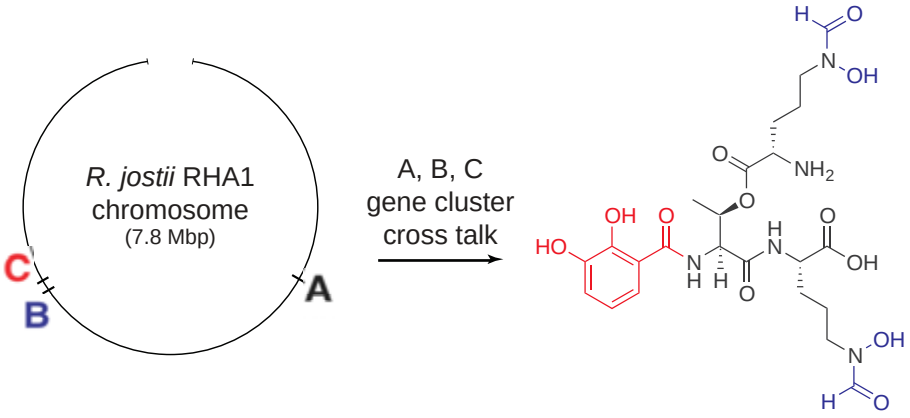
Structural characterization of the siderophore rhodochelin from *Rhodococcus jostii* RHA1 and elucidation of its biosynthetic machinery

---

Strukturelle Charakterisierung des Siderophors Rhodochelin aus *Rhodococcus jostii* RHA1 und Untersuchung seiner biosynthetischen Maschinerie



Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Biologie

der Philipps-Universität Marburg

vorgelegt von

Mattia Bosello

aus Bentivoglio, Italien

Marburg an der Lahn, 2012

Die Untersuchungen zur vorliegenden Arbeit wurden am Fachbereich Chemie der Philipps-Universität Marburg unter der Leitung von Herrn Prof. Dr. Mohamed A. Marahiel durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg als Dissertation angenommen am 19 Juni 2012.

Erstgutachter: Prof. Dr. Mohamed A. Marahiel (Philipps-Universität Marburg)
Zweitgutachter: Prof. Dr. Michael Bölker (Philipps-Universität Marburg)

Tag der mündlichen Prüfung am: 5 Juli 2012

The majority of the work presented herein has been published:

Mattia Bosello, Lars Robbel, Uwe Linne, Xiulan Xie, and Mohamed A. Marahiel
Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in *Rhodococcus jostii* RHA1
Journal of the American Chemical Society **2011** *133* (12), 4587-4595

Mattia Bosello, Andreas Mielcarek, Tobias W. Giessen, and Mohamed A. Marahiel
An enzymatic pathway for the biosynthesis of the formylhydroxyornithine required for rhodochelin iron coordination
Biochemistry **2012** *51* (14), 3059-3066

Additional publications:

Tobias W. Giessen, Kamila B. Franke, Thomas A. Knappe, Femke I. Kraas, Mattia Bosello, Xiulan Xie, Uwe Linne, and Mohamed A. Marahiel
Isolation, structure elucidation, and biosynthesis of an unusual hydroxamic acid ester-containing siderophore from *Actinosynnema mirum*
Journal of Natural Products **2012** *75* (5), 905-914

dedicated to my parents

Table of contents

List of abbreviations

A-domain	adenylation domain
aa	amino acid
Ac-CoA	acetyl coenzyme A
ArCP	aryl carrier protein
ACP	acyl carrier protein
α-KG	α-ketoglutarate
ACV	δ-aminoadipyl-cysteinyl-D-valine
ADP	adenosine diphosphate
AMP	adenosine monophosphate
AT	acetyltransferase
ATP	adenosine-5'-triphosphate
BLAST	Basic Local Alignment Search Tool
BSA	bovine serum albumine
Boc	*tert*-butoxycarbonyl
bp	base pairs
C-domain	condensation domain
Cy-domain	heterocyclization domain
CAS	chromazurol S
CDA	calcium-dependent antibiotic
CoA	coenzyme A
COSY	correlation spectroscopy
ddH_2O	double-distilled water
DHB	2,3-dihydroxybenzoic acid
DAD	diode-array detector
DMF	dimethyl formamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane sodium sulfonate
DTT	dithiothreitol
E-domain	epimerization domain
EDTA	ethylenediaminetetraacetic acid
EIC	extracted ion chromatogram
ESI	electron-spray ionization
F-domain	formylation domain
Fl-CoA	fluoresceinyl-CoA
fhOrn	δ-*N*-formyl-δ-*N*-hydroxyornithine
fOrn	δ-*N*-formylornithine
FTICR	fourier transform ion cyclotron resonance
FA	fatty acid
FAS	fatty acid synthase
FAD	flavin adenine dinucleotide
FDAA	*N*-α-(2,4-dinitro-5-fluorophenyl)-L-alaninamide
FMN	flavin mononucleotide
Fmoc	fluorenylmethyloxycarbonyl
FT-IR	Fourier transform infrared spectroscopy
FPLC	fast protein liquid chromatography
GARF	glycinamide ribonucleotide formyltransferase
H_4F	tetrahydrofolate
haOrn	δ-*N*-acetyl-δ-*N*-hydroxyornithine
haLys	ε-*N*-acetyl-ε-*N*-hydroxylysine
HEPES	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HMBC	heteronuclear multiple bond coherence
hLys	ε-*N*-hydroxylysine
hOrn	δ-*N*-hydroxyornithine
HPLC	high performance liquid chromatography
HR-MS	high-resolution mass spectrometry
HSQC	heteronuclear single-quantum correlation spectroscopy
ICL	isochorismate lyase, isochorismatase
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
LTQ	linear triple quadrupole
MCS	multiple cloning site
MT-domain	methyltransferase domain

mRNA	messenger ribonucleic acid
MS	mass spectrometry
N^5-fH$_4$F	N^5-formyl-tetrahydrofolate
N^5,N^{10}-methenylH$_4$F	N^5,N^{10}-methenyl-tetrahydrofolate
N^5-methylH$_4$F	N^5-methyl-tetrahydrofolate
N^5,N^{10}-methyleneH$_4$F	N^5,N^{10}-methylene-tetrahydrofolate
N^{10}-fH$_4$F	N^{10}-formyl-tetrahydrofolate;
n.a.	not applicable
NAD(P)H	nicotinamide adenine dinucleotide (phosphate)
NFPA	nonafluoropentanoic acid, n-perfluoropentanoic acid
NIS	NRPS independent siderophore
NMO	*N*-hydroxylating flavoprotein monooxygenases
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetase
NTA	nitrilotriacetic acid
NTP	nucleoside triphosphate
Ox-domain	oxidation domain
OD	optical density
ORF	open reading frame
Orn	ornithine
p.a.	per analysis
PCP	peptidyl-carrier-protein
PCR	polymerase chain reaction
PDB	protein data bank
PK	polyketide
PKS	polyketide synthase
ppan	4'-phosphopantetheine
PP$_i$	inorganic pyrophosphate
PPTase	4'-phosphopantetheine transferase
qTOF	quadrupole time-of-flight
R-domain	reductase domain
RNA	ribonucleic acid
ROESY	rotating frame nuclear Overhauser effect spectroscopy
ROS	radical oxygen species
RP	reversed-phase
RT	room temperature
SAM	*S*-denosylmethionine
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIC	selected (single) ion chromatogram
SIM	single ion mode
SOE	splicing overlap extension
spp.	species
T-domain	thiolation domain
TA	transaminase
TE	thioesterase domain
TEII	type II thioesterase
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIC	total ion chromatogram
TOCSY	total correlation spectroscopy
t_R	retention time
TRIS	tris-(hydroxymethyl)-aminomethane
tRNA	transfer ribonucleic acid
UDP	uridine diphosphate
v/v	volume/volume
w/	with
w/o	without
w/v	weight/volume
w.t.	*wild-type*

Table 1 Overview of the proteinogenic amino acids. The three- and one-letter codes are given for each amino acid, as well as the molecular weight.

amino acid	three letter code	one letter code	MW (Da)
alanine	Ala	A	89
arginine	Arg	R	174
asparagine	Asn	N	132
aspartic acid	Asp	D	133
cysteine	Cys	C	121
glutamic acid	Glu	E	147
glutamine	Gln	Q	146
glycine	Gly	G	75
histidine	His	H	155
isoleucine	Ile	I	131
leucine	Leu	L	131
lysine	Lys	K	146
methionine	Met	M	149
phenylalanine	Phe	F	165
proline	Pro	P	115
serine	Ser	S	105
threonine	Thr	T	119
tryptophan	Trp	W	204
tyrosine	Tyr	Y	181
valine	Val	V	117

Summary

Rhodococci represent an important genus of industrial interest, both because of their role in bioremediation and biocatalysis, as well as for their potential as producers of natural products. In this context, the genome sequencing of the biphenyl-degrading soil bacterium *Rhodococcus jostii* RHA1 represents the first attempt to harness the biosynthetic metabolic potential of the genus *Rhodococcus*, by enabling the systematic exploration of its natural product-producing capabilities. The genome of *R. jostii* RHA1 contains 23 secondary metabolite gene clusters, all considered to be orphan with respect to their product, including two clusters putatively involved in siderophore biosynthesis. In this study, the isolation, structural characterization and genetic analysis of the biosynthetic origin of rhodochelin, a unique mixed-type catecholate-hydroxamate siderophore isolated from *R. jostii* RHA1, which represents the first characterized NRPS-derived natural product of the strain, is reported. Structure elucidation of rhodochelin was accomplished via MS^n- and NMR-analysis and revealed the tetrapeptide to contain an unusual ester bond between an L-*δ*-*N*-formyl-*δ*-*N*-hydroxyornithine (L-fhOrn) moiety and the side chain of a threonine residue. Bioinformatic analysis of the *R. jostii* RHA1 genome revealed the enzymes responsible for siderophore biosynthesis to be encoded in three distant NRPS gene clusters. Single gene deletions within the three putative biosynthetic gene clusters abolished rhodochelin production, proving that the ORFs responsible for rhodochelin biosynthesis are located in different chromosomal loci. Biochemical characterization of the monooxygenase Rmo and the formyltransferase Rft established a route for the biosynthesis of the non-proteinogenic amino acid L-fhOrn, prior to its incorporation into the peptide scaffold by the NRPS-assembly line. The insights gained from the structural and functional characterization of rhodochelin, together with the genetic and biochemical characterization of the respective biosynthetic gene clusters, allowed the proposal of a biosynthetic model for rhodochelin assembly. Finally, the efficient and, in this work, first reported cross-talk between three distantly located secondary metabolite gene clusters provides deep insights into natural product biosynthesis that may facilitate future attempts to isolate new natural products.

Zusammenfassung

Bei den Rhodococci handelt es sich um ein bakterielles Genus von industrieller Relevanz, welches bei der biologischen Dekontaminierung und Biokatalyse zum Einsatz kommt und dessen Mitglieder großes Potential als Produzenten neuer Naturstoffe zeigen. In diesem Zusammenhang stellt die Genomsequenzierung des Biphenyl-abbauenden Bodenbakteriums *Rhodococcus jostii* RHA1 den ersten Versuch dar das biosynthetische Potential des Genus *Rhodococcus* auszuloten, da das Vorliegen der kompletten Genomsequenz die systematische Erforschung der Naturstoffproduktion erlaubt. Das Genom von *R. jostii* RHA1 enthält 23 Sekundärmetabolit-Gencluster, darunter zwei putative Siderophor-Biosynthesecluster, wobei alle als „orphan“ zu bezeichnen sind, da ihnen kein konkretes Produkt zugeordnet werden kann. In der vorliegenden Arbeit soll die Isolierung, strukturelle Charakterisierung und genetische sowie biochemische Analyse des biosynthetischen Ursprungs von Rhodochelin, dem ersten aus *R. jostii* RHA1 isolierten „mixed-type“ Catechol-Hydroxamat Siderophore, welches das erste charakterisierte NRPS-abhängige Naturprodukt dieses Stamms darstellt, behandelt werden. Zur Strukturaufklärung von Rhodochelin wurden sowohl MS^n- als auch NMR-Studien durchgeführt, welche ergaben, dass es sich um ein Tetrapeptid handelt, das eine ungewöhnliche Esterbindung zwischen einem L-*δ*-*N*-Formyl-*δ*-*N*-hydroxy-Ornithinrest (L-fhOrn) und einer Threoninseitenkette enthält. Eine bioinformatische Analyse des *R. jostii* RHA1 Genoms zeigte, dass die für die Biosynthese verantwortlichen Gene in drei unterschiedlichen, voneinander weit entfernten NRPS-Genclustern lokalisiert sind. Durch Einzelgendeletionen in den jeweiligen Clustern, durch welche die Rhodochelinproduktion komplett aufgehoben wurde, konnte eindeutig gezeigt werden, dass die für die Rhodochelin-Biosynthese verantwortlichen Gencluster in drei unterschiedlichen Loci auf dem *R. jostii* RHA1 Chromosom vorliegen. Durch die biochemische Charakterisierung der Monooxygenase Rmo und der Formyltransferase Rft konnte ein Biosyntheseweg für die nicht-proteinogene Aminosäure L-fhOrn etabliert werden, welche anschließend durch die NRPS-Maschinerie in das Peptidgerüst eingebaut wird. Mit Hilfe der durch die strukturelle und funktionelle Charakterisierung von Rhodochelin, sowie der genetischen und biochemischen Analysen der verantwortlichen Biosynthese-Gencluster gewonnen Einsichten, konnte eine Biosyntheseroute für das Siderophor Rhodochelin postuliert werden. Die aus der vorliegenden Arbeit gewonnenen Erkenntnisse bezüglich des effizienten und vormals unbekannten „cross-talks“ zwischen drei weit voneinander entfernten Sekundärmetabolit-Genclustern

erlauben neue Einblicke in die Organisation des bakteriellen Sekundärmetabolismus und tragen zu einem besseren Verständnis der Biosynthese von Naturstoffen bei. Des Weiteren können die gewonnenen Ergebnisse als Ausgangspunkt für eine zukünftige Isolierung neuer Naturstoffe dienen.

Chapter 1

Introduction

1.1 Siderophore-based iron acquisition

1.1.1 The biological role of iron

Under physiological conditions, iron exists in two redox forms: Fe^{2+} (ferrous iron) and Fe^{3+} (ferric iron), easily convertible into each other under acidic or basic conditions [E_0^{acid} (Fe^{2+}/Fe^{3+}) = 0.771 V, E_0^{basic} (Fe^{2+}/Fe^{3+}) = -0.690 V]. These redox properties put the element in a central and extremely versatile position for almost the entire spectrum of biological processes needing a redox potential between -0.5 and 0.6 V.[1] In fact, being the fourth most common element on Earth (and the second most common metal), iron takes part as an essential cofactor in many enzymes of cellular metabolism (photosynthesis, Krebs cycle, respiratory chain, nitrogen fixation and methanogenesis among others) since the early days of anaerobic life on the planet. Under aerobic conditions, soluble $Fe^{(II)}$ spontaneously oxidizes to $Fe^{(III)}$ which, in the presence of oxygen, water and at neutral pH, forms insoluble ferric oxide hydrate complexes, leading to a free $Fe^{(III)}$ concentration of only up to 10^{-18} M.[2] In addition, in the presence of molecular oxygen, iron can react to form extremely toxic reactive oxygen species (ROS), in the well-known Fenton reaction:

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^{\bullet} + OH^-$$
$$Fe^{3+} + H_2O_2 \rightarrow Fe^{2+} + OOH^{\bullet} + H^+$$

These radical species are subsequently able to damage cellular structures, *e.g.* nucleic acids and membranes, severely impairing cellular functions.[3] Thus, under oxidative conditions, iron is both rarely available and extremely toxic.

The development of high-affinity iron uptake systems and the precise regulation of the iron homeostasis is therefore an essential process to sustain cellular life. In particular, in order to cope with iron-limiting conditions, microbes have developed mechanisms for highly selective metal uptake.[4] The secretion of low-molecular weight organic chelators called siderophores is one of the main iron-mobilizing strategies used by both environmental and pathogenic strains to support their life under strict iron-liming conditions.[5]

1.1.2 Siderophore classification

The production and secretion of siderophores is the most efficient and widespread iron-scavenging strategy used by microorganisms (bacteria and fungi) to mobilize iron from iron-depleted environments.[6] Most siderophores display a molecular mass below 1 kDa, coordinate the ferric iron via six donor atoms as an octahedral

complex (in an ferric iron:siderophore ratio of 1:1) and have an extremely high affinity (K_f = 10^{22} - 10^{49} M^{-1}).[2] After secretion to the extracellular space and complexation of iron, the ferric iron-siderophore complex is subsequently selectively and actively imported into the intracellular space, the iron is released from the chelator-complex and channeled to the intracellular targets.[3,7,8] Although they perform the same biological function, siderophores are structurally diverse natural products and display great chemical diversity, in both iron coordination and their biosynthesis.[9] Siderophores may have a linear, exocyclic, endocyclic or tripodal structure.[10] On the other hand, on the basis of the chemical nature of the moieties involved in ferric iron coordination, they are usually classified into three main classes: catecholates (better termed as "aryl caps"), hydroxamates, and (α-hydroxy)-carboxylates. The most commonly known catecholate siderophores are the tris-cathecolates enterobactin[11] (*Escherichia coli*) and bacillibactin[12] (*Bacillus subtilis*) which possess the highest known affinity constants for the siderophore:metal complex, with (Fe:enterobactin)$^{3-}$ being 10^{49} M^{-1} and (Fe:bacillibactin)$^{3-}$ being 10^{47} M^{-1}, respectively. Tris-hydroxamate siderophores coelichelin[13] [*Streptomyces coelicolor* A3(2)] and erythrochelin[14] (*Saccharopolyspora erythraea*) contain iron-coordinating hydroxamic acid moieties, deriving from the subsequent tailoring of hydroxylated side chain amino groups of lysine or ornithine residues, via acetylation or formylation.[15] Moreover, amphibactins[16] and aquachelins[17] are additional examples of tris-hydroxamate amphiphilic siderophores isolated from the marine environment. Finally, carboxylate-type siderophores generally utilize α-hydroxycarboxylic acids as bidentate iron-chelating group. In the case of staphyloferrin A,[18] two citrate groups provide the iron-coordinating moieties.

However, the continuous discovery of new structures led to a more complex classification, due to the presence of at least two different coordinating groups within one molecule, resulting in "mixed-type" siderophores. A representative overview of the structural diversity of siderophores is shown in Figure 1.1.

1.1.3 Siderophore assembly strategies

In addition to the classification relying on the different iron-coordination functions, siderophores diversity can be further organized on the basis of their biosynthetic origin in NRPS-dependent (NRPS, non-ribosomal peptide synthetase) and NRPS-independent.[19,20] NRPS-dependent is the most prominent and extensively studied assembly strategy. Notable examples are the tris-catecholate enterobactin[11] and the tris-hydroxamate coelichelin,[13] whose genetic and biochemical analysis has allowed for the elucidation of alternatives strategies to the canonical linear NRPS assembly

logic (the detailed biosynthetic mechanism will be discussed in the sections 1.2.3.2 and 1.2.3.3).

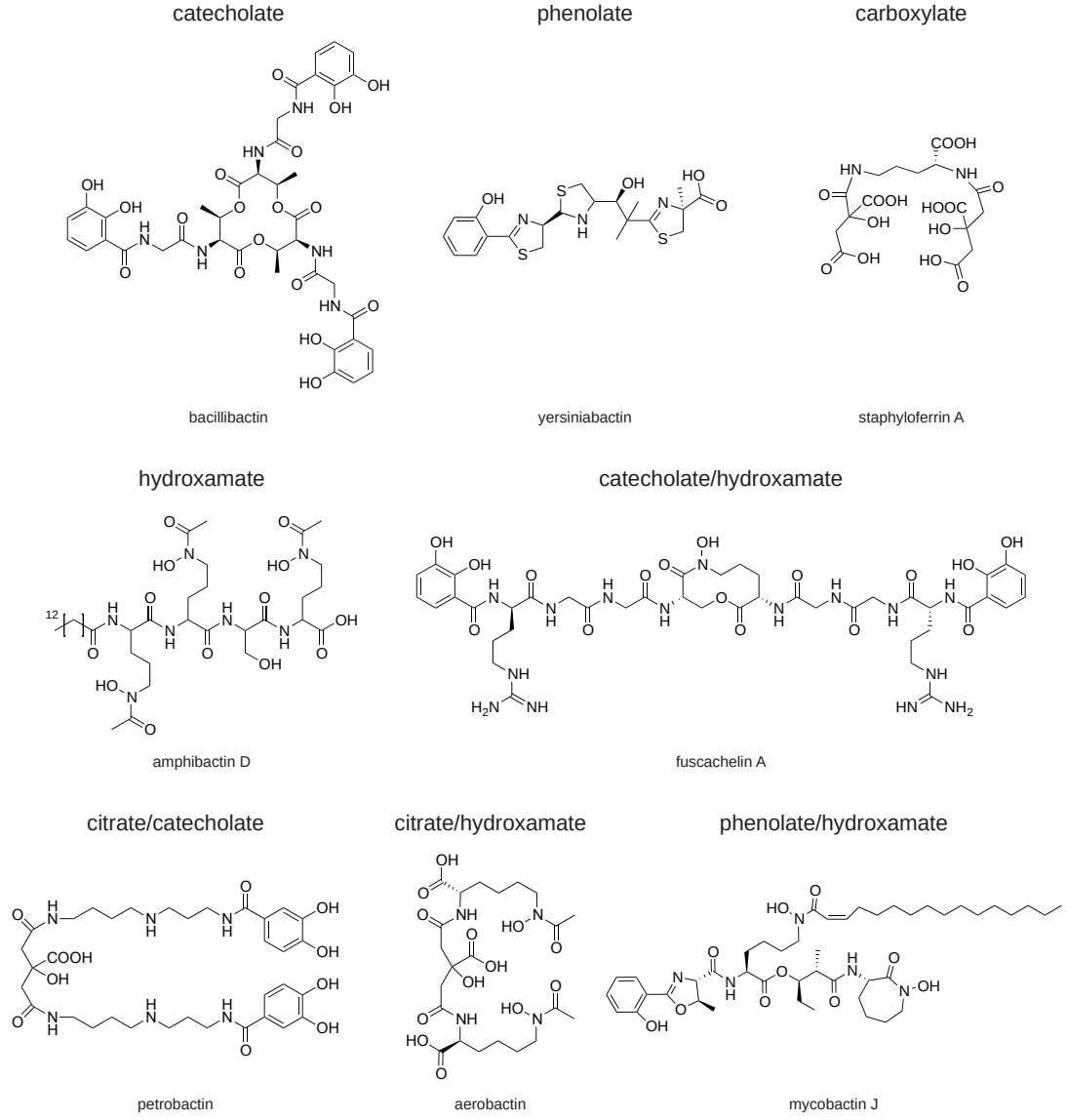


Figure 1.1 Structures of representative siderophores. The classification is based on the functional groups involved in octahedral coordination of ferric iron: catecholate/phenolate, carboxylate and hydroxamate (and the corresponding mixed structures). The siderophores are: bacillibactin (*Bacillus subtilis*), yersiniabactin (*Yersinia pestis*), staphyloferrin A (*Staphylococcus* spp.), amphibactin D (*Vibrio* sp. R-10), fuscachelin A (*Thermobifida fusca* YX), petrobactin (*Bacillus anthracis*), aerobactin (*Shigella flexneri* 5 str. 8401) and mycobactin J (*Mycobacterium tuberculosis*).

The second class of siderophores is synthesized following an NRPS-independent pathway. The first non-ribosomally independent siderophore (NIS) to be discovered was aerobactin, a mixed citrate-hydroxamate type siderophore isolated from different Gram-negative bacteria, including *E. coli*, *Shigella*, *Yersinia* and *Salmonella*

spp. The gene cluster responsible for the biosynthesis of aerobactin is composed of four genes (Figure 1.2).[21] Analysis of the aerobactin gene cluster revealed *iuc*D to encode an FAD-dependent monooxygenase that catalyzes ε-amino group hydroxylation of L-Lys to ε-*N*-L-hydroxylysine (L-hLys).[22] This hydroxylated intermediate is further tailored to ε-*N*-L-acetyl-*N*-L-hydroxylysine (L-haLys) by the acetyltransferase IucB, under the consumption of acetyl-CoA.[23] Then, in the postulated model for aerobactin biosynthesis, it is proposed that the synthetase IucA mediates the first acylation of the α-amino group of L-haLys with a prochiral carboxyl group of citric acid, subsequently followed by the second IucC-catalyzed acylation of the remaining citric acid carboxyl group with an additional L-haLys building block. The homology of IucA and IucC with other biochemically characterized NISs associated with the biosynthesis of desferrioxamines and putrebactin, further suggested these reactions to require nucleotide triphosphates (NTPs) as cosubstrates for the activation of carboxyl groups.[24,25] Furthermore, as IucA and IucC constitute the prototypical examples of NISs, every NIS-dependent biosynthetic pathway involves at least one synthetase homologous to them.[26]

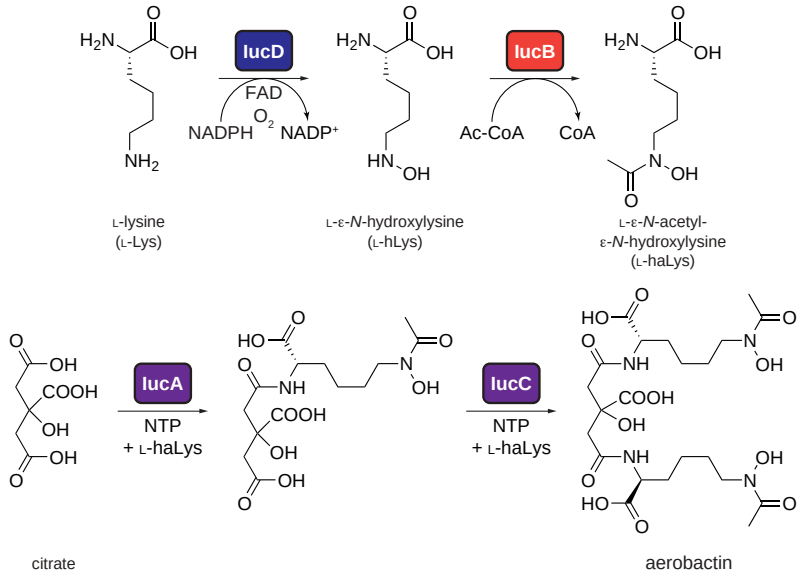


Figure 1.2 The postulated pathway for aerobactin biosynthesis. The FAD-dependent monooxygenase IucD hydroxylates the ε-amino group of L-Lys, subsequently acetylated to L-haLys by the IucB acetyltransferase. The tailored building block is condensed with a carboxyl group of citric acid by IucA. Analogously, IucC catalyzes the second condensation reaction to give rise siderophore aerobactin. In analogy with other NIS synthetases, it is proposed that the condensation reactions take require the consumption of NTPs for the activation of the citrate's carboxylic functions.

1.2 The non-ribosomal assembly of peptides

Non-ribosomal peptide synthetases (NRPSs) are large multimodular megaenzymes that catalyze the biosynthesis of biologically active peptides.[27] In contrast to ribosomal peptide synthesis, the assembly of the oligopeptide is carried out in an mRNA-independent function.[28] NRPS encoding genes are widely spread in nature, being mainly found in bacteria and fungi.[29] The best examined producing species of NRPS-assembling natural products are soil actinobacteria, who have already proved to be a rich source of new potent and pharmacologically-relevant molecules.[30,31]

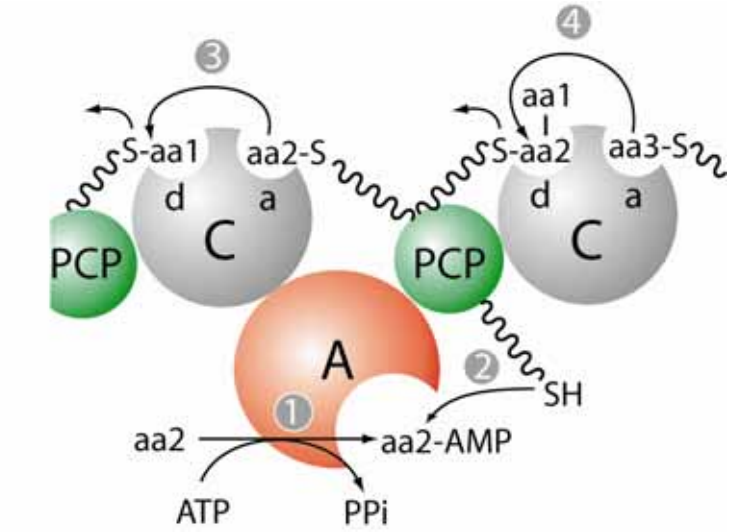


Figure 1.3 The catalytic steps of an elongation NRPS module are shown: (1) Substrate recognition and activation by the A-domain under ATP-consumption, (2) Substrate transfer onto the 4'-phosphopantetheine (ppan) cofactor covalently attached to an invariant Ser of a PCP-domain, (3) trapping of the thioesterified amino-acid in the acceptor site of the C-domain, followed by condensation with the incoming building block trapped in the donor-site, (4) trapping of the peptidyl-*S*-ppan moiety in the donor-site of the downstream C-domain.

The genetic and biochemical characterization of the NRPS assembly machinery revealed a multimodular organization, which can be further dissected into single catalytic domains.[32] Each module is responsible for the incorporation of a building block, catalyzing the elongation of the oligopeptide chain by one unit. In particular, a minimal module (Figure 1.3) contains all the essential units required for the recognition and activation of the monomer [adenylation (A)-domain], the formation of the peptide bond [condensation (C)-domain] and the translocation of the peptidyl intermediates to the subsequent module [peptidyl-carrier-protein (PCP)-domain]. A fourth catalytic NRPS domain, the thioesterase (TE)-domain is responsible for the release of the product from the NRPS machinery. Recently the crystal structure of SrfA-C, the termination module of the surfactin synthetase, was determined,

granting insights into unique inter-domains communication (Figure 1.4).[33] The C-domain and the core of the A-domain constitute a catalytic platform on which the PCP-domain and the smaller C-terminal A_{sub}-domain are located. The terminal TE-domain is in close contact with the upstream PCP-domain, via a short linker region.

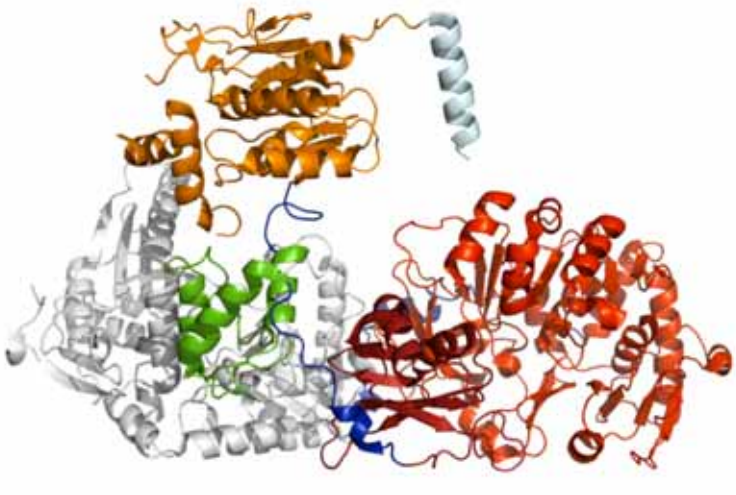


Figure 1.4 The crystal structure of the termination module (C-A-PCP-TE) of the surfactin NRPS SrfA-C (PDB code: 2VSQ). The C-domain (grey) and the A_{core}-domain (red) constitute the work bench on which the PCP-domain (green) and the smaller C-terminal A_{sub}-domain (dark red) are located. The terminal TE-domain (orange) is in close contact with the upstream PCP-domain. The inter-domain linker regions are highlighted in blue. The light-cyan C-terminal helix is the peptidic fusion-tag for protein purification.

1.2.1 The essential NRPS domains

1.2.1.1 The adenylation domain

The adenylation (A)-domain is responsible for the recognition and activation of the substrate.[27] Each A-domain usually consists of ~550 amino acids and comprises a larger N-terminal core (A_{core}, ~450 amino acids) and a smaller C-terminal subdomain (A_{sub}, ~100 amino acids). This domain organization is shared along the three known A-domain structures (PheA, DhbE and DtlA) and is also extended to other adenylate-generating enzymes, like the acetyl-CoA synthase ACS from *Salmonella enterica* and the firefly luciferase from *Photinus pyralis*.[34-38]

Prior catalyzing the activation reaction, the A-domain is responsible for substrate recognition. The analysis of the crystal structure of the dissected A-domain from the gramicidin S synthetase PheA gave insights into the relative position of the amino acid substrate within the binding pocket.[34] This allowed the definition of an A-domain specificity-conferring code (given by the ten amino acids surrounding the substrate binding pocket), and enabled the development of bioinformatic prediction

tools which facilitated the identification and the isolation of NRPS-derived natural products via genome mining approaches.[39-44]

After the recognition of the proper building block, in the presence of Mg^{2+} and ATP, the A-domain activates the amino acid substrate as an adenylate, with the subsequent release of pyrophosphate. Although this reaction is analogous to the one carried out by aminoacyl-tRNA synthetases during ribosomal protein synthesis, the enzymes do not share any sequential or structural homology. Furhtermore, A-domains lack a proof-reading mechanism that, combined with relaxed substrate specificity, often results in the synthesis of NRPs with different amino acid composition by one synthetase.[45]

1.2.1.2 The peptidyl-carrier-protein domain

The peptidyl-carrier-protein (PCP)-domain, also known as thiolation (T)-domain, is responsible for the covalent tethering of the monomeric building blocks and the translocation of the peptidyl intermediates.[46] It is usually located at the C-terminus of an A-domain and, despite its size of ~80 amino acids, it is of the greatest importance for the functionality of the NRPS assembly line. PCP-domains are post-translationally modified at a highly-conserved Ser redidue embedded in the core-T motif (GGxS) with a 4'-phospopantetheine (ppan) cofactor. In fact, during non-ribosomal peptide synthesis, the mobile ppan-arm delivers all the substrates and peptidyl intermediates as thioesters to the adjacent NRPS domains for the formation of the peptide bond, for modification of the PCP-bound substrate or for product release.

The conversion from the inactive *apo*-PCP to the active *holo*-PCP is mediated by phosphopantetheinyl-transferases (PPTases, *e.g.* Sfp), which catalyze the nucleophilic attack of the hydroxyl-group of the conserved Ser residue onto the β-phosphate of a donor coenzyme A molecule (and the subsequent release of 3',5'-adenosinediphosphate).[47,48]

PCPs share a high degree of sequential and structural homology with acyl-carrier-proteins (ACP) of fatty acid (FAS) and polyketide synthases (PKS).[49] NMR-based studies showed that the PCP-domain adopts a four-helix bundle structure and exists in three different conformational states: the *apo* (A), the *holo* (H) and the A / H form.[50] When the PCP is in the *apo*-state, both A and A / H coexist. On the other hand, when the PCP is in the *holo*-state, it slowly interconverts between the H and A / H states. These extensive conformational changes evidence the dynamic nature of the PCP carrying the ppan-bound substrates and intermediates to the adjacent domains. In particular, the terminal sulfhydryl-group of the ppan-arm is able to move

approximately 16 Å, confirming the long-proposed swinging mode of the ppan prostetic group during the non-ribosomal peptide synthesis.

1.2.1.3 The condensation domain

The condensation (C)-domain is the last essential domain of an NRPS module and carries out the peptide bond formation.[51] The C-domain contains an acceptor and a donor site, which harbor the nucleophilic aminoacyl-*S*-PCP substrate and the electrophilic peptidyl-*S*-PCP electrophilic substrate, respectively.[52,53] The formation of the peptide bond is initiated by the nucleophilic attack of the α-amino group of the aminoacyl-*S*-PCP substrate onto the thioester bond of the peptidyl-*S*-PCP. Upon the amid bond formation, the elongated peptide is transferred onto the downstream PCP-domain and serves as a donor substrate in the downstream condensation step. All C-domains have been found to operate unidirectionally, translocating the growing peptide chain towards the C-terminus.[53,54] In addition, following to the prototypical co-linearity assembly rule of non-ribosomal peptides, the number of C-domains is in agreement with the number of peptide bonds found in the mature NRP.

C-domains contain approximately 450 amino acids and are composed of two big similar subdomains arranged in a V-shaped canyon-like structure, of which the N-terminal one shares high sequence and structural homology with the chloramphenicol acetyltransferases.[55-57] This characteristic V-shaped structure allows the correct positioning of the up- and downstream PCP-domains at each opening (acceptor and donor site), with respect to the highly conserved catalytic His residue of the HHxxxDG motif, which remains at the bottom of the canyon. Althought the exact reaction mechanism has not yet been elucidated, it is suggested that the second His residue takes part in the deprotonation of the α-amino group of the aminoacyl-*S*-PCP substrate, enhancing the electron-donor character of the nucleophilic PCP-bound substrate, and therefore facilitating the reaction.

1.2.1.4 The thioesterase domain and the termination of non-ribosomal peptide assembly

The thioesterase (TE)-domain is the fourth essential domain of NRPSs and is usually located in the termination module of the assembly line. TE-domains catalyze the product release from the NRPS, resulting in a linear, cyclic or branched cyclic peptide.[58,59] These independently working domains contain approximately 230-270 amino acids arranged in an α/β hydrolase fold, similarly to serine proteases and lipases.[60-62] The catalytic Ser-His-Asp triad is located in a deep pocket, shielded from

solvent by an α-helical lid region, or by the peptide itself. The hydrolytic release of the template-bound NRP occurs in a two-step mechanism: first the formation of an acyl-*O*-intermediate with the active Ser, which is subsequently cleaved by the attack of a nucleophile.[63] The different nature of the nucleophilic group results in the different topology of the molecule: if an intramolecular attack occurs, a cyclic (or a branched cyclic) molecule is released. On the other hand, if a water molecule cleaves the intermediate, a linear peptide is released.[64-66]

The release of the peptide chain from the NRPS assembly machinery can also occur via different alternative strategies, for example via the reduction of the C-terminal carboxyl group to the corresponding aldehyde or alcohol, in an NAD(P)H dependent manner.[67,68] This reaction is catalyzed by a C-terminal reductase (R)-domain, which takes the place of the TE-domain. In addition, in several NRPSs, product release is carried out by a C-terminal C-domain, proposed to mediate the cyclorelease of the peptide. Furthermore, the existence of *in trans* acting TE-domains has been postulated for NRPSs that lack any C-terminal domain for the release of the NRP product.[13,69]

1.2.2 Additional NRPS domains and related enzymes

The structural and functional diversity of NRPS-derived natural products is usually extended by the addition of auxiliary domains to the essential core functions. *in cis* operating NRPS domains are responsible for the on-line modification of the peptidic backbone, whereas other stand-alone domains modify and tailor NRPS building blocks prior to the incorporation into the assembly line. Finally, a repair mechanism of the assembly machinery carries out in the regeneration of misprimed PCP-domains.

1.2.2.1 In cis *operating modification enzymes*

1.2.2.1.1 The epimerization domain

The presence of D-configured amino acids can be observed in numerous NRPS-derived compounds.[64,70] The incorporation of D-configured amino acids in the peptidic backbone is usually mediated by epimerization (E)-domains, located directly downstream of the adjacent PCPs, and represents one of the major differences between NRP and ribosomal peptide biosynthesis. E-domains epimerize L-configured amino acids immediately after their immobilization as aminoacyl-*S*-PCP intermediates. The E-domains of initiation modules generate a mixture of PCP-*S*-L/D-monomers, later correctly selected by the downstream C-domain prior to peptide bond formation. On the other hand, for E-domains embedded in elongation

modules, the epimerization occurs before the transfer to the subsequent module of the assembly line. Also in this case the downstream C-domains ensure the correct D-configured substrate specificity.[71,72]

1.2.2.1.2 The cyclization domain

The heterocyclization of cysteine, serine or threonine side chains resulting in the corresponding five-membered thiazoline, oxazoline or methyloxazoline heterocycles is a structural feature found in several NRPs. These modifications increase the diversity of the natural product, rigidifing the peptide backbone and providing either metal-chelating or intercalating properties.[73-76] The cyclization (Cy)-domains are responsible of these structural modifications and are variants of C-domains. First, Cy-domains catalyze the usual peptide bond formation and then carry out cyclization of the nucleophilic sidechain of cysteine or the hydroxyl sidechain of serine or threonine onto the newly formed peptide bond. The newly formed heterocycles are often associated with oxidation (Ox)-domains, that catalyze the FMN-dependent two-electron oxidation of the thiazoline or oxazoline ring structures to the thermodynamically more stable corresponding thiazoles or oxazoles.[73,77] Conversely, the reduction of thiazoline or oxazolines structures is mediated by *in trans* operating NAD(P)H-dependent reductases that recognize and directly reduce the PCP-bound substrate.[78]

1.2.2.1.3 The methylation domain

Methylation (MT)-domains catalyze the *in cis* transfer of a methyl group from a S-adenosylmethionine (SAM) donor to a carbon, nitrogen or oxygen atom of the NRP chain. Therefore, on the basis of the different site of methylation, MT-domains are classified as C-MT, N-MT or O-MT, respectively.[79] MT-domains share a bidomain structure, with the first subdomain containing the binding site for methyl group donor, while the second subdomain harbors the pocket for the acceptor substrate. In contrast to the domains described so far, MT-domains are usually embedded within the corresponding A-domains, between the core A_8- and A_9-motifs, separating the A_{core}, and A_{sub},subunits.

1.2.2.1.4 The formylation domain

The *N*-formylation of the N-terminal α-amino group of a non-ribosomal peptide chain is catalyzed by the formylation (F)-domain, through the transfer of a formyl group from the N^{10}-fH_4F or N^5-fH_4F donor cosubstrate to the α-amino function of the activated aminoacyl-*S*-PCP substrate. The linear gramicidin NRPS LgrA contains a

F-domain in the initiation module, upstream of the corresponding A-domain (Figure 1.5).[67] The F-domain catalyzes the reaction on the PCP-bound amino acid (L-Val) and its *N*-formylation is essential for the subsequent elongation of the peptide.[80]

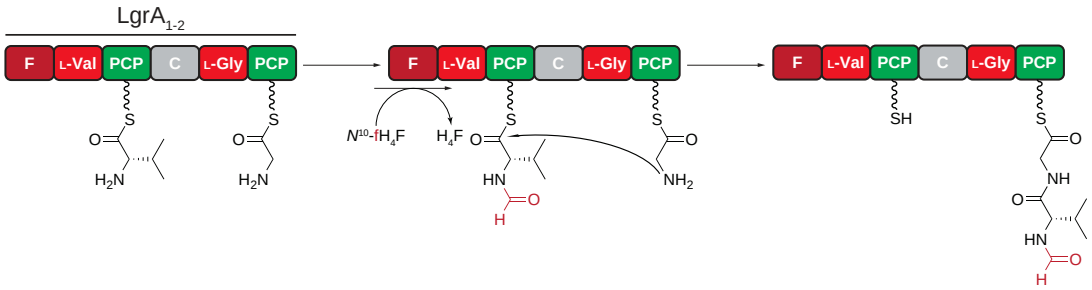


Figure 1.5 The F-domain of the linear gramicidin NRPS LgrA is located at the N-terminus of the initiation module. Upon activation of the L-Val substrate, the F-domain catalyzes the transfer of a formyl moiety from the N^{10}-fH_4F or N^5-fH_4F cosubstrate to the α-amino function of the activated L-Val-*S*-PCP. This reaction is required for the subsequent elongation step. In the picture, the dissected NRPS LgrA lacks the terminal E-domain.

