Genomic Analysis of Siderophore β-Hydroxylases Reveals Divergent Stereocontrol and Expands the Condensation Domain Family

Zachary L. Reitz, Clifford D. Hardy, Jaewon Suk, Jean Bouvet, Alison Butler*

Department of Chemistry & Biochemistry, University of California, Santa Barbara, CA 93106-9510

* to whom correspondence should be addressed; (805) 893-8178; butler@chem.ucsb.edu, orcid.org/0000-0002-3525-7864

PHYSICAL SCIENCES: Chemistry

Running Title: Genomic Analysis of Siderophore ß-Hydroxylases

Key words: siderophore, biosynthesis, hydroxyaspartate, non-heme iron, genomics, phylogenetics, nonribosomal peptide synthetase

Abstract (238 words; max 250 words)

Genome mining of biosynthetic pathways streamlines discovery of secondary metabolites, but can leave ambiguities in the predicted structures which must be rectified experimentally. Through coupling the reactivity predicted by biosynthetic gene clusters with verified structures, the origin of the β-hydroxyaspartic acid diastereomers in siderophores is reported herein. Two functional subtypes of non-heme Fe(II)/α-ketoglutarate-dependent aspartyl β-hydroxylases are identified in siderophore biosynthetic gene clusters, which differ in genomic organization—existing either as fused domains ($I\beta H_{Asp}$) at the carboxy-terminus of a nonribosomal peptide synthetase (NRPS) or as standalone enzymes ($T\beta H_{Asp}$)—and each directs opposite stereoselectivity of Asp β-hydroxylation. The predictive power of this subtype delineation is confirmed by the stereochemical characterization of β-OHAsp residues in pyoverdine GB-1, delftibactin, histicorrugatin, and cupriachelin. The L-*threo* (2*S*, 3*S*) β-OHAsp residues of alterobactin arise from hydroxylation by the β-hydroxylase domain integrated into NRPS AltH, while L-*erythro* (2*S*, 3*R*) β-OHAsp in delftibactin arises from the standalone β-hydroxylase DelD. Cupriachelin contains both L-*threo* and L-*erythro* β-OHAsp, consistent with the presence of both types of β-hydroxylases in the biosynthetic gene cluster. A third subtype of non-heme Fe(II)/α-ketoglutarate-dependent enzymes ($I\beta H_{His}$) hydroxylates histidyl residues with L-*threo* stereospecificity. A previously undescribed, non-canonical member of the 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.

Significance (114 words; max 120 words)

Bacteria produce siderophores to sequester iron(III). Genome mining of nonribosomal peptide synthetases predicts partial structures of peptidic siderophores; however, current tools cannot reliably predict which aspartate and histidine residues will be hydroxylated to form bidentate chelating groups, nor the resulting stereochemistry. We identified two functional subtypes of non-heme Fe(II)/α-ketoglutarate-dependent aspartyl β-hydroxylases in siderophore biosynthetic gene clusters and one type of histidyl β-hydroxylase. Standalone genes encode one class of aspartyl β-hydroxylases and the histidyl β-hydroxylases, while the second class of aspartyl β-hydroxylases is encoded within a domain of a NRPS gene. Each aspartyl β-hydroxylase subtype effects distinct diastereoselectivity. Mapping the β-OHAsp diastereomers in siderophores to the phylogenetic tree of β-hydroxylases enables prediction of β-OHAsp stereochemistry *in silico.*

Introduction

The vast majority of life requires iron as an enzyme cofactor. Because Fe(III) is nearly insoluble under aerobic physiological conditions, bacteria have evolved a variety of mechanisms to meet their metabolic iron needs, including the biosynthesis of small-molecule, high-affinity iron chelators called siderophores (1). Some 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 (SI Appendix, Fig. S1) 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 (SI Appendix, Fig. S1) (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 (SI Appendix, Fig. S1), which contain both β-OHHis and β-OHAsp (9-11). One β-hydroxyasparagine-containing siderophore has been reported, pyoverdine VLB120 from *Pseudomonas taiwanensis* VLB120 (SI Appendix, Fig. S1) (12, 13).

Peptidic siderophores are synthesized by large, multidomain nonribosomal peptide synthetase (NRPS) enzymes (14). The core domains of a NRPS are adenylation (A), thiolation (T), and condensation (C). The catalytic cycle begins with the activation of a select carboxylic acid by the A domain; this substrate, most often an L-amino acid, can be predicted by sequence

analysis (15). After activation, the acid adenylate is transferred to the phosphopantetheine (Ppant) tether of the T domain. Each peptide bond is formed by a C domain, which condenses two T-bound substrates. The growing peptide chain travels down the NRPS "assembly line", and the final peptide is released at the C-terminus of the NRPS, commonly by a thioesterase (Te) domain. Peptide bond-forming C domains are canonical members of the condensation domain superfamily, which also contains other NRPS domains that catalyze diverse reactions (16, 17). The most well studied is the epimerization (E) domain, which catalyzes the conversion of an L-amino acid to a D- amino acid. Other tailoring domains found in NRPS enzymes are responsible for formylation, halogenation, methylation, oxidation, and reduction (14). These modifications all add to the structural diversity of nonribosomal peptides.

