN90375.1	(41)	(41)
Thanamycin	*Pseudomonas sp.* SH-C52	ThaF	ALG65284.1	(42, 43)	(43)

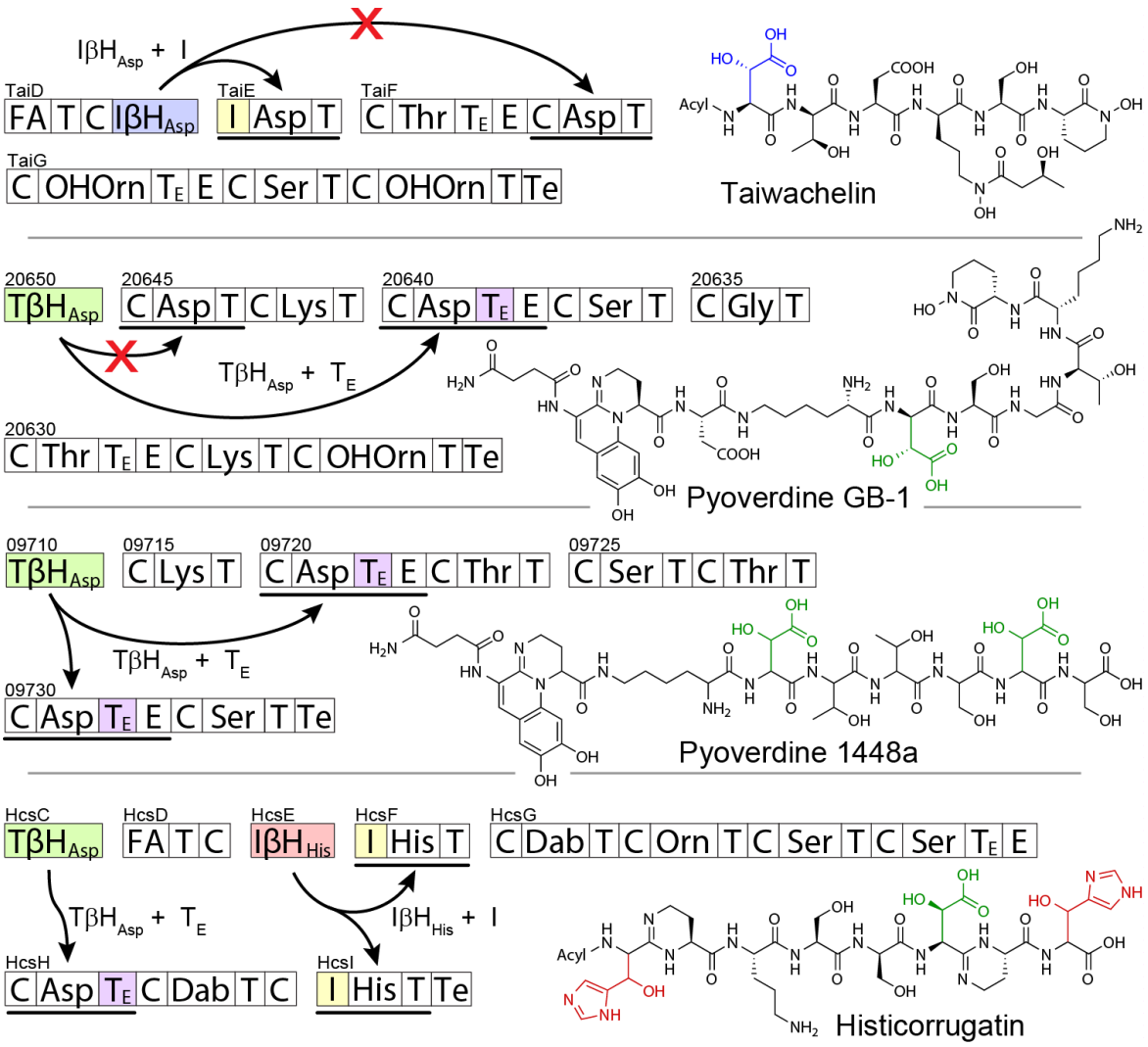


Figure 2.4. Representative reactivity of amino acid β-hydroxylases in siderophore biosynthesis. Aspartyl β-hydroxylases may be *interface*-associated (IβH$_{Asp}$), and interact only with Asp-loading modules with the architecture I-A-T, or *T$_E$*-associated (TβH$_{Asp}$), and interact only with modules containing a GGDSI motif in the thiolation domain. Siderophore histidyl β-hydroxylases are also interface-associated (IβH$_{His}$). The proposed reactivity partners explain the unmodified Asp residues in taiwachelin and pyoverdine GB-1. Genomic data were retrieved from RefSeq (Table 2.5). For NRPS domain abbreviations, see caption of Figure 2.2.

2.3.3. Reactivity by the standalone amino acid β-hydroxylase enzyme

The second class of aspartyl β-hydroxylases, encoded by standalone genes within the biosynthetic cluster, generally acts on Asp loaded by C-A-T-E modules, which leads to D-β-OHAsp in the peptidic siderophore. A sequence-level analysis of these targeted C-A-T-E modules shows a strictly conserved GGDSI (T_E) motif in the thiolation domain in place of the more common GGHSL (T_C) motif. The T_E motif is indicative of a T domain followed by an E domain (44). We therefore name this class of standalone enzymes the *T_E-associated* aspartyl β-hydroxylase ($T\beta H_{Asp}$) family. Several of the standalone $T\beta H_{Asp}$ enzymes hydroxylate Asp attached to modules which lack the E domain, thereby forming L-β-OHAsp; however, the T_E domain is still present. The pyoverdine GB-1 cluster encodes one $T\beta H_{Asp}$ enzyme (PputGB1_4087) and two Asp residues, loaded by C-A-T_C and C-A-T_E-E modules, of which only the latter is hydroxylated (Figure 2.4) (23). On the other hand, the pyoverdine 1448a gene cluster encodes a single $T\beta H_{Asp}$ enzyme (Pspph_RS09710) which hydroxylates Asp bound to two different C-A-T_E-E modules (Figure 2.4) (20).

Histidyl β-hydroxylases form a third functional subtype of amino acid β-hydroxylases. Only three sequenced examples were found in reported siderophores, all from *Pseudomonas* species (Table 2.3). These His β-hydroxylases are standalone enzymes, and they appear alongside an interface domain; therefore, they are herein named $I\beta H_{His}$ (interface-associated *histidyl* β-hydroxylase) enzymes. Histicorrugatin has two β-OHHis residues, both loaded by NRPS I-A-T modules, and the gene cluster only encodes a single $I\beta H_{His}$ enzyme (HcsE; Figure 2.4) (11).

All three siderophore amino acid β-hydroxylase subtypes—$I\beta H_{Asp}$, $T\beta H_{Asp}$, and $I\beta H_{His}$—functionally contrast a fourth subtype found in non-siderophore peptides, which includes the aspartyl β-hydroxylase SyrP of syringomycin biosynthesis (Figure 2.3, Table 2.3). Two

other *SyrP-like* β-hydroxylase ($S\beta H_{Asp}$) enzymes are encoded by orthologous phytotoxin biosynthetic gene clusters, *i.e.* ThaF and NupP of thanamycin and nunamycin biosyntheses, respectively (Figure 2.3, Table 2.3) (41, 43). SyrP, ThaF, and NupP are standalone enzymes that hydroxylate Asp loaded by C-A-T_C modules. They require neither an I domain nor the T_E motif, and a driver of residue selectivity could not be determined.

2.3.4. Phylogeny and sequence analyses of Asp and His β-hydroxylases

To corroborate the existence of distinct $I\beta H_{Asp}$, $T\beta H_{Asp}$, and $I\beta H_{His}$ amino acid β-hydroxylase functional subtypes, we reconstructed an unrooted phylogenetic tree of 30 known β-hydroxylase protein sequences from the 26 siderophore clusters (Figure 2.5). Three members of the SyrP-like family ($S\beta H_{Asp}$) were also included (Table 2.3). All four subtypes formed distinct clades with the exception of $I\beta H_{Asp}$, which is paraphyletic with respect to $S\beta H_{Asp}$ (Figure 2.5).

All of the siderophore amino acid β-hydroxylases belong to the TauD/TfdA family of non-heme Fe(II)/α-ketoglutarate-dependent dioxygenases (PFAM family: PF02668) (45). Crystallographic studies of taurine dioxygenase TauD, 2,4-dichlorophenoxyacetate monooxygenase TfdA, and related enzymes (46-50) have elucidated a 2-His-1-carboxylate facial triad Fe(II) binding motif (51, 52). The α-ketoglutarate cofactor, which coordinates Fe(II), is bound by strictly conserved Thr and Arg residues (53). A multiple sequence alignment of the 33 collected β-hydroxylases across the four subtypes ($I\beta H_{Asp}$, $T\beta H_{Asp}$, $I\beta H_{His}$, and $S\beta H_{Asp}$) reveals 31 residues (~11% of the total domain length) that are absolutely conserved, including the expected combined Fe(II)/α-ketoglutarate binding motif of **H**-N-**E**-X_{24}-**T**-$X_{164\text{-}170}$-**H**-X_8-**R** (Figure 2.6)

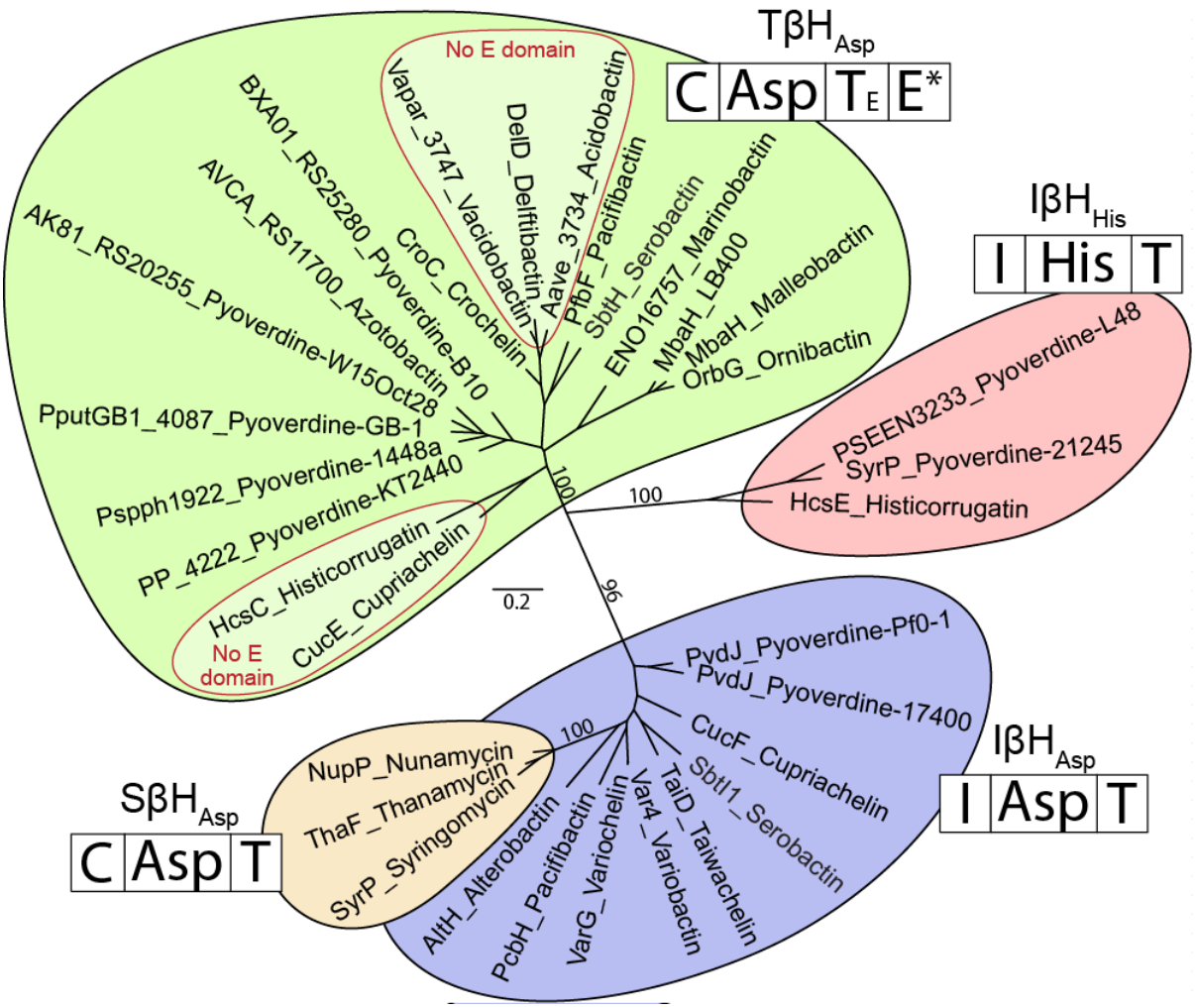


Figure 2.5. Maximum likelihood phylogenetic tree of 33 β-hydroxylases inferred from aligned amino acid sequences. The NRPS module that loads the residue to be hydroxylated is indicated next to each subtype. The two highlighted clades within the $T\beta H_{Asp}$ subtype do not act on E domain-containing modules. The phylogenetic tree was reconstructed in IQ-TREE using the LG+F+R5 model, as chosen by ModelFinder (Akaike information criterion) (54). Branch support was assessed by bootstrapping (100 bootstrap replicates); support values are only shown for inter-subtype branches. The scale bar indicates the average number of substitutions per site.

.

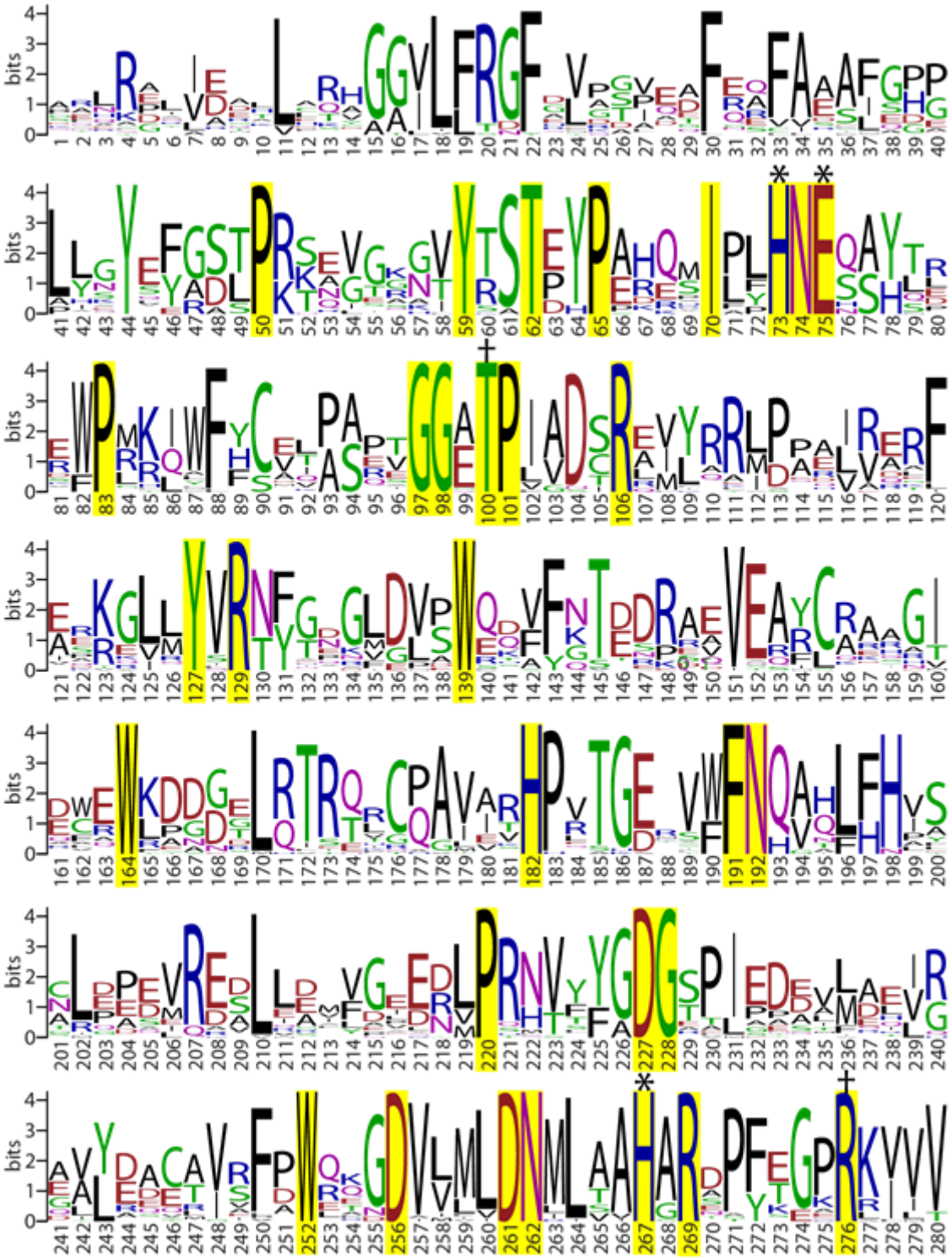


Figure 2.6. Sequence logo of Asp and His β-hydroxylases. Strictly conserved residues are highlighted in yellow; Fe(II) and α-ketoglutarate binding residues are marked with an asterisk (*) and a dagger (†), respectively.

2.3.5. Phylogeny and Sequence Analyses of Interface Domains

Interface domain amino acid sequences were collected from siderophore biosynthetic clusters, including two domains each from histicorrugatin (Figure 2.1) and alterobactin (Figure 2.1), resulting in a dataset of 10 $I\beta H_{Asp}$-associated and 4 $I\beta H_{His}$-associated interface domains (Table 2.4). All 14 of the I domains belong to the PFAM condensation domain family PF00668 (45). Rausch and coworkers previously delineated NRPS condensation domain subtypes and reconstructed their phylogenetic relationship (55). We aligned our extracted I domain sequences to their dataset and reconstructed a maximum-likelihood phylogenetic tree (Figure 2.7). The I domains form a well-supported clade, justifying the assignment of a separate subtype. The multiple sequence alignment shows strikingly few conserved residues; in fact, only nine residues are strictly conserved in the I domain family (Figure 2.8), less than three percent of the total domain length, and none are specific to the I domain. The canonical HHxxxDG active site is absent from I domains; only the Asp remains (Figure 2.7). Crystallographic studies of the condensation domains of VibH and EntF (NRPSs of vibriobactin and enterobactin, respectively) have revealed that this Asp residue is not catalytic, but rather structural, forming a salt bridge with an Arg residue (Figure 2.8: residue 228), and the pair is strictly conserved throughout the condensation domain superfamily (56, 57).

Table 2.4. Interface domains found in siderophore biosynthetic gene clusters.

Siderophore	NRPS Enzyme	Module Architecture	NCBI Accession	Corresponding β-Hydroxylase	βH-ase Type
Alterobactin	AltG	I-A(Asp)-T_C	WP_063365585.1	AltH	IβH_{Asp}
	AltI	I-A(Asp)-T_C-Te	WP_063365570.1		
Cupriachelin	CucG	I-A(Asp)-T_C	WP_082236112.1	CucF	IβH_{Asp}
Pacifibactin	PfbI	I-A(Asp)-T_C	WP_052269209.1	PfbH	IβH_{Asp}
Pyoverdine 17400	PvdK	I-A(Asp)-T_C	WP_029293154.1	PvdJ	IβH_{Asp}
Pyoverdine Pf0-1	PvdK	I-A(Asp)-T_C	WP_011333311.1	PvdJ	IβH_{Asp}
Taiwachelin	TaiE	I-A(Asp)-T_C	WP_012356046.1	TaiD	IβH_{Asp}
Variobactin	Var5	I-A(Asp)-T_C	ALG65340.1	Var4	IβH_{Asp}
Variochelin	VarH	I-A(Asp)-T_C	WP_062469880.1	VarG	IβH_{Asp}
Serobactin	SbtI2	I-A(Asp)-T_C	WP_048348543.1	SbtI1	IβH_{Asp}
Histicorrugatin	HcsF	I-A(His)-T_C	WP_053122086.1	HcsE	IβH_{His}
	HcsI	I-A(His)-T_C-Te	WP_053122092.1		
Pyoverdine pf244	PvdJ	I-A(His)-T_C	AJW67534.1	$SyrP_{21245}$	IβH_{His}
Pyoverdine L48	PSEEN_RS14935	I-A(His)-T_C	WP_011534377.1	PSEEN_RS14940	IβH_{His}

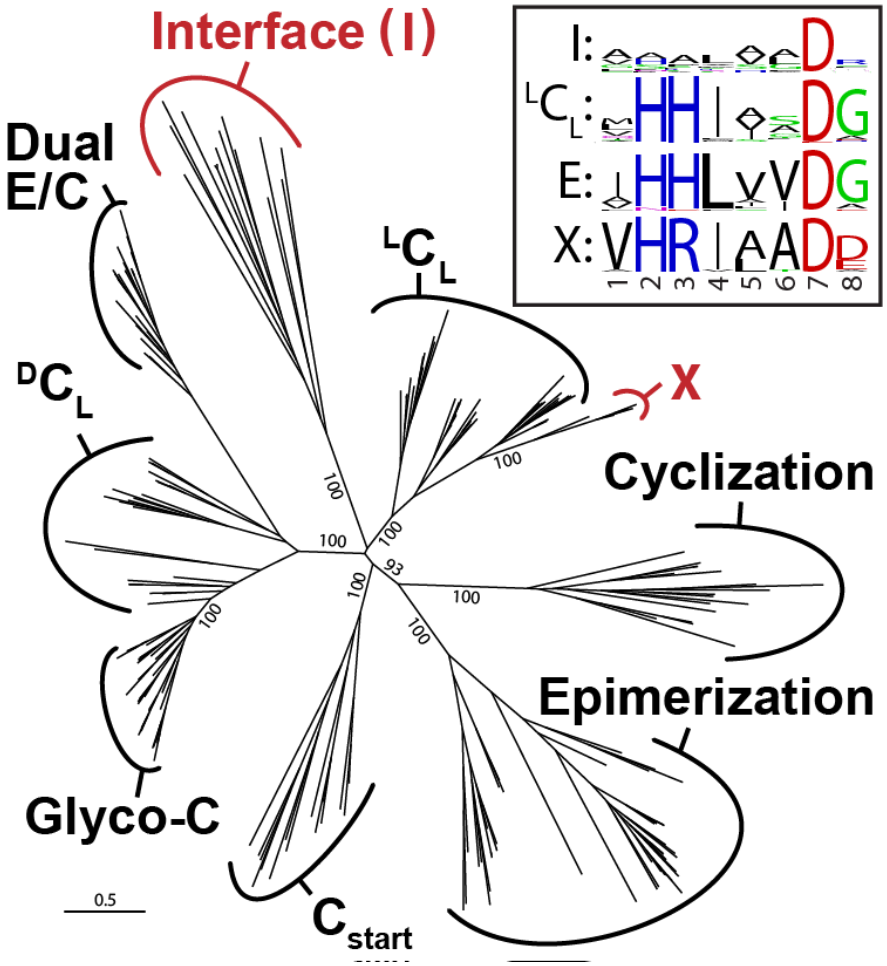


Figure 2.7. Maximum likelihood phylogenetic tree of the condensation domain superfamily and sequence logos of canonical condensation active sites. With the exception of the interface domains, amino acid sequences were originally collected and aligned by Rausch *et al* (55). The full Newick-formatted tree with sequence names is available in Dataset S1. The phylogenetic tree was reconstructed in PhyML v3.0 (58) through http://phylogeny.fr (59). The LG+I+F+G4 model was chosen by ModelFinder (Akaike information criterion)(54). Branch support was assessed by bootstrapping (100 bootstrap replicates). The scale bar indicates the average number of substitutions per site. Inset box: the traditional HHxxxDG condensation domain active site is poorly conserved in interface (I) domains. The epimerization (E) domain retains the motif, while the X domain has a conserved but modified HRxxxDD motif. Sequence logos were made with WebLogo (60), trimming any position with >50% gaps.

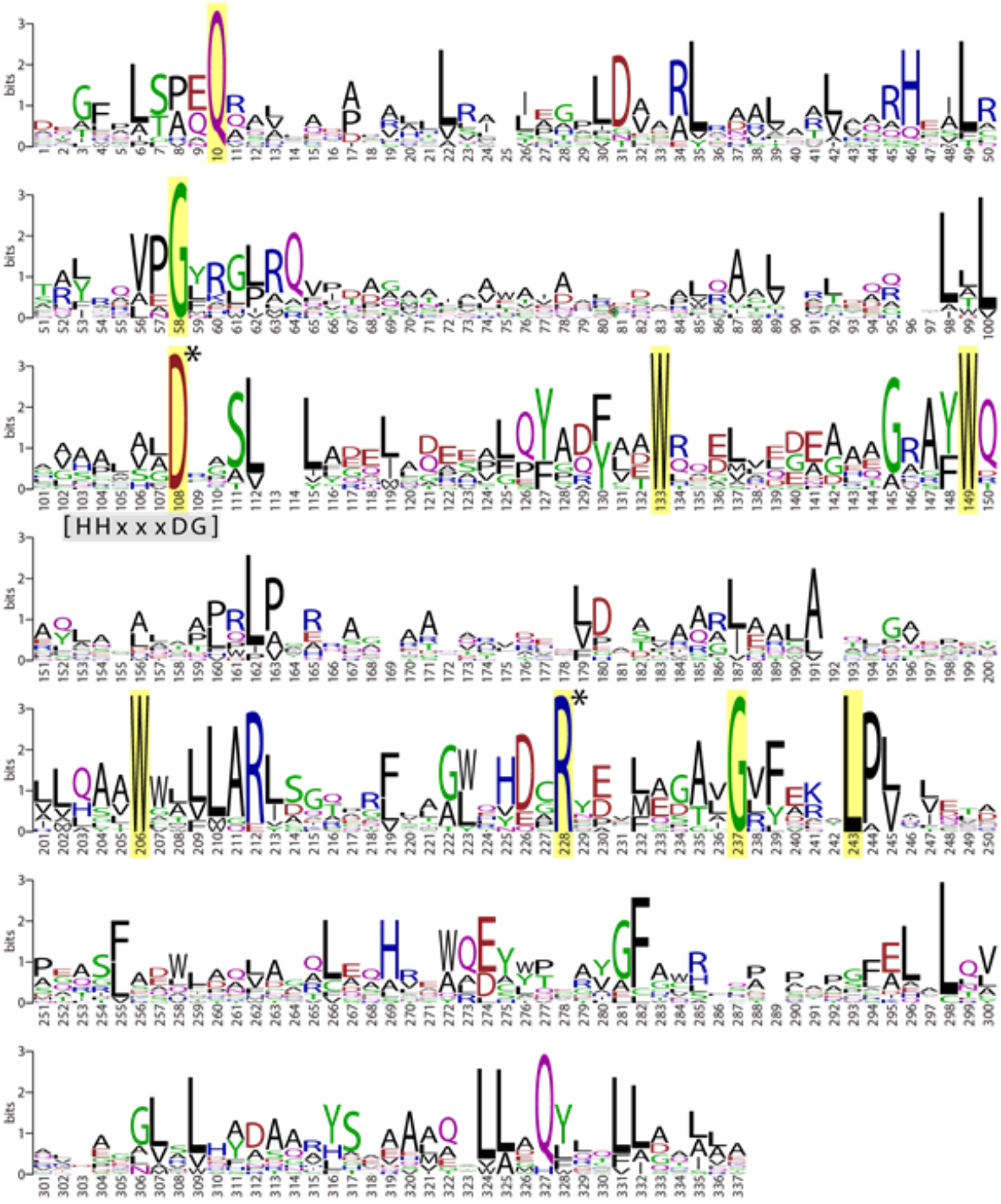


Figure 2.8. Sequence logo of interface domains extracted from siderophore biosynthetic gene clusters. Strictly conserved residues are highlighted in yellow. The canonical condensation domain active site motif, HHxxxDG, is not present. Starred residues Asp108 and Arg228 form a structural salt bridge that is strictly conserved within the condensation domain superfamily.

2.4. Discussion

2.4.1. Stereochemistry of β-OHAsp residues in siderophores

Many of the β-OHAsp-containing siderophores in Table 2.1 have been stereochemically characterized. The 2*S* or 2*R* C2 stereochemistry (*i.e.* L- or D-) of each β-OHAsp residue is consistent with the absence or presence of an epimerization domain in the NRPS (61), while the C3 stereochemistry is set during hydroxylation (Figure 2.9A). The products of seven of nine IβH_{Asp} domains (with the exception of pyoverdine Pf0-1 and variobactin) have reported β-OHAsp stereochemistry, all 3*S* (25, 26, 29, 30, 62-64). The target NRPS module of each IβH_{Asp} domain lacks an E domain, thereby producing the L-*threo* (2*S*, 3*S*) configuration of β-OHAsp (Table 2.1, Figure 2.9A). The stereochemistry of ten of 18 siderophores with TβH_{Asp}-mediated hydroxylation has been reported in the literature (3, 21, 22, 26, 28, 30, 33, 65-67). In contrast to the IβH_{Asp} clade, most of the TβH_{Asp} enzymes were reported to hydroxylate Asp to form the 3*R* product. The TβH_{Asp} target modules generally contain an E domain, resulting in the D-*threo* (2*R*, 3*R*) configuration of β-OHAsp (with exceptions below).

To clarify the reactivity of the TβH_{Asp} family and test the subtype-derived model of β-OHAsp stereochemistry, four representative siderophores from the β-hydroxylase phylogenetic tree in Figure 2.5 were selected for stereochemical characterization of the β-OHAsp residues. Clifford D. Hardy, a graduate student in the Alison Butler group, performed isolation and characterization of the siderophores, assisted by undergraduate researcher Jaewon Suk and visiting researcher Jean Bouvet. Pyoverdine from *Pseudomonas putida* GB-1 was determined to contain β-OHAsp in solely the D-*threo* configuration, while delftibactin from *Delftia acidovorans* DSM 39 and histicorrugatin from *Pseudomonas thivervalensis* DSM 13194 were observed to contain only L-*erythro* β-OHAsp (1, 68). A

reexamination of cupriachelin from *Cupriavidus necator* H16 found two β-OHAsp diastereomers, confirmed as L-*erythro* and L-*threo* (1, 68), in contrast to the previous report of solely L-*threo* β-OHAsp (21). Thus, all TβH$_{Asp}$ enzymes appear to selectively produce the 3*R* diastereomers L-*erythro* and D-*threo* β-OHAsp (Figure 2.9B).

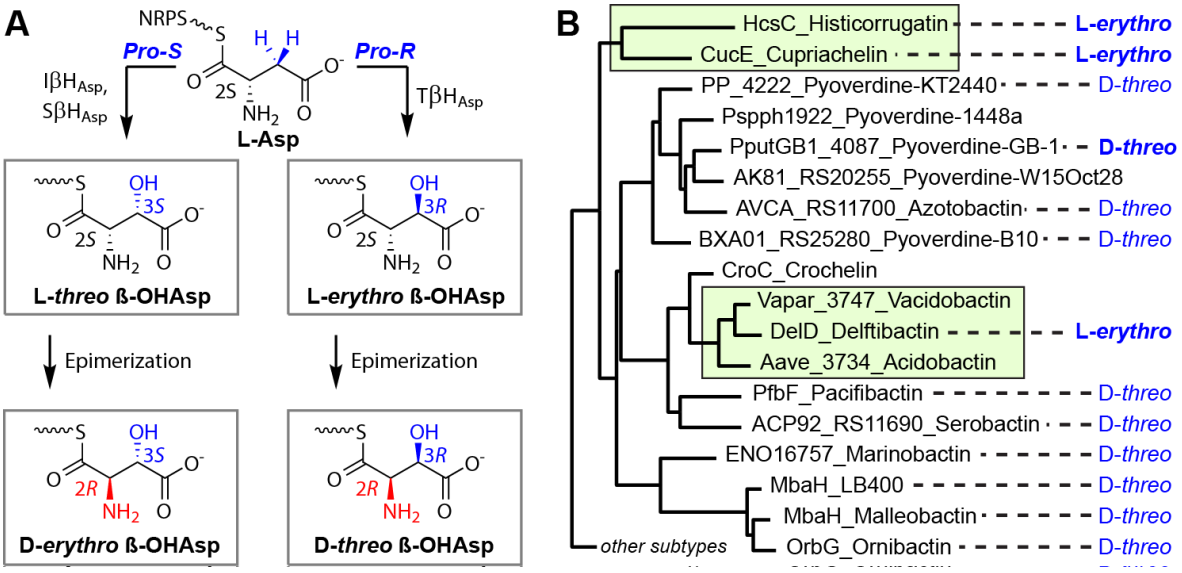
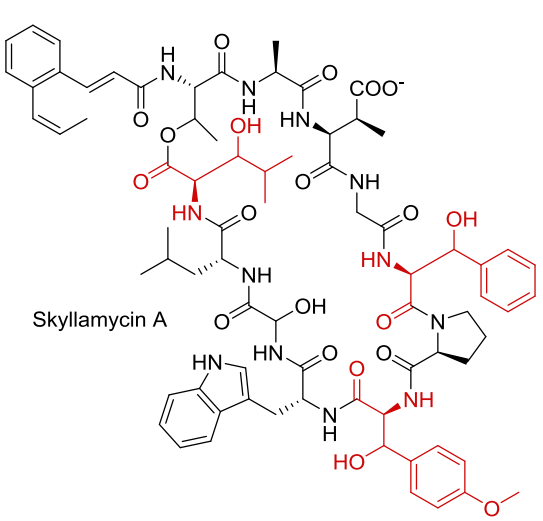


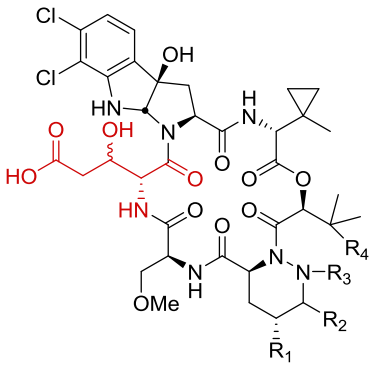
Figure 2.9. Divergent stereochemical products of aspartyl β-hydroxylases. (A) NRPS-bound L-Asp may be β-hydroxylated to give one of two diastereomers. L-*Threo* (2S, 3S) β-OHAsp is produced by the IβH$_{Asp}$ and SβH$_{Asp}$ subtypes. L-*Erythro* (2S, 3R) β-OHAsp is produced by the TβH$_{Asp}$ subtype, and is converted to D-*threo* (2R, 3R) β-OHAsp if an epimerization domain is present. D-*Erythro* β-OHAsp has not been observed in any stereochemically-characterized siderophore. Condensation to the upstream amino acid, which likely precedes epimerization (44), has been omitted for clarity. (B) Stereochemically characterized siderophores produced by members of the TβH$_{Asp}$ clade. The boxed subclades lack an epimerization domain and the hydroxylation product remains as L-*erythro* β-OHAsp. The phylogenetic tree was produced by midpoint rooting the phylogenetic tree from Figure 2.5.

SyrP of syringomycin (Figure 2.2) biosynthesis produces only the L-*threo* isomer *in vitro*, in accordance with the biosynthetic gene cluster, which lacks an E domain, and the structure of syringomycin (16); thus the entire $S\beta H_{Asp}/I\beta H_{Asp}$ clade likely produces L-*threo* β-OHAsp isomers (Figure 2.9A). The β-OHHis residue of pyoverdine pf244 (Figure 2.1) was reported to be in the L-*threo* configuration (5), and we therefore predict that the $I\beta H_{His}$ subtype is also L-*threo* selective. Corrugatin and ornicorrugatin (Figure 2.1), both reported in unsequenced Pseudomonads, were likewise determined to contain L-*threo* β-OHHis, and no other β-OHHis isomer has been reported in siderophores to date (9, 10).

β-Hydroxylation in siderophore biosynthesis is expected to be under strict stereochemical control. Many siderophore β-OHAsp residues have been stereochemically characterized (Table 2.1); all were reported as the L-*threo* (2*S*, 3*S*) or D-*threo* (2*R*, 3*R*) isomers until the recent report of L-*erythro* (2*S*, 3*R*) β-OHAsp in imaqobactin (Figure 2.1), produced by an unsequenced isolate (31). Both L-*erythro* and D-*threo* β-OHAsp share the 3*R* configuration; therefore, D-*threo* could arise from the epimerization of L-*erythro* (2*S*, 3*R*) β-OHAsp (Figure 2.9). The stereochemical determination of β-OHAsp in histicorrugatin and delftibactin (Figure 2.1), as well as the stereochemical re-assignment of one β-OHAsp residue in cupriachelin (Figure 2.1), supports this mechanism(1, 68). Each of these siderophore biosynthetic gene clusters encodes a $T\beta H_{Asp}$ enzyme that acts on a C-A-T_E module lacking an E domain, and each siderophore contains the L-*erythro* isomer. These results do not necessarily preclude the scenario where epimerization preceeds hydroxylation; however, they do show that epimerization is not required for $T\beta H_{Asp}$-mediated β-hydroxylation, and that the α-carbon configuration does not control the stereochemistry of β-hydroxylation. Stereochemical control can instead be attributed to—and be predicted by—the functional subtypes identified herein (Figure 2.9A).

The divergent stereochemistry of the Asp β-hydroxylases is consistent with the reactivity of other Fe(II)/α-ketoglutarate-dependent β-hydroxylases. Kutzneride (Figure 2.10) biosynthesis involves two Fe(II)/αKG *glutamyl* β-hydroxylases, KtzO and KtzP, both of which have been characterized *in vitro* (69). KtzO stereospecifically reacts with NRPS-bound L-Glu to produce L-*threo* (2*S*, 3*R*) β-OHGlu, which is then epimerized to the D-*erythro* (2*R*, 3*R*) isomer (69); KtzP has the opposite selectivity, first producing the L-*erythro* (2*S*, 3*S*) isomer before epimerization to D-*threo* (2*R*, 3*S*) β-OHGlu (69). The Fe_{NH}/αKG-dependent amino acid β-hydroxylation mechanism involves β-hydrogen abstraction by a reactive Fe(IV)-oxo species, followed by rebound hydroxylation to form the alcohol (2, 70). To achieve such stereospecificity, the amino acid substrate must be oriented in the active site with only one β-hydrogen (*pro-threo or pro-erythro*) oriented toward the active site. A comparison of the product-bound crystal structures of AsnO, a *threo*-selective L-Asn β-hydroxylase, and VioC, an *erythro*-selective L-Arg β-hydroxylase, revealed that the two homologs hold their substrates in different rotational conformations: in AsnO, the sidechain of L-Asn is held *trans*, pointed towards the center of the enzyme, while VioC holds L-Arg in a more strained *gauche*(–) conformation, pointing the sidechain towards the enzyme surface (49, 50). Similarly, we suspect that the contrasting β-OHAsp stereochemistry is caused by a difference in L-Asp positioning.





	β-OHGlu	R_1	R_2	R_3	R_4
Kutzneride 1	3S	H	H	H	Me
Kutzneride 2	3S	Cl	H	H	Me
Kutzneride 3	3R	H	H	H	Me
Kutzneride 4	3R	H	π-bond		Me
Kutzneride 5	3R	H	H	H	H
Kutzneride 6	3R	OH	π-bond		Me
Kutzneride 7	3S	H	H	H	H
Kutzneride 8	3R	Cl	H	H	Me
Kutzneride 9	3S	H	π-bond		Me

Figure 2.10. Select specialized metabolites with ß-hydroxylated amino acids. Hydroxylated residues are shown in red.

2.4.2. Application of β-hydroxylase subtypes towards improved siderophore structural predictions

In sum, the chemistry of siderophore aspartyl β-hydroxylation follows two divergent routes (Figures 2.4 and 2.9A, Table 2.1). Asp β-hydroxylase domains fused to the C-terminus of a NRPS enzyme selectively hydroxylate Asp bound to a NRPS module in the presence of an interface (I) domain, a member of the condensation domain superfamily newly described herein (Figure 2.7). The *interface-associated* aspartyl β-hydroxylase ($I\beta H_{Asp}$) domains exclusively produce the L-*threo* (2*S*, 3*S*) isomer of β-OHAsp (Table 2.1, Figure 2.9A). In the second pathway, Asp β-hydroxylases encoded by discrete genes in the biosynthetic cluster selectively act on Asp bound to thiolation domains with the T_E sequence motif (GGDSI), and are consequently named the *T_E-associated* aspartyl β-hydroxylase ($T\beta H_{Asp}$) enzymes. All of these standalone enzymes produce L-*erythro* (2*S*, 3*R*) or D-*threo*

(2*R*, 3*R*) β-OHAsp, requiring opposite stereospecificity of β-hydroxylation relative to $I\beta H_{Asp}$ domains (Table 2.1, Figure 2.9). Remarkably, the two drivers of residue specificity proposed here (*i.e.*, the interface domain, and the T_E-type thiolation domain, Figure 2.4) parallel selectivity seen in the cytochrome P450 family of enzymes involved in glycopeptide antibiotic and skyllamycin biosyntheses (71) (*vide infra*).

The reactivity of an undescribed siderophore aspartyl β-hydroxylase can now be predicted based solely on whether the β-hydroxylase is fused to the NRPS machinery ($I\beta H_{Asp}$) or a free-standing enzyme ($T\beta H_{Asp}$). The two drivers of residue specificity we propose here (*i.e.*, the interface domain and the T_E-type thiolation domain) then allow for quick prediction of hydroxylation sites. A more rigorous subtype determination can be made using profile hidden Markov models (pHMMs), probabilistic representations of the amino acid sequences. pHMMs for each subtype may be incorporated into domain and structure prediction workflows with HMMER3 (72). A sequence that poorly matches each of the four pHMMs (bitscore < 400) may belong to a new functional subtype distinct from the subtypes described herein. For example, glutamyl β-hydroxylases KtzO and KtzP of kutzneride (Figure 2.10) biosynthesis (69) best match the $I\beta H_{His}$ pHMM with bitscores of 340 and 224, respectively, well below the cutoff.

2.4.3. Parallels to glycopeptide antibiotic (GPA) biosynthesis

GPA biosynthesis provides an example of a specialized NRPS domain responsible for positioning a standalone hydroxylase. After NRPS-based assembly, the GPA peptide precursor, still attached to the NRPS, is extensively crosslinked by several cytochrome P450 (Oxy) enzymes (73-75). The final module of the NRPS Tcp12 of teicoplanin biosynthesis contains a member of the C domain superfamily called the X domain, which is missing the

HHxxxDG catalytic motif (Figure 2.7) (55). Haslinger *et al*. found through X-ray crystallography that the X domain is required to recruit the Oxy enzymes to the NRPS-bound peptide through protein-protein interaction (76). We propose the I domain function resembles that of the X domain, although this similarity is likely superficial, as the two subfamilies are distantly related phylogenetically (Figure 2.7) and interact with entirely different protein families (*i.e.*, non-heme Fe(II) dioxygenases and Fe-heme monooxygenases, respectively). Additionally, the X domain has a conserved, albeit modified, HRxxxD[DE] motif that blocks the traditional condensation active site (76), in contrast to the poorly-conserved active site of the I domain (Figure 2.7).

2.4.4. Parallels to skyllamycin biosynthesis.

The T_E motif GGDSI is generally found when an epimerization (E) domain follows a T domain, and contrasts the GGHSL (T_C) core motif usually found in T domains (44). Mutational studies showed that the T_E motif is required for proper interaction between the T and E domains (44). $T\beta H_{Asp}$ enzymes may also require the GGDSI motif to properly interact with the T-bound amino acid. Cytochrome $P450_{sky}$ of skyllamycin (Figure 2.10) biosynthesis similarly selects residues for β-hydroxylation by interacting only with specific T domains, as determined by X-ray crystallography (77). In contrast to the T_E domain, even non-interacting T domains within the skyllamycin pathway contain the key residues for $P450_{sky}$ interaction, and selectivity is believed to arise from minor changes in T domain tertiary structure that disrupt the NRPS/$P450_{sky}$ interface (71, 77). $T\beta H_{Asp}$ enzymes may similarly select modules for binding based on subtle structural changes, undetectable in the primary sequence of the NRPS. Regardless, the T_E motif serves as a useful predictor of selectivity in the $T\beta H_{Asp}$ family.

2.5. Conclusion

Genome mining is streamlining the discovery and structural elucidation of new specialized metabolites, but can leave ambiguities in the predicted structures that must be rectified experimentally. Comparing siderophore biosynthetic gene clusters to verified structures reveals the origin of the β-OHAsp diastereomers in siderophores, providing both predictive tools and avenues for future research. We have identified three functional subtypes of β-hydroxylases involved in siderophore biosynthesis (*i.e.*, $I\beta H_{Asp}$, $T\beta H_{Asp}$, and $I\beta H_{His}$), placing their reactivity in contrast to SyrP (16) and SyrP-like enzymes ($S\beta H_{Asp}$). These newly delineated subtypes, validated by phylogenetic reconstruction (Figure 2.5), show clear patterns in genomic organization (standalone enzymes or integrated NPRS tailoring domains), amino acid substrate (Asp or His), reactive NRPS partner (IAT, CAT_E, or CAT_C), and stereochemistry (L-*threo*, L-*erythro*, or D-*threo*) (Table 2.1).

With two stereocenters, β-OHAsp can exist as any of four diastereomers. Through mapping stereochemically-characterized β-OHAsp residues in siderophores to the phylogenetic tree of β-hydroxylases, we have developed the first method to predict β-OHAsp stereochemistry *in silico*. While $I\beta H_{Asp}$ domains consistently produce L-*threo* (2*S*, 3*S*) β-OHAsp, the $T\beta H_{Asp}$ subtype produces the L-*erythro* (2*S*, 3*R*) isomer, which is often epimerized to D-*threo* (2*R*, 3*R*) β-OHAsp (Figure 2.9A). The D-*erythro* (2*R*, 3*S*) stereoisomer has not been identified in any siderophore, but would be consistent with an $I\beta H_{Asp}$ domain paired with epimerization by an I-A-T-E NRPS architecture. The contrasting *pro-R* versus *pro-S* β-hydroxylation that we identified may have arisen from a reconfiguration of the active site, holding the aspartyl substrate in a different orientation. Future work will elucidate the structural basis of Asp β-hydroxylase stereospecificity.

2.6. Appendix

Table 2.5. Amino acid β-hydroxylases found in siderophore biosynthetic gene clusters.

Siderophore	Producing Strain	Hydroxylase	NCBI Accession	References		
				Structure	Cluster	Stereo
Alterobactin	*Alteromonas luteoviolacea* DSM 6061	AltH	WP_063365570.1	(62)	This work	(62)
Pyoverdine 17400	*Pseudomonas fluorescens* ATCC 17400	PvdJ	WP_081041737.1	(63)	(78)	(63)
Pyoverdine Pf0-1	*Pseudomonas fluorescens* Pf0-1	PvdJ	WP_041475189.1	(64)	(79)	-
Taiwachelin	*Cupriavidus taiwanensis* LMG 19424	TaiD	WP_012356045.1	(25)	(25)	(25)
Variobactin	*Variovorax paradoxus* P4B	Var4	ALG65339.1	(43)	(43)	-
Variochelin	*Variovorax boronicumulans* NBRC 103145	VarG	WP_062469881.1	(29)	(29)	(29)
Cupriachelin	*Cupriavidus necator* H16	CucF	WP_011617408.1	(21)	(21)	(21), This work
		CucE	WP_011617409.1			
Pacifibactin	*Alcanivorax pacificus* W11-5	PfbH	WP_008737432.1	(30)	(30)	(30)
		PfbF	WP_008737428.1			
Serobactin	*Herbaspirillum seropedicae* SmR1	SbtI1	WP_013234320.1	(26)	(26)	(26)
		SbtH	WP_053075621.1			
Acidobactin	*Acidovorax citrulli* AAC00-1	Aave_3734	WP_011796775.1	(43)	(43)	-
Azotobactin	*Azotobacter vinelandii* CA	AVCA_RS11700	WP_012701132.1	(65)	(80)	(65)
Crochelin	*Azotobacter chroococcum* NCIMB 8003	CroC	WP_039806858.1	(81)	(81)	-
Delftibactin	*Delftia acidovorans* SPH-1	DelD	WP_012206556.1	(82)	(82)	This work
LB400	*Burkholderia xenovorans* LB400	MbaH	WP_011492714.1	(28)	(28)	(28)
Malleobactin	*Burkholderia pseudomallei* K96243	MbaH	WP_004526940.1	(22)	(22)	(22)
Marinobactin	*Marinobacter nanhaiticus* D15-8W	ENO16757	WP_004583267.1	(33)	(24)	(33)
Ornibactin	*Burkholderia cenocepacia* J2315	OrbG	WP_006486082.1	(66)	(19)	(66)
Pyoverdine 1448a	*Pseudomonas syringae* 1448a	Pspph1922	WP_004664616.1	(20)	(20)	-
Pyoverdine B10	*Pseudomonas* sp. B10	BXA01_RS25280	WP_076567019.1	(3)	(83)	(3)
Pyoverdine GB-1	*Pseudomonas putida* GB-1	PputGB1_4087	WP_012273660.1	(23)	(23)	This work

Pyoverdine G4R	*Pseudomonas putida* KT2440	PP_4222	WP_010954977.1	(7, 12, 67)	(7, 79)	(67)
Pyoverdine W15Oct28	Pseudomonas putida W15Oct28	AK81_RS20255	WP_043201845.1	(84)	(84)	-
Vacidobactin	*Variovorax paradoxus* S110	Vapar_3747	WP_015866525.1	(43)	(43)	-
Histicorrugatin	*Pseudomonas thivervalensis* LMG 21626	HcsC	WP_082343559.1	(11)	(11)	This work
		HcsE	WP_053122094.1			
Pyoverdine pf224	*Pseudomonas taiwanensis* DSM 21245	$SyrP_{21245}$	AJW67533.1	(5, 8)	(8)	(5)
Pyoverdine L48	*Pseudomonas entomophila* L48	PSEEN_RS14940	WP_011534378.1	(7)	(7)	-

2.7. References

1. Reitz ZL, Hardy CD, Suk J, Bouvet J, & Butler A (2019) Genomic analysis of siderophore β-hydroxylases reveals divergent stereocontrol and expands the condensation domain family. *Proc. Natl. Acad. Sci. U. S. A.* 116(40):19805-19814.
2. Hardy CD & Butler A (2018) β-Hydroxyaspartic acid in siderophores: biosynthesis and reactivity. *J. Biol. Inorg. Chem.* 23(7):957-967.
3. Teintze M, Hossain MB, Barnes CL, Leong J, & van der Helm D (1981) Structure of ferric pseudobactin, a siderophore from a plant growth promoting Pseudomonas. *Biochemistry* 20(22):6446-6457.
4. Barbeau K, Rue EL, Bruland KW, & Butler A (2001) Photochemical cycling of iron in the surface ocean mediated by microbial iron(iii)-binding ligands. *Nature* 413(6854):409-413.
5. Hancock DK, *et al.* (1993) L - threo -β-hydroxyhistidine, an unprecedented iron(III) ion-binding amino acid in a pyoverdine-type siderophore from Pseudomonas fluorescens 244. *J. Chem. Soc. Chem. Commun.* 0(5):468-470.
6. Budzikiewicz H, Kilz S, Taraz K, & Meyer JM (1997) Identical Pyoverdines from Pseudomonas fluorescens 9AW and from Pseudomonas putida 9BW. *Zeitschrift für Naturforschung C* 52:721.
7. Matthijs S, *et al.* (2009) Siderophore-mediated iron acquisition in the entomopathogenic bacterium Pseudomonas entomophila L48 and its close relative Pseudomonas putida KT2440. *Biometals* 22(6):951-964.
8. Chen W-J, *et al.* (2016) Involvement of type VI secretion system in secretion of iron chelator pyoverdine in Pseudomonas taiwanensis. *Sci. Rep.* 6:32950.
9. Risse D, Beiderbeck H, Taraz K, Budzikiewicz H, & Gustine D (1998) Corrugatin, a lipopeptide siderophore from Pseudomonas corrugata. *Zeitschrift für Naturforschung C* 53c:295-304.
10. Matthijs S, Budzikiewicz H, Schäfer M, Wathelet B, & Cornelis P (2008) Ornicorrugatin, a New Siderophore from Pseudomonas fluorescens AF76. *Zeitschrift für Naturforschung C* 63(1-2).
11. Matthijs S, *et al.* (2016) Pyoverdine and histicorrugatin-mediated iron acquisition in Pseudomonas thivervalensis. *Biometals* 29(3):467-485.
12. Baune M, Qi Y, Scholz K, Volmer DA, & Hayen H (2017) Structural characterization of pyoverdines produced by Pseudomonas putida KT2440 and Pseudomonas taiwanensis VLB120. *Biometals* 30(4):589-597.
13. Scholz K, Tiso T, Blank LM, & Hayen H (2018) Mass spectrometric characterization of siderophores produced by Pseudomonas taiwanensis VLB120 assisted by stable isotope labeling of nitrogen source. *Biometals* 31(5):785-795.
14. Chen H, Hubbard BK, O'Connor SE, & Walsh CT (2002) Formation of beta-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. *Chem. Biol.* 9(1):103-112.
15. Makris TM, Chakrabarti M, Münck E, & Lipscomb JD (2010) A family of diiron monooxygenases catalyzing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 107(35):15391-15396.
16. Singh GM, Fortin PD, Koglin A, & Walsh CT (2008) beta-Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: characterization of two candidate hydroxylases AspH and SyrP in Pseudomonas syringae. *Biochemistry* 47(43):11310-11320.