1.2.2.2 Modifications through in trans *acting tailoring enzymes*

1.2.2.2.1 Methylation

Similar to *in cis* acting methyltransferases, stand-alone methyltransferases catalyze the *in trans* SAM-dependent (or N^5-methylH_4F-dependent) methylation of carbon, nitrogen, oxygen or sulphur atoms of building blocks, prior to their incorporation by the NRPS assembly line.[81] Interestingly, these enzymes generate the methylated amino acid by a two-step mechanism, by first methylating the corresponding α-ketoacid, followed by the transamination of the corresponding α-keto group, resulting in the correct building block (Figure 1.6).[82-84]

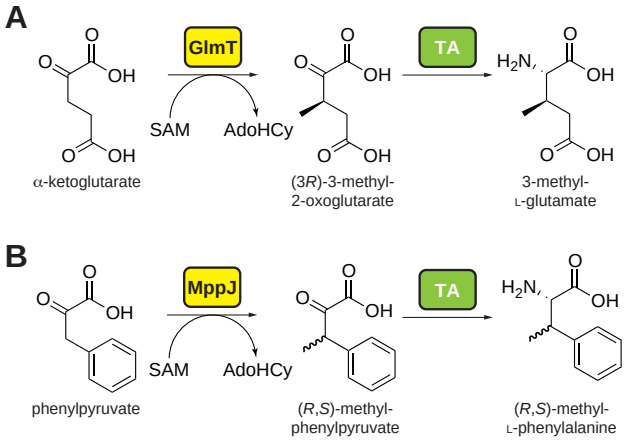


Figure 1.6 Side chain C-methylated amino acids are generated in a two-step reaction via the SAM-dependent modification of the corresponding α-ketoacid, followed by the transamination of the corresponding α-keto group. (A) GlmT, from the CDA-biosynthetic

gene cluster [*S. coelicolor* A3(2), A], catalyzes the stereospecific methylation of α-KG to (3*R*)-3-methyl-2-oxoglutarate followed by the subsequent conversion to 3-methyl-L-glutamate.[82,83] (B) MppJ (mannopeptimycin gene cluster, *Streptomyces hygroscopicus*) catalyzes the analogous reaction on phenylpyruvate but is not stereospecific.[84,85] The NRPS machinery solely incorporates the (2*S*,3*S*)-3-methyl-phenylalanine precursor, found in mannopeptimycin.

1.2.2.2.2 Hydroxylation

Three different classes of enzymes catalyze hydroxylation reactions: FAD-dependent monooxygenases, non-heme $Fe^{(II)}$/α-KG-dependent oxygenases and heme $Fe^{(II)}$-oxygenases, the former two will be discussed (Figure 1.7). FAD-dependent monooxygenases catalyzes a broad variety of oxygenation reactions (including epoxidations) and are often involved in the biosynthesis of iron-coordinating hydroxamate-groups of both NRPS-dependent and NRPS-independent siderophores.[15,86,87]

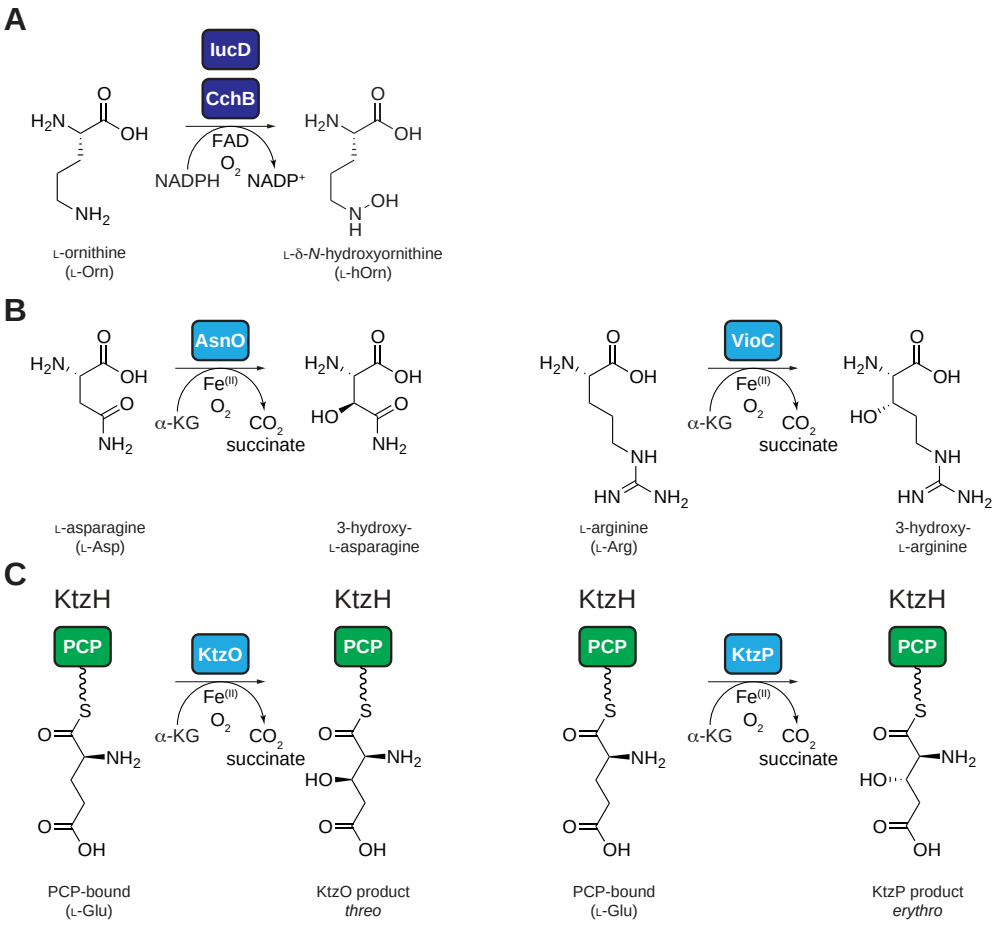


Figure 1.7 Examples for the hydroxylation of free building blocks or PCP-bound substrates. (A) The FAD-dependent monooxygenases CchB and IucD mediate the δ-*N*-hydroxylation of L-Orn during the assembly of the siderophores coelichelin and aerobactin.[22,88] (B) The non-heme $Fe^{(II)}$/α-KG-dependent oxygenases AsnO and VioC are responsible for the hydroxylation of free L-Asp and L-Arg substrates.[89,90] (C) During kutzeneride biosynthesis the two different

non-heme $Fe^{(II)}$/α-KG-dependent oxygenases KtzO and KtzP hydroxylate PCP-bound L-Glu generating the corresponding *threo* or *erythro* products, respectively.[91]

These enzymes require the reduction of the FAD-cofactor, with the needed electrons usually supplied by an NAD(P)H cosubstrate. $Fe^{(II)}$/α-KG-dependent oxygenases couple the oxidative conversion of the substrate with the decarboxylation of the cosubstrate α-ketoglutarate to succinate and carbon dioxide. Both enzyme classes acts as *in trans* hydroxylating catalysts, prior to the incorporation of the modified building block by the NRPS machinery.[89,90,92,93] In addition, $Fe^{(II)}$/α-KG-dependent oxygenases can either hydroxylate free or PCP-bound substrates.[91]

1.2.2.2.3 Acetylation and formylation

Acetylation and formylation reactions are often found as consecutive steps of the hydroxylation of the amino group of the side chain of lysine and ornithine moieties, giving rise to flexible iron-coordinating hydroxamate functionalities.[15] Very recently, the Mcd enzyme from the erythrochelin gene cluster has been characterized, demonstrating the transfer of an acetyl group from an acetyl- or malonyl-CoA donor to the side chain of an L-hOrn amino acid, establishing a pathway for the biosynthesis of the L-haOrn amino acid (Figure 1.8).[94] A similar study aimed at the characterization of the putative formyltransferase CchA, proposed to be involved in the formylation of L-hOrn in the coelichelin biosynthesis gene cluster, was not able to successfully confirm its function. Nevertheless, an analogous "hydroxylation first" model for the biosynthesis of the nonproteinogenic amino acid L-fhOrn has been inferred, based on the substrate specificity of the ornithine monooxygenase CchB.[88]

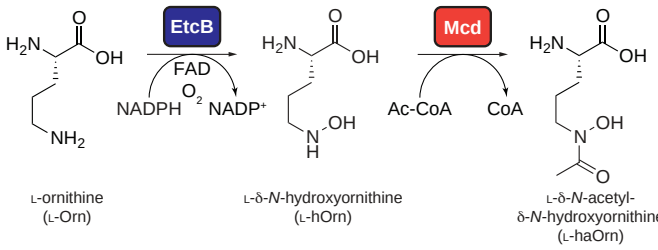


Figure 1.8 Coupled enzymatic biosynthesis of the L-haOrn building block.[94] EtcB converts L-Orn into L-hOrn and the hydroxylated intermediate serves as a substrate for the subsequent Mcd-dependent acetylation reaction, giving rise to the iron-coordinating L-haOrn.

1.2.2.3 NRPS repair mechanism: the type II thioesterase

Approximately 80% of CoA, the precursor of the ppan cofactor required during the priming of PCP-domains, is acetylated in bacteria; therefore, the incorporation within the NRPS assembly line of these misprimed ppan cofactors interrupts the

NRP biosynthesis. Consequently, in order to overcome this critical step, a second type of thioesterase (TEII) is often found in NRP gene clusters that ensures the deacylation of the misprimed PCP-domain.[95-97] These stand-alone domains display structural homology with canonical NRPS termination domains, but the overall greater accessibility of the catalytic pocket ensures the promiscuity of the TEII enzymes towards a broad range of short chain acyl-misprimed PCP-domain substrates.[98] For the same reason, TEIIs are not able to release peptidyl-*S*-PCP bound substrates.

1.2.3 Classification of non-ribosomal assembly line logic

1.2.3.1 Linear NRPS-assembly line logic

The prototypical and most extensively characterized NRPSs assemble the oligopeptide chain through a linear mechanism (type A). The classical modular organization (C-A-PCP) directly correlates, in both order and number of the building blocks, with the primary sequence of the assembled peptide. The release of the product is commonly mediated by a typical TE-domain located at the C-terminus of the last elongation module, that cleaves the peptide from the NRPS through hydrolysis or macrocyclization. Typical examples derived from type A NRPS assembly lines are surfactin, tyrocidine, daptomycin and the ACV precursor of the β-lactam antibiotics (penicillins and cephalosporins, Figure 1.9).[99-102]

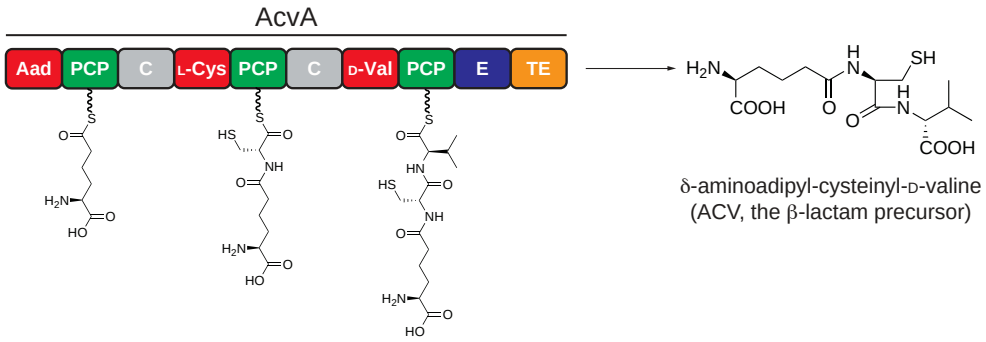


Figure 1.9 AcvA is a prototypical NRPS responsible for the assembly of the ACV tripeptide, the precursor of the β-lactam antibiotics (penicillins and cephalosporins).

1.2.3.2 Iterative NRPS-assembly line logic

The iteratively operating NRPS assembly lines (type B) repeatedly use the enzymatic template during the biosynthesis of the natural product. The modular domain organization resembles type A NRPSs, but, in this case, an iteratively working termination module is responsible for the covalent connection of the constitutive repetitive units and subsequent product release. Examples of iteratively

assembled NRPs are the macrocyclic decapeptide gramicidin and the trilactone siderophores enterobactin and bacillibactin.[12,67,103] In particular, enterobactin is a cyclic trimer composed of 2,3-DHB-L-Ser subunits, connected via three ester bonds between the side chain hydroxyl group of Ser and the carboxylic group of a second monomer (Figure 1.10). Coupled MS-analysis of the site-directed mutagenesis of the terminal TE-domain active site triad of the EntF NRPS allowed the postulation of an enterobactin biosynthetic model.[104] The monomeric 2,3-DHB-L-Ser unit bound to its cognate PCP-domain is subsequently transferred onto the active-site Ser of the terminal TE-domain. The monomer remains TE-bound until a second monomer is presented on the adjacent PCP. TE-mediated ligation affords the TE-bound dimer and a third ligation reaction gives rise to the TE-bound trimer, which is finally cyclized and released from the enzymatic machinery.

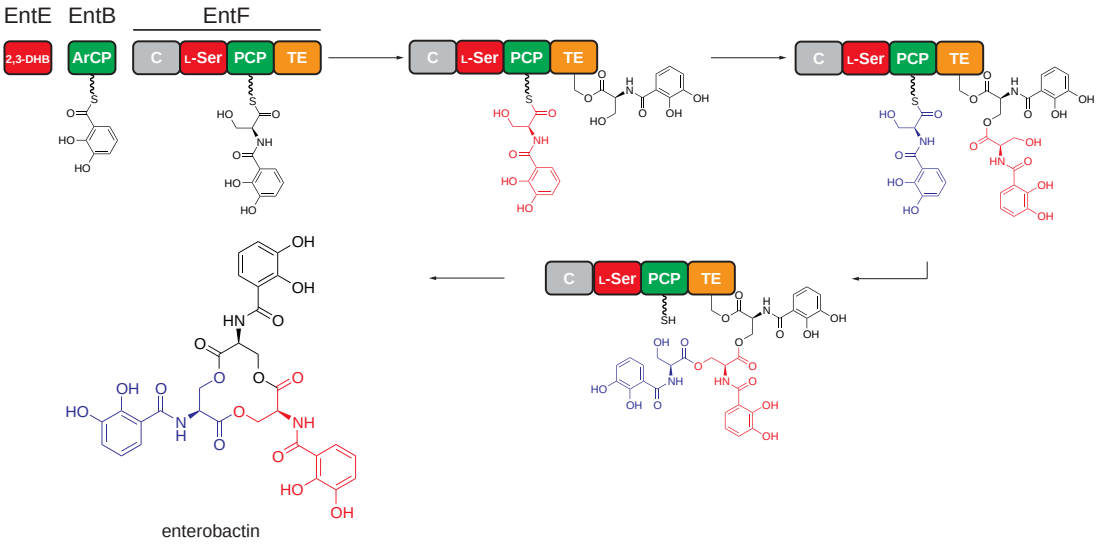


Figure 1.10 The postulated assembly mechanism of the siderophore enterobactin. The stand-alone AMP-ligase EntE activates 2,3-DHB, which is subsequently transferred to the synthetase EntF by the ArCP-domain of EntB. Assembly of the trilactone enterobactin is realized by iterative dimerization and trimerization of the 2,3-DHB-L-Ser units (coloured in black, red and blue).

1.2.3.3 Non-linear NRPS assembly line logic

In non-linear NRPSs (type C), the assembly of the peptide does not follow the typical linear elongation logic of the modular type A and type B synthetases. In fact, the number and organization of modules and/or domains does not reflect the primary sequence of the assembled product. In addition, lone standing individual domains are often involved in product assembly. Thus, in Type C NRPSs, the natural product structure prediction on the basis of the modular NRPS organization is often impeded or not correct. Notable examples of non-linear NRPS derived

natural products are the antibiotic congocidin (*Streptomyces ambofaciens*) and the tris-hydroxamate siderophore coelichelin.[13,105] In this latter case, a trimodular NRPS is responsible of the assembly of a tetrapeptide and the suggestion of the so-called "module skipping" mechanism allowed the postulation of a biosynthetic model (Figure 1.11).[13,106] Before the final release, with the corresponding tripeptide bound to the terminal PCP-domain, the initiation module of the CchH NRPS activates a second L-fhOrn building block that is subsequently incorporated into the peptide chain, to give rise to the final tetrapetide.

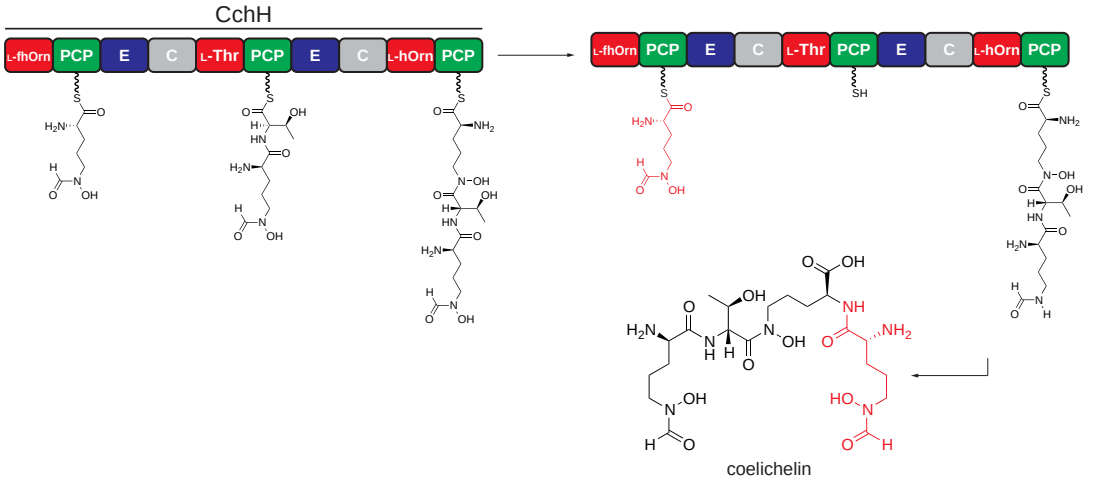


Figure 1.11 Coelichelin assembly is catalyzed by the trimodular NRPS CchH and follows a non-linear logic, due to the presence of four amino acids within the synthesized siderophore.[13] The biosynthesis of coelichelin is initiated by the FAD-dependent monooxygenase CchB that catalyzes the hydroxylation of L-Orn to L-hOrn,[88] which is either directly incorporated into the oligopeptide by the third module of the NRPS CchJ or further tailored resulting in L-δ-*N*-formyl-δ-*N*-hydroxyornithine (L-fhOrn) by the formyltransferase CchA (not shown).[107] The initiation of coelichelin assembly requires is the activation and the covalent tethering of L-fhOrn to the first module of CchH. The reaction proceeds with the condensation of Thr and the subsequent isopeptide linkage with L-hOrn, following (until this point) a linear NRPS assembly logic. Prior to hydrolytically release of the tetrapeptide, it is suggested that the first CchJ module additionally activates and tethers a second L-fhOrn building block, which is incorporated into the peptide by nucleophilic attack of the α-amino group of hOrn. The non-consecutive reuse of a NRPS module within the assembly of a single NRP molecule has therefore been termed "module skipping" mechanism.

1.3 Rational strategies for natural product discovery via genome mining

The discovery of natural products often relies on bioassay-guided fractionation of extracts from different natural sources and the subsequent isolation of the bioactive compound. This employed strategy has historically led to the discovery of many bioactive compounds, used in clinical therapy as antifungal, anticancer and immunosuppressive agents.[30] Lately, the discovery that the systematic cultivation of one species under several conditions strongly influences secondary metabolite production (OSMAC, One Strain - MAny Compounds approach) allowed the discovery of different natural products.[108] In recent years, with the increasing amount of information derived by huge advances in sequencing technologies, a plethora of sequenced microbial genomes has revealed a multitude of gene clusters associated with the biosynthesis of secondary metabolites.[109,110] This quantity of information, in combination with a substantial increase in the understanding of natural product biosynthesis has paved the way for the mining of genomes for bioactive compounds.[32]

Since many microbial natural products are assembled by multimodular synthases and synthetases (*e.g.* polyketides and non-ribosomal peptides), the development of bioinformatic tools for the analysis and the prediction of the modular organization and the substrate specificity of these assembly machineries has formed the basis for the subsequent development of genome mining approaches, allowing the isolation of new natural products solely on the basis of the genome sequence of the target organism.[32,40,111] The first microbial organism to be extensively analyzed for the production of unknown secondary metabolites was *S. coelicolor* A3(2). Within its genome, several gene clusters were identified to encode for new biosynthetic pathways, later associated to newly isolated natural products. Besides, the gene clusters that still remain "orphan" with respect to their cognate metabolites, were renamed "cryptic".[13,112-115]

During the past years, several approaches for microbial genome mining have been developed, leading to a successful characterization of new natural products (Figure 1.12). The first approach solely relies on the prediction of the physico-chemical properties of the target compound, and has led to the identification of salinilactam A (*Salinispora tropica*).[116] The sequence analysis of a modular PKS gene cluster putatively encoding a lysine-primed polyene macrolactam allowed the isolation and the structural characterization of salinilactam A solely on the basis of polyene UV-absorption properties. An alternative approach is represented by the *in vitro* reconstitution of natural product biosynthesis. This strategy uncouples the

biosynthetic enzymes from their native regulatory mechanism, through their recombinant production in a heterologous host. The two-component lantibiotic haloduracin (*Bacillus halodurans* C-125) has been isolated by the incubation of the biosynthetic enzymes with the ribosomally synthesized substrates *in vitro*.[117]

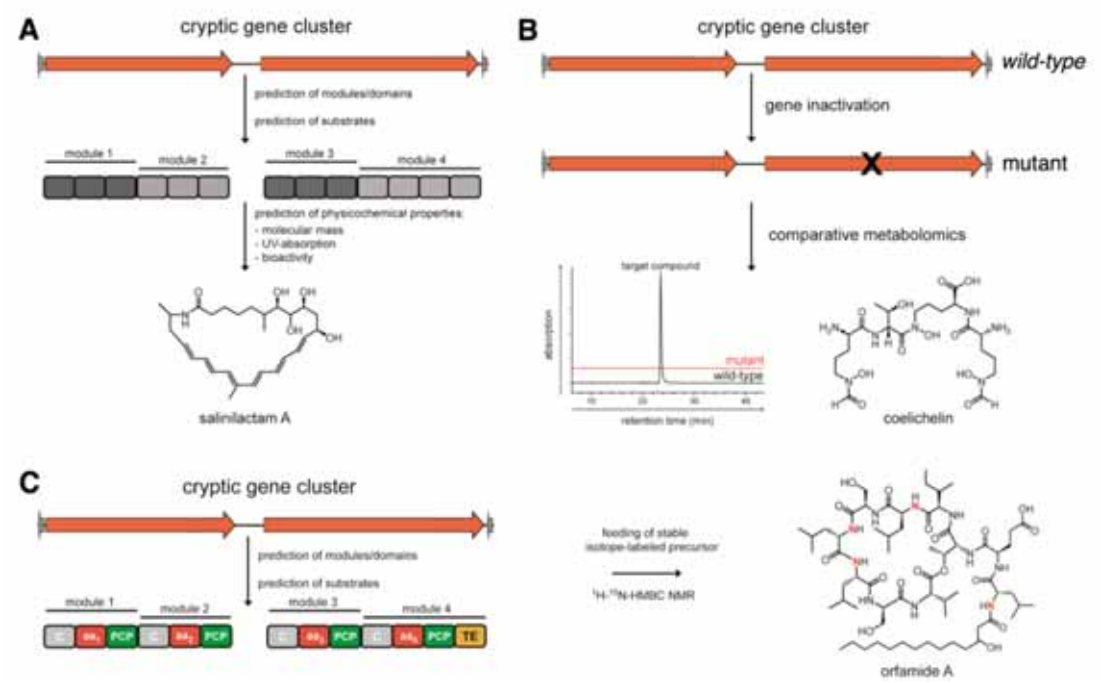


Figure 1.12 Genome mining strategies applied for the isolation of new natural products. (A) The isolation of salinilactam A (*S. tropica*) relied on the prediction of the physico-chemical properties of the cryptic natural product. (B) The inactivation of the natural product biosynthetic pathway, followed by the comparative metabolomic profiling, allowed the characterization of the siderophore coelichelin and its gene cluster [*S. coelicolor* A3(2)]. (C) A genomisotopic approach was employed for the identification of the lipopeptide orfamide A. On the basis of the bioinformatic analysis of the A-domain substrate specificities, ^{15}N-Leu was fed into the cultures of *Pseudomonas fluorescens* Pf-5; the subsequent ^{1}H-^{15}N-NMR experiments led to the identification and the isolation of the natural product.

The inactivation of gene cluster-associated genes and subsequent comparative metabolomic profiling of the *wild-type* and mutant strains has been successfully applied for the discovery of the hydroxamate-type siderophore coelichelin.[13] The main drawback of this strategy, which does not require the structure prediction of the secondary metabolite, is the genetic accessibility of the target strain. Additionally, the change in the metabolite profile of the *wild-type* and the mutant strain has to be detectable. Comparative metabolic profiling can also be applied to strategies that require the heterologous expression of cryptic biosynthetic gene cluster into a genetically transformed non-producing host strain. In this case, the profile comparison of the transformed strain with a control counter-part allows the natural product identification.[118-120]

The last more general strategy described for natural product discovery is the genomisotopic approach, which has been successfully applied for the characterization of the NRPS-derived cyclolipopeptide orfamide A (*Pseudomonas fluorescens* Pf-5).[121] This technique combines the bioinformatic predictions of the modular assembly machinery with the incorporation of a stable-isotope precursor added to the growing culture, allowing the labeling, detection and structural characterization of the assembled product. In the case of orfamide A, ^{15}N-labeled L-Leu was fed to cultures of *P. fluorescens* Pf-5 and HPLC fractions of the culture extracts were analyzed via ^{1}H-^{15}N-NMR experiments for the successful identification of isotope-labeled metabolites.

In conclusion, the choice of the best strategy for the successful characterization of new natural products via genome mining must take several aspects into account, among them, the culturing conditions and the genetic accessibility of the producing strain and the availability of the proper instrumentation for the isolation and structural characterization of the natural product.

1.4 Aim of the work

Rhodococcus spp. are extensively studied as extraordinary biocatalysts for steroid production and fossil fuel biodesulfurization and as tools suitable for bioremediation purposes.[122,123] This widespread biotechnological and industrial interest derive from their diverse metabolic capabilities. Furthermore, genome sequencing information revealed *Rhodococcus* spp. to possess a vast genetic potential for secondary metabolite production.[124] However, only very few natural products have been isolated from this genus, among them two siderophores, heterobactin A and rhodobactin, isolated from *Rhodococcus erythropolis* IGTS8 and *Rhodococcus rhodochrous* OFS, respectively (Figure 1.13).[125,126] They belong to the hydroxamate-catecholate mixed type family, with the common presence of 2,3-dihydroxybenzoic acid (2,3-DHB) and differently modified ornithine residues within their structures. In both cases, no gene clusters responsible for siderophore biosynthesis have been identified.

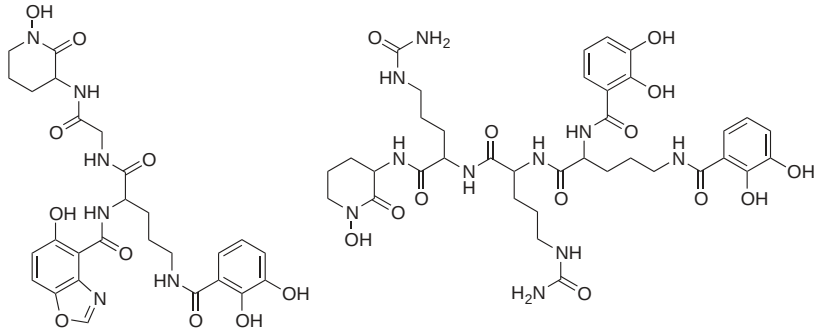


Figure 1.13 Chemical structures of representative siderophores isolated from *Rhodococcus* strains: heterobactin A and rhodobactin were isolated from *R. erythropolis* IGT8 and *R. rhodochrous* OFS, respectively.

Rhodococcus jostii RHA1 was isolated from a lindane-contaminated soil and is known for its ability to transform polychlorinated biphenyls and to utilize a wide range of aromatic compounds, carbohydrates, nitriles, and steroids as its sole energy source:[127] these features make it a species of significant industrial interest. In this context, its genome sequence has been published in 2006 and contains approximately 9.7 Mbp arranged in one linear chromosome and three additional linear plasmids. *R. jostii* RHA1 also contains more that 30 NRPS and PKS genes, 6 of which exceed 25 kbp, providing evidence of an extensive and uncharacterized secondary metabolism.[124]

On the basis of this knowledge, the isolation and the structural characterization of the endogenous siderophore of *R. jostii* RHA1 will confirm the metabolic capacity of the strain to produce secondary metabolites. Furthermore, the genome mining identification of the gene cluster responsible for the biosynthesis of the molecule will permit the rational construction of isogenic deletion mutant strains. Subsequently, the metabolic profile comparison between the *wild-type* and the mutant strains will undoubtedly connect the biosynthesis of the natural product with the corresponding genes. Finally, the biochemical characterization of the recombinantly-produced enzymes associated with the biosynthesis of the siderophore will integrate the genetic results and will allow the postulation of a model for the biosynthesis of the newly-discovered iron-scavenging compound.

Chapter 2

Material

2.1 Equipment

Table 2.1 List of the devices used in this study.

device	manufacturer and type
Autoclave	Tuttnauer 5075 ELV, Fedegari Autoclavi SPA FVA3/A1
Centrifuges	Sorvall RC 5B Plus and RC6+ (SS-34, SLC-300, SLC-4000 rotors), Heraeus Minifugue RF and Megafugue 1.0R, Eppendorf 5415 D, 5415 R and 5702 R
Clean bench	Antair BSK
Documentation system for DNA-electrophoresis gels	Cybertech CS1 camera, Mitsubishi video copy thermo printer
Electrophoresis chamber	Agarose gel chambers manufactured in-house (PUMa, Marburg), Bio-Rad Mini-PROTEAN 3 gel chamber
Electroporation	Bio-Rad Gene-Pulser II
Fast protein liquid chromatography (FPLC) system	Amersham Pharmacia Biotech Äktaprime and Äktapurifier: (pH/C-900, UV-900, P-900 and Frac-900 modules)
French-press	SLM Aminco French-Pressure Cell Press 5.1, Thermo Spectronic Standard Cell 40 KP
HPLC-systems	Agilent series 1100 (HPLC-system with DAD- and ESI-Quad-MS-detection, vacuum degasser, quarternary pump, auto sampler, preparative fraction collector, column thermostat, HP-ChemStation software) Thermo Scientific LTQ-FT Applied Biosystems API QStar Pulsar i
Incubators	New Brunswick Scientific Series 25, Innova 4300 Shaker, Infors HT Multitron II and Unitron
Lyophilizer	Christ Alpha 2-4 LSC
NMR-spectrometer	Bruker AV600
pH meter	Schott CG 840
Pipettes	Eppendorf Research series
Rotary Evaporator	Heidolph Laborota 4000
Scintillation counter	Packard Tri-Carb 2100-TR
Spectrophotometer	PEQLab Nanodrop ND-1000; Pharmacia Ultrospec 3000
Speed-Vac	Uniequip Univapo 150H
Thermal cycler	Eppendorf Mastercycler Personal
Thermomixer	Eppendorf Thermomixer comfort
Vortexer	Scientific Industries VortexGenie2
Water deionizer	Seral Seralpur Pro90CN

2.2 Chemicals, enzymes and consumables

All chemicals not listed in Table 2.2 were purchased from Sigma-Aldrich (Steinheim, Germany), Fluka (Steinheim, Germany) or Merck (Darmstadt, Germany) in p.a. quality and were used without further purification.

Table 2.2 Chemicals, enzymes and general materials and consumables.

manufacturer	product
Applichem (Darmstadt, Germany)	ampicillin, kanamycin, media components
Biomol (Ilvesheim, Germany)	DTT
Brand (Wertheim, Germany)	Plastbrand PS cuvettes
Eppendorf (Hamburg, Germany)	1.5 and 2.0 mL reaction tubes
Eurogentech (Seraing, Belgium)	agarose, electroporation cuvettes
GE Healthcare (Freiburg, Germany)	IPTG, FPLC Ni-NTA and Superdex 200 5/150 GL SEC columns, yeast extract, aldolase, ovalbumin, ribonuclease, aprotinin protein standards
Invitrogen (Karlsruhe, Germany)	*E. coli* strains (BL21, TOP10)
Macherey & Nagel (Düren, Germany)	C_8- and C_{18}-HPLC columns (Nucleosil, Nucleodur)
Merck4Biosciences - Novagen (Nottingham, UK)	pET28a(+)
Millipore (Schwalbach, Germany)	Dialysis membranes (pore size: 0.025 µM), Amicon Ultra-15 concentrators
MP Biomedicals (Illkirch, France)	coenzyme A trilithium salt
New England Biolabs (Frankfurt, Germany)	desoxyribonucleotides (dATP, dTTP, dGTP, dCTP), DNA ladders, protein size markers, restriction endonucleases, Phusion Hi-Fidelity DNA polymerase, T4 DNA ligase
Oxoid (Cambridge, UK)	agar nr. 1, tryptone
Perkin Elmer (Waltham, USA)	$Na_4{}^{32}P_2O_7$
Phenomenex (Torrance, USA)	Synergi Fusion RP-80 HPLC column
QIAgen (Hilden, Germany)	QIAquick Gel Extraction kit, Ni-NTA IMAC resin
Roth (Karlsruhe, Germany)	acrylamide solution, β-mercaptoethanol, ethidium bromide, scintillation fluid
Sarstedt (Nümbrecht, Germany)	Pipette tips, Falcon tubes (15 and 50 mL)
Schleicher & Schüll (Dassel, Germany)	Sterile filters (0.20 and 0.45 µm), Whatmann-3MM paper
Serva (Heidelberg, Germany)	bromophenol blue, Triton X-100
Thermo Scientific (Waltham, USA)	Hypercarb HPLC column

2.3 Oligonucleotides

All oligonucleotides listed in Table 2.3 were purchased from Sigma-Aldrich (Steinheim, Germany) in desalted quality grade and were used for PCR amplification of the targeted genes.

Table 2.3 List of oligonucleotides used in this work. Restriction sites inserted for cloning are underlined.

primer name	sequence (5'-3')	restriction site	plasmid	target gene
2319P1	CCCAAGCTTCGACTCGATGTCGAGGACGTGCAG	HindIII	pK18mobsacB	*rhc*B
2319P2	CGGACGTTTTCATCGATCTCCCTGTTCTGTCGAAAGGGACATCTAACAGGCACTCCTG			*rhc*B
2319P3	CAGGAGTGCCTGTTAGATGTCCCTTTCGACAGAACAGGGAGATCGATGAAAACGTCCG			*rhc*B
2319P4	GCTCTAGACTGGTCGGGTTCTGCGCCATGATG	XbaI	pK18mobsacB	*rhc*B
2319A1F	GCCGCTTCCATATGCCGGACAACATCGTCTCCGCCTTC	NdeI		*rhc*B
2319A1R	CATAGCTCGAGTCAGGTCCGCGGTGTGTCGCCG	XhoI		*rhc*B
4793P1	GCATACAAGCTTCAGGAGGTCGTCAGATGATGTTTCGCAG	HindIII	pK18mobsacB	*dhb*E
4793P2	GACTTTCGATGAGCACGAGCGTTCGCGTGACTTCGTAACCCGACCCCGCTATC			*dhb*E
4793P3	GATAGCGGGGTCGGGTTACGAAGTCACGCGAACGCTCGTGCTCATCGAAAGTC			*dhb*E
4793P4	GCATTCTCTAGAGCCACACTGCACAAGGTGGTACTCG	XbaI	pK18mobsacB	*dhb*E
4793INT1	CTCGCCGAGATACGGATCGGGC			*dhb*E
4793INT2	GTGGTGCAACTGCCGAACATCGTCG			*dhb*E
4716P1	GCATACAAGCTTCTTCTTCGACCTCGGCGGCAACTCC	HindIII	pK18mobsacB	*rmo*
4716P2	GGATTGGTGCTCATCTCGCCTCGCTTGTCGATTCACTCATAGCTCGTTGTCCGTCC			*rmo*
4716P3	GGACGGACAACGAGCTATGAGTGAATCGACAAGCGAGGCGAGATGAGCACCAATCC			*rmo*
4716P4	GCCTTCTCTAGACTCCATCGTGCTGTCCGCATACCG	XbaI	pK18mobsacB	*rmo*
4716INT1	CGATTCGGGTGGCATCCCGGGATG			*rmo*
4716INT2	CAGGAGCGACGACGTGATTCCGTG			*rmo*
4712P1	CCGGAATTCGATCGCGTCCGCCGCTGTGATCG	EcoRI	pK18mobsacB	*rft*

4712P2	GTCCAGGAGGACCGCGTTGAGAGTCTGACGGT CCCGCGCCGACACCAT			*rft*
4712P3	ATGGTGTCGGCGCGGGACCGTCAGACTCTCAA CGCGGTCCTCCTGGAC			*rft*
4712P4	GCTCTAGACGTCCCGGAAATGCACGACCAGCG	XbaI	pK18mobsacB	*rft*
4712INT1	CGATGACCATTCCGTCGCCTTCGTG			*rft*
4712INT2	GACCATGAGGTCGTCCTCGCGATC			*rft*
KANF	ATGGATTGCACGCAGGTTCTC			*kan*^{R}
KANR	CGATAGAAGGCGATGCGCT			*kan*^{R}
4793F	GGGAATTCCATATGAGCACGAGCGTTCGCGCT	NdeI	pET28a(+)	*dhb*E
4793R	CCCCAAGCTTTTACGAAGTCACGAACGTCTTC TCC	HindIII	pET28a(+)	*dhb*E
4712F	GGAATTCCATATGAGAGTCGCCACACTCGGAT ATC	NdeI	pET28a(+)	*rft*
4712R	ATAAGAATGCGGCCGCTCAGCTGAGGTAGCCG CCG	NotI	pET28a(+)	*rft*
4716F	GGAATTCCATATGAGTGAATCGCCGGAAACGG TCG	NdeI	pET28a(+)	*rmo*
4716R	ATAAGAATGCGGCCGCTCATCTCGCCTCGCTT GTCGCATAC	NotI	pET28a(+)	*rmo*
2322F	AAAAAACCATGGCTAGTTCCGCTTCCAGCACA GTTCC	NcoI	pET28a(+)	*rhc*E
2322R	AAAAAAAAGCTTGCGTGCTGCCGTCACCTCGA A	HindIII	pET28a(+)	*rhc*E
CchApCB^{128}	AAAAAAGAATTCATGCGGGTCGTCATGTTCGG CT	EcoRI	pCB28a(+)	*cch*A
CchApCB^{128}	AAAAAACTCGAGTCAGGGGCGGGCGGTCAG	XhoI	pCB28a(+)	*cch*A

2.4 Plasmids

2.4.1 pET28a(+) and pCB28a(+)

The pET28a(+) expression vector system was employed for the production of recombinant proteins in *E. coli* (Figure 2.1). The vector adds a hexahistidine (6xHis) fusion tag at the N- or C-termini of the recombinant protein, to allow subsequent Ni-NTA affinity chromatography purification. Transcription of the cloned genes is dependent on both T7 RNA polymerase activity and dissociation of the LacI repressor from its corresponding operator, upon IPTG induction. Plasmid selection and maintenance is permitted by the *kan*^{R} gene, which confers resistance to kanamycin.

The pCB28a(+) expression vector is a derivative of pET28a(+) with an alternative multiple cloning site (MCS).

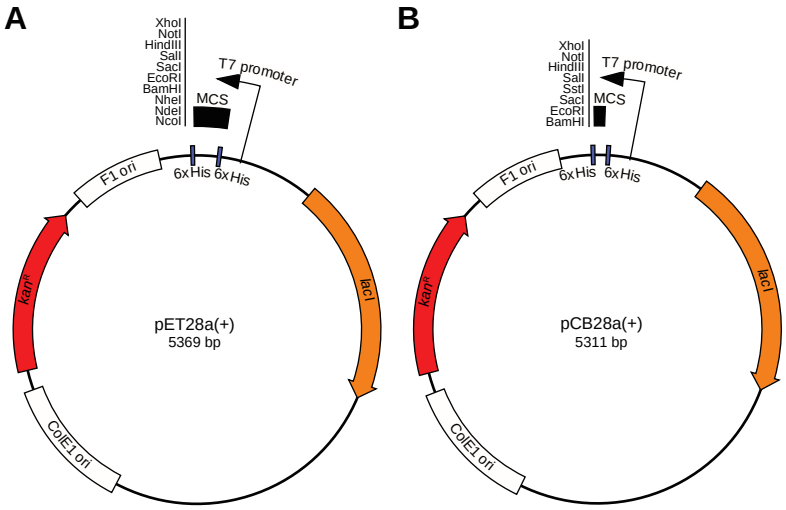


Figure 2.1 (A) Physical map of pET28a(+). (B) Physical map of pCB28a(+).

2.4.2 pK18mobsacB

The pK18mobsacB plasmid is a mobilizing cloning vector for the conjugative transfer of DNA into a recipient host.[129] Deriving from the *E. coli* pK18 plasmid,[130] it features the *lac*Zα fragment inserted within the MCS and the *kan*^{R} gene for selection and maintenance (Figure 2.2). The mobilizing machinery of the RP4 plasmid[131] confers broad host-specificity between different Gram-negative and Gram-positive species.[132-134] The *sac*B gene encodes for the *B. subtilis* levansucrase, which renders the recipient strain sensitive to sucrose, and thus is applicable as a negative-selection marker.[135-137] The plasmid was kindly provided by Dr. Robert van der Geize, University of Groningen, The Netherlands.