Several families of enzymes are responsible for the β-hydroxy amino acids found in nonribosomal peptides, namely Fe-heme monooxygenases (18), diiron monooxygenases (19), and non-heme Fe(II)/α-ketoglutarate-dependent dioxygenases (20). Only this last class of Fe_{NH}/αKG dioxygenases have been found to hydroxylate aspartic acid (20-22). OrbG, encoded in the biosynthetic gene cluster of the siderophore ornibactin (SI Appendix, Fig. S1), was the first enzyme predicted to be an aspartyl β-hydroxylase based on homology to Fe_{NH}/αKG dioxygenases (23). 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, Fig. 1) (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* (20). SyrP is involved in the biosynthesis of syringomycin (SI Appendix, Fig. S2), a phytotoxin produced by *Pseudomonas syringae* (20). 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 (20).

Based on homology to SyrP and OrbG, the existence of a number of β-OHAsp residues in siderophores has been rationalized (7, 11, 24-28) or predicted (29-33). Recently, Kurth and coworkers used *cucF*, a β-hydroxylase domain from the cupriachelin (Fig. 2) gene cluster, as a handle to scan genomes for photoactive Fe(III)-siderophores, leading to the discovery of variochelin (SI Appendix, Fig. S1) (33). Genome mining for β-OHAsp-containing siderophores has been quite successful, but current techniques can leave ambiguity in the predicted structure. Although the taiwachelin (Fig. 2) 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 (29). Some pyoverdines also retain an unmodified Asp (*e.g.*, pyoverdine G4R, Fig. 2) (7, 27). In contrast, both Asp residues are hydroxylated in cupriachelin, serobactin, pacifibactin, and alterobactin (Figures 1 and 2; SI Appendix, Fig. S1) (25, 30, 34).

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 (Fig. 2) (35). 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, we analyzed gene clusters responsible for the biosynthesis of structurally characterized siderophores containing β-OHAsp or β-OHHis. Functional subtypes of

β-hydroxylases emerged, which we corroborate 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. Stereochemical characterization of β-OHAsp residues in pyoverdine GB-1, delftibactin, histicorrugatin, and cupriachelin, reported herein, confirm the predictive power of subtype delineation. We also identify a previously undescribed member of the NRPS condensation domain superfamily which is associated with certain β-hydroxylase subtypes. We name this class of enzymes the *interface* (I) domain for a proposed role in positioning the β-hydroxylase and the NRPS-bound amino acid substrate prior to hydroxylation.

Results and Interpretation

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 (SI Appendix, Tables S1 and S2), of which 26 were isolated from strains with published genomes (SI Appendix, Table S1). The associated biosynthetic gene clusters were extracted and annotated, resulting in a final dataset of 30 β-hydroxylases (Table 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 1). The gene clusters for cupriachelin, serobactin, and pacifibactin each encode two putative aspartyl β-hydroxylases (*e.g.*, SbtH and SbtI1 in serobactin; Fig. 1). 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 1 and elaborated below.

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 (Fig. 3). 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 (17). 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 (Fig. 3) (29).

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 (36). 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 (Table 1). 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 (Fig. 3) (27). 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 (Fig. 3) (24).

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 1). 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; Fig. 3) (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 (SI Appendix, Fig. S2 and Table S3). 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 (SI Appendix, Fig. S2 and Table S3) (37, 38). 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.

Phylogeny and sequence analyses of Asp and His β-hydroxylases

To corroborate the existence of distinct IβH$_{Asp}$, TβH$_{Asp}$, and IβH$_{His}$ amino acid β-hydroxylase functional subtypes, we reconstructed an unrooted phylogenetic tree of 30 known β-hydroxylase protein sequences from the 26 siderophore clusters (Fig. 4). Three members of the SyrP-like family (SβH$_{Asp}$) were also included (SI Appendix, Table S3). All four subtypes formed distinct clades with the exception of IβH$_{Asp}$, which is paraphyletic with respect to SβH$_{Asp}$ (Fig. 4).

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) (39). Crystallographic studies of taurine dioxygenase TauD, 2,4-dichlorophenoxyacetate monooxygenase TfdA, and related enzymes (40-44) have elucidated a 2-His-1-carboxylate facial triad Fe(II) binding motif (45, 46). The α-ketoglutarate cofactor, which coordinates Fe(II), is bound by strictly conserved Thr and Arg residues (47). A multiple sequence alignment of the 33 collected β-hydroxylases across the four subtypes (IβH$_{Asp}$, TβH$_{Asp}$, IβH$_{His}$, and SβH$_{Asp}$) 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** (SI Appendix, Fig. S3).

Phylogeny and Sequence Analyses of Interface Domains

Interface domain amino acid sequences were collected from siderophore biosynthetic clusters, including two domains each from histicorrugatin (Fig. 3) and alterobactin (SI Appendix, Fig. S1), resulting in a dataset of 10 IβH$_{Asp}$-associated and 4 IβH$_{His}$-associated interface domains (Fig. 5; SI Appendix, Table S4). All 14 of the I domains belong to the PFAM condensation domain family PF00668 (39). Rausch and coworkers previously delineated NRPS condensation

domain subtypes and reconstructed their phylogenetic relationship (16). We aligned our extracted I domain sequences to their dataset and reconstructed a maximum-likelihood phylogenetic tree (Fig. 5, Dataset S1). 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 (SI Appendix, Fig. S4), 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 (Fig. 5). 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 (SI Appendix, Fig. S4: residue 228), and the pair is strictly conserved throughout the condensation domain superfamily (48, 49).