17. Ökesli A, Cooper LE, Fogle EJ, & van der Donk WA (2011) Nine Post-translational Modifications during the Biosynthesis of Cinnamycin. *Journal of the American Chemical Society* 133(34):13753-13760.
18. An L, *et al.* (2018) Substrate-assisted enzymatic formation of lysinoalanine in duramycin. *Nature Chemical Biology* 14(10):928-933.
19. Agnoli K, Lowe CA, Farmer KL, Husnain SI, & Thomas MS (2006) The ornibactin biosynthesis and transport genes of Burkholderia cenocepacia are regulated by an extracytoplasmic function sigma factor which is a part of the Fur regulon. *J. Bacteriol.* 188(10):3631-3644.
20. Owen JG & Ackerley DF (2011) Characterization of pyoverdine and achromobactin in Pseudomonas syringae pv. phaseolicola 1448a. *BMC Microbiol.* 11:218.
21. Kreutzer MF, Kage H, & Nett M (2012) Structure and Biosynthetic Assembly of Cupriachelin, a Photoreactive Siderophore from the Bioplastic Producer Cupriavidus necator H16. *J. Am. Chem. Soc.* 134(11):5415-5422.
22. Franke J, Ishida K, Ishida-Ito M, & Hertweck C (2013) Nitro versus hydroxamate in siderophores of pathogenic bacteria: effect of missing hydroxylamine protection in malleobactin biosynthesis. *Angew. Chem. Int. Ed Engl.* 52(32):8271-8275.
23. Parker DL, *et al.* (2014) Pyoverdine synthesis by the Mn(II)-oxidizing bacterium Pseudomonas putida GB-1. *Front. Microbiol.* 5:202.
24. Kem MP, Naka H, Iinishi A, Haygood MG, & Butler A (2015) Fatty acid hydrolysis of acyl marinobactin siderophores by Marinobacter acylases. *Biochemistry* 54(3):744-752.
25. Kreutzer MF & Nett M (2012) Genomics-driven discovery of taiwachelin, a lipopeptide siderophore from Cupriavidus taiwanensis. *Org. Biomol. Chem.* 10(47):9338-9343.
26. Rosconi F, *et al.* (2013) Identification and structural characterization of serobactins, a suite of lipopeptide siderophores produced by the grass endophyte Herbaspirillum seropedicae. *Environ. Microbiol.* 15(3):916-927.
27. Skinnider MA, *et al.* (2015) Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* 43(20):9645-9662.
28. Vargas-Straube MJ, *et al.* (2016) Genetic and Functional Analysis of the Biosynthesis of a Non-Ribosomal Peptide Siderophore in Burkholderia xenovorans LB400. *PLoS One* 11(3):e0151273.
29. Kurth C, Schieferdecker S, Athanasopoulou K, Seccareccia I, & Nett M (2016) Variochelins, Lipopeptide Siderophores from Variovorax boronicumulans Discovered by Genome Mining. *J. Nat. Prod.* 79(4):865-872.
30. Hardy CD & Butler A (2019) On the ambiguity of NRPS product structure predictions: four bidentate chelating groups in the siderophore pacifibactin. *J. Nat. Prod.* 82(4):990-998.
31. Robertson AW, *et al.* (2018) Isolation of Imaqobactin, an Amphiphilic Siderophore from the Arctic Marine Bacterium Variovorax Species RKJM285. *J. Nat. Prod.* 81(4):858-865.
32. Bloudoff K & Schmeing TM (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta* 1865(11 Pt B):1587-1604.
33. Martinez JS, *et al.* (2000) Self-assembling amphiphilic siderophores from marine bacteria. *Science* 287(5456):1245-1247.
34. Figueroa LO, Schwarz B, & Richards AM (2015) Structural characterization of amphiphilic siderophores produced by a soda lake isolate, Halomonas sp. SL01, reveals

cysteine-, phenylalanine- and proline-containing head groups. *Extremophiles* 19(6):1183-1192.
35. Homann VV, *et al.* (2009) Loihichelins A-F, a suite of amphiphilic siderophores produced by the marine bacterium Halomonas LOB-5. *J. Nat. Prod.* 72(5):884-888.
36. Richards AM (2007) Identification and Structural Characterization of Siderophores Produced by Halophilic and Alkaliphilic Bacteria. (Washington State University).
37. Kanoh K, Kamino K, Leleo G, Adachi K, & Shizuri Y (2003) Pseudoalterobactin A and B, new siderophores excreted by marine bacterium Pseudoalteromonas sp. KP20-4. *J. Antibiot.* 56(10):871-875.
38. Hermenau R, *et al.* (2018) Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. *Nat. Chem. Biol.* 14:841-843.
39. Fukuchi N, *et al.* (1992) Structure and stereochemistry of three phytotoxins, syringomycin, syringotoxin and syringostatin, produced by pseudomonas syringae pv. syringae. *J. Chem. Soc. Perkin 1* 0(9):1149-1157.
40. Zhang JH, Quigley NB, & Gross DC (1997) Analysis of the syrP gene, which regulates syringomycin synthesis by Pseudomonas syringae pv. syringae. *Appl. Environ. Microbiol.* 63(7):2771-2778.
41. Michelsen CF, *et al.* (2015) Nonribosomal peptides, key biocontrol components for Pseudomonas fluorescens In5, isolated from a Greenlandic suppressive soil. *MBio* 6(2):e00079.
42. Watrous J, *et al.* (2012) Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U. S. A.* 109(26):E1743-1752.
43. Johnston CW, *et al.* (2015) An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat. Commun.* 6:8421.
44. Linne U, Doekel S, & Marahiel MA (2001) Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. *Biochemistry* 40(51):15824-15834.
45. Finn RD, *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279-285.
46. Zhang Z, *et al.* (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* 7(2):127-133.
47. Clifton IJ, *et al.* (2003) Crystal structure of carbapenem synthase (CarC). *J. Biol. Chem.* 278(23):20843-20850.
48. Müller I, Stückl C, Wakeley J, Kertesz M, & Usón I (2005) Succinate complex crystal structures of the alpha-ketoglutarate-dependent dioxygenase AtsK: steric aspects of enzyme self-hydroxylation. *J. Biol. Chem.* 280(7):5716-5723.
49. Strieker M, Kopp F, Mahlert C, Essen L-O, & Marahiel MA (2007) Mechanistic and structural basis of stereospecific Cbeta-hydroxylation in calcium-dependent antibiotic, a daptomycin-type lipopeptide. *ACS Chem. Biol.* 2(3):187-196.
50. Helmetag V, Samel SA, Thomas MG, Marahiel MA, & Essen L-O (2009) Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *FEBS J.* 276(13):3669-3682.
51. Hegg EL & Que L, Jr. (1997) The 2-His-1-carboxylate facial triad—an emerging structural motif in mononuclear non-heme iron (II) enzymes. *Eur. J. Biochem.* 250(3):625-629.
52. Hogan DA, Smith SR, Saari EA, McCracken J, & Hausinger RP (2000) Site-directed Mutagenesis of 2,4-Dichlorophenoxyacetic Acid/α-Ketoglutarate Dioxygenase:

IDENTIFICATION OF RESIDUES INVOLVED IN METALLOCENTER FORMATION AND SUBSTRATE BINDING. *J. Biol. Chem.* 275(17):12400-12409.
53. Elkins JM, *et al.* (2002) X-ray crystal structure of Escherichia coli taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* 41(16):5185-5192.
54. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, & Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14(6):587-589.
55. Rausch C, Hoof I, Weber T, Wohlleben W, & Huson DH (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* 7:78.
56. Keating TA, Marshall CG, Walsh CT, & Keating AE (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* 9(7):522-526.
57. Roche ED & Walsh CT (2003) Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* 42(5):1334-1344.
58. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59(3):307-321.
59. Dereeper A, *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36(Web Server issue):W465-469.
60. Crooks GE, Hon G, Chandonia J-M, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res.* 14(6):1188-1190.
61. Luo L, *et al.* (2002) Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* 41(29):9184-9196.
62. Reid RT, Livet DH, Faulkner DJ, & Butler A (1993) A siderophore from a marine bacterium with an exceptional ferric ion affinity constant. *Nature* 366:455.
63. Demange P, Bateman A, Macleod JK, Dell A, & Abdallah MA (1990) Bacterial siderophores: unusual 3,4,5,6-tetrahydropyrimidine-based amino acids in pyoverdins frompseudomonas fluorescens. *Tetrahedron Lett.* 31(52):7611-7614.
64. Meyer J-M, *et al.* (2008) Siderotyping of fluorescent Pseudomonas: molecular mass determination by mass spectrometry as a powerful pyoverdine siderotyping method. *Biometals* 21(3):259-271.
65. Demange P, Bateman A, Dell A, & Abdallah MA (1988) Structure of azotobactin D, a siderophore of Azotobacter vinelandii strain D (CCM 289). *Biochemistry* 27(8):2745-2752.
66. Stephan H, *et al.* (1993) Ornibactins--a new family of siderophores from Pseudomonas. *Biometals* 6(2):93-100.
67. Atkinson RA, Salah El Din AL, Kieffer B, Lefevre JF, & Abdallah MA (1998) Bacterial iron transport: 1H NMR determination of the three-dimensional structure of the gallium complex of pyoverdin G4R, the peptidic siderophore of Pseudomonas putida G4R. *Biochemistry* 37(45):15965-15973.
68. Hardy CD (2020) Characterization of β-Hydroxyaspartate Siderophroes: Structures, Biosyntheses, Coordination Chemistry, and Photoreactivity. PhD (University of California, Santa Barbara, Santa Barbara, CA).

69. Strieker M, Nolan EM, Walsh CT, & Marahiel MA (2009) Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis. *J. Am. Chem. Soc.* 131(37):13523-13530.
70. Mitchell AJ, *et al.* (2017) Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* 139(39):13830-13836.
71. Peschke M, Gonsior M, Süssmuth RD, & Cryle MJ (2016) Understanding the crucial interactions between Cytochrome P450s and non-ribosomal peptide synthetases during glycopeptide antibiotic biosynthesis. *Curr. Opin. Struct. Biol.* 41:46-53.
72. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7(10):e1002195.
73. Zerbe K, *et al.* (2004) An oxidative phenol coupling reaction catalyzed by oxyB, a cytochrome P450 from the vancomycin-producing microorganism. *Angew. Chem. Int. Ed Engl.* 43(48):6709-6713.
74. Bischoff D, *et al.* (2005) The biosynthesis of vancomycin-type glycopeptide antibiotics--a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. *Chembiochem* 6(2):267-272.
75. Peschke M, Brieke C, Heimes M, & Cryle MJ (2018) The Thioesterase Domain in Glycopeptide Antibiotic Biosynthesis Is Selective for Cross-Linked Aglycones. *ACS Chem. Biol.* 13(1):110-120.
76. Haslinger K, Peschke M, Brieke C, Maximowitsch E, & Cryle MJ (2015) X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* 521(7550):105-109.
77. Haslinger K, *et al.* (2014) The structure of a transient complex of a nonribosomal peptide synthetase and a cytochrome P450 monooxygenase. *Angew. Chem. Int. Ed Engl.* 53(32):8518-8522.
78. Ye L, *et al.* (2014) Analysis of the draft genome of Pseudomonas fluorescens ATCC17400 indicates a capacity to take up iron from a wide range of sources, including different exogenous pyoverdines. *Biometals* 27(4):633-644.
79. Ravel J & Cornelis P (2003) Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends Microbiol.* 11(5):195-200.
80. Yoneyama F, Yamamoto M, Hashimoto W, & Murata K (2011) Azotobacter vinelandii gene clusters for two types of peptidic and catechol siderophores produced in response to molybdenum. *J. Appl. Microbiol.* 111(4):932-938.
81. Baars O, *et al.* (2017) Crochelins, siderophores with a novel iron-chelating moiety from the nitrogen-fixing bacterium Azotobacter chroococcum. *Angew. Chem. Int. Ed Engl.* 57:536-541.
82. Johnston CW, *et al.* (2013) Gold biomineralization by a metallophore from a gold-associated microbe. *Nat. Chem. Biol.* 9(4):241-243.
83. Ambrosi C, Leoni L, Putignani L, Orsi N, & Visca P (2000) Pseudobactin biogenesis in the plant growth-promoting rhizobacterium Pseudomonas strain B10: identification and functional analysis of the L-ornithine N(5)-oxygenase (psbA) gene. *J. Bacteriol.* 182(21):6233-6238.
84. Ye L, *et al.* (2013) A combinatorial approach to the structure elucidation of a pyoverdine siderophore produced by a Pseudomonas putida isolate and the use of pyoverdine as a taxonomic marker for typing P. putida subspecies. *Biometals* 26(4):561-575.

Chapter 3. Targeted genome mining of spacer-containing triscatechol siderophores

3.1. Introduction

The triscatechol siderophores trivanchrobactin, cyclic trichrysobactin, and turnerbactin (Figure 3.1) are an apparent expansion of the well-known enterobactin structure. Each features a different cationic spacer amino acid (D-Arg, D-Lys, and L-Orn, respectively) between 2,3-dihydroxybenzoate (DHB) and the triserine core. The elongated spacer-containing arms may prevent piracy by competing bacterial strains while only minimally affecting Fe(III) affinity.[1,2] However, the significance of the D- *versus* L- spacer configuration remains unknown. Upon Fe(III) coordination, these siderophores gain an additional stereocenter. Triscatecholate siderophores like enterobactin may adopt either a Δ or Λ configuration around the ferric coordination centre, with a *facial* configuration of the asymmetric 2,3-DHB moieties (Figure 3.2). Crucially, the "wrong" Fe(III)-center chirality has been implicated in reducing siderophore efficacy by hindering receptor recognition.[3-7]

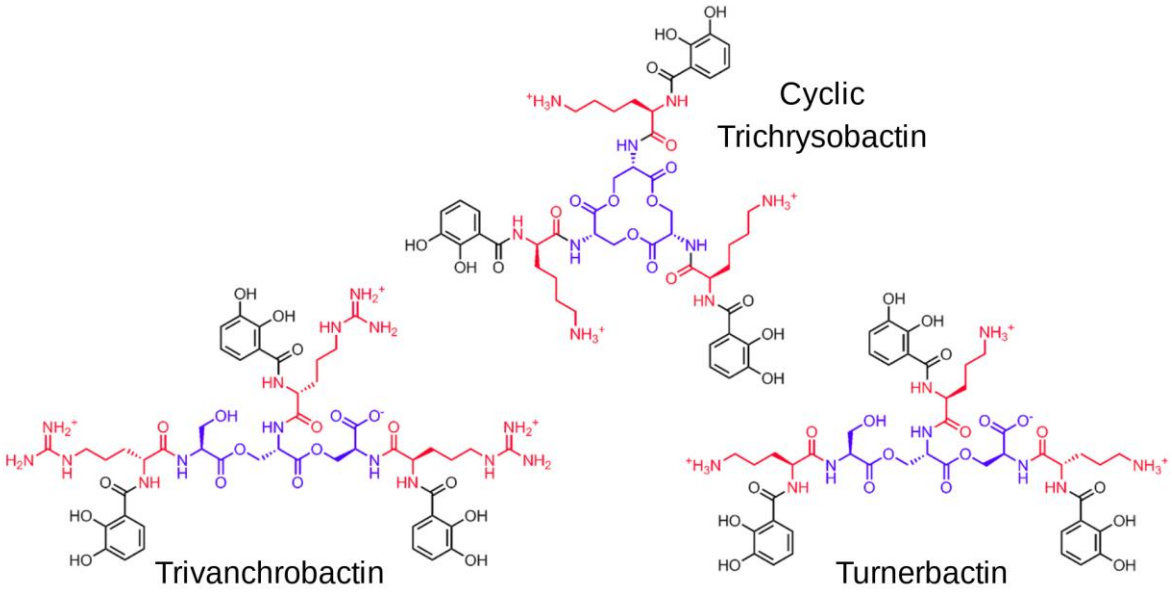


Figure 3.1. Previously reported DHB-CAA-Ser triscatechol siderophores.

The metal center chirality of several ferric triscatecholate complexes has been determined by circular dichroism (CD) spectroscopy and X-ray crystallography, allowing for the relationship between siderophore structure and metal center chirality to be explored (Table 3.1). The L-Ser core of enterobactin enforces the Δ Fe(III) complex, recently confirmed by the long-sought crystal structure of ferric enterobactin.[8-11] Ferric enantio-enterobactin, with a D-Ser macrolactone, has the opposite Λ configuration.[1,11] Interestingly, insertion of a glycine spacer reverses the stereochemistry around iron; synthetic Fe(III)–^{L}Ser-Gly-Cam has the Λ configuration, while Fe(III)–^{D}Ser-Gly-Cam has a Δ configuration.[1] Metal center chirality may also be enforced by a chiral spacer amino acid. Tren-Lys-Cam is a synthetic cyclic trichrysobactin mimic with an achiral polyamine core; Tren-^{L}Lys-Cam and Tren-^{D}Lys-Cam form Δ and Λ iron complexes, respectively.[2]

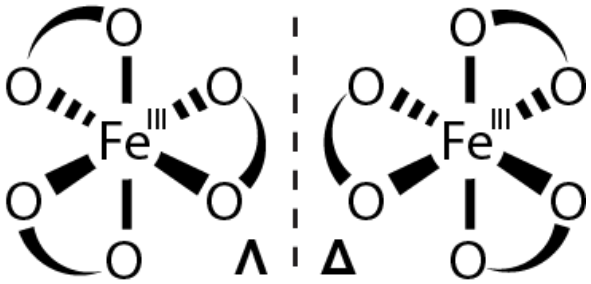


Figure 3.2. Enantiomers of Fe(III) coordinated by three bidentate ligands.

Trivanchrobactin and cyclic trichrysobactin, with stereocenters stemming from the L-Ser backbone and D- amino acid spacer, both form the Λ ferric complex in solution.[12] This configuration is determined by the L-Ser core (*cf.* ^{L}Ser-Gly-Cam) and/or the D- amino acid spacer (*cf.* Tren-^{D}Lys-Cam), each of which enforces the Λ stereoisomer when alone.[2,13] Is backbone chirality or spacer chirality the dominant factor in controlling the configuration of the metal center? The diastereomeric partners of trivanchrobactin and trichrysobactin, with

L-Ser and an L- amino acid spacer, would be affected by the opposing influences of the core (*cf.* ^{L}Ser-Gly-Cam, Λ^{13}) and the spacer (*cf.* Tren-^{L}Lys-Cam, Δ^{2}). The CD spectrum of such a complex could help parse the relative impact that triscatechol siderophore structure components have on metal center chirality; however, no pair of diastereomeric triscatechol siderophores has been reported to date.

Table 3.1. Fe(III) center chirality of triscatechol siderophores and synthetic mimics

Ferric complex	**Core**	**Spacer**	**Enantiomer**	**Reference**
Enterobactin	L-Ser	-	Δ	[11]
Enantioenterobactin	D-Ser	-	Λ	[2]
Linear enterobactin	L-Ser	-	Δ	[14]
Trivanchrobactin	L-Ser	D-Arg	Λ	[12]
Cyclic trichrysobactin	L-Ser	D-Lys	Λ	[12]
Bacillibactin	L-Thr	Gly	Λ	[13]
^{L}Ser–Gly–Cam	L-Ser	Gly	Λ	[13]
^{D}Ser–Gly–Cam	D-Ser	Gly	Δ	[1]
Tren–^{L}Lys–Cam	tren	L-Lys	Δ	[2]
Tren–^{D}Lys–Cam	tren	D-Lys	Λ	[2]

The discovery of novel siderophore structures has been greatly aided by scanning bacterial genomes for biosynthetic enzymes. Genome mining for siderophores generally takes one of three forms. In an "organism-first" approach, taxa of interest are scanned for any specialized metabolite biosynthetic gene clusters using workflows like PRISM or AntiSMASH. Alternatively, in an "enzyme-first" strategy, a desired class of siderophores may be found by searching for homologs of a biosynthetic enzyme using BLAST or HMMER, followed by manual or automatic structure prediction. A third "structure-first" strategy may be used when a siderophore is discovered and structurally characterized before

a genome of the producing strain is available. Based on existing knowledge of siderophore biosynthesis pathways, genes required for the biosynthesis of the orphan siderophore may be predicted. Once available, the genome of the producing strain can be searched for the relevant biosynthesis genes. The biosyntheses of DHB-CAA-Ser siderophores are sufficiently understood to make structural predictions from the genomic sequence. (For a full discussion of Ser-based triscatechol siderophore biosynthesis, see Chapter 1.) The variation of cationic spacer amino acids in trivanchrobactin, trichrysobactin, and turnerbactin come from the NRPSs (Figure 3.3). The first adenylation (A) domain selects which amino acid will be activated and incorporated as the spacer, and the epimerization (E) domain, if present, catalyzes formation of the D-amino acid spacer. Based on the structures and NRPS architectures of the previously reported siderophores, the NRPSs that could be responsible for the novel diastereomers may be predicted (Figure 3.3).

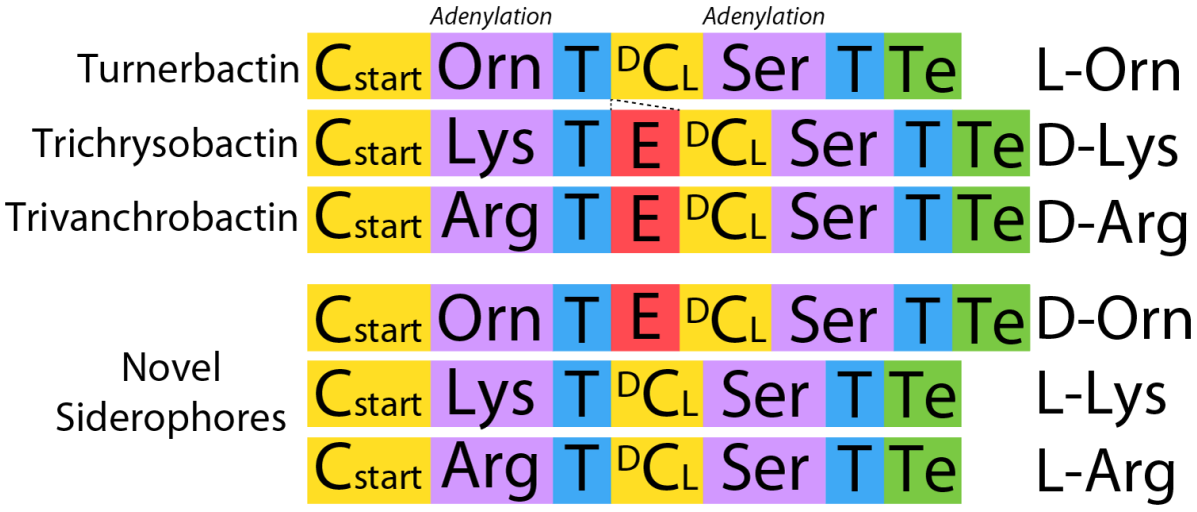


Figure 3.3. NRPS domain architectures of previously reported DHB-CAA-Ser triscatechol siderophores turnerbactin, (cyclic) trichrysobactin, and trivanchrobactin, as well as predicted architectures of novel siderophores with spacer amino acids not previously reported.

Both the organism-first approach and the enzyme-first approach cast broad nets and would give many false positives when looking for the novel triscatechol diastereomers. On the other hand, the highly-targeted structure-first strategy is only feasible when a producing strain is already known. The target siderophores were hypothetical, with no knowledge of which bacteria, if any, are capable of their production. Therefore, a hybrid strategy was developed, where a large set of bacterial genomes are successively filtered until the remaining genomes all putatively encode for the biosynthesis of a DHB-CAA-Ser siderophore. The final workflow, named the Catechol Siderophore Cluster Analysis (CatSCAn), was implemented in Python. Over 11,000 NCBI RefSeq Representative genomes were scanned with CatSCAn, resulting in only 99 genomes with the predicted gene cluster. The predictive models were tested with traditional siderophore isolation.

3.2. Methods

3.2.1. Development of a high-throughput genome mining workflow

The Catechol Siderophore Cluster Analysis (CatSCAn) workflow was primarily written in Python 3 using the Biopython package for protein sequence manipulation and file parsing. CatSCAN also relies on MUSCLE for multiple sequence alignments and HMMER3 for profile hidden Markov models (pHMMs). The full source code is available from GitHub (http://github.com/zreitz/catscan).

3.2.1.1. Profile hidden Markov models

Custom and previously-published profile hidden Markov models (pHMMs) were used to identify domains (Table 3.3). Amino acid sequences of domains involved in DHB

biosynthesis (EntA, EntB [isochorismatase domain], and EntC) and incorporation (EntE and ArCP) were collected from known catechol siderophore biosyntheses (Table 3.2). Amino acid sequences of the condensation domain superfamily members $^{L}C_{L}$, $^{D}C_{L}$, C_{start}, and E were obtained from Rausch *et al*. 2007. Collected sequences were aligned with MUSCLE, and pHMMs were constructed using the hmmbuild function from HMMER3. pHMMs for adenylation (PF00501), thiolation (PF00550), and thioesterase (PF00975) domains were obtained from Pfam and used without further modification. E-value cutoffs were determined manually (Table 3.3).

3.2.1.2. Identification of putative DHB-CAA-Ser siderophore biosynthesis clusters

RefSeq representative bacterial genomes were downloaded from NCBI (11064 as of April 1, 2020). Each genome was scanned for a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EntA) with the hmmscan function of HMMER3. If *entA* was present, genomes were then scanned with hmmscan for domains present in the biosyntheses of known DHB-CAA-Ser siderophores: DHB biosynthesis domains (EntA, EntB, and EntC), NRPS domains for DHB incorporation (ArCP, EntE, and C_{start}), and traditional NRPS domains (A, $^{D}C_{L}$, T, E, and Te). Genes less than 15 kbp apart were clustered by a greedy algorithm. Clusters were only kept if they contained one of the two NRPS domain architectures found across all DHB-CAA-Ser siderophores: C_{start}-A-T-$^{D}C_{L}$-A-T-Te or C_{start}-A-T-E-$^{D}C_{L}$-A-T-Te. Adenylation domain specificity was determined by comparison to a database of Stachelhaus codes adenylation domains with known specificity using NRPSPredictor2,[15] as implemented by SANDPUMA.[16] Siderophore structures were then predicted from any remaining gene clusters based on their NRPS domain architecture and A domain selectivity. No attempts were made to predict Ser oligomerization and cyclization,

as the governing factors are currently unknown. The [*Empedobacter*] *haloabium* genome GCF_008011715.1 was a representative genome on April 1, 2020, and is predicted to encode for turnerbactin biosynthesis; however, the genome has since been excluded from RefSeq ("untrustworthy as type"), and is therefore excluded from all genome counts.

3.2.2. General instrumental methods

Analytical reverse-phase high-pressure liquid chromatography (RP-HPLC) was performed on an analytical Waters YMC ODS-AQ C_{18} column (250 × 4.6-mm, 1 mL/min flow rate), coupled to a Waters 2998 photodiode array detector. Preparative RP-HPLC was performed on a semipreprative YMC ODS-AQ C_{18} column (250 × 20-mm, 7 mL/min flow rate). Ultraperformance liquid chromatography electrospray ionization mass spectrometry (UPLC-MS) was performed on a Waters Xevo G2-XS QTof coupled to a Waters Acquity H-Class UPLC system (Waters BEH C18 column).

3.2.3. Strains and culture conditions

Yersinia ruckeri YRB was obtained from D. Rozak (U.S. Army Medical Research Institute of Infectious Diseases) and maintained on LB agar plates. *Marinomonas* sp. TW1 was obtained from R. Paerl (North Carolina State University) and maintained on 2216 agar plates. *Chitinimonas koreensis* DSM 17726 was obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ) and maintained on LB agar plates. To induce siderophore production, *Y. ruckeri* and *C. koreensis* were cultured in casamino acid minimal medium containing 5 g/L Bacto low-iron casamino acids, 1.54 g/L K_2HPO_4, and 0.25 g/L $MgSO_4 \cdot 7H_2O$) *Marinomonas* sp. TW1 was cultured in M6 medium containing 15.5 g/L NaCl, 0.75 g/L KCl, 0.2 g/L $MgSO_4 \bullet 7H_2O$, 0.1 g/L $CaCl_2 \bullet 2H_2O$, 1 g/L NH_4Cl, 5 g/L sodium succinate, and

3 g/L K_2HPO_4, adjusted to pH 7.0. Strains were grown in 1 L of medium in a 2 L acid-washed Erlenmeyer flask on an orbital shaker (150 rpm) at ambient temperature. Growth was monitored by measuring the optical density at 600 nm, and the culture was tested for siderophore production by the ferric chrome azurol S (CAS) colorimetric assay.[17] For precursor directed biosynthesis, *Marinomonas* sp. TW1 was grown in 30 mL of medium in a 100 mL flask; potential substrates were added prior to sterilization.

3.2.4. Siderophore isolation and purification

Upon reaching late logarithmic or early stationary phase (3-6 days), cells were pelleted by centrifugation (6000 rpm, 30 min, 4 °C). Decanted supernatant was shaken with 10% v/v water-washed Amberlite XAD-2 (*Marinomonas* and *Y. ruckeri*) or XAD-4 (*C. koreensis*) resin for 3 hours. The resin was isolated by filtration and washed with 5 column volumes nanopure water and 5 column volumes 10% v/v methanol in nanopure water. Organic compounds were eluted with 3 column volumes 90% v/v methanol in nanopure water and concentrated *in vacuo*. Catechol species were identified in the resultant supernatant extract by analytical RP-HPLC, employing a linear gradient of 5-100% acetonitrile in ultrapure water (both with 0.05% trifluoroacetic acid) over 45 min. Siderophores were purified by preparative RP-HPLC, employing a linear gradient of 5-60% methanol in ultrapure water (both with 0.05% trifluoroacetic acid) over 55 min. Pooled fractions were concentrated *in vacuo* and lyophilized. Extracts and compounds were further analyzed by UPLC/ESIMS, employing a linear gradient of 0-60% acetonitrile in ddH_2O (both with 0.1% formic acid) over 10 min.

3.2.5. Amino acid analysis by Marfey's method

Purified siderophore (1 mg) was hydrolyzed in 6M HCl for 18 hours at 110°C in a sealed glass ampoule. HCl was removed by three evaporation-rehydration cycles. The sample was dissolved in 250 μL H_2O and derivatized with 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (FDAA; Marfey's reagent).[18] To 100 μL of hydrolysate solution was added 200 μL of a 1% w:w solution of FDAA in acetone and 40 μL of 1 M $NaHCO_3$. The solution was heated to 40 °C for 1 hr, then quenched with 20 μL 2 M HCl. Derivatized samples were analyzed via analytical RP-HPLC using a linear gradient of 10-40% acetonitrile in ultrapure water (both +0.05 trifluoroacetic acid) over 45 min, monitoring at 340 nm. Samples were compared to amino acid standards prepared identically.

3.3. Results and Interpretation

3.3.1. Development of the Catechol Siderophore Cluster Analysis (CatSCAn)

A comparative analysis of gene clusters responsible for the biosynthesis of previously reported DHB-CAA-Ser siderophores (Chapter 1) shows that each cluster encodes homologs of EntABCE, responsible for DHB biosynthesis and incorporation, and a two-module NRPS with the domain architectures C_{start}-A-T-$^{D}C_{L}$-A-T-Te or C_{start}-A-T-**E**-$^{D}C_{L}$-A-T-Te (Figure 3.3).[19] All biosynthetic enzymes and NRPS domains were identified by comparison to profile hidden Markov models (pHMMs). Existing protein families in the Pfam database were not specific enough to select for genes involved in siderophore biosynthesis. For example, the short chain dehydrogenase family (PF00106) includes not only functional homologs of EntA, but also 3-ketoacyl ACP reductases involved in fatty acid biosynthesis. To eliminate these false positives, custom pHMMs were built using the amino

acid sequences of EntABCE homologs and aryl carrier proteins (ArCP) collected from previously reported catechol siderophore biosynthetic gene clusters (Table 3.2).

Significance cutoffs (E-values) for each of EntABCE pHMMs were determined manually from a random set of 1000 complete bacterial genomes from NCBI RefSeq.[20] Of 428 genomes with matches to all four genes (E-value <1e-5), 181 contained a colocalized cluster of all four genes. The highest (least significant) E-values for each pHMM in these 181 genomes were used as the cutoff, with a conservative 0.85 multiplier to prevent false negatives. The resulting E-values (Table 3.3) successfully eliminated most enzymes annotated with undesired functions, particularly for EntA (Figure 3.4). Among 223 genomes in the test set with a significant match to the EntA pHMM, 222 (99.6%) also contained significant matches to each of EntBCE in the genome. Thus, the presence of an *entA* homolog alone is a reliable indicator that a genome encodes for the synthesis and activation of DHB.

Table 3.2. Catechol siderophores included in custom CatSCAn profile hidden Markov models

Siderophore	**Bacterium**	**Reference**
Amonabactin	*Aeromonas salmonicida* A449	[21]
Amphi-enterobactin	*Vibrio campbellii* BAA-1116	[22]
Bacillibactin	*Bacillus subtilis* 168	[23]
Chrysobactin	*Dickeya dadantii* 3937	[24]
Enterobactin	*Escherichia coli* K12	[25]
Fuscachelin	*Thermobifida fusca* YX	[26]
Griseobactin	*Streptomyces* sp. ATCC 700974	[27]
Heterobactin	*Rhodococcus erythropolis* PR4	[28]
Hyalachelin	*Hyalangium minutum*	[29]
Mirubactin	*Actinosynnema mirum*	[30]
Myxochelin	*Stigmatella aurantiaca* Sg a15	[31]
Paenibactin	*Paenibacillus elgii* B69	[32]
Protochelin	*Azotobacter vinelandii* CA	[33]
Rhodochelin	*Rhodococcus jostii* RHA1	[34]
Trivanchrobactin	*Vibrio campbellii* DS40M4	[35]
Turnerbactin	*Teredinibacter turnerae* T7901	[36]
Vanchrobactin	*Vibrio anguillarum* RV22	[37]
Acinetobactin	*Acinetobacter baumannii* ATCC 19606	[38]
Agrobactin	*Agrobacterium tumefaciens* B6	[39]
Anguibactin	*Vibrio anguillarum* 775(pJM1)	[40]
Brucebactin	*Brucella abortus*	[41]
Fimsbactin	*Acinetobacter* sp. ADP1	[42]
Photobactin	*Photorhabdus temperata* subsp. khanii NC19	[43]
Serratiochelins	*Serratia plymuthica* V4	[44]
Vibriobactin	*Vibrio cholerae* O395	[45]
Vulnibactin	*Vibrio vulnificus* CMCP6	[46]

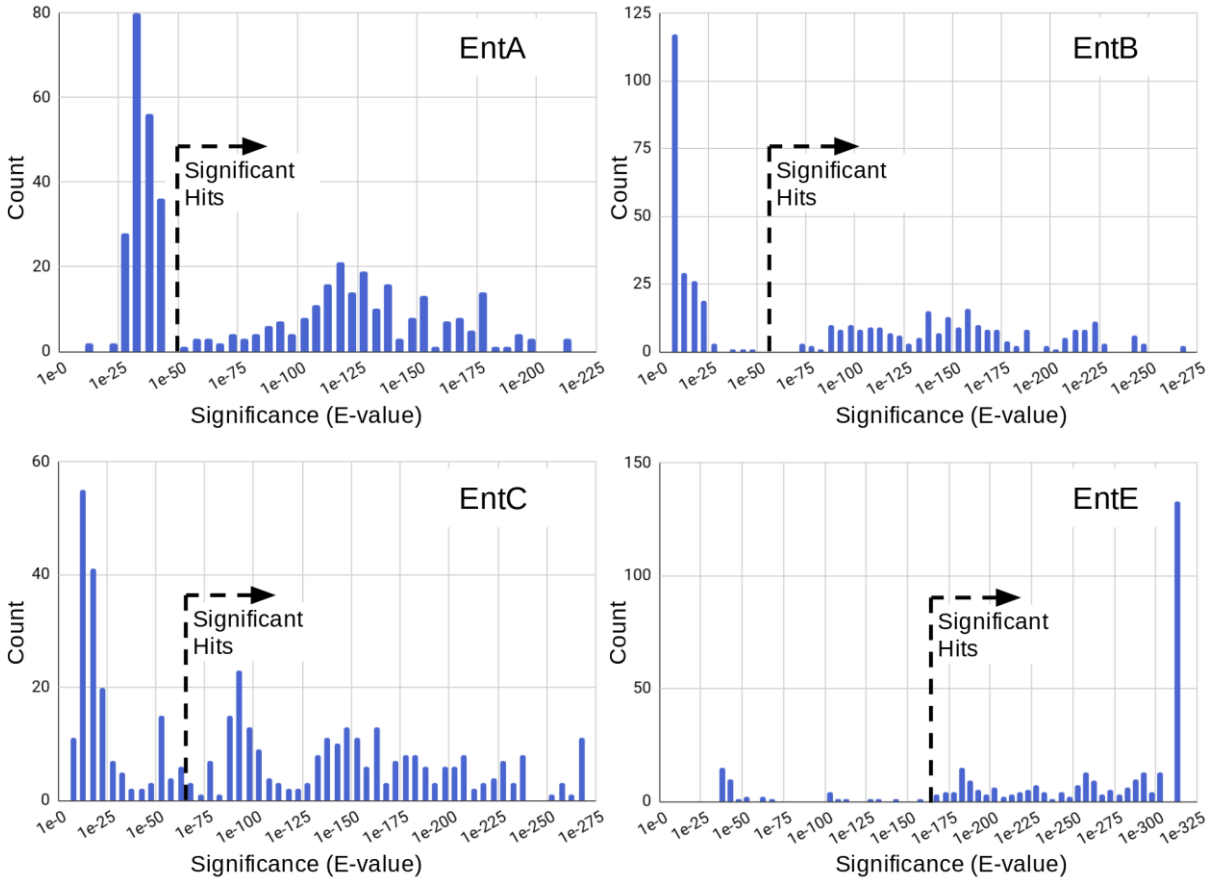


Figure 3.4. Significance cutoffs and histogram plots for EntABCE pHMMs. For each pHMM, hits were sorted into buckets by E-value order of magnitude. Plots are composed of the single most significant hit for each of 428 genomes with matches to all four genes (E-value <1e-5). Conservative significance cutoffs were determined as described in the text.

Table 3.3. Profile hidden Markov models incorporated in CatSCAn.			
Name	**Description**	**E-value cutoff**	**Source**
EntA	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	1e-65	this work
EntB	Isochorismatase	1e-60	this work
EntC	Isochorismate synthase	1e-90	this work
EntE	2,3-dihydroxybenzoate-AMP ligase	1e-165	this work
ArCP	Aryl carrier protein	1e-10	this work
A	Adenylation domain	1e-50	PF00501
T	Thiolation domain	1e-5	PF00550
Te	Thioesterase domain	1e-20	PF00975
LCL	$^{L}C_{L}$-type condensation domain	1e-50	[47]
DCL	$^{D}C_{L}$-type condensation domain	1e-50	[47]
Cstart	Aryl-selective condensation domain	1e-50	[47]
Cy	Heterocyclization domain	1e-50	[47]
E	Epimerization domain	1e-50	[47]

The genome mining workflow was automated in Python, resulting in a workflow named the Catechol Siderophore Cluster Analysis (CatSCAn, Figure 3.5). RefSeq Representative Genomes are downloaded from NCBI.[20] To reduce computing time, genomes are first filtered for the presence of a significant match to the EntA pHMM. The remaining strains, all putative 2,3-DHB producers, are then scanned for the full suite of biosynthetic enzymes (Table 3.3). A second round of filtering requires a single NRPS gene with one of the two domain architectures common to all three reported DHB-CAA-Ser siderophores: C_{start}-A-T-$^{D}C_{L}$-A-T-Te or C_{start}-A-T-**E**-$^{D}C_{L}$-A-T-Te (Chapter 1).[19] The absence or presence

of an E domain determines the stereochemistry of the first amino acid. The selectivity of the two adenylation domains are then predicted by NRPSpredictor2,[15] comparing the extracted 10AA Stachelhaus codes to a database of adenylation domains with known specificity.[48] A minimum percent identity cutoff was determined using 10AA codes from the catechol siderophore biosyntheses in Table 3.2. Within this dataset, a 90% or 100% match between codes was required to ensure two A domains load the same amino acid. For example, the first A domains from griseobactin and turnerbactin are an 80% match (D**A**WD**V**GLVDK *versus* D**S**WD**G**GLVDK) and load Arg and Orn, respectively. Supplementing the A domain database with catechol siderophores reported previously (Table 3.2) and herein improved amino acid prediction (*vide infra*).

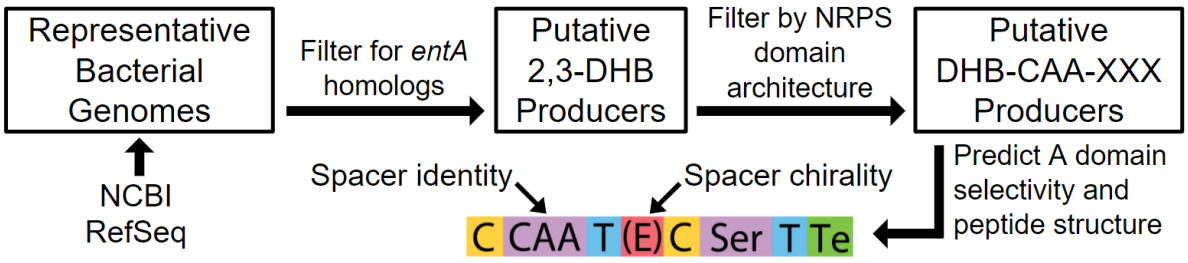


Figure 3.5. An overview of the CatSCAn workflow for the high-throughput discovery of triscatechol siderophore biosynthetic gene clusters.

3.3.2. Prevalence of triscatechol siderophore biosynthesis clusters

A total of 11064 bacterial genomes from the NCBI RefSeq Representative Genome database (April 1, 2020) were scanned for DHB-CAA-Ser biosynthetic gene clusters. Homologs of *entA* are present in 10.7% of genomes scanned, mostly among Proteobacteria, Firmicutes, and Actinobacteria (Figure 3.6A). Of the 1179 genomes with *entA*, only 99 contain a NRPS with the required NRPS architecture and a second module selective for Ser, thereby putatively capable of producing a DHB-CAA-Ser siderophore (Table 3.5). Nearly

all are proteobacteria, the sole exception being the Actinobacterium *Arthrobacter glacialis* (Figure 3.6B).

The second adenylation domain of each NRPS was confidently predicted as Ser using the Stachelhaus codes; each was a perfect match to known adenylation domains. However, the default NRPSPredictor2 training set could not predict the spacer amino acid for any of the 100 NRPS (Table 3.5). For example, turnerbactin of *Teredinibacter turnerae* (L-Orn) was predicted as Glu with eight of ten matching amino acids in the binding pocket, and the majority of adenylation domains had multiple closest AA codes with only 60 or 70% matches. Adding Stachelhaus codes from previously reported catechol siderophores (Table 3.2) greatly improved the predictions; 57 of the 100 adenylation domains had codes that were 90-100% matches to the updated training set (Table 3.5).

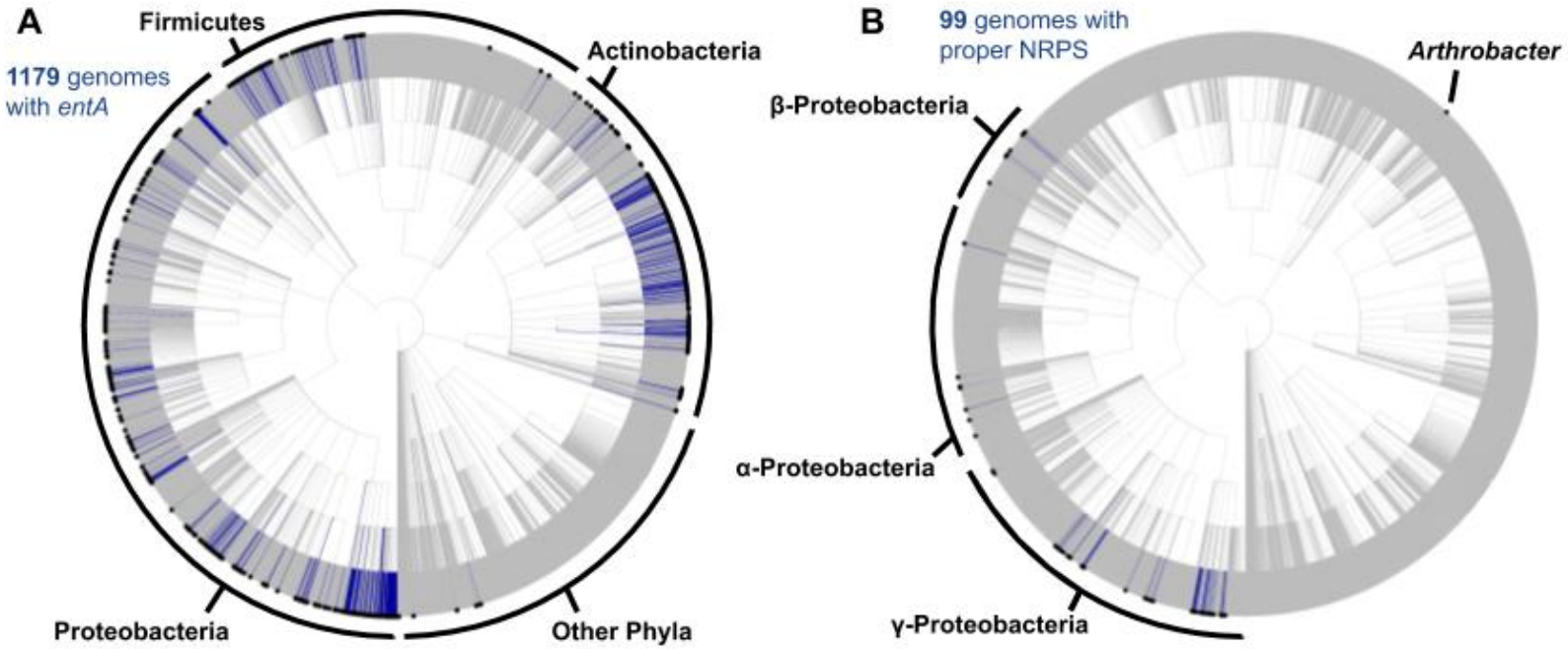


Figure 3.6. Distribution of catechol siderophore biosynthesis genes across bacterial strains. The cladogram of 11,064 representative bacterial genomes from NCBI RefSeq is based on classification in the NCBI taxonomy database (not direct phylogenetic information) and was created with the NCBI CommonTree tool. (A) Distribution of homologs of *entA*, mainly distributed across the phyla Proteobacteria, Firmicutes, and Actinobacteria. (B) Distribution of NRPSs with the architectures C_{start}-A-T-$^{D}C_{L}$-A-T-Te or C_{start}-A-T-E-$^{D}C_{L}$-A-T-Te, where the second A domain is predicted to be Ser selective. Except the Actinobacterium *Arthrobacter glacialis*, all positive strains are alpha-, beta-, or gammaproteobacteria.

3.3.3. Isolation and characterization of novel siderophores predicted by CatSCAn

3.3.3.1. Ruckerbactin from *Yersinia ruckeri* YRB

In addition to the previously reported siderophores vanchrobactin, chrysobactin, and turnerbactin (D-Arg, D-Lys, and L-Orn spacers, respectively, Figure 3.1), CatSCAn predicted the biosynthesis of novel siderophores with L-Arg and L-Lys spacer amino acids. The L-Arg spacer was only predicted in a single species, *Yersinia ruckeri*. The gene cluster identified by CatSCAn was previously discovered and was determined to encode for the biosynthesis and transport of the catechol siderophore *ruckerbactin*; however, no structural characterization was performed.[49] A targeted scan of RefSeq *Y. ruckeri* genomes revealed that the cluster is strictly conserved within the species. To test the predictive power of CatSCAn and determine the identity of ruckerbactin, *Y. ruckeri* YRB was obtained for further analysis.

Siderophores of *Yersinia ruckeri* YRB were extracted from the supernatant of a low-iron culture with Amberlite XAD-4 resin. RP-HPLC and RP-UPLC analysis of the methanolic eluent revealed three compounds with absorbance bands in the UV spectrum at approximately 250 and 315 nm, consistent with the presence of DHB (Figure 3.7). UPLC-ESIMS found these three compounds had identical masses to vanchrobactin (*m/z* 398 $[M+H]^+$), divanchrobactin (*m/z* 389 $[M+2H]^{2+}$), and trivanchrobactin (*m/z* 386 $[M+3H]^{3+}$);[50] a mass of 1200, consistent with Ti(IV)-bound trivanchrobactin, was also observed (Figure 3.8). ESIMS/MS revealed fragmentation consistent with losses of DHB, Arg, and Ser (Figures 3.9-3.11). These analyses, together with a lack of an E domain, are consistent with the ruckerbactin family of siderophores from *Y. ruckeri* YRB as the L-Arg diastereomers of vanchrobactin, divanchrobactin, and trivanchrobactin (Figure 3.12).

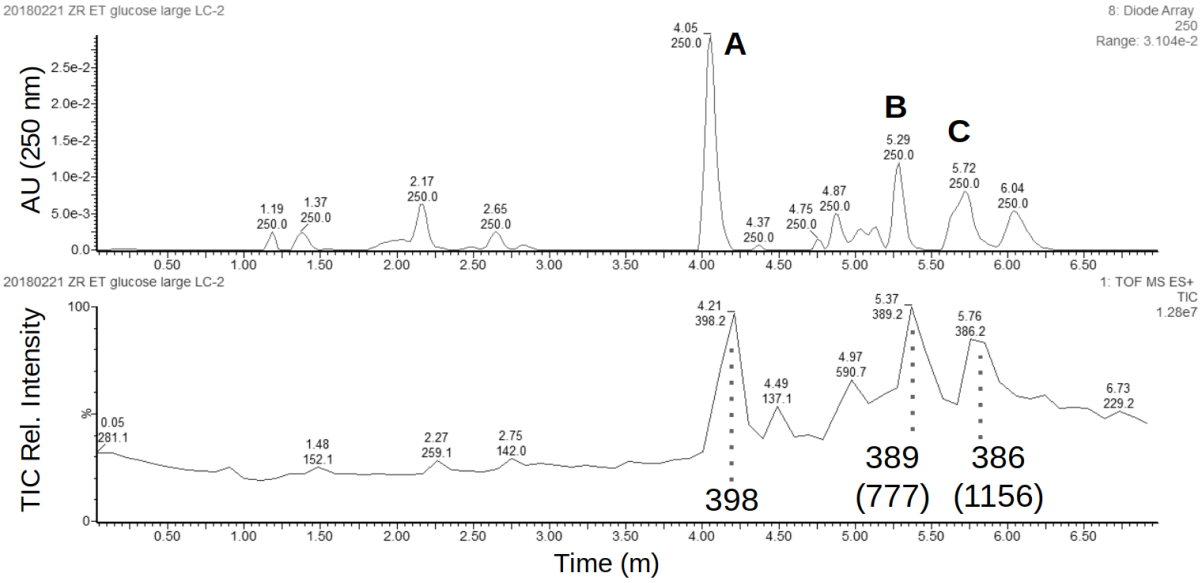


Figure 3.7. UPLC chromatogram of *Y. ruckeri* YRB low iron supernatant XAD extract. Peaks A, B, and C correspond to the ruckerbactin monomer, dimer, and trimer, respectively. Top: UPLC dual-wavelength detector output monitoring at 250 nm. Bottom: ESIMS total ion chromatogram. Multiply-charged ions are given with their corresponding singly-charged equivalents in parentheses.