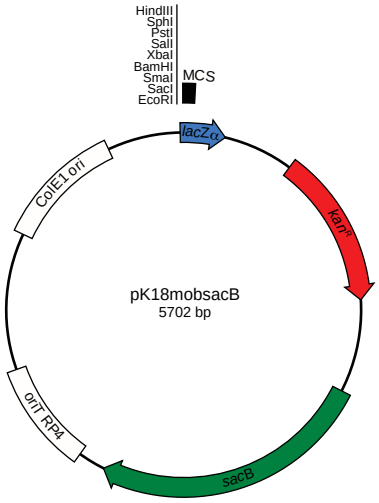


Figure 2.2 Physical map of pK18mobsacB.

2.5 Bacterial strains

2.5.1 *Rhodococcus jostii* RHA1

R. jostii RHA1 is a strain originally isolated from an insecticide-polluted soil sample. It is classified as a Gram-positive, non-sporulating and non-motile microorganism belonging to the Actinomicetales order.[127] Its completely sequenced genome comprises one linear chromosome and three linear plasmids, covering approximately 9.2 Mbp in total.[124] The strain has been cultured for the isolation of the siderophore rhodochelin, for the isolation of genomic DNA and for the construction of isogenic deletion mutants. The strain was kindly provided by Prof. Linday E. Eltis of the University of British Columbia, Vancouver, Canada.

2.5.2 *Escherichia coli* TOP10

The *E. coli* TOP10 strain was used as a natural host for cloning and plasmid propagation procedures. The genotype is: F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 ΔlacX74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G.

2.5.3 *Escherichia coli* BL21 (DE3)

The *E. coli* BL21 (DE3) strain was used as a general host for the heterologous expression of recombinant proteins using T7-promoter driven expression plasmids. It derives from the *E. coli* B strain and its genotype is: F^- *ompT* *hsd*S_B(r_B^-, m_B^-) *gal dcm* λ(DE3). The strain lacks the proteases Ion and OmpT to increase the half-life of the overproduced recombinant protein. In addition, the insertion of the λ prophage DE3 carrying the IPTG-inducible T7 RNA polymerase (under the control of the *lac*UV5 promoter) is essential for recombinant protein expression.[138]

2.5.4 *Escherichia coli* S17-1

The *E. coli* S17-1 strain (genotype recA pro hsdR RP4-2-Tc::Mu-Km::Tn7) was used as mobilizing strain for the transfer of the pK18mobsacB plasmid into the recipient *R. jostii* RHA1, through bacterial conjugation. The mobilizing elements of the RP4 plasmid are stably integrated within the chromosome, in order to avoid self-transfer of the original conjugating vector.[139]. The strain was kindly provided by Dr. Robert van der Geize, University of Groningen, The Netherlands.

2.6 Culture media

All media listed below were used for the cultivation and fermentation of the bacterial strains indicated above. If solid agar plates were required for the cultivation of the microorganism, agar nr. 1 was added to a final concentration of 1.5 % (m/w). All media were sterilized by autoclavation, (121 °C, 1.5 bar, 20 min). Antibiotics and other additional temperature-labile components were added after sterile-filtration, prior to use.

2.6.1 Lysogeny broth (LB-Miller)

LB broth (buffered at pH 7) has been routinely used for the cultivation and the maintenance of *R. jostii* RHA1 and *E. coli* strains.

component	quantity (1 L)
yeast extract	5 g
tryptone	10 g
NaCl	5 g

For the selection of *Rhodococcus* transconjugants sucrose was added to a final concentration of 10% (m/w). Where required, antibiotics were added at the following concentrations: nalidixic acid 30 µg/mL, kanamycin 50 µg/mL for derivatives of pET28a(+), pCB28a(+) and pK18mobsacB integrants in *Rhodococcus* host, and kanamycin 25 µg/mL for derivatives of pK18mobsacB in *E. coli*.

2.6.2 M9 minimal medium

M9 minimal medium was used for the isolation and the fermentative production of rhodochelin.

component	quantity (1 L)
$Na_2HPO_4 \cdot 2\ H_2O$	8.5 g
KH_2PO_4	10 g
NH_4Cl	1 g
NaCl	0.5 g

Prior to use, a concentrated trace element mix (100xMg/Ca/B1/Goodies) was prepared as follows and added to a final 1x concentration. Sterile glucose solution was used as a carbon source (final concentration: 4 g/L). Where required, Fe^{3+} was added from a sterile-filtered $FeCl_3$ solution.

100x Mg/Ca/B1/Goodies	
component	**quantity (100 mL)**
$MgSO_4$ 1 M	20 mL
$CaCl_2$ 1 M	1 mL
thiamine 10 mM	1 mL
Conc. goodies	25 mM
ddH_2O	top up to volume

Concentrated goodies	
component	**quantity (100 mL)**
Stock salts sol.	50 mL
$MgSO_4 \cdot 7\ H_2O$	3 g
ddH_2O	top up to volume

Stock salts solution[140]	
component	**quantity (1 L)**
$MgCl_2 \cdot 6\ H_2O$	22.94 g
$CaCO_3$	2.0 g
$ZnSO_4 \cdot 7\ H_2O$	1.44 g
$MnSO_4 \cdot H_2O$	0.85 g
$CuSO_4 \cdot 5\ H_2O$	0.25 g
$CoCl_2 \cdot 6\ H_2O$	0.24 g
H_3BO_3	0.06 g
HCl (conc.)	51.3 mL
ddH_2O	top up to volume

Chapter 3

Methods

3.1 Molecular biology techniques

General molecular biology material and methods (*e.g.* DNA agarose-gel electrophoresis, protein SDS-PAGE, the composition of routine buffers solutions, etc.) are not described in this section and were carried out according to established protocols.[141]

3.1.1 General strains maintenance

All *E. coli* strains were routinely maintained on LB-agar plates and incubated at 37 °C. *R. jostii* RHA1 and derivative strains were maintained on LB-agar plates at 30 °C. Liquid cultures were incubated under continuous shaking at 180 rpm. Antibiotics were added where required. For long-term storage, sterile glycerol was added to a liquid culture [final concentration: 20% (v/v)] and the resulting stocks were stored at -80 °C.

3.1.2 Preparation of genomic DNA

For genomic DNA isolation, 5 mL of liquid culture were harvested by centrifugation and the resulting pellet was washed with 1 mL of water. The cell pellet was resuspended in 500 µL lysis buffer [100 mM TRIS, 50 mM EDTA, 1% (w/v) SDS, pH 8] and acid-washed glass beads were added to a final volume of 1.25 mL. The mixture was vortexed for two minutes and the recovered supernatant was transferred into a new microfuge tube. 275 µL of 7 M ammonium acetate pH 7 was added and the solution was incubated for 5 min at 65 °C, and then for 5 min on ice. 500 µL of chloroform were added and the mixture was vortexed for 2 min. Following a centrifugation step at 13,000 rpm for 5 min, the recovered aqueous phase was added to 800 µL of ice-cold isopropanol. The genomic DNA was precipitated by centrifugation (5 min, 13,000 rpm), prior to 5 min incubation on ice. Subsequently, the pellet was washed with 500 µL of ice-cold 70% ethanol solution, dried and resuspended in 50 µL of EB buffer (10 mM TRIS, pH 8.5).

3.1.3 Preparation of plasmid DNA

Routine plasmid preparations were carried out according to the alkaline lysis with SDS procedure.[141] 5 mL of LB medium (supplemented with the respective antibiotic) was inoculated with a single colony and grown overnight. 2 mL of the stationary-phase culture were harvested by centrifugation (13,000 rpm, 1 min) and the cell pellet was resuspended in 300 µL of P1 buffer (50 mM TRIS, 10 mM EDTA, pH 8, RNAse A 100 µg/mL). 300 µL of P2 buffer [0.2 M NaOH, 1% (w/v) SDS] were

added to the cell suspension. The tube was mixed gently by inversion and incubated at RT for 5 min. Subsequently, 300 µL of P3 buffer (2.55 M potassium acetate, pH 4.8) were added, the content was mixed by inversion and centrifuged at RT for 15 min. The cleared supernatant (approx. 800 µL) was transferred into a new tube, where 700 µL of ice-cold isopropanol were added. The plasmid DNA was pelleted by centrifugation (13,000 rpm, 30 min, 4 °C) and washed once with 300 µL of cold 70% ethanol. Finally, the pellet was dried at 42 °C and resuspended in 50 µL of ddH_2O.

For sequencing purposes, high-purity plasmid preparations were carried out according to the Sigma GenElute HP Plasmid Miniprep kit manual.

3.1.4 PCR-based gene amplification

Target genes were amplified from genomic DNA using the primers listed in Table 2.3, with Phusion High-Fidelity DNA Polymerase, following the manufacturer's protocol for GC-rich DNA templates (5% DMSO final concentration). PCR fragments were purified from agarose gel slices, according to the instructions of the QIAquick Gel Extraction kit.

PCR fragments aimed at designing "markerless" gene deletions in *R. jostii* RHA1 were created by joining two ~1500 bp PCR products through splicing overlap extension (SOE) PCR.[142] First, the two amplicons flanking the desired gene mutation were amplified from the chromosome template using two pairs of primers P1, P2 and P3, P4. Then, because of the overlapping flanks of the P2 and P3 primers, the resulting *in-frame* deletion fragment (which still maintains the 5' and the 3' ends and the frame shift of the deleted gene) was obtained by combining the two PCR reactions (Figure 3.1).

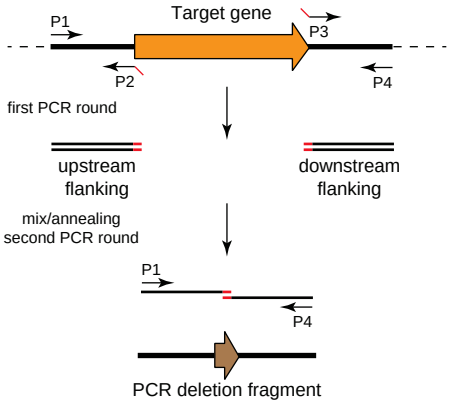


Figure 3.1 Scheme for the creation of *in-frame* gene deletion PCR fragments using splicing overlap extension (SOE) PCR. First, two PCR fragments, flanking the desired mutation, are amplified using two primer pairs. Then, the resulting PCR products are mixed and further

amplified resulting in the desired deletion fragment, which is subsequently cloned into the appropriate plasmid.

3.1.5 Vector construction

Purified PCR products were digested with the corresponding restriction endonucleases and cloned into their respective digested vector using T4 DNA ligase. Ligation products (1:3 plasmid to insert molar ratio) were used to transform (25 kV/cm^{-1} pulse, 200 Ω resistance, 25 μF capacitance) electrocompetent *E. coli* TOP10 cells. Transformants were plated and selected on LB-agar plates supplemented with the corresponding antibiotic. The correct fragment insertion was verified by analytical plasmid restriction and by dideoxy sequencing (GATC, Kostanz, Germany). Finally, the correct constructs were used to transform *E. coli* BL21 (DE3) or S17-1 strains. An overview of the constructs is given in Table 3.1.

Table 3.1 Overview of the plasmids created and used in this study.

constructs for protein expression			constructs for *Rhodococcus* mutagenesis	
name	**insert size**	**His-tag**	**name**	**insert size**
pET28a::*dhb*E	1655 bp	N-term	pK18mobsacB::RHA1Δ*rhc*B	3354 bp
pET28a::*rhc*E	245 bp	C-term	pK18mobsacB::RHA1Δ*dhb*E	3263 bp
pET28a::*rmo*	1350 bp	N-term	pK18mobsacB::RHA1Δ*rmo*	3176 bp
pET28a::*rft*	939 bp	N-term	pK18mobsacB::RHA1Δ*rft*	3299 bp
pCB28a::*cch*A[128]	954 bp	N-term		

3.1.6 Construction of *Rhodococcus* mutants

Rhodococcus "markerless" gene deletion mutants were generated according to the protocol described by Van der Geize *et al.*[143] This approach employs the transformation of the recipient *Rhodococcus* strain with the pK18mobsacB plasmid[129] carrying the desired mutation, via a conjugation step from the donor *E. coli* S17-1 strain.[139] A double selection strategy, carried out in a two-step procedure, allows the selection of the desired deletion mutant (Figure 3.2). In the first step, the clones that have integrated the suicide plasmid within the chromosome (through a intermolecular cross-over) are selected for the presence of the positive marker (*i.e.* the resistance to kanamycin). Subsequently, the transconjugants are propagated to allow the second rare recombination event (the intramolecular cross-over) and are selected by plating on counter-selective sucrose-supplemented medium. Under these conditions, the growth is only possible when the *sac*B marker was lost, which can further be distinguished by PCR using primers flanking the desired mutation.

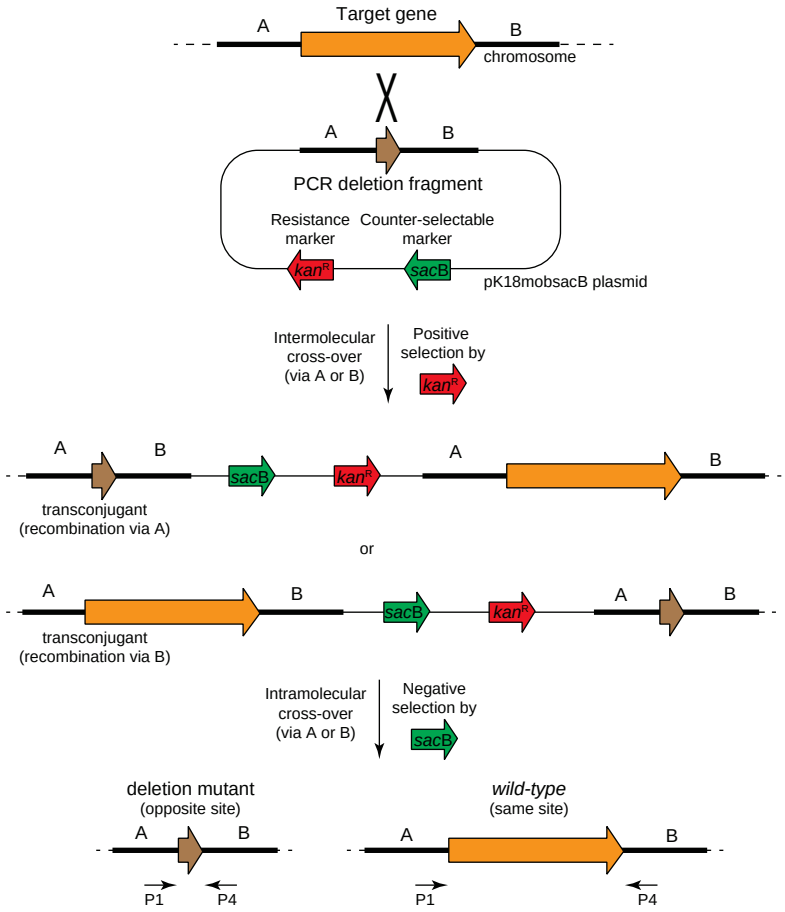


Figure 3.2 Overview of the double selection strategy used for the generation of *Rhodococcus* mutants. Upon transformation of the *wild-type* strain with the mutagenic plasmid, an intermolecular cross-over between one of the two homologous regions flanking the mutation occurs, integrating the plasmid into the chromosome. The transformants are then selected with the *kan*^{R} resistance marker. To allow the second rare intramolecular cross-over, the transconjugants are allowed to grow without antibiotic and subsequently selected by plating on counter-selective sucrose-supplemented medium. The second cross-over event, which ensures the correct gene plasmid excision (with the loss of the *sac*B marker) and results in the desired gene deletion, have to occur on the second homologous flanking region.

R. jostii RHA1 cells were grown on LB plates supplemented with nalidixic acid for 5 days, harvested, and resuspended in 2 mL of fresh LB broth. The same procedure was repeated with the overnight plates of derivative mutagenic *E. coli* S17-1 strains, additionally grown at RT for further 24 h. 750 µL of each cell suspension were mixed, incubated briefly at RT, pelleted, and resuspended in 2 mL of LB broth. 200 µL were spread on LB and incubated overnight at 30 °C. The following day, cells were harvested and resuspended in 2 mL of LB broth. Aliquots (150 µL each) were successively spread on LB plates supplemented with nalidixic acid and kanamycin and incubated at 30 °C for three days, until only *Rhodococcus* colonies

appeared. Transconjugants were grown in liquid medium and vector integration was checked by PCR of the kanamycin cassette and by replica plating on LB supplemented with kanamycin, or kanamycin and 10% sucrose. To force plasmid excision, single integrant colonies were inoculated in LB broth and subsequently plated on LB supplemented with 10% sucrose and grown at 30 °C until new colonies appeared. To confirm correct plasmid excision, single clones were tested for kanamycin sensitivity by replica plating and by PCR using different primer pairs: for the kanamycin cassette, for a flanking region and for the deleted gene.

3.2 Expression and purification of recombinant proteins

3.2.1 Gene expression

5 mL of an overnight culture of *E. coli* BL21 (DE3) carrying the desired expression construct was diluted inside a 2 L baffled flask containing 500 mL of fresh LB medium supplemented with 50 μg/mL kanamycin. Cells were grown at 25 °C, under continuous shaking at 230 rpm, until OD_{600} ~0.5 was reached. Gene expression was induced with IPTG (50 μM), followed by incubation at 25 °C for 4 h. Cells were harvested by centrifugation (6000 rpm, 15 min, 4 °C), resuspended in HEPES A buffer (50 mM HEPES, 300 mM NaCl, pH 8) and stored at -20 °C until further processing.

3.2.2 Protein purification

The frozen cell pellet was thawed on ice and lysed via French press. After a centrifugation step (17,000 rpm, 4 °C, 30 min), the cleared lysate was sterile-filtered and applied to a Ni-NTA IMAC column (equilibrated with HEPES A buffer) using a ÄktaPrime system with a flow rate of 0.8 mL/min. The elution was carried out using a linear gradient from 3 to 50% HEPES B buffer (50 mM HEPES, 250 mM imidazole, 300 mM NaCl, pH 8) over 30 min, followed by a linear increase to 100% B in 10 min at a flow rate of 1 mL/min, harvesting 2 mL fractions. Elution was monitored at 280 nm and protein-containing fractions were checked by qualitative Bradford assay[144] and further analyzed by SDS-PAGE. Pooled fractions were dialyzed against 25 mM HEPES, 100 mM NaCl, pH 7.5 buffer, and concentrated with Amicon Ultra-15 concentrators.

3.2.3 Protein quantification

The final concentration of the overproduced and purified protein was determined by Bradford colorimetric assay using a BSA calibration curve. Protein aliquots were flash-frozen in liquid nitrogen and stored at -80 °C.

3.3 Analytical methods

3.3.1 HPLC-MS

High performance liquid chromatography (HPLC) was used as a standard methodology for the characterization of compounds based on the retention time on a chromatography column. Reversed-phase (RP) chromatography relies on the hydrophobic interaction between compounds and the non-polar stationary phase of the column (porous graphitic carbon or C_8 or C_{18} coated silica gel). The elution of the compounds is mediated employing a gradient with methanol or acetonitrile, which shifts the interaction of the analyte for the column towards the mobile phase. The retention time of the compounds is monitored by UV-vis detection. The mobile phases were routinely supplied with either 0.1% TFA, or 0.05% formic acid, or 20 mM NFPA as ion pairing reagents to improve chromatographic separation and electrospray mass ionization (ESI-MS) of the liquid compounds at atmospheric pressure. Routine mass-spectrometry (MS) analysis were carried out on an Agilent 1100 MSD system. Acquisition of high resolution MS data and CID fragmentation analysis of rhodochelin were carried out on a Thermo Electron LTQ-MS Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS). Protein mass fingerprinting of tryptic digests and measurements of whole proteins were carried out on a Applied Biosystem API QStar Pulsar i system.

3.3.2 Peptide mass fingerprinting

The identity of recombinant proteins was verified by peptide mass fingerprinting. Gel bands of the correct molecular mass were excised after SDS-PAGE and incubated with 200 μL of wash solution [200 mM NH_4HCO_3, 50% (v/v) acetonitrile] for 30 min at 37 °C. The solvent was subsequently removed in a speed-vac manifold to complete dryness (37 °C, 30 min). *In-band* tryptic digestion was carried out by addition of 20 μL of a trypsin solution (0.02 μg/μL trypsin, 10% NH_4HCO_3, 10% acetonitrile, pH 8.1) at 37 °C for 45 min. The excess of trypsin was removed, followed by an additional incubation period of 16 h. Peptide fragments were eluted with 25 μL of a diffusion solution [1% (v/v) TFA, 10% acetonitrile, pH 8.1] and sonicated at RT for 45 min. The samples were analyzed on a nanospray-ESI-qTOF-MS system and the subsequent comparison of the peptide mass fingerprint with the MASCOT database allowed the correct protein identification. [145,146]

3.3.3 HPLC-ESI-qTOF-MS

The *apo-*, *holo-*, and the 2,3-DHB-loaded ArCP RhcE was analyzed by HPLC-MS using a QStar Pulsar i qTOF-MS coupled to a HPLC system equipped with a C_4 Nucleosil guard column (Macherey & Nagel, 10 x 3 mm, pore diameter of 300 Å, particle size 5 μm) with the following conditions: solvent A (water + 0.45% formic acid), solvent B (acetonitrile + 0.45% formic acid), flow rate 0.2 mL/min, temperature 45 °C with a gradient of 10-95% solvent B over 10 min, the gradient was then held for 7 min.

3.3.4 Natural product isolation

For the isolation of rhodochelin, *R. jostii* RHA1 was grown for two days in LB medium. Cells were harvested, washed once and resuspended in an equal amount of M9 medium. A 1/100 aliquot was used to inoculate fresh minimal medium (in polycarbonate flasks), and cultures were grown for two days, until a CAS positive reaction of the supernatant was observed.[147] The culture supernatant was extracted with 5 g/L of XAD-16 resin for 2 h, and after a washing step, the adsorbed compounds were eluted with methanol and immediately concentrated under vacuum to dryness. The eluate was resuspended in 2 mL of water and analyzed on an Agilent 1100 HPLC system equipped with a Nucleodur C_8(ec) column 125 × 2 mm, and coupled to an ESI-MS detector, utilizing the solvent gradient water + 0.05% formic acid (solvent A) and methanol + 0.04% formic acid (solvent B), with a linear gradient from 0% to 20% B within 40 min, followed by a linear increase to 95% B in 5 min and holding B for an additional 5 min. The flow rate was set to 0.3 mL/min and the column temperature to 40 °C. The gradient was also used to analyze comparative extractions of *R. jostii* RHA1 mutants. Large-scale purification was carried out by scaling-up the described protocol for 5 L cultures. The dried eluate was dissolved in 10 mL of water and separated on a preparative Agilent 1100 preparative HPLC system equipped with a Nucleodur C_8(ec) 250 × 21. Elution was performed with the same gradient described previously, using 215 and 280 nm as wavelengths for detection and a flow rate of 16 mL/min. Siderophore-containing fractions were analyzed via CAS assay and ESI-MS. Positive fractions were pooled according to their respective *m/z*, lyophilized, and subjected to further analysis.

3.3.5 UV-vis spectroscopy

UV-vis spectra were recorded on an Ultrospec 3000 (Pharmacia) spectrophotometer. Wavescan measurements were performed within a wavelength range of 200-800 nm and a scan rate of 750 nm/min. Absorption spectra of rhodochelin and *holo-*

rhodochelin as well as the *apo*- and *holo*-rhodochelin tripeptide were recorded in water at a final concentration of 400 μM. *holo*-complexes were obtained by incubating rhodochelin or the tripeptide (400 μM) with aqueous $FeCl_3$ (400 μM) for 10 min at RT prior to the scan. Extinction coefficients were derived from the UV-vis spectra. The following values were obtained. *holo*-rhodochelin: λ_{max} 330 nm, ε = 3842.5 M^{-1} cm^{-1}, λ_{max} 395 nm, ε = 2205.0 M^{-1} cm^{-1}, λ_{max} 525 nm, ε = 1567.5 M^{-1} cm^{-1}. *holo*-rhodochelin tripeptide: λ_{max} 327 nm, ε = 3372.5 M^{-1} cm^{-1}, λ_{max} 397 nm, ε = 1430.0 M^{-1} cm^{-1}, λ_{max} 580 nm, ε = 1567.5 M^{-1} cm^{-1}.

3.3.6 IR-spectroscopy

Purified rhodochelin in a KBr disk was subjected to FT-IR-spectroscopy on a Magna-IR 750 spectrometer (Nicolet). Main signals are the following: ν = 3367.2, 1749.5, 1660.5, 1586.1, 1534.1, 1448.0, 1381.8, 1204.0, 1137.4, 1066.9, 878.9, and 748.0 cm^{-1}.

3.3.7 NMR-spectroscopy

NMR-spectroscopic structure elucidation of rhodochelin was carried out in collaboration with Dr. Xiulan Xie (Department of chemistry - NMR facility, Philipps-Universität Marburg). About 8 mg of rhodochelin were dissolved in 0.7 mL of H_2O / D_2O (9:1). Measurements were carried out on a Bruker AV600 spectrometer with an inverse broadband probe installed with z gradient. The one-dimensional spectra ^{1}H and ^{13}C, the homonuclear two-dimensional spectra DQF-COSY, TOCSY, NOESY, and ROESY, the ^{1}H-^{13}C HSQC and HMBC, and the ^{1}H-^{15}N HSQC spectra were recorded with standard pulse programs at 283 K. The TOCSY spectrum was recorded with a mixing time of 80 ms, whereas NOESY and ROESY spectra were taken at 150 and 300 ms mixing times. The 1D spectra were acquired with 65,536 data points, whereas 2D spectra were collected using 4096 points in the F2 dimension and 512 increments in the F1 dimension. For 2D spectra, 16-32 transients were used, while the ^{13}C spectrum was recorded with 12 K scans. The relaxation delay was 2.5 s. The ^{1}H chemical shifts were referenced to 4,4-dimethyl-4-silapentane sodium sulfonate (DSS) in H_2O / D_2O (9:1) at 283 K externally, whereas those of ^{13}C and ^{15}N were referenced with spectrometer default calibration. The spectra were processed with Bruker Topspin 2.1.

3.3.8 Assignment of amino acid stereochemistry via FDAA-derivatization

Assignment of the amino acid stereochemistry was achieved through rhodochelin total acid hydrolysis followed by derivatization with *N*α-(2,4-dinitro-5-

fluorophenyl)-L-alaninamide (FDAA, Marfey's reagent).[148] Purified rhodochelin (3.76 mg) was hydrolyzed in 400 µL of 6 M HCl at 99 °C and 1000 rpm for 24 h. The lyophilized hydrolysate was resuspended in 50 µL of 1 M $NaHCO_3$, and 10 µL of this solution was added to 170 µL of 1% FDAA solution in acetone. The derivatization reaction was carried out for 1 h at 37 °C and terminated by addition of 20 µL of 1 M HCl. FDAA standard derivatives of amino acids (L/D-Thr, L/D-Orn, L-hOrn) were prepared by incubation of 25 µL of 50 mM amino acid solution with 50 µL of 1% FDAA solution and 10 µL of 1 M $NaHCO_3$ for 1 h. After lyophilization, all the products were resuspended in 200 µL of 1:1 water:acetonitrile solution prior to injection (10 µL) into a HPLC-MS system equipped with a Synergi Fusion RP-80 250 × 2.0 mm column utilizing the following solvent gradient: 0-30 min, 0-30% buffer A (10 mM ammonium formate, 1% methanol, 5% acetonitrile, pH 5.2) into buffer B (10 mM ammonium formate, 1% methanol, 60% acetonitrile, pH 5.2), followed by a linear increase to 95% buffer B in 2 min and holding 95% buffer B for an additional 5 min. The flow rate was set at 0.3 mL/min and the column temperature at 25 °C. Elution was monitored in single ion mode.

3.3.9 Analytical size-exclusion chromatography

Size-exclusion chromatography was employed to determine the quaternary structure of the recombinant Rmo and Rft. 25 µL of a 50 µM Rmo or Rft solution was analyzed using a Superdex 200 5/150 GL column combined with an ÄktaPurifier system (Amersham Pharmacia Biotechnology) equilibrated with 25 mM TRIS 150 mM NaCl pH 8 buffer. Protein elution was monitored at 280 nm. Aldolase (158 kDa), ovalbumin (43 kDa), ribonuclease (13.7 kDa) and aprotinin (6.5 kDa) were used as standards to determine the molecular weight of the proteins in solution.

3.4 Chemical synthesis

3.4.1 Synthesis of L-δ-*N*-hydroxylornithine (L-hOrn)‡

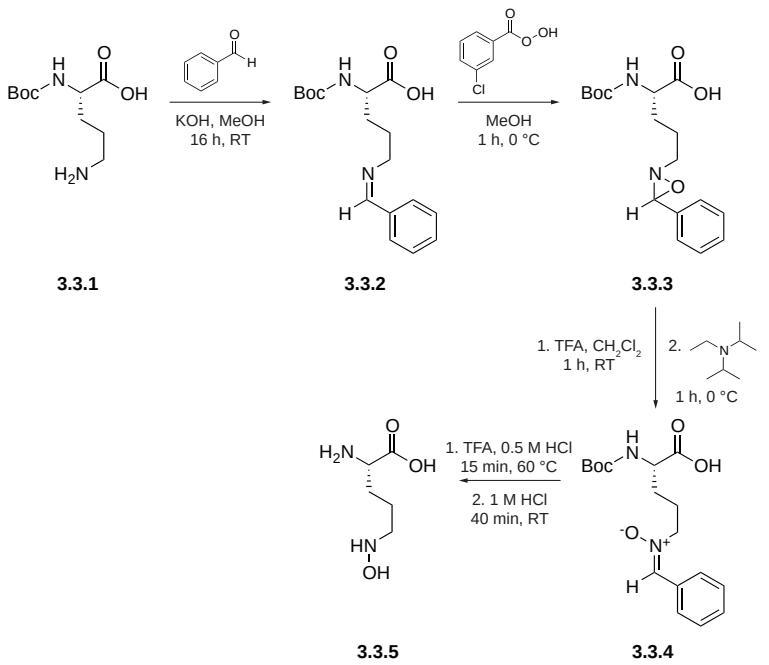


Figure 3.3 Reaction scheme of the synthesis of L-δ-*N*-hydroxyornithine (L-hOrn).

The synthesis of L-δ-*N*-hydroxyornithine was accomplished through an indirect oxidation approach, as reported by Y.-M. Lin and M. J. Jones (Figure 3.3).[149]

Step 1. Synthesis of the imine intermediate.

2 g (8.6 mmol) of the *N*α-Boc-protected L-ornithine **3.3.1** was added to a solution composed of 0.51 g KOH (9.0 mmol) dissolved in 17 mL of methanol. 0.92 mL benzaldehyde (9.0 mmol) and 3 Å molecular sieve were added and the solution was gently stirred over night at room temperature. Subsequently, the resin was filtered out and washed with methanol, whereas the filtrate was evaporated under reduced pressure resulting in a yellow foam corresponding to the imine **3.3.2**.

‡ The two non-proteinogenic amino acids L-δ-*N*-hydroxyornithine (L-hOrn) and L-δ-*N*-formylornithine (L-fOrn) were kindly provided by Dr. Verena Helmetag and their synthesis was previously published in the PhD thesis "V. Helmetag, Biochemische und strukturelle Untersuchungen der Biosynthese unnatürlicher Aminosäuren als Bausteine nicht-ribosomaler Peptide, Marburg, 2009". The synthesis of the two amino acids herein presented refers to that work, and it has only been included to improve the the sake of completeness.

Step 2. Oxidation of the imine **3.3.2** to oxaziridine.

2.2 g of the imine **3.3.2** (7.0 mmol) was dissolved in 13 mL of methanol and the solution was cooled to 0 °C. Then, a solution of 1.2 g of *meta*-chloroperbenzoic acid (7.0 mmol) in 5 mL of methanol was added over 15 min. The reaction was kept at 0 °C and allowed to proceed for 1 h, under continuous stirring. The resulting precipitate was filtered, washed with methanol and dried under reduced pressure. Subsequently, the solid compound was dissolved in 34 mL of a 1:1 ddH_2O:ethyl acetate solution, acidified to pH 2 with HCl 1 M and extracted with ethyl acetate. The recovered organic phases were washed with saturated NaCl solution, dried with Na_2SO_4 and concentrated under reduced pressure, resulting in the corresponding oxaziridine **3.3.3**.

Step 3. Synthesis of the stable nitrone intermediate **3.3.3**.

The unpurified oxaziridine precursor **3.3.3** was dissolved in 10 mL of trifluoroacetic acid and, after the addition of 10 mL of CH_2Cl_2, the reaction mixture was stirred for 1 h at room temperature. The volatile compounds were removed by evaporation and the resulting residue was dissolved in 30 mL of ethyl acetate. A solution of diisopropyethylamine in THF was added until pH 8-9 was reached. The reaction was cooled at 0 °C and allowed to proceed for 1 h, resulting in a pale-yellow precipitate which was filtered and freeze-dried. The crude product was dissolved in 10 mL of ddH_2O and purified by HPLC using a preparative C_{18}-column equilibrated with ddH_2O (0.1% TFA, solvent A). The compounds were eluted with a 20 mL / min flow rate, performing the following acetonitrile (0.1% TFA, solvent B) gradient: 0 % B for 5 min, 0-30 % B in 30 min and 30-95 % in 3 min. The elution was monitored at 215 nm and the resulting fractions were identified by ESI-MS, pooled and lyophilized, resulting in 945 mg of the corresponding nitrone **3.3.4**.

Step 4. Conversion of the nitrone **3.3.4** to the hydroxylamine **3.3.5**.

The final stage of the synthesis of L-hOrn was carried out from small quantities of the stable nitrone, as this intermediate is more stable than the final product. 0.1 g of the nitrone **3.3.4** (0.4 mmol) was diluted in 0.55 mL of hexane. Subsequently, 1.1 mL of 0.5 M HCl and 0.27 mL of TFA were added. The mixture was heated at 60 °C under reflux for 15 min. The volatile compounds were removed under reduced pressure and the resulting yellowish oil was mixed with 1.1 mL of CH_2Cl_2 and 1.62 mL of HCl 1 M. The mixture was heated at 40 °C to allow the compounds to dissolve and mixed for 40 min at RT. The organic phase was removed, whereas the

water phase was extracted with CH_2Cl_2 and hexane and concentrated, yielding L-δ-*N*-hydroxyornithine **3.3.5**, as a yellowish foam.

3.4.2 Synthesis of L-δ-*N*-formylornithine (L-fOrn)

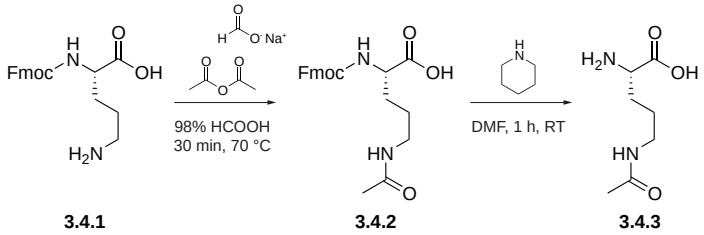


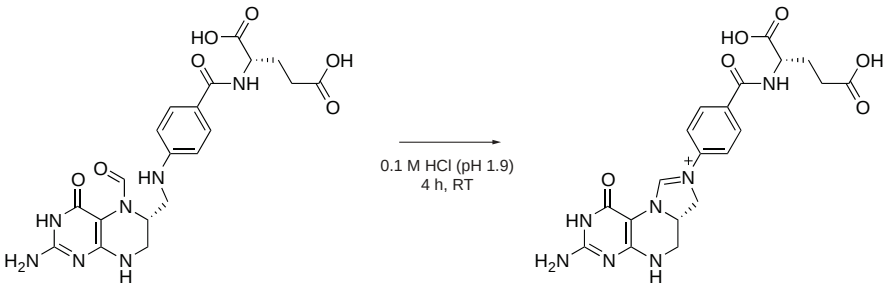
Figure 3.4 Reaction scheme of the synthesis of L-δ-*N*-formylornithine (L-fOrn).

The synthesis of L-δ-*N*-formylornithine was carried out according to the established protocol described by R. G. Jones (Figure 3.4).[150]

Step 1. Formylation of the side chain of the *N*-α-Fmoc-protected amino acid. 391 mg of *N*-α-Fmoc-protected L-ornithine **3.4.1** (1 mmol) was added to 1 mL of 98% formic acid and the mixture was heated until dissolved. After the addition of a hot solution of sodium formate (75 mg, 1.1 mmol) in 98% formic acid, the mixture was allowed to cool. Subsequently, 225 µL (2.4 mmol) of acetic anhydride was added and the reaction mixture was heated at 70 °C for 30 min. The excess of formic and acetic acid was removed under reduced pressure and the remaining residue was dissolved in 5 mL of acetone and the solvent further evaporated.

Step 2. Removal of the Fmoc-protecting group.
The formylated compound **3.4.2** was dissolved in 10 mL of DMF. 2.5 mL (25 mmol) of piperidine was added dropwise at 0 °C and the solution was stirred for 1 h at room temperature. The resulting product **3.4.3** was purified via flash-chromatography on a silica-gel column equilibrated with a 4:1 solution of methanol:ddH_2O. During the chromatography, the water content of the mobile phase was gradually increased to 100%. The product-containing fractions were analyzed by thin-layer chromatography (mobile phase methanol:ddH_2O 4:1) and freeze-dried.

3.4.3 Synthesis of the formyl-donor cosubstrate intermediate N^5,N^{10}-methenylH$_4$F



3.5.1 **3.5.2**

Figure 3.5 Reaction scheme of the synthesis of the formyl-donor cosubstrate intermediate N^5,N^{10}-methenylH$_4$F.

The formyl-donor cosubstrate intermediate N^5,N^{10}-methenylH$_4$F **3.5.2**, required for the generation of the N^{10}-fH$_4$F donor, was synthesized according to a protocol previously published, with minor modifications (Figure 3.5).[151]

In the first half-reaction, 7 mg of N^5-fH$_4$F **3.5.1** was dissolved in 1.5 mL of water and was converted to N^5,N^{10}-methenylH$_4$F **3.5.2** by drop-wise addition of 0.1 M HCl, until pH 1.9 was reached. The solution was then brought to a final volume of 2.2 mL using water and further incubated at room temperature for 4 h. The formation of N^5,N^{10}-methenylH$_4$F resulted in a color change of the solution, from colorless to pale yellow, and was further verified via ESI-MS measurements. The obtained compound (~ 6 mM final concentration) was stored at -20 °C until further usage.

3.5 Biochemical methods

3.5.1 ATP/PP_i exchange assay

The substrate specificity of the stand-alone DhbE domain was investigated by ATP/PP_i exchange. The reversibility of the 2,3-DHB-AMP formation is exploited to incorporate $^{32}PP_i$ into ATP in the backward reaction (Figure 3.6).[152,153]

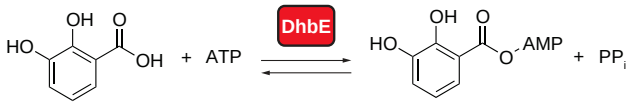


Figure 3.6 The reversible adenylation reaction exploited to investigate the substrate specificity of the adenylation domain DhbE.

Following incubation, the radioactively-labeled ATP is adsorbed on charcoal and the subsequent analysis of the total radioactivity count allows for a relative comparison of the incorporation rates among different amino acid substrates.

A 100 µL reaction was composed of the following: 50 mM Tris HCl pH 7.5 buffer, 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 5 mM $Na_4P_2O_7$, 10 mM amino acid. Prior to initiation of the reaction with 2 µM recombinant DhbE, 20 µL of a $Na_4{}^{32}P_2O_7$ solution (approx 100,000 counts) was added. The reaction was incubated at 25 °C for 30 min and subsequently quenched with 750 µL of charcoal suspension [100 mM $Na_4P_2O_7$, 600 mM $HClO_4$, 1.6% (w/v) charcoal]. After a washing step with water, the resuspended charcoal was combined with 3 mL of scintillation fluid, prior to counting with a Packard Tri-carb 2100TR liquid scintillation analyzer. All reactions were performed in triplicate.

3.5.2 Fluoresceinyl-CoA phospopantetheinylation assay

Purified recombinant PCP-domains are generally recovered in the inactive *apo* form, due to inefficient E. *coli* priming reaction (the transfer of the ppan group from the CoA donor to the conserved Ser of the core-T motif) ppan transferases during the heterologous production.[153] The *in vitro* conversion of the PCP-domain, from the inactive *apo*- to the active *holo*-enzyme, is catalyzed by the promiscous *B. subtilis* Sfp, which displays a broad substrate tolerance and is therefore suitable to catalyze the transfer of a different range CoA-derivatized substrates to different PCP-domains.[154,155] The use of fluoresceinyl-CoA, as a donor substrate, provides a direct measure of priming efficiency regarding the recombinant PCP (Figure 3.7). A typical assay was composed of 10 µM RhcE, 5 µM Sfp, 0.1 mM fluoresceinyl-CoA, 10 mM $MgCl_2$, in 50 mM TRIS buffer, pH 7.5. The reaction mixture was incubated at 37 °C

for 30 min and subjected to SDS-PAGE analysis. The labeled protein was visualized under UV-light, and the gel was subsequently stained with Comassie dye. Control reactions were carried out omitting Sfp.

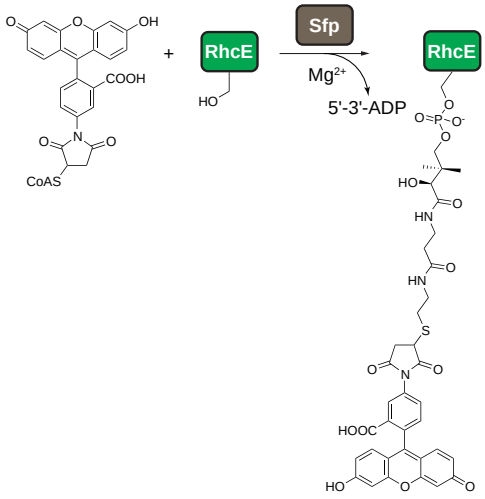


Figure 3.7 Sfp-mediated RhcE priming with fluoresceinyl-CoA.

3.5.3 RhcE priming and coupled reaction with DhbE

In order to verify the transfer of the activated 2,3-DHB-AMP to its cognate RhcE PCP, a two-step reaction assay was carried out (Figure 3.8). The first half-reaction required the priming of RhcE with the ppan arm.