Stereochemistry of β-OHAsp residues in siderophores

Many of the β-OHAsp-containing siderophores in Table 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 (50), while the C3 stereochemistry is set during hydroxylation (Fig. 6A). 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* (29, 30, 33, 34, 51-53). 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 1, Fig. 6A). The stereochemistry of ten of 18 siderophores with TβH_{Asp}-mediated hydroxylation has been reported in the literature (3, 25, 26, 30, 32, 34, 54-57). 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, we selected three representative siderophores from the β-hydroxylase phylogenetic tree in Fig. 4 and stereochemically characterized the β-OHAsp residues (Fig. 6B). The biosynthetic gene cluster of pyoverdine GB-1 encodes a TβH$_{Asp}$ enzyme, PputGB1_4087, that is paired with an E domain (Fig. 3). While β-OHAsp was previously confirmed in pyoverdine GB-1, its stereochemistry had not been reported (27). Stereochemical characterization of the pyoverdine from *Pseudomonas putida* GB-1 with Marfey's reagent (1-fluoro-2,4-dinitrophenyl-5-L-alanine amide, FDAA) (58) confirmed the sole presence of D-*threo* β-OHAsp (SI Appendix, Figures S5 – S7).

The hydroxylases from delftibactin, acidobactin, and vacidobactin biosyntheses (DelD, Aave_3734, and Vapar_3747, respectively) form a well-supported clade nested within the TβH$_{Asp}$ family (Figures 4 and 6B; SI Appendix, Fig. S1). None of the hydroxylated products had been stereochemically characterized; however, each of these enzymes act on modules with no E domain, suggesting that β-OHAsp must remain in the L-*erythro* configuration in these three siderophores. To test this prediction, the β-OHAsp residue in delftibactin from *Delftia acidovorans* DSM 39 was investigated and found to be L-*erythro* β-OHAsp (Figures S8 – S10). Histicorrugatin (Fig. 3) and cupriachelin (Fig. 2) TβH$_{Asp}$ enzymes HcsC and CucE likewise form a clade with no E domain (Figures 4 and 6B). Purified histicorrugatin from *Pseudomonas thivervalensis* DSM 13194 was also observed to contain only L-*erythro* β-OHAsp (SI Appendix, Figures S11 – S13). This result led us to re-examine the β-OHAsp stereochemistry in cupriachelin (25). The Marfey's derivatized hydrolysate of purified cupriachelin from

Cupriavidus necator H16 contained two β-OHAsp diastereomers, confirmed as L-*erythro* and L-*threo* (SI Appendix, Figures S14 – S16), in contrast to the previous report of solely L-*threo* β-OHAsp (25). Hydrolysis by DCl (59) confirmed that neither diastereomer was the result of epimerization during hydrolysis (SI Appendix, Fig. S16). Thus, all $T\beta H_{Asp}$ enzymes selectively produce the 3*R* diastereomers L-*erythro* and D-*threo* β-OHAsp (Fig. 6B).

SyrP of syringomycin (SI Appendix, Fig. S2) 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 (20); thus the entire $S\beta H_{Asp}/I\beta H_{Asp}$ clade likely produces L-*threo* β-OHAsp isomers (Fig. 6A). The β-OHHis residue of pyoverdine pf244 (SI Appendix, Fig. S1) 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 (SI Appendix, Fig. S1), 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).

Discussion

In sum, the chemistry of siderophore aspartyl β-hydroxylation follows two divergent routes (Figures 3 and 6, Table 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 (Fig. 5). 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 1, Fig. 6). 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 1, Fig. 6). Remarkably, the two drivers of residue specificity we propose here (*i.e.*, the interface domain, and the T_E-type thiolation domain, Fig. 3) parallel selectivity seen in the cytochrome P450 family of enzymes involved in glycopeptide antibiotic and skyllamycin biosyntheses (60). These similarities, as well as the stereochemistry of β-hydroxylation, bear further consideration.

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 (Fig. S17) (61-63). The final module of the NRPS Tcp12 contains a member of the C domain superfamily called the X domain (Fig. S17), which is missing the HHxxxDG catalytic motif (Fig. 5) (16). 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 (64). 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 (Fig. 5) 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 (64), in contrast to the poorly-conserved active site of the I domain (Fig. 5).

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 (36). Mutational studies show that the T_E motif is required for proper

interaction between the T and E domains (36), suggesting that TβH$_{Asp}$ enzymes may also require the GGDSI motif to properly interact with the T-bound amino acid. Cytochrome P450$_{sky}$ of skyllamycin (SI Appendix, Fig. S2) biosynthesis similarly selects residues for β-hydroxylation by interacting only with specific T domains, as determined by X-ray crystallography (65). 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 (60, 65). TβH$_{Asp}$ 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 TβH$_{Asp}$ family.

Stereochemistry. β-Hydroxylation in siderophore biosynthesis is expected to be under strict stereochemical control. Many siderophore β-OHAsp residues have been stereochemically characterized (Table 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 (Fig. 1), produced by an unsequenced isolate (35). 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 (Fig. 6). The stereochemical determination of β-OHAsp in histicorrugatin (Fig. 3) and delftibactin (SI Appendix, Fig. S1), as well as the stereochemical re-assignment of one β-OHAsp residue in cupriachelin (Fig. 1), supports this mechanism. Each of these siderophore biosynthetic gene clusters encodes a TβH$_{Asp}$ 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 TβH$_{Asp}$-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 (Fig. 6).