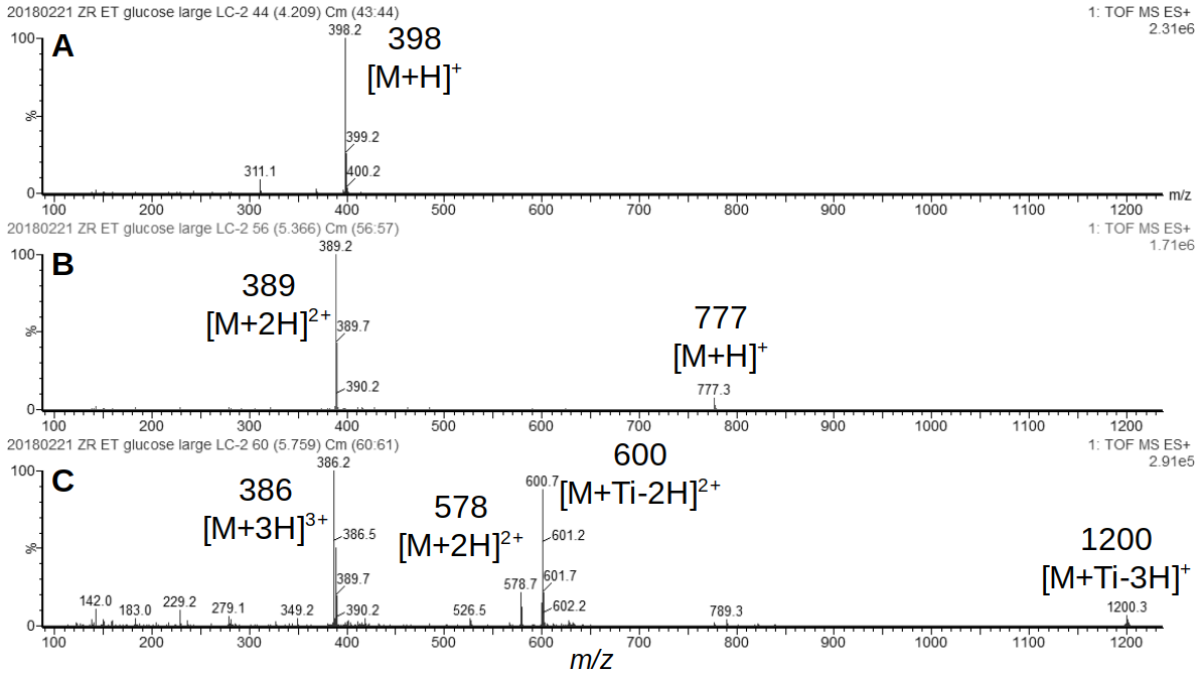


Figure 3.8. ESIMS of the ruckerbactin monomer (A), dimer (B), and trimer (C).

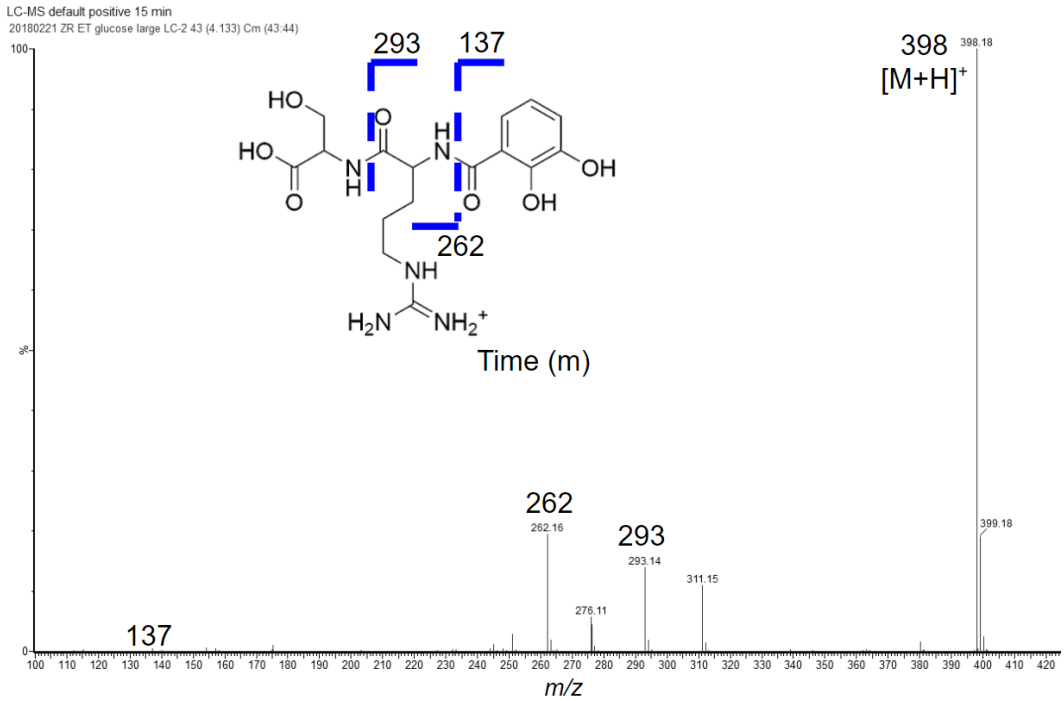


Figure 3.9. ESIMS/MS spectrum of the ruckerbactin monomer with proposed structure and fragmentation.

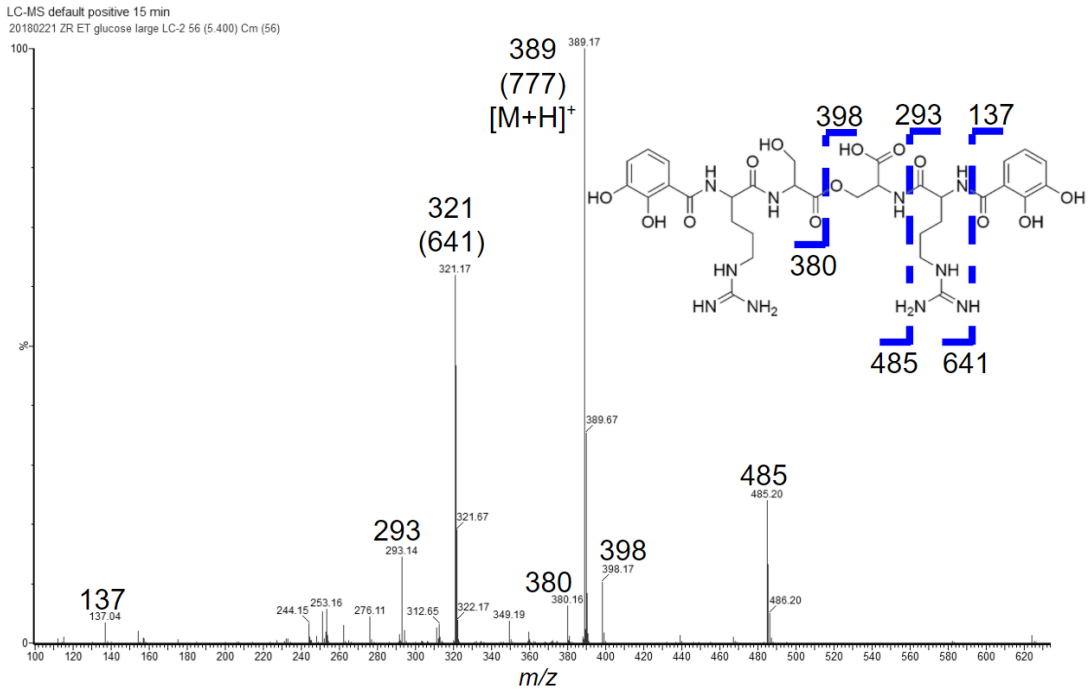


Figure 3.10. ESIMS/MS spectrum of the ruckerbactin dimer with proposed structure and fragmentation. Multiply-charged ions are given with their corresponding singly-charged equivalents in parentheses.

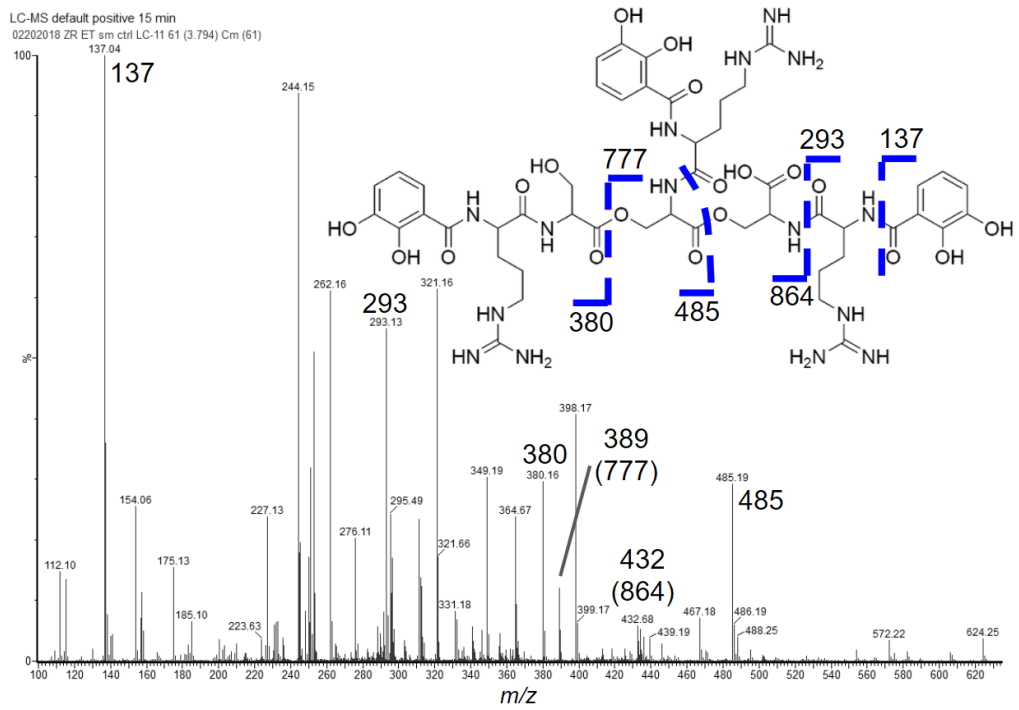


Figure 3.11. ESIMS/MS spectrum of ruckerbactin with proposed structure and fragmentation. Multiply-charged ions are given with their corresponding singly-charged equivalents in parentheses.

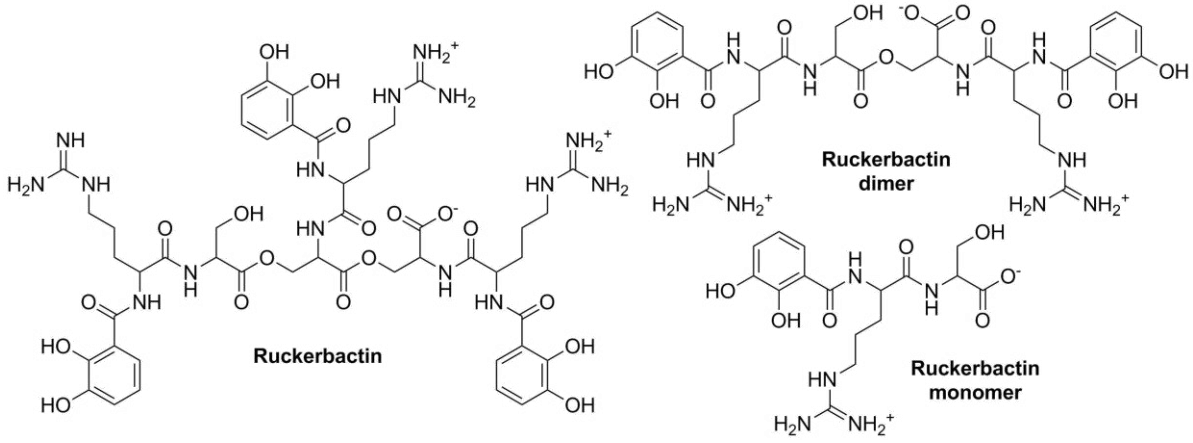


Figure 3.12. Proposed structures of the ruckerbactin family of siderophores.

3.3.3.2. Turnerbactin from *Acinetobacter nosocomialis* 8399 and *Azospirillum brasilense* Az39.

NRPSs from a diverse group of alpha-, beta-, and gammaproteobacteria have the identical Stachelhaus code DSLDLGLVDK (Table 3.5). CatSCAn was unable to definitively predict an amino acid, with closest matches Orn and Arg tied at 80% identity. A literature review revealed that two representatives of this group were previously reported to produce catechol siderophores. *Acinetobacter nosocomialis* 8399 (formerly *A. baumannii*) and *Azospirillum brasilense* az39 were each found to produce catechol siderophores that were never fully characterized.[51,52] The *A. brasilense* az39 siderophore, named spirillobactin, was determined to contain equimolar DHB, Orn, and Ser.[52] This partial characterization and the lack of an E domain led to the prediction that spirillobactin is identical to turnerbactin, $(DHB\text{-}^{L}Orn\text{-}^{L}Ser)_3$.

3.3.3.3. Trivanchrobactin from *Chitinimonas koreensis* DSM 17726

Of nine betaproteobacteria predicted to produce a DHB-CAA-Ser siderophore, only one, *Chitinimonas koreensis* DSM 17726, had a definitive spacer amino acid prediction: D-Arg (Table 3.5). The Stachelhaus code was 90% identical to the first A domain of FscG of fuscachelins biosynthesis. The two A domains have 45% identity across their length. The highest sequence similarity among catechol siderophores was with the Orn-selective A domain of *A. brasilense* az39 at 61% identity; however, the Stachelhaus codes only shared seven of ten residues. The XAD extract of a low-iron culture of *C. koreensis* DSM 17726 contained trace amounts of three catecholic compounds (Figure 3.13) with mass to charge ratios of 398, 777, and 386, consistent with DHB-Arg-Ser, $(DHB\text{-}Arg\text{-}Ser)_2$, and triply-charged $(DHB\text{-}Arg\text{-}Ser)_3$, respectively (Figure 3.14 and Figure 3.15). The most abundant

ions in the trimer ESIMS are consistent with Si-bound $(DHB\text{-}Arg\text{-}Ser)_3$ (*m/z* 1180, $[M+Si\text{-}3H]^+$, and 590, $[M+Si\text{-}2H]^{2+}$). Under our growth conditions, the compounds were not produced in sufficient quantities for isolation and full characterization. Nevertheless, the presence of an epimerization domain and the UPLC-ESIMS data lead us to conclude that the siderophores of *Chitinimonas koreensis* DSM 17726 likely have a D-Arg spacer amino acid and are identical to vanchrobactin, divanchrobactin, and trivanchrobactin (Figure 3.16).

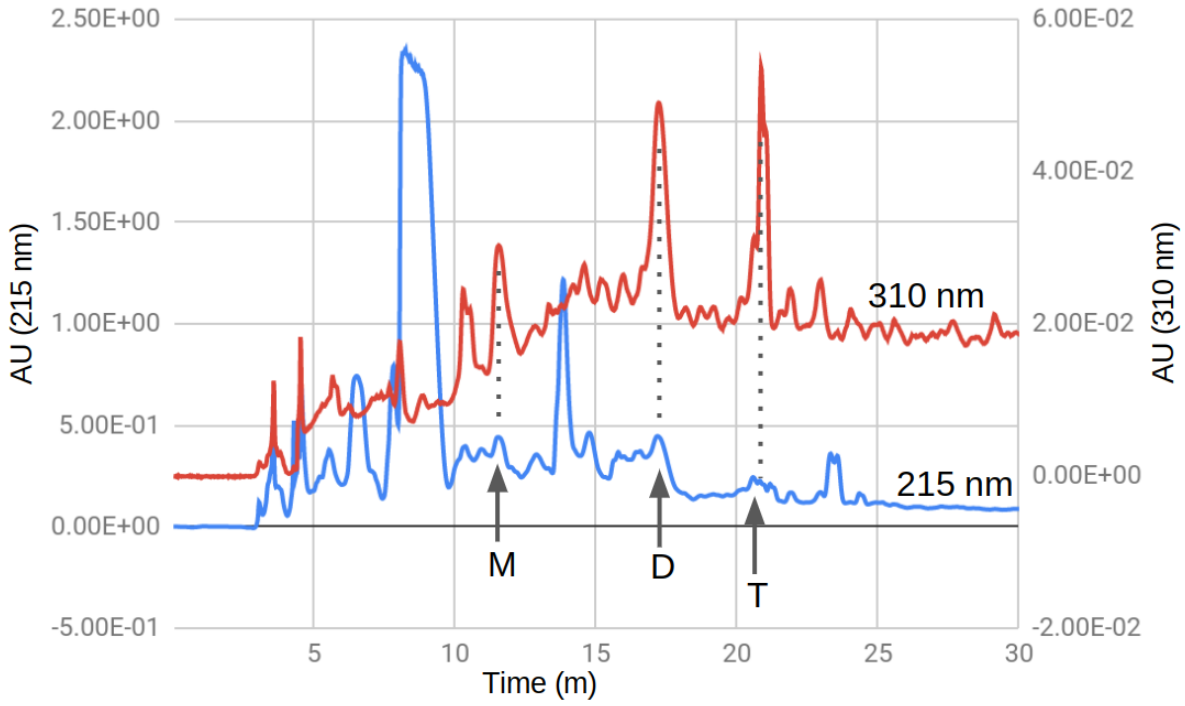


Figure 3.13. RP-HPLC–UV/Vis chromatogram of the Chitinomonas koreensis DSM 17726 low iron supernatant XAD extract. Column eluent was monitored at 215 nm (blue) and 310 nm (red). Peaks putatively corresponding to vanchrobactin (M), divanchrobactin (D), and trivanchobactin (T) are labeled.

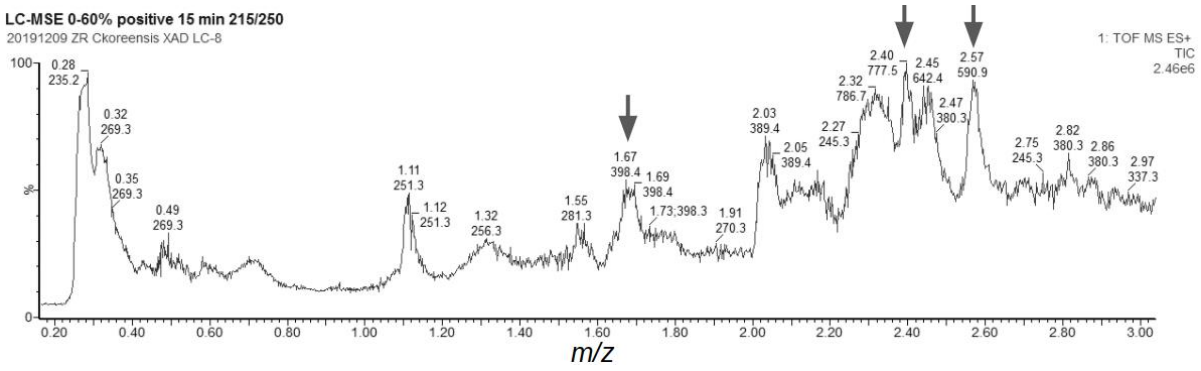


Figure 3.14. UPLC–ESIMS TIC chromatogram of the *Chitinomonas koreensis* DSM 17726 low iron supernatant XAD extract. Peaks putatively corresponding to vanchrobactin, divanchrobactin, and trivanchobactin are labeled.

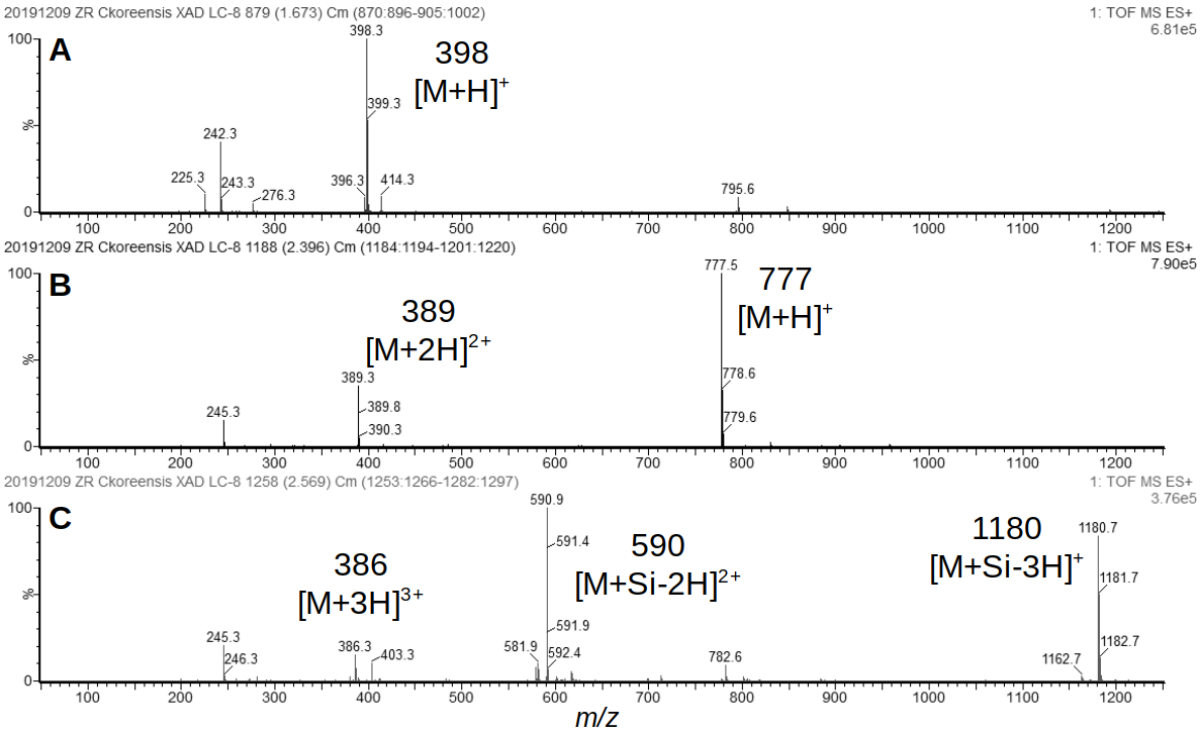


Figure 3.15. ESI mass spectrometry of the *Chitinomonas koreensis* DSM 17726 siderophores, putatively assigned as vanchrobactin (A), divanchrobactin (B), and trivanchrobactin (C).

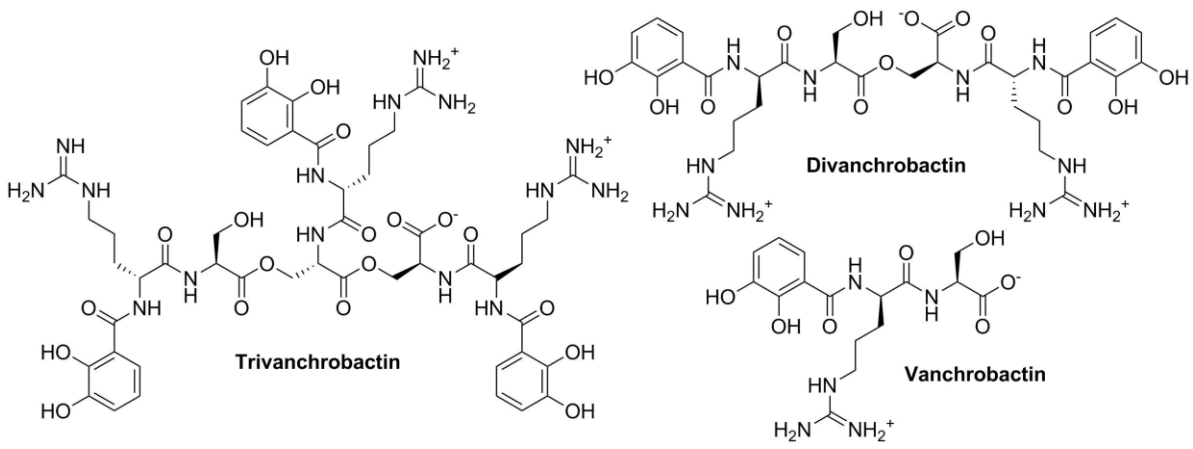


Figure 3.16. Proposed structures of vanchrobactin family of siderophores produced by *Chitinomonas koreensis* DSM 17726.

3.3.3.4. Mixed-spacer siderophores from *Marinomonas* sp. TW1

A phylogenetic tree of the siderophore NRPSs was reconstructed (see Chapter 4). A large clade with no isolated siderophore representative was primarily composed of *Marinomonas* strains. The NRPSs from two strains (*M. balearica* CECT 7378 and *M. mediterranea* MMB-1) were a 90% match to the Lys-selective A domain of *S. marcescens* (Table 3.5). Five other *Marinomonas* strains did not have a definitive A domain prediction; the majority had an 80% match to Glu as their closest known Stachelhaus code and were at least 90% identical to each other (Table 3.5). *Marinomonas* sp. TW1 was obtained for further analysis. HPLC analysis of the methanolic supernatant extract revealed five major compounds with characteristic catechol UV absorbance bands (Figure 3.17). UPLC-ESIMS molecular ions were consistent with monomeric and dimeric DHB-Orn-Ser (*m/z* 356 and 693), monomeric and dimeric DHB-Arg-Ser (*m/z* 398 and 777), and a mixed Orn/Arg dimer (DHB-Orn-Ser)(DHB-Arg-Ser) (*m/z* 735) (Figures 3.18-3.20). No trimeric $(DHB\text{-}Orn\text{-}Ser)_3$ or $(DHB\text{-}Arg\text{-}Ser)_3$ masses were observed.

All five compounds were isolated, and the constituent amino acids were established by Marfey's analysis.[18] Arg was solely present in the D- configuration, consistent with the presence of an E domain in the NRPS (Figure 3.22, Figure 3.24, Figure 3.25). The Orn monomer, Orn dimer, and mixed dimer each contained mostly D-Orn, but a significant portion of L-Orn was present (Figure 3.21, Figure 3.23, Figure 3.25). Partial racemization during acid hydrolysis may have occurred; DCl/D_2O hydrolysis will determine if this is the case.[53] However, during hydrolysis of the mixed dimer, Orn and Arg were subjected to the same conditions, and only Orn was found as a mixture of enantiomers (Figure 3.25). To test the effect of amino acid concentration on siderophore production, the growth medium was supplemented with either L-Orn or L-Arg (20 mM final concentration). Addition of L-Orn caused a slight increase in DHB-Orn-Ser and $(DHB\text{-}Orn\text{-}Ser)_2$ relative to the other catechol compounds, while addition of L-Arg nearly abolished the production of all three Orn-containing compounds (Figure 3.26). Furthermore, a new major species was present in the supernatant extract with an ESIMS spectrum consistent with apo and Si-bound linear $(DHB\text{-}Arg\text{-}Ser)_3$ (Figure 3.26 and Figure 3.27).

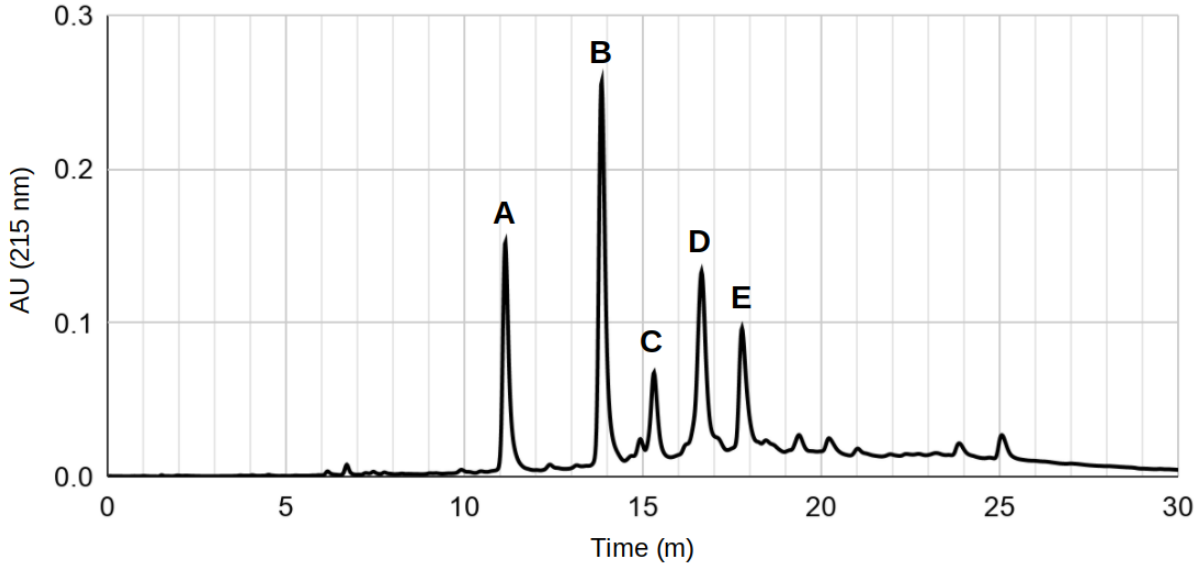


Figure 3.17. RP-HPLC–UV/Vis chromatogram of *Marinomonas* sp. TW1 low iron supernatant XAD extract. Labeled peaks correspond to the isolated siderophores DHB-Orn-Ser [A], DHB-Arg-Ser [B], $(DHB\text{-}Orn\text{-}Ser)_2$ [C], (DHB-Orn-Ser)(DHB-Arg-Ser) [4], and $(DHB\text{-}Arg\text{-}Ser)_2$ [E].

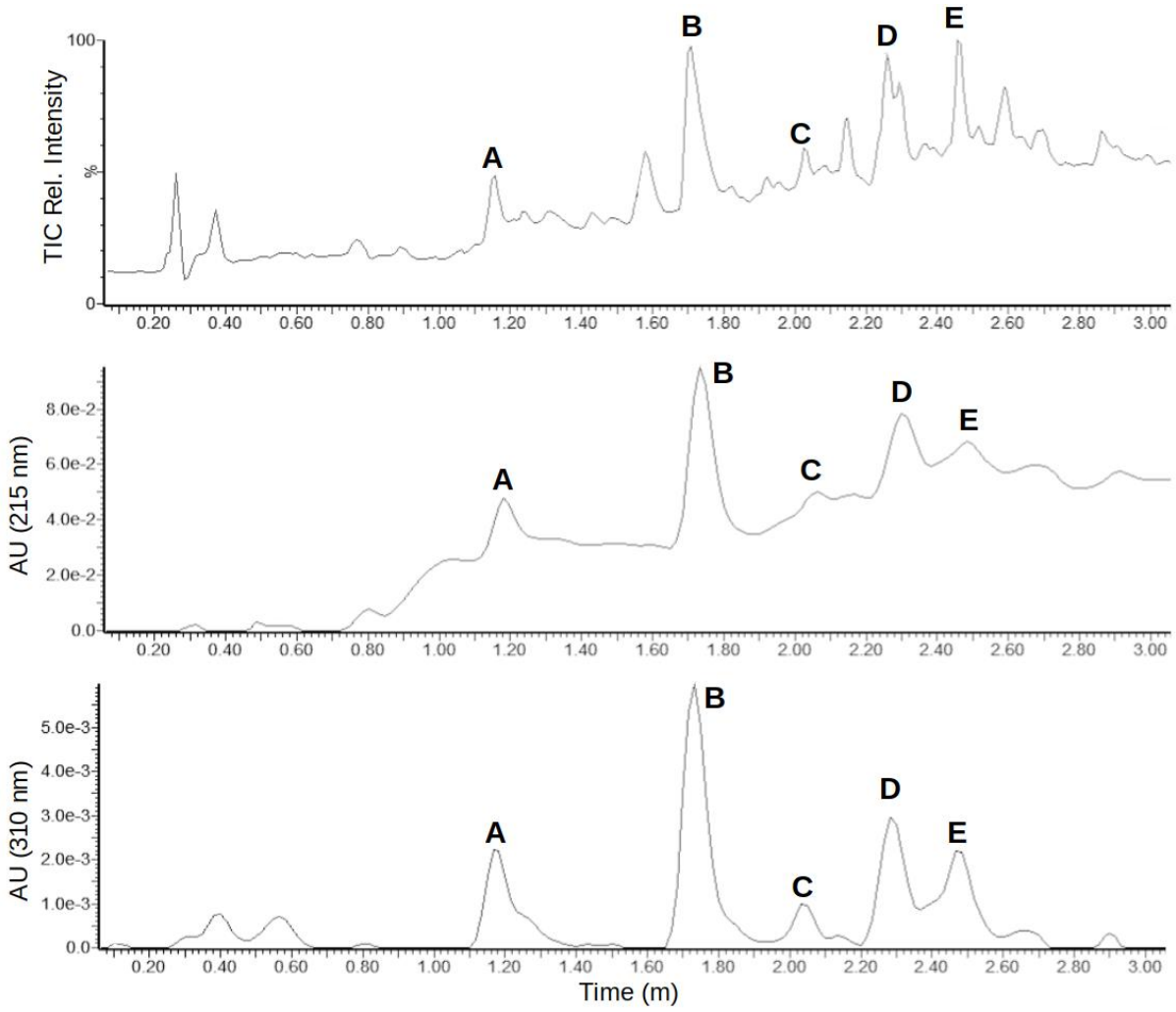


Figure 3.18. UPLC–ESIMS spectra of *Marinomonas* sp. TW1 low iron supernatant extract. Top: ESIMS total ion chromatogram. Middle: UPLC dual wavelength detector output monitoring at 215 nm. Bottom: UPLC dual wavelength detector output monitoring at 310 nm. Labeled peaks correspond to the isolated siderophores DHB-Orn-Ser [A], DHB-Arg-Ser [B], $(DHB\text{-}Orn\text{-}Ser)_2$ [C], (DHB-Orn-Ser)(DHB-Arg-Ser) [4], and $(DHB\text{-}Arg\text{-}Ser)_2$ [E].

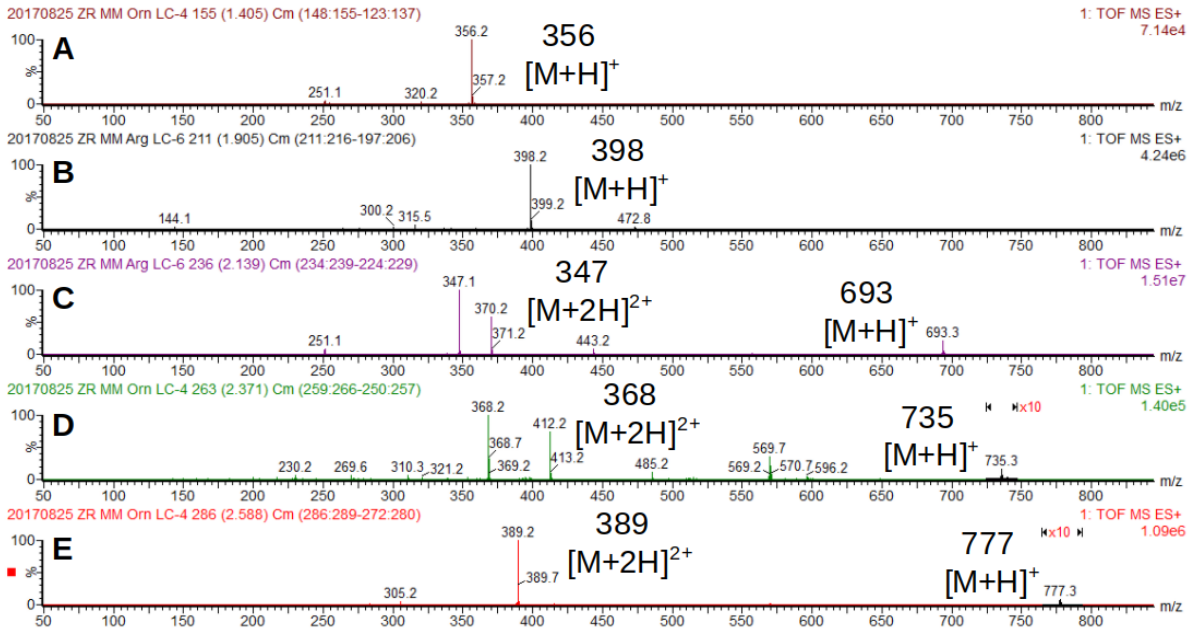


Figure 3.19. ESI mass spectrometry of the *Marinomonas* sp. TW1 siderophores DHB-Orn-Ser [A], DHB-Arg-Ser [B], $(DHB\text{-}Orn\text{-}Ser)_2$ [C], (DHB-Orn-Ser)(DHB-Arg-Ser) [4], and $(DHB\text{-}Arg\text{-}Ser)_2$ [E].

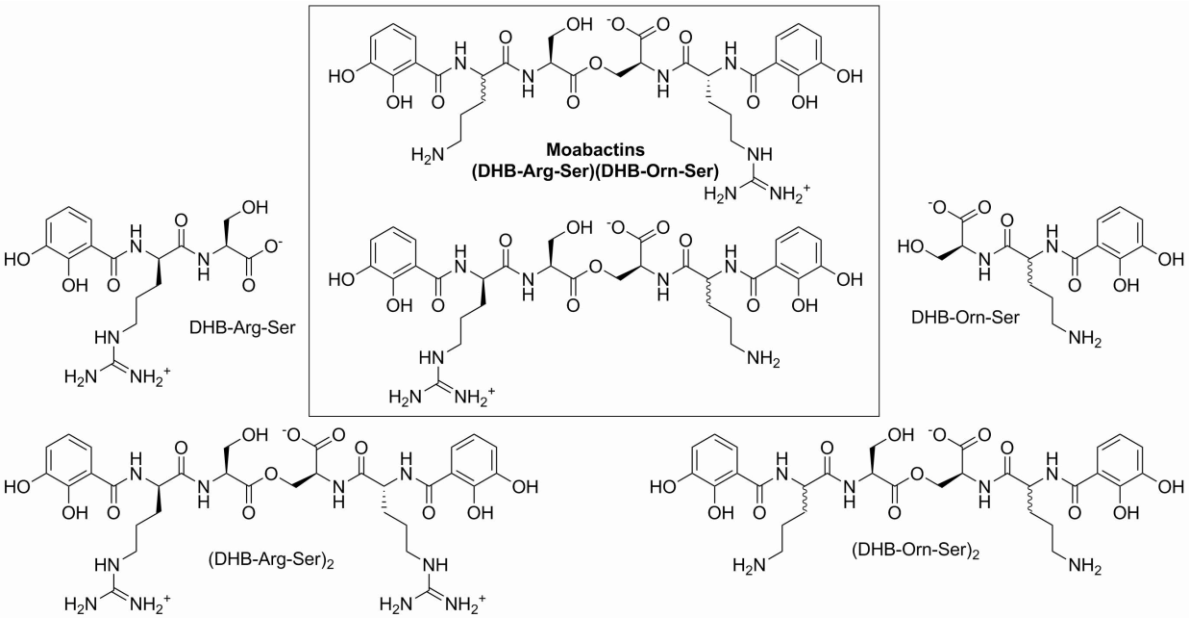


Figure 3.20. Proposed structures of siderophores from *Marinomonas* sp. TW1.

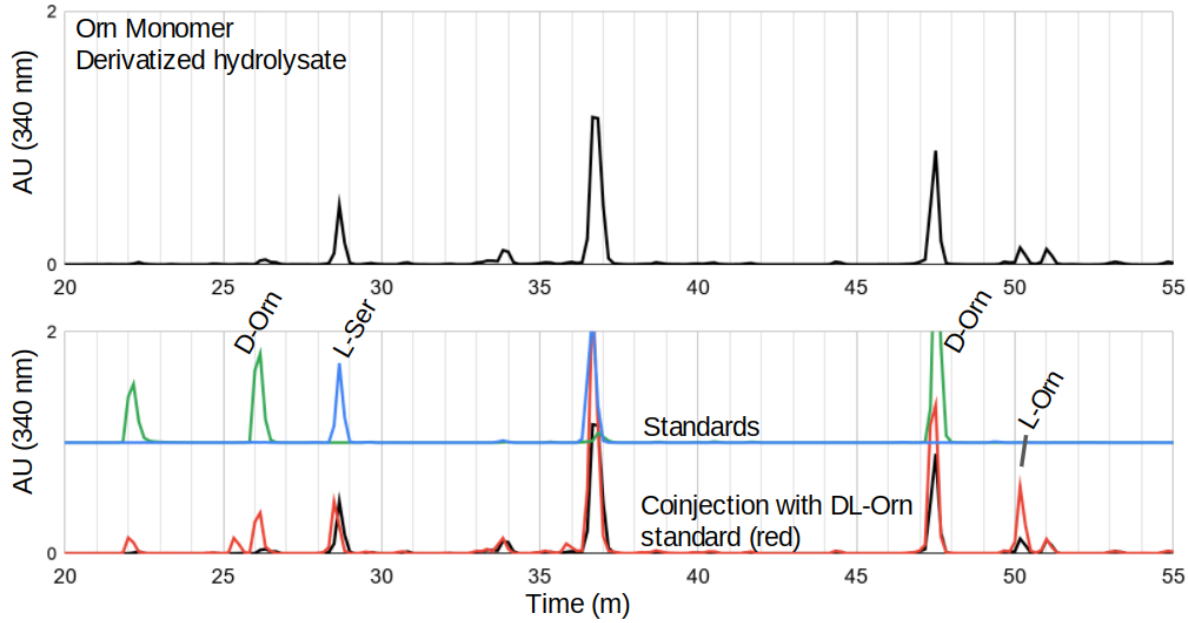


Figure 3.21. RP-HPLC–UV/Vis spectra of FDAA-derivatized DHB-Orn-Ser and amino acid standards.

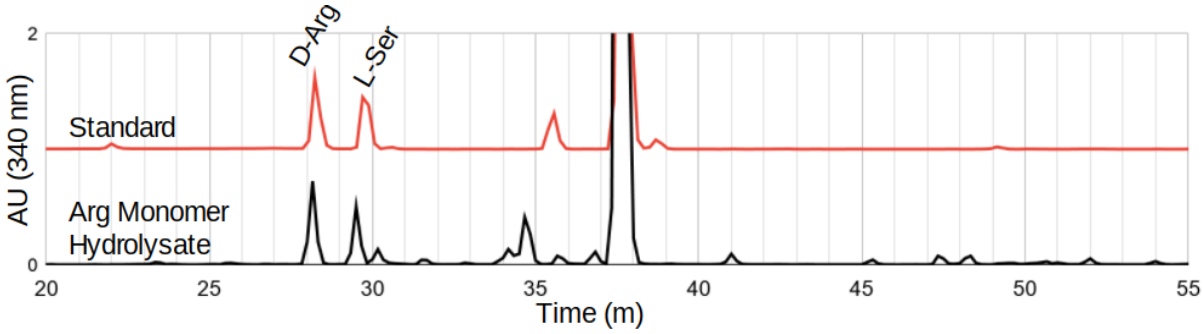


Figure 3.22. RP-HPLC–UV/Vis spectra of FDAA-derivatized DHB-Arg-Ser and mixed amino acid standard.

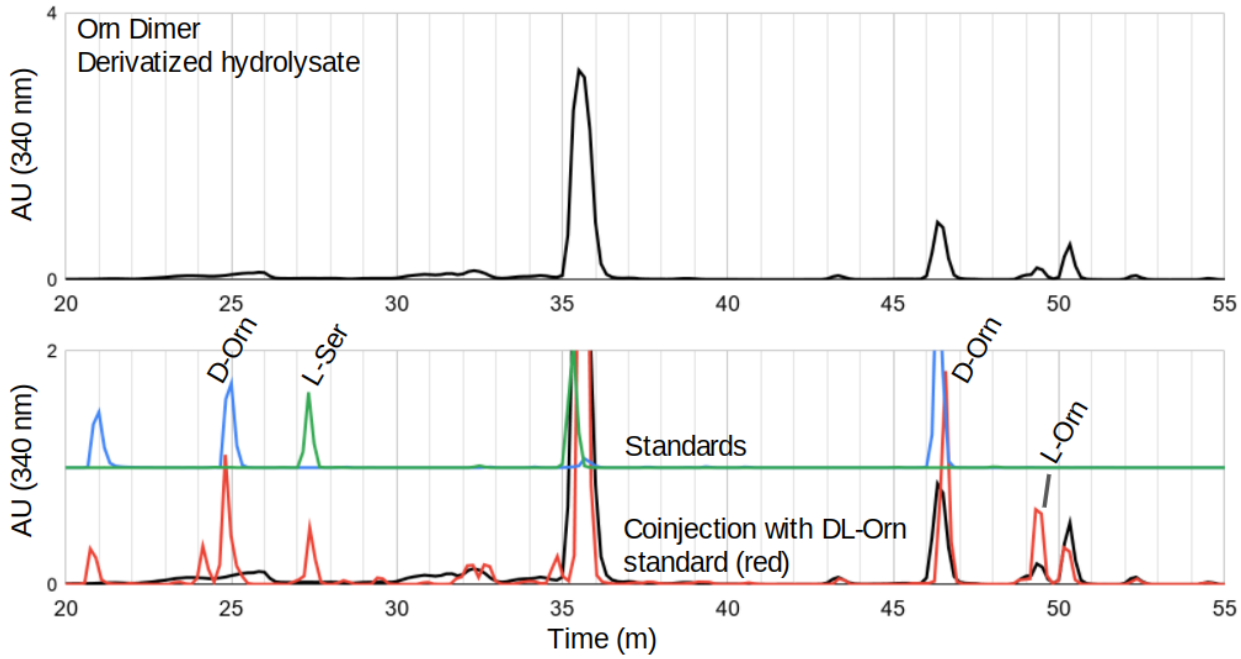


Figure 3.23. RP-HPLC–UV/Vis spectra of FDAA-derivatized (DHB-Orn-Ser)$_2$ and amino acid standards.

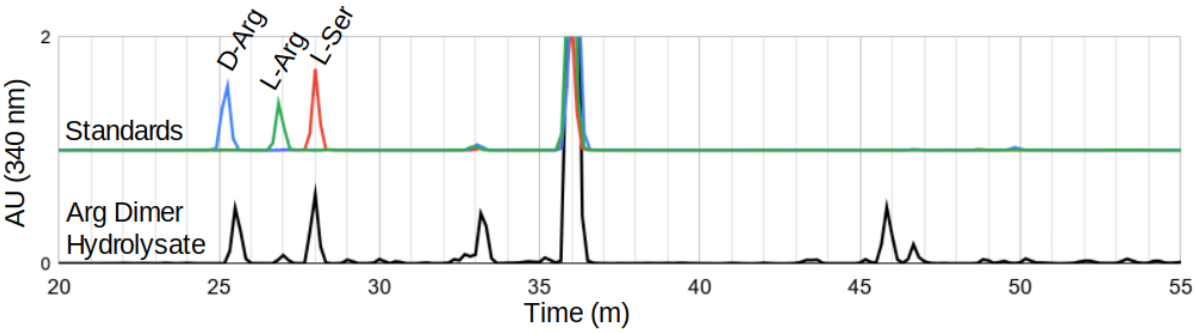


Figure 3.24. RP-HPLC–UV/Vis spectra of FDAA-derivatized (DHB-Arg-Ser)$_2$ and amino acid standards.

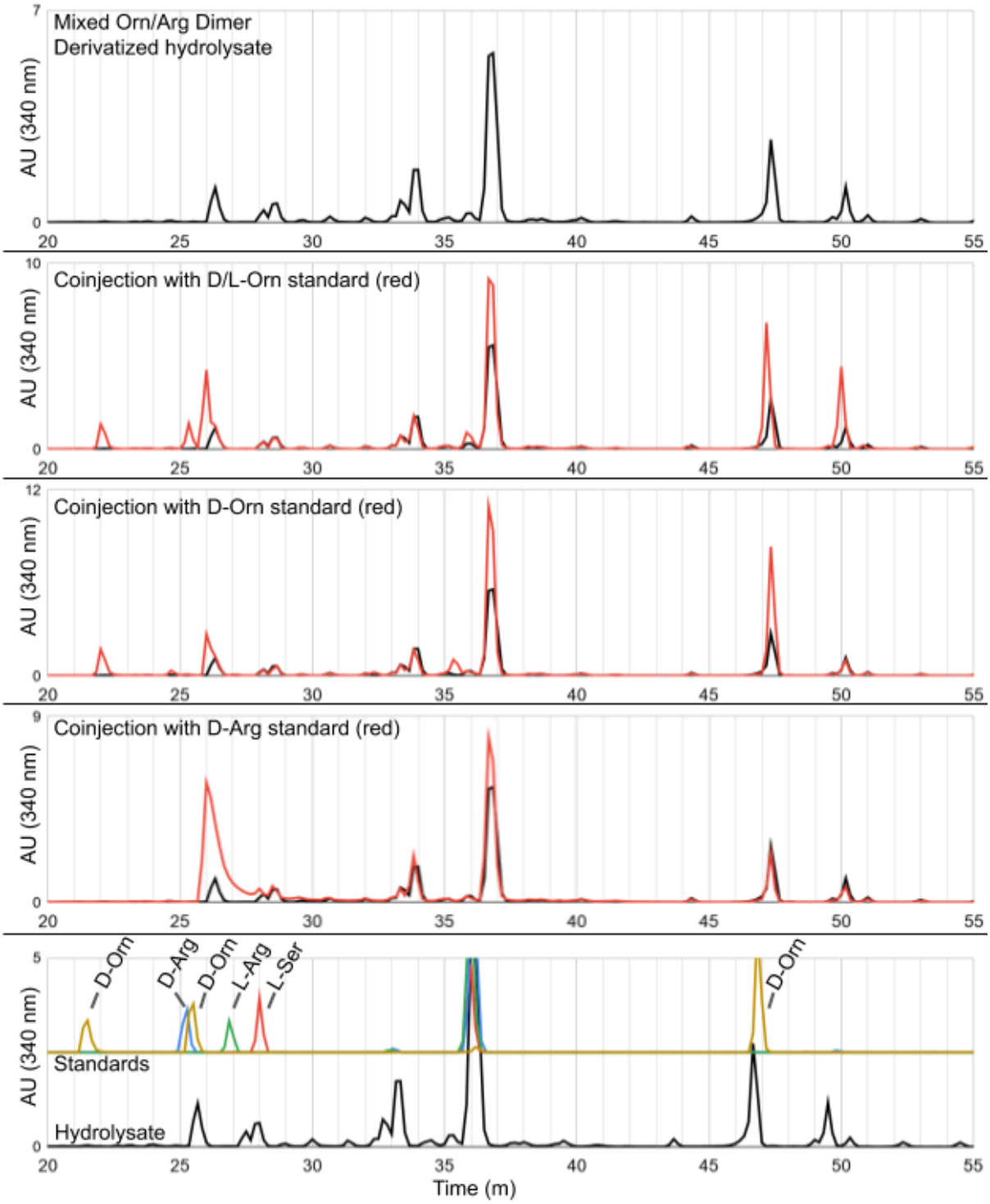


Figure 3.25. RP-HPLC–UV/Vis spectra of FDAA-derivatized (DHB-Orn-Ser)(DHB-Arg-Ser) and amino acid standards.