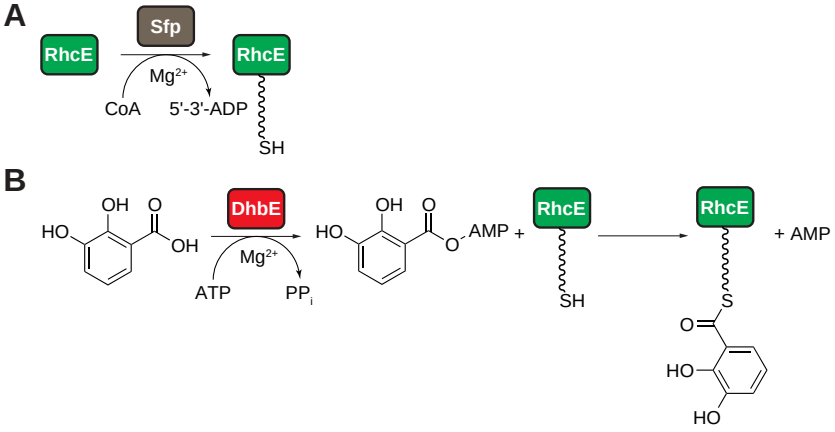


Figure 3.8 Reaction scheme of the coupled assay of DhbE and RhcE. (A) RhcE priming in the presence of Sfp, Mg^{2+} and CoA. (B) DhbE and RhcE coupled assay, where the 2,3-DHB substrate is first activated by DhbE and subsequently on the cognate primed ArCP RhcE.

A 50 μL mixture contained 50 μM RhcE, 5 μM Sfp, 0.5 mM CoA, 10 mM $MgCl_2$, in 50 mM TRIS buffer, pH 7.5 and was incubated at 37 °C for 30 min. Subsequently, 5 μM DhbE, 1 mM ATP and 1 mM 2,3-DHB were added to a final volume of 100 μL. The

mixture was further incubated at 37 °C for 30 min and analyzed via HPLC-ESI-TOF-MS.

3.5.4 Rmo-mediated L-Orn hydroxylation

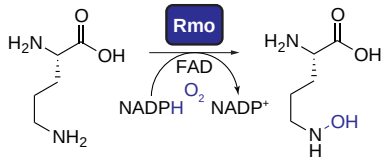


Figure 3.9 Reaction scheme of the Rmo-catalyzed FAD/NADPH-dependant δ-amino L-Orn hydroxylation.

The typical NMO Rmo is responsible for the hydroxylation of the δ-amino group of L-Orn, in the presence of the FAD cofactor and NADPH as a reducing cosubstrate (Figure 3.9). 50 μL reaction contained the following: 100 mM TRIS pH 8, 1 mM L-Orn, 2 mM NADPH, 25 μM FAD and 20 μM Rmo. Reaction controls were carried out in the absence of cosubstrate NADPH or enzyme. Different amino acids were tested as substrates to elucidate Rmo specificity. Reactions were incubated for 4 h at 30 °C and stopped by the addition of 2 μL formic acid and analyzed via HPLC-MS. Determination of the kinetic parameters for the Rmo-dependent hydroxylation was performed maintaining the NADPH and FAD concentration at 0.5 mM and 20 μM, respectively and varying L-Orn substrate concentration between 0.10 and 20 mM. Reactions were started by the addition of Rmo to a final concentration of 5 μM and stopped with 2 μL of formic acid after 2.5 min incubation. Product formation was quantified via HPLC-MS, using a L-hOrn calibration curve. All reactions were perfomed in triplicate.

3.5.5 *In situ* N^{10}-fH$_4$F conversion and L-hOrn formylation assay

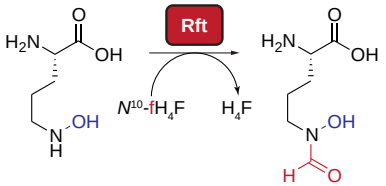


Figure 3.10 Reaction scheme of the Rft-catalyzed δ-amino L-hOrn formylation.

The N^{10}-fH$_4$F-dependent enzyme Rft catalyzes L-hOrn formylation (Figure 3.10). The reaction assay was set-up in a 50 μL volume, in the presence of 50 mM HEPES buffer pH 7.5, 1 mM L-hOrn, 1.5 mM N^5,N^{10}-methenylH$_4$F and 25 μM Rft (or CchA).

Prior to the addition of the amino acid and the enzyme, the cosubstrate intermediate was pre-incubated in the reaction buffer at 30 °C for 30 min to permit the final pH-dependent conversion to N^{10}-fH_4F. After the addition of the substrate and the enzyme, the reaction was allowed to proceed for 4 h at 30 °C, then stopped by the addition of 2 µL formic acid and analyzed via HPLC-MS.

The kinetic parameters for Rft-mediated transformylation were determined by maintaining the N^5,N^{10}-methenylH_4F concentration at 1.5 mM, and varying the L-hOrn concentration between 0.25 and 15 mM. Reactions were started by adding Rft to a final concentration of 5 µM and stopped with 2 µL formic acid after 5 min incubation. Product formation was quantified via HPLC-MS. All reactions were performed in triplicate.

3.5.6 Coupled L-Orn hydroxylation and formylation

A coupled hydroxylation and formylation assay was established to identify an enzymatic pathway for the generation of the iron-coordinating L-fhOrn. The reaction assay was performed in HEPES buffer, and prior to the addition of the amino acid and the enzymes, the cosubstrate intermediate was pre-incubated in the reaction buffer at 30 °C for 30 min to permit the final pH-dependent conversion to N^{10}-fH_4F. A typical assay in HEPES buffer contained 1 mM L-Orn, 2 mM NADPH, 1.5 mM N^5,N^{10}-methenylH_4F, 40 µM FAD, 25 µM Rmo and 25 µM Rft. The reaction was carried out for 4 h, stopped with 2 µL formic acid and analyzed via HPLC-MS.

3.5.7 HPLC-MS analysis of the L-Orn tailoring reactions

All Rmo- and Rft-catalyzed assays were analyzed by HPLC-MS on a Hypercarb column equilibrated with 20 mM aqueous NFPA (solvent A). Acetonitrile (solvent B) was used to employ a linear gradient from 0% to 15% B within 25 min, followed by a linear increase to 100% B in 2 min and finally holding B for an additional 3 min. The flow rate was set to 0.2 mL/min and the column temperature to 20 °C. The elution was monitored in single-ion mode.

3.6 Bioinformatic Methods

Web-browser-based and stand-alone programs were used to analyze nucleic acid and protein sequences. Genome analysis and operon visualization was performed with Artemis.[156] Given a genome sequence, ANTIsmash allowed a rapid identification of the major known secondary metabolites gene clusters.[157] The prediction of the adenylation domain substrate specificity was carried out with NRPSpredictor and its updated version NRPSpredictor2.[40,158] These tools analyze the physico-chemical properties of the residues surrounding the A-domain substrate binding pocket, extending the ten amino acid code defined by Stachelhaus *et al.*, and thus, improving the overall prediction.[39] Sequence-homology searches were executed using the BLAST algorithm, and multiple sequences alignments were conducted with ClustalW, ClustalO or Muscle.[159-162] Genome-BLAST comparisons were carried out using Mauve.[163] All sequences were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/).

Chapter 4

Results

4.1 Isolation and structural characterization of rhodochelin

4.1.1 Extraction and purification of rhodochelin

In bacteria, the biosynthesis of siderophores is a tightly regulated event. In order to force the microorganism to produce these types of secondary metabolites, *R. jostii* RHA1 was cultivated in M9 minimal medium under iron starvation conditions. The production of an iron-scavenging compound was confirmed via chromeazurol S (CAS) liquid assay and could be observed two days after the inoculum.[147] The supernatant was subsequently extracted with XAD-16 resin and analyzed via HPLC-MS. As shown in the chromatograms illustrated in Figure 4.1, the application of iron-deficient grow conditions resulted in two new major compounds. The first one (compound **4.1**, retention time, t_R 17.6 min) showed a *m/z* of 572.2201 ($[M+H]^+$), whereas the second one (compound **4.2**, t_R 21.4 min) displayed a *m/z* of 414.1510 ($[M+H]^+$).

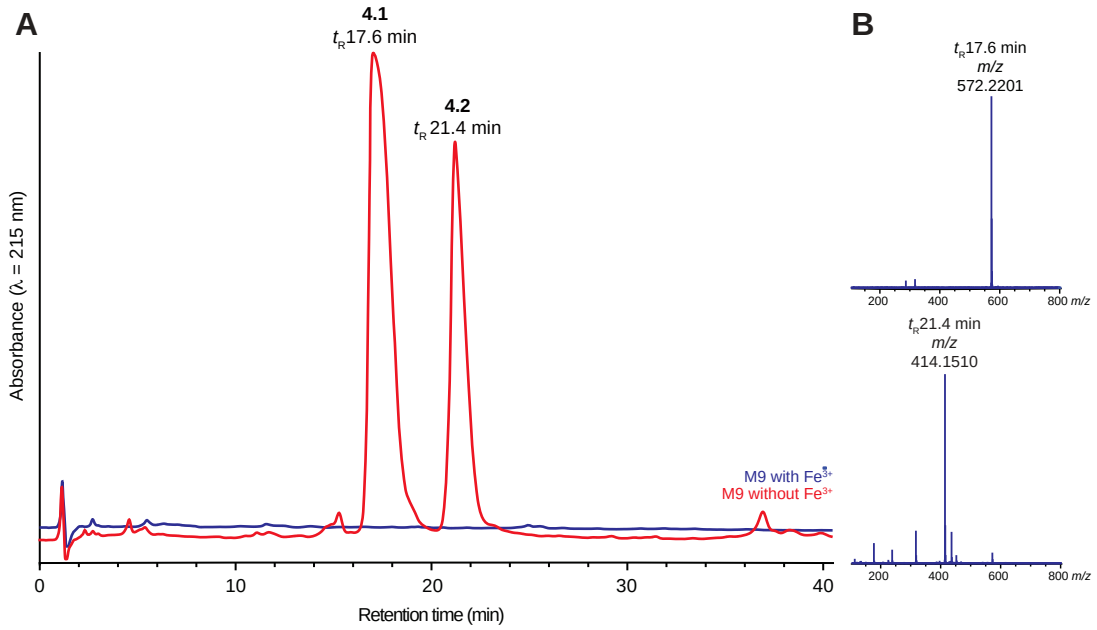


Figure 4.1 (A) HPLC-MS profiles of XAD-16 extracted culture supernatants of *R. jostii* RHA1 grown in M9 minimal medium in the absence (red line) or presence (blue line) of Fe^{3+}. The absence of ferric iron induces the production of two compounds which elute at t_R 17.6 min and t_R 21.4 min, respectively. (B) Detailed high-resolution mass spectra corresponding to the UV signals.

In order to obtain sufficient amounts of sample material required for structure elucidation, the bacterial culture was scaled-up to a volume of 5 L. Preparative HPLC showed a similar "two-peak" profile (Figure 4.2). Single fractions were analyzed via CAS liquid assay and ESI-MS and subsequently pooled according to

their *m/z* (bioassay-guided fractionation). Subsequent structure elucidation studies were performed using the compound **4.1** with a *m/z* of 572.2, which was obtained with a yield of to 13 mg/L. From now on, this compound will be referred to as rhodochelin. As will be shown later, comound **4.2** is a degradation product of rhodochelin and will thus be referred to as rhodochelin tripeptide.

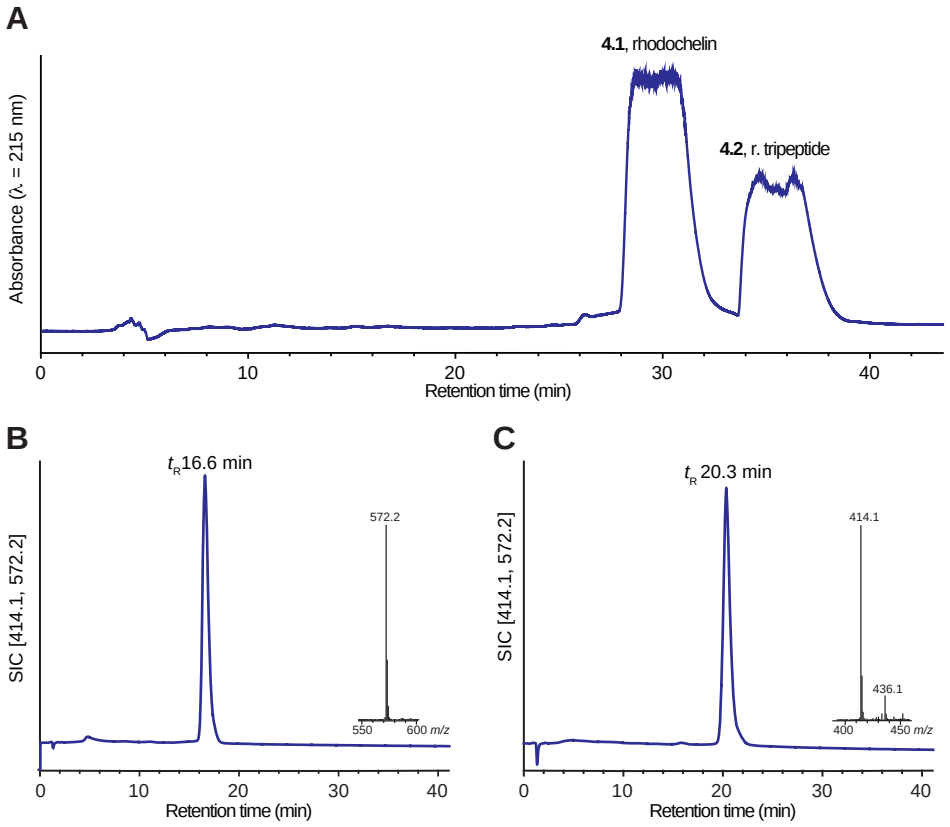


Figure 4.2 (A) Chromatogram of the preparative HPLC purification of the XAD-16 extracted culture supernatant of *R. jostii* RHA1 grown in iron-limited conditions. (B) HPLC-MS analysis and corresponding mass spectra of the purified compound 4.1 (***m/z* 572.2 [M+H]$^+$**). (C) HPLC-MS and corresponding mass spectra of the purified compound** 4.2 (***m/z* 414.1 [M+H]$^+$**, ***m/z* 436.1 [M+Na]$^+$**).**

4.1.2 MS^n analysis of rhodochelin composition

To acquire additional information about the individual building blocks of rhodochelin and their connectivity, collision induced dissociation (CID) experiments were carried out. MS^2 fragmentation of rhodochelin with a *m/z* of 572.2 ([M+H]$^+$) revealed a major fragment ion with a *m/z* of 396.1402 ([M+H]$^+$, calculated 396.1401), due to the loss of a δ-*N*-formyl-δ-*N*-hydroxyornithine (fhOrn) moiety. MS^3 experiments of the aforementioned fragment ion led to the formation of two additional ions with a *m/z* of 177.0870 and *m/z* of 238.0711. The first one could easily be associated with a second fhOrn residue ([M+H]$^+$, calculated 177.0870), whereas

the second one is indicative of a charged DHB-Thr dipeptide ($[M]^+$, calculated 238.0710) (Figure 4.3). In summary, the fragmentation studies led to the conclusion that rhodochelin consists of two fhOrn moieties, one 2,3-DHB residue, and one moiety of Thr (directly connected to the aryl residue). High resolution MS analysis confirmed an exact *m/z* of 572.2201 ($[M+H]^+$, calculated 572.2198), consistent with the molecular mass of a compound with a chemical formula of $C_{23}H_{33}N_5O_{12}$. High resolution MS experiments conducted on rhodochelin tripeptide revealed it to be a derivative of rhodochelin, lacking the second fhOrn moiety (*m/z* 414.1510 $[M+H]^+$, calculated 414.1507), suggesting that it represents a degradation product of the intact rhodochelin molecule.

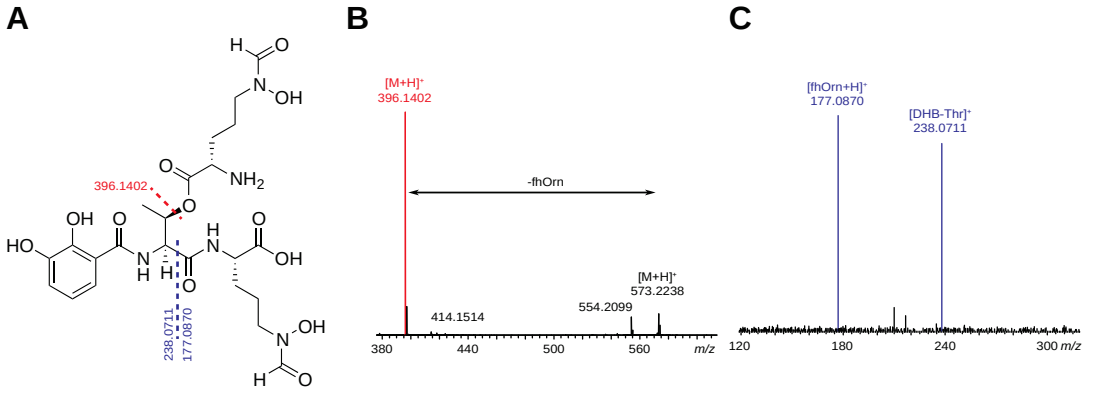


Figure 4.3 (A) Structure of rhodochelin highlighting the main fragments resulting from CID experiments. (B) Gas-phase fragmentation of rhodochelin (MS^2) resulted in the formation of a major fragment of *m/z* 396.1402 $[M+H]^+$, indicative of the loss of a fhOrn building block. (C) MS^3-fragmentation of the former daughter ion gave rise to the formation of a second fhOrn ion (*m/z* 177.0870 $[M+H]^+$) and a third charged fragment of *m/z* 238.0711 ($[M]^+$) coherent with a 2,3-DHB-Thr dipeptide.

4.1.3 Structure elucidation of rhodochelin via NMR

On the basis of the knowledge that our target rhodochelin contains one 2,3-DHB, one threonine (Thr), and two δ-*N*-formyl-δ-*N*-hydroxyornithine (fhOrn) moieties, the final structure was determined by NMR spectroscopy (Figure 4.4). The assigned ^1H, ^{13}C, and ^{15}N chemical shifts are listed in Table S1. The ^1H spectrum showed two doublets at 9.169 and 8.505 ppm for the amide protons of Thr^1 and $fhOrn^2$, respectively. Two singlets at 7.893 and 7.826 ppm were observed for the formamide protons of $fhOrn^2$ and $fhOrn^3$, respectively. A second set of peaks was also observed in this region, which corresponds to a minor stereoisomer of the siderophore in water at 283 K. Only chemical shifts of the major stereoisomer are listed in Table S1. Two cross-peaks were observed in the ^1H-^{15}N HSQC spectrum, which verified the presence of two amide bonds in the molecule. TOCSY cross-peaks additionally confirmed the presence of one threonine and two ornithines. NOE contacts between

the amide proton of Thr[1] (NH^1) and H^4 of 2,3-DHB and NH^2 and HR^1 were observed, which revealed the partial structure 2,3-DHB-Thr[1]-fhOrn[2]. Furthermore, long-range NOE contacts of Hβ^1 and Hγ^1 with HR^3 were detected. A long-range ^{1}H-^{13}C correlation was observed between Hβ^2 and the carbonyl carbon of fhOrn[3]. Combining all this information about the long-range connections, the presence of an ester bond between the carbonyl of fhOrn[3] and the side chain hydroxyl of Thr[1] was established (Figure 4.4), and further verified via IR-spectroscopy (Figure S1). The corresponding COSY, ROESY, and HMBC spectra are shown in Figures S2-S4.

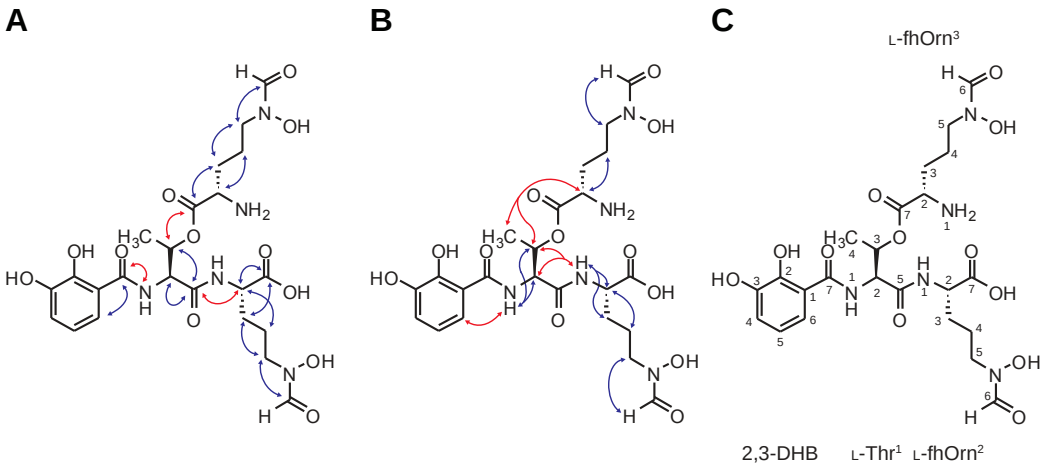


Figure 4.4 The structure of rhodochelin as determined by NMR. (A) Long-range ^{1}H-^{13}C correlations observed in H_2O/D_2O (9:1) at 283 K: blue arrows indicate intra-residue contacts, red arrows indicate long-range inter-residue contacts. (B) NOE contacts observed in H_2O/D_2O (9:1) at 283 K: blue arrows indicate intra-residue contacts, red arrows indicate long-range inter-residue contacts. (C) The structure of rhodochelin showing the numbering of the four building blocks.

4.1.4 Assignment of rhodochelin stereochemistry

The assignment of the amino acid stereochemistry was carried out by derivatization of the acid hydrolyzate of rhodochelin with FDAA (Marfey's reagent).[148] The derivatized mixture was then subjected to HPLC-MS analysis and compared with amino acid standards (L/D-Thr, L/D-Orn, L-hOrn), previously modified with the same derivatization reagent. Interestingly, during the acid hydrolization step, the loss of formyl moiety from fhOrn occurred, leading to the generation of hOrn. Furthermore, the reaction of this latter amino acid with FDAA in acetone resulted in the formation of a compound known to be its nitrone derivative, as previously reported during the synthesis of ε-*N*-acetyl-ε-*N*-hydroxylysine.[164] The comparison of the HPLC-MS chromatograms and the MS spectra of the derivatized acid hydrolysate with the synthetic amino acid standards revealed the sole presence of

DNPA-L-Thr and DNPA-L-hOrn as the constituents of the target compound (Figure 4.5).

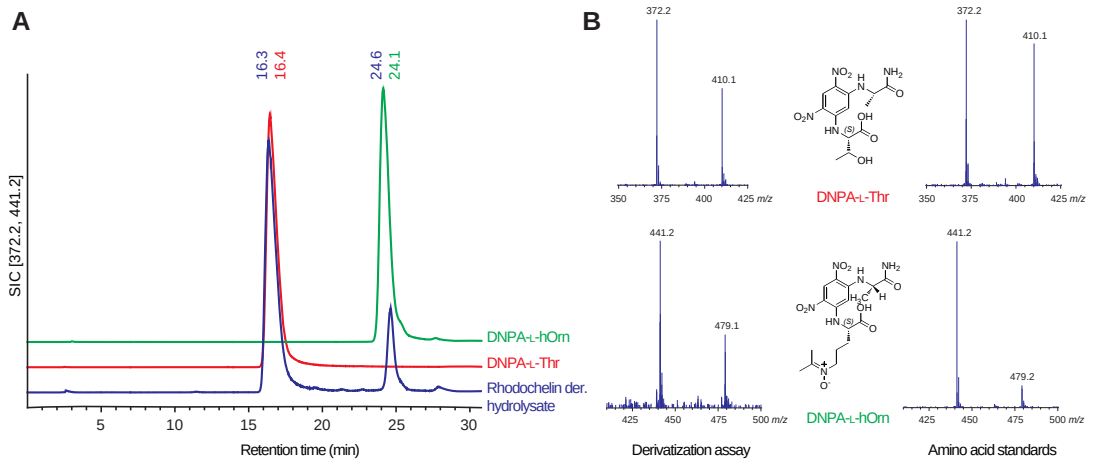


Figure 4.5 HPLC-MS profile of the FDAA-derivatized amino acid standards and the derivatized products of rhodochelin hydrolysis. (A) Rhodochelin was first hydrolyzed in HCl and the lyophilized products of the reaction were derivatized with FDAA. The reaction mixture was then analyzed via HPLC-MS and compared with synthetic derivatized amino acid standards (the respective retention time and MS-spectrum of each compound is given). (B) The chemical structures and the observed MS-spectra of the derivatized amino acid standards and the derivatized products of rhodochelin are given.

4.1.5 Physico-chemical properties of rhodochelin

In order to determine if rhodochelin and rhodochelin tripeptide possess a physiological function as iron chelating compounds, ferric iron complexes of rhodochelin and rhodochelin tripeptide were analyzed via UV-vis spectroscopy and HPLC-MS. Both compounds retained the capacity to complex Fe^{3+}, which is reflected both in the UV-vis spectra and HPLC-MS chromatograms. In particular, binding of the ferric ion altered the spectral properties of the siderophore, resulting in a shift of the absorption peak from 315 to 330 nm (corresponding to the $\pi \rightarrow \pi^*$ transition of the catechol group) and in the occurrence of two new absorption peaks at 395 and 525 nm, indicative of the *ferric*-hydroxamate and *ferric*-catecholate charge transfer, respectively.[165-167] Furthermore, the chelation of iron induced a strong shift in the chromatographic retention, suggesting that a drastic conformational going from *apo*- to *holo*-rhodochelin occurred, resulting in an altered hydrophilicity, and thus behavior, of the complex on the HPLC column (Figure 4.6). HR-MS analysis confirmed the identity and the 1:1 stoichiometry of the of *ferric*-rhodochelin complex, having a *m/z* of 625.1299 ($[M\text{-}H+Fe^{3+}\text{-}3H]^+$, calculated 625.1302). Besides, similar results were also obtained for rhodochelin tripeptide, the ester hydrolysis product of the siderophore, recovered during the extraction procedure. The

complete UV-vis and HPLC-MS analysis of the *apo*- and *holo*-rhodochelin tripeptide is shown in Figure S5.

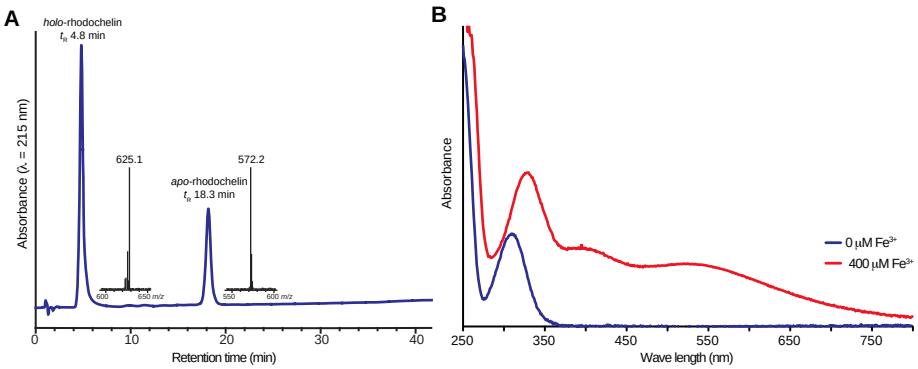


Figure 4.6 (A) HPLC-MS profile of Fe-loaded rhodochelin. (B) UV-vis spectrum of 400 μM *apo*-rhodochelin (blue trace) and *holo*-rhodochelin (red trace).

4.2 Identification of the rhodochelin biosynthetic gene clusters

4.2.1 Identification of the rhodochelin biosynthetic genes via genome mining

In order to associate a gene cluster with rhodochelin biosynthesis, a genome-mining analysis of the sequenced and annotated genome of *R. jostii* RHA1 identified three gene clusters putatively responsible for assembly, export, and subsequent uptake of a peptide siderophore (Figure 4.7).[124,168]

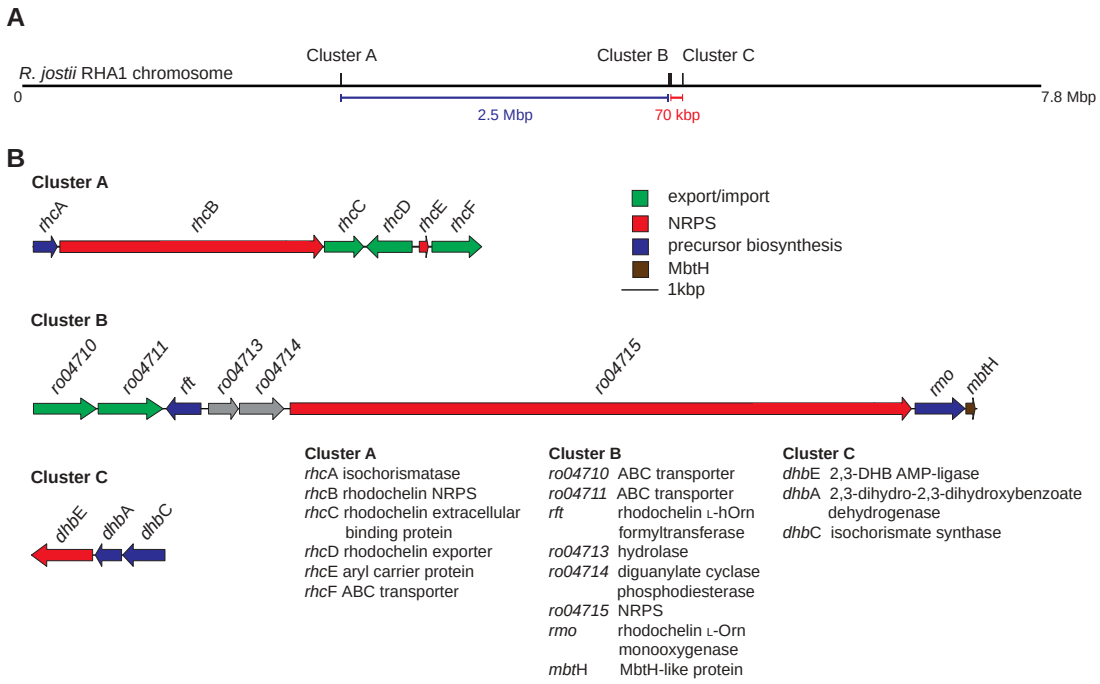


Figure 4.7 (A) Representation of the linear *R. jostii* RHA1 chromosome and localization of the three gene clusters involved in rhodochelin biosynthesis. (B) Schematic overview of the *rhc* gene cluster A that is responsible for rhodochelin assembly. Cluster C encodes for the tailoring enzymes required for the biosynthesis of the non-proteinogenic amino acid L-fhOrn. The *dhb* gene cluster C is responsible for the synthesis of 2,3-DHB. Genes are differently colored by proposed function. Putative functions of the proteins encoded within the operons are based on BLAST-analysis and are given in the figure.

The first cluster (cluster A) is composed of six genes and is located in a region covering approximately 12 kbp between ORFs *RHA1_ ro02318* and *RHA1_ro02323*; these genes have been renamed from *rhc*A to *rhc*F. *rhc*B is the largest gene (7.1 kbp) and encodes a typical modular NRPS, composed of two complete modules and a terminal thioesterase (TE) domain. Substrate-specificity prediction for the adenylation (A) domains suggests a preference for L-Thr for the first module, and L-Asp for the second module (Table 4.1 and Figure 4.8).[39]

Table 4.1 Comparison of the extracted active-site residues determining the adenylation domain specificity of RhcB with known adenylation domains. The substrate prediction for each A-domain as well as the product of the NRPS is given, whereas the differences are highlighted in red. CDA, calcium-dependent antibiotic.

A-domain	active site residues	substrate	product
RhcB-A_1	DFWNVGMVHK		
CDA PSI-A_2	DFWNVGMVHK	L-Thr	CDA
RhcB-A_2	DLWGMGAVNK		
CDA PSI-A_4	DLTKIGAVNK	L-Asp	CDA

The protein sequences of *rhc*A and *rhc*E show sequence homology to the single domains of DhbB found in the bacillibactin (*B. subtilis*) gene cluster. RhcA has been annotated as an isochorismatase (ICL), resembling the N-terminal domain of DhbB, whereas RhcE is homologous to the C-terminal aryl carrier protein (ArCP) domain.[103] The remaining three ORFs located in the cluster are involved in siderophore export (*rhc*D) and uptake: *rhc*C has been proposed encoding for an extracellular Fe^{3+}/hydroxamate binding protein, whereas *rhc*F encodes for an ABC type 2 transporter.

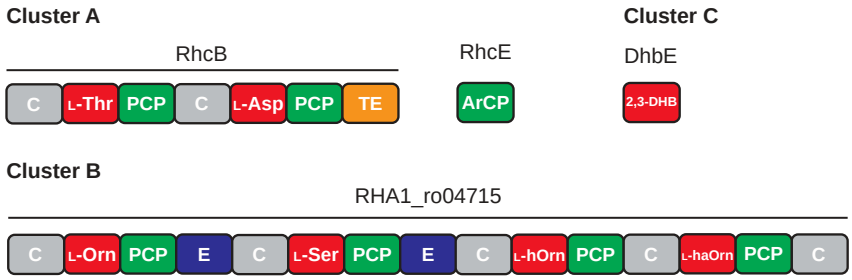


Figure 4.8 Representation of the modular organization of the NRPS RhcB and the stand-alone NRPS domains RhcE and DhbE that are responsible for rhodochelin biosynthesis. The NRPS RHA1_ro04715 located in cluster 2 is not involved in the biosynthesis of the siderophore. Neither its modular organization nor the adenylation domain substrate specificities are in accordance with the linear logic of the rhodochelin NRPS assembly line. The domains are colored differently by proposed function: grey for condensation (C), red for adenylation (A), green for peptidyl carrier protein (PCP), blue for epimerization (E) and orange for the thioesterase (TE) domains. The prediction of the substrate specificity of the adenylation domains is indicated within the module.

The absence of genes encoding for tailoring enzymes involved in the biosynthesis of the fhOrn moiety led to the identification of a second gene cluster responsible for siderophore biosynthesis located in another genomic region. Cluster B is composed of eight genes, spanning a DNA region of approximately 25 kbp, located between ORFs *RHA1_ro04710* and *RHA1_ro04717*. The largest gene is *RHA1_ro04715* (16.7 kbp) and encodes for a tetramodular NRPS that lacks the terminal TE-domain

(substituted by a C-domain). In addition, both the first two modules also contain an epimerization domain. This large NRPS gene shows high sequence homology (51% identity, 64% similarity) to EtcD, the NRPS responsible for erythrochelin biosynthesis, a siderophore isolated from *S. erythraea*.[14,169] The modular organization of the NRPS exactly matches EtcD, and, in addition, the comparison of the adenylation domain substrate specificity prediction (Table 5.2), together with the presence of two tailoring enzymes annotated as a formyltransferase (ORF *RHA1_ro04712*, renamed *rft*) and an ornithine monooxygenase (ORF *RHA1_ro04716*, renamed *rmo*), respectively, led to the hypothesis that this cluster could be responsible for the synthesis of a foroxymithine derivative (Figure 5.4).[170] In addition, the 3' end of *rmo* overlaps the 5' end of the following ORF (*RHA1_ro04717*) by four nucleotides. Due to the high sequence identity of the latter ORF with MbtH-like family proteins, it was renamed *mbtH*.[171] Interestingly, this gene is present as a single copy in the *R. jostii* RHA1 genome. ORFs *RHA1_ro04710* and *RHA1_ ro04711* are predicted to encode ABC-transporters, while *RHA1_ro04713* and *RHA1_ro04714* are thought not to be directly involved in rhodochelin biosynthesis and/or transport, as they share homology with hydrolases and diguanilate cyclase/phosphodiesterase. Taken together these results support the hypothesis that while the *rhc* cluster is responsible for the synthesis of the catecholate-hydroxamate siderophore rhodochelin and its export and uptake, the second cluster contains tailoring enzymes required for the synthesis of the L-fhOrn building block. In addition, cluster B contains the MbtH-like protein that was shown to be an essential cofactor for the amino acid activation in some other systems, especially under severe growth conditions, such as iron starvation.[172,173]

It is important to note that the incorporation of a 2,3-DHB building block into the NRPS assembly line of bacillibactin requires its activation as an adenylate:[174] this reaction is carried out by the aryl acid-activating domain DhbE.[35] Previous studies reported that the presence of 2,3-DHB moieties within catecholate siderophores requires additional genes necessary for its biosynthesis from the chorismate precursor.[175] In this study, no genes encoding enzymes involved in the chorismate pathway have been annotated in the surroundings of both gene clusters, except for the ICL RhcA. A BLAST search using DhbE as a query led to the identification of a homologue (*RHA1_ro04793*) in another locus of the *R. jostii* RHA1 genome (cluster C).[159] This gene was renamed to *dhb*E. In addition, two more genes located upstream of the adenylation domain, which where previously annotated as an isochorismate synthase (*dhb*C) and a 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (*dhb*A), have been found, identifying all the enzymes necessary for the biosynthesis of the

2,3-DHB moiety from its chorismate precursor.[176] A complete bioinformatic overview of the gene clusters involved in rhodochelin biosynthesis is presented in Table S2.

4.2.2 Construction of isogenic deletion mutants in *R. jostii* RHA1 and test for rhodochelin activity

In order to verify the hypothesis of distantly located gene clusters involved in the biosynthesis of the same natural product, gene deletion studies in *R. jostii* RHA1 were performed. *rhc*B, *dhb*E, *rmo*, and *rft* were chosen as targets because they were considered to be essential for the assembly of the natural product or the biosynthesis of its non-proteinogenic L-fhOrn precursor and, at the same time, were located in the three different chromosomal loci. Isogenic deletion mutants were constructed from the RHA1 *wild-type* strain employing a markerless *in-frame* gene deletion approach, which has the advantage to avoid any polar effects on downstream genes.[177] As a result, four new strains were generated: RHA1Δ*rhc*B, RHA1Δ*dhb*E, RHA1Δ*rmo*, and RHA1Δ*rft*. The correct *in-frame* deletion was verified by PCR amplification using primer pairs flanking the mutation, which resulted in a shorter amplicon in the mutant strain (Figure S6). Culture supernatants were analyzed for CAS activity, but no strain was capable of producing an iron-chelating compound (Figure S7). This result was further confirmed by comparing the HPLC-MS traces of extracted supernatants, which lacked the "two-peak" profile, typical for rhodochelin production (Figure 4.9).

RHA1Δ*rmo* and RHA1Δ*rft* extracts were furthermore analyzed via extracted ion chromatograms (EIC) for masses corresponding to rhodochelin derivatives lacking the *N*-hydroxy or *N*-formyl groups (data not shown). MS analysis of these supernatant extractions confirmed an abolished production of rhodochelin. In addition, no foroxymithine derivative has been detected. These results clearly demonstrate that these four genes are directly associated with rhodochelin biosynthesis, even if they are located in three different genomic regions within the bacterial chromosome.

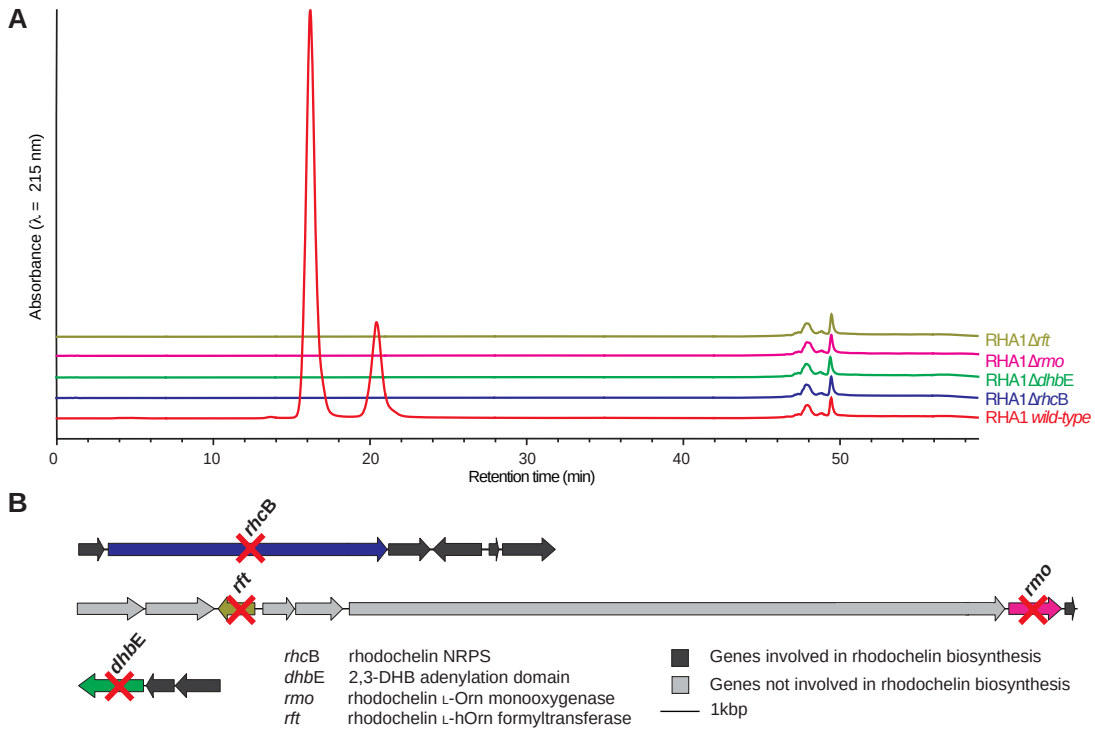


Figure 4.9 (A) Comparative HPLC-MS profiles obtained from the extracts of the culture supernatants of *R. jostii* RHA1 *wild-type* strain and nonproducing rhodochelin mutant strains. The single deletion of one of any four genes (which are located in three different chromosomal loci) resulted in an abolished rhodochelin production, implying an unprecedented cross-talk mechanism. (B) Schematic overview of the genes (and their relative location) that have been deleted from the bacterial chromosome.

4.3 Biochemical characterization of rhodochelin NRPS assembly-enzymes

4.3.1 DhbE ATP/PP_i exchange

In order to verify the adenylating activity and the substrate specificity of DhbE, the corresponding gene was cloned as an N-terminal His-tag fusion in pET28a(+), heterologously produced in *E. coli* and purified via affinity chromatography. The substrate specificity of the purified recombinant enzyme has been tested using several amino acids and two aryl acid isomers. In an ATP/PP_i exchange assay, following 30 min incubation in the presence of an excess of $Na_4{}^{32}P_2O_7$, DhbE showed a distinct preference for its cognate substrate 2,3-DHB, validating the bioinformatic prediction (Figure 4.10).