The divergent stereochemistry of the Asp β-hydroxylases is consistent with the reactivity of other Fe(II)/α-ketoglutarate-dependent β-hydroxylases. Kutzneride (SI Appendix, Fig. S2) biosynthesis involves two Fe(II)/αKG *glutamyl* β-hydroxylases, KtzO and KtzP, both of which have been characterized *in vitro* (66). 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 (66); 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 (66). 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, 67). 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 (43, 44). Similarly, we suspect that the contrasting β-OHAsp stereochemistry is caused by a difference in L-Asp positioning.

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. The pHMMs for each subtype (Dataset S2) may be incorporated into domain and structure prediction workflows with HMMER3 (68). 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 (SI Appendix, Fig. S2) biosynthesis (66) best match the IβH_{His} pHMM with bitscores of 340 and 224, respectively, well below the cutoff.

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.*, IβH_{Asp}, TβH_{Asp}, and IβH_{His}), placing their reactivity in contrast to SyrP (20) and SyrP-like enzymes (SβH_{Asp}). These newly delineated subtypes, validated by phylogenetic reconstruction (Fig. 4), 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 1).

A hallmark of NRPS-based biosynthesis is the presence of D-amino acids, which are coincident with epimerization domains. 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*. We further tested and confirmed our predictive methods with the stereochemical determination of β-OHAsp residues in pyoverdine GB-1, delftibactin, histicorrugatin, and cupriachelin (Fig. 6B). While IβH$_{Asp}$ domains consistently produce L-*threo* (2*S*, 3*S*) β-OHAsp, the TβH$_{Asp}$ subtype produces the L-*erythro* (2*S*, 3*R*) isomer, which is often epimerized to D-*threo* β-OHAsp (Fig. 6A) (2*R*, 3*R*). The D-*erythro* (2*R*, 3*S*) stereoisomer has not been identified in any siderophore, but would be consistent with an IβH$_{Asp}$ 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.

Materials and Methods

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 (SI Appendix, Table S1) (69). The collection covers siderophore structures published through June 2019. Siderophores produced by unsequenced strains (SI Appendix, Table S2) were excluded from this analysis, with the exception of alterobactins (SI Appendix, Fig. S1). 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.

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).

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, Fig. 4 (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.

Strains and growth conditions

Delftia acidovorans DSM 39 and *Pseudomonas thivervalensis* DSM 13194 were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ), *Pseudomonas putida* GB-1 was obtained from B. Tebo (Oregon Health and Science University), and *Cupriavidus necator* H16 was obtained from S. Parsons (University of California, Santa Barbara). Each was maintained on LB agar plates. For siderophore isolation, *D. acidovorans* was cultured in 1L of acidovorax complex medium (consisting of 0.5 g L^{-1} chelex-treated yeast extract, 1.0 g L^{-1} chelex-treated casamino acids, 2.0 g L^{-1} succinic acid, 2.0 g L^{-1} L-glutamic acid, 0.3 g L^{-1} KH_2PO_4, and 2.0 g L^{-1} MOPS buffer, pH adjusted to 7.2) for 48 h at 30 °C, shaken at 160 RPM. *P. thivervalensis* was cultured in 1 L of casamino acids minimal medium (consisting of 5 g L^{-1} chelex-treated casamino acids, 1.18 g L^{-1} K_2HPO_4, and 0.25 g L^{-1} $MgSO_4 \cdot 7H_2O$) for 144 h at 30 °C, shaken at 160 RPM. *P. putida* and *C. necator* were each cultured in 1 L of casamino acids minimal medium for 67 h at 30 °C, shaken at 160 RPM.

Siderophore isolation

Each culture was pelleted by centrifugation (6000 RPM, 30 min). The resultant supernatant was decanted into a clean 1 L flask, to which 100 g of water-washed Amberlite XAD-4 resin was added. The supernatant was shaken with the resin for 3-4 h at 4 °C, 150 RPM. The resin was then filtered from the supernatant and eluted with 75% methanol in ultrapure water (*P. putida*) or 90% methanol in ultrapure water (*D. acidovorans, P. thivervalensis, C. necator*). The eluent was concentrated *in vacuo* and analyzed by UPLC-ESIMS for the presence of siderophore. To obtain pure siderophore, the concentrated eluent was separated by semi-preparative RP-HPLC (250x20 mm YMC C18-AQ column, 7 mL/min flow rate), employing a gradient of methanol in ultrapure water (+0.05% trifluoroacetic acid): 10%-35% MeOH over 25 minutes (delftibactin), 10-40% MeOH over 30 min (histicorrugatin), 5-30% MeOH over 25 min (pyoverdine GB-1), or 40-80% MeOH over 40 min (cupriachelin).