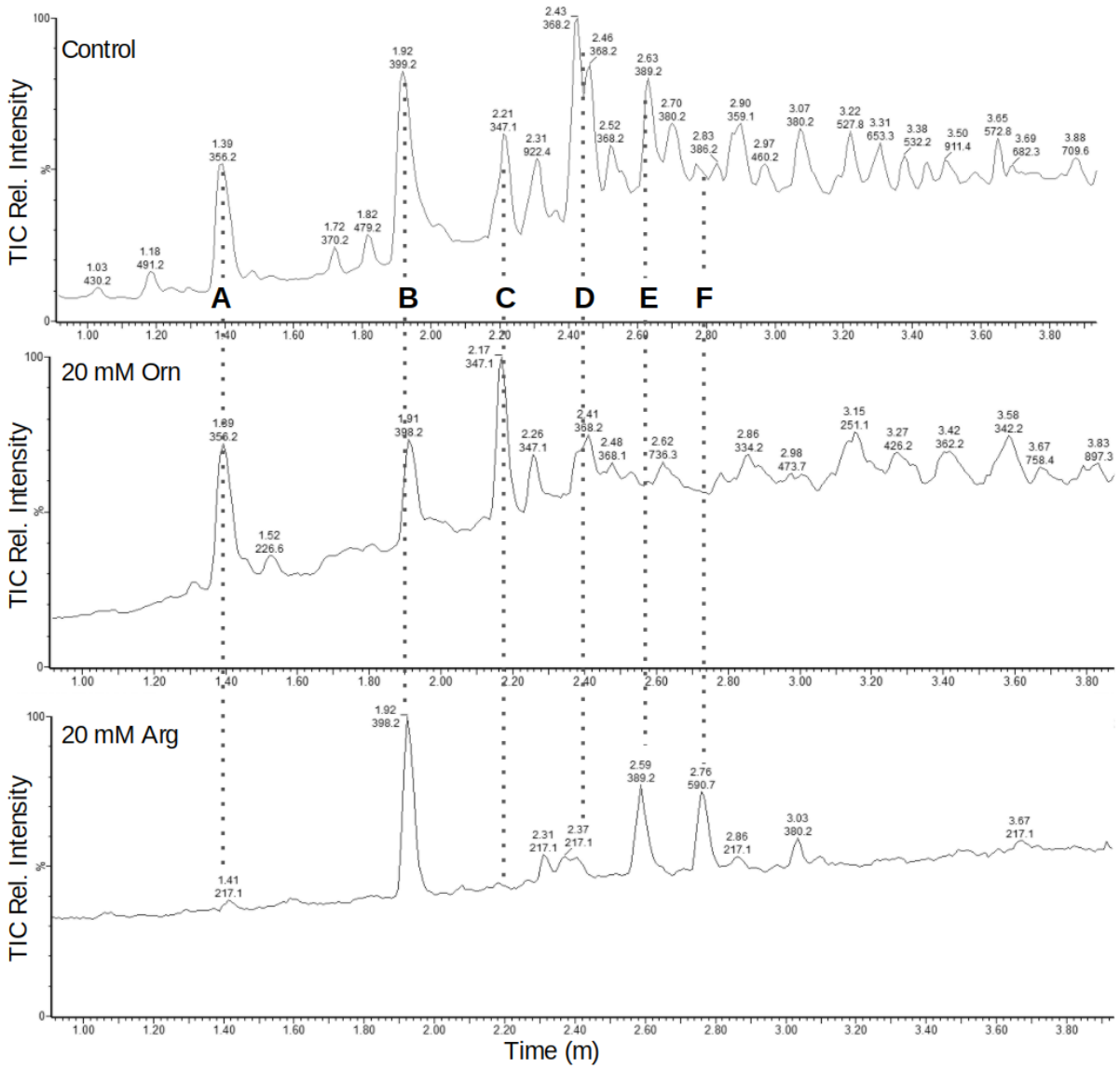


Figure 3.26. Amino acid supplementation studies in *Marinomonas* sp. TW1. UPLC-ESIMS total ion chromatograms of the supernatant extract for *Marinomonas* sp. TW1 grown with (top) no added substrate, (middle) 20 mM L-Orn, or (bottom) 20 mM L-Arg. Labeled peaks correspond to the isolated siderophores DHB-Orn-Ser [A], DHB-Arg-Ser [B], (DHB-Orn-Ser)$_2$ [C], (DHB-Orn-Ser)(DHB-Arg-Ser) [D], and (DHB-Arg-Ser)$_2$ [E], as well as (DHB-Arg-Ser)$_3$ [F], produced only under Arg supplementation.

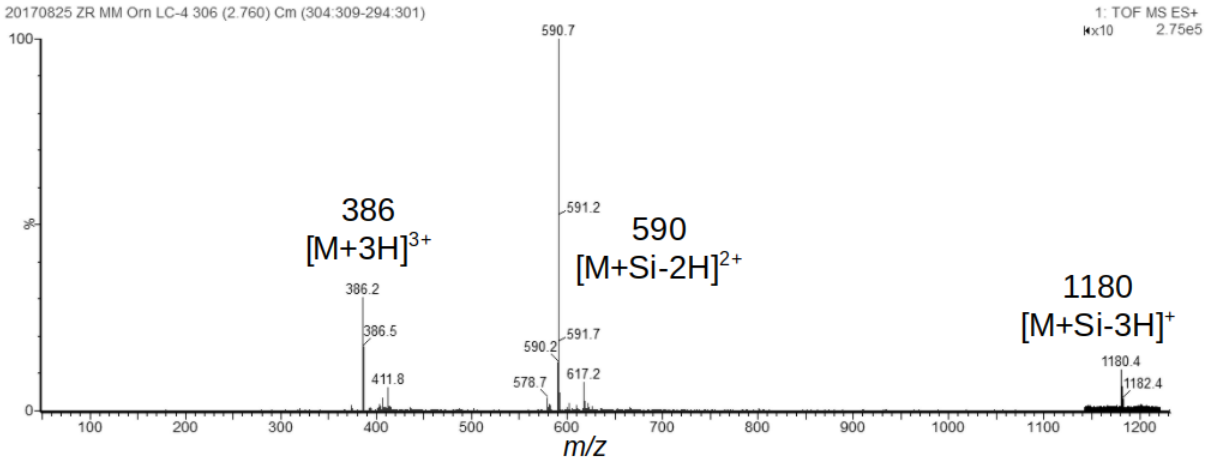


Figure 3.27. ESI mass spectrometry of the Marinomonas sp. TW1 siderophore (DHB-Arg-Ser)$_3$, produced only under Arg supplementation.

3.4. Discussion

3.4.1. CatSCAn-aided DHB-CAA-Ser siderophore discovery

The discovery of cationic amino acid (CAA) spacer-containing triscatechol siderophores trivanchrobactin, cyclic trichrysobactin, and turnerbactin with D-Arg, D-Lys, and L-Orn spacer amino acids, respectively (Figure 3.1), suggested the existence of diastereomeric novel DHB-CAA-Ser siderophores with L-Arg, L-Lys, and D-Orn spacers. Predicting what bacteria, if any, are capable of producing these hypothetical siderophores required a novel genome mining workflow that is not only high-throughput, but also precise enough to distinguish between the three CAAs. The Catechol Siderophore Cluster Analysis workflow (CatSCAn, Figure 3.5) successively filters genomes by the presence of (1) a homolog of *entA*, encoding the final step of DHB synthesis, (2) a two-module NRPS gene with the correct domain architecture, and (3) adenylation domains selective for a CAA and Ser, until only putative DHB-CAA-Ser producers remain. By choosing different enzyme

families or NRPS domain requirements, CatSCAn is easily adapted to search for other siderophore families.

CatSCAn searched over 11,000 representative bacterial genomes, revealing just 99 strains putatively capable of DHB-CAA-Ser siderophore biosynthesis (Figure 3.6 and Table 3.5). Of these strains, eight were predicted to incorporate a novel L-Lys spacer, and only a single strain, *Y. ruckeri* YRB, was predicted to incorporate L-Arg. Based on the results from CatSCAn, siderophores were isolated from several species. *Y. ruckeri* YRB was found to produce ruckerbactin, $(DHB\text{-}Arg\text{-}Ser)_3$. Further characterization of ruckerbactin, carried out by Emil Thomsen, a visiting Masters student in the Alison Butler group, found ruckerbactin to have a structure consistent with $(DHB\text{-}^{L}Arg\text{-}^{L}Ser)_3$.[54] Trivanchrobactin and ruckerbactin are the first reported pair of triscatechol siderophore diastereomers. Their characterization will allow for the exploration of the impact of spacer amino acid chirality on both the chemistry of iron chelation and the biology of siderophore uptake. Additionally, the NRPS from *Marinomonas* sp. TW1 was found to be unusually promiscuous, producing a suite of siderophores with D-Arg, D-Orn, and L-Orn spacer amino acids (Figure 3.20).

Parker R. Stow, a graduate student in the Alison Butler group, tested several more predictions made by CatSCAn.[55] *Y. frederiksenii* ATCC 33641 was found to produce a catechol siderophore consistent with the novel L-Lys diastereomer of linear trichrysobactin, named frederiksenibactin (Figure 3.28). Siderophores consistent with the turnerbactin family (Figure 3.29) were observed in cultures of *Azospirillum brasilense* Az39, *Acinetobacter nosocomialis* 8399, *A. nosocomialis* 216872, *A. nosocomialis* 1571545, and *A. pittii* 573719 (work in progress).[55]

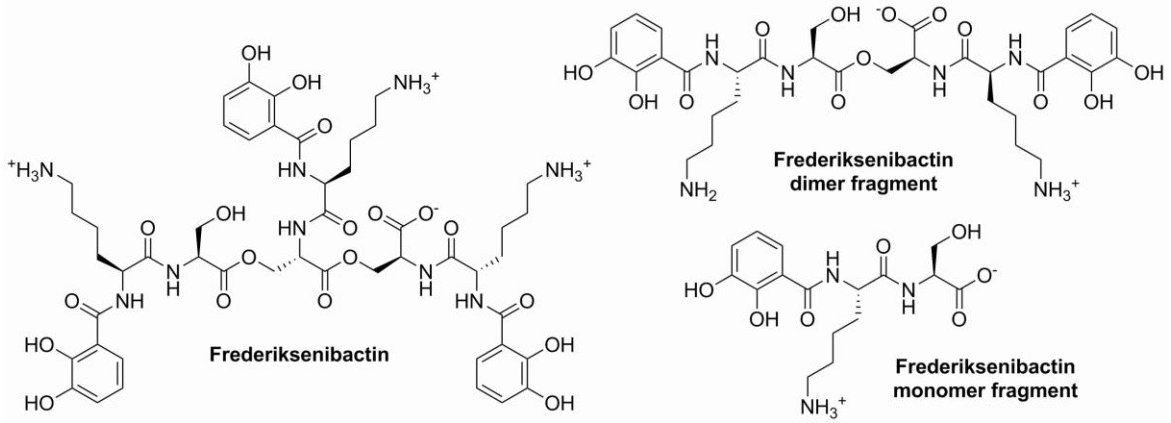


Figure 3.28. Proposed structures of the frederiksenibactin family of siderophores from *Y. frederiksenii* ATCC 33641

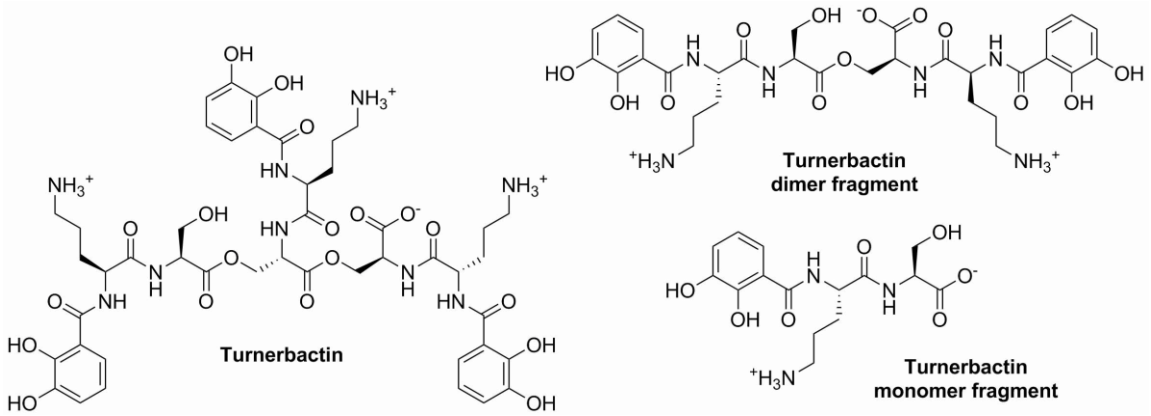


Figure 3.29. Proposed structures of the turnerbactin family of siderophores produced by *Azospirillum brasilense* and *Acinetobacter* spp.

In total, DHB-CAA-Ser siderophores were predicted across 32 genera (Table 3.4), resulting in a diverse set of potential targets for further exploration. With the structural elucidation of ruckerbactin and frederiksenibactin (L-Arg and L-Lys spacers, respectively), only one CAA spacer remains undiscovered: D-Orn. Unfortunately, none of the 100 putative siderophore producers were predicted to encode both an E domain and an Orn-selective A domain. If a siderophore with a D-Orn spacer amino acid exists, potentially promising taxa to investigate are those where CatSCAn could not make a trustworthy amino acid prediction: *Arthrobacter glacialis*, *Chromobacterium* spp., *Gynuella sunshinyii*,

Iodobacter fluviatilis, and *Pararhizobium polonicum* (Tables 3.4 and 3.5). In particular, a recent study of *Chromobacterium violaceum* ATCC 12472 used AntiSMASH to discover the same putative DHB-CAA-Ser biosynthetic gene cluster found by CatSCAn.[56] Knocking out the NRPS halted production of a catechol siderophore named viobactin; however, no structural characterization was performed. The NRPS Stachelhaus codes from *I. fluviatilis* and *G. sunshinyii* are 90% identical and isolation of one can inform the other; likewise with *P. polonicum* and *A. glacialis*.

Table 3.4. Predicted spacer amino acids among genera with DHB-CAA-Ser clusters.*

L-Arg	**L-Lys**	**L-Orn**	**L-CAA**
Yersinia †	*Marinomonas*	*Acinetobacter* †	*Catenovulum*
	Serratia	*Azospirillum* †	*Vibrio*
	Yersinia †	*Blastochloris*	
		Marinomonas	
D-Arg	**D-Lys**	*Massilia*	**D-CAA**
Chitinimonas †	*Brenneria*	*Mesorhizobium*	*Arthrobacter*
Hahella	*Dickeya* †	*Microbulbifer*	*Chromobacterium*
Vibrio †	*Erwinia*	*Notoacmeibacter*	*Gynuella*
	Lonsdalea	*Pseudoduganella*	*Iodobacter*
	Mixta	*Rugamonas*	*Pararhizobium*
	Pantoea	*Teredinibacter* †	
Mixed Arg/Orn	*Pectobacterium* †		**L-XXX**
Marinomonas †	*Serratia* †		*Methylocystis*
	Xenorhabdus		*Methylosinus*

***** Genera are listed by final spacer amino acid predictions made after the characterization of siderophores reported herein. Original predictions and a full list of strains are given in Table 3.5. Dagger symbol (†) indicates that a siderophore with the predicted spacer has been characterized in at least one strain belonging to the genus (**Table 3.2**, or as described in the text). *CAA* indicates that a specific cationic amino acid could not be predicted. Genomic analyses indicate *Methylocystis* and *Methylosinus* may incorporate an amino acid other than a CAA (See Chapter 4).

3.4.2. Development of the CatSCAn workflow

Predicting adenylation domain specificity based on peptide sequence remains an unsolved problem in NRPS research. The first method, developed by Stachelhaus *et al*., extracts from the A domain sequence 10 key residues that interact with the substrate to form a "specificity-conferring code". The "Stachelhaus codes" are still used today, primarily as a portion of NRPSPredictor2,[15] which serves as the A domain substrate prediction tool for AntiSMASH 5.[57] Several other tools have since been developed based on both active site residues and whole domain sequences.[16] Chevrette *et al.* recently developed an improved ensemble method, SANDPUMA, that was found to outperform individual algorithms;[16] unfortunately, the underlying dataset was flawed, leading to its removal from AntiSMASH.[58] Khayatt *et al*. developed an alternative prediction method based on substrate-specific pHMM ensembles.[59] Implementation into CatSCAn led to several mis-predictions, including the *Y. ruckeri* NRPS as Lys (found L-Arg) and the *C. koreensis* NRPS as Orn (found D-Arg). Chevrette *et al.* similarly found that active site-based prediction software out-performed whole domain pHMMs.[16]

The CatSCAn workflow incorporates NRPSPredictor2,[15] as implemented by AntiSMASH 5,[57] to predict A domain selectivity using the 10AA Stachelhaus codes and a minimum 90% identity cutoff. Because this method requires comparison to known A domain selectivities, performance is strongly dependent on the sequence space covered by the dataset. The original NRPSPredictor2 dataset did not contain a single Stachelhaus code that could be used to predict which CAA would be activated in the DHB-CAA-Ser family. Supplementing the dataset with the 10AA codes of previously reported siderophores (Table 3.2) allowed the selectivity of 57 NRPS to be predicted, including the correct prediction of novel siderophores from *Y. ruckeri* YRB and *Y. frederiksenii* ATCC 33641

(Table 3.5). Each structural characterization provides another A domain code that can be used to improve future predictions. Upon adding newly-discovered siderophores into CatSCAn's Stachelhaus code database, the number of strains without a 90 or 100% match dropped from 41 to 13 (Table 3.5). Among structurally characterized DHB-CAA-Ser siderophores, a 90% match is sufficient to differentiate amino acids; however, the Stachelhaus code from *Microbulbifer thermotolerans* DAU221 is 90% identical to codes for both Arg and Orn (Table 3.5), and thus a 90% match, while informative, cannot possibly be a guaranteed predictor.

CatSCAn uses successive filtering of genomes to reduce computing time (Figure 3.5). Further work is saved by using the presence of *entA* as the sole indicator for DHB synthesis, rather than searching for all of *entABCE*. Several other optimization methods were explored during the development of CatSCAn. In one iteration, the 50kbp region around an *entA* homolog was excised, and only that region was scanned for biosynthetic genes. Although this step significantly reduced computing time, siderophores produced by a NRPS in a cluster distinct from *entABCE* homologs were missed, including the previously reported siderophore trivanchrobactin. *Vibrio campellii* DS40M4 produces vanchrobactins, amphi-enterobactins, and anguibactin; homologs of *entA, entC, and entE* are only found in the amphi-enterobactin cluster. The current version does not require co-localization of genes.

3.4.3. Chelation of Ti and Si by DHB-CAA-Ser siderophores

ESIMS spectra for trimeric siderophores in *C. koreensis* DSM 17726 (Figure 3.15) and *Marinomonas sp.* TW1 (Figure 3.27) each contained masses 24 amu higher than expected for the apo siderophore, consistent with the Si(IV)-bound compound. Triscatechol

siderophores enterobactin and salmochelin were previously reported to form Si(IV) complexes.[60] ESIMS spectrum for trimeric ruckerbactin contains a peak (*m/z* 1200) consistent with the Ti(IV) complex (Figure 3.8). The putative Si and Ti complexes may be forming during siderophore extraction or UPLC-ESIMS rather than during growth; nevertheless, these ions hint at an affinity for tetravalent Si and Ti among the DHB-CAA-Ser siderophores that should be explored further.

3.5. Appendix

Table 3.5. Table of CatSCAn predictions

	Predicted		Stachelhaus	Spacer Amino Acid Predictions		
Organism	**CAA**	**RefSeq Assembly**	**Code**	**NRPSPredictor2**	**+ Reported Sids**	**+ Novel Sids**
Chitinimonas koreensis	D-Arg	GCF_000428465.1	DAWDLGCVDK	glu,gln @ 80%	arg @ 90%	arg @ 100%
Hahella chejuensis	D-Arg	GCF_000012985.1	DAWDLGFVDK	glu,gln @ 80%	glu,gln,arg @ 80%	arg @ 90%
Vibrio gazogenes	D-Arg	GCF_002196515.1	DAWDLGCVDK	glu,gln @ 80%	arg @ 90%	arg @ 100%
Vibrio rhizosphaerae	D-Arg	GCF_000711805.1	DAWDLGCVDK	glu,gln @ 80%	arg @ 90%	arg @ 100%
Vibrio ruber	D-Arg	GCF_900163965.1	DAWDLGCVDK	glu,gln @ 80%	arg @ 90%	arg @ 100%
Vibrio spartinae	D-Arg	GCF_900149295.1	DAWDLGCVDK	glu,gln @ 80%	arg @ 90%	arg @ 100%
Vibrio atypicus	D-Arg	GCF_009811315.1	DAWDLGFVDK	glu,gln @ 80%	glu,gln,arg @ 80%	arg @ 90%
Vibrio injenensis	D-Arg	GCF_001895205.1	DAWDLGFVDK	glu,gln @ 80%	glu,gln,arg @ 80%	arg @ 90%
Vibrio metschnikovii	D-Arg	GCF_000176155.1	DAWDLGFVDK	glu,gln @ 80%	glu,gln,arg @ 80%	arg @ 90%
Vibrio qinghaiensis	D-Arg	GCF_002257545.1	DAWDLGFVDK	glu,gln @ 80%	glu,gln,arg @ 80%	arg @ 90%
Vibrio alginolyticus	D-Arg	GCF_000354175.2	SFWDLGFVDK	glu,gln,asp @ 60%	arg @ 100%	arg @ 100%
Brenneria alni	D-Lys	GCF_003666245.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Brenneria nigrifluens	D-Lys	GCF_005484965.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Brenneria salicis	D-Lys	GCF_003315515.1	DSWDMGYVDK	asp,glu @ 70%	lys @ 90%	lys @ 100%
Dickeya chrysanthemi	D-Lys	GCF_000023565.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya dadantii	D-Lys	GCF_003049785.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya dianthicola	D-Lys	GCF_003403135.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya fangzhongdai	D-Lys	GCF_002812485.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya lacustris	D-Lys	GCF_003934295.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya paradisiaca	D-Lys	GCF_000023545.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya solani	D-Lys	GCF_001644705.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya undicola	D-Lys	GCF_000784735.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Dickeya zeae	D-Lys	GCF_000816045.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Erwinia persicina	D-Lys	GCF_001571305.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Erwinia piriflorinigrans	D-Lys	GCF_001050515.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Erwinia rhapontici	D-Lys	GCF_004364855.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Lonsdalea britannica	D-Lys	GCF_003515985.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%

Table 3.5 (cont). Table of CatSCAn predictions

Organism	Predicted CAA	RefSeq Assembly	Stachelhaus Code	Spacer Amino Acid Predictions NRPSPredictor2	+ Reported Sids	+ Novel Sids
Lonsdalea iberica	D-Lys	GCF_002111585.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Lonsdalea populi	D-Lys	GCF_003730035.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Lonsdalea quercina	D-Lys	GCF_000688655.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Marinomonas balearica	D-Lys	GCF_004362145.1	DSWDFGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 90%
Marinomonas mediterranea	D-Lys	GCF_000192865.1	DSWDFGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 90%
Mixta intestinalis	D-Lys	GCF_009914055.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Mixta theicola	D-Lys	GCF_009914075.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Pantoea coffeiphila	D-Lys	GCF_002976475.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Pantoea endophytica	D-Lys	GCF_002858935.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Pantoea rodasii	D-Lys	GCF_002811195.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Pectobacterium parmentieri	D-Lys	GCF_001742145.1	DSWDMGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 100%
Pectobacterium wasabiae	D-Lys	GCF_001742185.1	DSWDMGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 100%
Serratia marcescens	D-Lys	GCF_000513215.1	DSWDMGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 100%
Serratia nematodiphila	D-Lys	GCF_000738675.1	DSWDMGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 100%
Serratia rubidaea	D-Lys	GCF_001572725.1	DSWDMGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 100%
Serratia ureilytica	D-Lys	GCF_000988045.1	DSWDMGYVDK	asp.glu @ 70%	lys @ 90%	lys @ 100%
Xenorhabdus bovienii	D-Lys	GCF_000027225.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus cabanillasii	D-Lys	GCF_003386665.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus eapokensis	D-Lys	GCF_001908105.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus ehlersii	D-Lys	GCF_003610465.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus hominickii	D-Lys	GCF_001721185.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus khoisanae	D-Lys	GCF_001037465.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus miraniensis	D-Lys	GCF_002632615.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus thuongxuanensis	D-Lys	GCF_001908095.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Xenorhabdus vietnamensis	D-Lys	GCF_002127535.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
[Type-D symbiont]	D-Lys	GCF_001485335.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%

Table 3.5 (cont). Table of CatSCAn predictions

	Predicted		Stachelhaus	Spacer Amino Acid Predictions		
Organism	**CAA**	**RefSeq Assembly**	**Code**	**NRPSPredictor2**	**+ Reported Sids**	**+ Novel Sids**
Yersinia ruckeri	L-Arg	GCF_002192595.1	DAWDAGLVDK	arg,glu @ 80%	arg @ 90%	arg @ 100%
Yersinia canariae	L-Lys	GCF_009831415.1	DSWDMGFIDK	8AAs @ 60%	lys @ 90%	lys @ 100%
Yersinia frederiksenii	L-Lys	GCF_000834215.1	DSWDMGFIDK	8AAs @ 60%	lys @ 90%	lys @ 100%
Yersinia hibernica	L-Lys	GCF_004124235.1	DSWDMGFIDK	8AAs @ 60%	lys @ 90%	lys @ 100%
Yersinia kristensenii	L-Lys	GCF_000834865.1	DSWDMGFIDK	8AAs @ 60%	lys @ 90%	lys @ 100%
Yersinia massiliensis	L-Lys	GCF_003048255.1	DSWDMGFIDK	8AAs @ 60%	lys @ 90%	lys @ 100%
Serratia ficaria	L-Lys	GCF_900187015.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Yersinia entomophaga	L-Lys	GCF_001656035.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Yersinia nurmii	L-Lys	GCF_001112925.1	DSWDMGYIDK	8AAs @ 60%	lys @ 100%	lys @ 100%
Acinetobacter apis	L-Orn	GCF_900197575.1	DSIDLGLVDK	4AAs @ 70%	orn @ 80%	orn @ 90%
Acinetobacter calcoaceticus	L-Orn	GCF_002055515.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Acinetobacter genomosp	L-Orn	GCF_001990725.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Acinetobacter nosocomialis	L-Orn	GCF_000814165.3	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Acinetobacter oleivorans	L-Orn	GCF_000196795.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Acinetobacter seifertii	L-Orn	GCF_000368065.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Azospirillum brasilense	L-Orn	GCF_001315015.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Azospirillum doebereinerae	L-Orn	GCF_003989665.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Blastochloris viridis	L-Orn	GCF_001402875.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Marinomonas aquiplantarum	L-Orn	GCF_003314975.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Marinomonas polaris	L-Orn	GCF_900129155.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Massilia armeniaca	L-Orn	GCF_003028855.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Massilia buxea	L-Orn	GCF_009720835.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Massilia plicata	L-Orn	GCF_004421005.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Mesorhizobium zhangyense	L-Orn	GCF_011045115.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Notoacmeibacter ruber	L-Orn	GCF_003668555.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Pseudoduganella violaceinigra	L-Orn	GCF_000425385.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Rugamonas rubra	L-Orn	GCF_900114705.1	DSLDLGLVDK	6AAs @ 70%	orn,arg @ 80%	orn @ 100%
Teredinibacter turnerae	L-Orn	GCF_000023025.1	DSWDGGLVDK	glu @ 80%	orn @ 100%	orn @ 100%

Table 3.5 (cont). Table of CatSCAn predictions

	Predicted		Stachelhaus	Spacer Amino Acid Predictions		
Organism	**CAA**	**RefSeq Assembly**	**Code**	**NRPSPredictor2**	**+ Reported Sids**	**+ Novel Sids**
Marinomonas foliarum	D-Orn/Arg	GCF_003337275.1	DTWDYGYVDK	glu @ 80%	glu @ 80%	orn/arg @ 100%
Marinomonas posidonica	D-Orn/Arg	GCF_000214215.1	DTWDYGYVDK	glu @ 80%	glu @ 80%	orn/arg @ 100%
Marinomonas primoryensis	D-Orn/Arg	GCF_003285205.1	DTWDYGYIDK	glu @ 70%	lys @ 80%	orn/arg @ 90%
Marinomonas profundimaris	D-Orn/Arg	GCF_000508165.1	DTWDYGYVDK	glu @ 80%	glu @ 80%	orn/arg @ 100%
Marinomonas shanghaiensis	D-Orn/Arg	GCF_003259175.1	DTWDYGYVDK	glu @ 80%	glu @ 80%	orn/arg @ 100%
Catenovulum agarivorans	L-XXX	GCF_000281085.1	DTWANCIVYK	ala @ 60%	ala @ 60%	ala @ 60%
Methylocystis hirsuta	L-XXX	GCF_003722355.1	DVGNFIVVEK	ahp @ 60%	ahp @ 60%	ahp @ 60%
Methylocystis rosea	L-XXX	GCF_003855495.1	DVGNFIVVEK	ahp @ 60%	ahp @ 60%	ahp @ 60%
Methylosinus sporium	L-XXX	GCF_003113265.1	DVGNFIVVEK	ahp @ 60%	ahp @ 60%	ahp @ 60%
Microbulbifer thermotolerans	L-XXX	GCF_001617625.1	DSWDAGLVDK	glu @ 80%	orn @ 90%	orn,arg @ 90%
Vibrio palustris	L-XXX	GCF_900162645.1	DAWDMGFVDK	glu,gln,arg @ 70%	arg @ 80%	lys,arg @ 80%
Arthrobacter glacialis	D-XXX	GCF_002909445.1	DSWDLGFVDK	asp @ 80%	asp,orn,arg @ 80%	4AAs @ 80%
Chromobacterium sphagni	D-XXX	GCF_001855575.1	DGEDMGFISK	gly @ 60%	gly,lys @ 60%	lys @ 70%
Chromobacterium subtsugae	D-XXX	GCF_001676875.1	DGEDMGFISK	gly @ 60%	gly,lys @ 60%	lys @ 70%
Chromobacterium vaccinii	D-XXX	GCF_001855275.1	DGEDMGFISK	gly @ 60%	gly,lys @ 60%	lys @ 70%
Chromobacterium violaceum	D-XXX	GCF_000007705.1	DGEDMGFISK	gly @ 60%	gly,lys @ 60%	lys @ 70%
Gynuella sunshinyii	D-XXX	GCF_000940805.1	DTWDLGFIDK	glu,phe @ 70%	4AAs @ 70%	lys @ 80%
Iodobacter fluviatilis	D-XXX	GCF_900451195.1	DAWDLGFIDK	gln,glu,phe @ 70%	5AAs @ 70%	lys,arg @ 80%
Pararhizobium polonicum	D-XXX	GCF_001687365.1	DSWDQGFVDK	asp,glu @ 70%	orn @ 80%	orn,lys @ 80%

3.6. References

1. R. J. Abergel, A. M. Zawadzka, T. M. Hoette and K. N. Raymond, Enzymatic hydrolysis of trilactone siderophores: where chiral recognition occurs in enterobactin and bacillibactin iron transport, *J. Am. Chem. Soc.*, 2009, **131**, 12682-12692.
2. E. A. Dertz, J. Xu and K. N. Raymond, Tren-based analogues of bacillibactin: structure and stability, *Inorg. Chem.*, 2006, **45**, 5465-5478.
3. G. n. Winkelmann and V. Braun, Stereoselective recognition of ferrichrome by fungi and bacteria, *FEMS Microbiol. Lett.*, 1981, **11**, 237-241.
4. B. F. Matzanke, G. I. Müller and K. N. Raymond, Hydroxamate siderophore mediated iron uptake in E. coli: stereospecific recognition of ferric rhodotorulic acid, *Biochem. Biophys. Res. Commun.*, 1984, **121**, 922-930.
5. R. J. Bergeron, J. B. Dionis, G. T. Elliott and S. J. Kline, Mechanism and stereospecificity of the parabactin-mediated iron-transport system in Paracoccus denitrificans, *J. Biol. Chem.*, 1985, **260**, 7936-7944.
6. M. Münzinger, K. Taraz, H. Budzikiewicz, H. Drechsel, P. Heymann, G. Winkelmann and J.-M. Meyer, S, S-rhizoferrin (enantio-rhizoferrin) – a siderophore of Ralstonia (Pseudomonas) pickettii DSM 6297 – the optical antipode of R, R-rhizoferrin isolated from fungi, *Biometals*, 1999, **12**, 189-193.
7. Z. A. Youard, N. Wenner and C. Reimmann, Iron acquisition with the natural siderophore enantiomers pyochelin and enantio-pyochelin in Pseudomonas species, *Biometals*, 2011, **24**, 513-522.
8. T. B. Karpishin and K. N. Raymond, The First Structural Characterization of a Metal–Enterobactin Complex:[V(enterobactin)]2−, *Angew. Chem. Int. Ed Engl.*, 1992, **31**, 466-468.
9. T. B. Karpishin, T. M. Dewey and K. N. Raymond, Coordination chemistry of microbial iron transport. 49. The vanadium(IV) enterobactin complex: structural, spectroscopic, and electrochemical characterization, *J. Am. Chem. Soc.*, 1993, **115**, 1842-1851.
10. T. Baramov, K. Keijzer, E. Irran, E. Mösker, M.-H. Baik and R. Süssmuth, Synthesis and structural characterization of hexacoordinate silicon, germanium, and titanium complexes of the E. coli siderophore enterobactin, *Chemistry*, 2013, **19**, 10536-10542.
11. T. C. Johnstone and E. M. Nolan, Determination of the Molecular Structures of Ferric Enterobactin and Ferric Enantioenterobactin Using Racemic Crystallography, *J. Am. Chem. Soc.*, 2017, **139**, 15245-15250.
12. M. M. Sandy, University of California, Santa Barbara, 2011.
13. M. E. Bluhm, S. S. Kim, E. A. Dertz and K. N. Raymond, Corynebactin and enterobactin: related siderophores of opposite chirality, *J. Am. Chem. Soc.*, 2002, **124**, 2436-2437.
14. R. C. Scarrow, D. J. Ecker, C. Ng, S. Liu and K. N. Raymond, Iron(III) coordination chemistry of linear dihydroxyserine compounds derived from enterobactin, *Inorg. Chem.*, 1991, **30**, 900-906.

15. M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity, *Nucleic Acids Res.*, 2011, **39**, W362-W367.
16. M. G. Chevrette, F. Aicheler, O. Kohlbacher, C. R. Currie and M. H. Medema, SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria, *Bioinformatics*, 2017, **33**, 3202-3210.
17. B. Schwyn and J. B. Neilands, Universal Chemical Assay for the Detection and Determination of Siderophores, *Anal. Biochem.*, 1987, **160**, 47-56.
18. P. Marfey, Determination ofD-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene, *Carlsberg Res. Commun.*, 1984, **49**, 591.
19. Z. L. Reitz, M. Sandy and A. Butler, Biosynthetic considerations of triscatechol siderophores framed on serine and threonine macrolactone scaffolds, *Metallomics*, 2017, **9**, 824-839.
20. D. H. Haft, M. DiCuccio, A. Badretdin, V. Brover, V. Chetvernin, K. O'Neill, W. Li, F. Chitsaz, M. K. Derbyshire, N. R. Gonzales, M. Gwadz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, C. Zheng, F. Thibaud-Nissen, L. Y. Geer, A. Marchler-Bauer and K. D. Pruitt, RefSeq: an update on prokaryotic genome annotation and curation, *Nucleic Acids Res.*, 2018, **46**, D851-D860.
21. M. Balado, A. Souto, A. Vences, V. P. Careaga, K. Valderrama, Y. Segade, J. Rodríguez, C. R. Osorio, C. Jiménez and M. L. Lemos, Two Catechol Siderophores, Acinetobactin and Amonabactin, Are Simultaneously Produced by Aeromonas salmonicida subsp. salmonicida Sharing Part of the Biosynthetic Pathway, *ACS Chem. Biol.*, 2015, **10**, 2850-2860.
22. H. K. Zane, H. Naka, F. Rosconi, M. Sandy, M. G. Haygood and A. Butler, Biosynthesis of amphi-enterobactin siderophores by Vibrio harveyi BAA-1116: identification of a bifunctional nonribosomal peptide synthetase condensation domain, *J. Am. Chem. Soc.*, 2014, **136**, 5615-5618.
23. J. J. May, T. M. Wendrich and M. A. Marahiel, The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin, *J. Biol. Chem.*, 2001, **276**, 7209-7217.
24. T. Franza and D. Expert, The virulence-associated chrysobactin iron uptake system of Erwinia chrysanthemi 3937 involves an operon encoding transport and biosynthetic functions, *J. Bacteriol.*, 1991, **173**, 6874-6881.
25. K. N. Raymond, E. A. Dertz and S. S. Kim, Enterobactin: an archetype for microbial iron transport, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3584-3588.
26. E. J. Dimise, P. F. Widboom and S. D. Bruner, Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile Thermobifida fusca, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15311-15316.
27. S. I. Patzer and V. Braun, Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of Streptomyces sp. ATCC 700974, *J. Bacteriol.*, 2010, **192**, 426-435.
28. M. Bosello, M. Zeyadi, F. I. Kraas, U. Linne, X. Xie and M. A. Marahiel, Structural characterization of the heterobactin siderophores from Rhodococcus erythropolis PR4 and elucidation of their biosynthetic machinery, *J. Nat. Prod.*, 2013, **76**, 2282-2290.

29. S. Nadmid, A. Plaza, G. Lauro, R. Garcia, G. Bifulco and R. Müller, Hyalachelins A-C, unusual siderophores isolated from the terrestrial myxobacterium Hyalangium minutum, *Org. Lett.*, 2014, **16**, 4130-4133.
30. T. W. Giessen, K. B. Franke, T. A. Knappe, F. I. Kraas, M. Bosello, X. Xie, U. Linne and M. A. Marahiel, Isolation, structure elucidation, and biosynthesis of an unusual hydroxamic acid ester-containing siderophore from Actinosynnema mirum, *J. Nat. Prod.*, 2012, **75**, 905-914.
31. N. Gaitatzis, B. Kunze and R. Müller, In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11136-11141.
32. Y. Wen, X. Wu, Y. Teng, C. Qian, Z. Zhan, Y. Zhao and O. Li, Identification and analysis of the gene cluster involved in biosynthesis of paenibactin, a catecholate siderophore produced by Paenibacillus elgii B69, *Environ. Microbiol.*, 2011, **13**, 2726-2737.
33. O. Baars, X. Zhang, F. M. M. Morel and M. R. Seyedsayamdost, The Siderophore Metabolome of Azotobacter vinelandii, *Appl. Environ. Microbiol.*, 2015, **82**, 27-39.
34. M. Bosello, L. Robbel, U. Linne, X. Xie and M. A. Marahiel, Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in Rhodococcus jostii RHA1, *J. Am. Chem. Soc.*, 2011, **133**, 4587-4595.
35. M. Sandy, A. Han, J. Blunt, M. Munro, M. Haygood and A. Butler, Vanchrobactin and anguibactin siderophores produced by Vibrio sp. DS40M4, *J. Nat. Prod.*, 2010, **73**, 1038-1043.
36. A. W. Han, M. Sandy, B. Fishman, A. E. Trindade-Silva, C. A. G. Soares, D. L. Distel, A. Butler and M. G. Haygood, Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont Teredinibacter turnerae T7901, *PLoS One*, 2013, **8**, e76151.
37. M. Balado, C. R. Osorio and M. L. Lemos, Biosynthetic and regulatory elements involved in the production of the siderophore vanchrobactin in Vibrio anguillarum, *Microbiology*, 2008, **154**, 1400-1413.
38. T. Hasan, C. H. Choi and M. H. Oh, Genes Involved in the Biosynthesis and Transport of Acinetobactin in Acinetobacter baumannii, *Genomics Inform.*, 2015, **13**, 2-6.
39. H. Sonoda, K. Suzuki and K. Yoshida, Gene cluster for ferric iron uptake in Agrobacterium tumefaciens MAFF301001, *Genes Genet. Syst.*, 2002, **77**, 137-146.
40. H. Naka, L. A. Actis and J. H. Crosa, The anguibactin biosynthesis and transport genes are encoded in the chromosome of Vibrio harveyi: a possible evolutionary origin for the pJM1 plasmid–encoded system of Vibrio anguillarum?, *Microbiologyopen*, 2013, **2**, 182-194.
41. M. I. González Carreró, F. J. Sangari, J. Agüero and J. M. García Lobo, Brucella abortus strain 2308 produces brucebactin, a highly efficient catecholic siderophore, *Microbiology*, 2002, **148**, 353-360.
42. A. Proschak, P. Lubuta, P. Grün, F. Löhr, G. Wilharm, V. De Berardinis and H. B. Bode, Structure and biosynthesis of fimsbactins A-F, siderophores from Acinetobacter baumannii and Acinetobacter baylyi, *Chembiochem*, 2013, **14**, 633-638.

43. T. A. Ciche, M. Blackburn, J. R. Carney and J. C. Ensign, Photobactin: a catechol siderophore produced by Photorhabdus luminescens, an entomopathogen mutually associated with Heterorhabditis bacteriophora NC1 nematodes, *Appl. Environ. Microbiol.*, 2003, **69**, 4706-4713.
44. M. R. Seyedsayamdost, S. Cleto, G. Carr, H. Vlamakis, M. João Vieira, R. Kolter and J. Clardy, Mixing and matching siderophore clusters: structure and biosynthesis of serratiochelins from Serratia sp. V4, *J. Am. Chem. Soc.*, 2012, **134**, 13550-13553.
45. E. E. Wyckoff, J. A. Stoebner, K. E. Reed and S. M. Payne, Cloning of a Vibrio cholerae vibriobactin gene cluster: identification of genes required for early steps in siderophore biosynthesis, *J. Bacteriol.*, 1997, **179**, 7055-7062.
46. W. Tan, V. Verma, K. Jeong, S. Y. Kim, C.-H. Jung, S. E. Lee and J. H. Rhee, Molecular characterization of vulnibactin biosynthesis in Vibrio vulnificus indicates the existence of an alternative siderophore, *Front. Microbiol.*, 2014, **5**, 1.
47. C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution, *BMC Evol. Biol.*, 2007, **7**, 78.
48. T. Stachelhaus, H. D. Mootz and M. A. Marahiel, The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chem. Biol.*, 1999, **6**, 493-505.
49. L. Fernández, I. Márquez and J. A. Guijarro, Identification of specific in vivo-induced (ivi) genes in Yersinia ruckeri and analysis of ruckerbactin, a catecholate siderophore iron acquisition system, *Appl. Environ. Microbiol.*, 2004, **70**, 5199-5207.
50. M. Sandy and A. Butler, Chrysobactin siderophores produced by Dickeya chrysanthemi EC16, *J. Nat. Prod.*, 2011, **74**, 1207-1212.
51. C. W. Dorsey, M. E. Tolmasky, J. H. Crosa and L. A. Actis, Genetic organization of an Acinetobacter baumannii chromosomal region harbouring genes related to siderophore biosynthesis and transport, *Microbiology*, 2003, **149**, 1227-1238.
52. A. K. Bachhawat and S. Ghosh, Iron Transport in Azospirillum brasilense: Role of the Siderophore Spirilobactin, *Microbiology*, 1987, **133**, 1759-1765.
53. Z. L. Reitz, C. D. Hardy, J. Suk, J. Bouvet and A. Butler, Genomic analysis of siderophore β-hydroxylases reveals divergent stereocontrol and expands the condensation domain family, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 19805-19814.
54. E. Thomsen, Masters Thesis, University of Copenhagen, 2018.
55. P. R. Stow, personal communication, 2020.
56. B. B. Batista, R. E. R. d. S. Santos, R. Ricci-Azevedo and J. F. da Silva Neto, Production and Uptake of Distinct Endogenous Catecholate-Type Siderophores Are Required for Iron Acquisition and Virulence in Chromobacterium violaceum, *Infect. Immun.*, 2019, **87**.
57. K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema and T. Weber, antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline, *Nucleic Acids Res.*, 2019, **47**, W81-W87.
58. Sjshaw, Add back NRPS Predictor3 to the AD domain substrate prediction step - Issue #148 - antismash/antismash, https://github.com/antismash/antismash/issues/148#issuecomment-490140969).

59. B. I. Khayatt, L. Overmars, R. J. Siezen and C. Francke, Classification of the adenylation and acyl-transferase activity of NRPS and PKS systems using ensembles of substrate specific hidden Markov models, *PLoS One*, 2013, **8**, e62136.
60. T. Schmiederer, S. Rausch, M. Valdebenito, Y. Mantri, E. Mösker, T. Baramov, K. Stelmaszyk, P. Schmieder, D. Butz, S. I. Müller, K. Schneider, M.-H. Baik, K. Hantke and R. D. Süssmuth, The E. coli siderophores enterobactin and salmochelin form six-coordinate silicon complexes at physiological pH, *Angew. Chem. Int. Ed Engl.*, 2011, **50**, 4230-4233.

Chapter 4. Genomic diversity of spacer-containing triscatechol siderophores

4.1. Introduction

A series of three triscatechol siderophores have been reported with unique cationic amino acids (CAAs) between DHB and a Tri-Ser core (Figure 4.1): trivanchrobactin (CAA = D-Arg), trichrysobactin (CAA = D-Lys), and turnerbactin (CAA = L-Orn). The chemistry and biosynthesis of these three siderophores are reviewed in Chapter 1. Chapter 3 describes a high-throughput genome mining workflow, Catechol Siderophore Cluster Analysis (CatSCAn), which predicted the production of DHB-CAA-Ser siderophores in nearly 100 bacterial species with no characterized triscatechol siderophore. Three additional DHB-CAA-Ser siderophore families have been characterized (Figure 4.1): ruckerbactin (CAA = L-Arg), frederiksenibactin (CAA = L-Lys), and moabactin (CAA = D-Arg and DL-Orn). CatSCAn produced a rich dataset of newly identified biosynthetic gene clusters, paired with experimentally characterized siderophore structures, that may be used to study the chemistry and biology of the DHB-CAA-Ser siderophore family.

Eleven biosynthetic gene clusters encoding for characterized DHB-CAA-Ser siderophores were analyzed in detail, revealing shared aspects of DHB-CAA-Ser biosynthesis and transport. The turnerbactin pathway was found to involve a putative inner membrane reductase previously only reported among hydroxamate siderophores. The distribution of the newly-discovered siderophore frederiksenibactin (Figure 4.1) was explored; biosynthetic genes have been lost on at least two occasions in *Y. frederiksenii*. Mapping known and predicted siderophore CAAs onto a phylogenetic tree of NRPS proteins reveals repeated E domain loss and A domain selectivity switching during the evolution of DHB-CAA-Ser siderophores. Parsimony suggests that the ancestor of DHB-CAA-Ser

siderophores had a D-Arg spacer (vanchrobactin, Figure 4.1). A model of DHB-CAA-Ser siderophore evolution is presented where NRPS enzyme promiscuity allows for the biosynthesis of new siderophore structures.

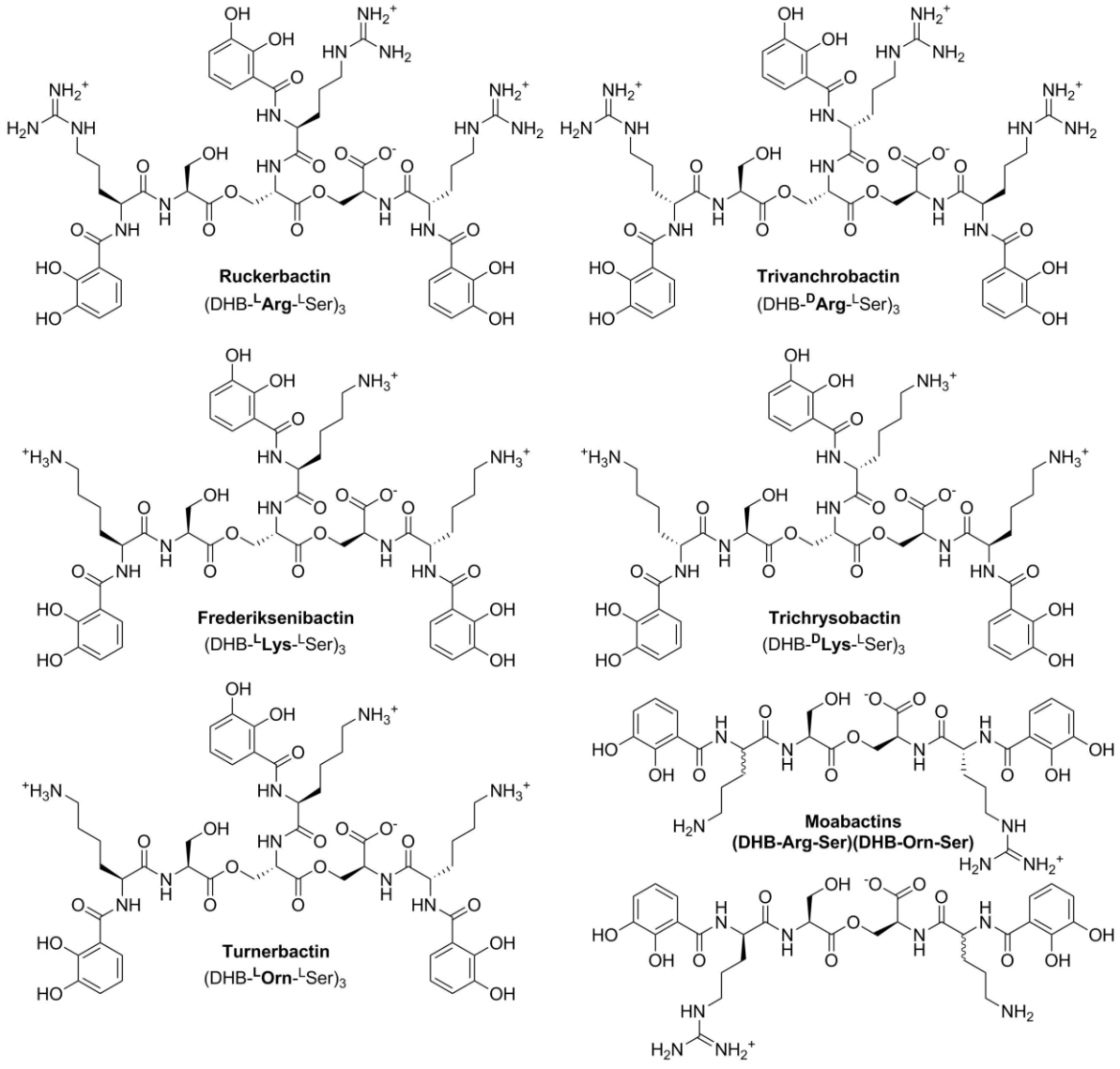


Figure 4.1. DHB-CAA-Ser siderophores.

4.2. Methods

4.2.1. Protein sequence manipulation

The function of enzymes encoded in siderophore biosynthetic gene clusters (BGCs) were predicted by homology to reported enzymes, found using HMMER3[1,2] and NCBI BLAST.[3] Putative DHB-CAA-Ser BGCs were found using CatSCAn (see Chapter 3), resulting in a dataset of 99 BGCs (Chapter 3, Appendix A). NRPS amino acid sequences were extracted from the clusters identified by CatSCAn aligned using MUSCLE.[4] Individual NRPS domains were excised from the multiple sequence alignment using SeaView[5] based on the pHMM envelope regions determined by CatSCAn. NRPS sequences were sorted into non-redundant clusters with CD-HIT.[6] Before phylogenetic analysis, alignments were cleaned with Gblocks[7] using less-stringent settings, as implemented by SeaView.[5]

4.2.2. Phylogenetic analyses

Maximum-likelihood phylogenetic trees were reconstructed from aligned sequences with PhyML[8] (NGPhylogeny.fr[9] webserver) using the appropriate evolutionary model, as chosen by ModelFinder (Bayesian information criterion).[10] Branch support was assessed by bootstrapping (100 bootstrap replicates). The Bayesian phylogenetic tree was reconstructed from the aligned sequences with MrBayes[11] (NGPhylogeny.fr[9] webserver; 10^5 generations, 4 chains, invgamma rates). Dendroscope was used to visualize trees, collapse branches with low support (ML: <75% bootstrap support; Bayesian: <90% posterior probability), and create tanglegrams.[12,13]

4.2.3. Whole genome phylogeny of *Y. frederiksenii*

Thirty *Y. frederiksenii* genome assemblies were retrieved from NCBI RefSeq.[14] The whole genomes were aligned with RealPhy[15] (default parameters) using *Y. enterocolitica* 8081 as the reference genome. A maximum-likelihood phylogenetic tree was reconstructed in IQ-TREE[16] from the resulting 516,076 nucleotide alignment using the TVM+F+I+G4 model, as chosen by ModelFinder (Bayesian information criterion).[10] *Y. enterocolitica* 8081 was used as the outgroup. Branch support was assessed by bootstrapping (100 bootstrap replicates). The phylogenetic tree was visualized in Dendroscope.[12]

4.3. Results and discussion

4.3.1. Common genes shared among Ser-based triscatechol siderophore biosynthetic gene clusters

The DHB-CAA-Ser siderophore biosynthetic gene clusters (BGCs) are similar to each other and to the well-studied enterobactin locus of *E. coli*, allowing for the assignment of homologous enzymes across each siderophore pathway (Table 4.1 and Figure 4.2). Full annotations of the biosynthetic gene clusters are given in Tables 4.2 to 4.12. Several clusters contain at least one gene where a siderophore-related function could not be proposed. For these genes, the encoded Pfam protein family is given in brackets, or the function is listed as "Hypothetical" if no protein family matches were found. These genes were not shared among clusters.