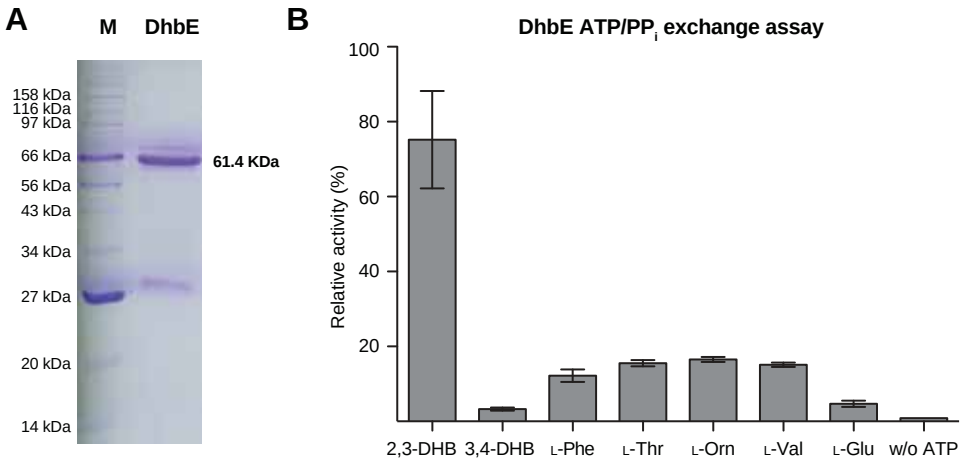


Figure 4.10 (A) SDS-PAGE of purified recombinant DhbE. (B) Relative activities obtained from the ATP/PP_i exchange assay for DhbE. To test the substrate specificity of the enzyme, DhbE was incubated in the presence of an excess of $Na_4{}^{32}P_2O_7$ and the different amino acids for 30 min. The reaction was quenched with charcoal and end-point radioactive-ATP was measured. Error bars represent standard deviations from three independent measurements.

4.3.2 Coupled assay of DhbE and RhcE

A coupled reaction assay was set up in order to verify the ability of the stand-alone ArCP *holo*-RhcE to accept the activated 2,3-DHB-AMP from the previously characterized DhbE-catalyzed adenylation reaction (Figure 4.11 A). RhcE was expressed as a C-terminal His-fusion protein in pET28a(+), heterologously produced in *E. coli* and purified via affinity chromatography. Initially, the *apo*-purified carrier protein was tested for the ability to be converted to its active *holo*-form, using the CoA analogue fuoresceinyl-CoA (Fl-CoA). In the presence of the phosphopantetheinyl transferase Sfp, the recombinant RhcE could be fluorescently labeled, as shown in the SDS-PAGE analysis of the enzymatic mixture under UV-light (Figure 4.11 B). Therefore, a coupled assay was set up in a similar fashion, in a

two-step reaction, the first being the priming of the *apo*-RhcE with CoA. As shown in Figure 4.11 C, the ESI-qTOF-MS measurement of the priming reaction displayed a mass shift of 341 Da, consistent with the covalent modification of the catalytically active Ser residue with the CoA ppan arm, which is indicative of the conversion from *apo*- to *holo*-RhcE. When the primed RhcE was incubated in the presence of ATP, 2,3-DHB and the stand-alone adenylation domain DhbE, a shift of 476 Da occurred, suggesting the transfer of the activated 2,3-DHB-AMP to the free thiol group of the ppan arm, thus resulting in RhcE-*S*-2,3-DHB (Figure 4.11 D).

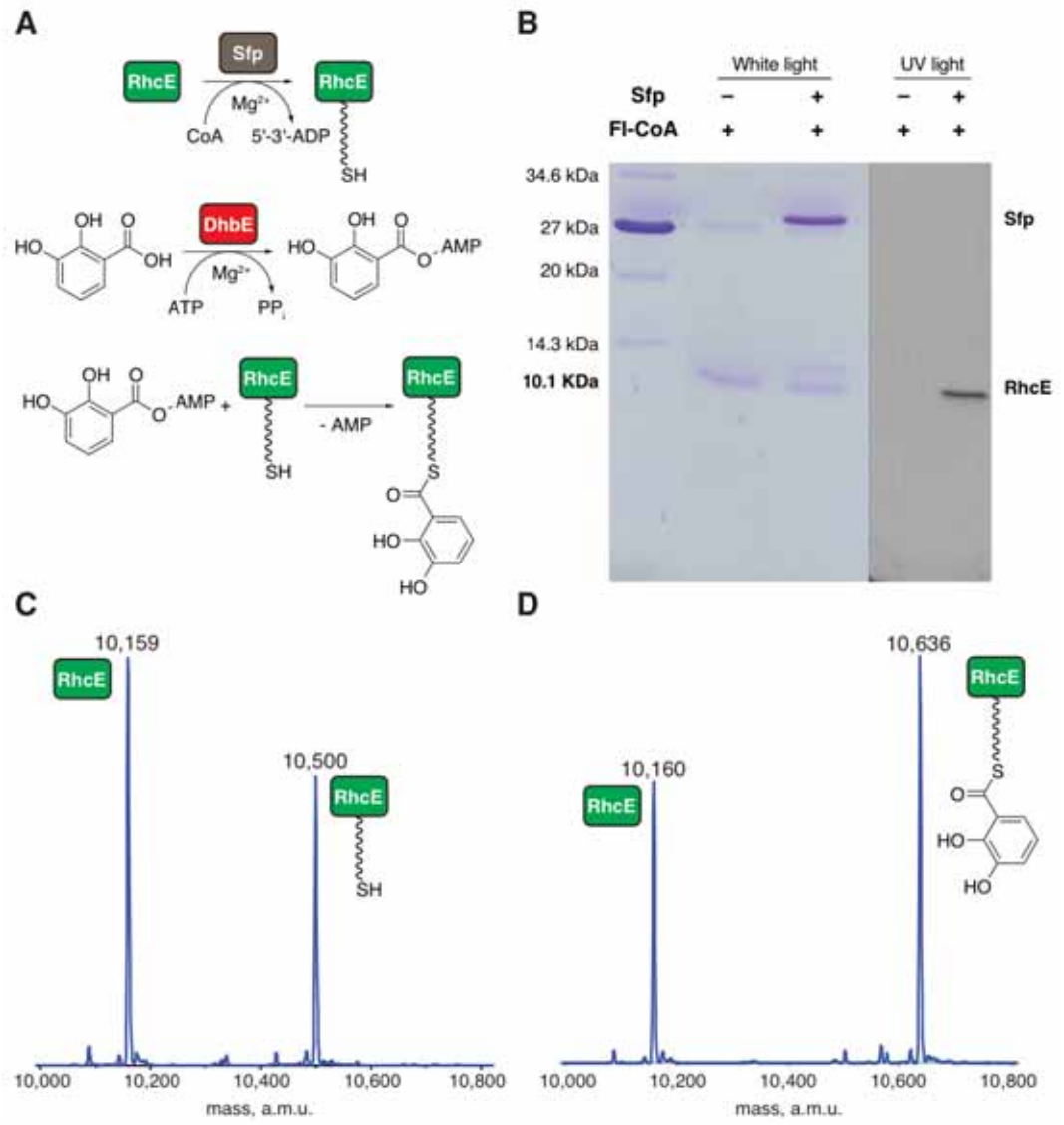


Figure 4.11 Coupled reaction of DhbE and RhcE. (A) Schematic overview of the conversion from *apo*- to *holo*-RhcE in the presence of Sfp, CoA and Mg^{2+} and subsequent coupled assay reaction with 2,3-DHB-AMP, catalyzed by DhbE. (B) SDS-PAGE analysis of the fluoresceinyl-labeled RhcE. When Sfp was present in the mixture, the transfer of the fluorescent group from the Fl-CoA donor to the catalytically active Ser residue of the PCP-domain occurred, resulting in a fluorescent SDS-PAGE band under UV-light. (C) Deconvoluted ESI-qTOF-MS

measurements of the Sfp-catalyzed RhcE priming reaction. The mass shift of 340 Da is indicative of the transfer of the ppan arm from the CoA donor to the catalytically active Ser residue of the PCP, resulting in the conversion from *apo*-enzyme to *holo*-enzyme. (D) When the primed RhcE was incubated in the presence of 2,3-DHB, ATP and DhbE, the coupled transfer of the DhbE-activated 2,3-DHB to the free thiol group of the ppan arm occurred, suggesting the biochemical cross-talk between the two proteins, that are encoded in two distantly-located genome regions.

4.4 Biochemical characterization of L-Orn tailoring enzymes

4.4.1 Biochemical characterization of Rmo L-Orn Monooxygenase

4.4.1.1 Bioinformatic analysis of the NMO Rmo

In the previous section, it has been demonstrated that *rmo* encodes for a putative L-Orn monooxygenase which is involved in the biosynthesis of rhodochelin, since its deletion from the chromosome resulted in an abolished siderophore production.

A bioinformatic analysis showed that Rmo belongs to the class of NAD(P)H/FAD-dependent monooxygenases and, when compared to already characterized homologues (CchB, EtcB, PvdA, SidA, and IucD), it displays high similarity along the overall sequence.[88,94,178-181] As depicted in the multiple sequence alignment shown in Figure 4.12, highly conserved motifs involved in substrate, NADPH, and FAD binding could be easily identified. The FAD cofactor-binding consensus GxGxxN is located at the N-terminus region of the protein in a conserved dinucleotide-binding βαβ-motif. A similar α/β-nucleotide-binding architecture is responsible for the binding of the NADPH cosubstrate, enclosed in a similar GxGQS motif. Lastly, the residues involved in binding the L-Orn substrate are located in a small domain located at the interface between the largest FAD and NADPH binding motifs (an homology model is presented in Figure 5.5)[15,182]

4.4.1.2 Recombinant production and purification of active apo*-Rmo*

rmo was amplified from the *R. jostii* RHA1 chromosome and cloned into the pET28a(+) expression vector. The recombinant protein was heterologously produced in *E. coli* as an N-terminal His-tag fusion and purified via Ni-NTA affinity chromatography. UV-vis spectroscopy analysis of the purified recombinant enzyme indicates Rmo to be in the *apo* form, without bound FAD cofactor (Figure 4.13).

Analytical size exclusion chromatography estimated the molecular mass of Rmo in solution to be equal to 220 kDa, suggesting the enzyme to adopt a tetrameric quaternary structure, as previously reported for SidA and PvdA (Figure 4.14).[178,179]

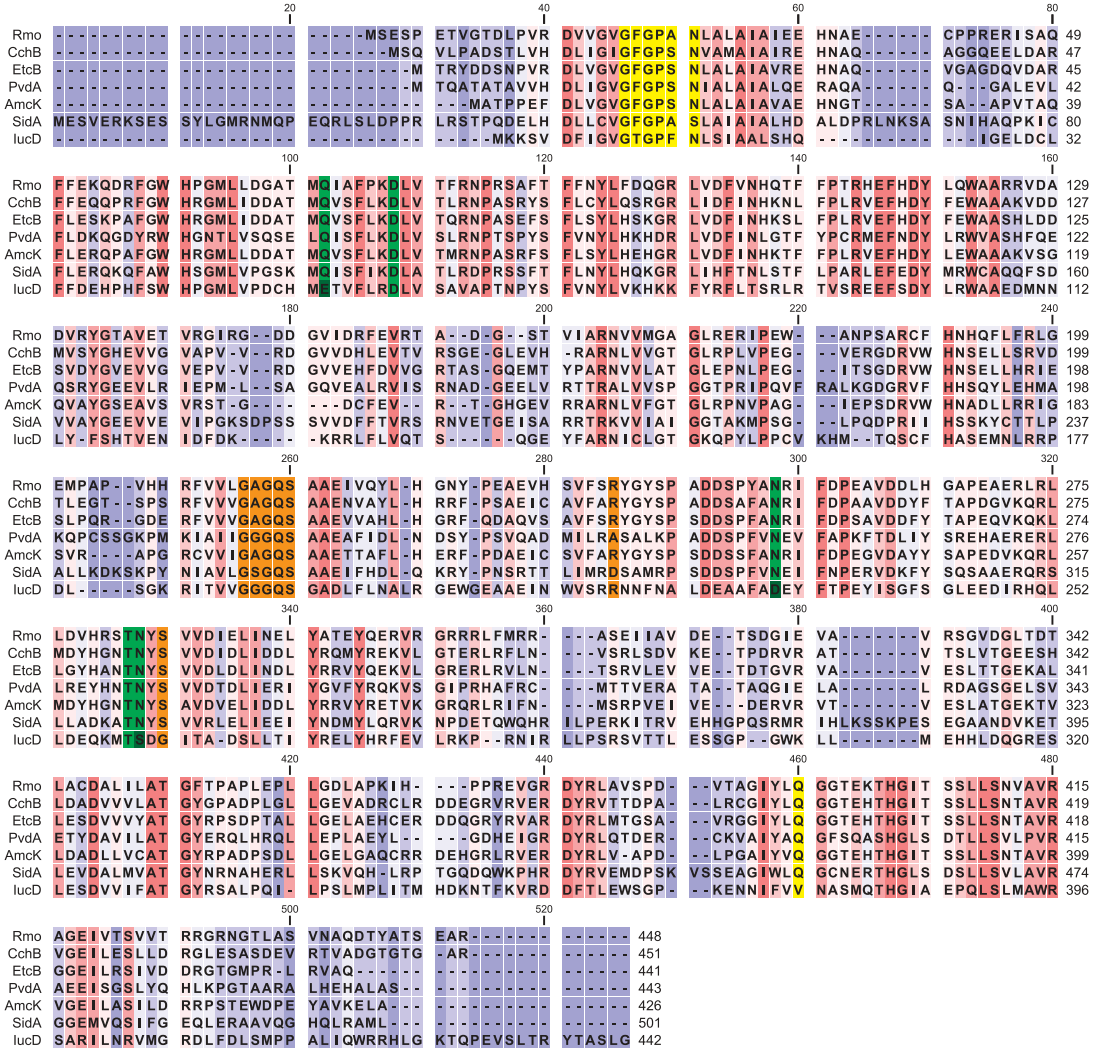


Figure 4.12 Multiple sequence alignment of different NMO homologous to Rmo (*R. jostii*) carried out by using the Clustal Omega algorithm.[161] The alignment includes: CchB (*S. coelicolor*, 48.2% sequence identity/61.6% sequence homology),[88] EtcB (*S. erythraea*, 50.9%/64.4%),[94] PvdA (*Pseudomonas aeruginosa*, 35.4%/51.3%),[178] AmcK (*Amycolatopsis* sp. AA4, 50.7%/63.2%),[183] SidA (*A. fumigatus*, 34.0%/47.8%)[179] and IucD (*E. coli* EN222, 25.5%/43.1%).[181] All proteins display substrate preference towards L-Orn, except for IucD (L-Lys). The degree of conservation is indicated by color: red indicates high and blue no agreement. According to Olucha *et al.*,[15] the residues involved in FAD cofactor binding are highlighted in yellow, the NADPH cosubstrate in orange and the amino acid substrate in green (dark green for the L-Lys coordinating residues in IucD).

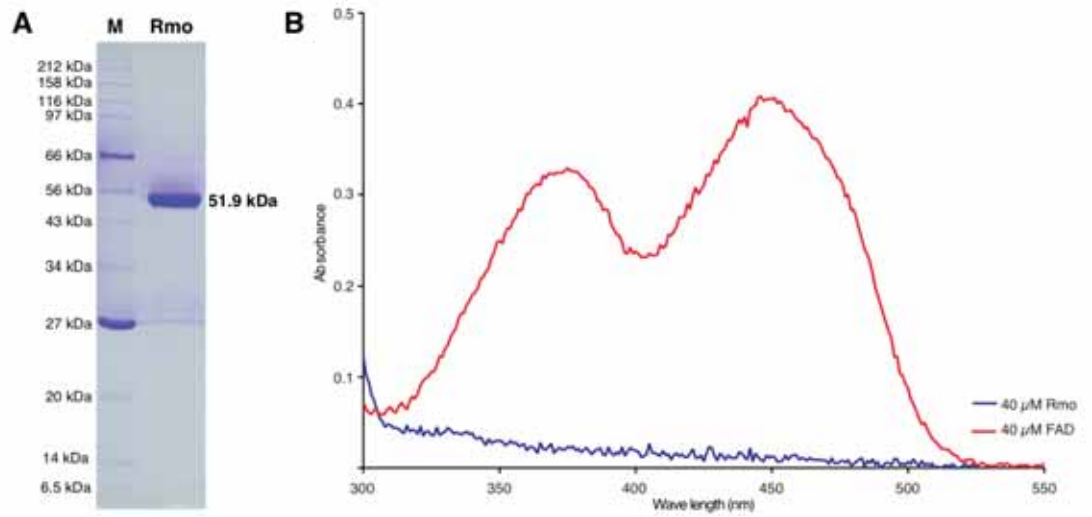


Figure 4.13 (A) Comassie blue staining of the SDS-PAGE of purified recombinant Rmo. (B) UV-vis spectra comparison between equimolar solutions of Rmo and free FAD. Typical absorption maxima for FAD are 370 and 450 nm.

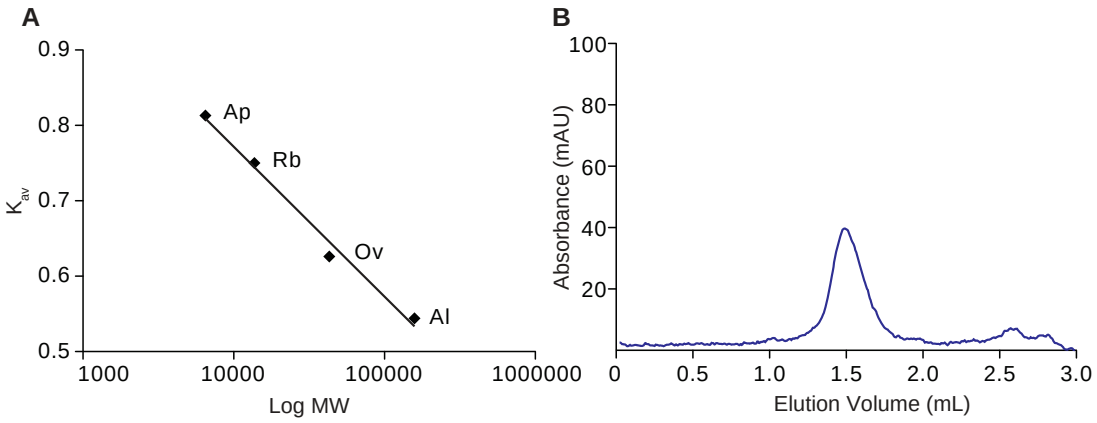


Figure 4.14 Analytical size exclusion chromatography of recombinant Rmo. (A) Column calibration curve: the gel-phase distribution coefficient K_{av} was calculated and plotted as a function of the logarithm of protein molecular mass. Aldolase (Al, 158 kDa, elution volume 1.62 mL), ovalbumin (Ov, 43 kDa, elution volume 1.83 mL), ribonuclease (Rb, 13.7 kDa, elution volume 2.22 mL) and aprotinin (Ap, 6.5 kDa, elution volume 2.43 mL) were used as standards. (B) Size-exclusion chromatogram of 50 μM Rmo displays an elution volume 1.49 mL, indicative of protein with an apparent molecular mass of 220 kDa. This result suggest Rmo to adopt a tetrameric quaternary structure, similar to the homologues SidA and PvdA.[178,179]

4.4.1.3 Biochemical characterization of Rmo

To investigate if Rmo is able to catalyze the conversion of L-Orn to L-hOrn, substrate and enzyme were incubated in the presence of the NADPH cosubstrate and the FAD cofactor. After a 4 h incubation and in the presence of the enzyme and molecular oxygen, HPLC-MS analysis revealed 65% conversion of L-Orn (t_R 12.1 min, *m/z* 133.1 $[M+H]^+$ observed, *m/z* 133.1 $[M+H]^+$ calculated) to L-hOrn (t_R 19.4 min, *m/z* = 149.1 $[M+H]^+$ observed, *m/z* 149.1 $[M+H]^+$ calculated, Figure 4.15). The identity of the enzymatic product of the reaction was verified by comparing its retention time and MS-spectra with the synthetic L-hOrn standard (data not shown) and, furthermore via HR-MS analysis (*m/z* 149.0916 $[M+H]^+$ observed, 149.0921 $[M+H]^+$ calculated). Besides, if in the reaction mixture Rmo or NADPH were omitted, the turnover of the substrate into the product was not observed.

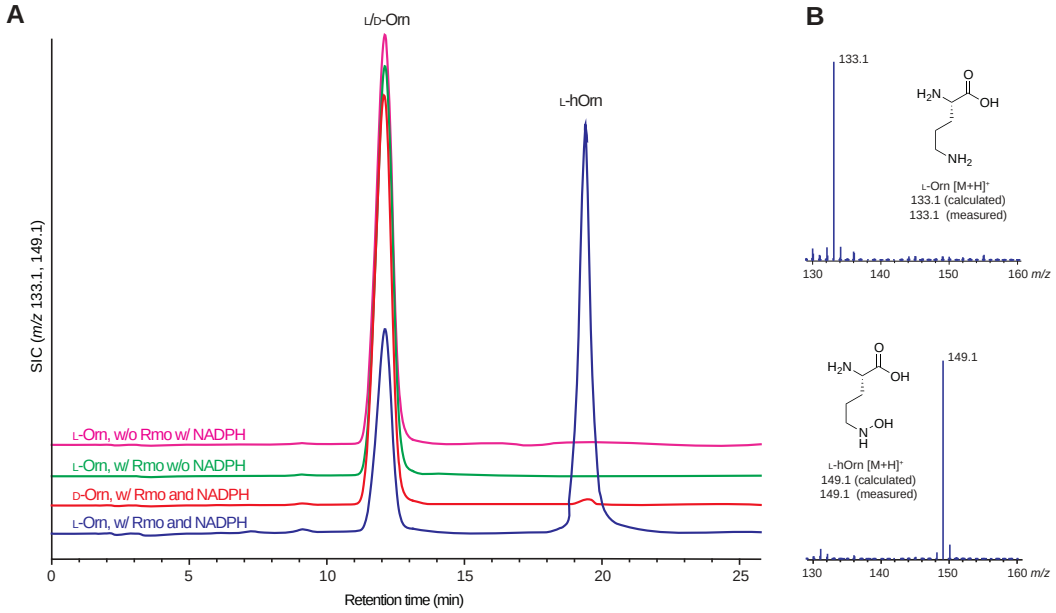


Figure 4.15 Rmo-mediated L-Orn hydroxylation. (A) HPLC-MS single-ion chromatogram (SIC) of the hydroxylation assays is shown: in the presence of the L-Orn substrate and the reducing cosubstrate NADPH, Rmo catalyzes the conversion of L-Orn to L-hOrn (blue trace). The control reactions evidence that Rmo is unable to hydroxylate D-Orn (red trace) and that the reaction does not proceed if either NADPH or the enzyme is missing (green and purple traces, respectively). (B) Chemical structures and observed ESI-MS spectra of the L-Orn substrate and the L-hOrn product.

To evaluate the substrate specificity of Rmo, the enzyme was incubated with a set of different amino acids and analyzed as described. Rmo displays an exclusive preference towards L-Orn as substrate and NADPH as reducing cosubstrate (Table 4.2).

Table 4.2 Overview of the evaluated substrate specificities for the Rmo-mediated L-Orn hydroxylation.

substrate	*m/z* $[M+H]^+$ substrate	*m/z* $[M+H]^+$ expected hydroxylation	*m/z* $[M+H]^+$ observed	hydroxylation
L-Orn	133.1	149.1	149.1	✔
D-Orn	133.1	149.1	133.1	✖
L-fOrn	161.1	177.1	161.1	✖
L-Lys	147.1	163.1	147.1	✖
L-Glu	148.1	164.1	148.1	✖
L-Gln	147.1	163.1	147.1	✖
L-Val	118.1	134.1	118.1	✖
L-Arg	175.1	191.1	175.1	✖
L-Orn (NADH)	133.1	149.1	133.1	✖

Kinetic parameters for Rmo-mediated L-Orn hydroxylation were determined monitoring the product formation and plotting the starting velocity of the reaction as a function of the increasing substrate concentration. The kinetic parameters were determined using a Michaelis-Menten equation plot and were calculated to an apparent K_M = 1.6 ± 0.2 mM and k_{cat} = 0.2331 ± 0.008 s^{-1}, resulting in a catalytic efficiency of k_{cat}/K_M = 0.15 s^{-1} mM^{-1} (Figure 4.16).

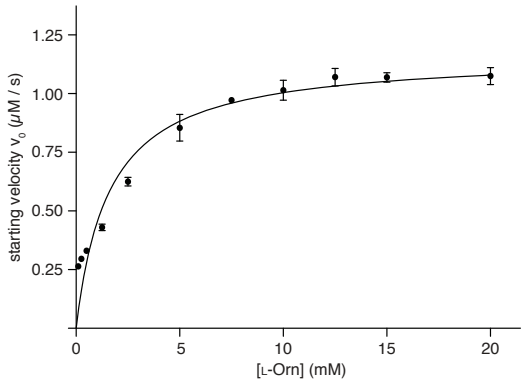


Figure 4.16 Michaelis-Menten kinetics for Rmo. NADPH concentration was fixed at 500 µM, FAD at 20 µM and Rmo at 5 µM. L-Orn substrate concentration was varied between 0.10 and 20 mM. Error bars represent standard deviations from three independent measurements.

In conclusion, Rmo represents a typical member of the NADPH/FAD-dependent monooxygenases required for the δ-*N*-hydroxylation of L-Orn or L-Lys side chains associated with the biosynthesis of hydroxamate-type siderophores.[15]

4.4.2 Biochemical characterization of the Rft L-hOrn formylatransferase

4.4.2.1 Bioinformatic analysis of Rft and other homologous formyltransferases

Similar to *rmo*, *rft* is essential for rhodochelin biosynthesis. A bioinformatic analysis showed that Rft exhibits an overall sequence conservation compared to the proposed formyltransferases CchA and AmcP, putatively involved in the generation of the formyl-derived iron-coordinating hydroxamate moieties in coelichelin and amychelin, respectively.[13,183] Additionally, the sequence homology also extendes to the N-terminal domain of ArnA (a bifunctional enzyme required for the generation of a lipid A analogue essential for polymixine resistance in *Escherichia* and *Salmonella* spp.) and to the endogenous and essential bacterial methionyl-tRNA^{fMet}-formyltransferase.[184,185] In a derived phylogenetic tree, these sequences clearly clusters into different clades, according to their different substrate specificities (Figure 4.17).

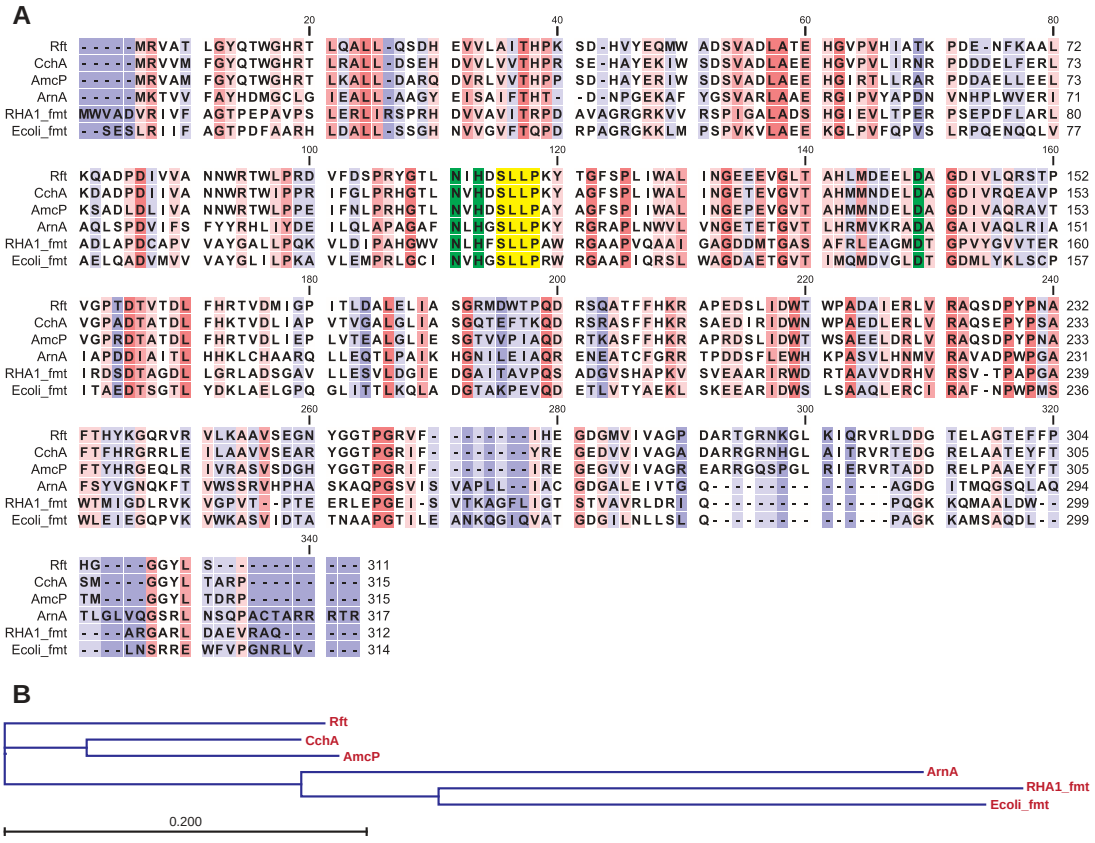


Figure 4.17 (A) Multiple sequence alignment of Rft's (*R. jostii* RHA1) closest homologues. The alignment includes: the L-hOrn formyltransferases CchA (*S. coelicolor*, 66.1% sequence identity/81.3% sequence homology)[88] and AmcP (*Amycolatopsis* sp. AA4, 62.2%/80.0%),[183] the N-terminal domain (res. 1-317) of ArnA (*E. coli*, 30.3%/45.0%)[184] and the endogenous methionyl-tRNA^{fMet}-formyltransferases (Fmt) of *R. jostii* (RHA1_fmt, 24.7%39.4%) and *E. coli* (Ecoli_fmt, 30.1%/48.1%).[124,185] The degree of conservation is indicated by color: red indicates high and blue no agreement. Residues involved in the *N*^{10}-fH$_4$F cosubstrate binding

are highlighted in yellow (SLLP motif), while the catalytic triad (Asn, His, Asp) is shown in green. (B) Phylogenetic tree derived from the alignment: the sequences cluster into different clades, according to their different substrate specificities: methionyl-tRNA^{fMet} for Fmt, UDP-4-amino-4-deoxy-L-arabinose (UDP-L-Ara4N) for ArnA and L-fhOrn for Rft, CchA and AmcP.

Additionally, Rft shares a bimodular organization with the above-mentioned enzymes (Figure S8).[186] The N-terminal subdomain displays typical elements for tetrahydrofolate binding enzymes: the catalytic Asn, His, Asp triad, and the N^{10}-fH$_4$F "SLLP" binding motif.

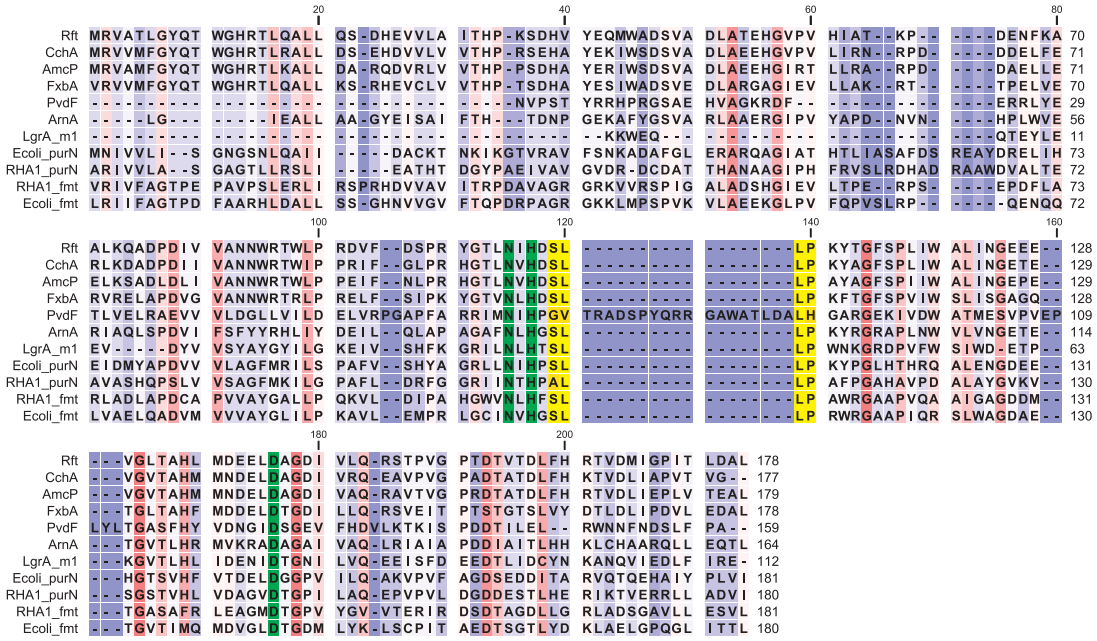


Figure 4.18 Multiple sequence alignment of the N-terminal catalytic formyltransferase domain carried out by using the MUSCLE algorithm.[162] Identity between Rft and the others sequences is indicated as follows: 68.2% CchA,[88] 64.8% AmcP,[183] 57.9% FxbA (exochelin L-hOrn formyltransferase *M. smegmatis*),[188] 12.8% PvdF (pyoverdine L-hOrn formyltransferase *P. aeruginosa*),[189] 31.3% ArnA,[184] 16.9% LgrA_m1 (formylation domain of the initiation module of linear gramicidin, *B. brevis*),[80] 24.6% Ecoli_purN (glycinamide ribonucleotide formyltransferase, *E. coli*),[190] 21.9% RHA1_purN (glycinamide ribonucleotide formyltransferase, *R. jostii* RHA1), 27.1% RHA1_fmt and 29.8% Ecoli_fmt.[124,185] Residues involved in the N^{10}-fH$_4$F cosubstrate binding are highlighted in yellow (SLLP motif), while the catalytic triad (Asn, His, Asp) is shown in green. The degree of conservation is indicated by color: red indicates high and blue no agreement.

These conserved residues are also found in the formyltransferases FxbA and PvdF (associated with the biosynthesis of L-fhOrn in exochelin and pyoverdine systems), the N-terminal formylation domain of the initiation module of the linear gramicidine NRPS LgrA, and the glycinamide ribonucleotide formyltransferases (GARF) family proteins, although in this latter case, additional structural differences and less sequence homology have already been reported (Figure 4.18).[80,185,187-189]

On the other hand, the overall sequence homology between Rft and its above-discussed closest homologues decreases through the C-terminal subdomain, which seems not to be involved in catalysis and, as demonstrated for the methionyl-tRNA^{fMet}-formyltransferase, could be associated with proper substrate recognition.[187]

4.4.2.2 Biochemical characterization of Rft and CchA L-hOrn formyltransferases

In order to investigate the role of Rft in L-fhOrn biosynthesis, the corresponding gene was amplified and cloned in the pET28a(+) expression vector. The recombinant protein was purified as an N-terminal His-tag fusion and tested for *in vitro* activity (Figure 4.19).

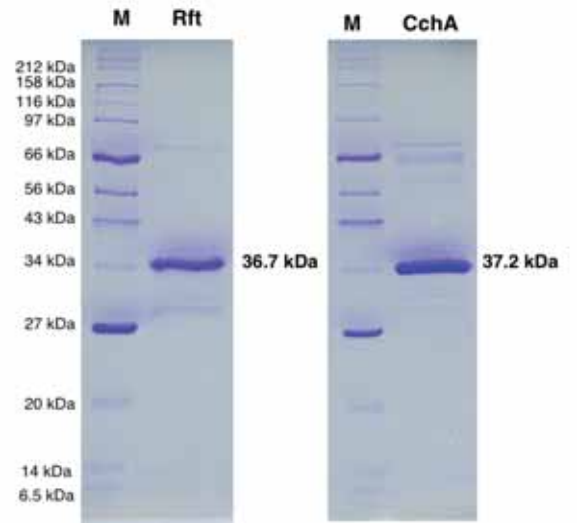


Figure 4.19 Comassie blue staining of the SDS-PAGE of purified recombinant Rft and CchA.

The required N^{10}-fH$_4$F cofactor was generated *in situ* from its N^5,N^{10}-methenylH$_4$F intermediate, through a 30 min preincubation in the assay buffer (Figure 4.20).[151,191] In the presence of L-hOrn and N^{10}-fH$_4$F, Rft was able to transfer the formyl group from the donor cosubstrate to the side chain of the hydroxylated amino acid. After a 4 h incubation, HPLC-MS analysis revealed 55% conversion of L-hOrn (t_R 18.9 min, *m/z* 149.1 $[M+H]^+$ observed, *m/z* 149.1 $[M+H]^+$ calculated) to L-fhOrn (t_R 12.9 min, *m/z* = 177.1 $[M+H]^+$ observed, *m/z* 177.1 $[M+H]^+$ calculated; Figure 4.21). L-fhOrn identity was confirmed by HR-MS analysis (*m/z* 177.0871 $[M+H]^+$ observed, 177.0870 $[M+H]^+$ calculated). No conversion was observed in the absence of enzyme or cosubstrate. In addition, in similar assay conditions, Rft is unable to formylate L-Orn or both L-Orn and L-hOrn if N^5-fH$_4$F was used as the donor cofactor (data not shown).

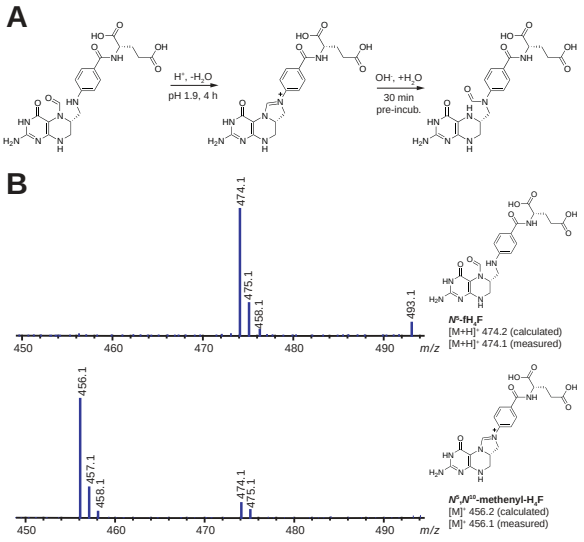


Figure 4.20 (A) Reaction scheme for the *in situ* generation of the N^{10}-fH$_4$F cofactor. (B) The chemical structures and the observed MS-spectra of the ESI-MS measurements of the N^5-fH$_4$F and the N^5,N^{10}-methenylH$_4$F reaction intermediate were carried out to evaluate the complete conversion of the substrate after 4h of incubation.

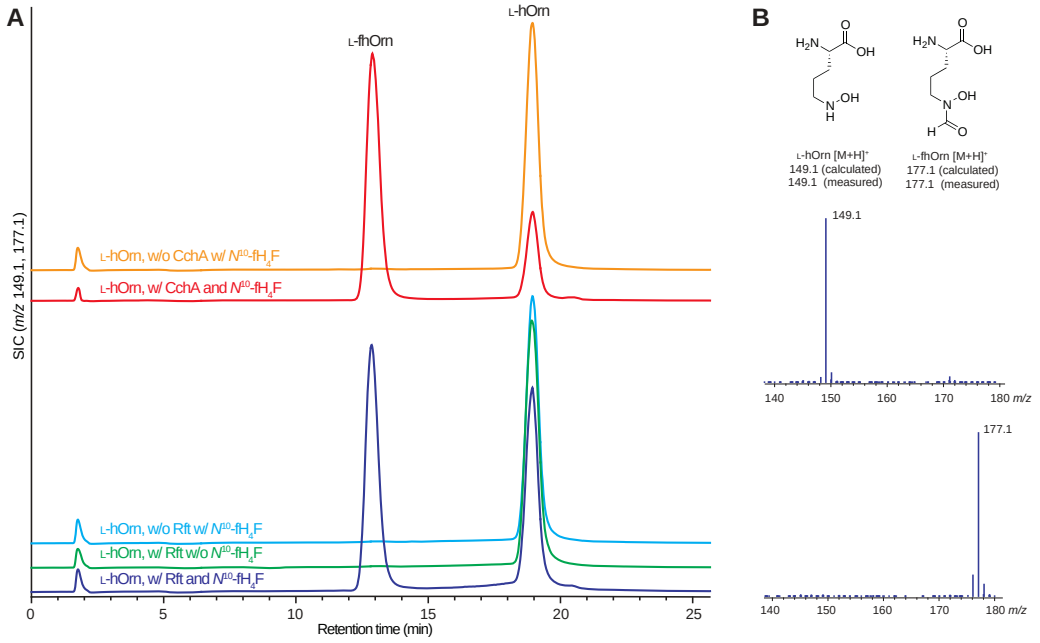


Figure 4.21 Rft- and CchA-mediated L-hOrn formylation. (A) Lower traces: HPLC-MS single-ion chromatogram (SIC) of the formyltransferase reaction assay is shown: in the presence of the L-hOrn substrate and the *in situ* generated N^{10}-fH$_4$F formyl-donor cosubstrate, Rft catalyzes the conversion of L-hOrn to L-fhOrn (blue trace). Control reactions were carried out in the absence of the donor cosubstrate or the enzyme (green and light blue traces, respectively). Upper traces: a similar assay performed in the presence or the absence of CchA (red and orange traces, respectively) rescues the enzymatic activity of this Rft homologue from the coelichelin biosynthesis gene cluster, which was previously reported to be inactive.[13,88] (B) Chemical structures and observed ESI-MS spectra of the L-hOrn substrate and the L-fhOrn product.

Previous attempts to characterize Rft using the N^{10}-fH$_4$F cosubstrate generated via a chemoenzymatic synthesis approach (chemical conversion of H$_4$F to N^5,N^{10}-methyleneH$_4$F in the presence of formaldehyde, followed by a FolD-catalyzed regiospecific oxidation/cyclohydrolyzation) always resulted in no detectable enzymatic activity (data not shown).[192] Similar results were obtained during the characterization of CchA, which left unclear whether the enzyme was purified in a soluble but inactive form or was rendered inactive by the incubation in the reaction assay.[88] Therefore, on the basis of the obtained results for Rft, CchA was expressed, purified (Figure 4.19), and assayed for enzymatic activity. As shown in Figure 4.21, CchA was able to catalyze the conversion of L-hOrn to L-fhOrn (75% after 4 h incubation).