Amino acid analysis

For each siderophore, ~1 mg was dissolved in 200 µL of ultrapure water. To the siderophore solution was added either 200 µL of 12 M HCl, 200 µL of 20% DCl in D_2O or 200 µL of 55% HI. Each acidified solution was then transferred to a glass ampoule, blanketed with Ar, and sealed. For HCl hydrolyses, ampoules were heated for 4 h at 100 °C. For DCl hydrolyses, ampoules were heated for 6 h at 100 °C. For HI hydrolyses, ampoules were heated for 22 h at 100 °C. After heating, ampoules were opened and crude hydrolysates were transferred to microcentrifuge tubes. Hydrolysates were evaporated and redissolved in ~700 µL ultrapure water 3 times to remove any acid, then brought to a final volume of 100 µL. Hydrolysates were reacted with 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (FDAA, Marfey's reagent) following standard conditions (58).

Derivatized hydrolysates were analyzed by RP-HPLC monitoring at 340 nm on a 250x4.6 mm YMC C18-AQ column, employing gradient elutions of either 15-50% acetonitrile (+0.05 trifluoroacetic acid) in ultrapure water (+0.05 trifluoroacetic acid) over 50 minutes, 1 mL/min flow rate (delftibactin, pyoverdine GB-1 hydrolysates); or 15-50% acetonitrile (no additives) in 50 mM triethylamine phosphate (pH 3.0) over 50 minutes, 1 mL/min flow rate (histicorrugatin hydrolysate). Amino acid identity was confirmed by positive ion mode ESI-MS analysis on a Waters Xevo G2-XS QTof coupled to an ACQUITY UPLC H-Class system. A Waters BEH C18 column was used with a linear gradient of 15-50% CH_3CN (0.1% formic acid) in ddH_2O (0.1% formic acid) over 10 min (delftibactin, histicorrugatin hydrolysates) or a linear gradient of 10-30% CH_3CN (0.1% formic acid) in ddH_2O (0.1% formic acid) over 10 min (pyoverdine GB-1, cupriachelin hydrolysates).

Acknowledgements

Support from NSF CHE-1710761 is gratefully acknowledged. This work made use of the MRL Shared Experimental Facilities supported by the MRSEC Program of the NSF, DMR 1121053. We thank B. Tebo and S. Parsons for providing bacterial strains, and R. Behrens for mass spectrometry support.

References

1. Hider RC & Kong X (2010) Chemistry and biology of siderophores. *Nat. Prod. Rep.* 27(5):637-657.
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. Süssmuth RD & Mainz A (2017) Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed Engl.* 56(14):3770-3821.
15. Chevrette MG, Aicheler F, Kohlbacher O, Currie CR, & Medema MH (2017) SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics* 33(20):3202-3210.
16. 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.
17. 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.
18. 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.

19. 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.
20. 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.
21. Ökesli A, Cooper LE, Fogle EJ, & van der Donk WA (2011) Nine Post-translational Modifications during the Biosynthesis of Cinnamycin. *J Am Chem Soc* 133(34):13753-13760.
22. An L*, et al.* (2018) Substrate-assisted enzymatic formation of lysinoalanine in duramycin. *Nature Chemical Biology* 14(10):928-933.
23. 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.
24. Owen JG & Ackerley DF (2011) Characterization of pyoverdine and achromobactin in Pseudomonas syringae pv. phaseolicola 1448a. *BMC Microbiol.* 11:218.
25. 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.
26. 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.
27. Parker DL*, et al.* (2014) Pyoverdine synthesis by the Mn(II)-oxidizing bacterium Pseudomonas putida GB-1. *Front. Microbiol.* 5:202.
28. 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.
29. Kreutzer MF & Nett M (2012) Genomics-driven discovery of taiwachelin, a lipopeptide siderophore from Cupriavidus taiwanensis. *Org. Biomol. Chem.* 10(47):9338-9343.
30. 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.
31. Skinnider MA*, et al.* (2015) Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* 43(20):9645-9662.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. Johnston CW*, et al.* (2015) An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nat. Commun.* 6:8421.

38. Michelsen CF*, et al.* (2015) Nonribosomal peptides, key biocontrol components for Pseudomonas fluorescens In5, isolated from a Greenlandic suppressive soil. *MBio* 6(2):e00079.
39. Finn RD*, et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279-285.
40. Zhang Z*, et al.* (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* 7(2):127-133.
41. Clifton IJ*, et al.* (2003) Crystal structure of carbapenem synthase (CarC). *J. Biol. Chem.* 278(23):20843-20850.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. Martinez JS*, et al.* (2000) Self-assembling amphiphilic siderophores from marine bacteria. *Science* 287(5456):1245-1247.

56. Stephan H*, et al.* (1993) Ornibactins--a new family of siderophores from Pseudomonas. *Biometals* 6(2):93-100.
57. 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.
58. Marfey P (1984) Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Research Communications* 49(6):591.
59. Hermenau R*, et al.* (2018) Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. *Nat. Chem. Biol.* 14:841-843.
60. 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.
61. 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.
62. 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.
63. 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.
64. 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.
65. 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.
66. 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.
67. 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.
68. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 7(10):e1002195.
69. O'Leary NA*, et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44(D1):D733-745.
70. Reid RT & Butler A (1991) Investigation of the mechanism of iron acquisition by the marine bacterium Alteromonas luteoviolaceus : Characterization of siderophore production. *Limnol. Oceanogr.* 36(8):1783-1792.
71. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-1797.
72. Gouy M, Guindon S, & Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27(2):221-224.
73. Crooks GE, Hon G, Chandonia J-M, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res.* 14(6):1188-1190.
74. Nguyen L-T, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32(1):268-274.