		Biosynthesis		Synthesis Accessory			Export		Import			Fe Release		Regulation	
Siderophore	**Strain**	**NRPS**	**DHB**	**DAHP**	**Ppant**	**MbtH**	**MFS**	**RND**	**OMR**	**PBP**	**ABC**	**Fes**	**PepSY**	**Fur**	**Other**
Ruckerbactin (L-Arg)	*Y. ruckeri* YRB	RucF	RucABCE	-	-	RucI	RucS	-	RupA	RupB	RupCDG	RucH	-	-	-
Vanchrobactin (D-Arg)	*C. koreensis* DSM 17726	VabF	VabABCE	VabG	VabD	VabI	VabS	VaxAB	FvtA	FvtB	FvtCDG	VabH	-	-	VabR
	V. anguillarum 96F	VabF	VabABCE	VabG	VabD	VabI	VabS	-	FvtA	-	-	VabH	-	-	-
	V. campbelli DS40M4	VabF	-	-	-	VabI	VabS	-	FvtA	-	-	VabH	-	-	-
Frederiksenibactin (L-Lys)	*Y. frederiksenii* ATCC 33641	FreF	FreABCE	-	-	FreI	FreS	-	FfuA	FfuB	FfuCDG	FreH	-	-	-
Chrysobactin (D-Lys)	*D. dadantii* 3937	CbsF	CbsABCE	-	-	CbsI	CbsS	-	FctA	FctB	FctCDG	CbsH	-	-	-
	S. marcescens SM6	CbsF	-	-	-	CbsI	CbsS	-	FctA	-	-	CbsH	-	-	-
Turnerbactin (L-Orn)	*T. turnerae*	TnbF	TnbABCE	-	-	TnbI	TnbS	TuxAB	FtbA	-	-	TnbH	FtbB	-	-
	A. nosocomialis 216872	TnbF	TnbABCE	-	TnbD	TnbI	TnbS	TuxB	FtbA	-	-	TnbH	FtbB	Fur	-
	A. brasilense Az39	TnbF	TnbABCE	-	-	TnbI	TnbS	TuxAB	FtbA	-	-	TnbH	FtbB	-	TnbR
Orn/Arg	*Marinomonas* sp. TW1	MmbF	MmbABCE	-	MmbD	MmbI	MmbS	-	FmtA	FmtB	FmtCDG	MmbH	-	-	-

Table 4.1. A summary of enzymes involved in DHB-CAA-Ser siderophore pathways. Only proteins encoded in the locus surrounding the NRPS gene are included; siderophore-related genes may be present in other loci, as described in the text. Full descriptions of each cluster are given in Tables 4.2 to 4.12. NRPS, nonribosomal peptide synthetase; DHB, 2,3-dihydroxybenzoate biosynthesis and activation; DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate synthase; Ppant, phosphopantetheine transferase; MbtH, MbtH-like NRPS accessory protein; MFS, major facilitator superfamily efflux pump; RND, resistance-nodulation-cell division efflux complex; OMR, TonB-dependent outer membrane receptor; PBP, periplasmic binding protein; ABC, ATP-binding cassette transporter; Fes, esterase; PepSY, inner membrane ferric reductase; Fur, ferric uptake regulator.

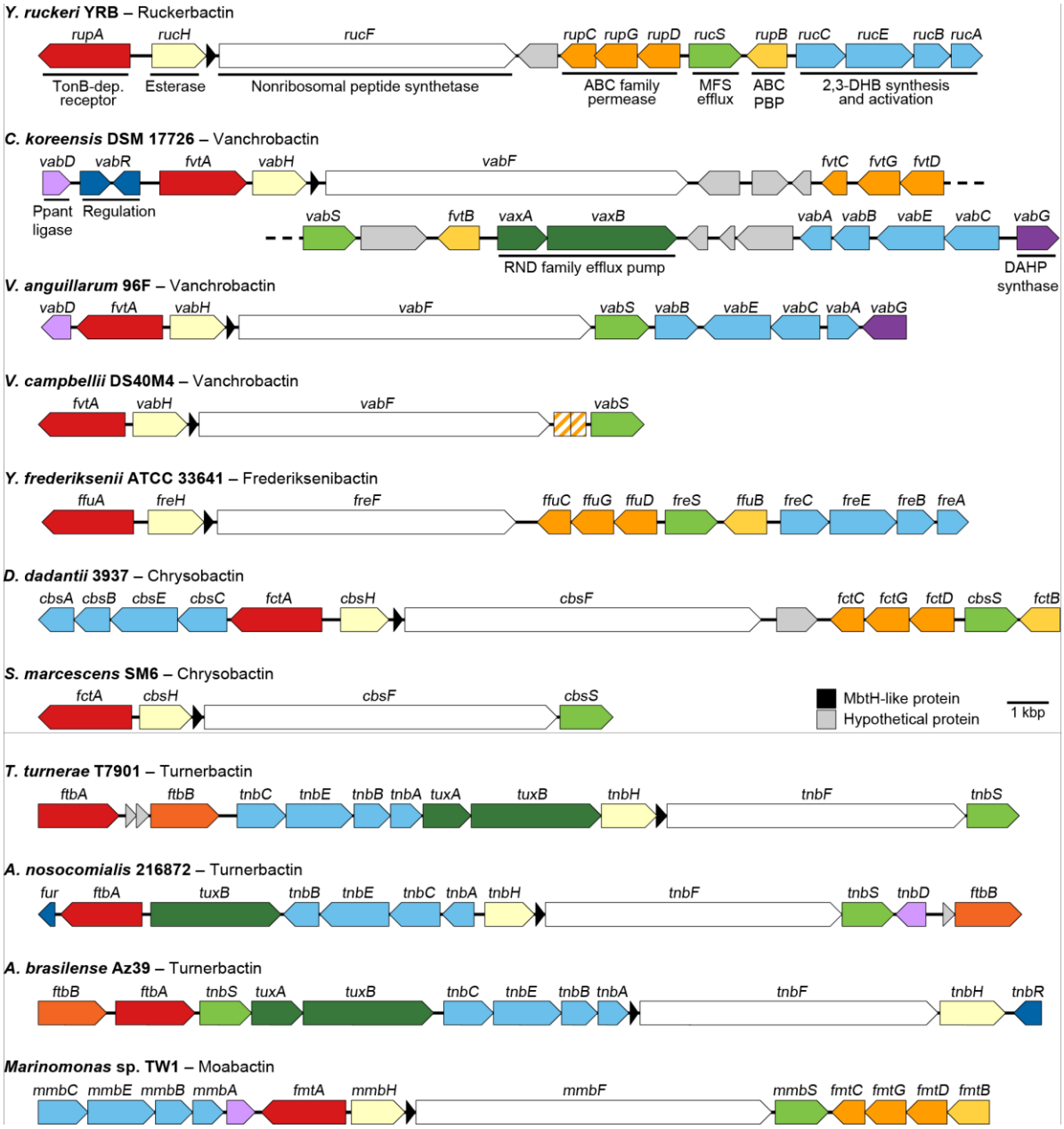


Figure 4.2. Gene clusters encoding for DHB-CAA-Ser siderophore pathways. Each cluster is drawn to scale; arrows represent the direction of transcription. Only the locus surrounding the NRPS gene is shown; siderophore-related genes may be present in other loci, as described in the text. Full descriptions of each cluster are given in Tables 4.2 to 4.12.

Table 4.2. The ruckerbactin biosynthetic gene cluster of *Yersinia ruckeri* YRB[a]

Name	**Locus tag**	**Accession**	**Putative function**
rucA	BD65_RS06575	WP_042528296.1	2,3-dihydro-2,3-DHB dehydrogenase
rucB	BD65_RS06580	WP_042525917.1	Isochorismatase / Aryl carrier protein
rucE	BD65_RS06585	WP_042525919.1	2,3-DHB–AMP ligase
rucC	BD65_RS06590	WP_042525922.1	Isochorismate synthase
rupB	BD65_RS06595	WP_042528298.1	ABC periplasmic binding protein
rucS	BD65_RS06600	WP_042525924.1	MFS transporter
rupD	BD65_RS06605	WP_042528300.1	ABC permease
rupG	BD65_RS06610	WP_042525926.1	ABC permease
rupC	BD65_RS06615	WP_042525928.1	ABC ATP-binding
-	BD65_RS06620	WP_042525930.1	*[Hypothetical]*
rucF	BD65_RS06625	WP_042525932.1	NRPS
rucI	BD65_RS06630	WP_072089264.1	MbtH-like protein
rucH	BD65_RS06635	WP_042525934.1	Esterase
rupA	BD65_RS06640	WP_042525936.1	TonB-dependent receptor

[a] The ruckerbactin locus was obtained from NCBI RefSeq (NZ_CP009539.1). Gene names reflect literature precedent and homology to other triscatechol biosynthetic gene clusters using the mnemonics *ruc* generally and *rup* for **r**uckerbactin **up**take.[19] Genes with no proposed siderophore-related function are left unnamed.

Table 4.3. The vanchrobactin biosynthetic gene cluster of *Chitinimonas koreensis* DSM 17726[a]			
Name	**Locus tag**	**Accession**	**Putative function**
vabD	F559_RS0115770	WP_028446931.1	Ppant transferase
vabR1	F559_RS0115775	WP_051319164.1	Response regulator
vabR2	F559_RS0115780	WP_028446932.1	Response regulator
fvtA	F559_RS0115785	WP_028446933.1	TonB-dependent receptor
vabH	F559_RS0115790	WP_028446934.1	Esterase
vabI	F559_RS0115795	WP_028446935.1	MbtH-like protein
vabF	F559_RS0115800	WP_169730210.1	NRPS
-	F559_RS0115805	WP_028446936.1	*[PF00487: Fatty acid desaturase]*
-	F559_RS0115810	WP_034607404.1	*[PF00497: ABC solute binding protein]*
-	F559_RS0115815	WP_051319165.1	*[Hypothetical]*
fvtC	F559_RS0115820	WP_051319166.1	ABC ATP-binding component
fvtG	F559_RS0115825	WP_028446939.1	ABC permease
fvtD	F559_RS0115830	WP_034607405.1	ABC permease
vabS	F559_RS0115835	WP_028446941.1	MFS transporter
-	F559_RS0115840	WP_028446942.1	[Hypothetical]
fvtB	F559_RS0115845	WP_157462062.1	ABC periplasmic binding protein
vaxA	F559_RS0115850	WP_051319167.1	RND periplasmic adapter
vaxB	F559_RS0115855	WP_028446944.1	RND pump
-	F559_RS0115860	WP_028446945.1	*[Hypothetical]*
-	F559_RS0115865	WP_028446946.1	*[PF04237: YjbR-like DNA binding protein]*
-	F559_RS0115870	WP_051319168.1	*[PF03572: Peptidase family S41]*
vabA	F559_RS0115875	WP_028446947.1	2,3-dihydro-2,3-DHB dehydrogenase
vabB	F559_RS0115880	WP_028446948.1	Isochorismatase / Aryl carrier protein
vabE	F559_RS0115885	WP_034607406.1	2,3-DHB–AMP ligase
vabC	F559_RS0115890	WP_157462036.1	Isochorismate synthase
vabG	F559_RS0115895	WP_051319241.1	DAHP synthase

[a] The vanchrobactin locus was obtained from NCBI RefSeq (NZ_KE386747.1). Gene names reflect homology to other triscatechol biosynthetic gene clusters using the mnemonics *vab* generally, *fvt* for **f**erric **v**anchrobactin **t**ransport, and *vax* for **va**nchrobactin efflu**x**.[20,51] Genes with no proposed siderophore-related function are left unnamed.

Table 4.4. The vanchrobactin biosynthetic gene cluster of *Vibrio anguillarum* 96F[a]

Name	Locus tag	Accession	Putative function
vabD	VAF_RS0102290	WP_013857278.1	Ppant transferase
fvtA	VAF_RS0102295	WP_013857277.1	TonB-dependent receptor
vabH	VAF_RS0102305	WP_019280823.1	Esterase
vabI	VAF_RS0102310	WP_013857274.1	MbtH-like protein
vabF	VAF_RS0102315	WP_019280824.1	NRPS
vabS	VAF_RS0102320	WP_019280825.1	MFS transporter
vabB	VAF_RS0102325	WP_094124597.1	Isochorismatase / Aryl carrier protein
vabE	VAF_RS0102330	WP_013857269.1	2,3-DHB–AMP ligase
vabC	VAF_RS0102335	WP_013857268.1	Isochorismate synthase
vabA	VAF_RS0102340	WP_013857267.1	2,3-dihydro-2,3-DHB dehydrogenase
vabG	VAF_RS0102345	WP_019280827.1	DAHP synthase

[a] The vanchrobactin locus was obtained from NCBI RefSeq (NZ_AEZA01000009.1). Gene names reflect literature precedent and homology to other triscatechol biosynthetic gene clusters using the mnemonics *vab* generally and *fvt* for **f**erric **v**anchrobactin **t**ransport.[20,51]

Table 4.5. The vanchrobactin biosynthetic gene cluster of *Vibrio campbellii* DS40M4[a]

Name	Locus tag	Accession	Putative function
fvtA	DSB67_RS23645	WP_029388829.1	TonB-dependent receptor
vabH	DSB67_RS23650	WP_050545982.1	Esterase
vabI	DSB67_RS23655	WP_010648831.1	MbtH-like protein
vabF	DSB67_RS23660	WP_010648829.1	NRPS
-	DSB67_RS23665	-	ABC permease (Fragment)
-	DSB67_RS23670	-	ABC ATP-binding (Fragment)
vabS	DSB67_RS23675	WP_010648822.1	MFS transporter

[a] The vanchrobactin locus was obtained from NCBI RefSeq (NZ_CP030789.1). Gene names reflect literature precedent and homology to other triscatechol biosynthetic gene clusters using the mnemonics *vab* generally and *fvt* for **f**erric **v**anchrobactin **t**ransport.[20,51]

Table 4.6. The frederiksenibactin biosynthetic gene cluster of *Yersinia frederiksenii* ATCC 33641[a]

Name	**Locus tag**	**Accession**	**Putative function**
freA	DJ58_RS13865	WP_032911814.1	2,3-dihydro-2,3-DHB dehydrogenase
freB	DJ58_RS13870	WP_004711933.1	Isochorismatase / Aryl carrier protein
freE	DJ58_RS13875	WP_032911801.1	2,3-DHB–AMP ligase
freC	DJ58_RS13880	WP_004711937.1	Isochorismate synthase
ffuB	DJ58_RS13885	WP_004711939.1	ABC periplasmic binding protein
freS	DJ58_RS13890	WP_004711941.1	MFS transporter
ffuD	DJ58_RS13895	WP_004711942.1	ABC permease
ffuG	DJ58_RS13900	WP_032911802.1	ABC permease
ffuC	DJ58_RS13905	WP_080544678.1	ABC ATP-binding component
freF	DJ58_RS13910	WP_050504533.1	NRPS
freI	DJ58_RS13915	WP_032911217.1	MbtH-like protein
freH	DJ58_RS13920	WP_080544642.1	Esterase
ffuA	DJ58_RS13925	WP_004709694.1	TonB-dependent receptor

[a] The frederiksenibactin locus was obtained from NCBI RefSeq (NZ_KN150731.1). Gene names reflect homology to other triscatechol biosynthetic gene clusters using the mnemonics *fre* generally and *ffu* for **f**erric **f**rederiksenibactin **u**ptake.

Table 4.7. The chrysobactin biosynthetic gene cluster of *Dickeya dadantii* 3937[a]			
Name	**Locus tag**	**Accession**	**Putative function**
cbsA	DDA3937_RS14720	WP_013318793.1	2,3-dihydro-2,3-DHB dehydrogenase
cbsB	DDA3937_RS14715	WP_013318792.1	Isochorismatase / Aryl carrier protein
cbsE	DDA3937_RS14710	WP_013318791.1	2,3-DHB–AMP ligase
cbsC	DDA3937_RS14705	WP_013318790.1	Isochorismate synthase
fctA	DDA3937_RS14700	WP_013318789.1	TonB-dependent receptor
cbsH	DDA3937_RS14695	WP_033111969.1	Esterase
cbsI	DDA3937_RS14690	WP_013318786.1	MbtH-like protein
cbsF	DDA3937_RS14685	WP_033112377.1	NRPS
-	DDA3937_RS14680	WP_013318784.1	*[PF01548/PF02371: Transposase]*
fctC	DDA3937_RS14675	WP_033111968.1	ABC ATP-binding component
fctG	DDA3937_RS14670	WP_033112376.1	ABC permease
fctD	DDA3937_RS14665	WP_013318781.1	ABC permease
cbsS	DDA3937_RS14660	WP_013318779.1	MFS transporter
fctB	DDA3937_RS14655	WP_033111967.1	ABC periplasmic binding protein
[a] The chrysobactin locus was obtained from NCBI RefSeq (NC_014500.1). Gene names reflect literature precedent and homology to other triscatechol biosynthetic gene clusters using the mnemonics *cbs* generally and *fct* for **f**erric **c**hrysobactin **t**ransport.[17,40] Genes with no proposed siderophore-related function are left unnamed.			

Table 4.8. The chrysobactin biosynthetic gene cluster of *S. marcescens* SM6[a]			
Name	**Locus tag**	**Accession**	**Putative function**
fctA	EG355_RS08760	WP_131164681.1	TonB-dependent receptor
cbsH	EG355_RS08755	WP_041922239.1	Esterase
cbsI	EG355_RS08750	WP_004937955.1	MbtH-like protein
cbsF	EG355_RS08745	WP_131164680.1	NRPS
cbsS	EG355_RS08740	WP_004937950.1	MFS transporter
[a] The chrysobactin locus was obtained from NCBI RefSeq (NZ_SDUW01000003.1). Gene names reflect literature precedent and homology to other triscatechol biosynthetic gene clusters using the mnemonics *cbs* generally and *fct* for **f**erric **c**hrysobactin **t**ransport.[17,40]			

Table 4.9. The turnerbactin biosynthetic gene cluster of *Teredinibacter turnerae* T7901[a]

Name	Locus tag	Accession	Putative function
ftbA	TERTU_RS18025	WP_015817255.1	TonB-dependent receptor
-	TERTU_RS18030	WP_015820466.1	*[Hypothetical]*
-	TERTU_RS18035	WP_015819999.1	*[Hypothetical]*
ftbB	TERTU_RS18040	WP_015819539.1	Ferric reductase
tnbC	TERTU_RS18045	WP_015819844.1	Isochorismate synthase
tnbE	TERTU_RS18050	WP_187148814.1	2,3-DHB–AMP ligase
tnbB	TERTU_RS18055	WP_015816823.1	Isochorismatase / Aryl carrier protein
tnbA	TERTU_RS18060	WP_015818365.1	2,3-dihydro-2,3-DHB dehydrogenase
tuxA	TERTU_RS18065	WP_015817978.1	RND periplasmic adapter
tuxB	TERTU_RS18070	WP_015817325.1	RND pump
tnbH	TERTU_RS18075	WP_015816819.1	Esterase
tnbI	TERTU_RS18080	WP_015818593.1	MbtH-like protein
tnbF	TERTU_RS18085	WP_041590315.1	NRPS
tnbS	TERTU_RS18090	WP_015819712.1	MFS transporter

[a] The turnerbactin locus was obtained from NCBI RefSeq (NC_012997.1). Gene names reflect literature precedent and homology to other triscatechol biosynthetic gene clusters using the mnemonics *tnb* generally, *ftb* for **f**erric **t**urner**b**actin, and *tux* for **tu**rnerbactin efflu**x**.[29] Genes with no proposed siderophore-related function are left unnamed.

Table 4.10. The turnerbactin biosynthetic gene cluster of *Acinetobacter nosocomialis* 216872[a]

Name	**Locus tag**	**Accession**	**Putative function**
fur	J633_RS08670	WP_004887245.1	Ferric uptake regulator
ftbA	J633_RS08675	WP_004887246.1	TonB-dependent receptor
tuxB	J633_RS08680	WP_004887247.1	RND pump
tnbB	J633_RS08685	WP_004887248.1	Isochorismatase / Aryl carrier protein
tnbE	J633_RS08690	WP_004887249.1	2,3-DHB–AMP ligase
tnbC	J633_RS08695	WP_004887250.1	Isochorismate synthase
tnbA	J633_RS08700	WP_004887251.1	2,3-dihydro-2,3-DHB dehydrogenase
tnbH	J633_RS08705	WP_004887252.1	Esterase
thbI	J633_RS08710	WP_004708805.1	MbtH-like
tnbF	J633_RS08715	WP_031952544.1	NRPS
tnbS	J633_RS08720	WP_004887256.1	MFS transporter
tnbD	J633_RS08725	WP_004887259.1	Ppant transferase
-	J633_RS08730	WP_004887261.1	*[Hypothetical]*
ftbB	J633_RS08735	WP_004887263.1	Ferric reductase

[a] The turnerbactin locus was obtained from NCBI RefSeq (NZ_JFEM01000004.1). Gene names reflect homology to other triscatechol biosynthetic gene clusters using the mnemonics *tnb* generally, *ftb* for **f**erric **t**urner**b**actin, and *tux* for **tu**rnerbactin efflu**x**.[18,29] Genes with no proposed siderophore-related function are left unnamed.

Table 4.11. The spirilobactin (turnerbactin) biosynthetic gene cluster of *Azospirillum brasilense* Az39[a]

Name	Locus tag	Accession	Putative function
ftbB	ABAZ39_RS25185	WP_040136709.1	Ferric reductase
ftbA	ABAZ39_RS25190	WP_040136711.1	TonB-dependent receptor
tnbS	ABAZ39_RS25195	WP_040136713.1	MFS transporter
tuxA	ABAZ39_RS25200	WP_040136715.1	RND periplasmic adapter
tuxB	ABAZ39_RS25205	WP_040136717.1	RND pump
tnbC	ABAZ39_RS25210	WP_040136720.1	Isochorismate synthase
tnbE	ABAZ39_RS25215	WP_040136723.1	2,3-DHB–AMP ligase
tnbB	ABAZ39_RS25220	WP_040136725.1	Isochorismatase / Aryl carrier protein
tnbA	ABAZ39_RS25225	WP_040136727.1	2,3-dihydro-2,3-DHB dehydrogenase
tnbI	ABAZ39_RS25230	WP_040136729.1	MbtH-like protein
tnbF	ABAZ39_RS25235	WP_051658519.1	NRPS
tnbH	ABAZ39_RS25240	WP_051658520.1	Esterase
tnbR	ABAZ39_RS25245	WP_040136731.1	GntR family regulator

[a] The turnerbactin locus was obtained from NCBI RefSeq (NZ_CP007795.1). Gene names reflect homology to other triscatechol biosynthetic gene clusters using the mnemonics *tnb* generally, *ftb* for **f**erric **t**urner**b**actin, and *tux* for **tu**rnerbactin efflu**x**.[29] Genes with no proposed siderophore-related function are left unnamed.

Table 4.12. The moabactin biosynthetic gene cluster of *Marinomonas* sp. TW1[a]			
Name	**Locus tag**	**Accession**	**Putative function**
mmbC	OA79_RS10010	WP_063333093	Isochorismate synthase
mmbE	OA79_RS10015	WP_063333094	2,3-DHB–AMP ligase
mmbB	OA79_RS10020	WP_063333095	Isochorismatase / Aryl carrier protein
mmbA	OA79_RS10025	WP_063333096	2,3-dihydro-2,3-DHB dehydrogenase
mmbD	OA79_RS10030	WP_063333097	Ppant transferase
fmtA	OA79_RS10035	WP_063333098	TonB-dependent receptor
mmbH	OA79_RS10040	WP_063333099	Esterase
mmbI	OA79_RS10045	WP_063333100	MbtH-like protein
mmbF	OA79_RS10050	WP_082825444	NRPS
mmbS	OA79_RS10055	WP_063333101	MFS Transporter
fmtC	OA79_RS10060	WP_063333102	ABC ATP-binding component
fmtG	OA79_RS10065	WP_063333103	ABC permease
fmtD	OA79_RS10070	WP_063333104	ABC permease
fmtB	OA79_RS10075	WP_063333105	ABC periplasmic binding protein
[a] The moabactin locus was obtained from NCBI RefSeq (NZ_JSEE01000009.1). Gene names reflect homology to other triscatechol biosynthetic gene clusters using the mnemonics *mmb* generally and *fmt* for **f**erric **m**oabactin **t**ransport.			

A comprehensive review of DHB-CAA-Ser siderophore biosynthesis is given in Chapter 1. The biosynthesis and activation of the DHB precursor is carried out by homologs of EntABCE,[17-20] which are encoded in most DHB-CAA-Ser siderophore clusters. However, one set of DHB genes may be shared among multiple catechol siderophores;[21,22] EntABCE homologs are located in separate loci in *V. campbellii* DS40M4 and *S. marcescens* SM6 (Figure 4.2). The DHB biosynthesis pathway begins with chorismate, part of the shikimate amino acid pathway. The vanchrobactin clusters in *V. anguillarum* and *C. koreensis* each encode a putative 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthase, which controls carbon flow into the shikimate pathway.[23] Bacteria have up to three copies of

DAHP synthases to allow for fine-tuned regulation controlled by multiple conditions.[23-25] In *V. anguillarum* RV22, the DAHP synthase gene *vabG* is required for optimal siderophore production despite the presence of a second homolog elsewhere in the genome.[25]

The first thiolation domain of each NRPS contains the T_E active site motif GGDSI, which is characteristic of a T domain preceding an E domain.[26] The second condensation domain belongs to the $^{D}C_L$ subfamily, which generally follows an E domain and enforces D-stereochemistry.[27] Together, the T_E and $^{D}C_L$ domains suggest that the NRPS responsible for turnerbactin, frederiksenibactin, and ruckerbactin lost an ancestral E domain (Chapter 1).[28-30] NRPS enzymes require the post-translational ligation of a phosphopantetheine (Ppant) arm to each T/ArCP domain for function;[31] genes encoding Ppant transferases are often located elsewhere in the genome.[30] The MbtH-like family of NRPS accessory proteins are often required for NRPS solubility or adenylation domain activity,[32] and they are strictly conserved across DHB-CAA-Ser siderophores.

Fully-synthesized enterobactin is first transported to the periplasm through the major facilitator superfamily (MFS) efflux pump EntS,[33] then exported into the environment by one of several promiscuous resistance-nodulation-cell division (RND) family efflux complexes.[34,35] All ten DHB-CAA-Ser siderophore clusters contain an *entS* homolog, while only three contain putative RND transport genes. The frederiksenibactin locus is also adjacent to a putative RND transport operon (DJ58_RS13935-50); however, the genes are predicted to encode a CusCFBA-like transporter responsible for heavy metal ion efflux.[35]

Fe(III)-bound siderophores are recognized and transported into the periplasm by TonB-dependent outer membrane receptors (OMRs), present in all clusters. Fe(III)-enterobactin is transported to the cytoplasm via the ATP-binding cassette (ABC) transporter complex FepCDG, assisted by periplasmic binding protein FepB. Only half of the clusters contain

fepBCDG homologs. Inner membrane ABC transporters are less specific than OMRs and may be shared across substrates;[36] *V. anguillarum* 96F and *V. campbellii* DS40M4 each encode FepBCDG homologs in the anguibactin cluster, while the *S. marcescens* 96F genome contains an ABC transport locus distinct from any biosynthesis. The vanchrobactin cluster of *V. campbellii* DS40M4 still contains remnants of an ancestral ABC transporter system (Figure 4.2, Table 4.5). None of the turnerbactin BGCs (from *T. turnerae* T7901, *A. nosocomialis* 216872, and *A. brasilense* Az39) encode putative FepBCDG homologs (*vide infra).* In addition to *fepBCDG* homologs, *C. koreensis* DSM 17726 contains an additional gene predicted to encode an ABC solute binding protein (F559_RS0115810). This gene has no homology to the PBP *fepB*, and belongs to a different Pfam (PF00497 vs PF01497 for PBPs). This Pfam corresponds to "family 3" of binding proteins, which are often specific for polar amino acids.[37] Therefore, F559_RS0115810 could be involved in importing the L-Arg precursor for vanchrobactin biosynthesis.

In the cytoplasm of *E. coli*, the trilactone core of ferric enterobactin is cleaved by the esterase Fes,[38] allowing iron release by the YqjH-mediated reduction of Fe(III) to Fe(II).[39] Fes homologs are strictly conserved across the DHB-CAA-Ser BGCs (Figure 4.2). Surprisingly, the Fes homolog CbsH of chrysobactin producer *D. dadantii* 3937 is not required for the utilization of iron from ferric chrysobactin, as determined by Mössbauer spectroscopy.[40] Nevertheless, in the absence of CbsH, growth is inhibited, and ferric chrysobactin accumulates in the cytoplasm. Rauscher *et al.* propose that CbsH is not an esterase, but rather a serine peptidase that cleaves the iron-free chrysobactin backbone to form DHB-Lys and serine, preventing toxic iron sequestration by chrysobactin.[40] No DHB-CAA-Ser Fes homolog has been investigated since the report of oligomeric vanchrobactin and chrysobactin, and cleaved DHB-Lys has not been reported in bacterial cultures. The

factors governing the varying ratios of monomer, dimer, linear trimer, and cyclic trimer remain unknown; further investigation of CbsH and other Fes homologs will provide key insights into whether monomer and dimer arise from incomplete biosynthesis or from hydrolysis.

Instead of an ABC transporter system homologous to FepBCDG, each turnerbactin BGC (from *T. turnerae* T7901, *A. nosocomialis* 216872, and *A. brasilense* Az39) encodes a PepSY_TM transmembrane helix protein (Pfam: PF03929) (Tables 4.9-4.11). The enzymes show homology to the inner membrane proteins FegB, FoxB, and VciB of *Bradyrhizobium japonicum*, *Pseudomonas aeruginosa*, and *Vibrio cholerae* respectively.[41-43] FegB and FoxB are required for hydroxamate siderophore utilization in their respective strains, although the precise functions of these enzymes remain unknown.[41,42] VciB was demonstrated to reduce periplasmic ferric iron *in vivo*, putatively using electrons from the electron transport chain.[43] Reductase activity required two His residues;[43] both are conserved across FegB, FoxB, and the turnerbactin PepSY_TM enzymes. Furthermore, *T. turnerae* T7901 and *A. nosocomialis* 216872 are the only DHB-CAA-Ser producers in Table 4.1 that do not encode a homolog of cytoplasmic reductase YqjH anywhere in their genomes. Thus, we propose an alternative pathway for iron acquisition from ferric turnerbactin: after FtbA-mediated outer membrane transport, PepSY_TM protein FtbB reduces periplasmic ferric turnerbactin, then the Feo system transports ferrous iron to the cytoplasm. (Figure 4.3). The turnerbactin BGCs encode the esterase TnbH, which is predicted to be cytoplasmic by SignalP.[44] If ferric turnerbactin does not cross the inner membrane, TnbH may cleave apo-turnerbactin directly after synthesis, similar to the salmochelin esterase IroE.[38]

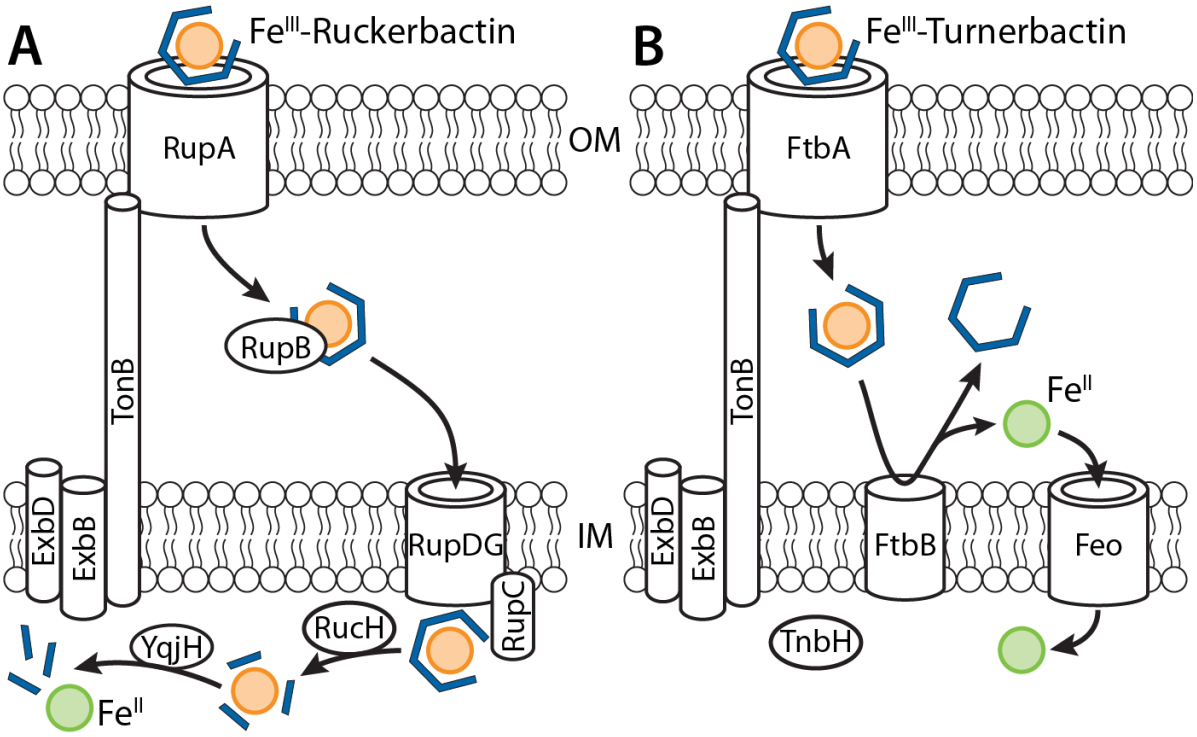


Figure 4.3. Proposed DHB-CAA-Ser siderophore-mediated iron acquisition pathways for (A) ruckerbactin and (B) turnerbactin. Both ferric siderophores are transported across the outer membrane (OM) by a TonB-ExbB-ExbD-powered outer membrane receptor. (A) Fe(III)-ruckerbactin is likely escorted through the periplasm by RupB and transported intact through the IM by RupCDG. Cytoplasmic RucH cleaves the Fe(III)-Sid backbone prior to Fe(III) reduction by a YqjH homolog located elsewhere in the genome. (B) In contrast to other DHB-CAA-Ser siderophores, periplasmic Fe(III)-turnerbactin is putatively reduced by inner membrane reductase FtbB. Ferrous iron is then transported to the cytoplasm by the Feo system. The fate of turnerbactin and the function of cytoplasmic esterase TnbH are unclear.

The ferric uptake regulator (Fur) protein is generally responsible for blocking transcription of iron acquisition genes under iron-replete conditions, ensuring that siderophores are only produced during iron starvation.[45,46] Only the *Acinetobacter* turnerbactin locus contains a Fur homolog, although Fur-mediated regulation of other DHB-

CAA-Ser siderophore pathways has been demonstrated.[25,40,47] The vanchrobactin cluster of *V. anguillarum* RV22 also encodes VabR, a putative LysR-family regulator which was found to promote the transcription of the DAHP synthase gene *vabG*, but no other vanchrobactin-related genes.[25] The vanchrobactin locus in *V. anguillarum* 96F does not contain a *vabR* homolog. The siderophore gene clusters in *C. koreensis* DSM 17726 and *A. brasilense* Az39 each contain putative regulatory genes (of the single-domain response regulator[48] and GntR[49,50] families, respectively).

4.3.2. Distribution of frederiksenibactin biosynthesis genes in *Y. frederiksenii*

Frederiksenibactin biosynthesis genes were detected in 25 out of 30 sequenced *Y. frederiksenii* strains. Using RealPhy and IQTREE,[15,16] a whole-genome phylogeny of these Y. frederiksenii strains was reconstructed (Figure 4.4). Four of the *fre* negative strains form a single clade, and no other siderophore biosynthesis genes were found in their genomes. *Y. frederiksenii* isolate BR166/97 has independently lost frederiksenibactin biosynthesis and export genes *freABCEFIS*, but retains inner membrane permease genes *ffuBCDG* and the esterase gene *freH* (ERS008531_02295-9), suggesting the strain may still be able to use pirated frederiksenibactin as a source of iron. BR166/97 and eight other strains contain a chromosomally-encoded NRPS-independent siderophore biosynthesis cluster (Figure 4.4), consistent with the report that some *Y. frederiksenii* strains produce aerobactin.[52]

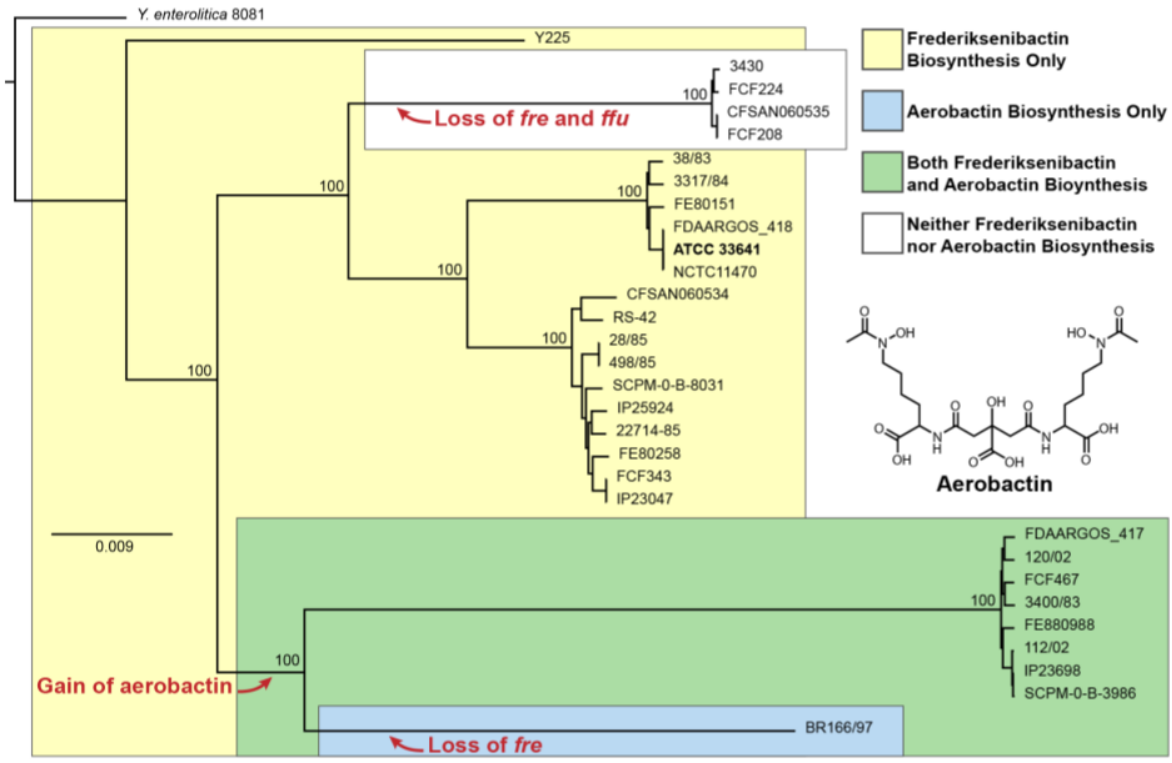


Figure 4.4. Distribution of siderophore biosynthesis genes in *Yersinia frederiksenii*. Two clades have independently lost frederiksenibactin biosynthesis genes (*fre*); strain BR166/97 retained the genes for ferric frederiksenibactin uptake (*ffu*). Strains with aerobactin biosynthesis genes form a single clade. The strain *Y. frederiksenii* ATCC 33641, from which frederiksenibactin was isolated, is shown in bold. *Y. enterolitica* 8081 was chosen as an outgroup and reference strain for the whole-genome phylogeny, reconstructed using RealPhy and IQTREE.[15,16] The scale bar indicates the average number of substitutions per site.

4.3.3. Evolution of NRPS genes encoding for the production of spacer-containing triscatechol siderophores

The spacer amino acid variation seen in DHB-CAA-Ser siderophores comes from the NRPS: the selectivity of the first A domain determines the identity of the amino acid, and the absence or presence of the E domain determines the chirality.[30] There are two major methods by which a change in A domain selectivity may occur.[53] Mutations in the NRPS sequence will naturally accumulate over time; sufficient genetic drift and selective pressure may enforce a new A domain selectivity. Alternatively, recombination may replace an existing A domain with a new domain that originated elsewhere. These two methods may be distinguished phylogenetically. In the former case, the entire NRPS is mutating in concert, and phylogenetic trees of individual domains will have topologies consistent with each other and with the NRPSs as a whole. A recombinant domain, in contrast, does not have the same evolutionary history as its neighbor.

4.3.3.1. Evidence of intra-NRPS recombination in methanotrophs

Amino acid sequences for the first and second A domains were extracted from the CatSCAn results (see Chapter 3) and combined to form a single maximum-likelihood phylogenetic tree. The A_1 and A_2 sequences largely form separate clades; however, A_1 domains from methanotrophs *Methylocystis* and *Methylosinus* are sister to the *Methylocystis* and *Methylosinus* A_2 domains (Figure 4.5). Thus, the ancestor of the methanotrophs' NRPSs underwent a relatively recent recombination event where A_2 was copied and replaced the original A_1. Accordingly, *Methylosinus sporium* NRPS has a high inter-module percent identity compared with other DHB-CAA-Ser NRPS (Figure 4.6). The neighboring C and T

domains in the *Methylocystis* NRPS do not have excess sequence identity, suggesting that the recombination event was limited to the A domain. The Stachelhaus code of the methanotroph A_1 domains (DVGNFIVVEK) is dissimilar to any known codes in our database, and shares only 50% identity with the Stachelhaus code of the methanotroph A_2 domain.

Phylogenetic trees of individual NRPS domains (C_{start}, A_1, E, $^{D}C_L$, A_2, and Te) were reconstructed and compared as tanglegrams in an attempt to detect inter-cluster recombination.[13] Unfortunately, the phylogenetic trees had poor branch support regardless of methods used (nucleotide *versus* amino acid sequences, MUSCLE *versus* hmmalign *versus* PROMALS3D alignment, and maximum likelihood *versus* Bayesian estimation). None of the domain tree pairs had any incongruencies after low-support branches were collapsed, including A_1 and A_2 (Figure 4.7), which should have different topologies due to the methanotroph intra-cluster recombination described above. Single NRPS domains may not have sufficient phylogenetic signal to support or refute inter-cluster recombination.

Figure 4.5 Phylogenetic reconstruction of DHB-CAA-Ser siderophore NRPS adenylation domain amino acid sequences. Adenylation domains from module 1 (A1) and module 2 (A2) form distinct clades, with the exception of *Methylosinus sporium* and *Methylocystis hirsuta* (highlighted), where the first adenylation domain is sister to the second. NRPS sequences from 99 CatSCAn genomes were sorted into non-redundant clusters with CD-HIT[6] at a 70% identity threshold, resulting in 45 representative strains. A maximum-likelihood phylogenetic tree was reconstructed from the aligned sequences with PhyML[8] (NGPhylogeny.fr[9] webserver) using the LG+I+F+G4 model, as chosen by ModelFinder (Bayesian information criterion).[10] Branch support was assessed by bootstrapping (100 bootstrap replicates).

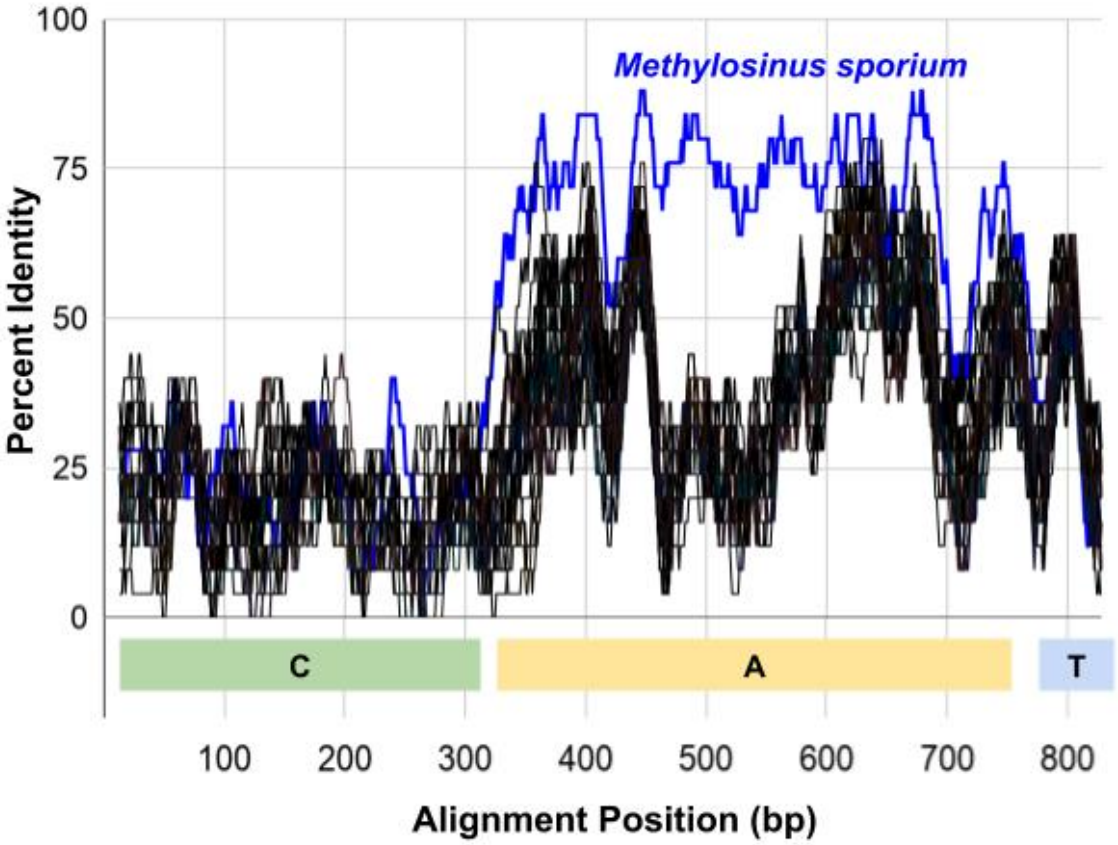


Figure 4.6 Inter-module amino acid identity across DHB-CAA-Ser siderophore NRPSs. Each series depicts the 25 amino acid trailing average percent identity between modules 1 and 2 of a single NRPS sequence. The two adenylation domains of the *Methylosinus sporium* NRPS (blue) share a high percent identity compared to the adenylation domains of other NRPSs, suggesting A domain recombination occurred. NRPS sequences from 99 CatSCAn genomes were sorted into non-redundant clusters with CD-HIT[6] at a 50% identity threshold, resulting in 30 representative strains. For each strain, the C_{start}-A_1-T_1 (module 1) and the $^{D}C_L$-A_2-T_2 (module 2) regions were excised from the NRPS and all 60 modules were aligned using MUSCLE.[4] All positions with gaps were removed before the calculation of rolling percent identity between modules.

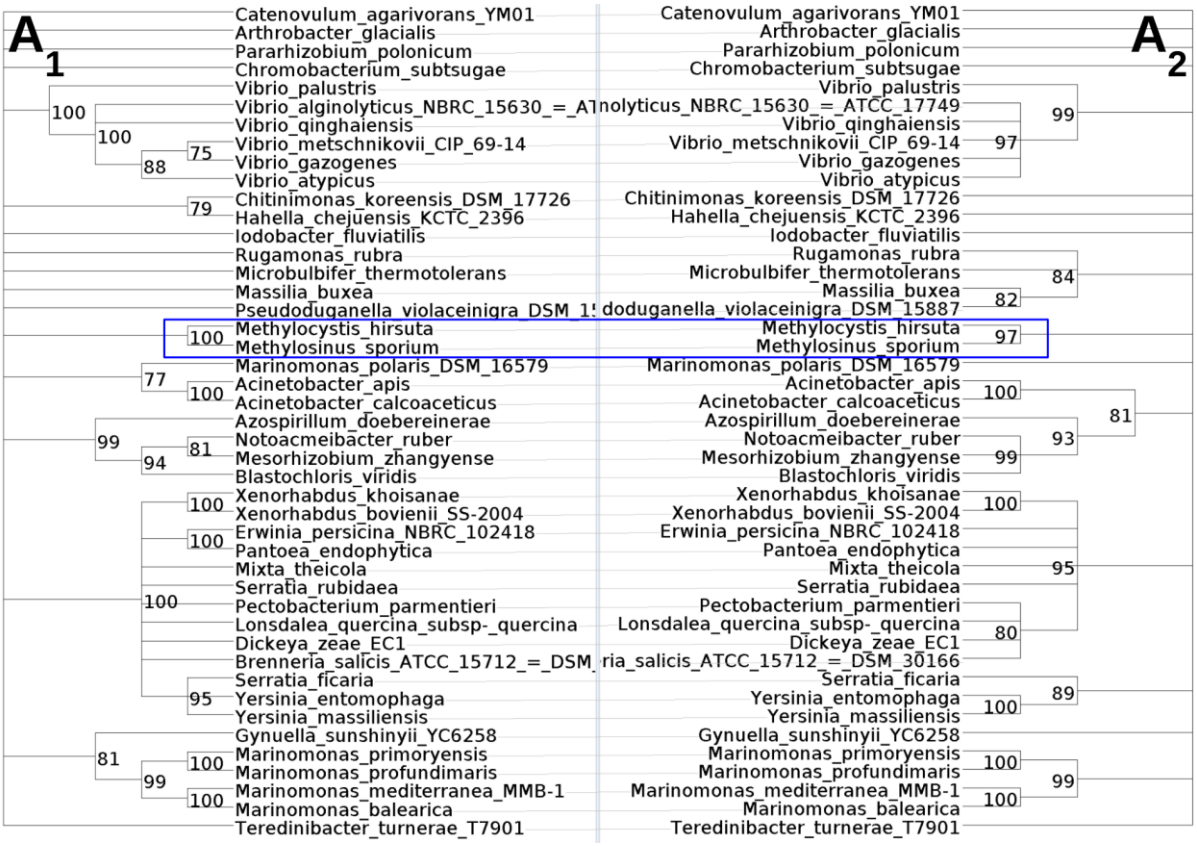


Figure 4.7 Tanglegram of reconstructed phylogenies of DHB-CAA-Ser siderophore NRPS amino acid sequences, A_1 and A_2. No topological discrepancies were found between the first and second A domains, suggesting no recombination occurred. The methanotroph (boxed in blue) intra-cluster recombination event (Figure 4.5 and Figure 4.6) was not detectable by this method. NRPS sequences from 99 CatSCAn genomes were sorted into non-redundant clusters with CD-HIT[6] at a 70% identity threshold, resulting in 45 representative strains. Adenylation domains were excised by trimming to the adenylation domain Pfam[54] envelope range determined by hmmsearch[1] and aligned with MUSCLE.[4] A maximum-likelihood phylogenetic tree was reconstructed from the each alignment with PhyML[8] (NGPhylogeny.fr[9] webserver) using the LG+I+F+G4 model, as chosen by ModelFinder (Bayesian information criterion).[10] Branch support was assessed by bootstrapping (100 bootstrap replicates), and branches with <75% support were collapsed to polytomies. The tanglegram was created with Dendroscope.[12]

4.3.3.2. Reconstruction of the NRPS phylogenetic tree

A whole-NRPS phylogenetic tree was reconstructed to explore the evolution of DHB-CAA-Ser siderophore diversity. The sequences of NRPS responsible for DHB-CAA-Ser siderophores characterized previously (Table 4.1) were included in the analysis. Sequences with recombinant regions may lead to erroneous topologies in phylogenetic analyses,[55] so the NRPSs from methanotrophs *Methylocystis* and *Methylosinus* were excluded. Nineteen genomes from *Nocardia* spp. were predicted by CatSCAn to have NRPS with the expected NRPS architectures C_{start}-A-T-$^{D}C_{L}$-A-T-Te or C_{start}-A-T-E-$^{D}C_{L}$-A-T-Te, but eliminated due to a second A domain not predicted to be Ser-selective. These *Nocardia* NRPSs were introduced back into the dataset as an outgroup to root the phylogeny. MUSCLE produced dubious alignments between sequences with and without an E domain, so the E domain was excised in addition to Gblocks processing. The maximum-likelihood phylogenetic tree of DHB-CAA-Ser NRPSs is presented in Figure 4.8. The ML tree showed excellent agreement with the Bayesian tree produced by MrBayes, and both methods failed to resolve similar polytomies (Figure 4.9).