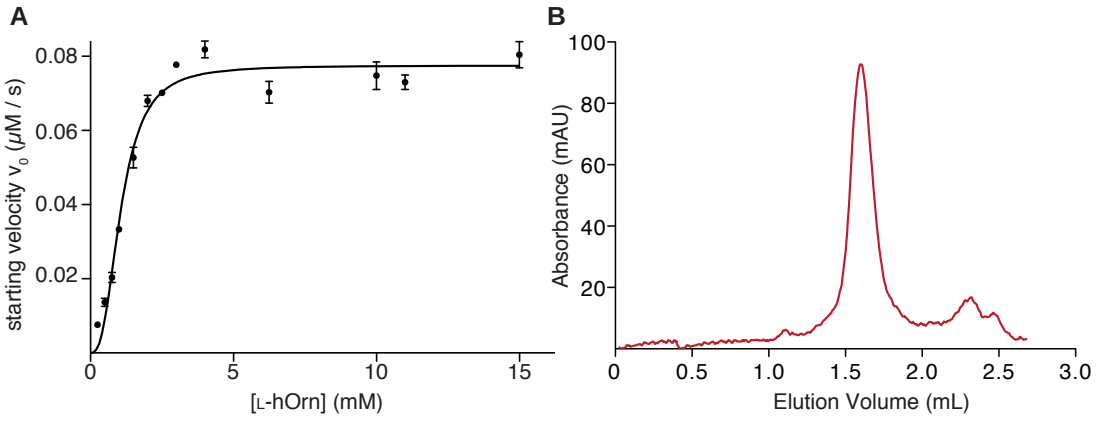


Figure 4.22 (A) Sigmoidal plot of Rft kinetic characterization. N^5,N^{10}-methenylH$_4$F concentration was fixed at 1.5 mM, while L-hOrn substrate concentration was varied between 0.25 and 15 mM. Error bars represent standard deviations from three independent measurements. (B) In order to elucidate the origin of this unusual cooperative behavior, analytical size exclusion chromatography of 50 μM Rft solution was carried out. Rft displays an elution volume 1.60 mL, indicative of protein with an apparent molecular mass of 146 kDa, suggesting the enzyme to adopt a tetrameric quaternary structure, likewise Rmo. The calibration curve for the analytical gel filtration is shared with the analytical gel filtration of the Rmo monooxygenase, which is illustrated in Figure 4.14.

The kinetic parameters of Rft tranformylation were determined monitoring the conversion of substrate to product and plotting the starting velocities as a function of the increasing substrate concentration. Surprisingly, the experimental data could be fitted best using a sigmoidal curve, indicative of an allosteric kinetic mechanism, rather than the typical rectangular hyperbola characteristic of classical Michaelis-Menten kinetics (Figure 4.22 A).[193] Normalized parameter values were calculated through the nonlinear Hill-fit regression and were found to be equal to a V_{max} = 0.078 ± 0.001 μmol L^{-1} s^{-1}, a $K_{0.5}$ = 1.2 ± 0.1 mM, and n = 2.7 ± 0.2. A value of the Hill parameter n greater than 1 is indicative of a positive cooperative mechanism. To

further clarify the origin of this cooperative behavior, analytical size exclusion chromatography was employed to elucidate the oligomeric state of Rft.[194] As shown in Figure 4.21 B, the estimated molecular weight of Rft was 146 kDa, suggesting the enzyme to adopt a tetrameric quaternary structure in solution.

4.4.3 L-fhOrn coupled enzymatic biosynthesis

To verify whether Rmo and Rft were able to act in tandem to generate L-fhOrn from the L-Orn substrate, a similar assay to the Rft-dependent L-hOrn transformylation was set up, where first the N^{10}-fH_4F cosubstrate was generated *in situ*, followed by the addition of all the remaining components needed for the enzymatic tandem conversion to L-fhOrn. Comparison of HPLC-MS traces showed the substrate conversion to the L-hOrn intermediate and the L-fhOrn product if both enzymes were present or, as expected, only to L-hOrn if Rft was omitted. On the other hand, if Rmo was missing, no substrate conversion was observed (Figure 4.23).

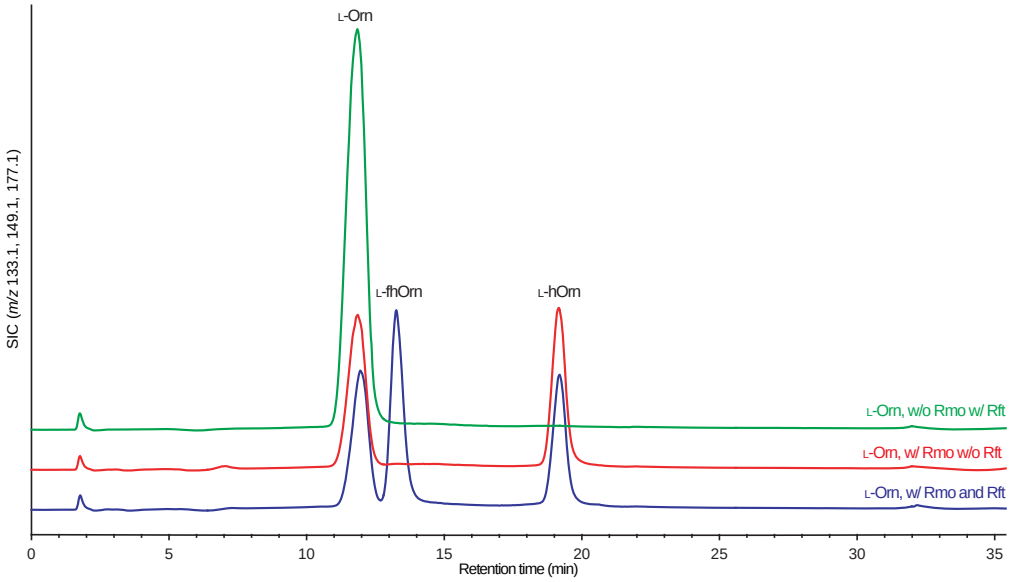


Figure 4.23 HPLC-MS single-ion chromatogram (SIC) of the coupled enzymatic biosynthesis of L-fhOrn from its L-Orn precursor in the presence of the Rmo monooxygenase and the Rft formyltransferase (blue trace). If Rft is omitted, the reaction stops generating only the L-hOrn intermediate (red trace); if Rmo is excluded, no conversion at all is observed (green trace).

Taking all these results of the biochemical characterization of the L-Orn tailoring enzymes Rmo and Rft together, it can be stated that both enzymes constitute an enzymatic pathway for the biosynthesis of the L-fhOrn building block required for the NRPS-dependent rhodochelin assembly.

Chapter 5

Discussion

5.1 Isolation and structural characterization of the siderophore rhodochelin

5.1.1 *Rhodococcus* spp. as a new source for secondary metabolites

In recent years, the increasing amount of sequenced microbial genomes has revealed the presence of an impressive number of secondary metabolite gene clusters, most of them considered "orphan" with respect to their natural product.[44] Furthermore, since the discovery of the capability of single strains to produce many natural products, the concept "One Strain - MAny Compounds" (OSMAC) has been introduced; thus, the interest to uncover new secondary metabolites in known species is increasing.[108] Despite the use of genome mining and its successful application, the identification of new natural products still remains challenging. In fact, without any experimental proof, it is difficult to define if an orphan cluster is silent because it is not functional, the metabolite cannot be detected due to analytical detection limits, or the laboratory cultivation methods are inappropriate for its biosynthesis.[43]

The continuous interest in the *Rhodococcus* genus as a bioremediation and bioconversion tool has shifted the focus to regard these species as natural product producing strains.[123] As streptomycetes, rhodococci belong to the actinomycetal order and thus are predisposed to a putative and underestimated secondary metabolism. In addition, they also offer experimental advantages that could ease the comprehension of the physiology of closely related species like *Mycobacterium* and *Streptomyces*, that are difficult to cultivate. Therefore, the complete genome sequence of *R. jostii* RHA1 provides an excellent opportunity that facilitate both the exploitation of the catabolic versatility of these industrially-important microorganisms and natural product discovery.[124] So far, the only secondary metabolites isolated from *Rhodococcus* spp. were the siderophores heterobactin A, rhodobactin and the ribosomally-derived lasso peptide lariatin.[125,126,195] In addition, with the exception of lariatin, the two siderophores still remain orphan with respect to their biosynthetic gene clusters.[196]

Genome analysis of the *R. jostii* RHA1 chromosome identified 23 secondary metabolite gene clusters, suggesting the possibility of an extensive secondary metabolism (Table 5.1).[124,157] Most of them encode for peptides of non-ribosomal origin, though PKS and hybrid PKS-NRPS systems are present as well. Moreover, six of the NRPS genes extent more than 25 kbp in length.[168]

Table 5.1 Secondary metabolite gene clusters in *R. jostii* RHA1. The list updates the prediction described by Nett *et al.* using the ANTIsmash bioinformatic tool.[157,168]

no.	cluster location	predicted product
1	*RHA1_ro00071-ro00073*	NRP
2	*RHA1_ro00136-ro00148*	NRP
3	*RHA1_ro00232-ro00235*	NRP
4	*RHA1_ro00429-ro00435*	NRP
5	*RHA1_ro01106-ro01115*	Terpene
6	*RHA1_ro01305-ro01307*	Ectoine
7	*RHA1_ro02207-ro02310*	PK-NRP hybrid
8	*RHA1_ro02318-ro02323*	rhodochelin
9	*RHA1_ro02391-ro02397*	NRP
10	*RHA1_ro02492-ro02494*	NRP
11	*RHA1_ro04063-ro04066*	Polyketide
12	*RHA1_ro04230-ro04231*	Polyketide
13	*RHA1_ro04382*	ε-poly-L-lysine[197]
14	*RHA1_ro04612-ro04614*	Butyrolactone
15	*RHA1_ro04713-ro04716*	NRP
16	*RHA1_ro05093-ro05103*	NRP
17	*RHA1_ro05430-ro05431*	NRP
18	*RHA1_ro05452-ro05468*	NRP
19	*RHA1_ro06098-ro06103*	NRP
20	*RHA1_ro06273-ro06307*	PK-NRP hybrid / n.a.
21	*RHA1_ro06663-ro06665*	NRP
22	*RHA1_ro07030*	NRP
23	*RHA1_ro07202-ro07203*	Terpene

5.1.2 Isolation and structural characterization of the siderophore rhodochelin

In this work, the isolation, structural characterization, and genetic and biochemical analysis of the biosynthetic origin of the siderophore rhodochelin is reported. The siderophore is also known to be the first secondary metabolite isolated from the producing strain, *R. jostii* RHA1. Rhodochelin production was triggered by growing the strain in minimal medium, under iron-limited conditions. This resulted in the isolation of a CAS-reactive compound that was further purified by HPLC. NMR and MS^n studies revealed the branched tetrapeptidic structure of the molecule, which is composed of a linear assembly of 2,3-DHB, L-Thr, and L-fhOrn. The fourth building block (an additional L-fhOrn moiety) is attached to the main tripeptide scaffold through an unusual and characteristic ester bond via the side chain hydroxyl group of the L-Thr residue. The comparison of the structure of rhodochelin with heterobactin A and rhodobactin highlights that the presence of 2,3-DHB and

modified ornithine residues is a shared iron-coordinating strategy among the different rhodoccocal siderophores isolated so far (Figure 5.1).[125,126]

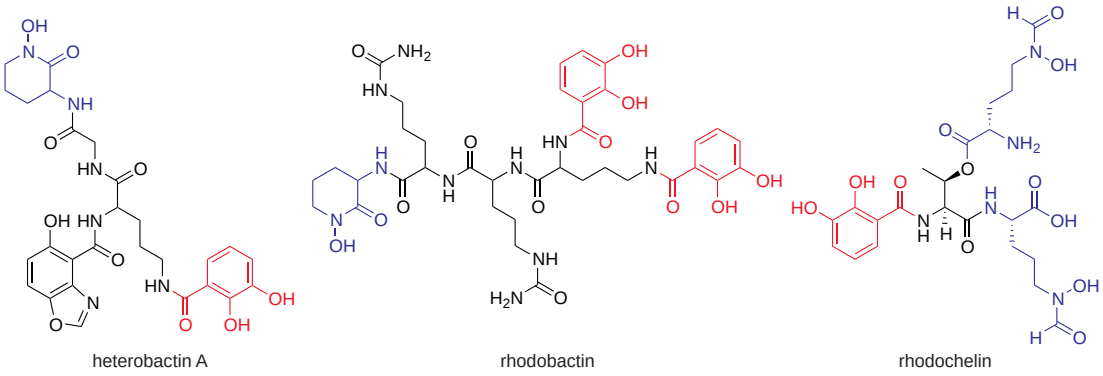


Figure 5.1 Chemical structures of heterobactin A, rhodobactin and rhodochelin with highlighted iron-coordinating functionalities: 2,3-DHB in red and tailored ornithines in blue. Ornithines moieties of heterobactin A and rhodobactin that do not take part in the iron-coordination are not highlighted.

Despite the fact that the complete elucidation of the mode and the affinity of the ferric-chelation remains to be elucidated, an iron-binding model could be proposed (Figure 5.2). Based on the evidence that characteristic UV-vis ligand-to-metal charge transfer absorption spectra appear upon the binding of the ferric ion (Figure 4.6), both the catecholate and the hydroxamate functionalities take part in the coordination.

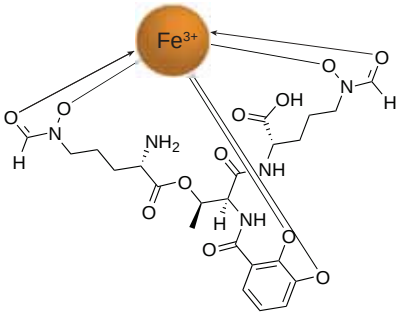


Figure 5.2 Model of rhodochelin iron-coordination.

During rhodochelin purification, a second CAS-reactive compound was isolated. MS-analysis revealed it to be a rhodochelin degradation product; more precisely it was identified as the tripeptide 2,3-DHB-L-Thr-L-fhOrn resulting from the hydrolysis of the characteristic ester bond of rhodochelin. This compound was thus renamed rhodochelin tripeptide and still retains the ability to complex ferric iron, as confirmed via UV-vis spectroscopy and HPLC-MS analysis (Figure S5).

In principle, after secretion out of the extracellular space, the ferric iron-siderophore is selectively and actively imported by specific import systems, different between

Gram-positive and Gram-negative bacteria.[169,198-201] Thus, as resulting from *in vitro* studies carried out with the *holo*-rhodochelin and the *holo*-rhodochelin tripeptide, the cognate siderophore binding protein RhcC solely recognizes the ferric-rhodochelin complex, whereas the ferric-rhodochelin tripeptide does not display any affinity for the protein.[202] In addition, if purified rhodochelin was incubated at room temperature in M9 minimal medium, for a period of time corresponding to the growing conditions used for its isolation (~48 h), the spontaneous hydrolysis of the ester bond occurs. These results suggest that the hydrolysis of the ester bond takes place spontaneously, probably due to the pH of the culture's medium (Figure 5.3).

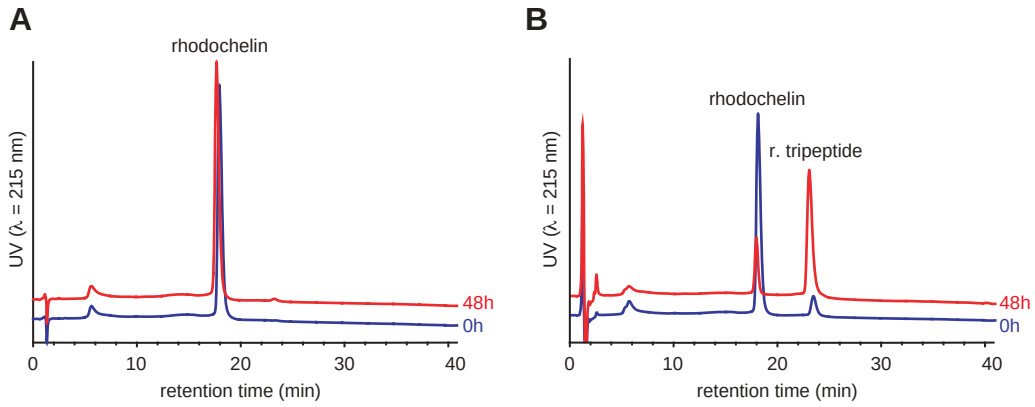


Figure 5.3 Stability assays of rhodochelin solutions. (A) HPLC analysis rhodochelin solution in ddH_2O, incubated at room temperature for 48 h. (B) HPLC analysis of a rhodochelin solution in M9 medium incubated at room temperature for 48 h, revealing the spontaneous hydrolyzation of rhodochelin's ester bond occurs.

Similarly, the isolation of siderophore analogues that differ from the main molecule by the hydrolysis of a specific bond was already reported for fuscachelin and heterobactin,[125,203] supporting the hypothesis that spontaneous hydrolyzation can occur under laboratory growing conditions. Unfortunately, no additional investigations were carried out or discussed. It cannot be excluded that the *ferric*-rhodochelin tripeptide retains a biological function and is imported into the intracellular space by an alternative system.

5.2 The biosynthesis of rhodochelin requires NRPS cross-talk

5.2.1 Identification of the gene set associated with rhodochelin biosynthesis

The complete set of gene clusters for the biosynthesis of rhodochelin was identified by a genome mining approach. The *rhc* cluster (cluster A) contains the complete bimodular NRPS synthetase *rhc*B, the genes involved in rhodochelin export and import (*rhc*C, *rhc*D, and *rhc*F) and two additional ORFs (*rhc*A and *rhc*E), homologues to the two distinct domains of DhbB, ICL, and ArCP, respectively (Figure 4.7).[204] Interestingly, the first three genes of this cluster (*rhc*A, *rhc*B, and *rhc*C) have previously been annotated to belong to an orphan siderophore gene cluster.[168]

The presence of 2,3-DHB within the rhodochelin structure led to the investigation of the *R. jostii* RHA1 genome to identify all the genes involved in the biosynthesis of the aryl moiety.[176] One of the enzymes, an isochorismatase, has already been found to be encoded in the *rhc* cluster (*rhc*A). Since the biosynthesis of the aryl-capped siderophore bacillibactin requires the activation of 2,3-DHB by the stand-alone A-domain DhbE prior to NRPS-catalyzed assembly, a gene homologous to *dhb*E (cluster C) was identified in a different genomic region, along with two other genes involved in 2,3-DHB biosynthesis, namely *dhb*C and *dhb*A.[35,175] These three genes are arranged in an operon-like way and, together with *rhc*A, cover the entire 2,3-DHB pathway, from the chorismate precursor to its activation as adenylate.

Table 5.2 Substrate specificity prediction for the adenylation domains of the putative NRPS RHA1_ro04715. An overview of the modular organization of the synthetase is presented in Figure 4.8. The first adenylation domain possesses the same specificity-conferring code found in the first A-domain of CchH, involved in coelichelin biosynthesis,[13,88] suggesting that also RHA1_ro04715 activates L-fhOrn instead of the predicted L-Orn. Variations in the consensus are highlighted in red.

A-domain	active site residues	substrate	product
4715-A_1	DINYWGGIGK		
Coe-A_1	DINYWGGIGK	L-Orn	coelichelin
4715-A_2	DVLHSSLVDK		
ArfB-A_4	DVWHMSLVDK	L-Ser	arthrofactin
4715-A_3	DMENLGLINK		
Ery-A_3	DMENLGLINK	L-hOrn	erythrochelin
4715-A_4	DVFILGAVNK		
Ery-A_4	DVFALGAVNK	L-haOrn	erythrochelin

It is important to note that an additional putative siderophore gene cluster has been identified in the *R. jostii* RHA1 genome (cluster B, Figure 4.7, no. 15 in Table 5.1). Because of the overall homology of the tetramodular NRPS RHA1_ro04715 with

EtcD[14,169] (supported by the comparison of the adenylation domain substrate specificity prediction, and the contextual presence of two tailoring enzymes Rmo and Rft), it is suggested that this cluster could be responsible for the synthesis of a foroxymithine / erythrochelin analogue (Figure 5.4).[170] Bioinformatic analysis of the NRPS RHA1_ro04715 revealed that the first adenylation domain possesses the same specificity-conferring code found in the first A-domain of CchH, involved in coelichelin biosynthesis (Table 5.2),[13,88] suggesting that RHA1_ro04715 also activates L-fhOrn as a cognate substrate, which is subsequently incorporated into the putative natural product.

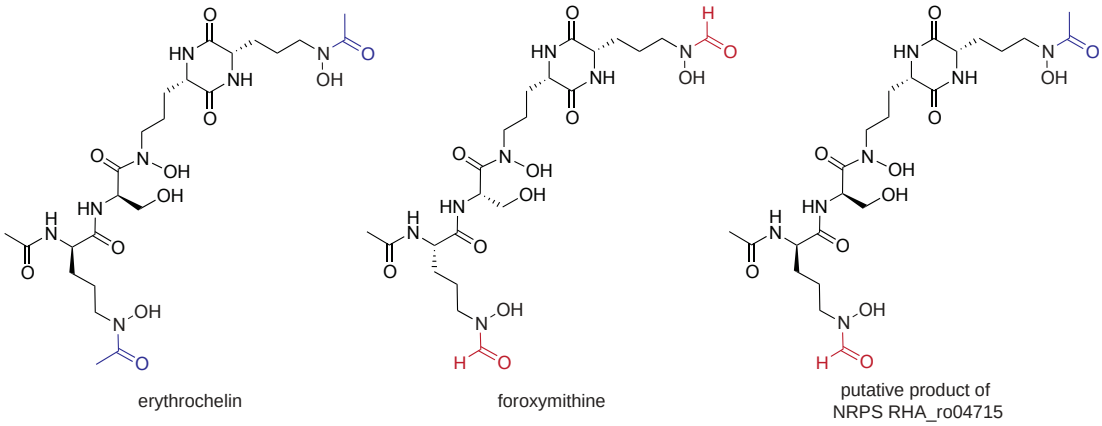


Figure 5.4 Comparison of the chemical structures of erythrochelin[14] (*S. erythraea* NRRL2338) and foroxymithine[170] (*Streptomyces nitrosporeus*) reveals a high degree of structural similarity; the molecules differ in the stereochemistry of the first two amino acids and the tailoring of the hydroxyornithine sidechain of the first and the last amino acid (acetylation is highlighted in blue, formylation in red). On the right, the structure of the putative product of the NRPS RHA1_ro04715 is presented. The structure is based on both the modular organization of the NRPS (Figure 4.8) and the A-domains substrate specificity prediction (Table 5.2), implying a similar assembly strategy like the one proposed for erythrochelin.[14,169]

According to the results of the genome mining, the existence of three distantly located gene clusters involved in rhodochelin assembly was proposed. This hypothesis was verified employing a gene disruption approach, showing that single key gene deletions in different clusters (*rhc*B, *dhb*E, *rmo*, and *rft* were targeted) were sufficient to inhibit siderophore biosynthesis, and that the dispersed genes along the three clusters are equally required for the biosynthesis of rhodochelin.

5.2.2 Genome comparison between sequenced *Rhodococcus* strains

Genome comparison of *R. jostii* RHA1 with other sequenced *Rhodococcus* strains revealed that only the closely-related *Rhodococcus opacus* B4[205,206] shares an identical gene arrangement within all the three gene clusters associated with rhodochelin biosynthesis. The RhcB NRPS synthetase and its *R. opacus* B4 homologue revealed

100% identity for the A-domain specificity-conferring codes, suggesting that also rhodochelin is the endogenous siderophore of the latter strain (Table S3). Interestingly, the additional sequenced strains analyzed (*Rhodococcus equii* 103S[207] and *R. erythropolis* PR4) share the contextual presence of cluster B and cluster C, despite the presence of minor genetic rearrangements (gene integrations and deletions, Table S4).

Table 5.3 Comparison of the adenylation domains substrate specificity predictions for the cryptic NRPS RHA1_ro04715 in the other *Rhodococcus* strains analyzed. Variations in the consensus are highlighted in red.

strain	% id./sim.	A_1 – L-fhOrn	A_1 – L-Ser	A_3 – L-hOrn	A_4 – L-haOrn
R. jostii RHA1		DINYWGGIGK	DVLHSSLVDK	DMENLGLINK	DVFILGAVNK
R. opacus B4	91 / 95	DINYWGGIGK	DVLHSSLVDK	DMENLGLINK	DVFILGAVNK
R. equi 103S	68 / 78	DINYWGGIGK	DVPHASLVEK	DMENLGLINK	DVFILGAVNK
R. erythropolis PR4	65 / 77	DINYWGGIGK	DVPHNSLVDK	DMENLGLINK	DVFILGAVNK

In addition, similar results were also obtained for the cryptic NRPS RHA1_ro04715, which displayed high sequence similarity among all the homologues in the three strains analyzed (Table 5.3). As shown in Figure 5.4, the putative product of the NRPS RHA1_ro04715 only includes hydroxamate iron-coordinating functionalities, while rhodochelin contains both catecholate and hydroxamate groups. Catecholates are known to bind iron more tightly than hydroxamates.[10] This would lead to a siderophore with increased iron-binding affinity, providing an evolutionary advantage to the organism's capability to thrive under iron-limited conditions. A second possible explanation is that the evolution of an alternative siderophore (with mixed iron-coordinating functionalities) could confer an additional selective advantage due to the fact that siderophore uptake systems usually employ receptors capable of recognizing only certain types of ferric-siderophore complexes. This would results in a situation where less competing organisms would be able to hijack the iron-loaded siderophore in question.[208,209]

5.2.3 Genome cluster cross-talk associated with the production of microbial secondary metabolites

Usually the genes required for the biosynthesis of secondary metabolites are clustered together within one chromosomal locus, allowing their coordinated expression and regulation.[210,211] This concept, together with the increasing number of sequenced genomes, accounted for the successful development of different strategies aimed for successful new natural product discovery.[43,212] The hypothesis

that natural product gene clusters are among the most diverse and rapidly-evolving genetic elements of a bacterium's chromosome has been recently proposed, suggesting them to be the ideal candidates for the evolutionary study of bacterial collectives.[213] In this context, the discovery of different examples of cross-talk mechanisms between the secondary metabolite gene clusters can be representative of the bacterium's genome plasticity towards the natural selective pressure, continuously remodeling its elements (and thus its metabolome), fueling the fitness for the environment.

Similarly to rhodochelin, the catecholate siderophore vibriobactin (*Vibrio cholerae*), harbors its biosynthetic genes within two different genome regions. The genes for the siderophore transport and utilization, *viu*A and *viu*B, and the biosynthetic NRPS *vib*F[214-216] are separated from those associated with the biosynthesis and activation of 2,3-DHB and for the periplasmic binding and ABC-dependent intracellular import of the iron-siderophore complex.[217-219] On the other hand, to our knowledge, rhodochelin is the first example where the genes associated with the biosynthesis of the aryl moiety are not clustered together.[175] Another siderophore, erythrochelin, is synthetized by genes dispersed between two diverse gene clusters:[14,169] the *erc*/*etc* cluster encodes the enzymes responsible for precursor biosynthesis and siderophore assembly and export, whereas the δ-*N*-L-acetyltransferase *mcd* (also essential for the biosynthesis of erythromycin and the red pigment flaviolin[220]) is located in the nonfunctional *nrps1* cluster.[221] Moreover, PPTases, the enzymes responsible for the *apo*- to *holo*- conversion of carrier protein domains of FAS, PKS and NRPS, are an additional example of enzymes being shared by different biosynthetic pathway, sometimes displaying a high degree of substrate tolerance towards the carrier protein, like *B. subtilis* Sfp.[154,155]

Recently, a few additional examples of functional secondary metabolism cross-talk mechanisms have been reported, demonstrating that a more complex biosynthetic network exists. For example, either one of two essential malonyl-CoA acyltransferases of the endogenous fatty acid synthase complex are required for the biosynthesis of the anticancer depsipeptide FK288 in *Chromobacterium violaceum* 968, since the two modules DepB and DepC of the PKS lacks the essential AT function which is replaced by an AT-docking domain. Additionally, in the same species, the activation of the FK288 PKS ACP domains requires a broad-substrate-range Sfp-type PPTase which is encoded in a different locus.[222] Similar to the last discussed example, the initiation reaction of the assembly of the quinoxaline antibiotics requires the transfer of the activated chromophore quinoxaline-2-carboxylic acid from the cognate stand-alone A-domain TrsI to the first elongation NRPS TrsII. This

step is carried out by an ACP from the primary FAS II, to form an initiation module, showing a functional interaction between the primary and the secondary metabolism.[49]

Cross-talk between different secondary metabolites gene clusters can also occur on the regulatory level. For example, in *P. fluorescens* Pf-5, phloroglucinol act either as a precursor, an inducer and a repressor of the biosynthesis of the antibiotics pyoluteorin and 2,4-diacetylphloroglucinol. At nanomolar concentration, the presence of the 2,4-diacetylphloroglucinol precursor phloroglucinol is required to stimulate the biosynthesis of pyoluteorin, which is inhibited at higher concentrations.[223] Another similar cross-talk mechanism has also been observed in fungi. In *Aspergillus nidulans*, the overexpression of a putative pathway-specific NRPS-regulatory gene *scp*R induces the activation and, thus, the formation of a cryptic NRP, which further downstream upregulates a second transcription factor, *afo*A. This latter one activates the biosynthesis of the polyketide asperfuranone through its specific biosynthetic pathway. In addition, these results represent the first report of regulatory cross-talk between two biosynthetic gene clusters located on different fungal chromosomes.[224]

5.3 Biosynthesis of the non-proteinogenic amino acid L-fhOrn

The use of hydroxamate groups as iron-coordinating moieties is a shared strategy employed by both NRPS-dependent and NRPS-independent pathways for siderophores biosynthesis.[20] In both cases, it generally requires the hydroxylation of the lateral amino group of a basic amino acid (lysine or ornithine), followed by the additional transfer of an acetyl or formyl group to the secondary amine intermediate, generating the functional hydroxamate moiety. This building block is subsequently incorporated by the NRPS-dependent or NRPS-independent assembly enzymes into the siderophore peptide scaffold.[15,26]

5.3.1 Characterization of the L-Orn monooxygenase Rmo

On the basis of a bioinformatic analysis, Rmo was predicted to belong to the *N*-hydroxylating flavoprotein monooxygenases (NMO), of whom different homologues have previously been characterized through extensive biochemical and structural studies. [88,94,178-181,225]

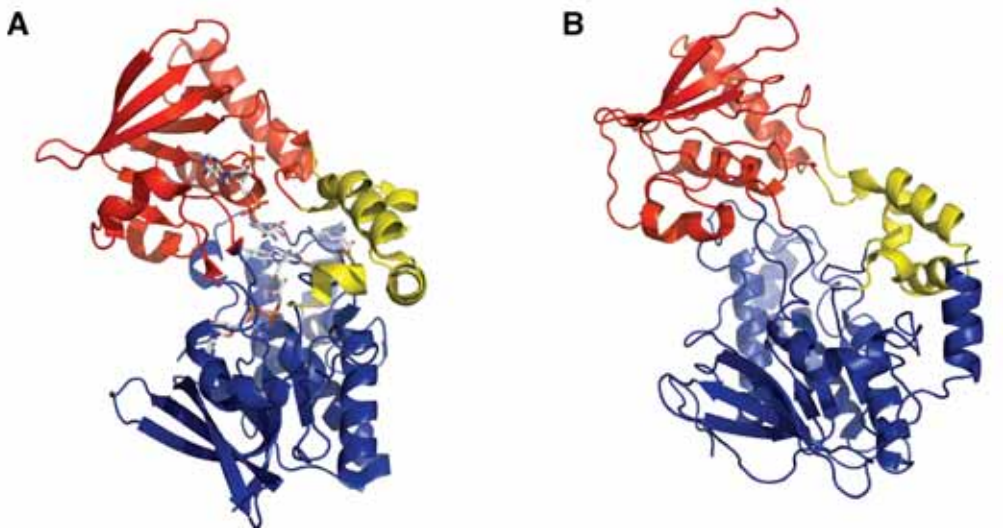


Figure 5.5 The comparison between the crystal structure of the *P. aeruginosa* L-Orn monooxygenase PvdA (PDB code: 36S1, A) and the I-TASSER[226] structure prediction of Rmo (B, prediction C-score 0.90). The FAD-binding domain is colored in blue, the NADPH-binding domain in red and the L-Orn binding domain in yellow. The catalytic pocket is located at the interface of the three domains. In the PvdA structure FAD, NADPH and L-Orn are shown as CPK-colored sticks.

In a multiple sequence alignment, Rmo shares a high degree of sequence identity with the aforementioned homologues, especially around the sites involved in the coordination of the substrates and the FAD cofactor. In model prediction, the Rmo structure reflects the overall characteristic three-domains folding of the PvdA's

crystal structure (Rmo homologue from *P. aeruginosa*, Figure 5.5). The NADPH-binding domain is inserted within the largest FAD-binding domain and the active site pocket is formed between the two and the opposite smallest L-Orn binding domain.[182] Spectrophotometric analysis of the recombinant enzyme confirmed Rmo to be purified in the *apo* form, without bound FAD cofactor. Again, according to the crystal structure of PvdA, the FAD binding site is located in close proximity to the surface of the enzyme. Therefore, the loss of the flavin cofactor during purification is not unexpected.[182] In the presence of molecular oxygen, NADPH and FAD, the Rmo monooxygenase was able to convert L-Orn into L-hOrn. As other members of the NMO family, Rmo shows an exclusive preference toward its cognate amino acid substrate and the reducing cosubstrate NADPH. In addition, similar to SidA and PvdA, Rmo adopts a tetrameric quaternary structure in solution, which does not alter the catalytic properties of the single subunits, resulting in classical Michaelis–Menten behavior.[178,179] When compared with other L-Orn monooxygenase homologues, the determined kinetic parameter are in accordance with previously published results, despite the fact that Rmo does not possess a similar high catalytic efficiency as EtcB or VbsO (Table 5.4).[94,225]

Table 5.4 Kinetic parameter for Rmo-mediated L-Orn hydroxylation and comparison with others homologues L-Orn monooxygenases involved in siderophore biosynthesis.

enzyme	siderophore	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
Rmo[a]	rhodochelin	1.6 ± 0.2	0.2331 ± 0.008	0.15
EtcB[b,94]	erythrocelin	0.286 ± 0.035	0.3267 ± 0.0005	1.14
CchB[b,88]	coelichelin	3.6 ± 0.58	0.290 ± 0.01	0.081
VbsO[b,225]	vicibactin	0.305 ± 0.024	1.80 ± 0.03	5.90
PvdA[a,178]	pyoverdin	0.60 ± 0.07	0.400 ± 0.05	0.67
SidA[a,179]	ferrichrome	1.70 ± 0.06	0.4833 ± 0.005	0.284
SidA[a,180]	ferrichrome	0.58 ± 0.07	0.611	1.0

a steady state parameters for hydroxylated product formation.
b steady state parameters for coupled NADPH oxidation assay.

Extensive kinetic analyses of PvdA and SidA have allowed the postulation of a reaction mechanism for the L-Orn hydroxylation (Figure 5.6).[180,227] Upon reduction of the flavin cofactor by NADPH, the subsequent binding of molecular oxygen results into two different intermediates. For PvdA, it is hypothesized to be a peroxyflavin which is subsequently converted to the active and long-lived hydroperoxyflavin by the donation of a proton from the amino group of the substrate side chain.[178] For SidA, it is a hydroperoxyflavin species that is stabilized by the presence of $NADP^+$.

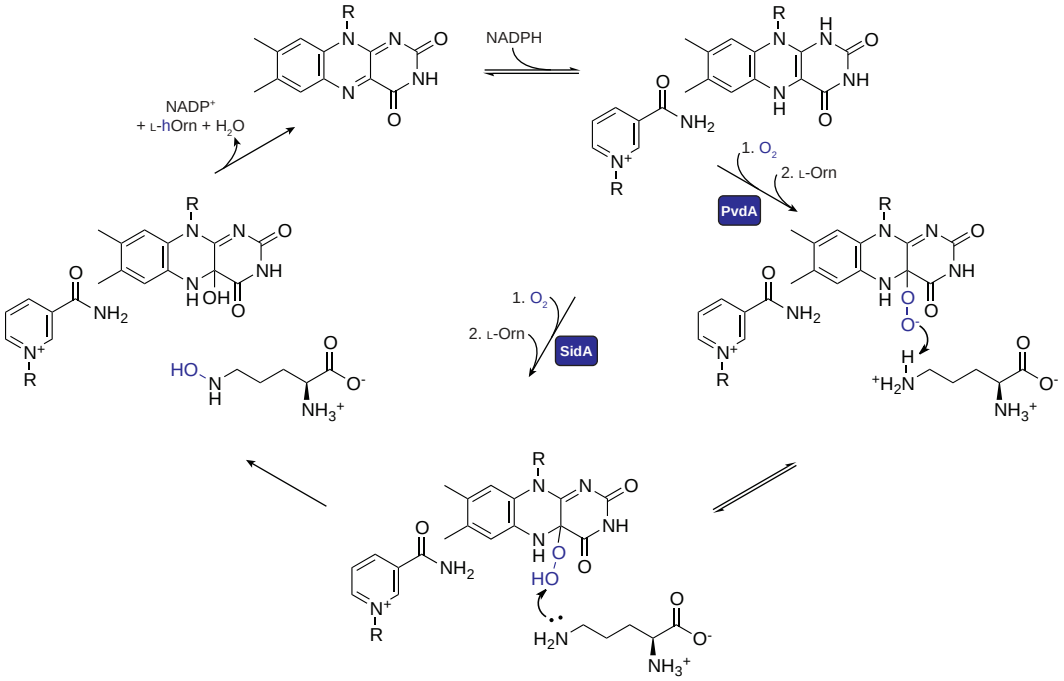


Figure 5.6 Mechanism for the *N*-hydroxylation flavoprotein monooxygernases based on kinetic data for PvdA and SidA.[178-180,182,227]

In both cases, the subsequent binding of the L-Orn substrate (which is therefore uncoupled from the NADPH oxidation) enhances the rate of formation of the reactive hydroperoxyflavin intermediate that donates its distal oxygen to the amino group of the L-Orn side chain, giving the L-hOrn product. The cycle ends with the dehydration of the hydroxyflavin to flavin (regenerating the cofactor) and the dissociation of the L-hOrn and $NADP^+$.

In conclusion, Rmo represents a typical member of the NADPH/FAD-dependent monooxygenases required for the δ-*N*-hydroxylation of L-Orn or L-Lys side chains associated with the biosynthesis of hydroxamate-type siderophores.[15]

5.3.2 Characterization of the L-hOrn formyltransferase Rft

The biochemical characterization of Rft represents the first *in vitro* study of a tailoring formyltransferase involved in the biosynthesis of formyl-based iron-coordinating hydroxamate moieties. In fact, previous attempts to elucidate the role of homologous and analogous enzymes (Ccha, PvdF, and FxbA) were not successful or relied on the use of genetic strategies employing indirect detection methods.[88,188,189] Rtf catalyzes the N^{10}-fH_4F-depented formylation reaction of the L-hOrn intermediate, establishing a route for the L-fhOrn biosynthesis, prior to its use by the NRPS assembly line.

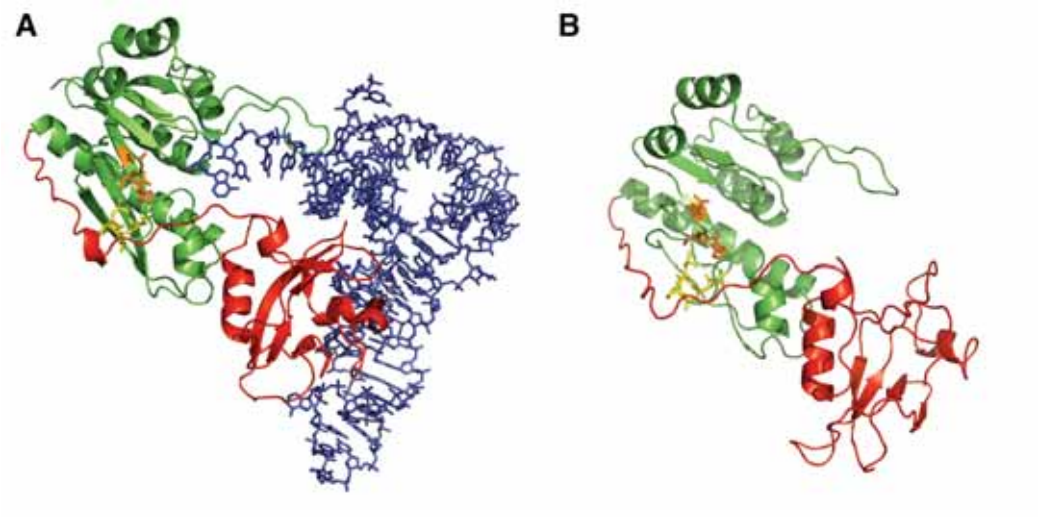


Figure 5.7 The comparison between the crystal structure of the *E. coli* methionyl-tRNA^{fMet}-formyltransferase (PBD code: 2FMT, A) and the I-TASSER[226] structure prediction of Rft (B, prediction C-score 1.50). The catalytic N-terminal subdomain is colored in green, the C-terminal subdomain in red; in the *E. coli* formyltransferase, the latter subdomain is involved in the proper recognition and binding of the methionyl-tRNA^{fMet} substrate (blue).[187] The SLLP motif, required for the coordination of the N^{10}-fH$_4$F cofactor, and the catalytic triad are shown as yellow and orange sticks, respectively.