75. 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.
76. 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.
77. Dereeper A*, et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36(Web Server issue):W465-469.
78. Linne U & Marahiel MA (2000) Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* 39(34):10439-10447.

Fig. 1. 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 adapted from reference (30).

Fig. 2. Representative siderophores with β-OHAsp residues. Some Asp residues are not hydroxylated in β-OHAsp-containing siderophores. Variable β-OHAsp stereochemistry has led to further ambiguity in genomic predictions. The L-*erythro* β-OHAsp diastereomer in cupriachelin is reported herein.

Fig. 3. Representative reactivity of amino acid β-hydroxylases in siderophore biosynthesis. Aspartyl β-hydroxylases may be *interface*-associated ($I\beta H_{Asp}$), and interact only with Asp-loading modules with the architecture I-A-T, or T_E-associated ($T\beta H_{Asp}$), and interact only with modules containing a GGDSI motif in the thiolation domain. Siderophore histidyl β-hydroxylases are also interface-associated ($I\beta H_{His}$). The proposed reactivity partners explain the

unmodified Asp residues in taiwachelin and pyoverdine GB-1. Genomic data were retrieved from RefSeq (SI Appendix, Table S1). For NRPS domain abbreviations, see caption of Figure 1.

Fig. 4. 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 (74) using the LG+F+R5 model, as chosen by ModelFinder (Akaike information criterion) (75). 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.

Fig. 5. 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* (16). The full Newick-formatted tree with sequence names is available in Dataset S1. The phylogenetic tree was reconstructed in PhyML v3.0 (76) through http://phylogeny.fr (77). The LG+I+F+G4 model was chosen by ModelFinder (Akaike information criterion) (75). 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 (73), trimming any position with >50% gaps.

Fig. 6. 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 (78), 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. Stereochemistries in bold are reported herein. The phylogenetic tree was produced by midpoint rooting the phylogenetic tree from Figure 4.