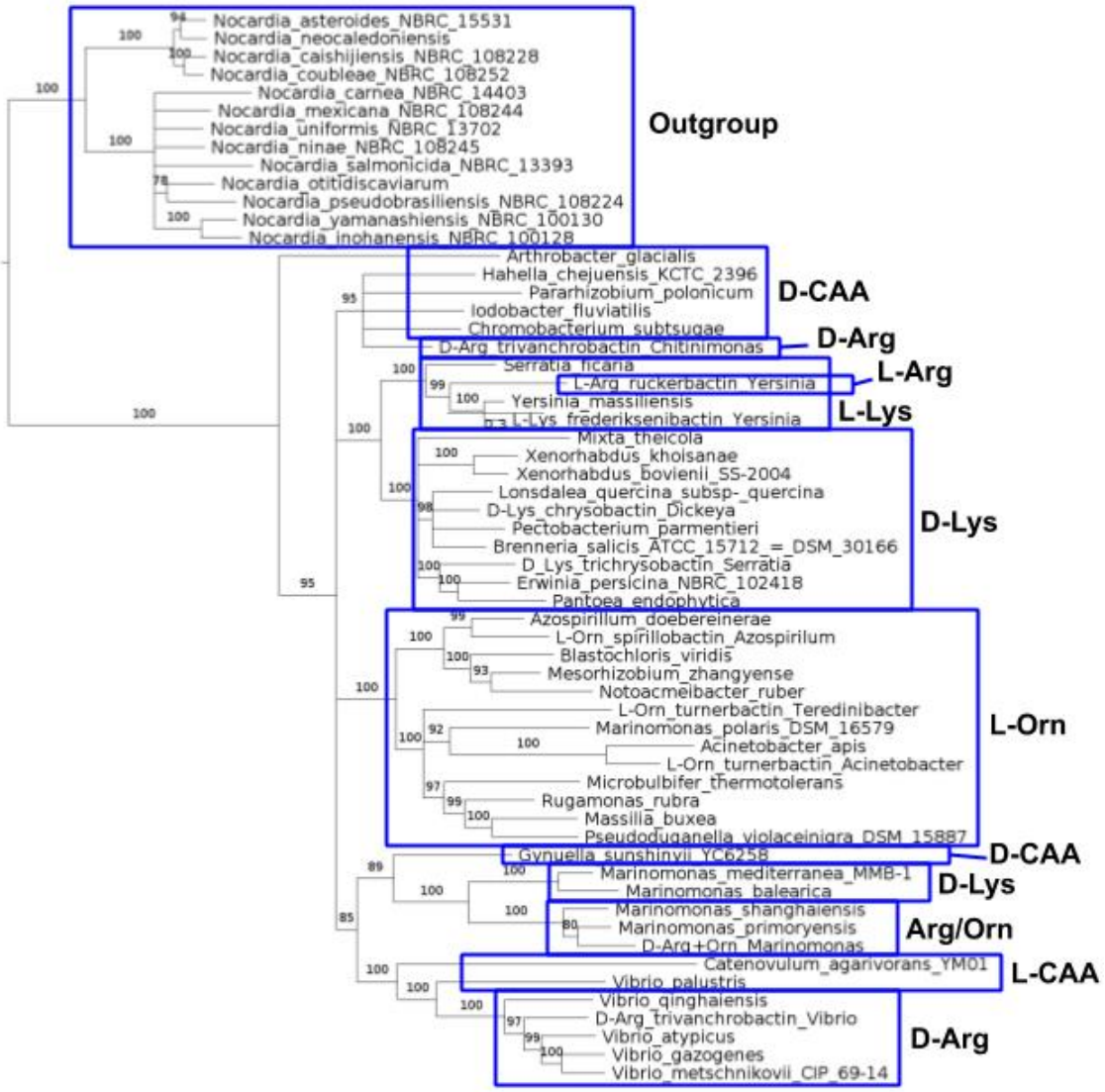


Figure 4.8 Reconstructed phylogeny of DHB-CAA-Ser siderophore NRPS amino acid sequences. Taxa are boxed based on the predicted spacer amino acid identity and chirality. Taxa in bold produce characterized siderophores. After excluding the recombinant NRPSs from *Methylosinus sporium* and *Methylocystis* spp., NRPS sequences from 118 CatSCAn genomes and 10 genomes with known siderophores were sorted into non-redundant clusters with CD-HIT-2D[6] at a 70% identity threshold, resulting in 55 representative strains. A maximum-likelihood phylogenetic tree was reconstructed from the aligned sequences using the LG+I+F+G4 model. Branch support was assessed by bootstrapping (100 bootstrap replicates), and branches with <75% support were collapsed to polytomies.

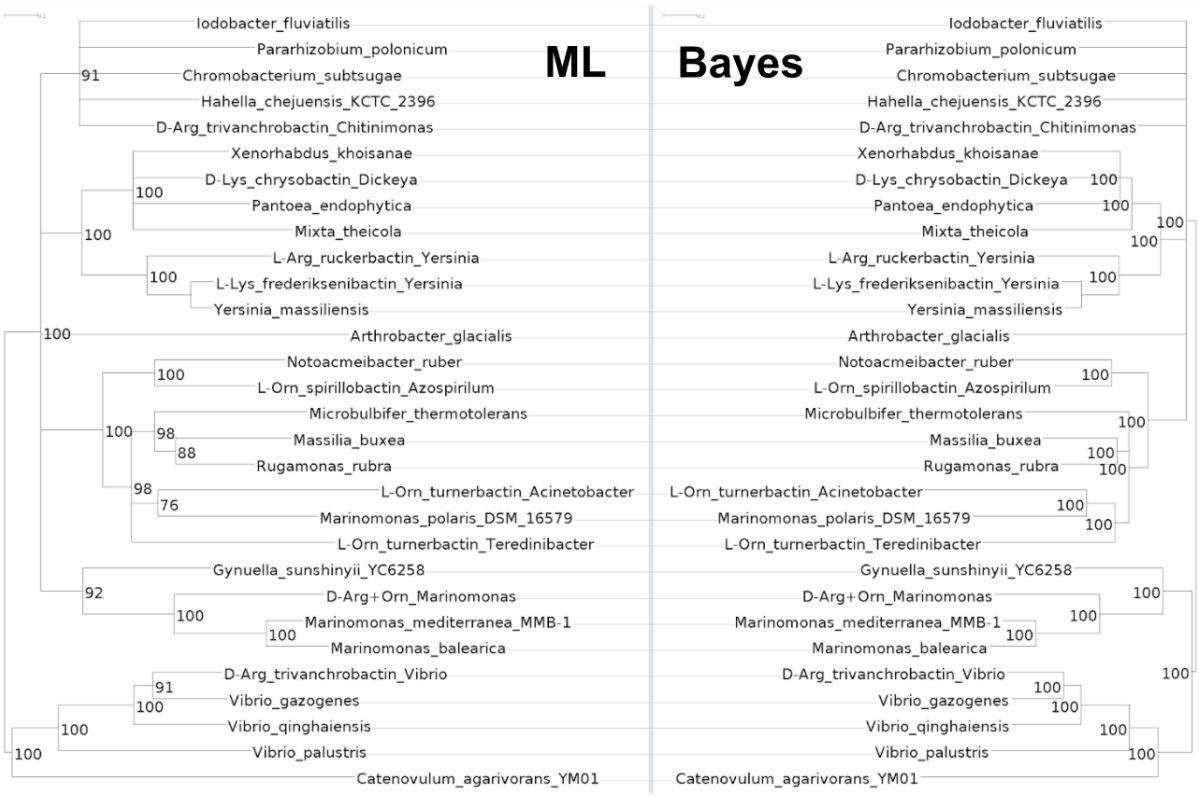


Figure 4.9 Tanglegram of DHB-CAA-Ser siderophore NRPS amino acid sequence phylogenies reconstructed using maximum likelihood (ML) and Bayesian (Bayes) methods. The two methods give consistent topologies. After excluding the recombinant NRPSs from *Methylosinus sporium* and *Methylocystis* spp., NRPS sequences from 118 CatSCAn genomes and 10 genomes with known siderophores were sorted into non-redundant clusters with CD-HIT[6] at a 50% identity threshold, resulting in 30 representative strains. NRPS sequences were aligned with MUSCLE,[4] and trimmed with Gblocks,[7] then the epimerization domain region was removed. The scale bar indicates the average number of substitutions per site. The tanglegram was created with Dendroscope.[12] **Left:** A maximum-likelihood phylogenetic tree was reconstructed from the aligned sequences with PhyML[8] (NGPhylogeny.fr[9] webserver) using the LG+I+F+G4 model, as chosen by ModelFinder (Bayesian information criterion).[10] Branch support was assessed by bootstrapping (100 bootstrap replicates), and branches with <75% support were collapsed to polytomies. **Right:** A Bayesian phylogenetic tree was reconstructed from the aligned sequences with MrBayes (NGPhylogeny.fr[9] webserver; 10^5 generations, 4 chains, invgamma rates). Branches with <90% posterior probability were collapsed to polytomies.

NRPS genes with no epimerization domain each contain a thiolation domain with a GGDSI (T_E) motif and a $^{D}C_L$-type condensation domain. Both of these domain subtypes are generally associated with epimerization domains and provide evidence of an excised E domain (See Chapter 1).[26,28,30] Collapsing the ML tree based on the absence or presence of an epimerization domain (Figure 4.10) shows four separate domain deletion events in the DHB-CAA-Ser siderophore family. E domain deletion has also been reported in β-OHAsp siderophores and glycopeptide antibiotics;[27,56] similar phylogenetic analyses in those families may determine if the events similarly reoccurred.

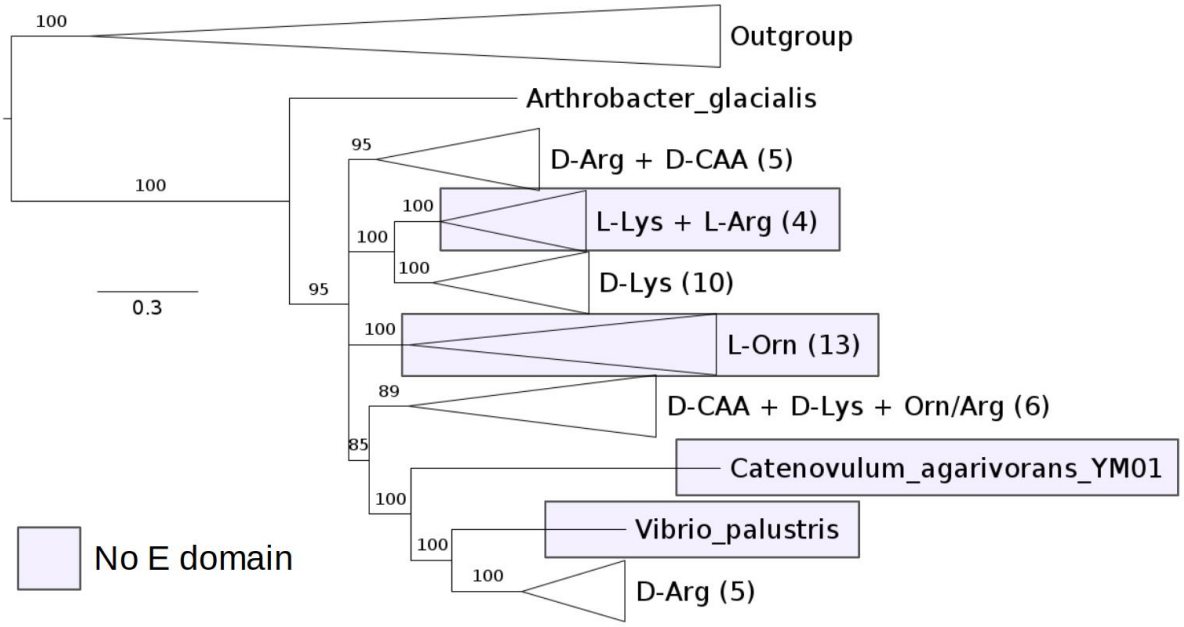


Figure 4.10. Evolution of spacer amino acid chirality in the DHB-CAA-Ser siderophore family. The phylogenetic tree in Figure 4.8 was simplified by collapsing clades where all members share the same predicted spacer chirality. Taxa highlighted blue lack an epimerization domain in the NRPS, and are predicted to produce siderophores with L- amino acid spacers.

The whole-NRPS phylogeny in Figure 4.8 was simplified by collapsing clades with the same CAA identity (*i.e.* Arg, Lys, or Orn) as observed experimentally or predicted by CatSCAn (Figure 4.11). Each clade was "anchored" by the existence of at least one experimentally characterized siderophore, with the exception of a small clade predicted to produce D-Lys-containing siderophores. All putative turnerbactin (L-Orn) NRPSs form a single clade. Frederiksenibactin (L-Lys) and chrysobactin (D-Lys) NRPSs are sister to each other, suggesting they both descend from a single change in CAA selectivity. Trivanchrobactin (D-Arg) NRPSs from *C. koreensis* DSM 17726 and *Vibrio* spp. do not form a clade.

The moabactin (mixed Orn/Arg) family of siderophores of *Marinomonas* sp. TW1 contains a mixture of D- and L-Orn, but only D-Arg (not L-Arg) (Chapter 3). Furthermore, amino acid supplementation resulted in trimeric DHB-Arg-Ser formation, but no Orn-containing trimer was observed (Chapter 3). Together, these suggest that the ancestor of the *Marinomonas* sp. TW1 NRPS was optimized for D-Arg, but has lost selectivity (*vide infra*). D-Arg selectivity is the most dispersed in the phylogenetic tree (Figure 4.8); thus, the most parsimonious evolutionary history of DHB-CAA-Ser siderophore biosynthesis includes a shared ancestor that was selective for D-Arg (*i.e.* a vanchrobactin producer). Surprisingly, ruckerbactin (L-Arg) biosynthesis did not evolve directly from trivanchrobactin (D-Arg) biosynthesis by loss of an E domain. Instead, the ruckerbactin NRPS lies within the frederiksenibactin (L-Lys) clade.

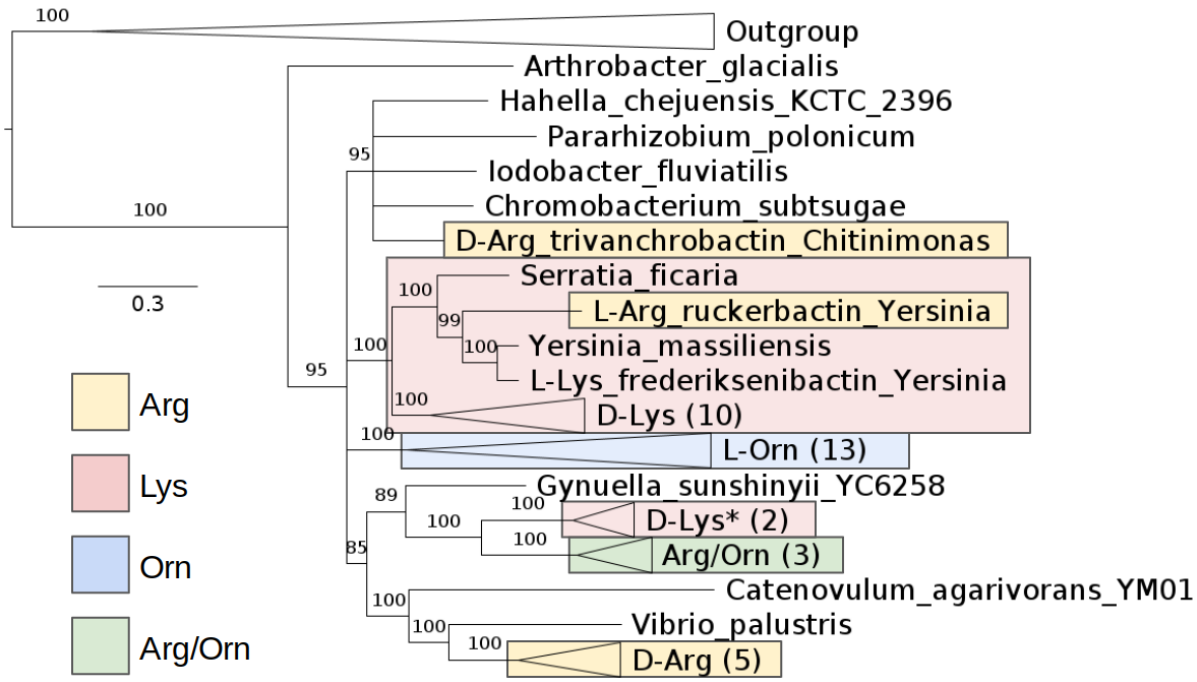


Figure 4.11. Evolution of spacer amino acid identity in the DHB-CAA-Ser siderophore family. The phylogenetic tree in Figure 4.8 was simplified by collapsing clades where all members share the same observed or predicted CAA identity and chirality. Taxa are highlighted based on the predicted spacer amino acid identity: yellow, Arg; red, Lys; blue, Orn; and green, mixed Orn/Arg. Uncolored clades did not have an amino acid prediction. The D-Lys clade marked with an asterisk does not have a characterized representative.

4.3.3.3. Enzyme substrate promiscuity in moabactin biosynthesis

The genome of *Marinomonas* sp. TW1 contains a single NRPS gene, *mmbF*. In low-iron medium, *Marinomonas* sp. TW1 was found to produce five different catechol compounds: DHB-Orn-Ser and $(DHB\text{-}Orn\text{-}Ser)_2$ (also referred to as the Orn monomer and dimer), DHB-Arg-Ser and $(DHB\text{-}Arg\text{-}Ser)_2$ (also referred to as the Arg monomer and dimer), and the mixed Orn/Arg dimer [(DHB-Orn-Ser)(DHB-Arg-Ser)], named moabactin. Furthermore, the Orn-containing compounds were found to exist as a mixture of D- and L- Orn.

This remarkable mixture of siderophores requires not only relaxed amino acid specificity for Arg and Orn in the first A domain of MmbF, but also flexibility in the downstream E domain, $^{D}C_{L}$ domain, and Te domain, all of which have been observed to act as secondary gatekeepers to ensure the correct natural product, with correct stereochemistry, is produced by the NRPS.[57-59] The A and E domains are evidently non-specific for amino acid identity, as both Arg and Orn are observed, although the final product ratio may be influenced by differential rates of Arg/Orn adenylation and epimerization. In contrast, the Te domain and $^{D}C_{L}$ domain both appear to have retained some selectivity, limiting the siderophores observed.

The Te domain is responsible for the esterification of DHB-CAA-Ser to produce dimers and trimers, but the factors governing the extent of oligomerization and cyclization are currently unknown. *Marinomonas* sp. TW1 naturally produces the Arg monomer and dimer, the Orn monomer and dimer, and the mixed Arg/Orn dimer moabactin. Upon the addition of 20 mM Arg to the growth medium, the Arg trimer (trivanchrobactin) was produced, but no Orn trimer was detected during Orn supplementation. The thioesterase of MmbF appears to tolerate Orn substrates for dimerization, but not trimerization. *In vitro* studies suggest

thioesterase promiscuity varies across biosyntheses;[60-62] other triscatechol siderophore thioesterases may be more permissive to spacer amino acid substitutions.

Epimerization domains produce an equilibrium mixture of D- and L- amino acids. The strict D- stereospecificity seen in most NRPSs with E domains comes from the $^{D}C_{L}$ domain, a clade of C domains which only accepts D- amino acids into the upstream donor active site. Some domains within the $^{D}C_{L}$ clade are capable of forming peptide bonds between two L- amino acids, as seen in triscatechol NRPSs TnbF, RucF, and FreF (of turnerbactin, ruckerbactin, and frederiksenibactin biosyntheses, respectively) and in the biosynthesis of glycopeptide antibiotics.[27] The $^{D}C_{L}$ domain of *Marinomonas* NRPS MmbF faithfully selects D-Arg over L-, but does not fully exclude L-Orn from the active site (Figure 4.12A). Similarly variable stereoselectivity was observed during *in vitro* studies of the C_5 domain of tyrocidine biosynthesis. Domain C_5 forms a peptide bond between the T-bound substrates (D-Phe-L-Pro-L-Phe)-D-Phe and L-Asn (Figure 4.12B). The pentapeptide is only formed when the final Phe residue of the tetrapeptide is in the D- configuration.[63] In contrast, when the upstream T domain is loaded with only Phe, dipeptides L-Phe-L-Asn and D-Phe-L-Asn are produced in a 2:1 mixture.[58] The aminoacyl Phe substrate has more flexibility in the binding pocket than the natural tetrapeptidyl Phe substrate, allowing both L- and D-Phe to be condensed to Asn.[63]

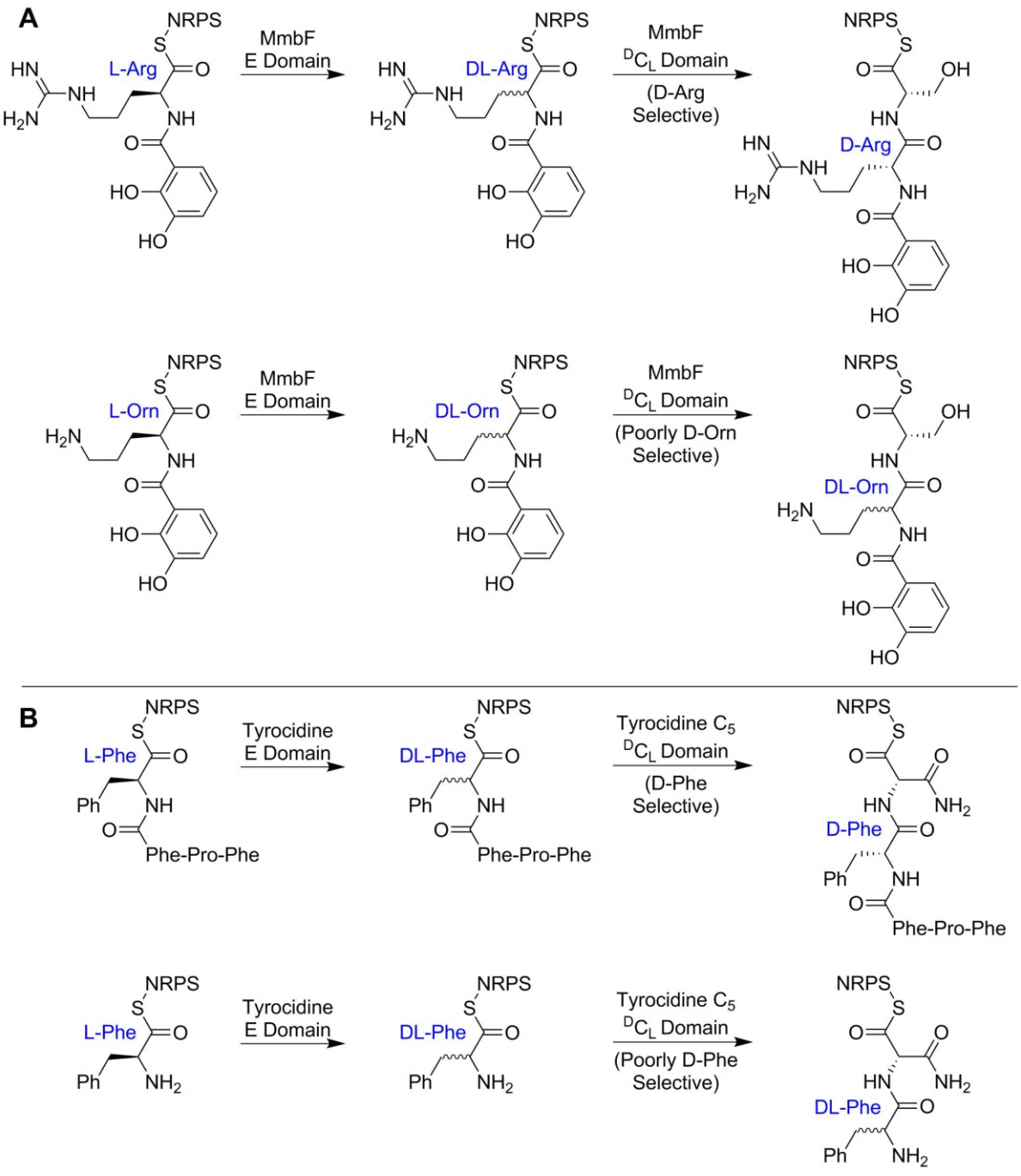


Figure 4.12. Variable stereoselectivity of $^{D}C_L$ domains involved in moabactin and tyrocidine biosyntheses. **(A)** The $^{D}C_L$ domain of MmbF is selective for D-Arg over L-Arg, but not D-Orn over L-Orn. **(B)** Similarly, domain C_5 of tyrocidine biosynthesis is selective for D-Phe in the case of the native tetrapeptide, but not in the case of the single amino acid.

4.3.3.4. A proposed model of NRPS evolution through enzyme promiscuity

The MmbF $^{D}C_{L}$ domain is only able to enforce a strict D- stereochemistry for Arg, and trimerization by the Te domain is only observed for the D-Arg substrate. These observations support the whole-NRPS phylogeny hypothesis (Section 4.3.3.2) that the ancestor of MmbF produced trivanchrobactin before the first A domain evolved to accept both Orn and Arg (Figure 4.13A). MmbF produces siderophores with three observed spacer CAAs (*i.e.* D-Arg, D-Orn, and L-Orn). One may envision pathways by which MmbF continues to evolve, becoming more specific for any single CAA as A domain and/or $^{D}C_{L}$ domain selectivities shift (Figure 4.13A and B).

Alternatively, a sudden deletion of the MmbF epimerization domain should result in the sole production of the turnerbactin (L-Orn) family (Figure 4.14A). With no mechanism to produce D- amino acids, L-Orn would be the only substrate compatible with the $^{D}C_{L}$ domain. Adenylation and thiolation of L-Arg would be a dead end, and the first A domain would be under selective pressure to disfavor the L-Arg substrate. In this scenario, true stereospecificity in the $^{D}C_{L}$ domain would not be necessary. Indeed, C_{7} of teicoplanin biosynthesis, a member of the $^{D}C_{L}$ clade with $^{L}C_{L}$-like activity, is able to accept a D- amino acid donor *in vitro* with only a 20% loss in yield.[64] TnbF, RucF, or FreF engineered to reintroduce a compatible E domain may still be capable of producing mixtures of the respective D- and L- diastereomers if their $^{D}C_{L}$ domain was non-selective at the point of E domain deletion.

The phylogenies reconstructed herein suggest that the DHB-CAA-Ser siderophore NRPS family has undergone multiple changes in adenylation domain selectivity and repeated loss of the epimerization domain. The NRPS MmbF of *Marinomonas* sp. TW1, which incorporates D-Arg, D-Orn, and L-Orn into the final natural products, may be a snapshot of

NRPS evolution in progress. Promiscuous enzymes with multiple products serve as starting points for functional divergence, allowing for an exploration of the fitness landscape without losing the central function (in this case, siderophore activity).[53,65,66] Particularly, stereochemical promiscuity in the $^{D}C_{L}$ domain may be required before E domain deletion can occur; deletion in a strictly D- amino acid selective module would likely be a deleterious mutation that halts all siderophore formation (Figure 4.14B). In *Marinomonas* sp. TW1, relaxation of A_1 domain specificity has allowed for L-Orn to bypass the $^{D}C_{L}$ domain (Figure 4.14A). NRPSs for turnerbactin (L-Orn) and frederiksenibactin (L-Lys) may have both evolved through an MmbF-like intermediate. Intriguingly, the ruckerbactin (L-Arg) NRPS is not directly descended from the D-Arg ancestral NRPS in the phylogenetic tree, but sits within the frederiksenibactin clade. A potential evolutionary history of frederiksenibactin, ruckerbactin, and cyclic trichrysobactin based on their phylogenetic relationship is depicted in Figure 4.15. The structural diversity of DHB-CAA-Ser siderophores is borne of a relatively simple two-module NRPS; studying their evolutionary history may provide broader insights into how siderophores and other non-ribosomal peptides evolve.

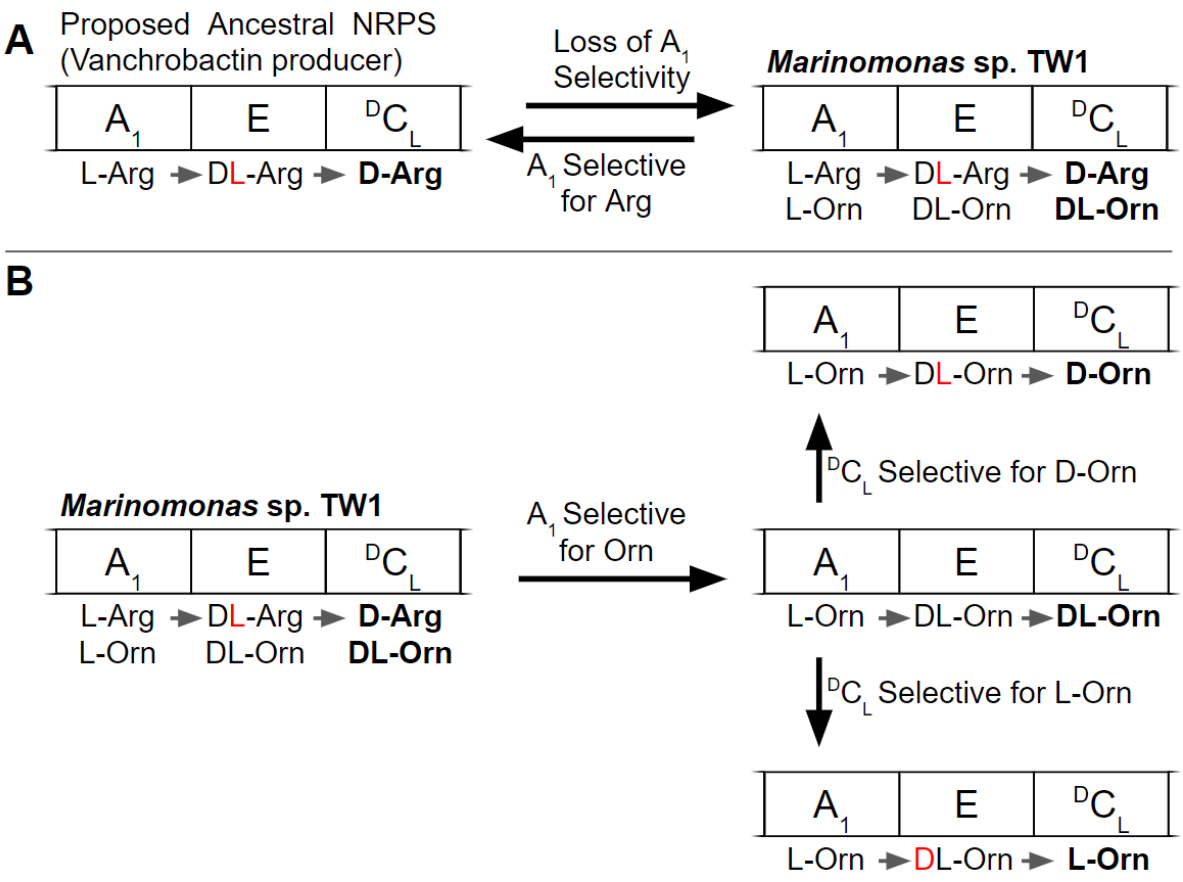


Figure 4.13. Hypothetical pathways for the evolution of amino acid identity and stereochemistry in DHB-CAA-Ser siderophores. Only the first adenylation (A_1) domain, epimerization (E) domain, and $^{D}C_{L}$-type condensation domain are shown. Amino acids are first selected by the A_1 domain (in the L- configuration), condensed to DHB, then epimerized to a DL- mixture by the E domain.[58] The $^{D}C_{L}$ domain controls which stereoisomer(s) will be condensed to the downstream L-Ser. Species excluded from entering the $^{D}C_{L}$ domain are shown in red and species incorporated into the siderophore are shown in bold. **(A)** The whole-NRPS phylogeny supports an ancestral D-Arg-selective NRPS (Section 4.3.4.2). MmbF (*Marinomonas* sp. TW1) contains an A_1 domain with relaxed selectivity to allow both Arg and Orn adenylation. The $^{D}C_{L}$ domain maintains D- selectivity only for Arg. A reinforcement of Arg selectivity in A_1 would produce vanchrobactin. **(B)** Selective pressures could shift A_1 specificity in MmbF to Orn, resulting in a mixture of solely D- and L-Orn siderophores. The $^{D}C_{L}$ domain could regain traditional $^{D}C_{L}$ activity to produce the (hereto unknown) D-Orn siderophore, or could instead become selective for L-Orn, resulting in turnerbactin production.

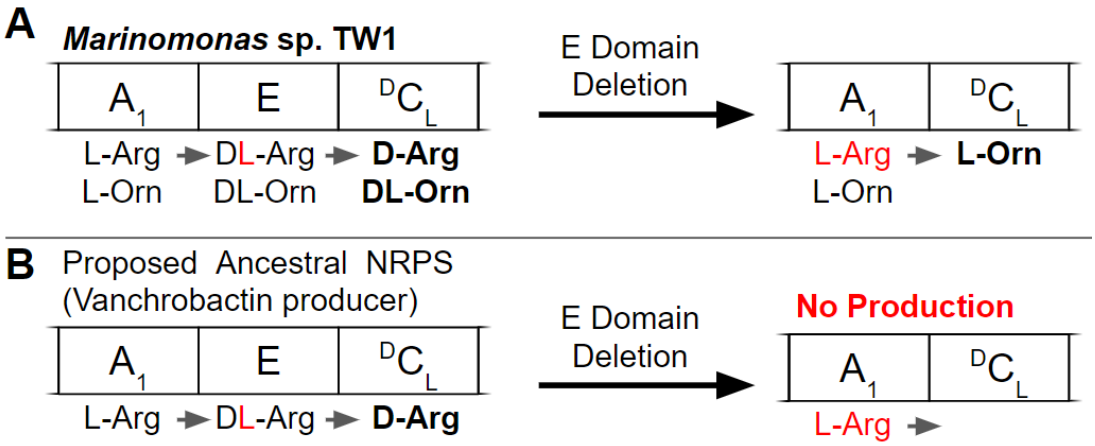


Figure 4.14. Hypothetical E domain deletion events in DHB-CAA-Ser siderophores. Only the first adenylation (A_1) domain, epimerization (E) domain, and $^{D}C_{L}$-type condensation domain are shown. Amino acids are first selected by the A_1 domain (in the L- configuration), condensed to DHB, then epimerized to a DL- mixture by the E domain.[58] The $^{D}C_{L}$ domain controls which stereoisomer(s) will be condensed to the downstream L-Ser. Species excluded from entering the $^{D}C_{L}$ domain are shown in red and species incorporated into the siderophore are shown in bold. **(A)** A deletion of the E domain in MmbF would eliminate D-amino acid production. L-Arg would not be able to enter the active site of the $^{D}C_{L}$ domain; however, turnerbactin (L-Orn) could still be produced due to the flexibility of the $^{D}C_{L}$ domain. **(B)** If the vanchrobactin NRPS $^{D}C_{L}$ domain strictly enforces D- stereochemistry, a sudden deletion of the E domain would stop siderophore production, as no L-Arg could access the binding pocket of the $^{D}C_{L}$ domain.

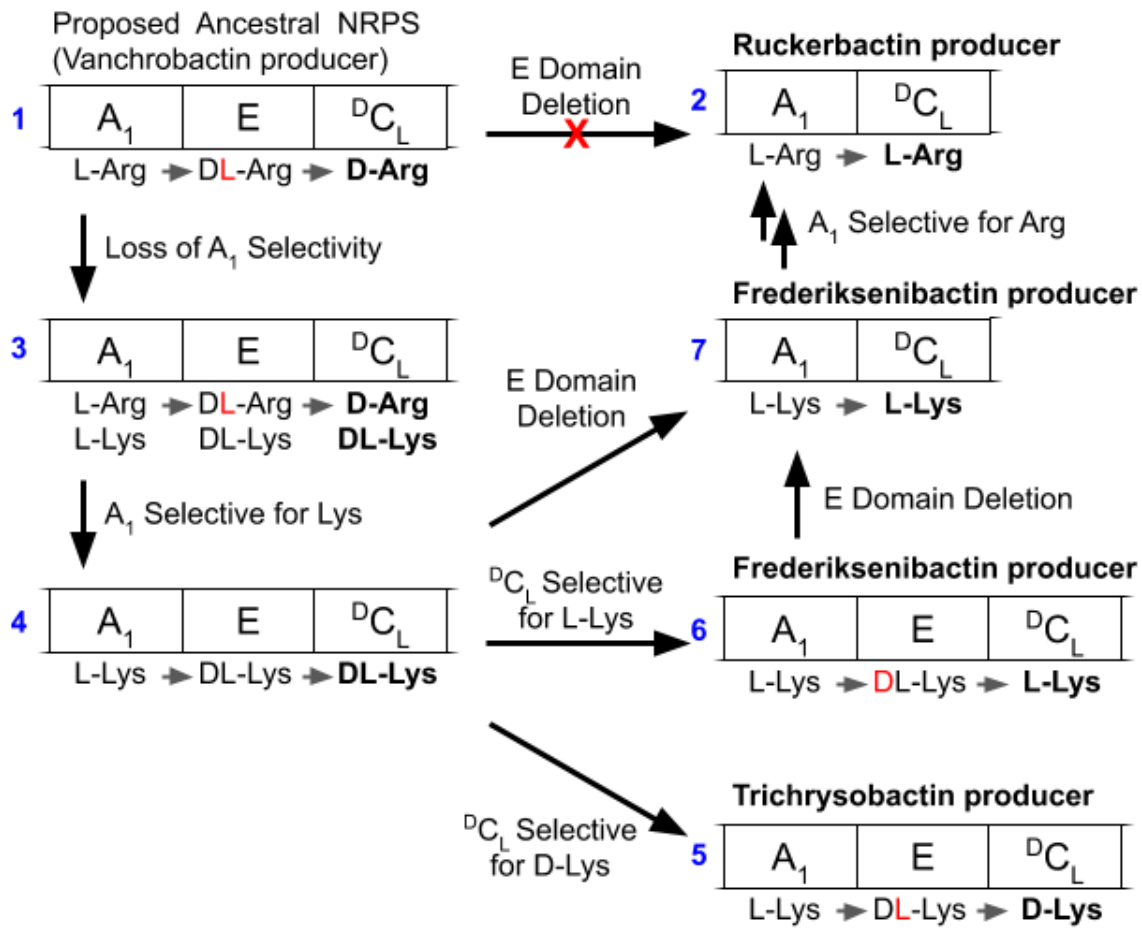


Figure 4.15. A proposed evolutionary history of NRPSs responsible for ruckerbactin (L-Arg), cyclic trichrysobactin (D-Lys), and frederiksenibactin (L-Lys). The whole-NRPS phylogeny supports an ancestral D-Arg selective NRPS (**1**) (Section 4.3.4.2), but not the direct evolution of L-Arg selectivity (**2**) by E domain deletion. Instead, the A_1 domain selectivity relaxed to include Lys, giving a promiscuous NRPS (**3**) similar to MmbF. Strict selection of Lys by A_1 would then result in a mixture of D- and L- Lys incorporation (**4**). The $^{D}C_L$ domain of NRPS **4** could regain traditional $^{D}C_L$ activity to produce cyclic trichrysobactin (**5**), or could instead become selective for L-Lys, resulting in frederiksenibactin production (**6**). The modern-day frederiksenibactin NRPS architecture (**7**) lacks the E domain, which could have been deleted from NRPS **6** or directly from NRPS **4**. Finally, another shift in A_1 selectivity from Lys to Arg resulted in the production of ruckerbactin (**2**). Other possible transformations include single-mutation changes in A_1 selectivity (*i.e.*, from NRPSs **1** to **4** or NRPSs **7** to **2**) that bypass the promiscuous intermediates.

4.4. References

1. S. R. Eddy, Accelerated Profile HMM Searches, *PLoS Comput. Biol.*, 2011, **7**, e1002195.
2. S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez and R. D. Finn, HMMER web server: 2018 update, *Nucleic Acids Res.*, 2018, **46**, W200-W204.
3. M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T. L. Madden, NCBI BLAST: a better web interface, *Nucleic Acids Res.*, 2008, **36**, W5-9.
4. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, 2004, **32**, 1792-1797.
5. M. Gouy, S. Guindon and O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building, *Mol. Biol. Evol.*, 2010, **27**, 221-224.
6. W. Li and A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics*, 2006, **22**, 1658-1659.
7. G. Talavera and J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, *Syst. Biol.*, 2007, **56**, 564-577.
8. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, *Syst. Biol.*, 2010, **59**, 307-321.
9. F. Lemoine, D. Correia, V. Lefort, O. Doppelt-Azeroual, F. Mareuil, S. Cohen-Boulakia and O. Gascuel, NGPhylogeny.fr: new generation phylogenetic services for non-specialists, *Nucleic Acids Res.*, 2019, **47**, W260-W265.
10. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler and L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, *Nat. Methods*, 2017, **14**, 587-589.
11. F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, *Syst. Biol.*, 2012, **61**, 539-542.
12. D. H. Huson, D. C. Richter, C. Rausch, T. Dezulian, M. Franz and R. Rupp, Dendroscope: An interactive viewer for large phylogenetic trees, *BMC Bioinformatics*, 2007, **8**, 460.
13. C. Scornavacca, F. Zickmann and D. H. Huson, Tanglegrams for rooted phylogenetic trees and networks, *Bioinformatics*, 2011, **27**, i248-256.
14. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy and K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation, *Nucleic Acids Res.*, 2016, **44**, D733-745.

15. F. Bertels, O. K. Silander, M. Pachkov, P. B. Rainey and E. van Nimwegen, Automated reconstruction of whole-genome phylogenies from short-sequence reads, *Mol. Biol. Evol.*, 2014, **31**, 1077-1088.
16. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler and B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, *Mol. Biol. Evol.*, 2015, **32**, 268-274.
17. T. Franza and D. Expert, The virulence-associated chrysobactin iron uptake system of Erwinia chrysanthemi 3937 involves an operon encoding transport and biosynthetic functions, *J. Bacteriol.*, 1991, **173**, 6874-6881.
18. C. W. Dorsey, M. E. Tolmasky, J. H. Crosa and L. A. Actis, Genetic organization of an Acinetobacter baumannii chromosomal region harbouring genes related to siderophore biosynthesis and transport, *Microbiology*, 2003, **149**, 1227-1238.
19. L. Fernández, I. Márquez and J. A. Guijarro, Identification of specific in vivo-induced (ivi) genes in Yersinia ruckeri and analysis of ruckerbactin, a catecholate siderophore iron acquisition system, *Appl. Environ. Microbiol.*, 2004, **70**, 5199-5207.
20. M. Balado, C. R. Osorio and M. L. Lemos, A gene cluster involved in the biosynthesis of vanchrobactin, a chromosome-encoded siderophore produced by Vibrio anguillarum, *Microbiology*, 2006, **152**, 3517-3528.
21. M. Balado, A. Souto, A. Vences, V. P. Careaga, K. Valderrama, Y. Segade, J. Rodríguez, C. R. Osorio, C. Jiménez and M. L. Lemos, Two Catechol Siderophores, Acinetobactin and Amonabactin, Are Simultaneously Produced by Aeromonas salmonicida subsp. salmonicida Sharing Part of the Biosynthetic Pathway, *ACS Chem. Biol.*, 2015, **10**, 2850-2860.
22. H. Naka, Z. L. Reitz, A. L. Jelowicki, A. Butler and M. G. Haygood, Amphi-enterobactin commonly produced among Vibrio campbellii and Vibrio harveyi strains can be taken up by a novel outer membrane protein FapA that also can transport canonical Fe(III)-enterobactin, *J. Biol. Inorg. Chem.*, 2018, **23**, 1009-1022.
23. L. M. Weaver and K. M. Herrmann, Cloning of an aroF allele encoding a tyrosine-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, *J. Bacteriol.*, 1990, **172**, 6581-6584.
24. H. Kloosterman, G. I. Hessels, J. W. Vrijbloed, G. J. Euverink and L. Dijkhuizen, (De)regulation of key enzyme steps in the shikimate pathway and phenylalanine-specific pathway of the actinomycete Amycolatopsis methanolica, *Microbiology*, 2003, **149**, 3321-3330.
25. M. Balado, C. R. Osorio and M. L. Lemos, Biosynthetic and regulatory elements involved in the production of the siderophore vanchrobactin in Vibrio anguillarum, *Microbiology*, 2008, **154**, 1400-1413.
26. U. Linne, S. Doekel and M. A. Marahiel, Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases, *Biochemistry*, 2001, **40**, 15824-15834.
27. C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution, *BMC Evol. Biol.*, 2007, **7**, 78.
28. A. W. Han, PhD Ph.D. Thesis, Oregon Health & Science University, 2011.
29. A. W. Han, M. Sandy, B. Fishman, A. E. Trindade-Silva, C. A. G. Soares, D. L. Distel, A. Butler and M. G. Haygood, Turnerbactin, a novel triscatecholate

siderophore from the shipworm endosymbiont Teredinibacter turnerae T7901, *PLoS One*, 2013, **8**, e76151.
30. Z. L. Reitz, M. Sandy and A. Butler, Biosynthetic considerations of triscatechol siderophores framed on serine and threonine macrolactone scaffolds, *Metallomics*, 2017, **9**, 824-839.
31. M. M. Nakano, N. Corbell, J. Besson and P. Zuber, Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in Bacillus subtilis, *Mol. Gen. Genet.*, 1992, **232**, 313-321.
32. R. A. Schomer and M. G. Thomas, Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase, *Biochemistry*, 2017, **56**, 5380-5390.
33. J. L. Furrer, D. N. Sanders, I. G. Hook-Barnard and M. A. McIntosh, Export of the siderophore enterobactin in Escherichia coli: involvement of a 43 kDa membrane exporter, *Mol. Microbiol.*, 2002, **44**, 1225-1234.
34. T. Horiyama and K. Nishino, AcrB, AcrD, and MdtABC multidrug efflux systems are involved in enterobactin export in Escherichia coli, *PLoS One*, 2014, **9**, e108642.
35. J. Anes, M. P. McCusker, S. Fanning and M. Martins, The ins and outs of RND efflux pumps in Escherichia coli, *Front. Microbiol.*, 2015, **6**, 587.
36. W. Köster, ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12, *Res. Microbiol.*, 2001, **152**, 291-301.
37. R. Tam and M. H. Saier, Jr., Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria, *Microbiol. Rev.*, 1993, **57**, 320-346.
38. H. Lin, M. A. Fischbach, D. R. Liu and C. T. Walsh, In vitro characterization of salmochelin and enterobactin trilactone hydrolases IroD, IroE, and Fes, *J. Am. Chem. Soc.*, 2005, **127**, 11075-11084.
39. M. Miethke, J. Hou and M. A. Marahiel, The siderophore-interacting protein YqjH acts as a ferric reductase in different iron assimilation pathways of Escherichia coli, *Biochemistry*, 2011, **50**, 10951-10964.
40. L. Rauscher, D. Expert, B. F. Matzanke and A. X. Trautwein, Chrysobactin-dependent Iron Acquisition inErwinia chrysanthemi : FUNCTIONAL STUDY OF A HOMOLOG OF THE ESCHERICHIA COLIFERRIC ENTEROBACTIN ESTERASE, *J. Biol. Chem.*, 2002, **277**, 2385-2395.
41. H. P. Benson, E. Boncompagni and M. L. Guerinot, An iron uptake operon required for proper nodule development in the Bradyrhizobium japonicum-soybean symbiosis, *Mol. Plant. Microbe. Interact.*, 2005, **18**, 950-959.
42. P. Ó Cuív, D. Keogh, P. Clarke and M. O'Connell, FoxB of Pseudomonas aeruginosa Functions in the Utilization of the Xenosiderophores Ferrichrome, Ferrioxamine B, and Schizokinen: Evidence for Transport Redundancy at the Inner Membrane, *J. Bacteriol.*, 2007, **189**, 284-287.
43. E. D. Peng and S. M. Payne, Vibrio cholerae VciB Mediates Iron Reduction, *J. Bacteriol.*, 2017, **199**.
44. L. Käll, A. Krogh and E. L. L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server, *Nucleic Acids Res.*, 2007, **35**, W429-432.
45. J. Yang, I. Sangwan and M. R. O'Brian, The Bradyrhizobium japonicum Fur protein is an iron-responsive regulator in vivo, *Mol. Genet. Genomics*, 2006, **276**, 555-564.

46. M. F. Fillat, The FUR (ferric uptake regulator) superfamily: diversity and versatility of key transcriptional regulators, *Arch. Biochem. Biophys.*, 2014, **546**, 41-52.
47. G. Kumar, K. Hummel, M. Ahrens, S. Menanteau-Ledouble, T. J. Welch, M. Eisenacher, E. Razzazi-Fazeli and M. El-Matbouli, Shotgun proteomic analysis of Yersinia ruckeri strains under normal and iron-limited conditions, *Vet. Res.*, 2016, **47**, 100.
48. U. Jenal and M. Y. Galperin, Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics, *Curr. Opin. Microbiol.*, 2009, **12**, 152-160.
49. I. A. Suvorova, Y. D. Korostelev and M. S. Gelfand, GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution, *PLoS One*, 2015, **10**, e0132618.
50. T. Tanabe, A. Isshiki, K. Miyamoto, H. Tsujibo, S. Yamamoto and T. Funahashi, Transcriptional regulation of the ferric aerobactin receptor gene by a GntR-like repressor IutR in Vibrio furnissii, *FEMS Microbiol. Lett.*, 2018, **365**.
51. M. Balado, C. R. Osorio and M. L. Lemos, FvtA is the receptor for the siderophore vanchrobactin in Vibrio anguillarum: utility as a route of entry for vanchrobactin analogues, *Appl. Environ. Microbiol.*, 2009, **75**, 2775-2783.
52. S. J. Stuart, J. K. Prpic and R. M. Robins-Browne, Production of aerobactin by some species of the genus Yersinia, *J. Bacteriol.*, 1986, **166**, 1131-1133.
53. M. H. Medema, P. Cimermancic, A. Sali, E. Takano and M. A. Fischbach, A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis, *PLoS Comput. Biol.*, 2014, **10**, e1004016.
54. R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate and A. Bateman, The Pfam protein families database: towards a more sustainable future, *Nucleic Acids Res.*, 2016, **44**, D279-285.
55. D. Posada and K. A. Crandall, The effect of recombination on the accuracy of phylogeny estimation, *J. Mol. Evol.*, 2002, **54**, 396-402.
56. Z. L. Reitz, C. D. Hardy, J. Suk, J. Bouvet and A. Butler, Genomic analysis of siderophore β-hydroxylases reveals divergent stereocontrol and expands the condensation domain family, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 19805-19814.
57. L. Luo, M. D. Burkart, T. Stachelhaus and C. T. Walsh, Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase, *J. Am. Chem. Soc.*, 2001, **123**, 11208-11218.
58. U. Linne and M. A. Marahiel, Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization, *Biochemistry*, 2000, **39**, 10439-10447.
59. M. Peschke, C. Brieke, M. Heimes and M. J. Cryle, The Thioesterase Domain in Glycopeptide Antibiotic Biosynthesis Is Selective for Cross-Linked Aglycones, *ACS Chem. Biol.*, 2018, **13**, 110-120.
60. R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel and C. T. Walsh, Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases, *Biochemistry*, 2001, **40**, 7099-7108.

61. C. C. Tseng, S. D. Bruner, R. M. Kohli, M. A. Marahiel, C. T. Walsh and S. A. Sieber, Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase, *Biochemistry*, 2002, **41**, 13350-13359.
62. K. M. Hoyer, C. Mahlert and M. A. Marahiel, The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization, *Chem. Biol.*, 2007, **14**, 13-22.
63. S. L. Clugston, S. A. Sieber, M. A. Marahiel and C. T. Walsh, Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst, *Biochemistry*, 2003, **42**, 12095-12104.
64. M. Schoppet, M. Peschke, A. Kirchberg, V. Wiebach, R. D. Süssmuth, E. Stegmann and M. J. Cryle, The biosynthetic implications of late-stage condensation domain selectivity during glycopeptide antibiotic biosynthesis, *Chem. Sci.*, 2019, **10**, 118-133.
65. A. Aharoni, L. Gaidukov, O. Khersonsky, S. McQ Gould, C. Roodveldt and D. S. Tawfik, The 'evolvability' of promiscuous protein functions, *Nat. Genet.*, 2005, **37**, 73-76.
66. M. G. Chevrette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson and F. Barona-Gómez, Evolutionary dynamics of natural product biosynthesis in bacteria, *Nat. Prod. Rep.*, 2020, **37**, 566-599.

Chapter 5. Polyamine catechol siderophores of Acinetobacter bouvetii DSM 14964

This chapter was reproduced from Reference 1 with permission from The Royal Society of Chemistry.