In contrast to the endogenous methionyl-tRNA^{fMet}-formyltransferase, ArnA, and the F-domain embedded within the LgrA initiation module, Rft catalyzes the formylation of a secondary amine, whereas the previously mentioned enzymes modify a primary amino group.[80,184,185] On the basis of the sequence similarities of Rft with ArnA and the other N^{10}-fH$_4$F-depented formyltransferases, a mechanism for the L-hOrn transformylation reaction is proposed (Figures 5.7 and 5.8).[228-231] His$_{105}$ and Asn$_{103}$ activate the carbonyl carbon of the formyl group, which is attacked by the secondary amine of L-hOrn (the putative intermediate is stabilized by both His$_{105}$ and Asn$_{103}$). A water molecule, properly positioned in the active site by the Asp$_{141}$ side chain, mediates the proton transfer from the amine group of the L-hOrn side chain to the tetrahydrofolate.[232] Subsequently, the reaction ends with the decomposition of the tetrahedral intermediate and the release of the L-fhOrn product and tetrahydrofolate. The use of an alternative strategy for the generation of the unstable, and thus commercially unavailable, N^{10}-fH$_4$F cosubstrate, namely the *in situ* transformation of N^5,N^{10}-methenylH$_4$F to the desired N^{10}-fH$_4$F, through preincubation in the assay buffer (Figure 5.9), has proved to be successful, resulting in enzymatically active cofactor. Likewise, this strategy was applied to the hitherto inactive CchA, promoting the conversion of L-hOrn to L-fhOrn, implying a similar biosynthetic mechanism for the same iron-coordinating group in the siderophore coelichelin.[88]

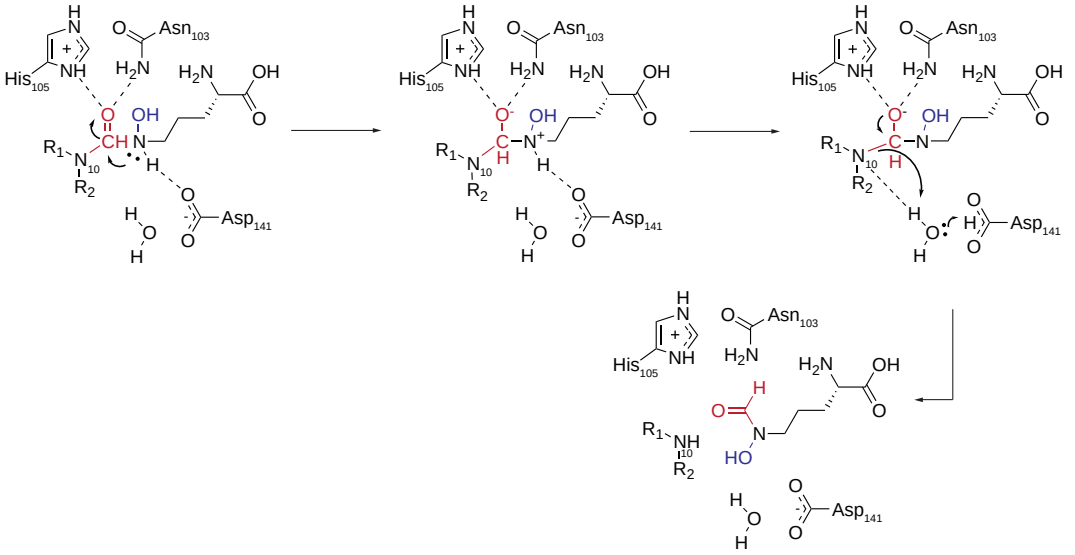


Figure 5.8 Suggested catalytic mechanism for the Rft-dependent L-hOrn transformylation reaction, based on the reaction mechanism proposed by Gatzeva-Topalova *et al*.[228]

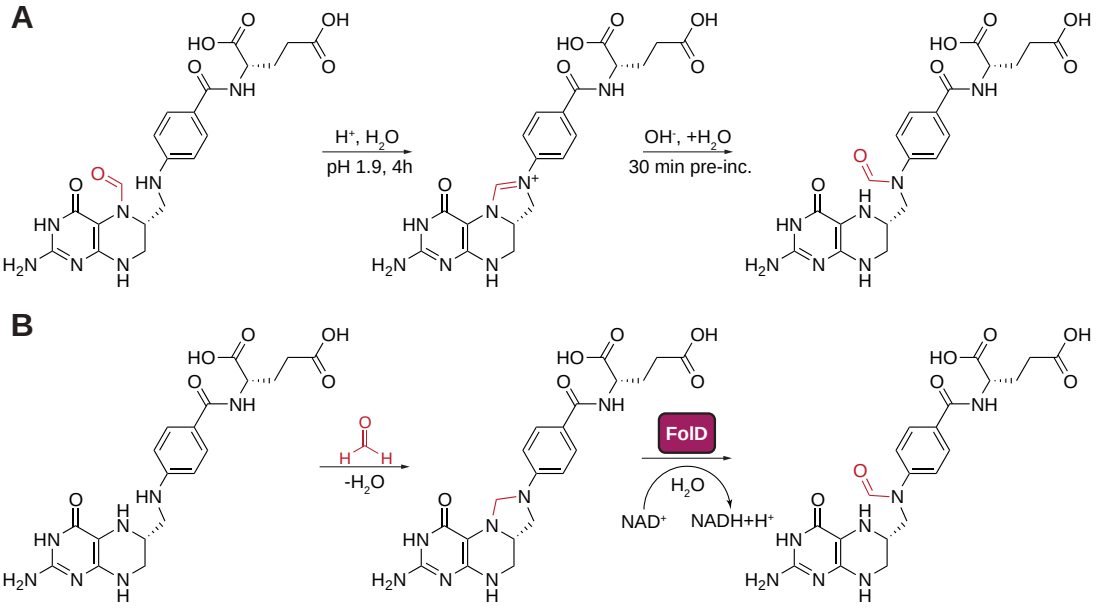


Figure 5.9 (A) Synthesis of the N^{10}-fH$_4$F cofactor via the chemical conversion of the N^5-fH$_4$F substrate to the to N^5,N^{10}-methenylH$_4$F intermediate. Prior to the addition of the enzyme, the preincubation of the intermediate in the reaction buffer (pH 7.5) allows the final conversion to N^{10}-fH$_4$F. (B) Chemoenzymatic synthesis of N^{10}-fH$_4$F. The H$_4$F substrate is spontaneously converted to the N^5,N^{10}-methyleneH$_4$F intermediate by the addition of formaldehyde. Then, the FolD dehydrogenase/cyclohydrolase catalyzes the final conversion to N^{10}-fH$_4$F.[192] Solely the first approach generated active N^{10}-fH$_4$F cofactor for the characterization of the L-hOrn formyltransferase activities of Rft and CchA.

Rft kinetic characterization indicates a positive cooperative effect, possibly resulting from the fact that, in solution, the enzyme adopts a tetrameric quaternary structure. In this context, the conserved presence of the C-terminal subdomain (which does not feature any catalytic residues) could be seen as a modification of the original tRNA binding function of the methionyl-tRNA^{fMet}-formyltransferase to an oligomerization and/or allosteric role in Rft.[187] In addition, the positive cooperative control mechanism could improve the overall biosynthesis of the non-proteinogenic L-fhOrn amino acid (and thus its incorporation by the NRPS assembly line), providing an enhancement for siderophore biosynthesis under the restrictive iron-limiting growing conditions. Moreover, it is not to be excluded that both Rmo and Rft tetramers could undergo an additional oligomerization, raising up a functional octamer that further enhance L-fhOrn biosynthesis, in a similar fashion the EntE-dependent 2,3-DHB adenylation is stimulated by the protein-protein interaction of a EntA tetramer.[233]

In conclusion, on the basis of the results of the coupled assay and the additional inability of Rmo to hydroxylate L-fOrn and of Rft to transformyate L-Orn, a model for the biosynthesis of the formyl-based hydroxamate-containing siderophores could be proposed, according to the so-called "hydroxylation first" mechanism, recently described for the acetyl-based hydroxamates (Figure 5.10).[94] Initially, the L-Orn side chain amino group is hydroxylated by a NMO enzyme, and then the newly modified L-hOrn could be either incorporated by the NRPS assembly line (coelichelin) or further modified by formylation, leading to the generation of iron-coordinating L-fhOrn (coelichelin or rhodochelin).

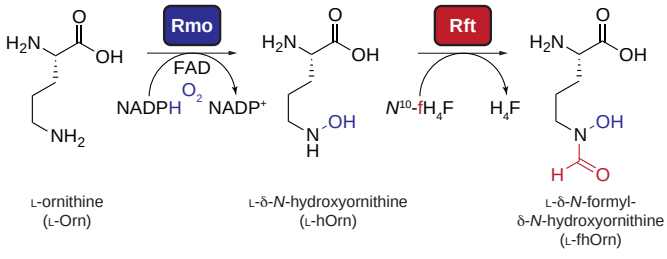


Figure 5.10 Coupled enzymatic biosynthesis of L-fhOrn. Rmo converts L-Orn into L-hOrn. This hydroxylated intermediate serves as a substrate for the subsequent N^{10}-fH_4F Rtf-catalyzed formylation reaction, resulting in the non-proteinogenic L-fhOrn, subsequently incorporated in the peptide scaffold by the RhcB NRPS.

5.4 Biochemical and genetic model for rhodochelin biosynthesis

5.4.1 A biosynthetic model for rhodochelin assembly

The results obtained in this study allow the postulation of a model for rhodochelin biosynthesis (Figure 5.11). Rhodochelin assembly is initiated by DhbE, which activates 2,3-DHB that is subsequently transferred to its cognate stand-alone aryl carrier protein RhcE. RhcB assembles the tripeptide DHB-L-Thr-L-fhOrn following the classical linear logic of NRPS assembly lines, similar to the fashion in which the tripeptide DHB-L-Gly-L-Thr of bacillibactin is formed.[35] From the RhcB-PCP_2 the newly assembled tripeptide is transferred to the conserved catalytic Ser of the TE-domain. Subsequently, a second L-fhOrn is proposed to be activated by the respective domain (following a module skipping mechanism, similar to coelichelin assembly) and tethered to RhcB-PCP_2.[13] With the tripeptide and the monomer residing in adjacent positions, the nucleophilic attack of the hydroxyl group in the L-Thr side chain on the L-fhOrn thioester could occur. This mechanism is consistent with an iterative TE-domain following a "forward" mechanism, like the proposed lactonization mechanism of DHB-L-Gly during enterobactin biosynthesis.[104] Then, the branched tetrapeptide, still attached to the TE, is hydrolytically released and exported to the extracellular space to carry out its biological function.

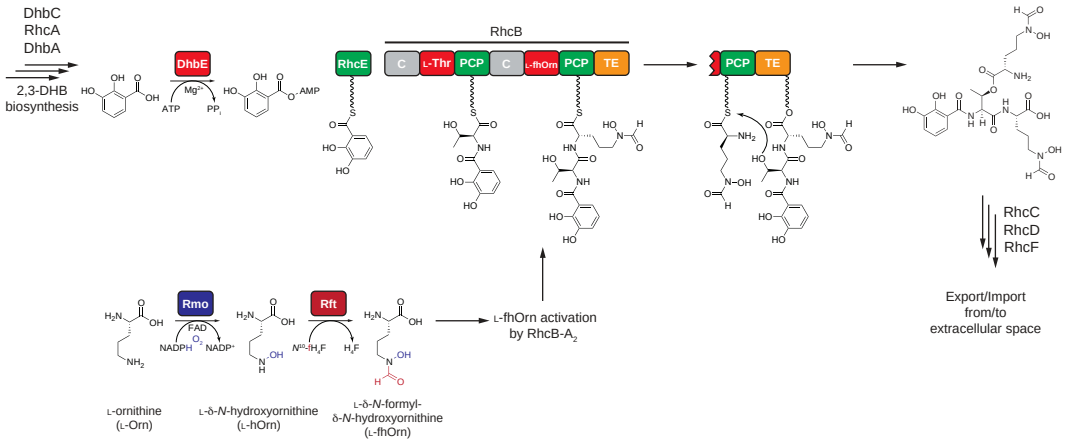


Figure 5.11 Proposed biosynthetic pathway for rhodochelin assembly. The two non-proteinogenic building blocks 2,3-DHB and L-fhOrn are synthesized by the corresponding pathways and channeled to the synthetase RhcB. Prior to the TE-catalyzed release, a second L-fhOrn building block is activated, following a "module skipping" mechanism. A detail of the ester bond formation between L-fhOrn and the side chain of L-Thr is presented.

The tripeptide found in the culture supernatant also displays iron-scavenging properties (Figure S5). In addition, while it was demonstrated that the hydrolysis of the labile ester bond occur spontaneously in the culture medium, with regards to

the postulated "forward" mechanism, it cannot be completely excluded that rhodochelin tripeptide is the result of the premature release from the TE-domain during the assembly of the peptide.

5.4.2 Putative regulation of the rhodochelin biosynthesis

In GC-rich Gram-positive bacteria, control of the expression of the genes involved in iron metabolism is usually achieved by DmdR (divalent metal-dependent regulatory protein). This system was previously known as DtxR, where it was first reported to control the expression of the diphtheria toxin gene *tox*.[234] The *tox* gene promoter contains a palindromic consensus that is recognized by a DtxR dimer, upon binding of a divalent metal ion (Fe^{2+}, Co^{2+}, Ni^{2+}, Cd^{2+}).[235-241] DmdR/DtxR homologues were found in *M. tuberculosis*, *Streptomyces pilosus*, *Streptomyces lividans*, *S. coelicolor*, and *R. equii*.[242-246] In particular, in *S. coelicolor*, two DmdR homologues (namely DmdR1 and DmdR2) were reported to bind the 19-bp iron boxes consensus located upstream of the *tox* and the *des*A genes, forming two different complexes: whereas DmdR1 only forms dimers, DmdR2 is also capable of forming tetramers. In addition, seven different iron-boxes were found in ten different locations in the genome of *S. coelicolor*, most of them located upstream of putative genes involved in siderophore biosynthesis. Lastly, the comparison of the iron-boxes sequence allowed the definition of a consensus for the *Streptomyces* species (TTAGGTTAGGCTCACCTAA).[245]

On the basis of this knowledge, two DmdR homologues were identified in the genome of *R. jostii* RHA1. The putative protein product of the ORF *RHA1_06810* displayed 60% sequence identity (73% of sequence similarity) with DmdR1, whereas the putative product of the ORF *RHA1_ro06652* showed 29% sequence identity (48% of sequence similarity), respectively.

Figure 5.12 Comparison between the genomic regions of *R. jostii* RHA1 and *C. diphtheriae* where the iron-dependent repressor *dtx*R is located.[247] *gal*E is predicted to encode a putative UDP-galactose 4-epimerase,[248] *sig*B a sigma transcription factor with homology to σ^{70}-type family, *ro06811* an acetoin dehydrogenase and *ro06812* an acetyl-CoA synthetase.

In addition, *RHA1_06810* is located in a genomic region that displays a similar genic organization with respect to the chromosomal locus of *C. diphtheriae* where *dtx*R is

located (Figure 5.12).[247] More interestingly, the use of the *S. coelicolor* iron-box consensus sequence allowed the identification of six different iron-boxes within the *R. jostii* RHA1 chromosome, all located in the three genic cluster responsible for the biosynthesis of rhodochelin, approximately 100 bp upstream of the start codon of the following genes: *rhc*A, *rhc*B, *rhc*D / *rhc*E, *RHA1_ro04715*, *rmo* and *dhb*C (Figure 5.13). Analogously, a putative iron-box consensus for *R. jostii* RHA1 was proposed (TTAGGTTAGGCTANCCTTA).

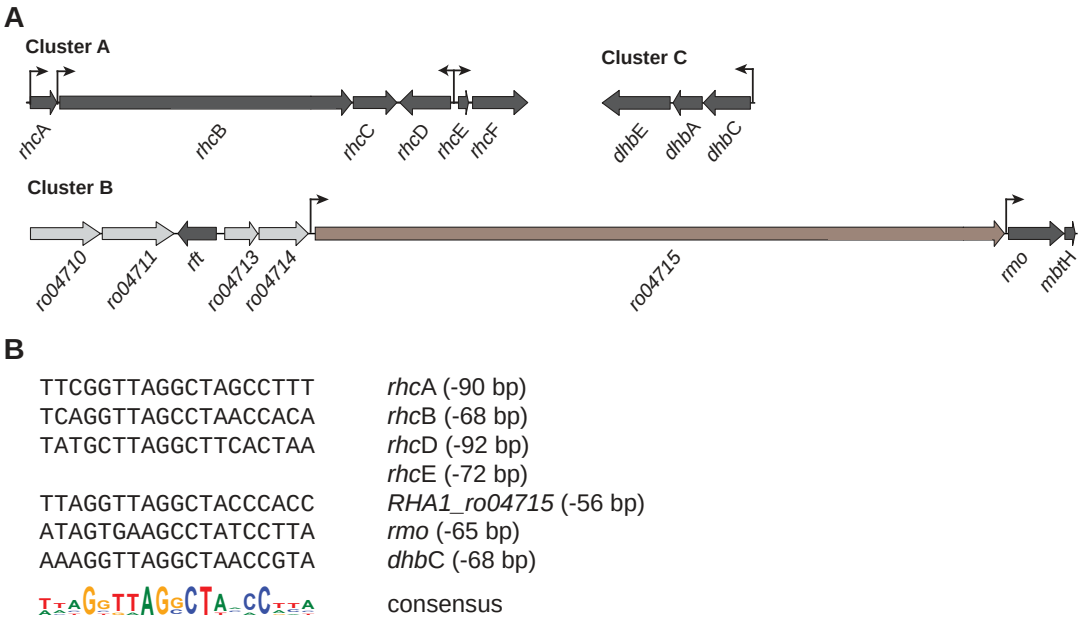


Figure 5.13 (A) Location of the six iron-boxes within the three rhodochelin gene clusters. The sequences were identified using consensus identified using the Virtual Footprint tool of the PRODORIC database.[249] (B) Multiple sequence alignment of the iron-boxes located upstream the rhodochelin biosynthetic genes. The distance to the closest gene is indicated in parenthesis. In addition, a putative iron-box consensus for *R. jostii* RHA1 is proposed (TTAGGTTAGGCTANCCTTA).[250]

The current results suggest that, similarly to *S. coelicolor*, an analogous system for the global iron-dependent transcriptional regulation could be present in *R. jostii* RHA1.[245] Under iron-limited conditions, the absence of the ferrous iron in the intracellular space promotes the dissociation of the DmdR repressor dimer from its regulatory sequence, allowing the transcription of the genes associated with rhodochelin biosynthesis. Five iron-boxes are located directly upstream of all the genes responsible for the precursor biosynthesis, the siderophore assembly and its subsequent export to and uptake from the extracellular space, while the sixth is located upstream the cryptic NRPS gene *RHA1_ro04715*. Therefore, although the cryptic NRPS gene seems to be under the same transcriptional control as the genes for rhodochelin's biosynthesis, the evidence that no product has been detected does

not exclude the synthetase to be functional, to be silent (other unknown repressors or activators could play a role at a different hierarchical transcriptional level) or the growing conditions applied are inadequate to allow detection. On the other hand, the absence on the iron-box consensus upstream of the *rft* gene still does not exclude a different transcriptional control, or, that the biosynthesis of the nonproteinogenic L-fhOrn amino acid is simply enhanced by the positive cooperative mechanism of the Rft tetramer *in vivo*.

5.5 Perspective and outlook

Microorganisms produces a multitude of small molecules that are not used in the primary metabolism but take part in important physiological and ecological roles, mediating the interactions among other species of the community and/or with multicellular organisms.[213] The discovery that the genes responsible for the biosynthesis of natural products are often clustered together in a genome facilitated the important genetic and the biochemical findings during the last two decades. Conversely, the presented discovery of cross-talk mechanisms between three gene clusters for the biosynthesis of a single secondary metabolite in *R. jostii* RHA1,[251] supports an unexpected level of complexity, especially for a molecule with a relative simple (but still important) function, like a siderophore. In this context, the hypothesis that natural product gene clusters are among the most diverse and rapidly evolving genetic elements of a chromosome reflects the genome plasticity towards the natural selective pressure, continuously remodeling its elements to best fit the environment. In addition, understanding the molecules and the mechanisms that microorganism use to interact with each other can lead to a deeper understanding of their physiology, ecology and evolution and, lately, to a new and better drug discovery for human therapy.[252]

The existence of approximately 20 secondary metabolite gene clusters in *R. jostii* RHA1 that remains still orphan with respect to their biosynthesized molecule put the strain into account for the isolation and the structural characterization of new natural products. Several strategies have already been reported being successful and can be employed to elicit cryptic pathways in other species; for example, triggering the global nutritional regulators of primary metabolism under phosphate or nitrogen starvation[253], using of small chemical as elicitors,[254] or with the co-cultivation in the presence of other species of the same ecosystem, in order to attempt to simulate the natural habitat and prompt the inter-species communication.[255] Apart from culture-dependent techniques, the availability of different genetic tools specifically designed for *Rhodoccocus* spp. allows the genetic engineering of the strain and thus the identification of new natural products through comparative metabolic profiling.[44] In particular, the use of plasmid systems specifically engineered for heterologous gene expression in *Rhodoccocus* spp.[256] can be employed to activate cryptic biosynthetic gene clusters through the inducible and/or constitutive manipulation of pathway-specific activators, as recently demonstrated for *Streptomyces ambofaciens* and *Burkholderia* spp.[257-259]

References

1. Pierre, J. L., Fontecave, M. & Crichton, R. R. Chemistry for an essential biological process: the reduction of ferric iron. *Biometals* **15**, 341–346 (2002).
2. Miethke, M. & Marahiel, M. A. Siderophore-based iron acquisition and pathogen control. *Microbiol Mol Biol Rev* **71**, 413–451 (2007).
3. Andrews, S. C., Robinson, A. K. & Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**, 215–237 (2003).
4. Wandersman, C. & Delepelaire, P. Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol* **58**, 611–647 (2004).
5. Schaible, U. E. & Kaufmann, S. H. E. Iron and microbial infection. *Nat Rev Microbiol* **2**, 946–953 (2004).
6. Winkelmann, G. & Drechsel, H. Microbial Siderophores. *Biotechnology* **8**, 199–246 (2008).
7. Miethke, M. *et al.* Ferri-bacillibactin uptake and hydrolysis in Bacillus subtilis. *Mol Microbiol* **61**, 1413–1427 (2006).
8. Peuckert, F., Miethke, M., Albrecht, A. G., Essen, L.-O. & Marahiel, M. A. Structural basis and stereochemistry of triscatecholate siderophore binding by FeuA. *Angew Chem Int Ed Engl* **48**, 7924–7927 (2009).
9. Sandy, M. & Butler, A. Microbial iron acquisition: marine and terrestrial siderophores. *Chem Rev* **109**, 4580–4595 (2009).
10. Crumbliss, A. L. & Harrington, J. M. Iron sequestration by small molecules: Thermodynamic and kinetic studies of natural siderophores and synthetic model compounds. *Advances in Inorganic Chemistry* **Volume 61**, 179–250 T2 – (2009).
11. Raymond, K. N., Dertz, E. A. & Kim, S. S. Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci USA* **100**, 3584–3588 (2003).
12. May, J. J., Wendrich, T. M. & Marahiel, M. A. 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* **276**, 7209–7217 (2001).
13. Lautru, S., Deeth, R. J., Bailey, L. M. & Challis, G. L. Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. *Nat Chem Biol* **1**, 265–269 (2005).
14. Robbel, L., Knappe, T. A., Linne, U., Xie, X. & Marahiel, M. A. Erythrochelin--a hydroxamate-type siderophore predicted from the genome of Saccharopolyspora erythraea. *FEBS J* **277**, 663–676 (2010).
15. Olucha, J. & Lamb, A. L. Mechanistic and structural studies of the N-hydroxylating flavoprotein monooxygenases. *Bioorg. Chem.* **39**, 171–177 (2011).
16. Martinez, J. S. *et al.* Structure and membrane affinity of a suite of amphiphilic siderophores produced by a marine bacterium. *Proc Natl Acad Sci USA* **100**, 3754–3759 (2003).
17. Martinez, J. S. *et al.* Self-assembling amphiphilic siderophores from marine bacteria. *Science* **287**, 1245–1247 (2000).
18. Meiwes, J. *et al.* Isolation and characterization of staphyloferrin A, a compound with siderophore activity from Staphylococcus hyicus DSM 20459. *FEMS Microbiol Rev* **55**, 201–205 (1990).
19. Oves-Costales, D., Kadi, N. & Challis, G. L. The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chem. Commun. VL* - 6530–6541 (2009).doi:10.1039/b913092f
20. Barry, S. M. & Challis, G. L. Recent advances in siderophore biosynthesis. *Current Opinion in Chemical Biology* **13**, 205–215 (2009).
21. de Lorenzo, V., Bindereif, A., Paw, B. H. & Neilands, J. B. Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in Escherichia coli K-12. *J Bacteriol* **165**, 570–578 (1986).
22. Thariath, A., Socha, D., Valvano, M. A. & Viswanatha, T. Construction and biochemical characterization of recombinant cytoplasmic forms of the IucD protein (lysine:N6-hydroxylase) encoded by the pColV-K30 aerobactin gene cluster. *J Bacteriol* **175**, 589–596 (1993).
23. Coy, M., Paw, B. H., Bindereif, A. & Neilands, J. B. Isolation and properties of N epsilon-hydroxylysine:acetyl coenzyme A N epsilon-transacetylase from Escherichia coli pABN11. *Biochemistry* **25**, 2485–2489 (1986).
24. Kadi, N., Oves-Costales, D., Barona-Gómez, F. & Challis, G. L. A new family of ATP-dependent oligomerization-macrocyclization biocatalysts. *Nat Chem Biol* **3**, 652–656 (2007).
25. Kadi, N., Arbache, S., Song, L., Oves-Costales, D. & Challis, G. L. Identification of a gene cluster that directs putrebactin biosynthesis in Shewanella species: PubC catalyzes cyclodimerization of N-hydroxy-N-succinylputrescine. *J Am Chem Soc* **130**, 10458–10459 (2008).
26. Challis, G. L. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chembiochem* **6**, 601–611 (2005).
27. Schwarzer, D. & Marahiel, M. A. Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* **88**, 93–101 (2001).

28. Nolan, E. M. & Walsh, C. T. How nature morphs peptide scaffolds into antibiotics. *Chembiochem* **10**, 34–53 (2009).
29. Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol* **58**, 453–488 (2004).
30. Newman, D. J., Cragg, G. M. & Snader, K. M. The influence of natural products upon drug discovery. *Nat Prod Rep* **17**, 215–234 (2000).
31. Hopwood, D. A. *Streptomyces in Nature And Medicine*. 250 (Oxford University Press, USA: 2007).
32. Walsh, C. T. & Fischbach, M. A. Natural products version 2.0: connecting genes to molecules. *J Am Chem Soc* **132**, 2469–2493 (2010).
33. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663 (2008).
34. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174–4183 (1997).
35. May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci USA* **99**, 12120–12125 (2002).
36. Yonus, H. *et al.* Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *Journal of Biological Chemistry* **283**, 32484–32491 (2008).
37. Hegg, E. L. Unraveling the structure and mechanism of acetyl-coenzyme A synthase. *Acc Chem Res* **37**, 775–783 (2004).
38. Conti, E., Lloyd, L. F., Akins, J., Franks, N. P. & Brick, P. Crystallization and preliminary diffraction studies of firefly luciferase from Photinus pyralis. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 876–878 (1996).
39. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493–505 (1999).
40. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic acids research* **33**, 5799–5808 (2005).
41. Bachmann, B. O. & Ravel, J. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Meth Enzymol* **458**, 181–217 (2009).
42. Weber, T. *et al.* CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology* **140**, 13–17 (2009).
43. Challis, G. L. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154**, 1555–1569 (2008).
44. Zerikly, M. & Challis, G. L. Strategies for the discovery of new natural products by genome mining. *Chembiochem* **10**, 625–633 (2009).
45. Francklyn, C. S. DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression. *Biochemistry* **47**, 11695–11703 (2008).
46. Stachelhaus, T., Hüser, A. & Marahiel, M. A. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem Biol* **3**, 913–921 (1996).
47. Lambalot, R. H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol* **3**, 923–936 (1996).
48. Mootz, H. D., Finking, R. & Marahiel, M. A. 4'-phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis. *Journal of Biological Chemistry* **276**, 37289–37298 (2001).
49. Schmoock, G. *et al.* Functional cross-talk between fatty acid synthesis and nonribosomal peptide synthesis in quinoxaline antibiotic-producing streptomycetes. *Journal of Biological Chemistry* **280**, 4339–4349 (2005).
50. Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273–276 (2006).
51. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1999).
52. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *Journal of Biological Chemistry* **273**, 22773–22781 (1998).
53. Linne, U. & Marahiel, M. A. Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39**, 10439–10447 (2000).
54. Luo, L. *et al.* Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry* **41**, 9184–9196 (2002).

55. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
56. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
57. Leslie, A. G. Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 A resolution. *J Mol Biol* **213**, 167–186 (1990).
58. Kopp, F. & Marahiel, M. A. Where chemistry meets biology: the chemoenzymatic synthesis of nonribosomal peptides and polyketides. *Curr. Opin. Biotechnol.* **18**, 513–520 (2007).
59. Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat Prod Rep* **24**, 735–749 (2007).
60. Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301–310 (2002).
61. Samel, S. A., Wagner, B., Marahiel, M. A. & Essen, L.-O. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J Mol Biol* **359**, 876–889 (2006).
62. Pazirandeh, M., Chirala, S. S., Huang, W. Y. & Wakil, S. J. Characterization of recombinant thioesterase and acyl carrier protein domains of chicken fatty acid synthase expressed in Escherichia coli. *Journal of Biological Chemistry* **264**, 18195–18201 (1989).
63. Tseng, C. C. *et al.* Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* **41**, 13350–13359 (2002).
64. Hubbard, B. K. & Walsh, C. T. Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl* **42**, 730–765 (2003).
65. Chiu, H. T. *et al.* Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Natl Acad Sci USA* **98**, 8548–8553 (2001).
66. Miller, D. A., Luo, L., Hillson, N., Keating, T. A. & Walsh, C. T. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chem Biol* **9**, 333–344 (2002).
67. Kessler, N., Schuhmann, H., Morneweg, S., Linne, U. & Marahiel, M. A. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *Journal of Biological Chemistry* **279**, 7413–7419 (2004).
68. Kopp, F., Mahlert, C., Grünewald, J. & Marahiel, M. A. Peptide macrocyclization: the reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine. *J Am Chem Soc* **128**, 16478–16479 (2006).
69. Magarvey, N. A., Haltli, B., He, M., Greenstein, M. & Hucul, J. A. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. *Antimicrobial Agents and Chemotherapy* **50**, 2167–2177 (2006).
70. Debono, M. *et al.* Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J Antibiot* **41**, 1093–1105 (1988).
71. Stein, D. B., Linne, U., Hahn, M. & Marahiel, M. A. Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis. *Chembiochem* **7**, 1807–1814 (2006).
72. Konz, D. & Marahiel, M. A. How do peptide synthetases generate structural diversity? *Chem Biol* **6**, R39–48 (1999).
73. Schneider, T. L., Shen, B. & Walsh, C. T. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**, 9722–9730 (2003).
74. Gehring, A. M., Mori, I., Perry, R. D. & Walsh, C. T. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis. *Biochemistry* **37**, 11637–11650 (1998).
75. Shen, B. *et al.* Cloning and Characterization of the Bleomycin Biosynthetic Gene Cluster from StreptomycesverticillusATCC15003 1. *J Nat Prod* **65**, 422–431 (2002).
76. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4**, 927–937 (1997).
77. Du, L., Chen, M., Zhang, Y. & Shen, B. BlmIII and BlmIV nonribosomal peptide synthetase-catalyzed biosynthesis of the bleomycin bithiazole moiety involving both in cis and in trans aminoacylation. *Biochemistry* **42**, 9731–9740 (2003).
78. Reimmann, C. *et al.* Essential PchG-Dependent Reduction in Pyochelin Biosynthesis of Pseudomonas aeruginosa. (2001).
79. Ansari, M. Z., Sharma, J., Gokhale, R. S. & Mohanty, D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinformatics* **9**, 454 (2008).
80. Schoenafinger, G., Schracke, N., Linne, U. & Marahiel, M. A. Formylation Domain: An Essential Modifying Enzyme for the Nonribosomal Biosynthesis of Linear Gramicidin. *J Am Chem Soc* **128**, 7406–7407 (2006).

81. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends in Biochemical Sciences* **28**, 329–335 (2003).
82. Hojati, Z. *et al.* Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyces coelicolor. *Chem Biol* **9**, 1175–1187 (2002).
83. Mahlert, C., Kopp, F., Thirlway, J., Micklefield, J. & Marahiel, M. A. Stereospecific enzymatic transformation of alpha-ketoglutarate to (2S,3R)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J Am Chem Soc* **129**, 12011–12018 (2007).
84. Huang, Y.-T. *et al.* In vitro characterization of enzymes involved in the synthesis of nonproteinogenic residue (2S,3S)-beta-methylphenylalanine in glycopeptide antibiotic mannopeptimycin. *Chembiochem* **10**, 2480–2487 (2009).
85. He, H. *et al.* Mannopeptimycins, novel antibacterial glycopeptides from Streptomyces hygroscopicus, LL-AC98. *J Am Chem Soc* **124**, 9729–9736 (2002).
86. van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *Journal of Biotechnology* **124**, 670–689 (2006).
87. Kopp, F., Linne, U., Oberthür, M. & Marahiel, M. A. Harnessing the chemical activation inherent to carrier protein-bound thioesters for the characterization of lipopeptide fatty acid tailoring enzymes. *J Am Chem Soc* **130**, 2656–2666 (2008).
88. Pohlmann, V. & Marahiel, M. A. Delta-amino group hydroxylation of L-ornithine during coelichelin biosynthesis. *Org Biomol Chem* **6**, 1843–1848 (2008).
89. Strieker, M., Kopp, F., Mahlert, C., Essen, L.-O. & Marahiel, M. A. Mechanistic and Structural Basis of Stereospecific Cβ-Hydroxylation in Calcium-Dependent Antibiotic, a Daptomycin-Type Lipopeptide. *ACS Chem Biol* **2**, 187–196 (2007).
90. Helmetag, V., Samel, S. A., Thomas, M. G., Marahiel, M. A. & Essen, L.-O. Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *FEBS J* **276**, 3669–3682 (2009).
91. Strieker, M., Nolan, E. M., Walsh, C. T. & Marahiel, M. A. Stereospecific Synthesis of threo- and erythro-β-Hydroxyglutamic Acid During Kutzneride Biosynthesis. *J Am Chem Soc* **131**, 13523–13530 (2009).
92. Bruijnincx, P. C. A., van Koten, G. & Klein Gebbink, R. J. M. Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem Soc Rev* **37**, 2716–2744 (2008).
93. Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem Rev* **104**, 939–986 (2004).
94. Robbel, L., Helmetag, V., Knappe, T. A. & Marahiel, M. A. Consecutive enzymatic modification of ornithine generates the hydroxamate moieties of the siderophore erythrochelin. *Biochemistry* **50**, 6073–6080 (2011).
95. Linne, U., Schwarzer, D., Schroeder, G. N. & Marahiel, M. A. Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *Eur J Biochem* **271**, 1536–1545 (2004).
96. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc Natl Acad Sci USA* **99**, 14083–14088 (2002).
97. Yeh, E., Kohli, R. M., Bruner, S. D. & Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *Chembiochem* **5**, 1290–1293 (2004).
98. Schneider, A. & Marahiel, M. A. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. *Arch Microbiol* **169**, 404–410 (1998).
99. Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol* **179**, 6843–6850 (1997).
100. Cosmina, P. *et al.* Sequence and analysis of the genetic locus responsible for surfactin synthesis in Bacillus subtilis. *Mol Microbiol* **8**, 821–831 (1993).
101. Miao, V. *et al.* Daptomycin biosynthesis in Streptomyces roseosporus: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507–1523 (2005).
102. Smith, D. J. *et al.* Beta-lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. *EMBO J* **9**, 741–747 (1990).
103. Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648–2659 (1998).
104. Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chem Biol* **6**, 385–400 (1999).
105. Juguet, M. *et al.* An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in Streptomyces ambofaciens. *Chem Biol* **16**, 421–431 (2009).
106. Wenzel, S. C. *et al.* Structure and biosynthesis of myxochromides S1-3 in Stigmatella aurantiaca: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *Chembiochem* **6**, 375–385 (2005).

107. Bosello, M., Mielcarek, A., Giessen, T. W. & Marahiel, M. A. An Enzymatic Pathway for the Biosynthesis of the Formylhydroxyornithine Required for Rhodochelin Iron Coordination. *Biochemistry* **51**, 3059–3066 (2012).
108. Bode, H., Bethe, B., Höfs, R. & Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem* (2002).
109. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
110. Donadio, S., Monciardini, P. & Sosio, M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* **24**, 1073–1109 (2007).
111. Ansari, M. Z., Yadav, G., Gokhale, R. S. & Mohanty, D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic acids research* **32**, W405–13 (2004).
112. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **417**, 141–147 (2002).
113. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* **100**, 1541–1546 (2003).
114. Song, L. *et al.* Type III polyketide synthase beta-ketoacyl-ACP starter unit and ethylmalonyl-CoA extender unit selectivity discovered by Streptomyces coelicolor genome mining. *J Am Chem Soc* **128**, 14754–14755 (2006).
115. Corre, C., Song, L., O'Rourke, S., Chater, K. F. & Challis, G. L. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. *Proc Natl Acad Sci USA* **105**, 17510–17515 (2008).
116. Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proc Natl Acad Sci USA* **104**, 10376–10381 (2007).
117. McClerren, A. L. *et al.* Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc Natl Acad Sci USA* **103**, 17243–17248 (2006).
118. Fortin, P. D., Walsh, C. T. & Magarvey, N. A. A transglutaminase homologue as a condensation catalyst in antibiotic assembly lines. *Nature* **448**, 824–827 (2007).
119. Watanabe, K. *et al.* Escherichia coli allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product. *J Am Chem Soc* **131**, 9347–9353 (2009).
120. Watanabe, K., Oguri, H. & Oikawa, H. Diversification of echinomycin molecular structure by way of chemoenzymatic synthesis and heterologous expression of the engineered echinomycin biosynthetic pathway. *Current Opinion in Chemical Biology* **13**, 189–196 (2009).
121. Gross, H. *et al.* The genomisotopic approach: a systematic method to isolate products of orphan biosynthetic gene clusters. *Chem Biol* **14**, 53–63 (2007).
122. A, B., R, S. & U, B. The nitrile-degrading enzymes: current status and future prospects. *Applied microbiology …* **60**, 33–44 (2002).
123. van der Geize, R. & Dijkhuizen, L. Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. *Curr Opin Microbiol* **7**, 255–261 (2004).
124. McLeod, M. P. *et al.* The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci USA* **103**, 15582–15587 (2006).
125. Carrano, C. J., Jordan, M., Drechsel, H., Schmid, D. G. & Winkelmann, G. Heterobactins: A new class of siderophores from Rhodococcus erythropolis IGTS8 containing both hydroxamate and catecholate donor groups. *Biometals* **14**, 119–125 (2001).
126. Dhungana, S. *et al.* Purification and characterization of rhodobactin: a mixed ligand siderophore from Rhodococcus rhodochrous strain OFS. *Biometals* **20**, 853–867 (2007).
127. Seto, M. *et al.* A Novel Transformation of Polychlorinated Biphenyls by Rhodococcus sp. Strain RHA1. *Applied and …* **61**, 3353–3358 (1995).
128. Helmetag, V. *Biochemische und strukturelle Untersuchungen der Biosynthese unnatürlicher Aminosäuren als Bausteine nicht-ribosomaler Peptide*. (Philipps-Universität Marburg: Marburg an der Lahn, 2009).
129. Schäfer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**, 69–73 (1994).
130. Pridmore, R. D. New and versatile cloning vectors with kanamycin-resistance marker. *Gene* **56**, 309–312 (1987).
131. Datta, N., Hedges, R., Shaw, E. J., Sykes, R. & Richmond, M. Properties of an R factor from Pseudomonas aeruginosa. *J Bacteriol* **108**, 1244–1249 (1971).
132. Trieu-Cuot, P., Carlier, C., Martin, P. & Courvalin, P. Plasmid transfer by conjugation from Escherichia coli to Gram-positive bacteria. *FEMS Microbiol Rev* **48**, 289–294 (1987).
133. Mazodier, P., Petter, R. & Thompson, C. Intergeneric conjugation between Escherichia coli and Streptomyces species. *J Bacteriol* **171**, 3583–3585 (1989).
134. Schäfer, A., Kalinowski, J., Simon, R., Seep-Feldhaus, A. H. & Pühler, A. High-frequency conjugal plasmid transfer from gram-negative Escherichia coli to various gram-positive coryneform bacteria. *J Bacteriol* **172**, 1663–1666 (1990).

135. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C. I. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* **164**, 918–921 (1985).
136. Jäger, W., Schäfer, A., Pühler, A., Labes, G. & Wohlleben, W. Expression of the Bacillus subtilis sacB gene leads to sucrose sensitivity in the gram-positive bacterium Corynebacterium glutamicum but not in Streptomyces lividans. *J Bacteriol* **174**, 5462–5465 (1992).
137. Steinmetz, M., Le Coq, D., Djemia, H. B. & Gay, P. Genetic analysis of sacB, the structural gene of a secreted enzyme, levansucrase of Bacillus subtilis Marburg. *Mol. Gen. Genet.* **191**, 138–144 (1983).
138. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130 (1986).
139. Simon, R., Priefer, U. & Pühler, A. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. *Nature Biotechnology* (1983).
140. Bauchop, T. & Elsden, S. R. The growth of micro-organisms in relation to their energy supply. *J Gen Microbiol* **23**, 457–469 (1960).
141. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press: 2001).
142. Horton, R. M., Cai, Z. L., Ho, S. N. & Pease, L. R. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *BioTechniques* **8**, 528–535 (1990).
143. van der Geize, R., Hessels, G., van Gerwen, R., van der Meijden, P. & Dijkhuizen, L. Unmarked gene deletion mutagenesis of kstD, encoding 3-ketosteroid Delta1-dehydrogenase, in Rhodococcus erythropolis SQ1 using sacB as counter-selectable marker. *FEMS Microbiol Rev* **205**, 197–202 (2001).
144. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254 (1976).
145. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
146. Koenig, T. *et al.* Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J Proteome Res* **7**, 3708–3717 (2008).
147. Schwyn, B. & Neilands, J. B. Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160**, 47–56 (1987).
148. Bhushan, R. & Brückner, H. Marfey's reagent for chiral amino acid analysis: a review. *Amino Acids* **27**, 231–247 (2004).
149. Lin, Y. & Miller, M. Practical Synthesis of Hydroxamate-Derived Siderophore Components by an Indirect Oxidation Method and Syntheses of a DIG− Siderophore Conjugate and a Biotin …. *J. Org. Chem* (1999).
150. Jones, R. G. Studies on imidazoles. II. The Synthesis of 5-imidazolecarboxylates from glycine and substituted glycine esters. *J Am Chem Soc* **71**, 644–647 (1949).
151. Blanquet, S., Dessen, P. & Kahn, D. Properties and specificity of methionyl-tRNAfMet formyltransferase from Escherichia coli. *Meth Enzymol* **106**, 141–152 (1984).
152. Santi, D. V., Webster, R. W. & Cleland, W. W. Kinetics of aminoacyl-tRNA synthetases catalyzed ATP-PPi exchange. *Meth Enzymol* **29**, 620–627 (1974).
153. Linne, U. & Marahiel, M. A. Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases. *Meth Enzymol* **388**, 293–315 (2004).
154. Reuter, K., Mofid, M. R., Marahiel, M. A. & Ficner, R. Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily. *EMBO J* **18**, 6823–6831 (1999).
155. La Clair, J. J., Foley, T. L., Schegg, T. R., Regan, C. M. & Burkart, M. D. Manipulation of carrier proteins in antibiotic biosynthesis. *Chem Biol* **11**, 195–201 (2004).
156. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
157. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research* (2011).doi:10.1093/nar/gkr466
158. Röttig, M. *et al.* NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic acids research* **39**, W362–7 (2011).
159. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
160. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**, 4673–4680 (1994).
161. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
162. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).

163. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
164. Hu, J. & Miller, M. J. A New Method for the Synthesis of N-ε-Acetyl-N-ε-hydroxy-L-lysine, the Iron-Binding Constituent of Several Important Siderophores. *The Journal of Organic Chemistry* **59**, 4858–4861 (1994).
165. Patzer, S. I. & Braun, V. Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of Streptomyces sp. ATCC 700974. *J Bacteriol* **192**, 426–435 (2010).
166. Plowman, J. E., Loehr, T. M., Goldman, S. J. & Sanders-Loehr, J. Structure and siderophore activity of ferric schizokinen. *J Inorg Biochem* **20**, 183–197 (1984).
167. Persmark, M. & Neilands, J. B. Iron(III) complexes of chrysobactin, the siderophore of Erwinia chrysanthemi. *Biometals* **5**, 29–36 (1992).
168. Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* **26**, 1362–1384 (2009).
169. Lazos, O. *et al.* Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. *Chem Biol* **17**, 160–173 (2010).
170. Umezawa, H. *et al.* Foroxymithine, a new inhibitor of angiotensin-converting enzyme, produced by actinomycetes. *J Antibiot* **38**, 1813–1815 (1985).
171. Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* **5**, 631–645 (1998).
172. Felnagle, E. A. *et al.* MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. *Biochemistry* **49**, 8815–8817 (2010).
173. McMahon, M. D., Rush, J. S. & Thomas, M. G. Analyses of MbtB, MbtE, and MbtF Suggest Revisions to the Mycobactin Biosynthesis Pathway in Mycobacterium tuberculosis. *J Bacteriol* (2012).doi:10.1128/JB.00088-12
174. Miethke, M. *et al.* Inhibition of aryl acid adenylation domains involved in bacterial siderophore synthesis. *FEBS J* **273**, 409–419 (2006).
175. Crosa, J. H. & Walsh, C. T. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **66**, 223–249 (2002).
176. Rowland, B. M., Grossman, T. H., Osburne, M. S. & Taber, H. W. Sequence and genetic organization of a Bacillus subtilis operon encoding 2,3-dihydroxybenzoate biosynthetic enzymes. *Gene* **178**, 119–123 (1996).
177. Raynal, A., Karray, F., Tuphile, K., Darbon-Rongère, E. & Pernodet, J.-L. Excisable cassettes: new tools for functional analysis of Streptomyces genomes. *Applied and …* **72**, 4839–4844 (2006).
178. Meneely, K. M. & Lamb, A. L. Biochemical characterization of a flavin adenine dinucleotide-dependent monooxygenase, ornithine hydroxylase from Pseudomonas aeruginosa, suggests a novel reaction mechanism. *Biochemistry* **46**, 11930–11937 (2007).
179. Chocklett, S. W. & Sobrado, P. Aspergillus fumigatus SidA is a highly specific ornithine hydroxylase with bound flavin cofactor. *Biochemistry* **49**, 6777–6783 (2010).
180. Mayfield, J. A. *et al.* Comprehensive spectroscopic, steady state, and transient kinetic studies of a representative siderophore-associated flavin monooxygenase. *J Biol Chem* **285**, 30375–30388 (2010).
181. Macheroux, P., Plattner, H. J., Romaguera, A. & Diekmann, H. FAD and substrate analogs as probes for lysine N6-hydroxylase from Escherichia coli EN 222. *Eur J Biochem* **213**, 995–1002 (1993).
182. Olucha, J., Meneely, K. M., Chilton, A. S. & Lamb, A. L. Two structures of an N-hydroxylating flavoprotein monooxgenase: the ornithine hydroxylase from Pseudomonas aeruginosa. *J Biol Chem* (2011).doi:10.1074/jbc.M111.265876
183. Seyedsayamdost, M. R., Traxler, M. F., Zheng, S.-L., Kolter, R. & Clardy, J. Structure and Biosynthesis of Amychelin, an Unusual Mixed-Ligand Siderophore from Amycolatopsis sp. AA4. *J Am Chem Soc* **133**, 11434–11437 (2011).
184. Breazeale, S. D., Ribeiro, A. A., McClerren, A. L. & Raetz, C. R. H. A formyltransferase required for polymyxin resistance in Escherichia coli and the modification of lipid A with 4-Amino-4-deoxy-L-arabinose. Identification and function oF UDP-4-deoxy-4-formamido-L-arabinose. *Journal of Biological Chemistry* **280**, 14154–14167 (2005).
185. Schmitt, E., Blanquet, S. & Mechulam, Y. Structure of crystalline Escherichia coli methionyl-tRNA(f)Met formyltransferase: comparison with glycinamide ribonucleotide formyltransferase. *EMBO J* **15**, 4749–4758 (1996).
186. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**, 5857–5864 (1998).
187. Schmitt, E., Panvert, M., Blanquet, S. & Mechulam, Y. Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J* **17**, 6819–6826 (1998).
188. Fiss, E. H., Yu, S. & Jacobs, W. R. Identification of genes involved in the sequestration of iron in mycobacteria: the ferric exochelin biosynthetic and uptake pathways. *Mol Microbiol* **14**, 557–569 (1994).

189. McMorran, B. J., Shanta Kumara, H. M., Sullivan, K. & Lamont, I. L. Involvement of a transformylase enzyme in siderophore synthesis in Pseudomonas aeruginosa. *Microbiology* **147**, 1517–1524 (2001).
190. Chen, P. *et al.* Crystal structure of glycinamide ribonucleotide transformylase from Escherichia coli at 3.0 A resolution. A target enzyme for chemotherapy. *J Mol Biol* **227**, 283–292 (1992).
191. Breazeale, S. D., Ribeiro, A. A. & Raetz, C. R. H. Oxidative decarboxylation of UDP-glucuronic acid in extracts of polymyxin-resistant Escherichia coli. Origin of lipid a species modified with 4-amino-4-deoxy-L-arabinose. *Journal of Biological Chemistry* **277**, 2886–2896 (2002).
192. Buchenau, B. & Thauer, R. K. Tetrahydrofolate-specific enzymes in Methanosarcina barkeri and growth dependence of this methanogenic archaeon on folic acid or p-aminobenzoic acid. *Arch Microbiol* **182**, 313–325 (2004).
193. Hammes, G. G. & Wu, C. W. Kinetics of allosteric enzymes. *Annu. Rev. Biophys. Bioeng.* **3**, 1–33 (1974).
194. Porter, C. M. & Miller, B. G. Cooperativity in monomeric enzymes with single ligand-binding sites. *Bioorg. Chem.* (2011).doi:10.1016/j.bioorg.2011.11.001
195. Iwatsuki, M. *et al.* Lariatins, antimycobacterial peptides produced by Rhodococcus sp. K01-B0171, have a lasso structure. *J Am Chem Soc* **128**, 7486–7491 (2006).
196. Inokoshi, J., Matsuhama, M., Miyake, M., Ikeda, H. & Tomoda, H. Molecular cloning of the gene cluster for lariatin biosynthesis of Rhodococcus jostii K01-B0171. *Applied microbiology …* (2012).doi:10.1007/s00253-012-3973-8
197. Yamanaka, K., Maruyama, C., Takagi, H. & Hamano, Y. Epsilon-poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. *Nat Chem Biol* **4**, 766–772 (2008).
198. Hider, R. C. & Kong, X. Chemistry and biology of siderophores. *Nat Prod Rep* **27**, 637–657 (2010).
199. Braun, V. & Braun, M. Iron transport and signaling in Escherichia coli. *FEBS Lett* **529**, 78–85 (2002).
200. Chakraborty, R., Storey, E. & van der Helm, D. Molecular mechanism of ferricsiderophore passage through the outer membrane receptor proteins of Escherichia coli. *Biometals* **20**, 263–274 (2007).
201. Krewulak, K. D. & Vogel, H. J. Structural biology of bacterial iron uptake. *Biochim. Biophys. Acta* **1778**, 1781–1804 (2008).
202. Franke, K. *Klonierung, Expression und biochemische Charakterisierung von Siderophor-Bindungsproteinen grampositiver Bakterien* Philipps-Universität Marburg: Marburg an der Lahn, 2011.
203. Dimise, E. J., Widboom, P. F. & Bruner, S. D. Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile Thermobifida fusca. *Proc Natl Acad Sci USA* **105**, 15311–15316 (2008).
204. Gehring, A. M., Bradley, K. A. & Walsh, C. T. 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* **36**, 8495–8503 (1997).
205. Klatte, S., Kroppenstedt, R. M. & Rainey, F. A. Rhodococcus opacus sp.nov., An Unusual Nutritionally Versatile Rhodococcus-species. *Systematic and Applied Microbiology* **17**, 355–360 (1994).
206. Gürtler, V., Mayall, B. C. & Seviour, R. Can whole genome analysis refine the taxonomy of the genus Rhodococcus? *FEMS Microbiol Rev* **28**, 377–403 (2004).
207. Letek, M. *et al.* The genome of a pathogenic rhodococcus: cooptive virulence underpinned by key gene acquisitions. *PLoS Genet* **6**, (2010).
208. Ghosh, A. *et al.* Iron transport-mediated drug delivery using mixed-ligand siderophore-beta-lactam conjugates. *Chem Biol* **3**, 1011–1019 (1996).
209. Schobert, R., Stangl, A. & Hannemann, K. Mixed catechol-hydroxamate and catechol-(o-hydroxy) phenacyl siderophores: synthesis and uptake studies with receptor-deficient Escherichia coli mutants. *Tetrahedron* **64**, 1711–1720 (2008).
210. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106**, 3468–3496 (2006).
211. Flatt, P. M. & Mahmud, T. Biosynthesis of aminocyclitol-aminoglycoside antibiotics and related compounds. *Nat Prod Rep* **24**, 358–392 (2007).
212. Challis, G. L. Genome mining for novel natural product discovery. *J Med Chem* **51**, 2618–2628 (2008).
213. Fischbach, M. A., Walsh, C. T. & Clardy, J. The evolution of gene collectives: How natural selection drives chemical innovation. *Proc Natl Acad Sci USA* **105**, 4601–4608 (2008).
214. Butterton, J. R. & Calderwood, S. B. Identification, cloning, and sequencing of a gene required for ferric vibriobactin utilization by Vibrio cholerae. *J Bacteriol* **176**, 5631–5638 (1994).

215. Butterton, J. R., Stoebner, J. A., Payne, S. M. & Calderwood, S. B. Cloning, sequencing, and transcriptional regulation of viuA, the gene encoding the ferric vibriobactin receptor of Vibrio cholerae. *J Bacteriol* **174**, 3729–3738 (1992).
216. Butterton, J. R., Choi, M. H., Watnick, P. I., Carroll, P. A. & Calderwood, S. B. Vibrio cholerae VibF is required for vibriobactin synthesis and is a member of the family of nonribosomal peptide synthetases. *J Bacteriol* **182**, 1731–1738 (2000).
217. Wyckoff, E. E., Valle, A. M., Smith, S. L. & Payne, S. M. A multifunctional ATP-binding cassette transporter system from Vibrio cholerae transports vibriobactin and enterobactin. *J Bacteriol* **181**, 7588–7596 (1999).
218. Wyckoff, E. E., Stoebner, J. A., Reed, K. E. & Payne, S. M. Cloning of a Vibrio cholerae vibriobactin gene cluster: identification of genes required for early steps in siderophore biosynthesis. *J Bacteriol* **179**, 7055–7062 (1997).
219. Wyckoff, E. E., Smith, S. L. & Payne, S. M. VibD and VibH are required for late steps in vibriobactin biosynthesis in Vibrio cholerae. *J Bacteriol* **183**, 1830–1834 (2001).
220. Hsieh, Y. J. & Kolattukudy, P. E. Inhibition of erythromycin synthesis by disruption of malonyl-coenzyme A decarboxylase gene eryM in Saccharopolyspora erythraea. *J Bacteriol* **176**, 714–724 (1994).
221. Oliynyk, M. *et al.* Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nature Biotechnology* **25**, 447–453 (2007).
222. Wesener, S. R., Potharla, V. Y. & Cheng, Y.-Q. Reconstitution of the FK228 biosynthetic pathway reveals cross talk between modular polyketide synthases and fatty acid synthase. *Applied and …* **77**, 1501–1507 (2011).
223. Kidarsa, T. A., Goebel, N. C., Zabriskie, T. M. & Loper, J. E. Phloroglucinol mediates cross-talk between the pyoluteorin and 2,4-diacetylphloroglucinol biosynthetic pathways in Pseudomonas fluorescens Pf-5. *Mol Microbiol* **81**, 395–414 (2011).
224. Bergmann, S. *et al.* Activation of a silent fungal polyketide biosynthesis pathway through regulatory cross talk with a cryptic nonribosomal peptide synthetase gene cluster. *Applied and …* **76**, 8143–8149 (2010).
225. Heemstra, J. R., Walsh, C. T. & Sattely, E. S. Enzymatic tailoring of ornithine in the biosynthesis of the Rhizobium cyclic trihydroxamate siderophore vicibactin. *J Am Chem Soc* **131**, 15317–15329 (2009).
226. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
227. Meneely, K. M., Barr, E. W., Bollinger, J. M. & Lamb, A. L. Kinetic mechanism of ornithine hydroxylase (PvdA) from Pseudomonas aeruginosa: substrate triggering of O2 addition but not flavin reduction. *Biochemistry* **48**, 4371–4376 (2009).
228. Gatzeva-Topalova, P. Z., May, A. P. & Sousa, M. C. Crystal structure and mechanism of the Escherichia coli ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry* **44**, 5328–5338 (2005).
229. Warren, M. S., Marolewski, A. E. & Benkovic, S. J. A rapid screen of active site mutants in glycinamide ribonucleotide transformylase. *Biochemistry* **35**, 8855–8862 (1996).
230. Su, Y. *et al.* A pH-dependent stabilization of an active site loop observed from low and high pH crystal structures of mutant monomeric glycinamide ribonucleotide transformylase at 1.8 to 1.9 A. *J Mol Biol* **281**, 485–499 (1998).
231. Caperelli, C. A. N10-substituted 5,8-dideazafolate inhibitors of glycinamide ribonucleotide transformylase. *J Med Chem* **30**, 1254–1256 (1987).
232. Klein, C. *et al.* Towards structure-based drug design: crystal structure of a multisubstrate adduct complex of glycinamide ribonucleotide transformylase at 1.96 A resolution. *J Mol Biol* **249**, 153–175 (1995).
233. Khalil, S. & Pawelek, P. D. Enzymatic adenylation of 2,3-dihydroxybenzoate is enhanced by a protein-protein interaction between Escherichia coli 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (EntA) and 2,3-dihydroxybenzoate-AMP Ligase (EntE). *Biochemistry* **50**, 533–545 (2011).
234. Boyd, J. & Murphy, J. R. Analysis of the diphtheria tox promoter by site-directed mutagenesis. *J Bacteriol* **170**, 5949–5952 (1988).
235. Boyd, J., Oza, M. N. & Murphy, J. R. Molecular cloning and DNA sequence analysis of a diphtheria tox iron-dependent regulatory element (dtxR) from Corynebacterium diphtheriae. *Proc Natl Acad Sci USA* **87**, 5968–5972 (1990).
236. Fourel, G., Phalipon, A. & Kaczorek, M. Evidence for direct regulation of diphtheria toxin gene transcription by an Fe2+-dependent DNA-binding repressor, DtoxR, in Corynebacterium diphtheriae. *Infect. Immun.* **57**, 3221–3225 (1989).
237. Tao, X. & Murphy, J. R. Binding of the metalloregulatory protein DtxR to the diphtheria tox operator requires a divalent heavy metal ion and protects the palindromic sequence from DNase I digestion. *Journal of Biological Chemistry* **267**, 21761–21764 (1992).
238. Schmitt, M. P., Twiddy, E. M. & Holmes, R. K. Purification and characterization of the diphtheria toxin repressor. *Proc Natl Acad Sci USA* **89**, 7576–7580 (1992).

239. Tao, X. & Murphy, J. R. Cysteine-102 is positioned in the metal binding activation site of the Corynebacterium diphtheriae regulatory element DtxR. *Proc Natl Acad Sci USA* **90**, 8524–8528 (1993).
240. Tao, X., Zeng, H. Y. & Murphy, J. R. Transition metal ion activation of DNA binding by the diphtheria tox repressor requires the formation of stable homodimers. *Proc Natl Acad Sci USA* **92**, 6803–6807 (1995).
241. Schmitt, M. P. & Holmes, R. K. Analysis of diphtheria toxin repressor-operator interactions and characterization of a mutant repressor with decreased binding activity for divalent metals. *Mol Microbiol* **9**, 173–181 (1993).
242. Doukhan, L. *et al.* Genomic organization of the mycobacterial sigma gene cluster. *Gene* **165**, 67–70 (1995).
243. Schmitt, M. P., Predich, M., Doukhan, L., Smith, I. & Holmes, R. K. Characterization of an iron-dependent regulatory protein (IdeR) of Mycobacterium tuberculosis as a functional homolog of the diphtheria toxin repressor (DtxR) from Corynebacterium diphtheriae. *Infect. Immun.* **63**, 4284–4289 (1995).
244. Günter-Seeboth, K. & Schupp, T. Cloning and sequence analysis of the Corynebacterium diphtheriae dtxR homologue from Streptomyces lividans and S. pilosus encoding a putative iron repressor protein. *Gene* **166**, 117–119 (1995).
245. Flores, F. J. & Martín, J.-F. Iron-regulatory proteins DmdR1 and DmdR2 of Streptomyces coelicolor form two different DNA-protein complexes with iron boxes. *Biochem J* **380**, 497–503 (2004).
246. Boland, C. A. & Meijer, W. G. The iron dependent regulatory protein IdeR (DtxR) of Rhodococcus equi. *FEMS Microbiol Rev* **191**, 1–5 (2000).
247. Oram, D. M., Jacobson, A. D. & Holmes, R. K. Transcription of the contiguous sigB, dtxR, and galE genes in Corynebacterium diphtheriae: evidence for multiple transcripts and regulation by environmental factors. *J Bacteriol* **188**, 2959–2973 (2006).
248. Cerdeño-Tárraga, A.-M. *et al.* The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. *Nucleic acids research* **31**, 6516–6523 (2003).
249. Münch, R. *et al.* PRODORIC: prokaryotic database of gene regulation. *Nucleic acids research* **31**, 266–269 (2003).
250. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
251. Bosello, M., Robbel, L., Linne, U., Xie, X. & Marahiel, M. A. Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in Rhodococcus jostii RHA1. *J Am Chem Soc* **133**, 4587–4595 (2011).
252. Phelan, V. V., Liu, W.-T., Pogliano, K. & Dorrestein, P. C. Microbial metabolic exchange-the chemotype-to-phenotype link. *Nat Chem Biol* **8**, 26–35 (2011).
253. Martín, J. F. *et al.* Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in Streptomyces. *Microbial Biotechnology* **4**, 165–174 (2011).
254. Pettit, R. K. Small-molecule elicitation of microbial secondary metabolites. *Microbial Biotechnology* (2010).doi:10.1111/j.1751-7915.2010.00196.x
255. Brakhage, A. A. & Schroeckh, V. Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal Genetics and Biology* **48**, 15–22 (2011).
256. Nakashima, N. & Tamura, T. Isolation and characterization of a rolling-circle-type plasmid from Rhodococcus erythropolis and application of the plasmid to multiple-recombinant-protein expression. *Applied and …* **70**, 5557–5568 (2004).
257. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens. *Proc Natl Acad Sci USA* **108**, 6258–6263 (2011).
258. Biggins, J. B., Liu, X., Feng, Z. & Brady, S. F. Metabolites from the Induced Expression of Cryptic Single Operons Found in the Genome of Burkholderia pseudomallei. *J Am Chem Soc* **133**, 1638–1641 (2011).
259. Biggins, J. B., Gleber, C. D. & Brady, S. F. Acyldepsipeptide HDAC Inhibitor Production Induced in Burkholderia thailandensis. *Org. Lett.* **13**, 1536–1539 (2011).

Supplementary section

Supporting tables

Table S1 Table of chemical shifts (ppm) of rhodochelin in H_2O/D_2O (9:1) at 283 K.

Position		$\delta^{13}C$	$\delta^{1}H$ (multi, *J* Hz)	$\delta^{15}N$	HMBC correlation
2,3-DHB	1	117.13			
	2	146.06			
	3	144.68			
	4	119.82	7.088 (d: 8.3 Hz)		H4 to C2, C6
	5	119.67	6.876 (t: 8.4 Hz)		H5 to C1, C3
	6	120.12	7.347 (d: 8.8 Hz)		H6 to C2, C4, C7
	7	169.18			
	1		9.169 (d: 8.6 Hz)	114.4	H1 to C7 (2,3-DHB)
	2	56.15	4.978 (m)		H2 to C5
L-Thr[1]	3	73.19	5.573 (m)		H3 to C5, C7 (L-fhOrn[3])
	4	15.84	1.396 (d: 6.4 Hz)		H4 to C2
	5	169.71			
	1		8.505 (d: 7.5 Hz)	125.0	H1 to C5 (L-Thr[1])
	2	54.72	4.19 (m)		H2 to C4, C5 (L-Thr[1]), C7
	3	27.81	1.77 (m)		H3 to C5
L-fhOrn[2]	4	21.70	1.77 (m)		H4 to C2
	5	49.87	3.53 (m)		H5 to C3, C6
	6	159.28	7.893 (s)		H6 to C5
	7	177.71			
	1				
	2	52.30	4.131 (m)		H2 to C4
	3	26.51	1.88, 1.69 (m)		H3 to C7
L-fhOrn[3]	4	22.63	1.88, 1.69 (m)		H4 to C2
	5	49.34	3.47 (m)		H5 to C3, C6
	6	159.28	7.826 (s)		H6 to C5
	7	168.38			

Table S2 Bioinformatic overview of the gene clusters involved in rhodochelin biosynthesis.

locus name	refseq accession	gene name	gene lenght (bp)	protein size (aa)	proposed function	sequence similarity, organism	% id./sim
RHA1_ro02318	YP_702282	*rhcA*	648	215	isochorismatase	DhbB, *B. subtilis* 168	52 72
RHA1_ro02319	YP_702283	*rhcB*	7101	2366	putative NRPS	ROP_20370, *R. opacus* B4	92 95
RHA1_ro02320	YP_702284	*rhcC*	1065	354	iron-siderophore ABC transporter substrate-binding protein	SrosN15_01010028435, *S. roseosporus* NRRL 15998	37 57
RHA1_ro02321	YP_702285	*rhcD*	1236	411	enterobactin exporter EntS	SeD_A0689 , *S. enterica ssp. enterica* CT_02021853	41 59
RHA1_ro02322	YP_702286	*rhcE*	243	80	aryl carrier protein domain	Svir_20840, *S. viridis* DSM 43017	51 62
RHA1_ro02323	YP_702287	*rhcF*	1344	447	ABC type 2 transporter family protein	MAV_5003, *M. avium* 104	48 68
RHA1_ro04710	YP_704654		1704	567	ABC transporter transmembrane protein	SCO0491, *S. coelicolor* A3(2)	47 64
RHA1_ro04711	YP_704655		1749	582	ABC transporter transmembrane protein	SCO0493, *S. coelicolor* A3(2)	51 67
RHA1_ro04712	YP_704656	*rft*	936	311	L-hOrn formyltransferase	SCO0499, *S. coelicolor* A3(2)	67 82
RHA1_ro04713	YP_704657		804	267	hydrolase	SCO3233, *S. coelicolor* A3(2)	51 65
RHA1_ro04714	YP_704658		1200	399	diguanylate cyclase / phosphodiesterase	SMa1548, *S. meliloti* 1021	36 50
RHA1_ro04715	YP_704659		16779	5592	putative NRPS	EtcD (Sace_3035), *S. erythraea* NRRL 2338	51 64
RHA1_ro04716	YP_704660	*rmo*	1347	448	L-Orn monooxygenase	MSMEG_0022, *M. smegmatis* MC2 155	56 68
RHA1_ro04717	YP_704661	*mbtH*	252	83	MbtH protein	MbtH, *M. tuberculosis* CDC1551	79 91
RHA1_ro04793	YP_704736	*dhbE*	1653	550	2,3-dihydroxybenzoate-AMP ligase	DhbE, *B. subtilis* 168	66 78
RHA1_ro04794	YP_704737	*dhbA*	771	256	2,3-dihydroxybenzoate-2,3-dehydrogenase	DhbA,*B. subtilis 168*	51 63
RHA1_ro04795	YP_704738	*dhbC*	1143	380	isochorismate synthase	DhbC, *B. subtilis* 168	42 58

Table S3 Comparison of the RhcB synthetase adenylation domains code between *R. jostii* RHA1 and *R. opacus* B4.

strain	% id./sim.	A_1	A_2
R. jostii RHA1		DFWNVGMVHK	DLWGMGAVNK
R. opacus B4	92 / 95	DFWNVGMVHK	DLWGMGAVNK

Table S4 Summarizing overview of rhodochelin biosynthetic gene clusters in other *Rhodococcus* strains as found via bioinformatic analysis.

	Cluster A	Cluster B	Cluster C
R. opacus B4	✔	✔	✔
R. equi 103S	✖	✔[a]	✔[c]
R. erythropolis PR4	✖	✔[b]	✔[d]

✔ all genes were found in the cluster.

✔ majority of the genes were found within the cluster (differences are described below).

✖ gene cluster not found.

a gene insertion between *RHA1_ro04711* and rft homologues; the gene cluster lacks *RHA1_ro04714* homologue.

b gene insertion between *RHA1_ro04713* and *RHA1_ro04715* homologues; the gene cluster lacks *RHA1_ro04714* homologue.

c presence of *rhc*A homologue (isochorismate lyase) at the 3' end of the cluster.

d additional presence of *rhc*E homologue (aryl carrier protein) located between *dhb*A and *dhb*C (different genomic orientation).

Supporting figures

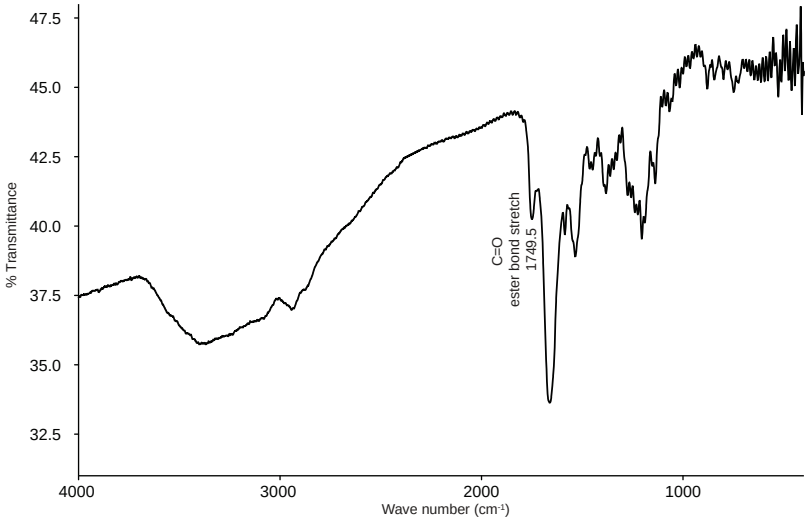


Figure S1 FT-IR spectrum of rhodochelin.

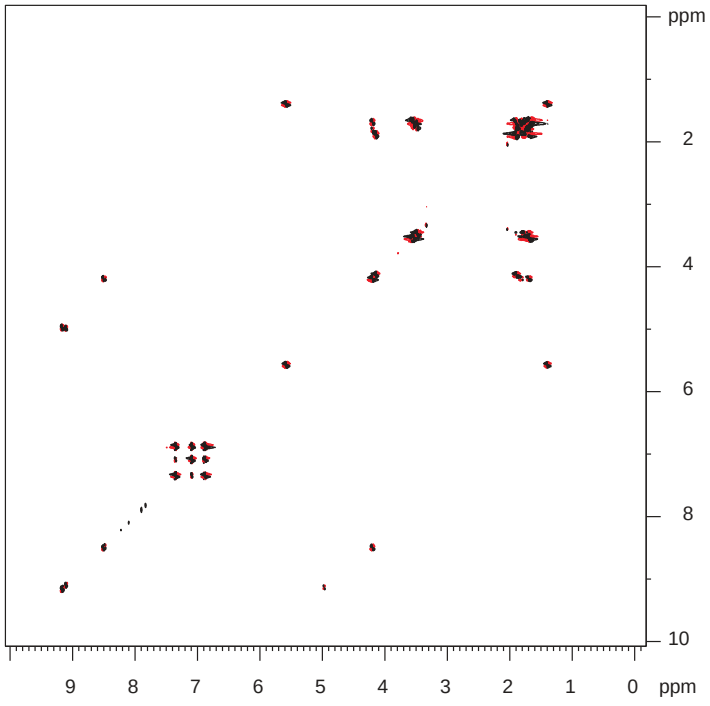


Figure S2 DQF-COSY spectrum of rhodochelin in in H_2O/D_2O (9:1) at 283 K.

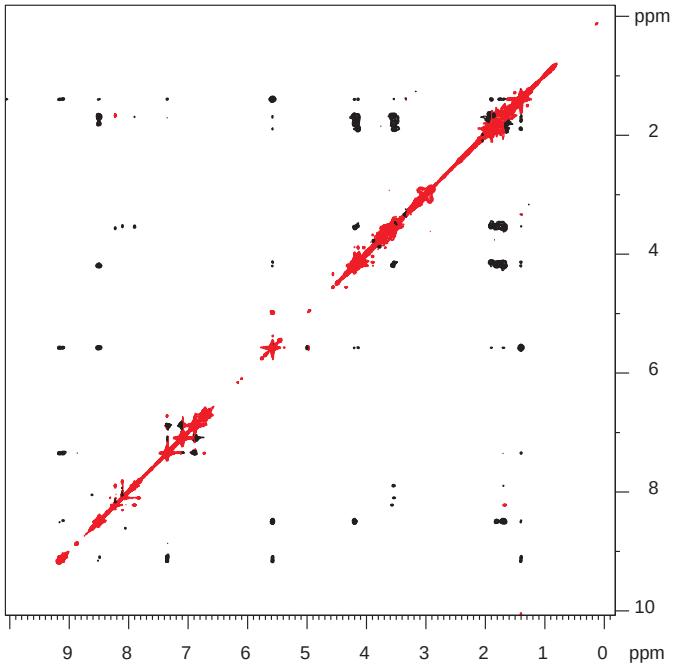


Figure S3 ROESY spectrum of rhodochelin in in H_2O/D_2O (9:1) at 283 K with mixing time 300 ms.

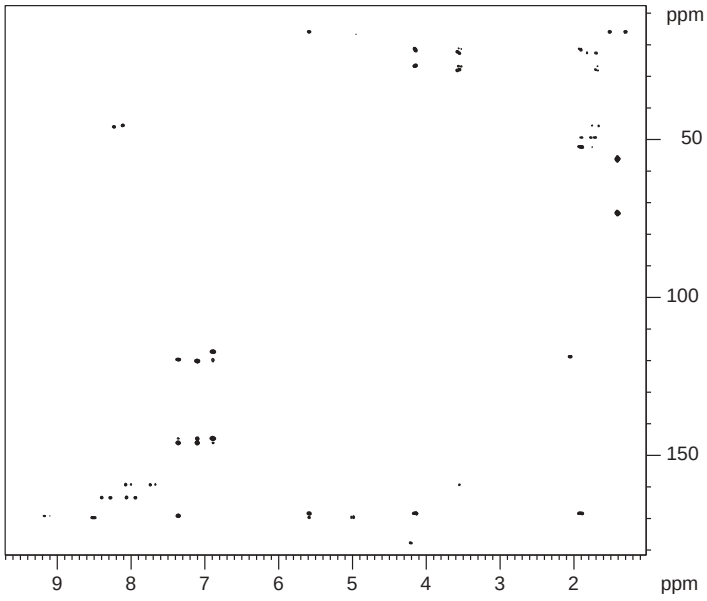


Figure S4 HMBC spectrum of rhodochelin in in H_2O/D_2O (9:1) at 283 K.

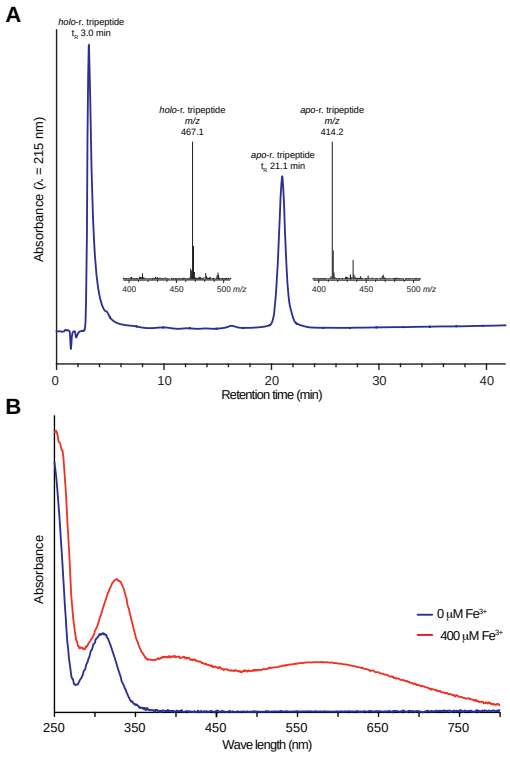


Figure S5 (A) HPLC-MS profile of the Fe-loaded rhodochelin tripeptide. (B) UV-VIS spectrum of 400 μM *apo*-rhodochelin tripeptide (blue line) and *holo*-rhodochelin tripeptide (red line).

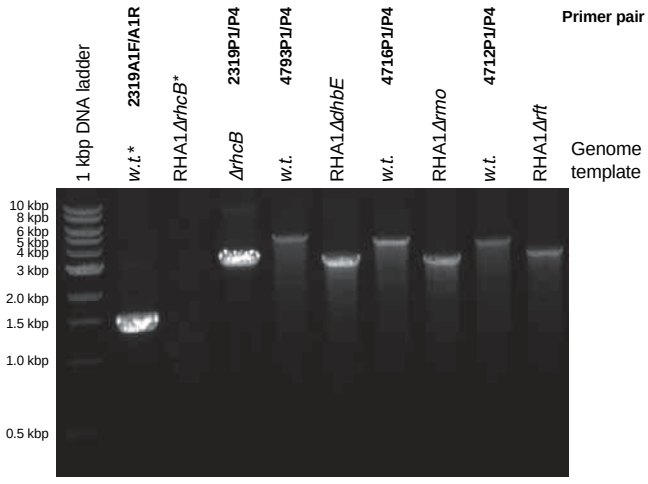
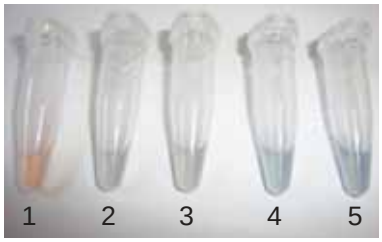


Figure S6 Agarose gel electrophoresis of PCR reactions amplified with external primer pair P1/P4, showing the resulting gene deletion compared to the corresponding *wild-type* strain (on the left). Due to the length of *rhcB* (> 7 kbp), lanes marked with an asterisk were amplified with a different primer pair, annealing inside *rhcB*.



1 RHA1 *wild-type*
2 RHA1Δ*rhc*B
3 RHA1Δ*dhb*E
4 RHA1Δ*rmo*
5 RHA1Δ*rft*

Figure S7 Comparison of CAS activity assays of the culture supernatant of *wild-type R. jostii* RHA1 and the deletion strains.

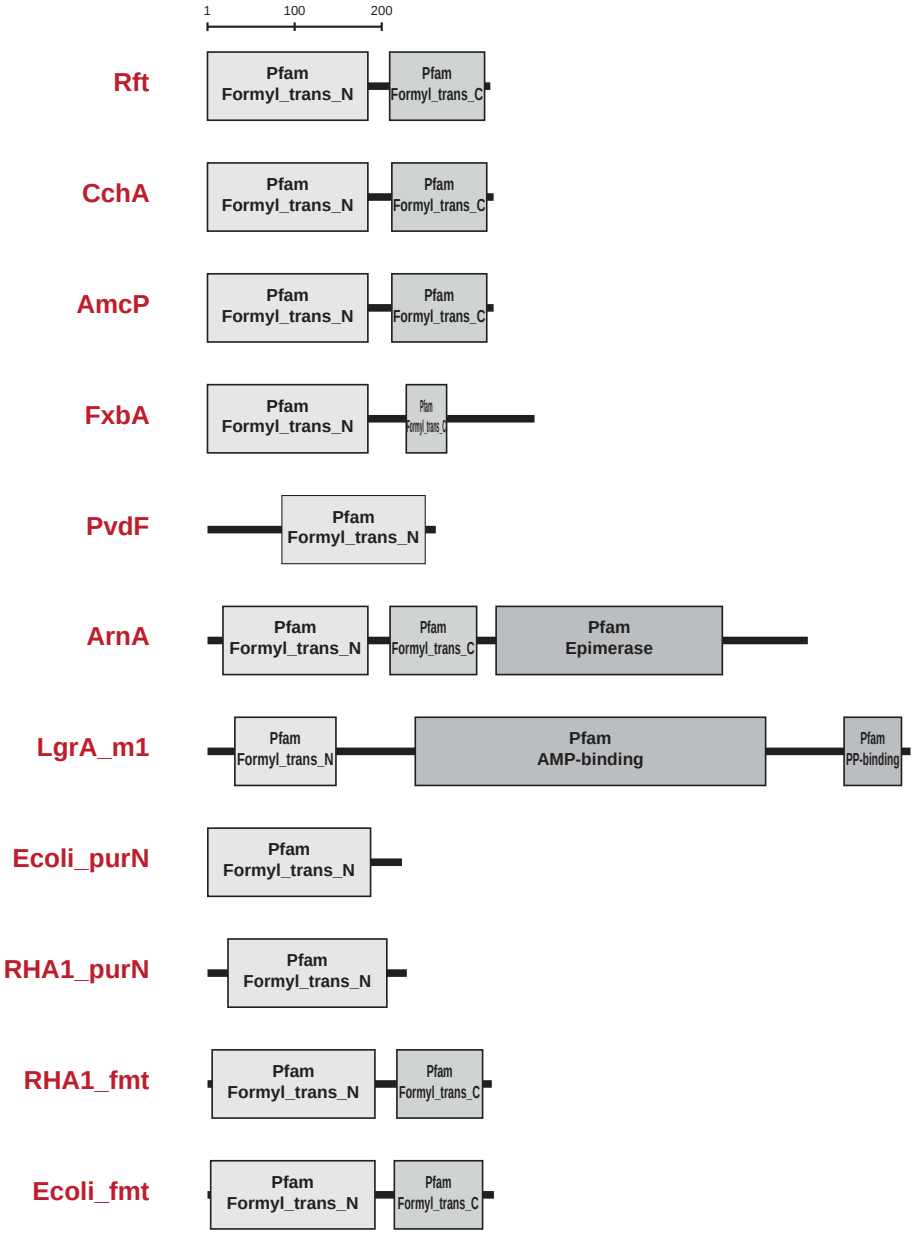


Figure S8 SMART domain analysis and modular organization of the different N^{10}-fH_4F-dependent formyltransferases discussed in the text.

Acknowledgements

First and foremost, I would like to thank Prof. Dr. M. A. Marahiel for letting me join his research group and for providing scientific guidance and constant support during my Ph. D. thesis. I gratefully acknowledge his open-mindedness towards new ideas and enthusiasm about science, which was a constant source of motivation and resulted in great benefit for my work. I am also grateful for the opportunity he gave me to attend the "Annual Conference of the Association for General and Applied Microbiology" in Karlsruhe and the "4th FEMS congress" in Geneve, Switzerland. These meeting have truly been outstanding experiences that expanded my scientific and personal horizons.

I gratefully thank Prof. Dr. M. Bölker for reviewing this thesis and, together with Prof. Dr. E. Bremer and Dr. S.V. Albers, for being part of the thesis committee.

I would also like to thank the Max Planck Institute for Terrestrial Microbiology for supporting my studies with the fellowship and the constructive activities held by the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology. I am also grateful to Dr. C. van der Does for being part of the IMRPS committee. I am thankful to Roswita Roller-Müller and Susanne Rommel for their administrative help.

I would like to thank Dr. Uwe Linne for his excellent support with HPLC and MS and for his expert advices. I am deeply indebt to Dr. Xiulan Xie for conducting extensive and challenging NMR analyses for rhodochelin structure elucidation.

A special thank you goes to everyone in the Marahiel group for their support, the fruitful discussions and the great time in the lab. I would like to thank current and former colleagues of lab 4710 for the nice working atmosphere: Dr. Lars Robbel, Dr. Florian Peuckert, Dr. Alexander Albrecht, Andreas Mielcarek, Mustafa Zeyadi, Kamila Franke and Christin Schönfeld.

Antje Schäfer, Anke Botthof, Christiane Bomm and Gabriele Schimpff-Weiland are gratefully acknowledged for the technical assistance through the years and Tobias Gießen and Andreas Mielcarek for the careful proofreading of this thesis.

Special thanks go to my friends Luca, Anita, Simone and Emiliano, as well as all for anyone who supported myself from Italy.

No words can describe how thankful I am to Rossella: our friendship grew over distance (and through the internet) since the early days of my stay in Germany and helped us support each other in the choices we made day by day.

Finally, I am completely grateful to my parents, who have been a continuous support throughout my whole life. Their constant encouragement has made all of this possible. For this, I dedicate this thesis to them.

Erklärung

Ich versichere, dass ich meine Dissertation „Structural characterization of the siderophore rhodochelin from *Rhodococcus jostii* RHA1 and elucidation of its biosynthetic machinery″ selbständig, ohne erlaubte Hilfe angefertigt und mich dabei keiner als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.
Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, 19 Juni 2012 _____