Figure 1

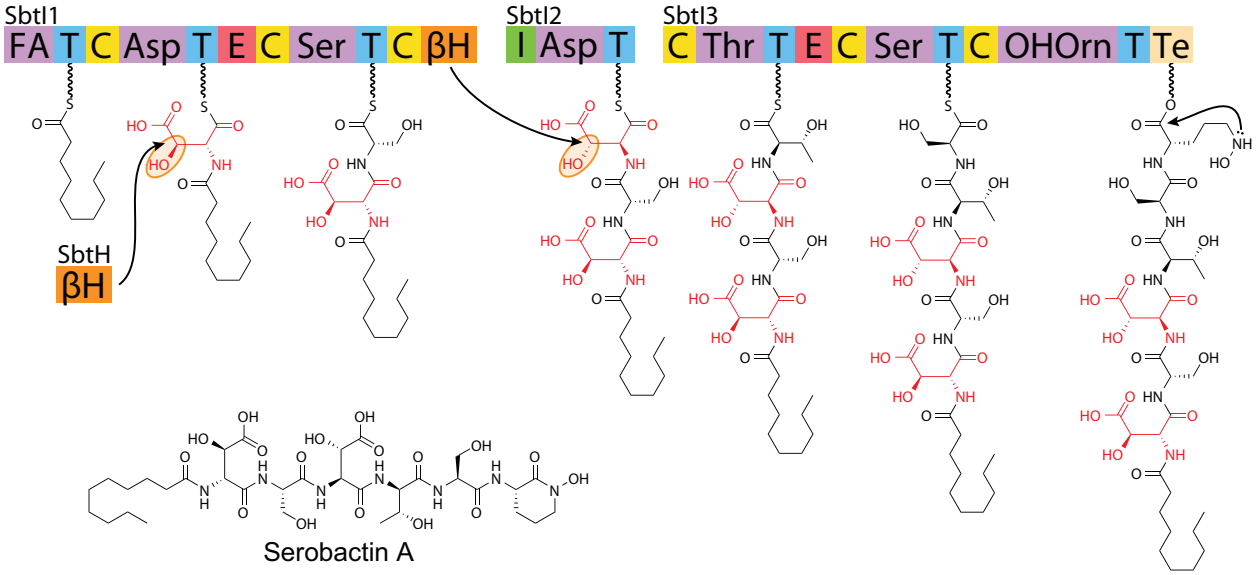

Figure 2

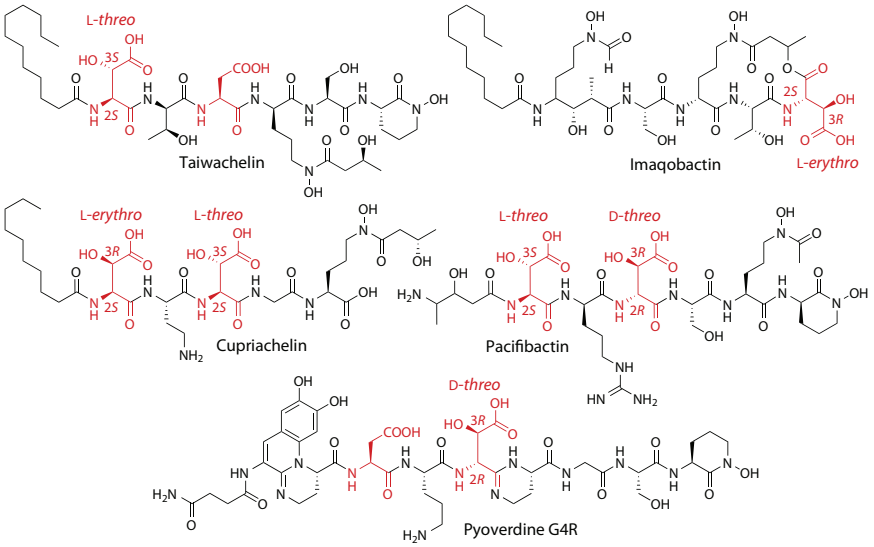

Figure 3

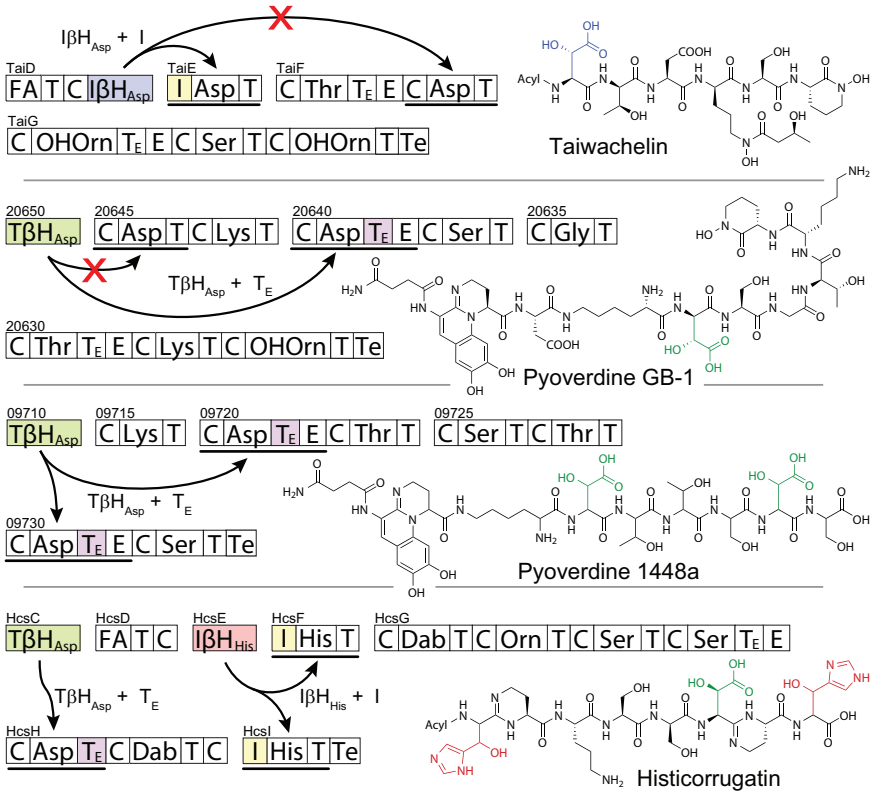

Figure 4