5.1. Introduction

Many bacteria acquire iron from their environment by producing small-molecule, high-affinity iron(III) chelators called siderophores. Siderophores containing the catechol 2,3-dihydroxybenzoic acid, a common ligand in siderophores, are synthesized by nonribosomal peptide synthetases (NRPSs), large multi-domain enzymes that function in an assembly-line manner. 2,3-DHB is synthesized in three steps from chorismate, activated by a 2,3-DHB–AMP ligase, and transferred onto an aryl carrier protein (ArCP).[2] A condensation domain then catalyzes amide bond formation between 2,3-DHB and an amine, most often a NRPS-bound amino acid. VibH of vibriobactin (Figure 5.1) biosynthesis is an unusual standalone condensation domain that condenses ArCP-bound 2,3-DHB and norspermidine, forming the amide linkage.[3,4] Several other siderophore biosyntheses feature VibH-like enzymes putatively responsible for the condensation of 2,3-DHB to diamines and polyamines, including brucebactin (Figure 5.1), fluvibactin, and agrobactin, among others.[5-7]

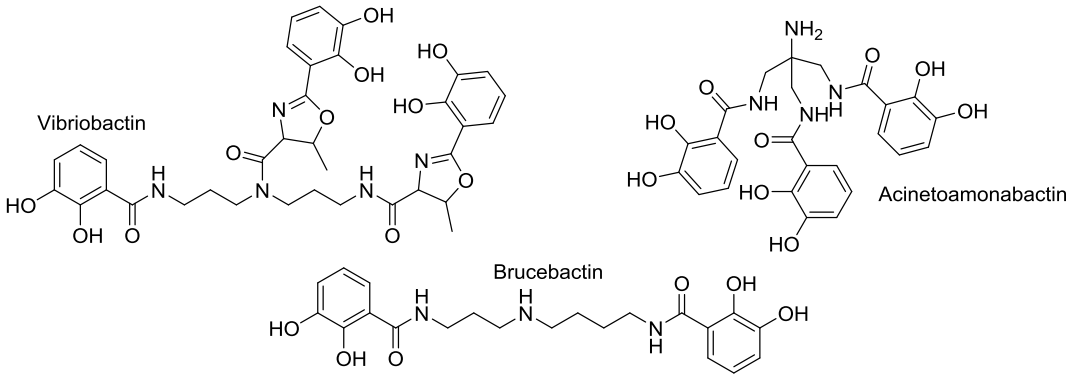


Figure 5.1. Structures of vibriobactin (based on norspermidine), brucebactin (based on spermidine), and acinetoamonabactin (based on tris(aminomethyl)-methylamine).

Precursor directed biosynthesis (PDB) was one of the first methods used to generate novel NRPS products.[8] Non-natural substrates, when added to the growth medium of a bacterial culture, are incorporated into the final molecule by the native biosynthetic pathway.[9-11] PDB can access novel secondary metabolites while bypassing the complexities of synthetic chemistry or synthetic biology. Cleto and Lu combined PDB and heterologous expression to generate a variety of analogs, demonstrating the versatility of VibH by producing several non-natural 2,3-DHB-diamine and 2,3-DHB-polyamine conjugates.[12] However, the reactivity of VibH is limited to installing a single DHB moiety and no bis-catechol product was reported.[3,12] In contrast, the VibH homolog involved in the biosynthesis of brucebactin (Figure 5.1) twice adds 2,3-DHB to spermidine.[7] Similarly the unsequenced strain *Acinetobacter soli* MTCC 5918, which produces acinetoamonabactin (Figure 5.1) may also possess a homolog of VibH capable of condensing three 2,3-DHB moieties to tris(aminomethyl)methylamine, a new polyamine scaffold in siderophores.[13] Catecholic compounds have been shown to adhere to surfaces in aqueous conditions,[14 15] as well as to promote the adhesion of *E. coli* to titania via the siderophore enterobactin.[16] Thus a VibH homolog capable of iterative 2,3-DHB conjugation is an attractive candidate for further PDB investigations directed towards biosynthesis of new catechol materials.

Using VibH as a genomic handle, we identified several *Acinetobacter* species that could potentially produce polyamine-based catechol siderophores similar to acinetoamonabactin (Figure 5.1). We report herein three novel biscatechol siderophores present in the low-iron culture supernatant of *A. bouvetii* DSM 14964, named propanochelin (**1**), butanochelin (**2**), and pentanochelin (**3**). Precursor directed biosynthesis demonstrates

that *A. bouvetii* DSM 14964 can incorporate non-natural amines *in vivo* to produce a variety of siderophore analogs. We found that PDB using *A. bouvetii* DSM 14964 is particularly effective at condensing 2,3-DHB to propargylamine and allylamine, forming catechol products which would be suitable for click-directed chemical modifications in the design of potential new wet adhesive catechol materials.

5.2. Methods

5.2.1. Genome mining

The protein sequence for VibH was retrieved from UniProtKB (A0A0H3AII5). Homologs were found with phmmer (EMBL-EBI web server), targeting the UniProtKB database and restricting taxonomy to *Acinetobacter*. Results were filtered by domain to find standalone condensation domains.

5.2.2. Culture conditions

Acinetobacter bouvetii DSM 14964 was obtained from DSMZ GmbH, Braunschweig, Germany. For siderophore isolation, *A. bouvetii* DSM 14964 was cultured in 1 L casamino acid minimal medium (5 g/L Bacto low-iron casamino acids, 1.54 g/L K_2HPO_4, and 0.25 g/L $MgSO_4$) in a 2L acid-washed Erlenmeyer flask. For the precursor-directed biosynthesis investigations, *A. bouvetii* DSM 14964 was cultured in 100 mL casamino acid medium in a 250 mL acid-washed Erlenmeyer flask; potential substrates were added directly after autoclave sterilization. Cultures were grown on an orbital shaker (180 rpm) at ambient temperature. Growth was monitored by measuring the optical density at 600 nm.

5.2.3. Siderophore isolation

After 72 h, the bacterial cell culture was centrifuged at 10000 rpm for 25 min to pellet the cells. Decanted supernatant was shaken with Amberlite XAD-4 resin (10% v/v) for 3 hours. The resin was isolated by filtration and washed with 5 column volumes nanopure water followed by 5 column volumes 10% v/v MeOH/H_2O. Organic compounds were eluted with 5 column volumes 90% v/v MeOH/H_2O and concentrated *in vacuo*. Siderophores were purified by reverse-phase high-pressure liquid chromatography (RP-HPLC) on a semi-preparative C_4 column using a constant 7 mL/min flow rate and a linear methanol gradient (5 – 60%, 2%/min). Pooled fractions were concentrated *in vacuo*.

5.2.4. Siderophore characterization

Extracts were analyzed through positive ion mode ESI-MS or ESI-MS^E on a Waters Xevo G2-XS QTof coupled to a Waters Acquity H-Class UPLC system. A Waters BEH C18 column was used with a gradient of 0% to 60% acetonitrile/water (both with 0.1% w/v formic acid) over 12 minutes. GC-CI accurate mass data was obtained on a Waters GCT Premier Time of Flight mass spectrometer. Gas chromatography was performed using an Agilent 7890A GC equipped with a J&W Scientific DB-5MS narrow bore column (30m, 0.250mm ID, 0.25μm film thickness). Chemical ionization was achieved using methane reagent gas. Positive mode accurate mass data was calibrated using known fragments of perfluorotributylamine as an internal standard. ESI accurate mass data was obtained on a Waters LCT Premier Time of Flight mass spectrometer. LC-MS grade methanol (Fisher Optima) was used as background eluent and solvent with no additives. Samples were directly infused using an Alliance 2695 Separations Module as solvent pump and autosampler. Positive mode accurate mass data were calibrated using polyethylene glycol or

polyethylene glycol monomethyl ether internal standards as appropriate. NMR spectroscopy was carried out on Varian Unity Inova 500 MHz and 600 MHz spectrometers or a Bruker Avance NEO 500 MHz spectrometer equipped with a Prodigy coldprobe. Chemical shifts were referenced through residual DMSO-d_6 solvent peaks at 2.50 (^{1}H) or 39.51 (^{13}C) ppm.

5.3. Results and interpretation

5.3.1. Identification and annotation of the *Acinetobacter bouvetii* DSM 14964 siderophore biosynthetic gene cluster

VibH homologs were found in *Acinetobacter* genomes by searching the UniProtKB protein database with phmmer.[17] Filtering the results to standalone condensation domains eliminated traditional condensation domains that are generally fused to other NRPS domains. Homologs were found in *A. bouvetii*, *A. larvae*, *A. gandensis*, *A. piscicola*, *A. pragensis*, *A. puyangensis*, and several unspecified strains (Table 5.1). Each VibH homolog is part of an identical gene cluster, which contains genes putatively encoding 2,3-DHB biosynthesis and incorporation enzymes, a TonB-dependent outer membrane receptor protein, a Fe(III)-siderophore reductase, and a PLP-dependent enzyme annotated as a Orn/Lys decarboxylase (Figure 5.2, Table 5.2). Together, these genes are predicted to encode for the biosynthesis and utilization of a catechol siderophore with a diamine backbone.

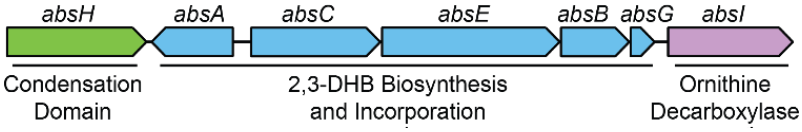
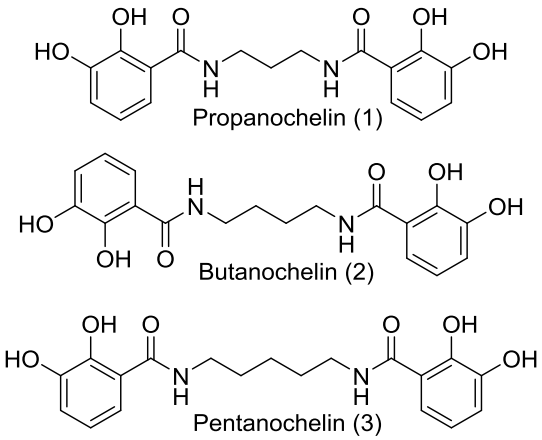


Figure 5.2. The biosynthetic gene cluster in *A. bouvetii* DSM 14964 putatively responsible for siderophore production. The cluster is drawn to scale; arrows represent the direction of transcription. A full description of the gene cluster can be found in Table 5.2.

5.3.2. Structural characterization of propanochelin, butanochelin, and pentanochelin

To identify the natural catechol siderophores putatively produced by the VibH homolog AbsH, *Acinetobacter bouvetii* DSM 14964 was cultured in a low-iron medium (See SI for culture and isolation conditions). The methanolic supernatant extract contained three peaks with a UV-Vis absorbance band near 310 nm and an ESI-MS^E fragment of 137 m/z, each consistent with catechol-containing compounds (Figures 5.3, 5.6, and 5.7). The three peaks had base molecular ions at 347, 361, and 375 *m/z*, respectively, suggesting a suite of compounds each differing by a methylene unit (Figures 5.8-5.10). Each is capable of binding Fe(III), as indicated by ions corresponding to $[L+Fe]^+$ and $[L_2+Fe]^+$ (Figures 5.8-5.10). The observed masses and fragmentation patterns are consistent with biscatechol siderophores based on diaminopropane, putrescine, and cadaverine; herein named propanochelin (**1**), butanochelin (**2**), and pentanochelin (**3**), respectively.



5.3.3. Precursor directed biosynthesis of unnatural siderophores

The flexibility and fidelity of the biosynthesis pathway was investigated by precursor directed biosynthesis (Table 5.3). The ratios of the three natural siderophores could

be shifted by the addition of diaminopropane, putrescine, or cadaverine to the growth medium, although diaminopropane had only a slight effect on production of propanochelin, **1** (Figure 5.3). No 2,3-DHB-diamine single addition product was identified by ESI-MS. In contrast, supplementation with ethylenediamine resulted in only the single 2,3-DHB addition product, while 1,6-diaminohexane resulted in a mixture of single and double 2,3-DHB addition products (Figure 5.3, 5.11, and 5.12). Thus, the biosynthetic pathway can accommodate a wider variety of diamines for the addition of the first 2,3-DHB than the second 2,3-DHB.

Next we investigated whether functionalized amines could be accepted as substrates in the biosynthetic pathway of *A. bouvetii* DSM 14964 (Table 5.3). No condensation products were formed with the amino acids L-diaminobutyric acid, L-ornithine, or L-lysine; however, the addition of L-Orn caused an increase in butanochelin (**2**) production (Figure 5.4). Therefore, the PLP-dependent enzyme AbsI is likely responsible for the decarboxylation of ornithine to putrescine. The polyamine diethylenetriamine is an isostere of cadaverine (1,5-diaminopentane) and was likewise incorporated as the biscatechol product **4** (Figures 5.5 and 5.13-5.18). Norspermidine (1 mM final concentration) and the branched polyamine tris(2-aminoethyl)amine (tren; 1 mM) both inhibited detectible growth of *A. bouvetii* DSM 14964. Histidine (1 mM) was not incorporated, and instead partially inhibited natural siderophore production (Figure 5.19).

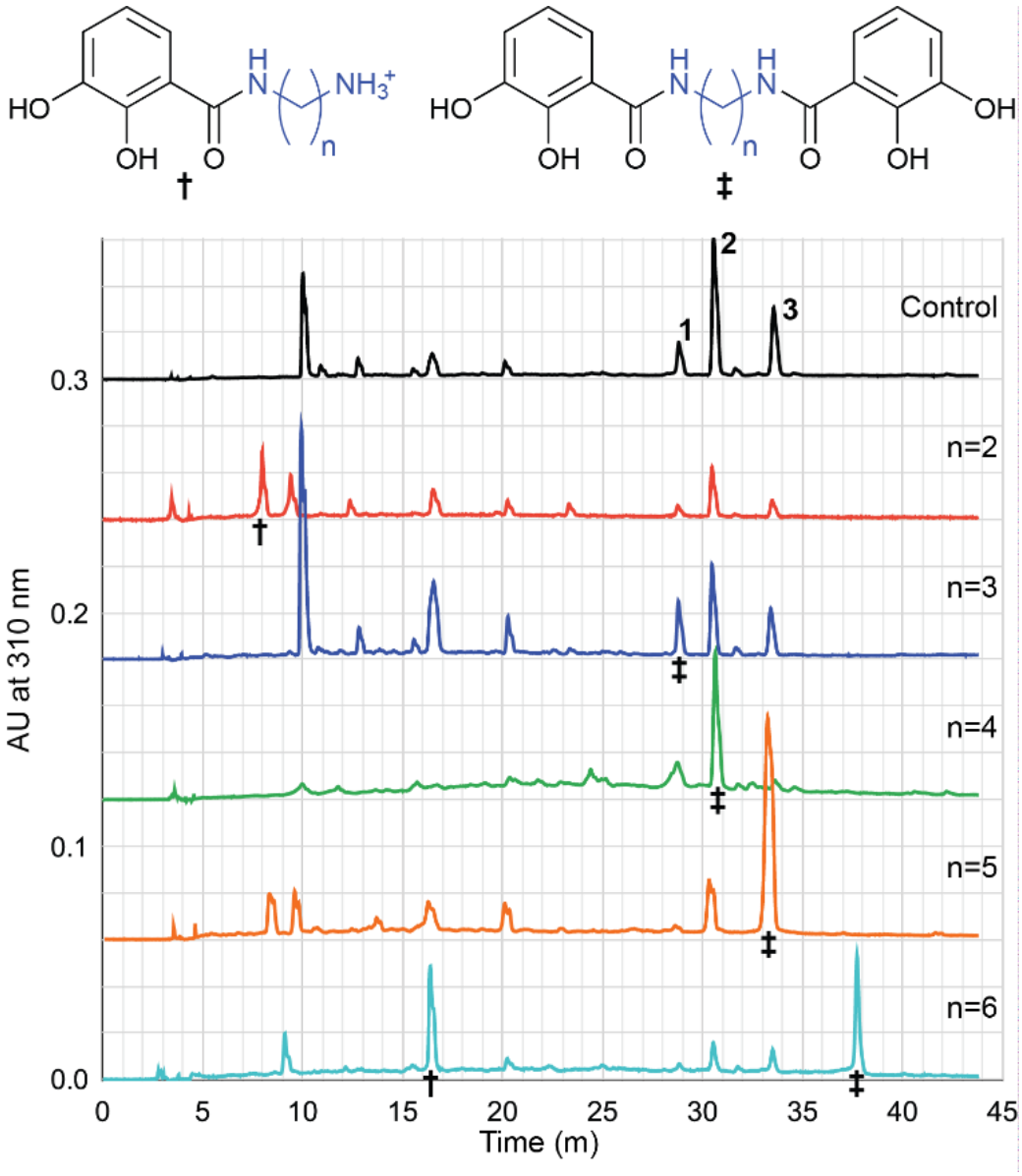


Figure 5.3. Precursor directed biosynthesis of natural and unnatural diamine conjugates in growth of *A. bouvetii* DSM 14964. Peaks corresponding to mono-2,3-DHB (†) and di-2,3-DHB (‡) conjugates are marked. The methanol eluent from the XAD column of 72 h cultures were monitored by HPLC-UV/Vis at 310 nm to detect catechol-containing compounds. Casamino acid minimal medium (control) was supplemented with diamines (1 mM final concentration for each) prior to inoculation. The di-2,3-DHB conjugates for n=3, n=4, and n=5 correspond to the natural siderophores 1, 2, and 3, respectively. Conjugate identities were confirmed by UPLC/MS (Figure 5.8-5.12).

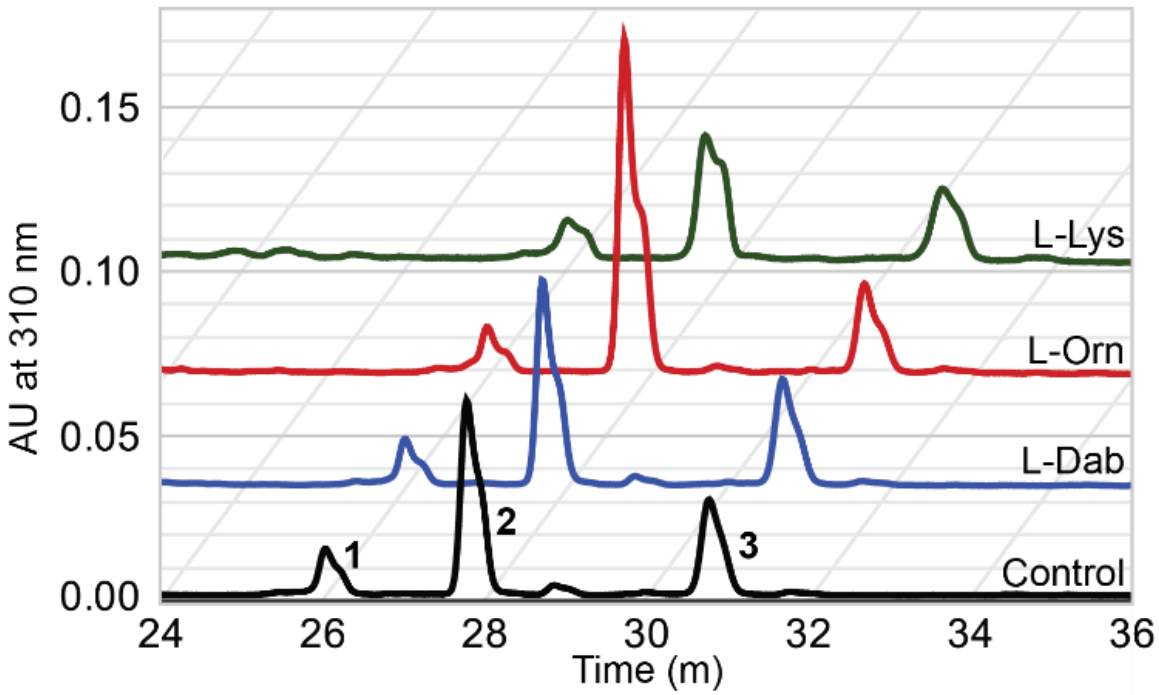


Figure 5.4. HPLC trace of catechol-containing products resulting from addition of L-Lys, L-Orn and L-Dab to cultures of *A. bouvetii* DSM 14964. The methanol eluent from the XAD column of 72 h cultures were monitored by HPLC-UV/Vis at 310 nm to detect catecholic compounds. Peaks corresponding to natural siderophores 1, 2, and 3 are labeled. Casamino acid minimal medium (control) was supplemented with the indicated amino acids (1 mM final concentration for each) prior to inoculation.

Propargylamine showed the most efficient coupling of any of the amines tested, nearly abolishing native siderophore production in favor of the 2,3-DHB-propargylamine conjugate **5** (Figures 5.5 and 5.20-5.23). The 2,3-DHB-allylamine conjugate **6** was also produced in greater quantities than the natural siderophore (Figures 5.5 and 5.24-5.27). Increasing the medium concentration of allylamine to 10 mM suppressed native compounds to levels similar to 1 mM propargylamine; however, the production of the desired conjugate **6** also decreased relative to 1 mM allylamine addition.

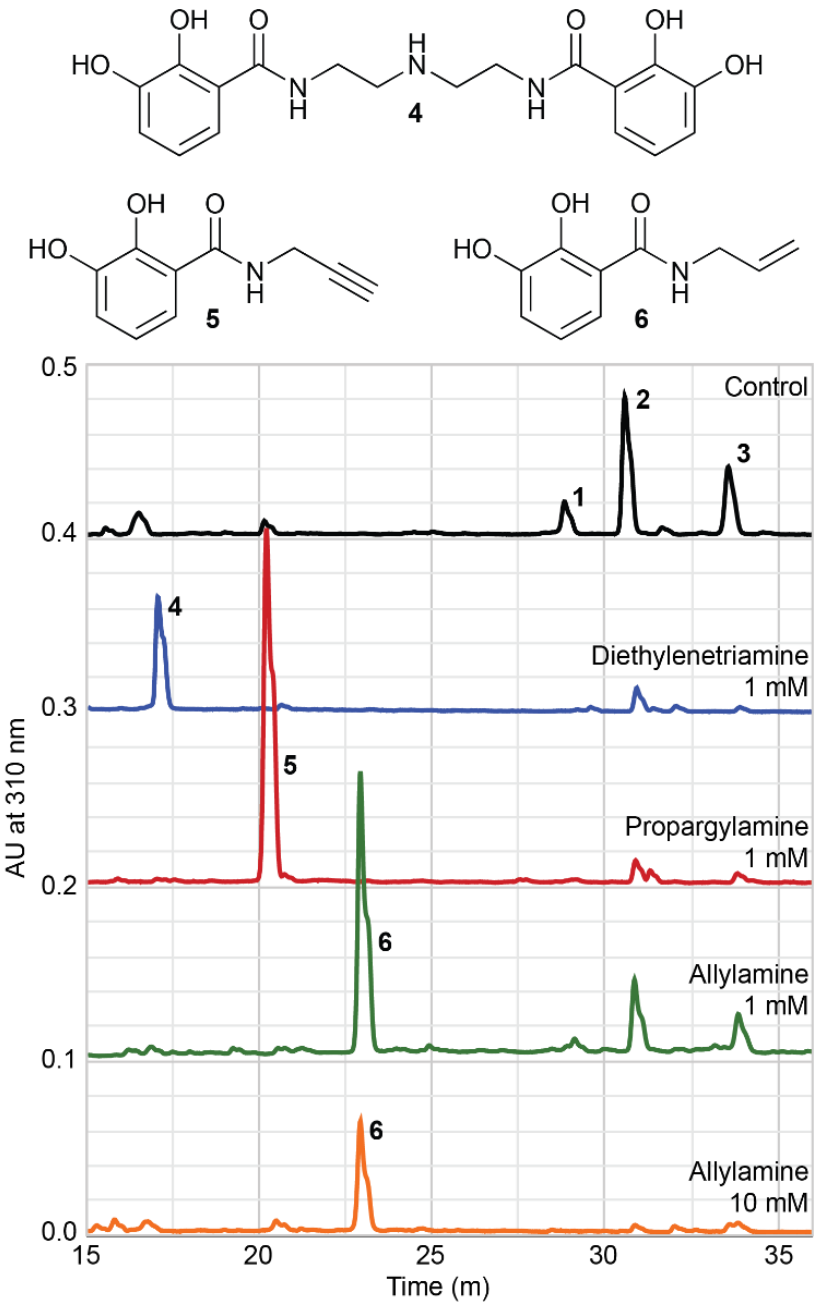


Figure 5.5. HPLC trace of catechol-containing products resulting from addition of diethylaminetriamine, allylamine, and propargylamine to cultures of *A. bouvetii* DSM 14964. The methanol eluent from the XAD column of 72 h cultures were monitored by HPLC-UV/Vis at 310 nm, indicative of catecholic compounds. Peaks corresponding to natural siderophores 1, 2, and 3 are labeled. Conjugate identities were confirmed by MS and NMR analyses (Figures 5.13-5.18 and 5.20-5.27). Casamino acid minimal medium (control) was supplemented with the indicated amines prior to inoculation.

5.4. Discussion

In sum, the genomes of several species of *Acinetobacter* were found to contain homologs of *vibH*, alongside biosynthetic genes putatively encoding the biosynthesis of catechol siderophores with an amine core. *A. bouvetii* DSM 14964 was found to produce three novel biscatechol siderophores: propanochelin (**1**), butanochelin (**2**), and pentanochelin (**3**). The amide bond-forming enzyme, putatively assigned as AbsH, is able to append a second 2,3-DHB molecule to diamines and polyamines, albeit with less flexibility than the first addition. Precursor directed biosynthesis revealed that the biosynthetic machinery has a relaxed specificity for the diamine substrate, allowing for the biosynthesis of a variety of non-natural siderophore analogs. Of particular significance, *A. bouvetii* DSM 14964 condenses 2,3-DHB to allylamine and propargylamine, producing catecholic compounds which bind iron(III) and may be further modified via thiol-ene or azide-alkyne click chemistry.[18,19] Propargylamine was particularly well incorporated, outcompeting the natural diamine substrates at 1 mM concentration. With the exceptions of tren and norspermidine, growth of *A. bouvetii* DSM 14964 was not significantly affected by diamine supplementation. Future enzymatic studies *in vitro* or in a heterologous system will provide further evidence that AbsH is responsible for amide bond formation. AbsH and other VibH homologs may prove useful in the synthesis of catechol compounds tailored for robust adhesion in salty aqueous conditions, such as those with attendant amine functionalities mimicking mussel foot proteins.[15,20]

5.5. Appendix

5.5.1. Characterization of novel compounds

N,N'-(azanediylbis(ethane-2,1-diyl))bis(2,3-dihydroxybenzamide) **(4). ^{1}H NMR** (600 MHz, DMSO) δ 12.31 (s, 2H), 9.23 (br s, 2H), 8.93 (br t, J = 5.6 Hz, 2H), 8.72 (br s, 1H), 7.28 (dd, J = 8.2, 1.5 Hz, 2H), 6.93 (dd, J = 7.8, 1.4 Hz, 2H), 6.69 (t, J = 7.9 Hz, 2H), 3.59 (q, J = 5.9 Hz, 4H), 3.17 (m, 4H). **^{13}C NMR** (126 MHz, DMSO) δ 170.76, 149.84, 146.68, 119.48, 118.52, 118.10, 115.58, 46.91, 36.07. **HRMS** for $C_{18}H_{22}N_3O_6^+$: calcd. $[M+H]^+$ 376.1508, found: 376.1507.

2,3-dihydroxy-N-(prop-2-yn-1-yl)benzamide **(5). ^{1}H NMR** (500 MHz, DMSO) δ 12.31 (s, 1H), 9.21 (s, 1H), 9.14 (br t, J = 5.5 Hz, 1H), 7.26 (dd, J = 8.1, 1.5 Hz, 1H), 6.91 (dd, J = 7.9, 1.5 Hz, 1H), 6.68 (t, J = 8.0 Hz, 1H), 4.06 (dd, J = 5.6, 2.5 Hz, 2H), 3.15 (t, J = 2.5 Hz, 1H). **^{13}C NMR** (126 MHz, DMSO) δ 169.79, 149.85, 146.68, 119.53, 118.66, 117.86, 115.33, 81.18, 73.68, 28.76. **HRMS** for $C_{10}H_{10}NO_3^+$: calcd. $[M+H]^+$ 192.0661, found: 192.0654.

N-allyl-2,3-dihydroxybenzamide **(6). ^{1}H NMR** (500 MHz, DMSO) δ 12.66 (br s, 1H), 9.16 (br s, 1H), 8.96 (br t, J = 5.9 Hz, 1H), 7.32 (d, J = 8.0 Hz, 1H), 6.92 (d, J = 7.7 Hz, 1H), 6.69 (t, J = 8.0 Hz, 1H), 5.90 (ddt, J = 16.0, 10.5, 5.3 Hz, 1H), 5.18 (d, J = 17.2 Hz, 1H), 5.12 (d, J = 10.3 Hz, 1H), 3.92 (t, J = 5.6 Hz, 2H). **^{13}C NMR** (126 MHz, DMSO) δ 170.01, 150.07, 146.68, 135.22, 119.30, 118.46, 117.68, 116.00, 115.41, 41.67. **HRMS** for $C_{10}H_{12}NO_3^+$: calcd. $[M+H]^+$ 194.0817, found: 194.0818.

5.5.2. Supplementary tables

Table 5.1. *Acinetobacter* strains putatively containing a siderophore biosynthetic gene cluster with a VibH homolog.	
Protein Accession	Organism
A0A4Q7ARR8_9GAMM	*Acinetobacter bouvetii*
N9DJ67_9GAMM	*Acinetobacter bouvetii* DSM 14964 = CIP 107468
A0A1A7R8J8_9GAMM	*Acinetobacter gandensis*
A0A1B2M479_9GAMM	*Acinetobacter larvae*
A0A4Q4H0F1_9GAMM	*Acinetobacter piscicola*
A0A151Y3Z5_9GAMM	*Acinetobacter pragensis*
A0A240ECV6_9GAMM	*Acinetobacter puyangensis*
A0A241WES3_9GAMM	*Acinetobacter* sp. ANC 3813
A0A4R0ECX9_9GAMM	*Acinetobacter* sp. ANC 4249
A0A1H7NGA5_9GAMM	*Acinetobacter* sp. DSM 11652
N9M6P9_9GAMM	*Acinetobacter* sp. NIPH 713
A0A031LTQ2_9GAMM	*Acinetobacter* sp. Ver3
A0A3B7LX09_9GAMM	*Acinetobacter* sp. WCHAc010005
A0A3B7PT29_9GAMM	*Acinetobacter* sp. WCHAc010034
A0A3B7M868_9GAMM	*Acinetobacter* sp. WCHAc010052

Table 5.2. The putative siderophore biosynthetic gene cluster of *A. bouvetii* DSM 14964			
Name	**Locus tag**	**Accession**	**Putative function**
bsuB	B090_RS0102660	WP_005009566.1	Ferric siderophore reductase
absH	B090_RS0102665	WP_005009568.1	VibH-like condensation domain
absA	B090_RS0102670	WP_005009570.1	2,3-dihydro-2,3-DHB dehydratase
absC	B090_RS0102675	WP_005009576.1	Isochorismate synthase
absE	B090_RS0102680	WP_005009579.1	2,3-DHB–AMP ligase
absB	B090_RS0102685	WP_005009581.1	Isochorismatase
absG	B090_RS0102690	WP_005009583.1	Aryl carrier protein
absI	B090_RS0102695	WP_005009584.1	Ornithine decarboxylase
bsuA	B090_RS0102700	WP_026253326.1	TonB-dependent receptor

Table 5.3. Natural and non-natural amines used in precursor directed biosynthesis studies.*		
Substrate	**Single Addition Product**	**Double Addition Product**
Diamines		
Ethylenediamine	**Major**	N.D.†
1,3-Diaminopropane	N.D.	**Major**
Putrescine (1,4-diaminobutane)	N.D.	**Major**
Cadaverine (1,5-diaminopentane)	N.D.	**Major**
1,6-Diaminohexane	**Major**	**Major**
Amino acids		
L-Diaminobutyric acid	N.D.	N.D.
L-Ornithine	N.D.	N.D.
L-Lysine	N.D.	N.D.
Polyamines		
Diethylenetriamine	N.D.	**Major**
Norspermidine	*No growth detected*	
Tris(2-aminoethyl)amine (Tren)	*No growth detected*	
Other amines		
Histamine	N.D.	-
Allylamine	**Major**	-
Propargylamine	**Major**	-
* Substrates were added to casamino acid minimal medium prior to inoculation at a final concentration of 1 mM. † N.D., not detected.		

5.5.3. Supplementary figures

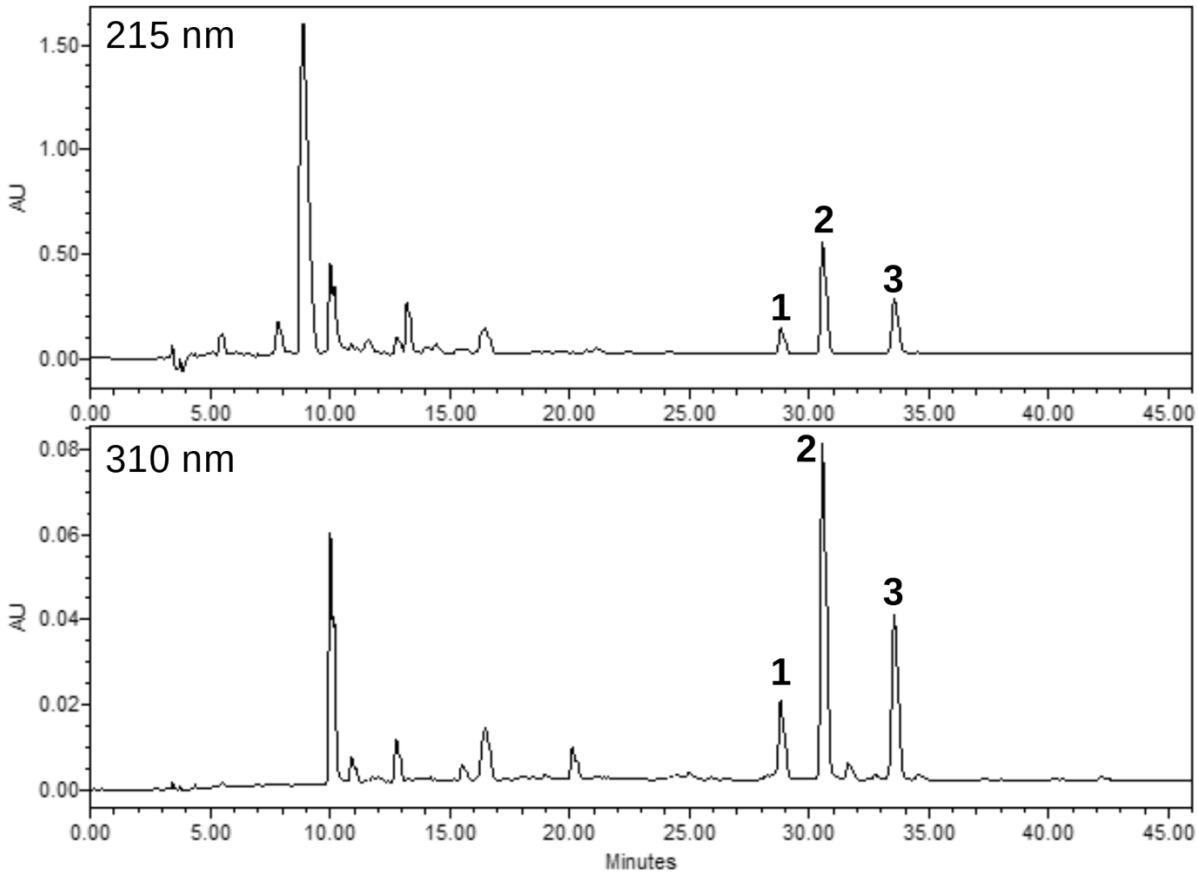


Figure 5.6. HPLC–UV/Vis chromatogram of *A. bouvetii* DSM 14964 supernatant extract from growth in low iron medium. Top: chromatogram monitored at 215 nm. Bottom: chromatogram monitored at 310 nm. Peaks corresponding to natural siderophores **1**, **2**, and **3** are labeled.

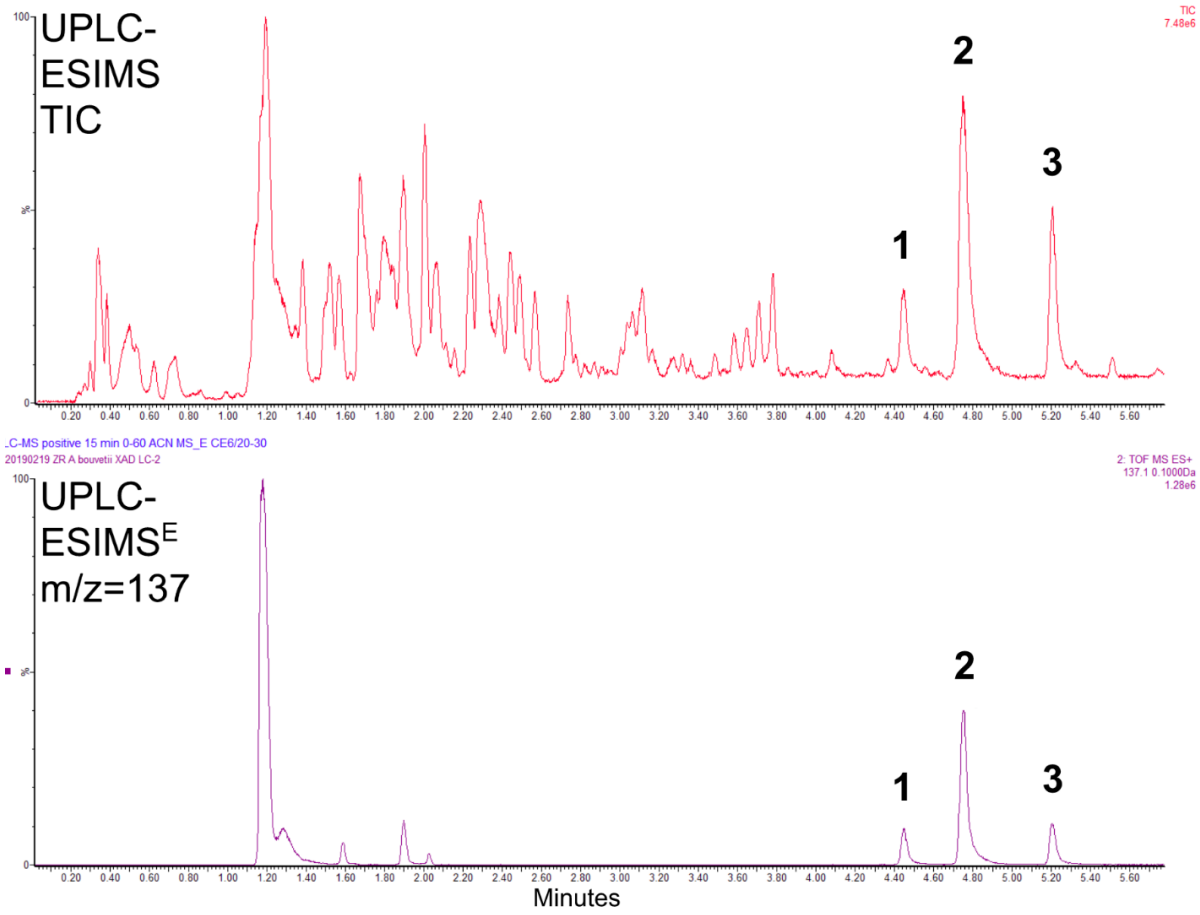


Figure 5.7. UPLC–ESIMS spectra of *A. bouvetii* DSM 14964 supernatant extract from a low iron medium. Top: Total ion chromatogram. Bottom: MS^E high-collision-energy extracted ion chromatogram for 137 m/z. Peaks corresponding to natural siderophores **1**, **2**, and **3** are labeled.

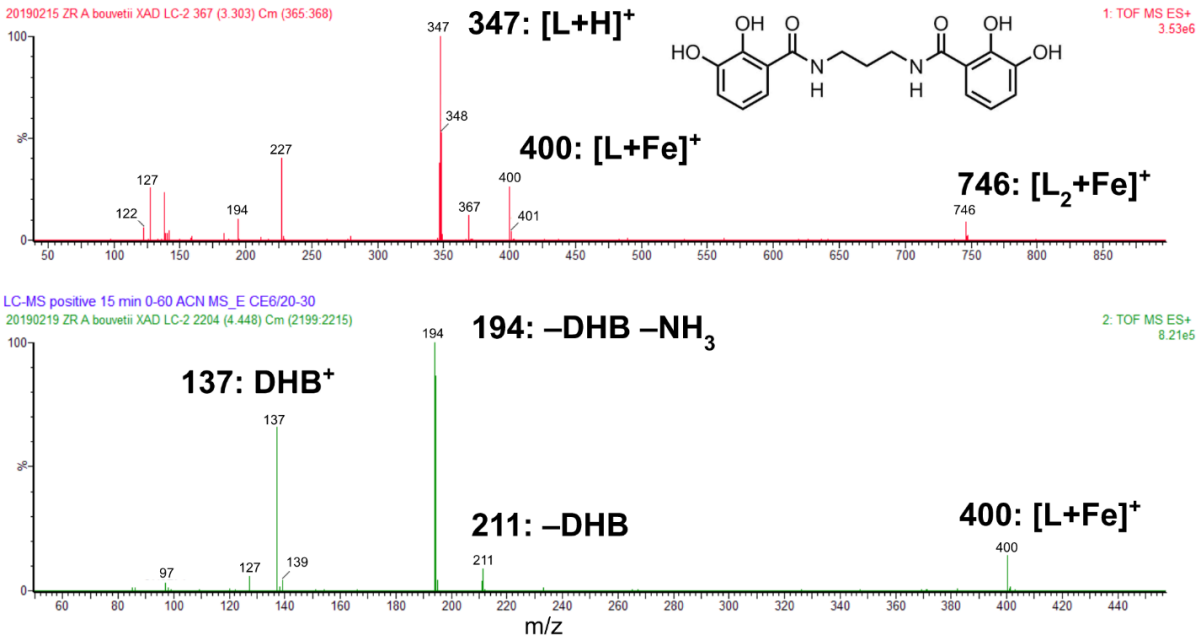


Figure 5.8. Mass spectrometry and fragmentation of propanochelin (**1**). Top: ESIMS spectrum. Bottom: high-collision-energy MS^E fragmentation.

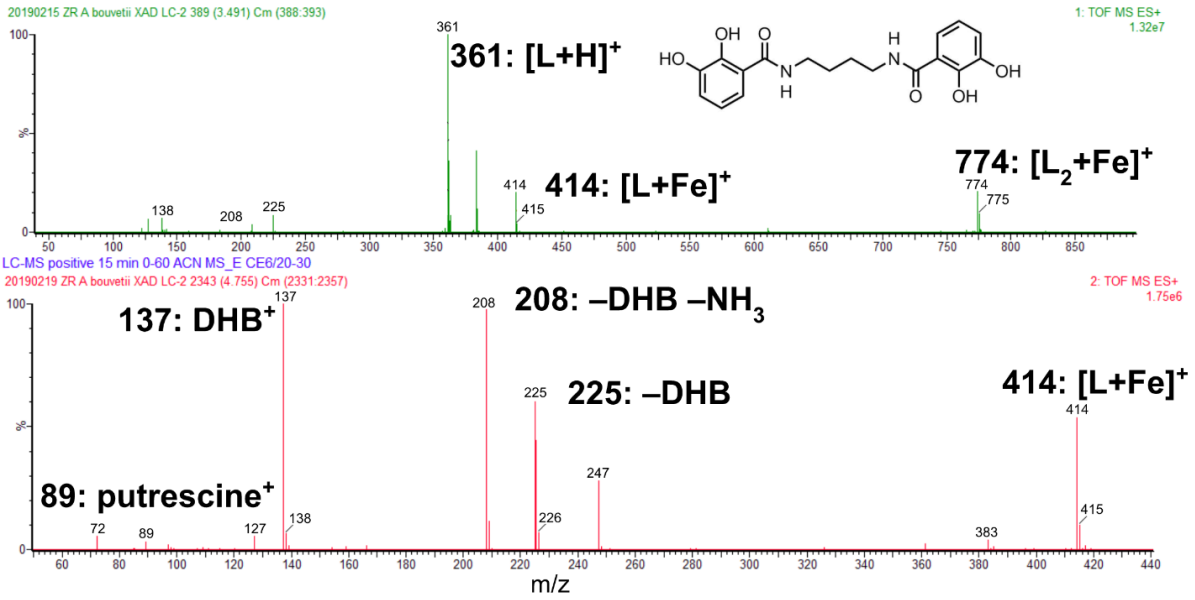


Figure 5.9. Mass spectrometry and fragmentation of butanochelin (**2**). Top: ESIMS spectrum. Bottom: high-collision-energy MS^E fragmentation.

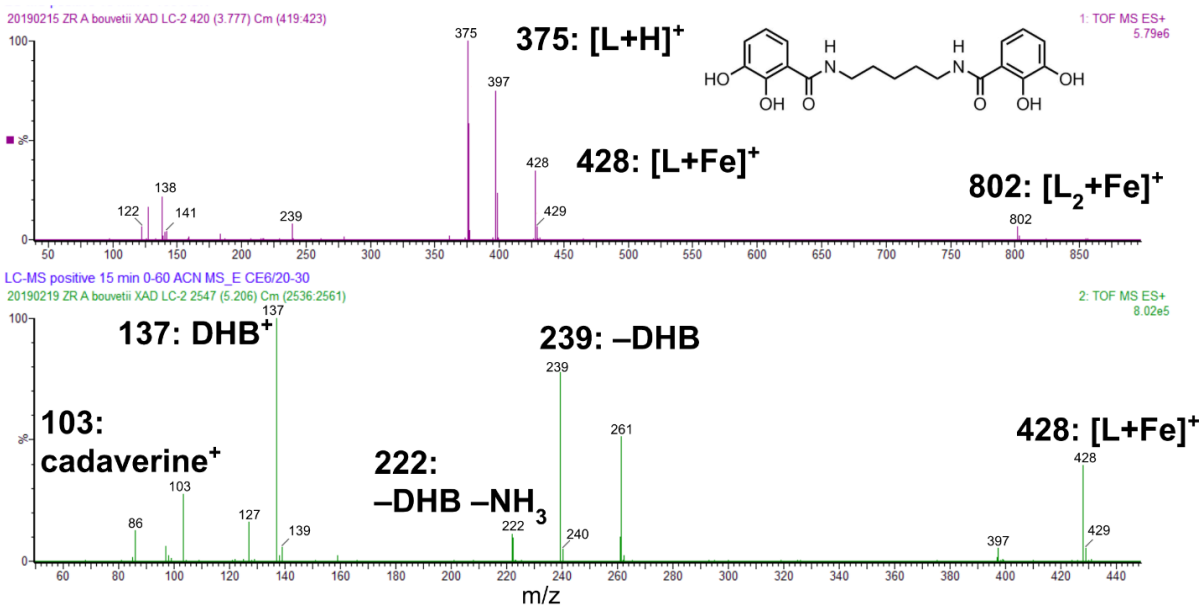


Figure 5.10. Mass spectrometry and fragmentation of pentanochelin (**3**). Top: ESIMS spectrum. Bottom: high-collision-energy MS^E fragmentation.

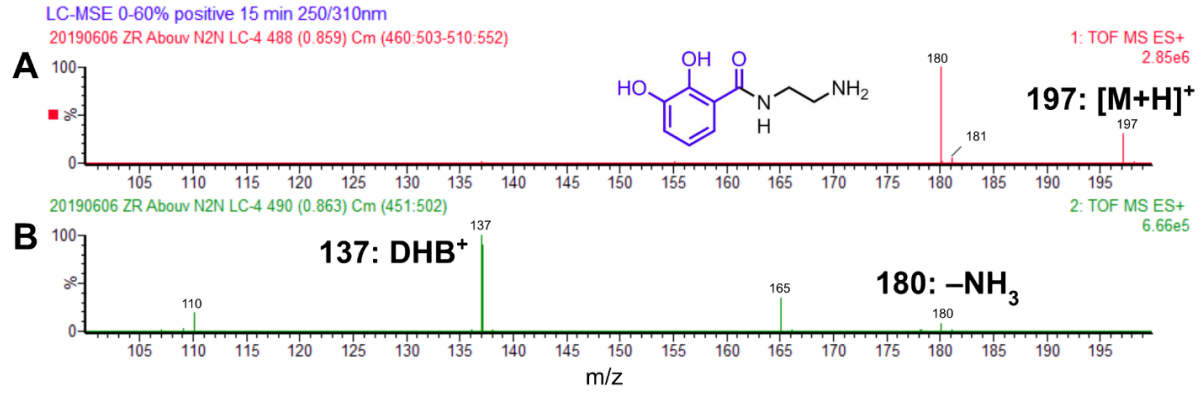


Figure 5.11. Mass spectrometry and fragmentation of the catecholic product unique to ethylenediamine supplementation in the growth medium. A: ESIMS spectrum and proposed structure. B: high-collision-energy MS^E fragmentation.

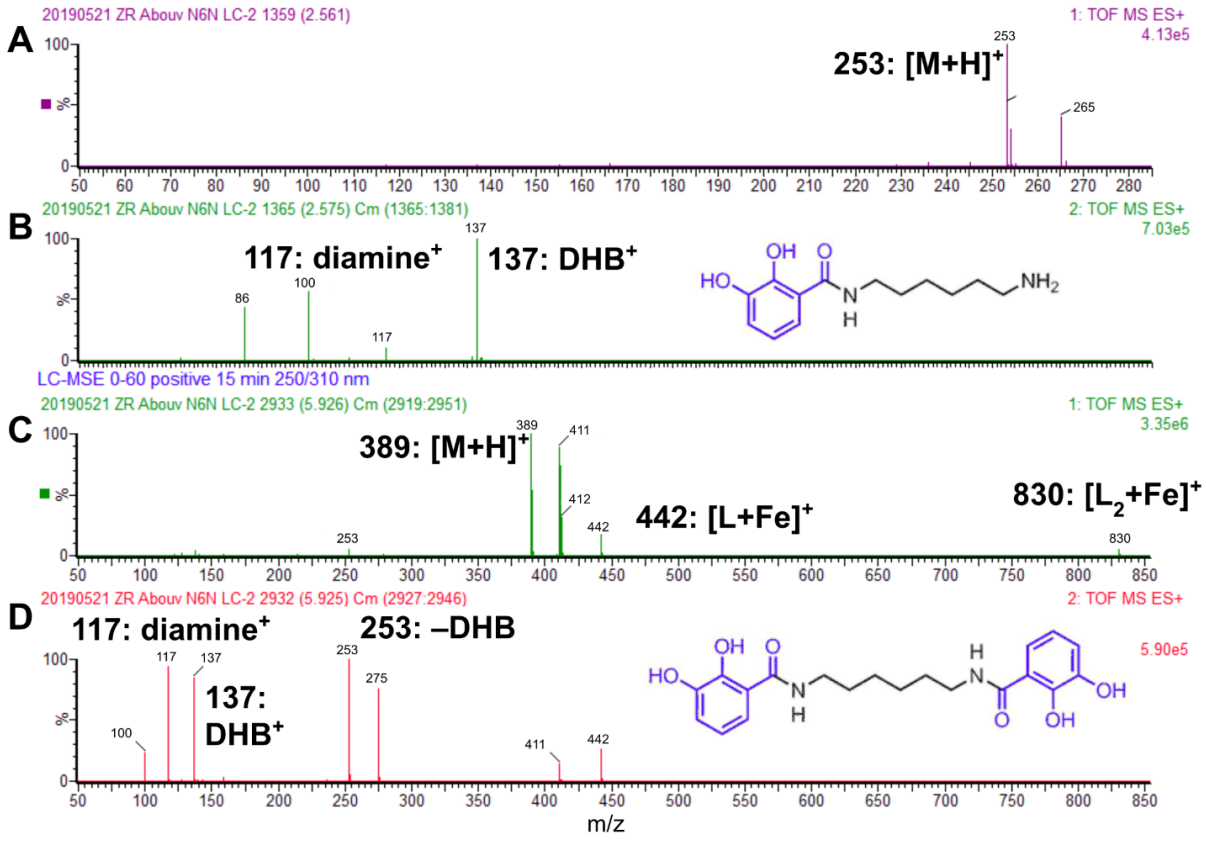


Figure 5.12. Mass spectrometry and fragmentation of the catecholic products unique to 1,6-diaminohexane supplementation in the growth medium. A: ESIMS spectrum of the single 2,3-DHB addition product. B: high-collision-energy MS^E fragmentation and proposed structure. C: ESIMS spectrum of the double 2,3-DHB addition product. D: high-collision-energy MS^E fragmentation and proposed structure.

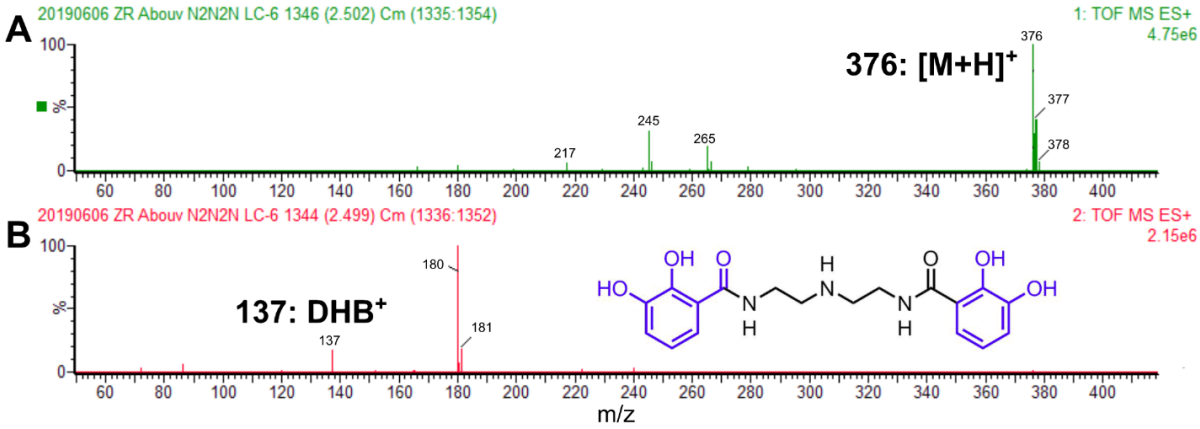


Figure 5.13. Mass spectrometry and fragmentation of the catecholic product **4** unique to diethylene triamine supplementation. A: ESIMS spectrum. B: high-collision-energy MS^E fragmentation and proposed structure.

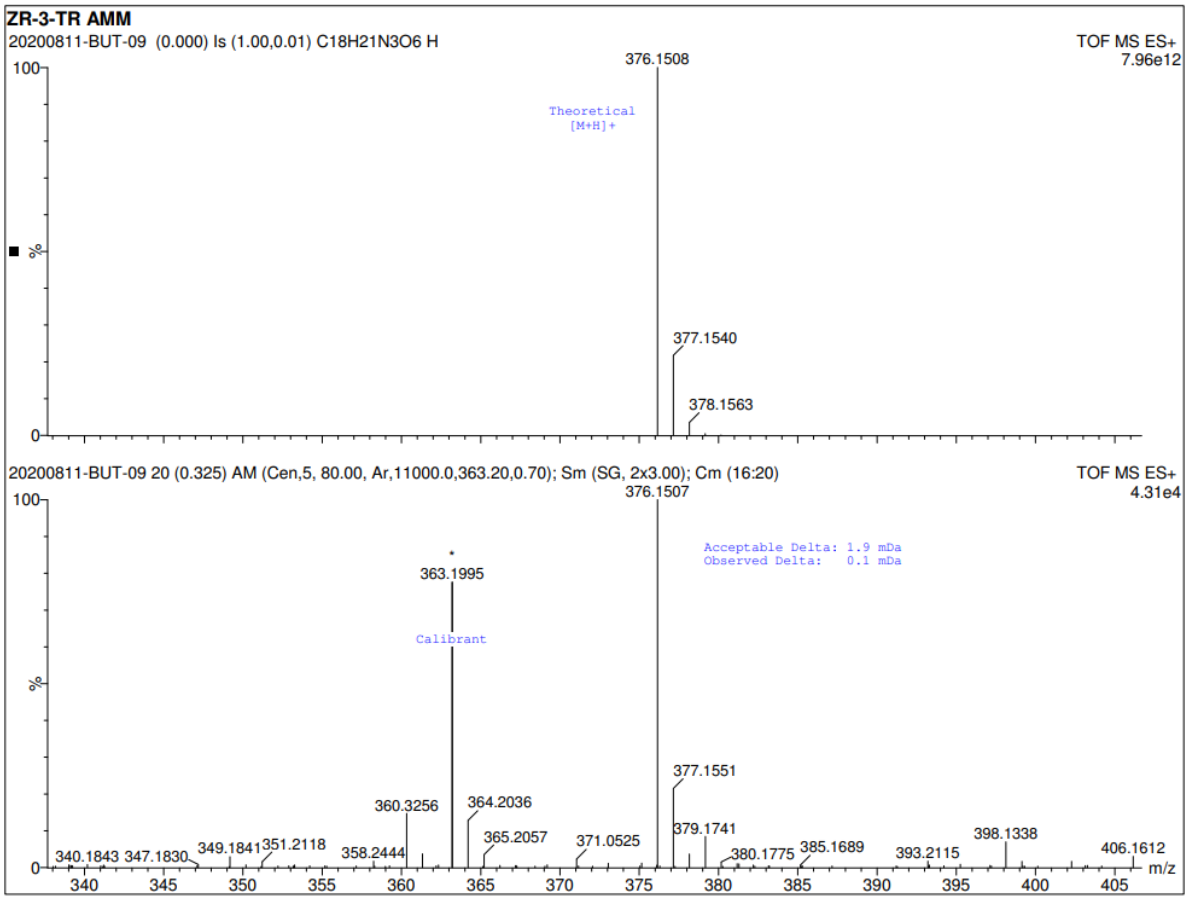


Figure 5.14. High-resolution mass spectrometry (positive-mode ESI) of the catecholic **4** product unique to diethylene triamine supplementation.

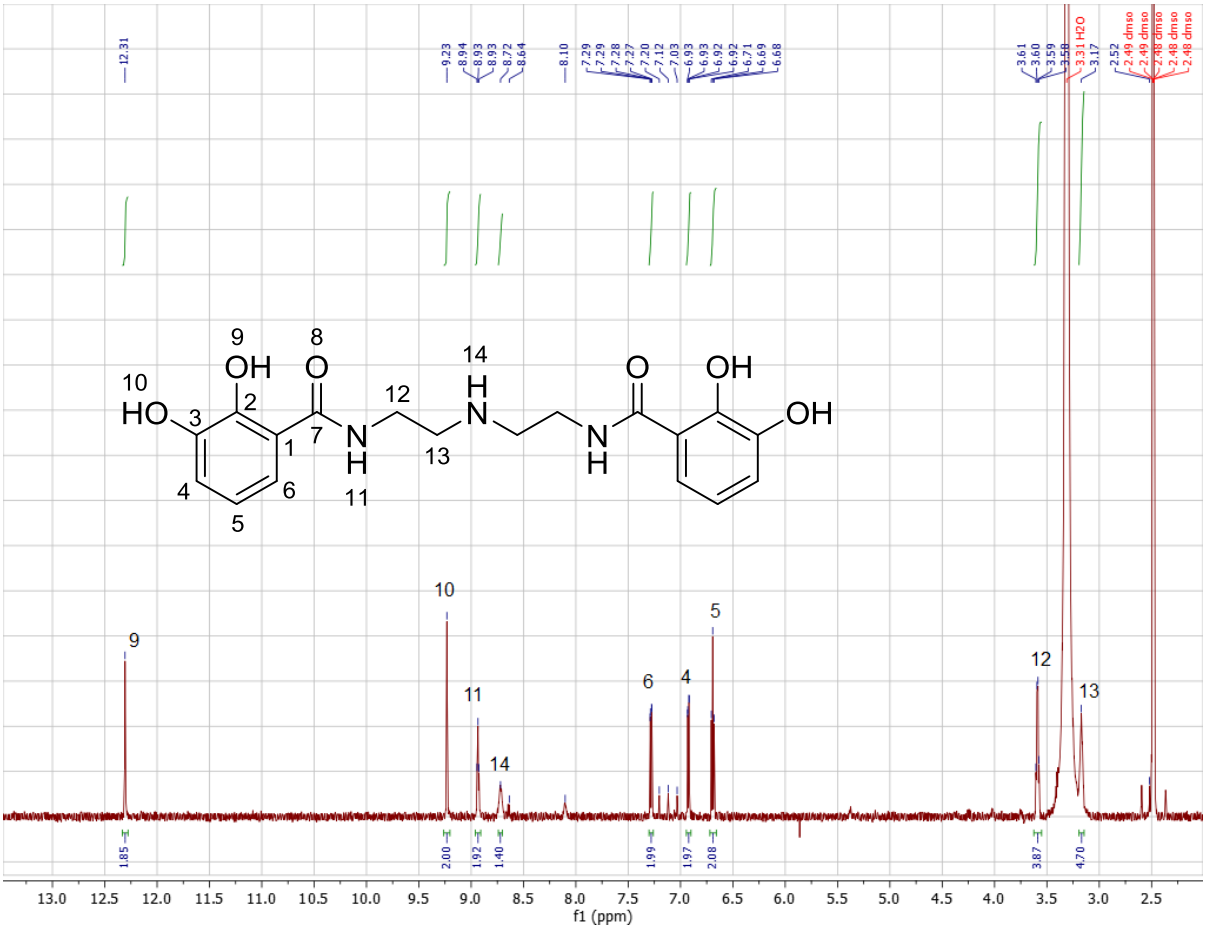


Figure 5.15. ^{1}H NMR (DMSO-d_6; 500 MHz) of the catecholic product **4** unique to diethylene triamine supplementation.

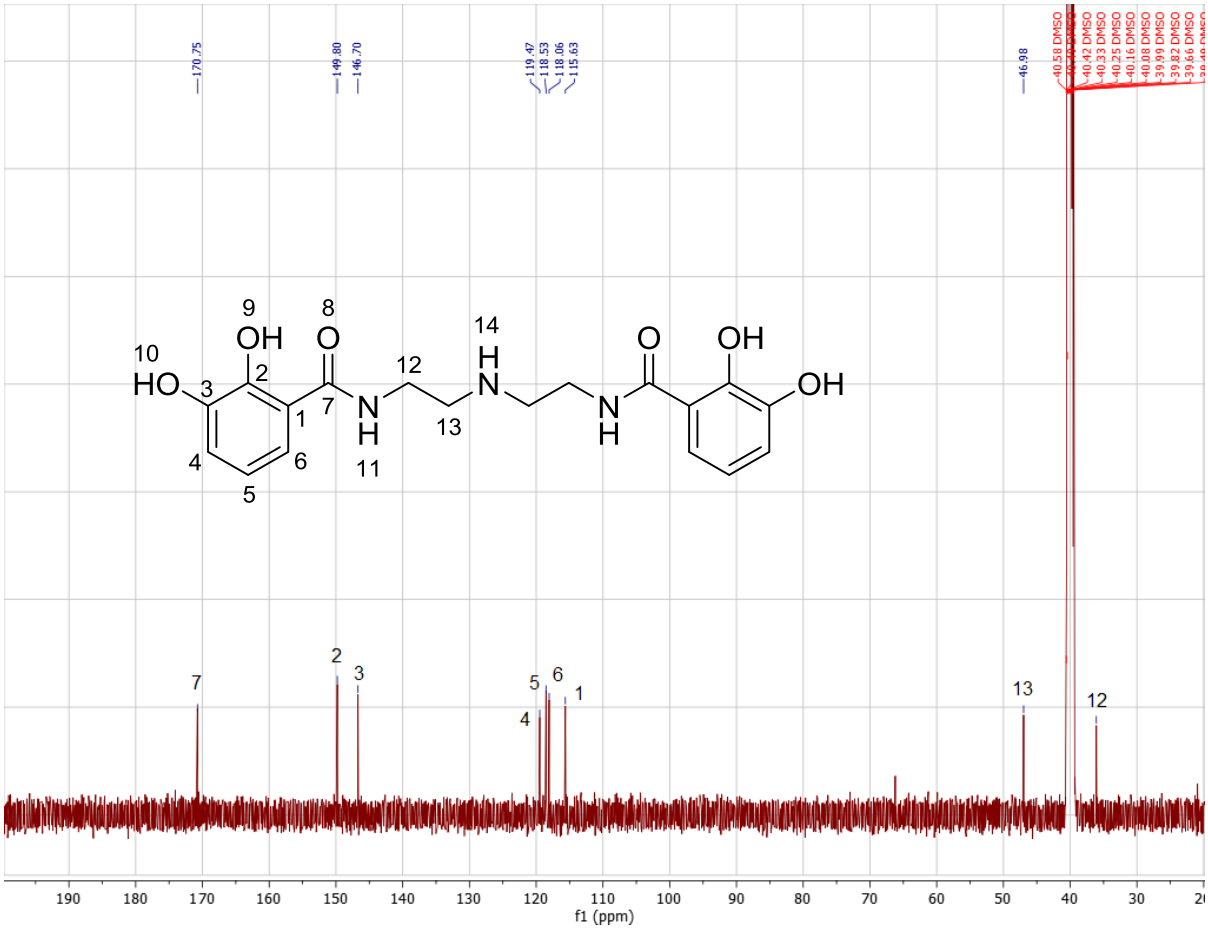


Figure 5.16. ^{13}C NMR (DMSO-d_6; 126 MHz) of the catecholic product **4** unique to diethylene triamine supplementation.

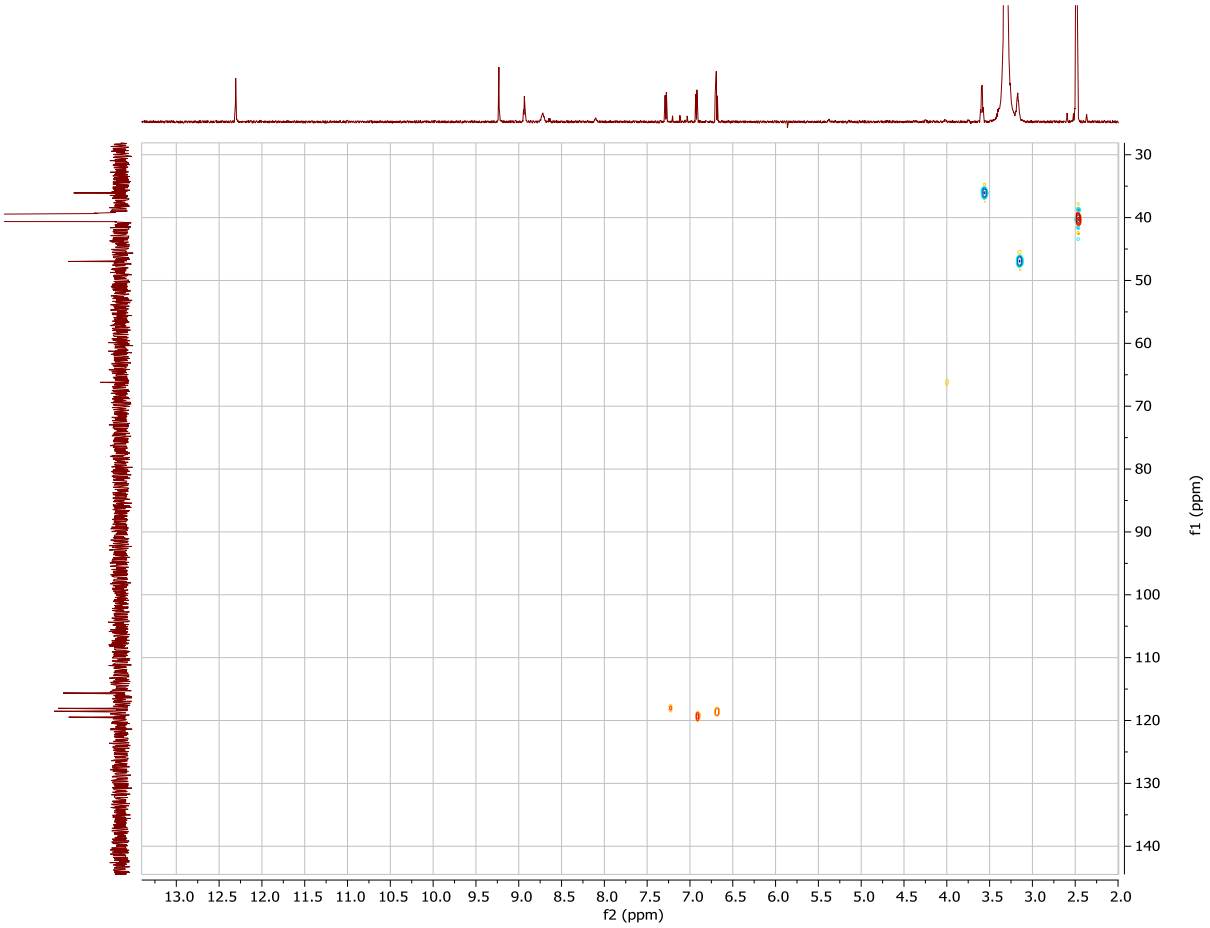


Figure 5.17. ^{1}H-^{13}C HSQC NMR (DMSO-d_6; 600 MHz) of the catecholic product **4** unique to diethylene triamine supplementation.

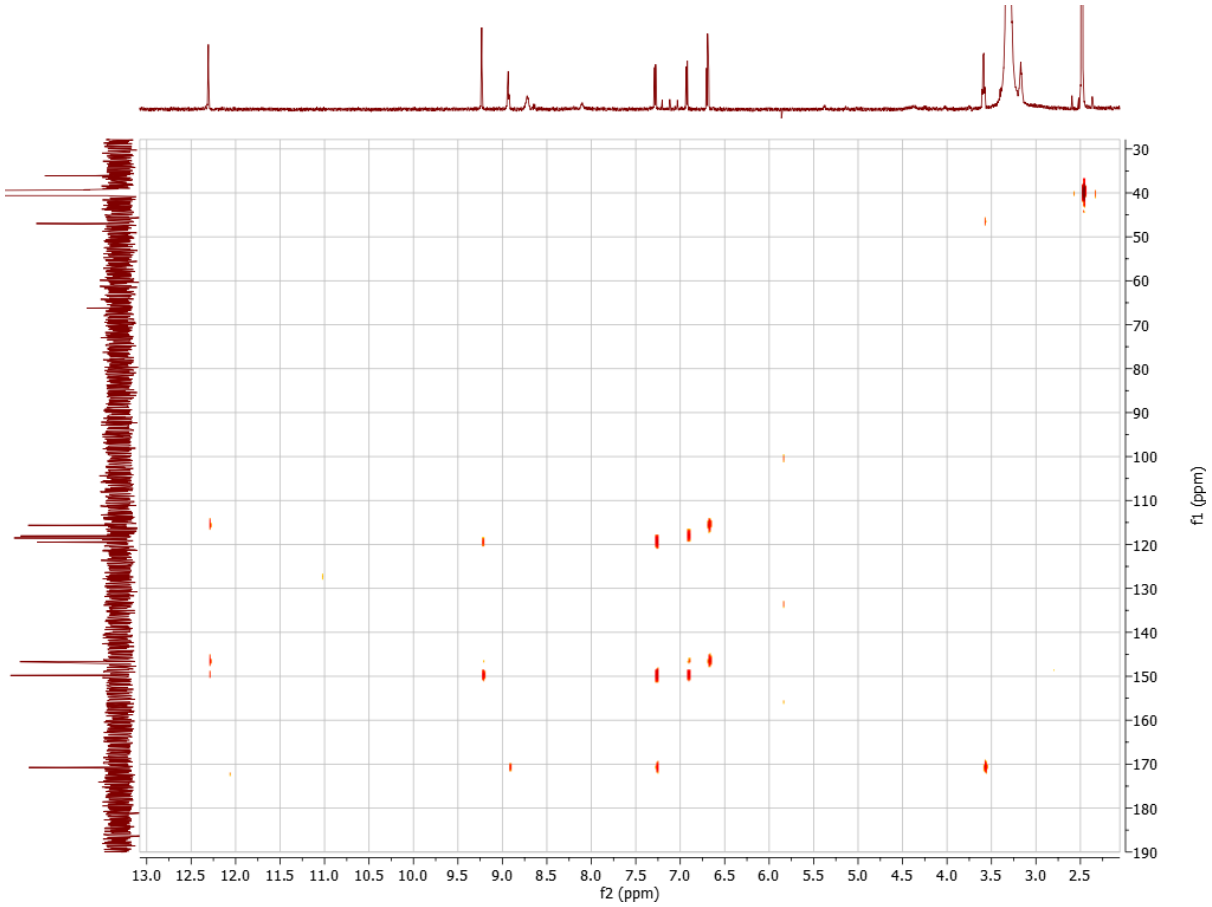


Figure 5.18. ^{1}H-^{13}C HMBC NMR (DMSO-d_6; 600 MHz) of the catecholic product **4** unique to diethylene triamine supplementation.

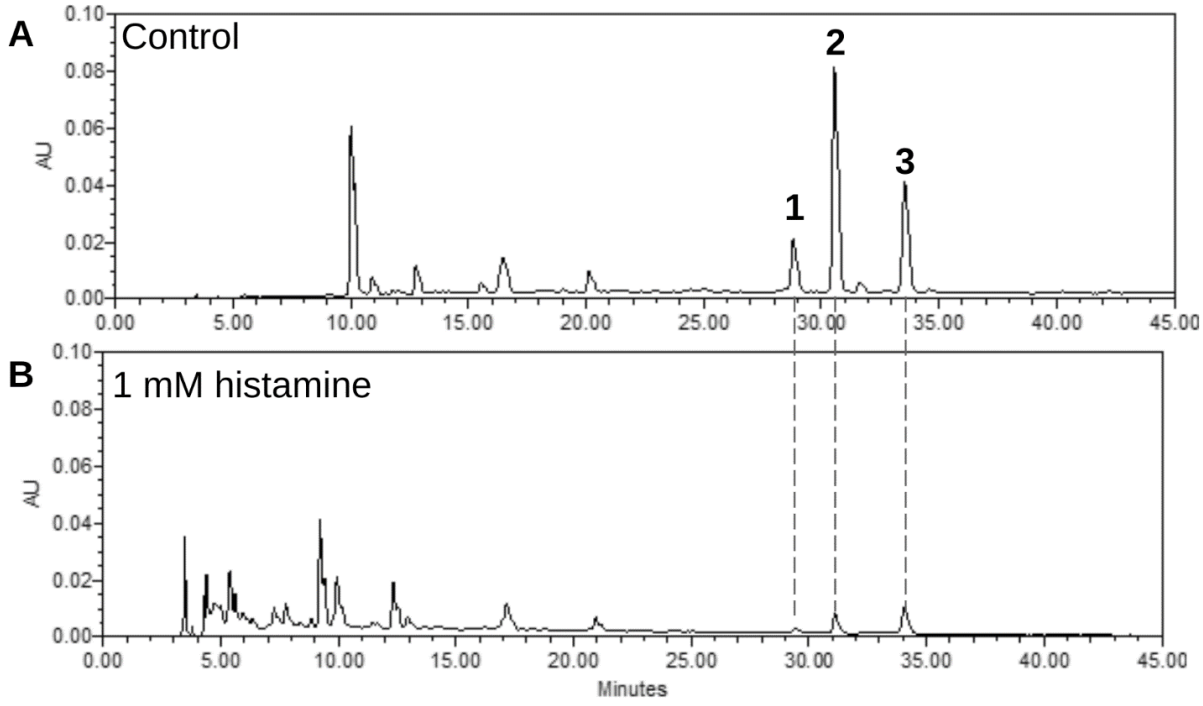


Figure 5.19. Supplementation of *A. bouvetii* DSM 14964 growth with functionalized amines. HPLC of the bacterial supernatant monitored at 310 nm of *A. bouvetii* DSM 14964 grown with (A) no added substrate or (B) 1 mM histamine. Peaks corresponding to natural siderophores **1**, **2**, and **3** are labeled.

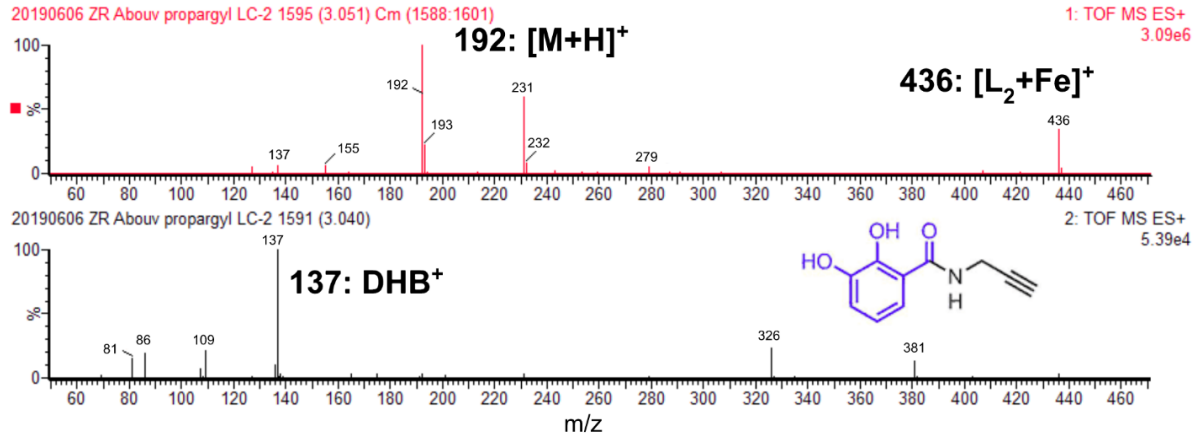


Figure 5.20. Mass spectrometry and fragmentation of the catecholic product **5** unique to propargylamine supplementation. A: ESIMS spectrum. B: high-collision-energy MS^E fragmentation and proposed structure.

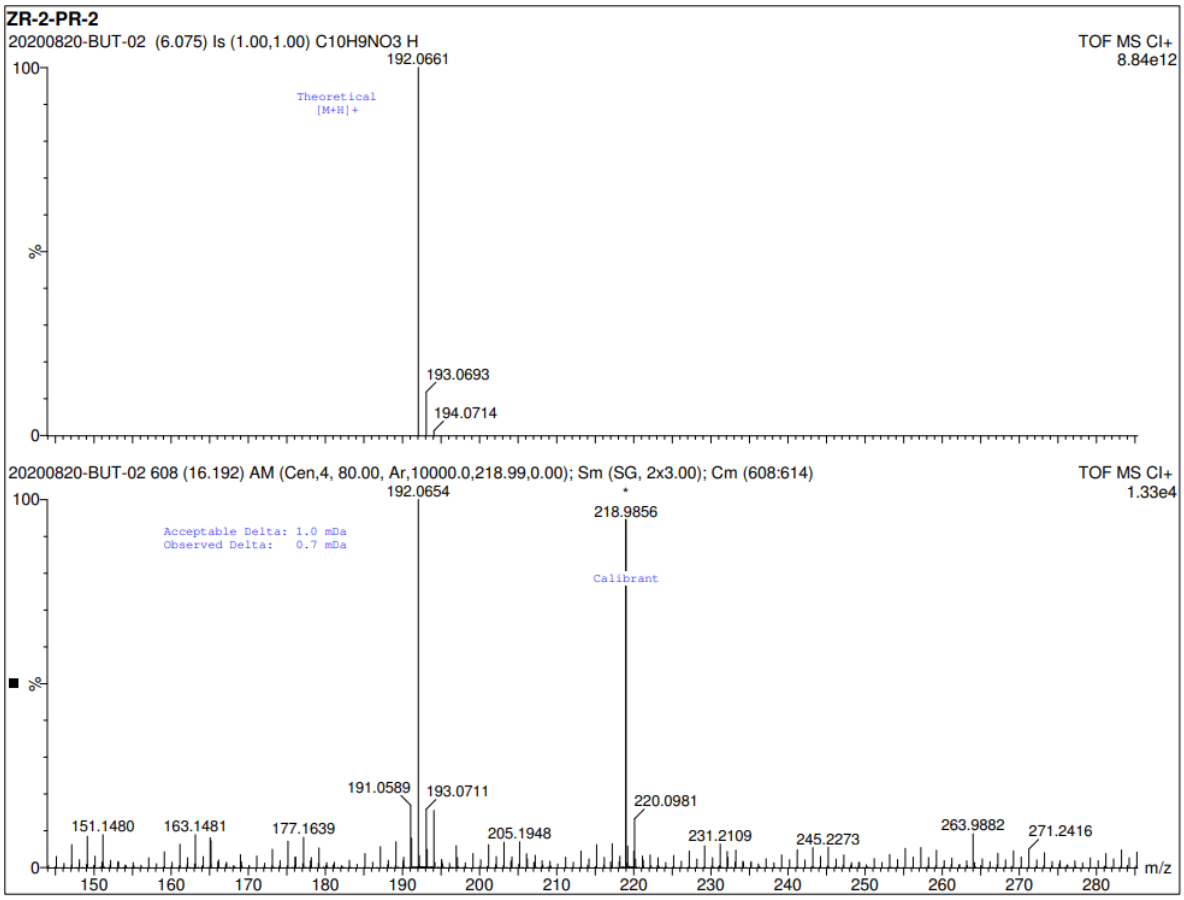


Figure 5.21. High-resolution mass spectrometry (positive-mode CI) of the catecholic product **5** unique to propargylamine supplementation.

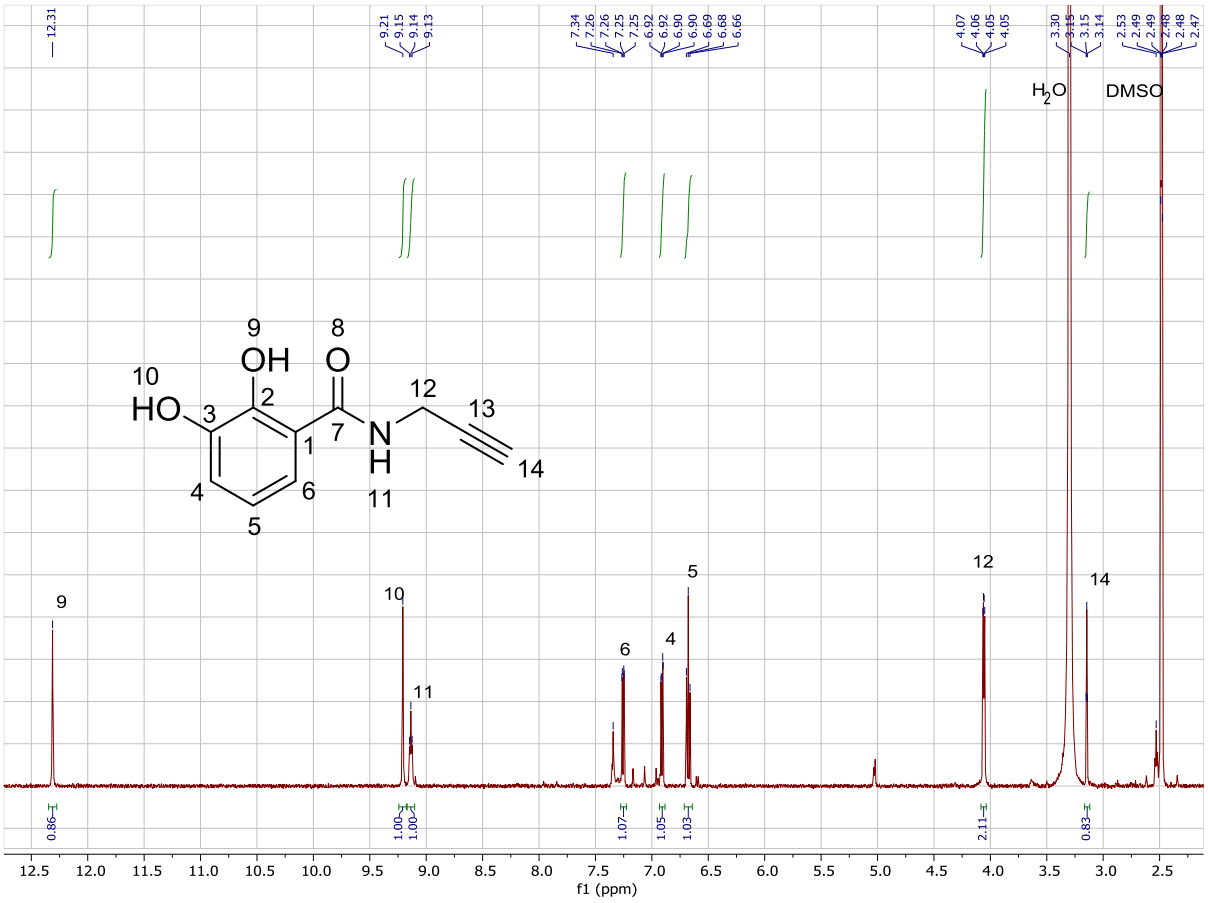


Figure 5.22. ^{1}H NMR (DMSO-d_6; 500 MHz) of the catecholic product **5** unique to propargylamine supplementation.

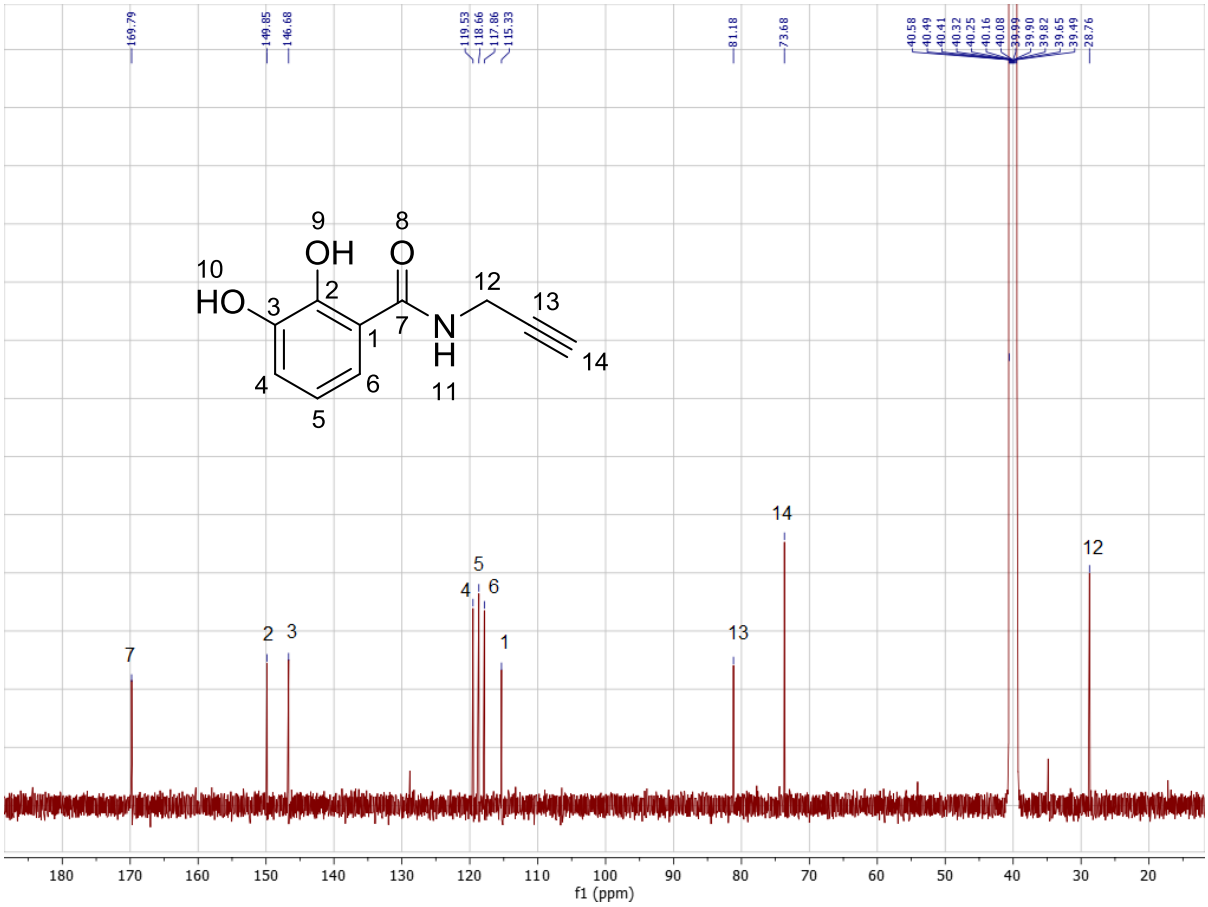


Figure 5.23. ^{13}C NMR (DMSO-d_6; 126 MHz) of the catecholic product **5** unique to propargylamine supplementation.

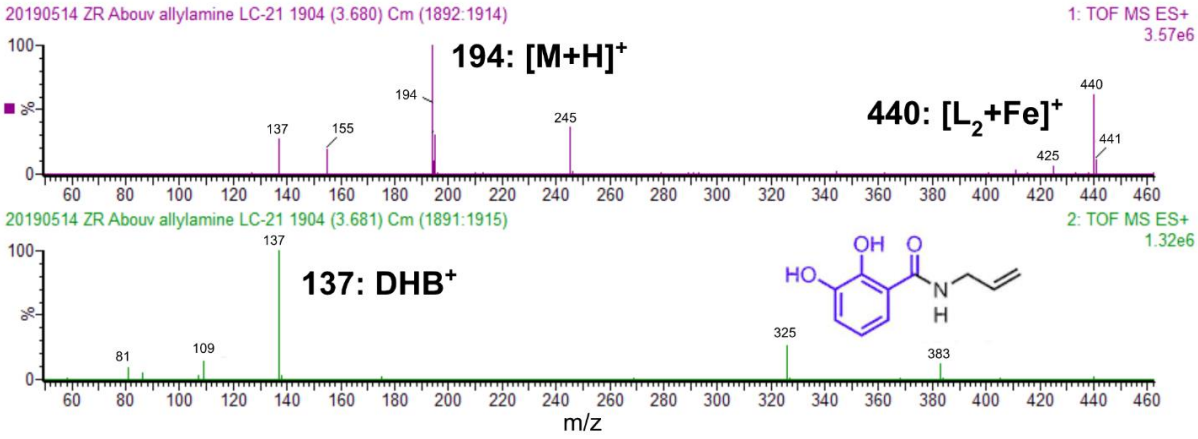


Figure 5.24. Mass spectrometry and fragmentation of the catecholic product **6** unique to allylamine supplementation. A: ESIMS spectrum. B: high-collision-energy MS^E fragmentation and proposed structure.

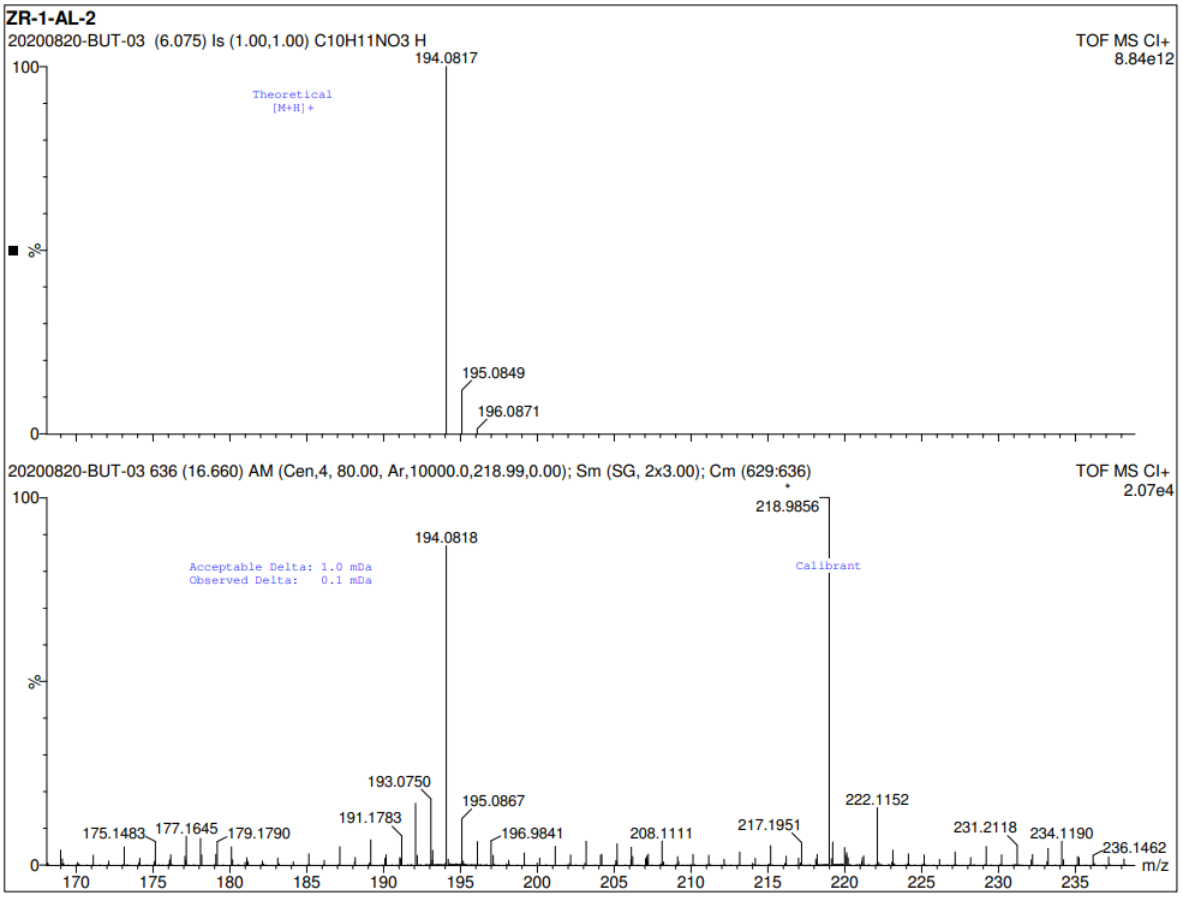


Figure 5.25. High-resolution mass spectrometry (positive-mode CI) of the catecholic product **6** unique to allylamine supplementation.

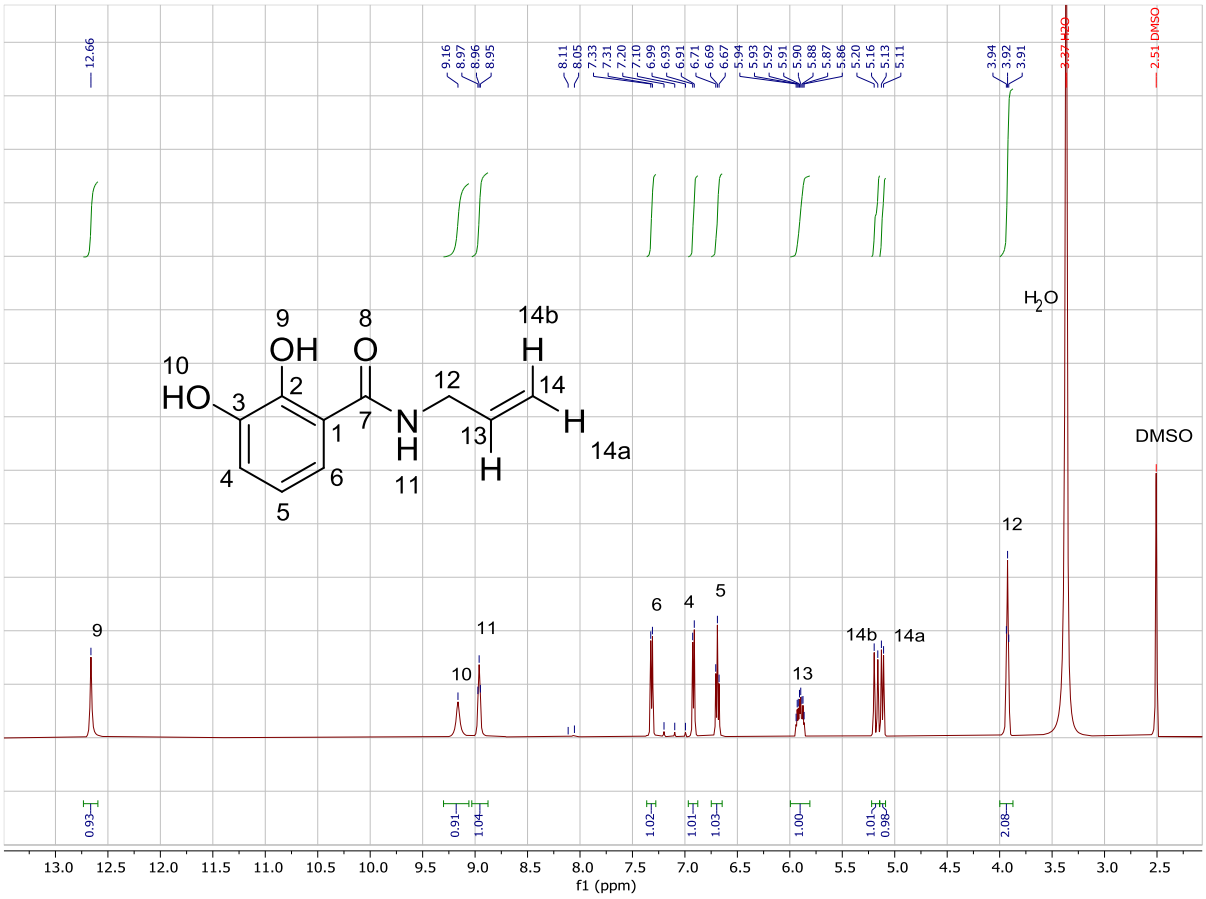


Figure 5.26. ^{1}H NMR (DMSO-d_6; 500 MHz) of the catecholic product **6** unique to allylamine supplementation.

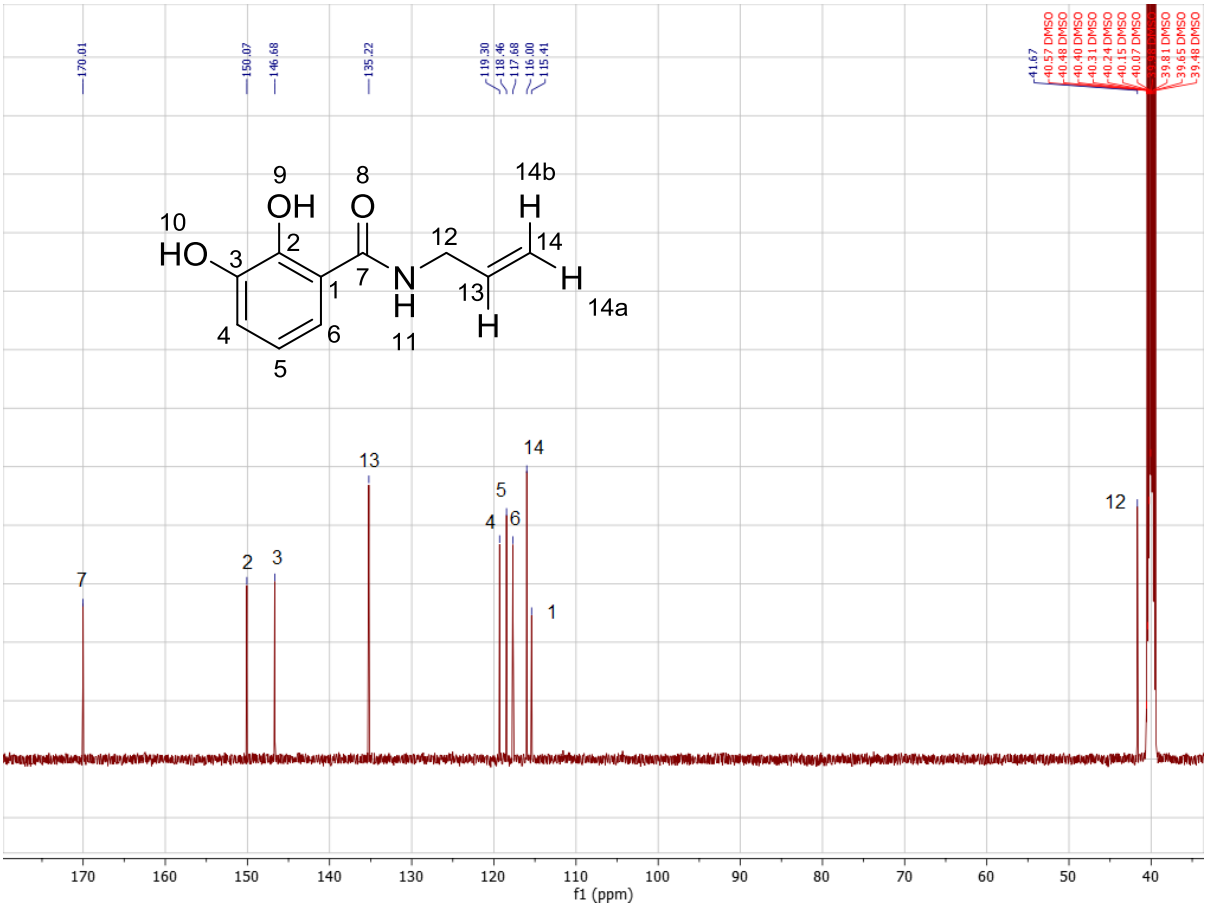


Figure 5.27. ^{13}C NMR (DMSO-d_6; 126 MHz) of the catecholic product **6** unique to allylamine supplementation.

5.6. References

1. Z. L. Reitz and A. Butler, Precursor-directed biosynthesis of catechol compounds in Acinetobacter bouvetii DSM 14964, *Chem. Commun.*, 2020, **56**, 12222-12225.
2. A. M. Gehring, K. A. Bradley and C. T. Walsh, Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate, *Biochemistry*, 1997, **36**, 8495-8503.
3. T. A. Keating, C. G. Marshall and C. T. Walsh, Vibriobactin biosynthesis in Vibrio cholerae: VibH is an amide synthase homologous to nonribosomal peptide synthetase condensation domains, *Biochemistry*, 2000, **39**, 15513-15521.

4. T. A. Keating, C. G. Marshall, C. T. Walsh and A. E. Keating, The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains, *Nat. Struct. Biol.*, 2002, **9**, 522-526.
5. M. R. Seyedsayamdost, S. Cleto, G. Carr, H. Vlamakis, M. João Vieira, R. Kolter and J. Clardy, Mixing and matching siderophore clusters: structure and biosynthesis of serratiochelins from Serratia sp. V4, *J. Am. Chem. Soc.*, 2012, **134**, 13550-13553.
6. O. Baars, X. Zhang, F. M. M. Morel and M. R. Seyedsayamdost, The Siderophore Metabolome of Azotobacter vinelandii, *Appl. Environ. Microbiol.*, 2015, **82**, 27-39.
7. R. M. Roop II, A. E. Elhassanny, M. A. Almirón, E. S. Anderson and X. J. Atkinson, in *Metals and the Biology and Virulence of Brucella*, eds. R. M. Roop Ii and C. C. Caswell, Springer International Publishing, Cham, 2017, pp. 9-39.
8. M. Winn, J. K. Fyans, Y. Zhuo and J. Micklefield, Recent advances in engineering nonribosomal peptide assembly lines, *Nat. Prod. Rep.*, 2016, **33**, 317-347.
9. C. Z. Soe and R. Codd, Unsaturated Macrocyclic Dihydroxamic Acid Siderophores Produced by Shewanella putrefaciens Using Precursor-Directed Biosynthesis, *ACS Chemical Biology*, 2014, **9**, 945-956.
10. T. Richardson-Sanchez and R. Codd, Engineering a cleavable disulfide bond into a natural product siderophore using precursor-directed biosynthesis, *Chem Commun (Camb)*, 2018, **54**, 9813-9816.
11. S. Rütschlin and T. Böttcher, Dissecting the Mechanism of Oligomerization and Macrocyclization Reactions of NRPS-Independent Siderophore Synthetases, *Chemistry – A European Journal*, 2018, **24**, 16044-16051.
12. S. Cleto and T. K. Lu, An Engineered Synthetic Pathway for Discovering Nonnatural Nonribosomal Peptides in Escherichia coli, *MBio*, 2017, **8**, e01474-01417.
13. R. Mudhulkar, S. Rajapitamahuni, S. Srivastava, S. V. V. Bharadwaj, V. P. Boricha, S. Mishra and P. B. Chatterjee, Identification of a New Siderophore Acinetoamonabactin Produced by a Salt-Tolerant Bacterium Acinetobacter Soli, *ChemistrySelect*, 2018, **3**, 8207-8211.
14. H. Lee, N. F. Scherer and P. B. Messersmith, Single-molecule mechanics of mussel adhesion, *Proc Natl Acad Sci U S A*, 2006, **103**, 12999-13003.
15. G. P. Maier, M. V. Rapp, J. H. Waite, J. N. Israelachvili and A. Butler, BIOLOGICAL ADHESIVES. Adaptive synergy between catechol and lysine promotes wet adhesion by surface salt displacement, *Science*, 2015, **349**, 628-632.
16. H. G. Upritchard, J. Yang, P. J. Bremer, I. L. Lamont and A. J. McQuillan, Adsorption of enterobactin to metal oxides and the role of siderophores in bacterial adhesion to metals, *Langmuir*, 2011, **27**, 10587-10596.
17. S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez and R. D. Finn, HMMER web server: 2018 update, *Nucleic Acids Res*, 2018, **46**, W200-W204.
18. M. J. Kade, D. J. Burke and C. J. Hawker, The power of thiol-ene chemistry, *J. Polym. Sci. A Polym. Chem.*, 2010, **48**, 743-750.
19. L. Li and Z. Zhang, Development and Applications of the Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC) as a Bioorthogonal Reaction, *Molecules*, 2016, **21**.
20. M. Sandy and A. Butler, Chrysobactin siderophores produced by Dickeya chrysanthemi EC16, *J Nat Prod*, 2011, **74**, 1207-1212.