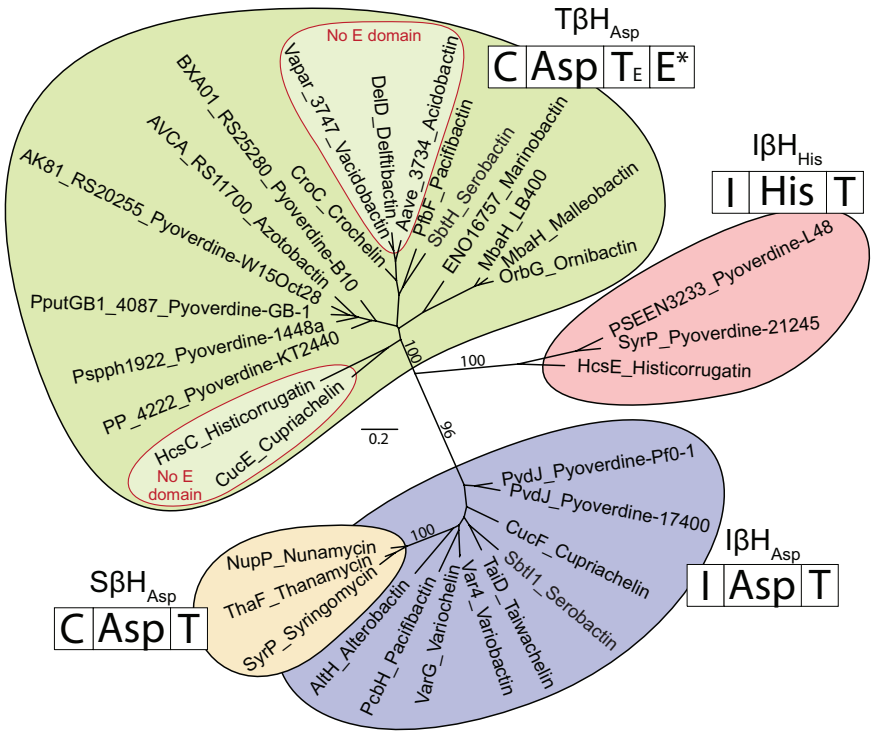

Figure 5

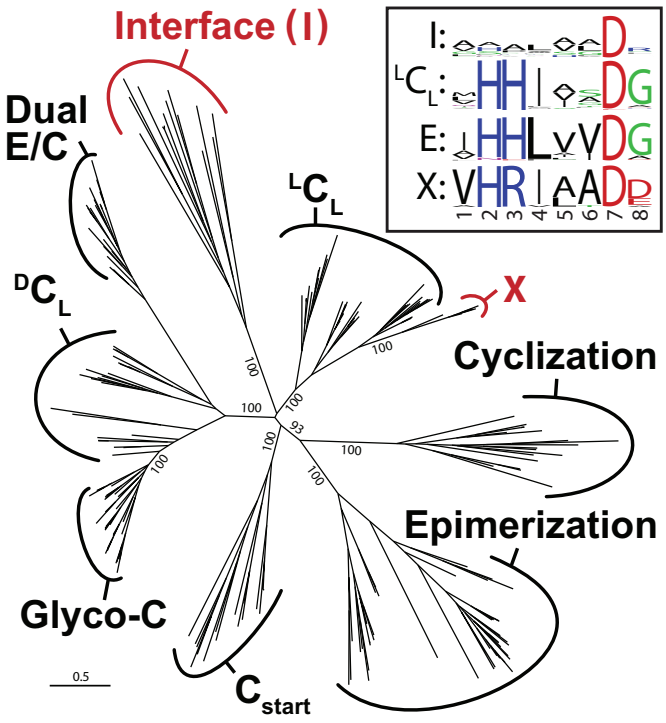

Figure 6

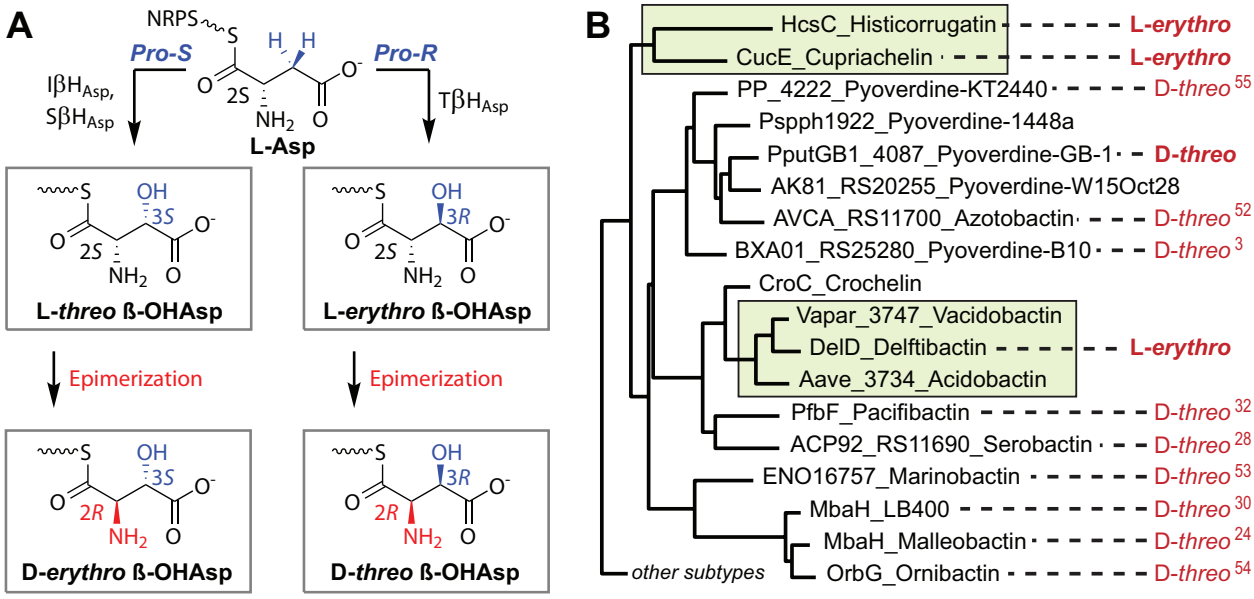
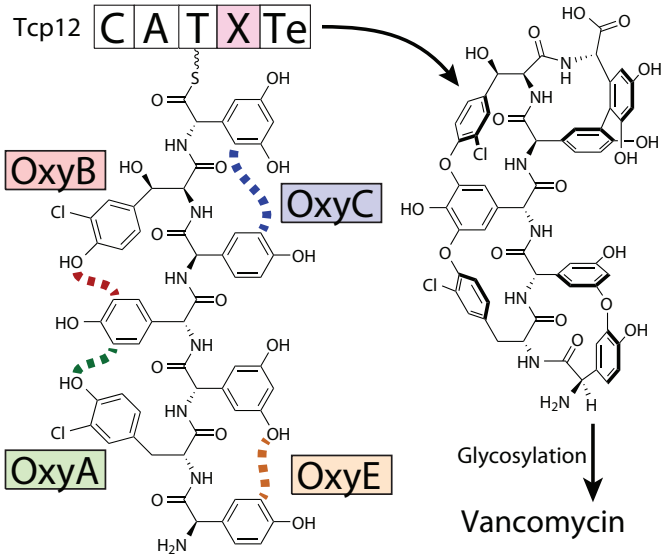
Figure 7